

# Data Cleaning: Employee Survey Results

The goal of this project is to clean and transform the dataset in a form suitable for further exploratory data analysis

## Objectives:

- To identify errors and inconsistencies and understand how to clean and transform the dataset for further analysis
- Perform an overall and column-wise inspection
- Document all the findings and analyses
- Implement a function that will perform all the cleaning and transformations steps

## Importing Libraries

```
In [1]: from collections import Counter  
  
from IPython.display import HTML, display  
  
import numpy as np  
  
import pandas as pd  
  
import os
```

```
In [2]: pd.set_option("display.max_columns", None)
```

## Getting the Data

```
In [3]: file_dir = r"C:\Python Programs\datasets"  
file_name = "survey_results_public.csv"  
full_path = os.path.join(file_dir, file_name)  
full_path
```

```
Out[3]: 'C:\\\\Python Programs\\\\datasets\\\\survey_results_public.csv'
```

```
In [4]: df = pd.read_csv(full_path)  
display(HTML(f"<h3>Data Shape: {df.shape}</h3>"))  
df.head()
```

Data Shape: (73268, 79)

Out[4]:

	RespondId	MainBranch	Employment	RemoteWork	CodingActivities	EdLevel	LearnCode
0	1	None of these	NaN	NaN	NaN	NaN	NaN
1	2	I am a developer by profession	Employed, full-time	Fully remote	Hobby;Contribute to open-source projects	NaN	NaN
2	3	I am not primarily a developer, but I write co...	Employed, full-time	Hybrid (some remote, some in-person)	Hobby	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Books / Physical media;Friend or family member...
3	4	I am a developer by profession	Employed, full-time	Fully remote	I don't code outside of work	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Books / Physical media;School (i.e., Universit...
4	5	I am a developer by profession	Employed, full-time	Hybrid (some remote, some in-person)	Hobby	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Other online resources (e.g., videos, blogs, f...

## Data Dictionary

- Provides a description of all the columns in the dataset

In [5]:

```
file_dir = r"C:\Python Programs\datasets"
file_name = "survey_results_schema.csv"
full_path = os.path.join(file_dir, file_name)
full_path
```

Out[5]:

```
'C:\\\\Python Programs\\\\datasets\\\\survey_results_schema.csv'
```

In [6]:

```
dictionary = (pd
    .read_csv(full_path)
    .query("selector != 'TB'")
    .loc[:, ["qname", "question"]]
    .set_axis(["column", "description"], axis=1)
    .iloc[1:]
    .set_index("column"))
dictionary
```

Out[6]:

column	description
<b>MainBranch</b>	Which of the following options best describes ...
<b>Employment</b>	Which of the following best describes your cur...
<b>RemoteWork</b>	Which best describes your current work situation?
<b>CodingActivities</b>	Which of the following best describes the code...
<b>EdLevel</b>	Which of the following best describes the high...
...	...
<b>Frequency_2</b>	Interacting with people outside of your immedi...
<b>Frequency_3</b>	Encountering knowledge silos (where one indivi...
<b>TrueFalse_1</b>	Are you involved in supporting new hires durin...
<b>TrueFalse_2</b>	Do you use learning resources provided by your...
<b>TrueFalse_3</b>	Does your employer give you time to learn new ...

70 rows × 1 columns

## Meta-data

In [7]:

`df.info()`

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 73268 entries, 0 to 73267
Data columns (total 79 columns):
 #   Column           Non-Null Count Dtype  
--- 
 0   ResponseId      73268 non-null  int64   
 1   MainBranch       73268 non-null  object  
 2   Employment       71709 non-null  object  
 3   RemoteWork       58958 non-null  object  
 4   CodingActivities 58899 non-null  object  
 5   EdLevel          71571 non-null  object  
 6   LearnCode         71580 non-null  object  
 7   LearnCodeOnline   50685 non-null  object  
 8   LearnCodeCoursesCert 29389 non-null  object  
 9   YearsCode         71331 non-null  object  
 10  YearsCodePro     51833 non-null  object  
 11  DevType          61302 non-null  object  
 12  OrgSize          51039 non-null  object  
 13  PurchaseInfluence 50969 non-null  object  
 14  BuyNewTool       67963 non-null  object  
 15  Country          71771 non-null  object  
 16  Currency          51264 non-null  object  
 17  CompTotal         38422 non-null  float64 
 18  CompFreq          44425 non-null  object  
 19  LanguageHaveWorkedWith 70975 non-null  object  
 20  LanguageWantToWorkWith 67027 non-null  object  
 21  DatabaseHaveWorkedWith 60121 non-null  object  
 22  DatabaseWantToWorkWith 51014 non-null  object  
 23  PlatformHaveWorkedWith 49924 non-null  object  
 24  PlatformWantToWorkWith 40415 non-null  object  
 25  WebframeHaveWorkedWith 53544 non-null  object  
 26  WebframeWantToWorkWith 46122 non-null  object  
 27  MiscTechHaveWorkedWith 44992 non-null  object  
 28  MiscTechWantToWorkWith 36810 non-null  object  
 29  ToolsTechHaveWorkedWith 54171 non-null  object  
 30  ToolsTechWantToWorkWith 46566 non-null  object  
 31  NEWCollabToolsHaveWorkedWith 70347 non-null  object  
 32  NEWCollabToolsWantToWorkWith 64108 non-null  object  
 33  OpSysProfessional use    65503 non-null  object  
 34  OpSysPersonal use      70963 non-null  object  
 35  VersionControlSystem   71379 non-null  object  
 36  VCInteraction          68156 non-null  object  
 37  VCHostingPersonal use   0 non-null   float64 
 38  VCHostingProfessional use 0 non-null   float64 
 39  OfficeStackAsyncHaveWorkedWith 46223 non-null  object  
 40  OfficeStackAsyncWantToWorkWith 32072 non-null  object  
 41  OfficeStackSyncHaveWorkedWith 62128 non-null  object  
 42  OfficeStackSyncWantToWorkWith 47688 non-null  object  
 43  Blockchain            71071 non-null  object  
 44  NEWSOSites             71365 non-null  object  
 45  SOVisitFreq            70961 non-null  object  
 46  SOAccount              71572 non-null  object  
 47  SOPartFreq             58229 non-null  object  
 48  SOComm                 71408 non-null  object  
 49  Age                    70946 non-null  object  
 50  Gender                  70853 non-null  object  
 51  Trans                  70315 non-null  object  
 52  Sexuality               66565 non-null  object  
 53  Ethnicity               69474 non-null  object  
 54  Accessibility           67244 non-null  object

```

```
data_cleaning_notebook
55 MentalHealth          66447 non-null  object
56 TBranch                52670 non-null  object
57 ICorPM                 36283 non-null  object
58 WorkExp                36769 non-null  float64
59 Knowledge_1             35804 non-null  object
60 Knowledge_2             34973 non-null  object
61 Knowledge_3             35133 non-null  object
62 Knowledge_4             35097 non-null  object
63 Knowledge_5             35014 non-null  object
64 Knowledge_6             34991 non-null  object
65 Knowledge_7             34977 non-null  object
66 Frequency_1             35371 non-null  object
67 Frequency_2             35344 non-null  object
68 Frequency_3             34515 non-null  object
69 TimeSearching            36198 non-null  object
70 TimeAnswering            36022 non-null  object
71 Onboarding               35679 non-null  object
72 ProfessionalTech         34906 non-null  object
73 TrueFalse_1              35819 non-null  object
74 TrueFalse_2              35715 non-null  object
75 TrueFalse_3              35749 non-null  object
76 SurveyLength              70444 non-null  object
77 SurveyEase                70508 non-null  object
78 ConvertedCompYearly       38071 non-null  float64
dtypes: float64(5), int64(1), object(73)
memory usage: 44.2+ MB
```

In [8]:

```
(df
 .dtypes
 .value_counts())
```

Out[8]:

object	73
float64	5
int64	1
dtype:	int64

In [9]:

```
df.isna().sum()
```

Out[9]:

ResponseId	0
MainBranch	0
Employment	1559
RemoteWork	14310
CodingActivities	14369
...	
TrueFalse_2	37553
TrueFalse_3	37519
SurveyLength	2824
SurveyEase	2760
ConvertedCompYearly	35197
Length:	79, dtype: int64

In [10]:

```
df.memory_usage(deep=True).sum()
```

Out[10]:

375549705

In [11]:

```
375549705 / 1024 / 1024
```

Out[11]:

358.15210819244385

## Observations:

- The dataset has 73,268 observations
- There're totally 79 variables:
  - `int` : 1
  - `float` : 5
  - `object` : 73
- The dataset utilizes 375,549,705 bytes of memory
  - Roughly 358 mega-bytes of memory
- There're several columns with missing values
  - `VCHostingPersonal use` and `VCHostingProfessional use` have 0 non-null values
  - Values are missing most likely because the users chose to skip/ignore the questions
- Many columns contain inconsistent characters and potentially incorrect data types

## Steps:

- The column names need to be renamed for better readability
- Irrelevant columns to be dropped:
  - `ResponseId`
  - `VCHostingPersonal use`
  - `VCHostingProfessional use`
- Check for duplicates
- Inspect each column individually for errors and inconsistencies
- The missing values can be replaced with `unanswered`

## Checking Duplicates

```
In [12]: df.duplicated().sum()
```

```
Out[12]: 0
```

```
In [13]: df.ResponseId.duplicated().sum()
```

```
Out[13]: 0
```

```
In [14]: df.duplicated(subset=["ResponseId"]).sum()
```

```
Out[14]: 0
```

- There seem to be no duplicates in the dataset

## Column-wise Inspection

## Convenience Functions

In [15]:

```
def col_value_counts(name):
    """
    This function accepts the name of a categorical column
    and returns the distribution of its categories with the
    counts and normalized percentages

    Parameters:
    -----
    name: str
        The name of the categorical column
    """

    return (df
            .loc[:, name]
            .value_counts()
            .pipe(lambda ser: pd.concat([ser,
                                         df
                                         .loc[:, name]
                                         .value_counts(normalize=True)],
                                         axis=1))
            .set_axis(["count", "pct"], axis=1)
            .rename_axis(index=name))
```

In [16]:

```
def col_memory_usage(name, category=False):
    """
    This function accepts the name of a categorical column
    and returns the amount of memory utilized in bytes

    Parameters:
    -----
    name: str
        The name of the categorical column

    category: bool
        Whether to convert the variable to type 'category' or not
    """

    if category:
        return (df
                .loc[:, name]
                .astype("category")
                .memory_usage(deep=True))
    else:
        return (df
                .loc[:, name]
                .memory_usage(deep=True))
```

In [17]:

```
def col_strip_nested(name):
    """
    This function accepts the name of a nested categorical column
    and returns the unique values
    """

    result_set = dict()
    for i, entry in df[name].items():
```

```

if isinstance(entry, str):
    for value in entry.split(","):
        result_set[value] = result_set.get(value, 0) + 1
display(HTML("<h4>Unique Values:</h4>"))
return sorted(result_set.items(),
              key=lambda val: val[1],
              reverse=True)

```

In [18]:

```

def col_description(name):
    """
    This function accepts the name of a column
    and returns its description as per the data dictionary
    """

    display(HTML("<h4>Description:</h4>"))
    print(dictionary
          .loc[name]
          .values[0])

```

## MainBranch

In [19]:

```
col_description("MainBranch")
```

### Description:

Which of the following options best describes you today? Here, by "developer" we mean "someone who writes code." <b>\*</b>

In [20]:

```
df.MainBranch
```

Out[20]:

0	None of these
1	I am a developer by profession
2	I am not primarily a developer, but I write co...
3	I am a developer by profession
4	I am a developer by profession
	...
73263	I am a developer by profession
73264	I am a developer by profession
73265	I am not primarily a developer, but I write co...
73266	I am a developer by profession
73267	I used to be a developer by profession, but no...

Name: MainBranch, Length: 73268, dtype: object

In [21]:

```
df.MainBranch.unique()
```

Out[21]:

```
array(['None of these', 'I am a developer by profession',
       'I am not primarily a developer, but I write code sometimes as part of my wor
k',
       'I code primarily as a hobby', 'I am learning to code',
       'I used to be a developer by profession, but no longer am'],  

       dtype=object)
```

In [22]:

```
col_value_counts("MainBranch")
```

Out[22]:

MainBranch	count	pct
I am a developer by profession	53507	0.730292
I am learning to code	6309	0.086109
I am not primarily a developer, but I write code sometimes as part of my work	5794	0.079080
I code primarily as a hobby	4865	0.066400
None of these	1497	0.020432
I used to be a developer by profession, but no longer am	1296	0.017688

In [23]:

```
(df
    .MainBranch
    .rename("coding_proficiency")
    .replace(["I am a developer by profession",
              "I am learning to code",
              "I am not primarily a developer, but I write code sometimes as part of my work",
              "I code primarily as a hobby",
              "None of these",
              "I used to be a developer by profession, but no longer am"],
              ["developer",
               "learning",
               "work_partly",
               "hobby",
               "other",
               "former_developer"])
    .astype("category"))
```

Out[23]:

```
0          other
1        developer
2      work_partly
3        developer
4        developer
...
73263      developer
73264      developer
73265      work_partly
73266      developer
73267  former_developer
Name: coding_proficiency, Length: 73268, dtype: category
Categories (6, object): ['developer', 'former_developer', 'hobby', 'learning', 'other', 'work_partly']
```

In [24]:

```
col_memory_usage("MainBranch")
```

Out[24]:

6583633

In [25]:

```
col_memory_usage("MainBranch", category=True)
```

Out[25]:

74134

In [26]:

6583633 / 74134

Out[26]:

88.80720047481587

## Observations

- There're 6 unique categories
- Values are valid

## Steps

- The column can be renamed
- The categories can be shortened
- Since cardinality is less, the type can be converted to `category`
  - Will utilize about 89 times less memory

## Employment

```
In [27]: df.Employment
```

```
Out[27]: 0                      NaN
1          Employed, full-time
2          Employed, full-time
3          Employed, full-time
4          Employed, full-time
...
73263      Employed, full-time
73264      Employed, full-time
73265      Employed, full-time
73266      Employed, full-time
73267  Independent contractor, freelancer, or self-em...
Name: Employment, Length: 73268, dtype: object
```

```
In [28]: # df.Employment.unique()
```

```
In [29]: col_value_counts("Employment").head(20)
```

Out[29]:

		count	pct
Employment			
	<b>Employed, full-time</b>	42962	0.599116
	<b>Student, full-time</b>	6756	0.094214
	<b>Independent contractor, freelancer, or self-employed</b>	4978	0.069419
	<b>Employed, full-time;Independent contractor, freelancer, or self-employed</b>	3486	0.048613
	<b>Not employed, but looking for work</b>	1831	0.025534
	<b>Student, full-time;Employed, part-time</b>	1168	0.016288
	<b>Employed, part-time</b>	1132	0.015786
	<b>Student, part-time</b>	1045	0.014573
	<b>Employed, full-time;Student, full-time</b>	972	0.013555
	<b>Employed, full-time;Student, part-time</b>	946	0.013192
	<b>Student, full-time;Not employed, but looking for work</b>	729	0.010166
	<b>Student, part-time;Employed, part-time</b>	721	0.010055
	<b>Not employed, and not looking for work</b>	717	0.009999
	<b>I prefer not to say</b>	611	0.008521
	<b>Student, full-time;Independent contractor, freelancer, or self-employed</b>	441	0.006150
	<b>Independent contractor, freelancer, or self-employed;Employed, part-time</b>	338	0.004713
	<b>Retired</b>	305	0.004253
	<b>Student, full-time;Not employed, and not looking for work</b>	274	0.003821
	<b>Not employed, but looking for work;Independent contractor, freelancer, or self-employed</b>	257	0.003584
	<b>Employed, full-time;Employed, part-time</b>	228	0.003180

In [30]: `col_strip_nested("Employment")`

## Unique Values:

Out[30]: `[('Employed, full-time', 49199),  
 ('Student, full-time', 10932),  
 ('Independent contractor, freelancer, or self-employed', 10721),  
 ('Employed, part-time', 4154),  
 ('Student, part-time', 3722),  
 ('Not employed, but looking for work', 3381),  
 ('Not employed, and not looking for work', 1244),  
 ('I prefer not to say', 611),  
 ('Retired', 396)]`In [31]: `# (df  
# .Employment  
# .dropna()  
# .loc[Lambda ser: ser.str.contains("employed")]  
# .unique())`

```
In [32]: df
    .Employment
    .str.contains("Retired")
    .sum()
```

Out[32]: 396

```
In [33]: df
    .Employment
    .fillna("unanswered")
    .str.lower()
    .replace({"i prefer not to say": "unanswered"})
    .str.replace("independent contractor, freelancer, or self-employed", "freelancer")
    .str.replace("not employed, but looking for work", "unemployed")
    .str.replace("not employed, and not looking for work", "unemployed")
    .str.replace(", ", "_")
    .str.replace("-", "_"))
```

```
Out[33]: 0           unanswered
1           employed_full_time
2           employed_full_time
3           employed_full_time
4           employed_full_time
...
73263     employed_full_time
73264     employed_full_time
73265     employed_full_time
73266     employed_full_time
73267         freelancer
Name: Employment, Length: 73268, dtype: object
```

## Observations

- There're 103 unique categories
- About 90 categories occur in less than 1% of the total observations
  - These could be considered as rare
- Values are nested, separated by a semicolon

## Steps

- The categories can be shortened
- The value I prefer not to say can be renamed unanswered
- The missing values could be imputed with unanswered because it makes sense to think that this field has values missing because the user chose to skip this question
- This variable can be handled in various ways:
  - The rare categories (<1%) can be grouped together
  - The categories can be shortened based on presence of keywords:
    - Employed
    - Not employed
    - Student
    - Retired
  - Multiple binary columns can be created:

- is\_employed
- is\_student
- is\_freelancer
- These steps should be kept in mind for performing feature engineering during exploratory analysis

## RemoteWork

In [34]: `df.RemoteWork`

```
Out[34]: 0           NaN
          1           Fully remote
          2   Hybrid (some remote, some in-person)
          3           Fully remote
          4   Hybrid (some remote, some in-person)
          ...
          73263      Fully remote
          73264      Full in-person
          73265  Hybrid (some remote, some in-person)
          73266  Hybrid (some remote, some in-person)
          73267      Fully remote
Name: RemoteWork, Length: 73268, dtype: object
```

In [35]: `df.RemoteWork.unique()`

```
Out[35]: array([nan, 'Fully remote', 'Hybrid (some remote, some in-person)',
       'Full in-person'], dtype=object)
```

In [36]: `col_value_counts("RemoteWork")`

	count	pct
<b>RemoteWork</b>		
<b>Fully remote</b>	25341	0.429814
<b>Hybrid (some remote, some in-person)</b>	25021	0.424387
<b>Full in-person</b>	8596	0.145799

In [37]: `df.RemoteWork.isna().sum()`

```
Out[37]: 14310
```

In [38]: `col_memory_usage("RemoteWork")`

```
Out[38]: 5143846
```

In [39]: `col_memory_usage("RemoteWork", category=True)`

```
Out[39]: 73737
```

In [40]: `(df
 .RemoteWork
 .rename("work_type"))`

```
.replace(["Fully remote",
         "Hybrid (some remote, some in-person)",
         "Full in-person"],
        ["remote",
         "hybrid",
         "in-person"])
.astype("category"))

Out[40]: 0      NaN
1      remote
2     hybrid
3      remote
4     hybrid
...
73263    remote
73264   in-person
73265    hybrid
73266    hybrid
73267    remote
Name: work_type, Length: 73268, dtype: category
Categories (3, object): ['hybrid', 'in-person', 'remote']
```

## Observations

- There're 3 unique categories
- Cardinality is less

## Steps

- The column can be renamed to `work_type`
- The data type can be converted to `category` to save memory

## Coding Activities

```
In [41]: df.CodingActivities

Out[41]: 0                      NaN
1      Hobby;Contribute to open-source projects
2                               Hobby
3      I don't code outside of work
4                               Hobby
...
73263      Freelance/contract work
73264                               Hobby
73265      Hobby;School or academic work
73266                               Hobby
73267 Hobby;Contribute to open-source projects;Boots...
Name: CodingActivities, Length: 73268, dtype: object
```

```
In [42]: df.CodingActivities.unique()
```

```
Out[42]: array(['Hobby;Contribute to open-source projects', 'Hobby',  
   'I don\'t code outside of work',  
   'Hobby;Contribute to open-source projects;Bootstrapping a business',  
   'Hobby;Contribute to open-source projects;Freelance/contract work',  
   'Hobby;Freelance/contract work', 'Hobby;Bootstrapping a business',  
   'Other (please specify):', 'Contribute to open-source projects',  
   'Hobby;Other (please specify):',  
   'Hobby;Contribute to open-source projects;Bootstrapping a business;Freelance/c  
ontract work',  
   'Bootstrapping a business', 'Freelance/contract work',  
   'Hobby;Bootstrapping a business;Freelance/contract work',  
   'Bootstrapping a business;Freelance/contract work',  
   'Hobby;Contribute to open-source projects;Other (please specify):',  
   'Contribute to open-source projects;Freelance/contract work',  
   'Hobby;Freelance/contract work;Other (please specify):',  
   'Contribute to open-source projects;Bootstrapping a business',  
   'Contribute to open-source projects;Bootstrapping a business;Freelance/contrac  
t work',  
   'Hobby;Bootstrapping a business;Freelance/contract work;School or academic wor  
k',  
   'Hobby;School or academic work',  
   'Hobby;Contribute to open-source projects;Bootstrapping a business;School or a  
cademic work',  
   'Hobby;Contribute to open-source projects;Freelance/contract work;School or ac  
ademic work',  
   'Hobby;Bootstrapping a business;School or academic work',  
   'Contribute to open-source projects;School or academic work',  
   'Hobby;Contribute to open-source projects;School or academic work',  
   'Hobby;Freelance/contract work;School or academic work',  
   'Contribute to open-source projects;Other (please specify):',  
   'Hobby;Contribute to open-source projects;Bootstrapping a business;Freelance/c  
ontract work;School or academic work',  
   'School or academic work',  
   'Contribute to open-source projects;Freelance/contract work;School or academic  
work',  
   'Bootstrapping a business;Freelance/contract work;School or academic work',  
   'Hobby;Bootstrapping a business;Other (please specify):',  
   'Bootstrapping a business;Other (please specify):',  
   'Bootstrapping a business;School or academic work',  
   'Hobby;Contribute to open-source projects;Bootstrapping a business;Other (plea  
se specify):',  
   'Freelance/contract work;School or academic work',  
   'Hobby;Other (please specify):;School or academic work',  
   'Other (please specify):;School or academic work',  
   'Freelance/contract work;Other (please specify):;School or academic work',  
   'Hobby;Contribute to open-source projects;Freelance/contract work;Other (plea  
se specify):;School or academic work',  
   'Contribute to open-source projects;Bootstrapping a business;School or academi  
c work',  
   'Hobby;Contribute to open-source projects;Freelance/contract work;Other (plea  
se specify):',  
   'Contribute to open-source projects;Freelance/contract work;Other (please spec  
ify):',  
   'Freelance/contract work;Other (please specify):',  
   'Hobby;Contribute to open-source projects;Other (please specify):;School or ac  
ademic work',  
   'Contribute to open-source projects;Bootstrapping a business;Other (please spe  
cify):',  
   'Hobby;Bootstrapping a business;Freelance/contract work;Other (please specif  
y):'],
```

```
'Hobby;Bootstrapping a business;Other (please specify);School or academic work',
'Hobby;Bootstrapping a business;Freelance/contract work;Other (please specify);School or academic work',
'Hobby;Freelance/contract work;Other (please specify);School or academic work',
'Hobby;Contribute to open-source projects;Bootstrapping a business;Freelance/contract work;Other (please specify):',
'Bootstrapping a business;Freelance/contract work;Other (please specify):',
'Contribute to open-source projects;Bootstrapping a business;Freelance/contract work;School or academic work',
'Contribute to open-source projects;Other (please specify);School or academic work',
'Hobby;Contribute to open-source projects;Bootstrapping a business;Freelance/contract work;Other (please specify);School or academic work',
'Contribute to open-source projects;Bootstrapping a business;Freelance/contract work;Other (please specify):',
'Contribute to open-source projects;Bootstrapping a business;Other (please specify);School or academic work',
'Contribute to open-source projects;Bootstrapping a business;Freelance/contract work;Other (please specify);School or academic work',
'Bootstrapping a business;Other (please specify);School or academic work',
'Hobby;Contribute to open-source projects;Bootstrapping a business;Other (please specify);School or academic work'],
dtype=object)
```

In [43]: `col_value_counts("CodingActivities").head(20)`

Out[43]:

		count	pct
CodingActivities			
	<b>Hobby</b>	18118	0.307611
	<b>I don't code outside of work</b>	7311	0.124128
	<b>Hobby;Contribute to open-source projects</b>	6549	0.111190
	<b>Hobby;Freelance/contract work</b>	3554	0.060341
	<b>Hobby;School or academic work</b>	3016	0.051206
	<b>Freelance/contract work</b>	2189	0.037165
	<b>Hobby;Bootstrapping a business</b>	2136	0.036265
	<b>Hobby;Contribute to open-source projects;Freelance/contract work</b>	2028	0.034432
	<b>Hobby;Contribute to open-source projects;School or academic work</b>	1254	0.021291
	<b>School or academic work</b>	1208	0.020510
	<b>Bootstrapping a business</b>	1163	0.019746
	<b>Hobby;Contribute to open-source projects;Bootstrapping a business</b>	1133	0.019236
	<b>Hobby;Bootstrapping a business;Freelance/contract work</b>	929	0.015773
	<b>Contribute to open-source projects</b>	917	0.015569
	<b>Hobby;Contribute to open-source projects;Bootstrapping a business;Freelance/contract work</b>	824	0.013990
	<b>Hobby;Freelance/contract work;School or academic work</b>	767	0.013022
	<b>Other (please specify):</b>	735	0.012479
	<b>Hobby;Other (please specify):</b>	674	0.011443
	<b>Contribute to open-source projects;Freelance/contract work</b>	528	0.008964
	<b>Hobby;Contribute to open-source projects;Freelance/contract work;School or academic work</b>	526	0.008931

In [44]: `col_strip_nested("CodingActivities")`

### Unique Values:

Out[44]: `[('Hobby', 42922), ('Contribute to open-source projects', 15378), ('Freelance/contract work', 13305), ('School or academic work', 8561), ('Bootstrapping a business', 8401), ('I don't code outside of work', 7311), ('Other (please specify):', 2179)]`

In [45]:

```
(df
 .CodingActivities
 .dropna()
 .loc[lambda ser: ser.str.contains("I don't code outside of work")]
 .unique())
```

In [45]: array(['I don't code outside of work'], dtype=object)

In [46]:

```
(df
 .CodingActivities
 .rename("coding_activity")
 .fillna("unanswered")
 .str.lower()
 .str.replace("contribute to open-source projects", "open_source_contribution")
 .str.replace("freelance/contract work", "freelancing")
 .str.replace("school or academic work", "academics")
 .str.replace("bootstrapping a business", "startup")
 .str.replace("i don't code outside of work", "only_work")
 .str.replace("other \pleas e specify\):", "other", regex=True)
 .astype("category"))
```

Out[46]:

0	unanswered
1	hobby;open_source_contribution
2	hobby
3	only_work
4	hobby
...	
73263	freelancing
73264	hobby
73265	hobby;academics
73266	hobby
73267	hobby;open_source_contribution;startup

Name: coding\_activity, Length: 73268, dtype: category  
Categories (64, object): ['academics', 'freelancing', 'freelancing;academics', 'freelancing;other', ..., 'startup;freelancing;other', 'startup;other', 'startup;other;academics', 'unanswered']

## Observations

- There're 63 unique categories
- About 45 categories occur in less than 1% of the total observations
  - These could be considered as rare
- Values are nested, separated by a semicolon
- Bootstrapping a business means beginning a new business with one's own savings

## Steps

- The categories can be shortened
- This variable can be handled in various ways:
  - The rare categories (<1%) can be grouped together
  - The categories can be shortened based on presence of keywords
  - Multiple binary columns can be created
- These steps should be kept in mind for performing feature engineering during exploratory analysis

## EdLevel

In [47]: df.EdLevel

```
Out[47]: 0           NaN
          1           NaN
          2    Master's degree (M.A., M.S., M.Eng., MBA, etc.)
          3    Bachelor's degree (B.A., B.S., B.Eng., etc.)
          4    Bachelor's degree (B.A., B.S., B.Eng., etc.)
          ...
          73263   Bachelor's degree (B.A., B.S., B.Eng., etc.)
          73264   Master's degree (M.A., M.S., M.Eng., MBA, etc.)
          73265   Bachelor's degree (B.A., B.S., B.Eng., etc.)
          73266   Bachelor's degree (B.A., B.S., B.Eng., etc.)
          73267   Bachelor's degree (B.A., B.S., B.Eng., etc.)
Name: EdLevel, Length: 73268, dtype: object
```

In [48]: `df.EdLevel.unique()`

```
Out[48]: array([nan, 'Master's degree (M.A., M.S., M.Eng., MBA, etc.)',
       'Bachelor's degree (B.A., B.S., B.Eng., etc.)',
       'Secondary school (e.g. American high school, German Realschule or Gymnasium,
etc.)',
       'Some college/university study without earning a degree',
       'Something else', 'Primary/elementary school',
       'Other doctoral degree (Ph.D., Ed.D., etc.)',
       'Associate degree (A.A., A.S., etc.)',
       'Professional degree (JD, MD, etc.)'], dtype=object)
```

In [49]: `col_value_counts("EdLevel")`

EdLevel	count	pct
<b>Bachelor's degree (B.A., B.S., B.Eng., etc.)</b>	30276	0.423020
<b>Master's degree (M.A., M.S., M.Eng., MBA, etc.)</b>	15486	0.216373
<b>Some college/university study without earning a degree</b>	9326	0.130304
<b>Secondary school (e.g. American high school, German Realschule or Gymnasium, etc.)</b>	7904	0.110436
<b>Associate degree (A.A., A.S., etc.)</b>	2236	0.031242
<b>Other doctoral degree (Ph.D., Ed.D., etc.)</b>	2169	0.030306
<b>Primary/elementary school</b>	1806	0.025234
<b>Something else</b>	1247	0.017423
<b>Professional degree (JD, MD, etc.)</b>	1121	0.015663

In [50]: `(df
 .EdLevel
 .replace(["Bachelor's degree (B.A., B.S., B.Eng., etc.)",
 "Master's degree (M.A., M.S., M.Eng., MBA, etc.)",
 "Some college/university study without earning a degree",
 "Secondary school (e.g. American high school, German Realschule or Gymnasium",
 "Associate degree (A.A., A.S., etc.)",
 "Other doctoral degree (Ph.D., Ed.D., etc.)",
 "Primary/elementary school",
 "Something else",
 "Professional degree (JD, MD, etc.) "],
 ["bachelors",
 "masters",
 "postgraduates",
 "highschool",
 "associate",
 "otherdoctoral",
 "primary",
 "somethingelse",
 "professional"])`

```

    "college",
    "secondary_school",
    "associate",
    "doctorate",
    "primary_school",
    "other",
    "professional"])
.astype("category"))

```

Out[50]:

	EdLevel
0	NaN
1	NaN
2	masters
3	bachelors
4	bachelors
...	
73263	bachelors
73264	masters
73265	bachelors
73266	bachelors
73267	bachelors

Name: EdLevel, Length: 73268, dtype: category  
Categories (9, object): ['Professional degree (JD, MD, etc.)', 'associate', 'bachelors', 'college', ..., 'masters', 'other', 'primary\_school', 'secondary\_school']

## Observations

- There're 9 unique categories
- the categories are unnecessarily lengthy
- Less cardinality

## Steps

- The categories can be shortened
- The type can be made `category` for less memory usage

## LearnCode

In [51]: `df.LearnCode`

Out[51]:

	LearnCode
0	NaN
1	NaN
2	Books / Physical media;Friend or family member...
3	Books / Physical media;School (i.e., Universit...
4	Other online resources (e.g., videos, blogs, f...
...	
73263	Books / Physical media;Other online resources ...
73264	Other online resources (e.g., videos, blogs, f...
73265	Books / Physical media;Other online resources ...
73266	Books / Physical media;On the job training
73267	Books / Physical media;Friend or family member...

Name: LearnCode, Length: 73268, dtype: object

In [52]: `df.LearnCode.unique()`

```
Out[52]: array([nan,
   'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc)',
   'Books / Physical media;School (i.e., University, College, etc)',
   'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training',
   'Other online resources (e.g., videos, blogs, forum)',
   'Online Courses or Certification',
   'On the job training;Coding Bootcamp',
   'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc)',
   'School (i.e., University, College, etc)',
   'Books / Physical media',
   'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Colleague',
   'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification',
   'Books / Physical media;School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp',
   'Books / Physical media;Coding Bootcamp;Colleague',
   'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training',
   'Other (please specify):',
   'Other online resources (e.g., videos, blogs, forum);Colleague;Other (please specify):',
   'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification',
   'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague',
   'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Colleague;Other (please specify):',
   'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training',
   'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification',
   'Friend or family member;Other online resources (e.g., videos, blogs, forum)',
   'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification',
   'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum)',
   'Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification',
   'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification',
   'Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification',
   'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Colleague',
   'Books / Physical media;Other online resources (e.g., videos, blogs, forum)',
   'Other online resources (e.g., videos, blogs, forum);Online Courses or Certification',
   'Other online resources (e.g., videos, blogs, forum);Other (please specify):',
   'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc)',
   'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc)',
   'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification'])
```

ine Courses or Certification',

'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Hackathons (virtual or in-person)',

'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Other (please specify):',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify):',

'School (i.e., University, College, etc);Other (please specify):',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification',

'School (i.e., University, College, etc);Colleague',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Colleague',

'Other online resources (e.g., videos, blogs, forum);On the job training',

'Books / Physical media;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Boot camp',

'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Hackathons (virtual or in-person)',

'Books / Physical media;Friend or family member',

'On the job training;Online Courses or Certification;Coding Bootcamp',

'Books / Physical media;Other (please specify):',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Hackathons (vir tual or in-person)',

'Books / Physical media;Friend or family member;Other (please specify):',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Colleague',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training',

'School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',

'Books / Physical media;School (i.e., University, College, etc);On the job tra ining',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certifica tion;Hackathons (virtual or in-person)',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Col league',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Hackathons (virtual or in-person)',

'School (i.e., University, College, etc);On the job training',

'Books / Physical media;Friend or family member;School (i.e., University, Coll ege, etc)',

'Books / Physical media;On the job training',

'School (i.e., University, College, etc);Online Courses or Certification',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp',  
'School (i.e., University, College, etc);On the job training;Online Courses or Certification',  
'Friend or family member;School (i.e., University, College, etc);On the job training',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Colleague',  
'Books / Physical media;On the job training;Online Courses or Certification;Colleague',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague',  
'Books / Physical media;Friend or family member;Online Courses or Certification;Colleague',  
'Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp',  
'Friend or family member',  
'Books / Physical media;Online Courses or Certification;Coding Bootcamp',  
'Books / Physical media;School (i.e., University, College, etc);Online Courses or Certification',  
'Books / Physical media;Online Courses or Certification',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Hackathons (virtual or in-person)',  
'Colleague',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Colleague',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Coding Bootcamp',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Hackathons (virtual or in-person)',  
'School (i.e., University, College, etc);Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;School (i.e., University, College, etc);On the job training;Colleague',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp',  
'Books / Physical media;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Colleague',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Colleague',  
'Other online resources (e.g., videos, blogs, forum);Coding Bootcamp;Colleague',  
'School (i.e., University, College, etc);On the job training;Colleague',  
'Books / Physical media;School (i.e., University, College, etc);On the job training;Online Courses or Certification',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On

the job training;Colleague',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);Other (please specify):',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Colleague',  
        'Friend or family member;School (i.e., University, College, etc)',  
        'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Colleague',  
        'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification',  
        'Books / Physical media;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
        'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
        'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
        'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
        'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Colleague',  
        'Coding Bootcamp',  
        'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Hackathons (virtual or in-person)',  
        'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp',  
        'On the job training',  
        'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification',  
        'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp',  
        'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Hackathons (virtual or in-person)',  
        'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
        'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training',  
        'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Online Courses or Certification',  
        'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Colleague',  
        'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Other (please specify):',  
        'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training',  
        'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Other (please specify):',  
        'Other online resources (e.g., videos, blogs, forum);On the job training;Colleague',  
        'Friend or family member;Coding Bootcamp;Hackathons (virtual or in-person)',  
        'Books / Physical media;On the job training;Coding Bootcamp',  
        'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification',  
        'Other online resources (e.g., videos, blogs, forum);School (i.e., University,

College, etc);On the job training;Online Courses or Certification;Colleague',  
'School (i.e., University, College, etc);Online Courses or Certification;Other  
(please specify):',  
'Friend or family member;Coding Bootcamp',  
'Books / Physical media;Friend or family member;School (i.e., University, Coll  
ege, etc);On the job training;Other (please specify):',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);S  
chool (i.e., University, College, etc);On the job training',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);Online Courses or Certification',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);S  
chool (i.e., University, College, etc);On the job training;Online Courses or Certific  
ation;Colleague',  
'Books / Physical media;Friend or family member;School (i.e., University, Coll  
ege, etc);On the job training;Online Courses or Certification;Coding Bootcamp',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);On the job training;Online Courses or Certification;Colleague',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certi  
fication;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);S  
chool (i.e., University, College, etc);On the job training;Colleague;Hackathons (virt  
ual or in-person)',  
'Books / Physical media;Friend or family member;On the job training;Online Cou  
rses or Certification',  
'Books / Physical media;On the job training;Colleague',  
'On the job training;Online Courses or Certification;Colleague',  
'Books / Physical media;On the job training;Online Courses or Certification',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Ot  
her (please specify):',  
'Friend or family member;On the job training',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);School (i.e., University, College, etc);On the job training;Onl  
ine Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Sc  
hool (i.e., University, College, etc);On the job training;Online Courses or Certifica  
tion;Coding Bootcamp;Colleague',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On  
the job training;Online Courses or Certification;Colleague',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University,  
College, etc);On the job training;Other (please specify):',  
'Books / Physical media;On the job training;Online Courses or Certification;Co  
ding Bootcamp',  
'Online Courses or Certification;Other (please specify):',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);School (i.e., University, College, etc);On the job training;Onl  
ine Courses or Certification;Colleague;Other (please specify):',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);S  
chool (i.e., University, College, etc);Online Courses or Certification',  
'School (i.e., University, College, etc);Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;School (i.e., University, Coll  
ege, etc);Coding Bootcamp',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Colleague',  
'Books / Physical media;Coding Bootcamp',  
'Books / Physical media;School (i.e., University, College, etc);Other (please  
specify):',  
'Other online resources (e.g., videos, blogs, forum);Online Courses or Certifi  
cation;Coding Bootcamp',  
'Books / Physical media;School (i.e., University, College, etc);On the job tra

ining;Colleague;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);Colleague',  
'Books / Physical media;School (i.e., University, College, etc);Coding Bootcamp',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Other (please specify):',  
'School (i.e., University, College, etc);On the job training;Coding Bootcamp',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);Hackathons (virtual or in-person)',  
'Books / Physical media;School (i.e., University, College, etc);On the job training;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;On the job training;Colleague',  
'Books / Physical media;On the job training;Coding Bootcamp;Colleague',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Other (please specify):',  
'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Colleague',  
'On the job training;Online Courses or Certification',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Other (please specify):',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Colleague',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Hackathons (virtual or in-person)',  
'Friend or family member;School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Colleague;Other (please specify):',  
'Other online resources (e.g., videos, blogs, forum);Coding Bootcamp',  
'Books / Physical media;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp',  
'Friend or family member;School (i.e., University, College, etc);Online Courses or Certification',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Other (please specify):',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp;Other (please specify):',  
'Books / Physical media;School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Coding Bootcamp;Colleague;Hack

athons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Colleague',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague',  
    'On the job training;Other (please specify):',  
    'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Colleague;Other (please specify):',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Colleague;Hackathons (virtual or in-person)',  
    'Friend or family member;Other (please specify):',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
    'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Colleague;Other (please specify):',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
    'Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp;Colleague',  
    'Books / Physical media;School (i.e., University, College, etc);Colleague',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp',  
    'School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp',  
    'Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Hackathons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Colleague',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Colleague;Other (please specify):',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Online Courses or Certification;Hackathons (virtual or in-person)',  
    'Books / Physical media;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Other (please specify):',  
    'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp;Colleague',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague'

hool (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague',  
'On the job training;Colleague;Other (please specify):',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);School (i.e., University, College, etc);On the job training;Onl  
ine Courses or Certification;Coding Bootcamp;Colleague',  
'Books / Physical media;School (i.e., University, College, etc);On the job tra  
ining;Other (please specify):',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);School (i.e., University, College, etc);On the job training;Onl  
ine Courses or Certification;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);S  
chool (i.e., University, College, etc);Colleague;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University,  
College, etc);On the job training;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Ot  
her (please specify);Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);On the job training;Online Courses or Certification;Coding Boot  
camp;Colleague',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On  
the job training;Online Courses or Certification;Coding Bootcamp;Colleague',  
'School (i.e., University, College, etc);Coding Bootcamp',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On  
the job training;Coding Bootcamp',  
'Friend or family member;School (i.e., University, College, etc);On the job tr  
aining;Coding Bootcamp',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);School (i.e., University, College, etc);On the job training;Oth  
er (please specify):',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);O  
n the job training',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);Online Courses or Certification;Coding Bootcamp',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On  
line Courses or Certification;Other (please specify):',  
'Other online resources (e.g., videos, blogs, forum);On the job training;Onlin  
e Courses or Certification;Colleague',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);S  
chool (i.e., University, College, etc);Online Courses or Certification;Colleague',  
'School (i.e., University, College, etc);On the job training;Colleague;Hackath  
ons (virtual or in-person)',  
'School (i.e., University, College, etc);On the job training;Online Courses or  
Certification;Coding Bootcamp',  
'Books / Physical media;School (i.e., University, College, etc);Online Courses  
or Certification;Other (please specify):',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);School (i.e., University, College, etc);On the job training;Col  
league;Other (please specify);Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;School (i.e., University, Coll  
ege, etc);Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);On the job training;Onlin  
e Courses or Certification;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);School (i.e., University, College, etc);On the job training;Col  
league;Hackathons (virtual or in-person)',  
'Books / Physical media;School (i.e., University, College, etc);Online Courses  
or Certification;Coding Bootcamp;Colleague',  
'Friend or family member;Online Courses or Certification',  
'Other online resources (e.g., videos, blogs, forum);On the job training;Other  
(please specify):',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Colleague;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Colleague;Other (please specify):',  
'Books / Physical media;School (i.e., University, College, etc);Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Other (please specify):',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Colleague',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Other (please specify):',  
'School (i.e., University, College, etc);On the job training;Other (please specify):',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify):;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
'On the job training;Online Courses or Certification;Other (please specify):',  
'Friend or family member;School (i.e., University, College, etc);On the job training;Colleague',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Other (please specify):;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Online Courses or Certification',  
'Online Courses or Certification;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);On the job training;Colleague;Other (please specify):',  
'Other online resources (e.g., videos, blogs, forum);Coding Bootcamp;Other (please specify):',  
'School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Other (please specify):',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'School (i.e., University, College, etc);Other (please specify):;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Colleague'

chool (i.e., University, College, etc);Other (please specify):',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On  
the job training;Other (please specify):',  
    'Friend or family member;Online Courses or Certification;Colleague',  
    'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);School (i.e., University, College, etc);On the job training;Hac  
kathons (virtual or in-person)',  
    'Friend or family member;School (i.e., University, College, etc);Colleague',  
    'Other online resources (e.g., videos, blogs, forum);School (i.e., University,  
College, etc);Coding Bootcamp',  
    'Books / Physical media;Colleague',  
    'Books / Physical media;On the job training;Other (please specify):',  
    'Other online resources (e.g., videos, blogs, forum);Online Courses or Certifi  
cation;Other (please specify):',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);C  
olleague',  
    'Books / Physical media;Friend or family member;On the job training',  
    'Books / Physical media;School (i.e., University, College, etc);Online Courses  
or Certification;Colleague',  
    'Books / Physical media;Friend or family member;School (i.e., University, Coll  
ege, etc);Online Courses or Certification;Hackathons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);S  
chool (i.e., University, College, etc);Hackathons (virtual or in-person)',  
    'Hackathons (virtual or in-person)',  
    'Online Courses or Certification;Coding Bootcamp',  
    'School (i.e., University, College, etc);Online Courses or Certification;Cordin  
g Bootcamp',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Sc  
hool (i.e., University, College, etc);Online Courses or Certification;Coding Bootcam  
p;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Colleague;Hackathons (virtual  
or in-person)',  
    'Other online resources (e.g., videos, blogs, forum);On the job training;Onlin  
e Courses or Certification;Coding Bootcamp',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);S  
chool (i.e., University, College, etc);Colleague;Other (please specify):',  
    'School (i.e., University, College, etc);Online Courses or Certification;Colle  
ague',  
    'Other online resources (e.g., videos, blogs, forum);School (i.e., University,  
College, etc);On the job training;Coding Bootcamp',  
    'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);School (i.e., University, College, etc);On the job training;Onl  
ine Courses or Certification;Other (please specify):',  
    'School (i.e., University, College, etc);Coding Bootcamp;Hackathons (virtual  
or in-person)',  
    'Books / Physical media;Friend or family member;School (i.e., University, Coll  
ege, etc);Colleague',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Sc  
hool (i.e., University, College, etc);Online Courses or Certification;Coding Bootcam  
p;Colleague',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);S  
chool (i.e., University, College, etc);On the job training;Online Courses or Certific  
ation;Coding Bootcamp',  
    'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certifi  
cation;Other (please specify):',  
    'Online Courses or Certification;Colleague',  
    'Books / Physical media;School (i.e., University, College, etc);On the job tra  
ining;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-perso  
n)',

'Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Colleague',  
'Online Courses or Certification;Coding Bootcamp;Colleague',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify):',  
'Other online resources (e.g., videos, blogs, forum);Colleague;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Friend or family member;Coding Bootcamp;Colleague',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Colleague',  
'Friend or family member;School (i.e., University, College, etc);Other (please specify):',  
'On the job training;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp',  
'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp;Colleague',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Colleague',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Colleague',  
'Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Colleague',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague',  
'Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;On the job training;Online Courses or Certification;Coding Bootcamp;Colleague',  
'Books / Physical media;On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Other (please specify):',  
'Books / Physical media;School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague',  
'Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g.,

videos, blogs, forum);On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',

'School (i.e., University, College, etc);On the job training;Hackathons (virtual or in-person)',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',

'On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',

'Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',

'Friend or family member;On the job training;Online Courses or Certification;Coding Bootcamp;Colleague',

'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague',

'School (i.e., University, College, etc);On the job training;Other (please specify):;Hackathons (virtual or in-person)',

'On the job training;Colleague',

'Books / Physical media;Online Courses or Certification;Colleague',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',

'Coding Bootcamp;Colleague',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Colleague;Other (please specify):;Hackathons (virtual or in-person)',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Hackathons (virtual or in-person)',

'Friend or family member;School (i.e., University, College, etc);Online Courses or Certification;Colleague',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Coding Bootcamp',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Other (please specify):',

'Books / Physical media;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague',

'Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',

'Friend or family member;On the job training;Online Courses or Certification;Colleague',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Colleague;Other (please specify):',

'Books / Physical media;Online Courses or Certification;Other (please specify):',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Colleague',

'Books / Physical media;Friend or family member;Coding Bootcamp',

'Books / Physical media;Coding Bootcamp;Hackathons (virtual or in-person)',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Hackathons (virtual or in-person)',

'School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Other (please specify):',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Colleague;Other (please specify):',

'Books / Physical media;Friend or family member;On the job training;Online Courses or Certification;Colleague',

'Friend or family member;Colleague',

'Other online resources (e.g., videos, blogs, forum);Other (please specify):;Hackathons (virtual or in-person)',

'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);Colleague;Other (please specify):',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Other (please specify):;Hackathons (virtual or in-person)',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Other (please specify):;Hackathons (virtual or in-person)',

'School (i.e., University, College, etc);Coding Bootcamp;Colleague',

'Friend or family member;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',

'Books / Physical media;School (i.e., University, College, etc);Colleague;Hackathons (virtual or in-person)',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Colleague;Other (please specify):',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Hackathons (virtual or in-person)',

'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Online Courses or Certification;Colleague',

'On the job training;Online Courses or Certification;Coding Bootcamp;Colleague',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify):',

'School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague',

'Books / Physical media;Hackathons (virtual or in-person)',

'Books / Physical media;School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Coding Bootcamp;Other (please specify):',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp',

chool (i.e., University, College, etc);On the job training;Online Courses or Certification;Other (please specify):',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Colleague;Hackathons (virtual or in-person)',

'Friend or family member;Online Courses or Certification;Other (please specify):',

'Books / Physical media;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',

'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague',

'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Hackathons (virtual or in-person)',

'Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',

'Friend or family member;Online Courses or Certification;Coding Bootcamp',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Colleague',

'Coding Bootcamp;Hackathons (virtual or in-person)',

'On the job training;Coding Bootcamp;Colleague',

'Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',

'School (i.e., University, College, etc);Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',

'School (i.e., University, College, etc);Colleague;Other (please specify):',

'Books / Physical media;Friend or family member;Colleague',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification',

'Friend or family member;On the job training;Online Courses or Certification',

'Other online resources (e.g., videos, blogs, forum);On the job training;Hackathons (virtual or in-person)',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Hackathons (virtual or in-person)',

'Books / Physical media;Friend or family member;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Colleague;Other (please specify);Hackathons (virtual or in-person)',

'Books / Physical media;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',

'Books / Physical media;School (i.e., University, College, etc);On the job training;Coding Bootcamp',

'Coding Bootcamp;Other (please specify):',

'Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',

'School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague',

'Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Other (please specify);',  
'Books / Physical media;School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Other (please specify);',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Colleague;Hackathons (virtual or in-person)',  
'School (i.e., University, College, etc);Online Courses or Certification;Hackathons (virtual or in-person)',  
'School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;On the job training;Online Courses or Certification;Hackathons (virtual or in-person)',  
'Books / Physical media;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Other (please specify);',  
'Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Other (please specify);',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify);',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Hackathons (virtual or in-person)',  
'School (i.e., University, College, etc);On the job training;Online Courses or Certification;Other (please specify);',  
'Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Friend or family member;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;On the job training;Online Courses or Certification;Coding Bootcamp',  
'Books / Physical media;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Hackathons (virtual or in-person)',  
'On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',

'Books / Physical media;On the job training;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'School (i.e., University, College, etc);On the job training;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);Coding Bootcamp;Colleague',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Other (please specify):',  
'Friend or family member;School (i.e., University, College, etc);Online Courses or Certification;Hackathons (virtual or in-person)',  
'Books / Physical media;Other (please specify):;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify):',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Other (please specify):;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Colleague;Other (please specify):',  
'Friend or family member;School (i.e., University, College, etc);Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Other (please specify):',  
'Books / Physical media;On the job training;Other (please specify):;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Online Courses or Certification;Coding Bootcamp',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify):;Hackathons (virtual or in-person)',  
'Friend or family member;School (i.e., University, College, etc);Colleague;Hackathons (virtual or in-person)',  
'School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;On the job training;Online Courses or Certification;Colleague;Other (please specify):',  
'School (i.e., University, College, etc);On the job training;Online Courses or Certification;Hackathons (virtual or in-person)',  
'Books / Physical media;On the job training;Colleague;Other (please specify):',  
'Books / Physical media;On the job training;Online Courses or Certification;Other (please specify):',  
'Books / Physical media;Online Courses or Certification;Coding Bootcamp;Colleague',  
'Friend or family member;Coding Bootcamp;Other (please specify):',  
'Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',

'Books / Physical media;Friend or family member;On the job training;Colleague;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp',  
'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Other (please specify):',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Other (please specify):;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Colleague;Other (please specify):',  
'School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Colleague;Other (please specify):',  
'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Colleague;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);Coding Bootcamp;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Other (please specify):',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Colleague',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague',  
'Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague',  
'Friend or family member;On the job training;Coding Bootcamp',  
'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Other (please specify):',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Other (please specify):',  
'Books / Physical media;School (i.e., University, College, etc);Other (please specify):;Hackathons (virtual or in-person)',  
'Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Other (please specify):',  
'Books / Physical media;Friend or family member;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Colleague;Other (please specify):',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify):;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',

'Friend or family member;School (i.e., University, College, etc);On the job training;Other (please specify);Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);On the job training;Other (please specify);Hackathons (virtual or in-person)',  
'Books / Physical media;On the job training;Hackathons (virtual or in-person)',  
'Books / Physical media;School (i.e., University, College, etc);Coding Bootcamp;Other (please specify):',  
'Books / Physical media;School (i.e., University, College, etc);On the job training;Colleague;Other (please specify):',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;School (i.e., University, College, etc);Colleague;Other (please specify):',  
'Friend or family member;School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague',  
'On the job training;Online Courses or Certification;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Friend or family member;School (i.e., University, College, etc);Coding Bootcamp;Other (please specify):',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);Coding Bootcamp',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Other (please specify):',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Hackathons (virtual or in-person)',  
'On the job training;Online Courses or Certification;Coding Bootcamp;Other (please specify):',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Friend or family member;Online Courses or Certification;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Other (please specify):',  
'Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp',  
'Books / Physical media;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Other (please specify):',  
'Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Colleague;Other (please specify):',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Colleague;Other (please specify):;Hackathons (virtual or in-person)',  
'Books / Physical media;Online Courses or Certification;Colleague;Other (please specify):',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University,

College, etc);On the job training;Online Courses or Certification;Colleague;Other (please specify):',  
    'Friend or family member;On the job training;Colleague',  
    'Colleague;Other (please specify):',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague;Other (please specify):',  
    'Online Courses or Certification;Colleague;Other (please specify):',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Other (please specify):',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Other (please specify):',  
    'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp',  
    'Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;On the job training;Other (please specify):',  
    'Online Courses or Certification;Coding Bootcamp;Other (please specify):',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Other (please specify);Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',  
    'Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Other (please specify);Hackathons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Colleague;Other (please specify):',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Other (please specify);Hackathons (virtual or in-person)',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Coding Bootcamp;Colleague',  
    'Friend or family member;On the job training;Colleague;Hackathons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Colleague;Other (please specify):',  
    'Friend or family member;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Coding Bootcamp;Hackathons (virtual or in-person)',

'Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Other (please specify):;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Colleague;Other (please specify):',  
'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague',  
'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague',  
'Books / Physical media;Friend or family member;Coding Bootcamp;Colleague',  
'Books / Physical media;Colleague;Other (please specify):',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Other (please specify):;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Other (please specify):',  
'Books / Physical media;School (i.e., University, College, etc);Coding Bootcamp;Colleague',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Colleague;Other (please specify):;Hackathons (virtual or in-person)',  
'School (i.e., University, College, etc);Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Other (please specify):',  
'Books / Physical media;School (i.e., University, College, etc);On the job training;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Hackathons (virtual or in-person)',  
'Books / Physical media;School (i.e., University, College, etc);Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Friend or family member;On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Colleague;Other (please specify):',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Other (please specify):',  
'Friend or family member;On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Other (please specify):;Hackathons (virtual or in-person)',  
'Books / Physical media;Colleague;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Friend or family member;On the job training;Coding Bootcamp;Hackathons (virtual or in-person)',

'Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Colleague;Other (please specify);Hackathons (virtual or in-person)',

'Friend or family member;School (i.e., University, College, etc);Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Colleague;Other (please specify):',

'School (i.e., University, College, etc);On the job training;Colleague;Other (please specify);Hackathons (virtual or in-person)',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Colleague;Other (please specify):',

'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',

'Books / Physical media;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify);Hackathons (virtual or in-person)',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Hackathons (virtual or in-person)',

'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Hackathons (virtual or in-person)',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Other (please specify):',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Colleague;Other (please specify):',

'Other online resources (e.g., videos, blogs, forum);On the job training;Colleague;Hackathons (virtual or in-person)',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',

'School (i.e., University, College, etc);Coding Bootcamp;Other (please specify):',

'Books / Physical media;Friend or family member;Online Courses or Certification;Hackathons (virtual or in-person)',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Colleague;Other (please specify);Hackathons (virtual or in-person)',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Colleague;Other (please specify):',

'Books / Physical media;School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify):',

'Books / Physical media;Online Courses or Certification;Coding Bootcamp;Other (please specify):',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify);Hackathons (virtual or in-person)',

'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',

'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',

'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify):',

'Friend or family member;School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',

'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On

the job training;Online Courses or Certification;Colleague;Other (please specify);;Hackathons (virtual or in-person)',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Other (please specify);',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Colleague;Other (please specify);',  
    'Friend or family member;School (i.e., University, College, etc);Coding Bootcamp',  
    'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Other (please specify);',  
    'Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Online Courses or Certification;Other (please specify);;Hackathons (virtual or in-person)',  
    'Friend or family member;School (i.e., University, College, etc);On the job training;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
    'Friend or family member;On the job training;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify);;Hackathons (virtual or in-person)',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Other (please specify);;Hackathons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Other (please specify);;Hackathons (virtual or in-person)',  
    'School (i.e., University, College, etc);On the job training;Colleague;Other (please specify);',  
    'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Coding Bootcamp;Colleague;Other (please specify);',  
    'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Other (please specify);;Hackathons (virtual or in-person)',  
    'Books / Physical media;School (i.e., University, College, etc);Coding Bootcamp;Hackathons (virtual or in-person)',  
    'On the job training;Hackathons (virtual or in-person)',  
    'School (i.e., University, College, etc);Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Other online resources (e.g.,

videos, blogs, forum);Coding Bootcamp;Colleague',  
'Friend or family member;School (i.e., University, College, etc);On the job training;Other (please specify):',  
'Colleague;Hackathons (virtual or in-person)',  
'School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Online Courses or Certification;Coding Bootcamp;Colleague',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Colleague;Other (please specify):',  
'Books / Physical media;Friend or family member;Online Courses or Certification;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;School (i.e., University, College, etc);Online Courses or Certification;Colleague;Other (please specify):',  
'On the job training;Online Courses or Certification;Colleague;Other (please specify):',  
'Books / Physical media;School (i.e., University, College, etc);On the job training;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Other (please specify):;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;On the job training;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Colleague',  
'Friend or family member;Online Courses or Certification;Coding Bootcamp;Colleague',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Other (please specify):',  
'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Friend or family member;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Other (please specify):;Hackathons (virtual or in-person)',  
'Friend or family member;Colleague;Other (please specify):',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Colleague;Other (please specify):;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;Online Courses or Certification;Colleague;Other (please specify):',  
'Books / Physical media;Friend or family member;On the job training;Hackathons (virtual or in-person)',  
'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague',  
'Friend or family member;School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
'Books / Physical media;On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Colleague',  
'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Hackathons (virtual or in-person)'

on)',  
    'Friend or family member;On the job training;Online Courses or Certification;Coding Bootcamp',  
    'Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;On the job training;Coding Bootcamp',  
    'Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'On the job training;Coding Bootcamp;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;On the job training;Coding Bootcamp;Colleague',  
    'Friend or family member;On the job training;Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',  
    'Friend or family member;Hackathons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify);Hackathons (virtual or in-person)',  
    'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Colleague;Other (please specify);',  
    'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp;Colleague;Other (please specify);',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Coding Bootcamp;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;On the job training;Online Courses or Certification;Coding Bootcamp;Colleague',  
    'Friend or family member;Online Courses or Certification;Colleague;Other (please specify);',  
    'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',  
    'Books / Physical media;School (i.e., University, College, etc);Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Friend or family member;On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Friend or family member;Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Colleague;Other (please specify);Hackathons (virtual or in-person)',  
    'Friend or family member;School (i.e., University, College, etc);Coding Bootcamp;Colleague',  
    'Books / Physical media;On the job training;Online Courses or Certification;Coding Bootcamp;Other (please specify);',  
    'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Coding Bootcamp',  
    'School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Other (please specify);Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;On the job training;Online Courses or Certification'

rses or Certification;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Online Courses or Certification;Other (please specify):',  
        'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Other (please specify);Hackathons (virtual or in-person)',  
            'Books / Physical media;Friend or family member;Coding Bootcamp;Hackathons (virtual or in-person)',  
                'Friend or family member;Other online resources (e.g., videos, blogs, forum);Other (please specify);Hackathons (virtual or in-person)',  
                    'On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify):',  
                        'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Other (please specify);Hackathons (virtual or in-person)',  
                            'Friend or family member;School (i.e., University, College, etc);On the job training;Hackathons (virtual or in-person)',  
                                'Friend or family member;On the job training;Online Courses or Certification;Hackathons (virtual or in-person)',  
                                    'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
                                        'Friend or family member;On the job training;Other (please specify):',  
                                            'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Coding Bootcamp;Other (please specify):',  
                                                'Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Colleague;Other (please specify):',  
                                                    'Coding Bootcamp;Other (please specify);Hackathons (virtual or in-person)',  
                                                        'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Other (please specify):',  
                                                            'Books / Physical media;Coding Bootcamp;Other (please specify):',  
                                                                'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Other (please specify):',  
                                                                    'Books / Physical media;School (i.e., University, College, etc);On the job training;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
                                                                        'Books / Physical media;Friend or family member;On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
                                                                    'Books / Physical media;Friend or family member;Other (please specify);Hackathons (virtual or in-person)',  
                                                                        'School (i.e., University, College, etc);Coding Bootcamp;Colleague;Other (please specify):',  
                                                                            'Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Colleague;Other (please specify);Hackathons (virtual or in-person)',  
                                                                                'Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Other (please specify):',  
                                                                            'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Other (please specify);Hackathons (virtual or in-person)',  
                                                                                'Books / Physical media;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
                                                                                    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Other (please specify):',  
                                                                                    'Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Colleague;Other (please specify);',  
                                                                                'Books / Physical media;Friend or family member;School (i.e., University, College, etc)'

ege, etc);Coding Bootcamp;Colleague',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Other (please specify);Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify);Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Coding Bootcamp;Other (please specify):',  
    'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify);Hackathons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;School (i.e., University, College, etc);Coding Bootcamp;Hackathons (virtual or in-person)',  
    'Friend or family member;On the job training;Online Courses or Certification;Other (please specify):',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify);Hackathons (virtual or in-person)',  
    'Books / Physical media;School (i.e., University, College, etc);Online Courses or Certification;Other (please specify);Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;On the job training;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)',  
    'Books / Physical media;On the job training;Online Courses or Certification;Colleague;Other (please specify):',  
    'Books / Physical media;Friend or family member;School (i.e., University, College, etc);On the job training;Coding Bootcamp;Hackathons (virtual or in-person)',  
    'Friend or family member;School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Hackathons (virtual or in-person)',  
    'Friend or family member;School (i.e., University, College, etc);Coding Bootcamp;Hackathons (virtual or in-person)',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Other (please specify):',  
    'Friend or family member;Online Courses or Certification;Coding Bootcamp;Other (please specify):',  
    'Books / Physical media;Friend or family member;Online Courses or Certification;Other (please specify):',  
    'Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Other (please specify):',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);On the job training;Colleague;Other (please specify);Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Coding Bootcamp;Other (please specify):',  
    'Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp;Colleague;Other (please specify):',  
    'Other online resources (e.g., videos, blogs, forum);On the job training;Online Courses or Certification;Coding Bootcamp;Colleague;Other (please specify):',  
    'On the job training;Other (please specify);Hackathons (virtual or in-person)',  
    'Other online resources (e.g., videos, blogs, forum);Colleague;Other (please s

```
pecify);;Hackathons (virtual or in-person)',  
    'Books / Physical media;On the job training;Coding Bootcamp;Colleague;Hackatho  
ns (virtual or in-person)',  
    'Books / Physical media;School (i.e., University, College, etc);On the job tra  
ining;Coding Bootcamp;Other (please specify):',  
    'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);On the job training;Coding Bootcamp;Colleague',  
    'Friend or family member;Other online resources (e.g., videos, blogs, forum);S  
chool (i.e., University, College, etc);Coding Bootcamp;Other (please specify):;Hackat  
hons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;On the job training;Coding Boo  
tcamp;Hackathons (virtual or in-person)',  
    'Other online resources (e.g., videos, blogs, forum);Coding Bootcamp;Colleagu  
e;Hackathons (virtual or in-person)',  
    'Books / Physical media;Friend or family member;Other online resources (e.g.,  
videos, blogs, forum);Online Courses or Certification;Coding Bootcamp;Other (please s  
pecify):;Hackathons (virtual or in-person)',  
    'Books / Physical media;Other online resources (e.g., videos, blogs, forum);Co  
ding Bootcamp;Colleague;Other (please specify):',  
    'Friend or family member;School (i.e., University, College, etc);On the job tr  
aining;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual  
or in-person)',  
    'Other online resources (e.g., videos, blogs, forum);On the job training;Cordin  
g Bootcamp;Other (please specify):'],  
    dtype=object)
```

In [53]: col\_value\_counts("LearnCode")

Out[53]:

	count	pct
LearnCode		
School (i.e., University, College, etc)	3669	0.051257
Other online resources (e.g., videos, blogs, forum)	3292	0.045991
Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc)	2873	0.040137
Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc)	2697	0.037678
Books / Physical media;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification	2392	0.033417
...	...	...
Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);On the job training;Online Courses or Certification;Coding Bootcamp;Other (please specify);Hackathons (virtual or in-person)	1	0.000014
Friend or family member;Online Courses or Certification;Coding Bootcamp;Colleague;Hackathons (virtual or in-person)	1	0.000014
Books / Physical media;Friend or family member;Other online resources (e.g., videos, blogs, forum);School (i.e., University, College, etc);Online Courses or Certification;Coding Bootcamp;Other (please specify):	1	0.000014
Books / Physical media;School (i.e., University, College, etc);On the job training;Coding Bootcamp;Hackathons (virtual or in-person)	1	0.000014
Other online resources (e.g., videos, blogs, forum);On the job training;Coding Bootcamp;Other (please specify):	1	0.000014

737 rows × 2 columns

In [54]: `col_strip_nested("LearnCode")`

### Unique Values:

Out[54]: `[('Other online resources (e.g., videos, blogs, forum)', 50756), ('School (i.e., University, College, etc)', 44506), ('Books / Physical media', 38994), ('Online Courses or Certification', 33379), ('On the job training', 28523), ('Colleague', 13188), ('Friend or family member', 9987), ('Coding Bootcamp', 7731), ('Hackathons (virtual or in-person)', 5269), ('Other (please specify):', 3558)]`In [55]: `(df
 .LearnCode
 .rename("learnt_coding")
 .fillna("unanswered")
 .str.replace("Other online resources \(\e.g., videos, blogs, forum\)", "online_resources")
 .str.replace("School \(\i.e., University, College, etc\)", "academia", regex=True)
 .str.replace("Books / Physical media", "books")
 .str.replace("Online Courses or Certification", "online_courses"))`

```
.str.replace("On the job training", "job_training")
.str.replace("Friend or family member", "friends_family")
.str.replace("Hackathons \(\virtual or in-person\)", "hackathons", regex=True)
.str.replace("Other \(\please specify\):", "other", regex=True))
```

```
Out[55]: 0                               unanswered
          1                               unanswered
          2           books;friends_family;online_resources;academia
          3                           books;academia
          4           online_resources;academia;job_training
          ...
          73263   books;online_resources;job_training;online_cou...
          73264   online_resources;academia;job_training;online_...
          73265       books;online_resources;academia;online_courses
          73266                   books;job_training
          73267   books;friends_family;online_resources;academia...
Name: learnt_coding, Length: 73268, dtype: object
```

## Observations

- There're 737 unique categories
- Many values occur in less than 1% of the total observations
- Even the most frequent value occurs only in 5% of observations
- Values are nested, separated by a semicolon

## Steps

- The categories can be shortened
- This variable can be handled in various ways:
  - The rare categories (<1%) can be grouped together
  - The categories can be shortened based on presence of keywords
  - Multiple binary columns can be created
- These steps should be kept in mind for performing feature engineering during exploratory analysis

## LearnCodeOnline

```
In [56]: df.LearnCodeOnline
```

```
Out[56]: 0                               NaN
          1                               NaN
          2           Technical documentation;Blogs;Programming Game...
          3                           NaN
          4           Technical documentation;Blogs;Stack Overflow;0...
          ...
          73263   Technical documentation;Blogs;Written Tutorial...
          73264   Technical documentation;Blogs;Written Tutorial...
          73265   Technical documentation;Programming Games;Stac...
          73266                   NaN
          73267   Technical documentation;Blogs;Programming Game...
Name: LearnCodeOnline, Length: 73268, dtype: object
```

```
In [57]: df.LearnCodeOnline.unique()
```

```
Out[57]: array([nan,
   'Technical documentation;Blogs;Programming Games;Written Tutorials;Stack Overflow',
   'Technical documentation;Blogs;Stack Overflow;Online books;Video-based Online Courses;Online challenges (e.g., daily or weekly coding challenges)',
   ...,
   'Written Tutorials;Online books;Video-based Online Courses;How-to videos;Written-based Online Courses;Coding sessions (live or recorded);Certification videos',
   'Programming Games;Stack Overflow;Video-based Online Courses;Online challenges (e.g., daily or weekly coding challenges);How-to videos;Written-based Online Courses;Interactive tutorial;Coding sessions (live or recorded);Certification videos',
   'Technical documentation;Programming Games;Stack Overflow;Online books;Video-based Online Courses;How-to videos;Written-based Online Courses;Coding sessions (live or recorded);Certification videos'],
   dtype=object)
```

```
In [58]: col_value_counts("LearnCodeOnline").head(20)
```

Out[58]:

		count	pct
LearnCodeOnline			
<b>Technical documentation;Blogs;Written Tutorials;Stack Overflow</b>	715	0.014107	
<b>Technical documentation;Blogs;Written Tutorials;Stack Overflow;Online forum</b>	706	0.013929	
<b>Technical documentation;Blogs;Stack Overflow</b>	589	0.011621	
<b>Technical documentation;Blogs;Written Tutorials;Stack Overflow;Online books;Online forum</b>	530	0.010457	
<b>Technical documentation;Blogs;Written Tutorials;Stack Overflow;Online books</b>	447	0.008819	
<b>Technical documentation;Blogs;Written Tutorials;Stack Overflow;Online forum;How-to videos</b>	430	0.008484	
<b>Technical documentation;Blogs;Stack Overflow;How-to videos</b>	398	0.007852	
<b>Technical documentation;Blogs;Stack Overflow;Online forum</b>	374	0.007379	
<b>Technical documentation;Blogs;Written Tutorials;Stack Overflow;How-to videos</b>	370	0.007300	
<b>Technical documentation;Blogs;Stack Overflow;Video-based Online Courses;How-to videos</b>	365	0.007201	
<b>Technical documentation;Blogs;Stack Overflow;Online books</b>	312	0.006156	
<b>Technical documentation;Written Tutorials;Stack Overflow</b>	308	0.006077	
<b>Technical documentation;Blogs;Written Tutorials;Stack Overflow;Online books;Online forum;How-to videos</b>	261	0.005149	
<b>Technical documentation;Blogs;Written Tutorials;Stack Overflow;Video-based Online Courses;How-to videos</b>	256	0.005051	
<b>Technical documentation;Blogs;Stack Overflow;Online forum;How-to videos</b>	244	0.004814	
<b>Technical documentation;Stack Overflow</b>	234	0.004617	
<b>Technical documentation;Blogs;Programming Games;Written Tutorials;Stack Overflow;Online books;Video-based Online Courses;Online challenges (e.g., daily or weekly coding challenges);Online forum;How-to videos;Written-based Online Courses;Auditory material (e.g., podcasts);Interactive tutorial;Coding sessions (live or recorded);Certification videos</b>	229	0.004518	
<b>Technical documentation;Blogs;Written Tutorials;Stack Overflow;Online forum;Written-based Online Courses</b>	219	0.004321	
<b>Technical documentation;Blogs;Stack Overflow;Video-based Online Courses</b>	212	0.004183	
<b>Technical documentation;Blogs;Written Tutorials;Stack Overflow;Online books;Online forum;Written-based Online Courses</b>	210	0.004143	

In [59]: `col_strip_nested("LearnCodeOnline")`**Unique Values:**

```
Out[59]: [('Technical documentation', 44669),
           ('Stack Overflow', 43658),
           ('Blogs', 38192),
           ('How-to videos', 30371),
           ('Written Tutorials', 29436),
           ('Video-based Online Courses', 26064),
           ('Online books', 22238),
           ('Online forum', 20446),
           ('Written-based Online Courses', 17424),
           ('Coding sessions (live or recorded)', 14626),
           ('Interactive tutorial', 13287),
           ('Online challenges (e.g., daily or weekly coding challenges)', 12723),
           ('Certification videos', 7541),
           ('Programming Games', 6752),
           ('Auditory material (e.g., podcasts)', 3652),
           ('Other (Please specify):', 1028)]
```

```
In [60]: df
    .LearnCodeOnline
    .fillna("unanswered")
    .str.replace("Other \Please specify\:", "other", regex=True)
    .str.replace("Technical documentation", "documentation")
    .str.replace("How-to videos", "videos")
    .str.replace("Video-based Online Courses", "courses")
    .str.replace("Online books", "books")
    .str.replace("Online forum", "forums")
    .str.replace("Written-based Online Courses", "courses")
    .str.replace("Coding sessions \live or recorded\:", "coding_sessions", regex=True)
    .str.replace("Written Tutorials", "tutorials")
    .str.replace("Interactive tutorial", "tutorials")
    .str.replace("Certification videos", "videos")
    .str.replace("Auditory material \e.g., podcasts\:", "podcasts", regex=True)
    .str.replace("Online challenges \e.g., daily or weekly coding challenges\:", "coding_challenges")
    .str.lower()
    .str.replace(" ", "_")
    .value_counts()
```

```
Out[60]: unanswered
22583
documentation;blogs;tutorials;stack_overflow
715
documentation;blogs;tutorials;stack_overflow;forums
706
documentation;blogs;stack_overflow
589
documentation;blogs;tutorials;stack_overflow;books;forums
530

...
documentation;blogs;programming_games;tutorials;stack_overflow;books;courses;other
1
documentation;blogs;programming_games;tutorials;courses;videos;courses;other
1
documentation;blogs;tutorials;stack_overflow;books;coding_challenges;videos;podcasts;
tutorials;coding_sessions;videos      1
blogs;tutorials;courses;coding_challenges;videos;coding_sessions;videos
1
documentation;programming_games;stack_overflow;books;courses;videos;courses;coding_se
ssions;videos          1
Name: LearnCodeOnline, Length: 6927, dtype: int64
```

## Observations

- There're 7192 unique entries
- All the entries occur in around 1% of the total observations
  - These could be considered as `rare`
- Values are nested, separated by a semicolon

## Steps

- The categories can be shortened
- This variable can be handled in various ways:
  - The rare categories (<1%) can be grouped together
  - The categories can be shortened based on presence of keywords
  - Multiple binary columns can be created
- These steps should be kept in mind for performing feature engineering during exploratory analysis

## LearnCodeCoursesCert

```
In [61]: df.LearnCodeCoursesCert
```

```
Out[61]: 0                      NaN
          1                      NaN
          2                      NaN
          3                      NaN
          4                      NaN
          ...
          73263                 Udemy
          73264      Coursera;Udemy;Udacity
          73265  Udemy;Codecademy;Pluralsight;edX
          73266                  NaN
          73267      Udemy;Pluralsight
Name: LearnCodeCoursesCert, Length: 73268, dtype: object
```

```
In [62]: df.LearnCodeCoursesCert.unique()
```

```
Out[62]: array(['nan', 'Coursera;Udemy', 'Udemy;Codecademy', 'Coursera;Pluralsight',  
   'Coursera;Udemy;Codecademy;edX;Udacity',  
   'Coursera;Udemy;Pluralsight;edX', 'Udemy', 'Other',  
   'Coursera;Udemy;Udacity', 'Udemy;Pluralsight',  
   'Coursera;Udemy;Pluralsight', 'Codecademy', 'Coursera',  
   'Coursera;Udemy;edX', 'Udemy;Other', 'Pluralsight',  
   'Coursera;Udemy;Codecademy', 'Codecademy;Pluralsight',  
   'Coursera;edX', 'Udemy;Codecademy;Pluralsight',  
   'Pluralsight;Udacity', 'Coursera;Udemy;Other',  
   'Codecademy;Pluralsight;Other',  
   'Udemy;Codecademy;Pluralsight;Other', 'Udemy;Pluralsight;Udacity',  
   'Coursera;Udemy;Codecademy;Udacity', 'Udemy;edX',  
   'Coursera;Udemy;edX;Udacity',  
   'Coursera;Pluralsight;edX;Udacity;Other', 'edX',  
   'Coursera;Codecademy', 'Coursera;Other', 'Codecademy;Other',  
   'Udemy;Codecademy;Pluralsight;edX;Udacity', 'Coursera;Udacity',  
   'Udemy;Pluralsight;Other', 'Coursera;Codecademy;Pluralsight;edX',  
   'Coursera;Udemy;Other;Skillsoft', 'Udemy;Codecademy;Udacity;Other',  
   'Coursera;Codecademy;edX;Other', 'Codecademy;Skillsoft',  
   'Coursera;Udemy;edX;Udacity;Other',  
   'Coursera;Udemy;Codecademy;edX', 'Udemy;Codecademy;Other',  
   'Coursera;Pluralsight;edX',  
   'Coursera;Udemy;Codecademy;Pluralsight', 'Udemy;Skillsoft',  
   'Coursera;Udemy;Codecademy;Pluralsight;edX;Udacity',  
   'Pluralsight;Other', 'Udemy;Udacity', 'Coursera;Codecademy;Other',  
   'Coursera;Pluralsight;Udacity',  
   'Udemy;Codecademy;Pluralsight;Skillsoft',  
   'Udemy;Codecademy;Udacity', 'Pluralsight;edX',  
   'Coursera;Udemy;Pluralsight;edX;Udacity',  
   'Coursera;Udemy;Pluralsight;Other', 'Udemy;edX;Other',  
   'Udemy;Codecademy;Pluralsight;Udacity', 'Udacity',  
   'Coursera;Udemy;edX;Other', 'Coursera;Codecademy;edX',  
   'Coursera;Codecademy;Udacity', 'Codecademy;Pluralsight;Skillsoft',  
   'Coursera;Codecademy;Pluralsight;edX;Other', 'Coursera;edX;Other',  
   'Udemy;Codecademy;edX',  
   'Coursera;Udemy;Codecademy;edX;Udacity;Other', 'Codecademy;edX',  
   'Skillsoft', 'Udemy;Pluralsight;Skillsoft',  
   'Coursera;Codecademy;Pluralsight', 'Coursera;edX;Udacity',  
   'Coursera;edX;Udacity;Other', 'Coursera;Udemy;Pluralsight;Udacity',  
   'edX;Other', 'Udemy;Codecademy;edX;Udacity', 'Codecademy;Udacity',  
   'Coursera;Skillsoft', 'Udemy;Codecademy;Pluralsight;edX',  
   'Udemy;Codecademy;edX;Other', 'Codecademy;Udacity;Other',  
   'Coursera;Udemy;Codecademy;Skillsoft',  
   'Coursera;Codecademy;Udacity;Other',  
   'Coursera;Udemy;Pluralsight;edX;Other',  
   'Coursera;Codecademy;edX;Udacity', 'Codecademy;edX;Udacity',  
   'Codecademy;edX;Other', 'Pluralsight;Udacity;Other',  
   'Coursera;Pluralsight;Other',  
   'Coursera;Udemy;Codecademy;Pluralsight;edX',  
   'Coursera;Udemy;Codecademy;Pluralsight;Other',  
   'Coursera;Codecademy;Pluralsight;edX;Udacity;Other',  
   'Codecademy;Udacity;Skillsoft', 'Coursera;Pluralsight;Skillsoft',  
   'Udemy;Udacity;Other',  
   'Coursera;Udemy;Pluralsight;Other;Skillsoft',  
   'Coursera;Udemy;Codecademy;Pluralsight;Udacity',  
   'Coursera;Udemy;Pluralsight;Udacity;Other', 'Udacity;Other',  
   'Coursera;Udemy;Codecademy;Pluralsight;edX;Udacity;Skillsoft',  
   'Udemy;Codecademy;Pluralsight;Udacity;Other',  
   'Udemy;Pluralsight;edX', 'Coursera;Udemy;Udacity;Other',  
   'Udemy;Pluralsight;edX;Skillsoft', 'Coursera;Udacity;Other',
```

'Coursera;Udemy;Codecademy;Other', 'Coursera;Udemy;Skillsoft',  
'Codecademy;Pluralsight;edX',  
'Coursera;Udemy;Pluralsight;Skillsoft',  
'Coursera;Pluralsight;edX;Other', 'Pluralsight;Skillsoft',  
'Coursera;Pluralsight;edX;Udacity',  
'Coursera;Udemy;Codecademy;Other;Skillsoft',  
'Udemy;Codecademy;Pluralsight;edX;Other',  
'Coursera;Udemy;Codecademy;Pluralsight;edX;Udacity;Other',  
'Coursera;Udemy;Pluralsight;edX;Udacity;Skillsoft',  
'Coursera;Udemy;Codecademy;Pluralsight;Skillsoft',  
'Coursera;Udemy;Codecademy;Pluralsight;edX;Udacity;Other;Skillsoft',  
'Coursera;Codecademy;Udacity;Skillsoft',  
'Coursera;Udemy;Codecademy;Udacity;Other', 'Pluralsight;edX;Other',  
'edX;Udacity', 'Coursera;Udemy;edX;Skillsoft',  
'Coursera;Udemy;edX;Udacity;Skillsoft',  
'Coursera;Codecademy;Skillsoft',  
'Coursera;Udemy;Codecademy;Pluralsight;Udacity;Other',  
'Coursera;Udemy;Codecademy;edX;Skillsoft',  
'Udemy;Pluralsight;Udacity;Other',  
'Udemy;Pluralsight;Other;Skillsoft', 'Udemy;edX;Udacity',  
'Codecademy;Pluralsight;Udacity', 'Pluralsight;edX;Udacity',  
'Codecademy;edX;Udacity;Other', 'Udemy;Pluralsight;edX;Udacity',  
'Udemy;Codecademy;Skillsoft', 'Udemy;Codecademy;Other;Skillsoft',  
'Coursera;Codecademy;Pluralsight;Other',  
'Udemy;Pluralsight;edX;Udacity;Other',  
'Coursera;Udemy;Codecademy;edX;Other',  
'Codecademy;Pluralsight;Other;Skillsoft',  
'Udemy;Codecademy;Pluralsight;Other;Skillsoft',  
'Coursera;Codecademy;Pluralsight;Udacity', 'edX;Other;Skillsoft',  
'Udemy;Pluralsight;edX;Other', 'Udemy;edX;Skillsoft',  
'Codecademy;Pluralsight;Udacity;Other', 'Coursera;Other;Skillsoft',  
'Coursera;Udemy;Pluralsight;edX;Udacity;Other', 'Other;Skillsoft',  
'Coursera;Pluralsight;Udacity;Other',  
'Codecademy;Pluralsight;edX;Other',  
'Coursera;Udemy;Pluralsight;Udacity;Skillsoft',  
'Udemy;Codecademy;edX;Skillsoft', 'edX;Skillsoft',  
'Coursera;Udemy;Codecademy;Pluralsight;edX;Other',  
'Udemy;edX;Udacity;Other',  
'Coursera;Pluralsight;edX;Udacity;Skillsoft',  
'Coursera;Udemy;Codecademy;Udacity;Skillsoft',  
'Udemy;Codecademy;Pluralsight;Udacity;Other;Skillsoft',  
'Coursera;Udemy;Udacity;Skillsoft',  
'Coursera;Codecademy;Other;Skillsoft',  
'Coursera;Udemy;Codecademy;Pluralsight;edX;Skillsoft',  
'edX;Udacity;Other', 'Pluralsight;edX;Udacity;Other',  
'Coursera;Udemy;Codecademy;Pluralsight;Udacity;Skillsoft',  
'Coursera;Codecademy;edX;Udacity;Other',  
'Udemy;Pluralsight;Udacity;Skillsoft', 'Udacity;Skillsoft',  
'Udemy;Codecademy;edX;Udacity;Other',  
'Coursera;Udemy;Codecademy;edX;Other;Skillsoft',  
'Coursera;Codecademy;Pluralsight;Skillsoft',  
'Coursera;Codecademy;Pluralsight;edX;Udacity',  
'Udemy;Codecademy;Udacity;Skillsoft',  
'Coursera;Codecademy;Pluralsight;edX;Skillsoft',  
'Coursera;Udemy;Codecademy;edX;Udacity;Other;Skillsoft',  
'Coursera;Udemy;Pluralsight;edX;Skillsoft',  
'Coursera;Codecademy;edX;Skillsoft', 'Codecademy;edX;Skillsoft',  
'Pluralsight;Other;Skillsoft',

```
'Coursera;Pluralsight;Udacity;Skillsoft',
'Coursera;Udemy;Codecademy;edX;Udacity;Skillsoft',
'Coursera;Pluralsight;edX;Skillsoft',
'Codecademy;Pluralsight;edX;Udacity;Other',
'Coursera;Pluralsight;edX;Other;Skillsoft',
'Pluralsight;edX;Other;Skillsoft',
'Udemy;Codecademy;Pluralsight;edX;Other;Skillsoft',
'Pluralsight;edX;Skillsoft',
'Udemy;Pluralsight;edX;Udacity;Other;Skillsoft',
'Coursera;Udemy;Codecademy;Pluralsight;Other;Skillsoft',
'Codecademy;Pluralsight;edX;Udacity', 'Udemy;Udacity;Skillsoft',
'Coursera;Codecademy;Pluralsight;edX;Udacity;Skillsoft',
'Coursera;Udemy;Pluralsight;edX;Other;Skillsoft',
'Coursera;edX;Skillsoft', 'Codecademy;Other;Skillsoft',
'Udemy;Codecademy;Pluralsight;Udacity;Skillsoft',
'Coursera;Codecademy;Pluralsight;Udacity;Other;Skillsoft',
'Udacity;Other;Skillsoft',
'Coursera;Codecademy;Pluralsight;Udacity;Other',
'Coursera;Codecademy;edX;Udacity;Skillsoft',
'Udemy;Codecademy;Pluralsight;edX;Udacity;Skillsoft',
'Udemy;Udacity;Other;Skillsoft',
'Coursera;Pluralsight;Other;Skillsoft'], dtype=object)
```

In [63]: `col_value_counts("LearnCodeCoursesCert")`

Out[63]:

		count	pct
<b>LearnCodeCoursesCert</b>			
<b>Udemy</b>	5643	0.192011	
<b>Other</b>	2594	0.088264	
<b>Coursera;Udemy</b>	1893	0.064412	
<b>Udemy;Codecademy</b>	1472	0.050087	
<b>Udemy;Pluralsight</b>	1402	0.047705	
...	...	...	...
<b>Udemy;Codecademy;edX;Udacity;Skillsoft</b>	1	0.000034	
<b>Coursera;Other;Skillsoft</b>	1	0.000034	
<b>Codecademy;edX;Udacity;Skillsoft</b>	1	0.000034	
<b>Pluralsight;edX;Udacity;Other</b>	1	0.000034	
<b>Coursera;Pluralsight;Other;Skillsoft</b>	1	0.000034	

206 rows × 2 columns

In [64]: `col_strip_nested("LearnCodeCoursesCert")`

**Unique Values:**

```
Out[64]: [('Udemy', 19540),
           ('Coursera', 10261),
           ('Codecademy', 7712),
           ('Pluralsight', 6594),
           ('Other', 6528),
           ('edX', 4590),
           ('Udacity', 3995),
           ('Skillsoft', 553)]
```

```
In [65]: (df
          .LearnCodeCoursesCert
          .fillna("unanswered")
          .str.lower()
          .value_counts())
```

```
Out[65]: unanswered                43879
udemy                      5643
other                      2594
coursera;udemy              1893
udemy;codecademy            1472
...
edx;other;skillsoft          1
codecademy;pluralsight;other;skillsoft    1
udemy;codecademy;other;skillsoft          1
coursera;codecademy;udacity;skillsoft      1
coursera;pluralsight;other;skillsoft        1
Name: LearnCodeCoursesCert, Length: 207, dtype: int64
```

## Observations

- There're 206 unique entries
- Entries are nested, separated by a semicolon

## Steps

- The entries can be shortened
- This variable can be handled in various ways:
  - The rare categories (<1%) can be grouped together
  - The categories can be shortened based on presence of keywords

## YearsCode

```
In [66]: df.YearsCode
```

```
Out[66]: 0      NaN
1      NaN
2      14
3      20
4       8
...
73263     8
73264     6
73265    42
73266    50
73267    16
Name: YearsCode, Length: 73268, dtype: object
```

```
In [67]: df.YearsCode.unique()
```

```
Out[67]: array([nan, '14', '20', '8', '15', '3', '1', '6', '37', '5', '12', '22',
   '11', '4', '7', '13', '36', '2', '25', '10', '40', '16', '27',
   '24', '19', '9', '17', '18', '26', 'More than 50 years', '29',
   '30', '32', 'Less than 1 year', '48', '45', '38', '39', '28', '23',
   '43', '21', '41', '35', '50', '33', '31', '34', '46', '44', '42',
   '47', '49'], dtype=object)
```

```
In [68]: col_value_counts("YearsCode")
```

Out[68]:

YearsCode	count	pct
<b>10</b>	5217	0.073138
<b>5</b>	5193	0.072801
<b>6</b>	4651	0.065203
<b>4</b>	4480	0.062806
<b>7</b>	4237	0.059399
<b>8</b>	4227	0.059259
<b>3</b>	4122	0.057787
<b>2</b>	3351	0.046978
<b>12</b>	2995	0.041987
<b>15</b>	2962	0.041525
<b>20</b>	2659	0.037277
<b>9</b>	2581	0.036183
<b>11</b>	1819	0.025501
<b>14</b>	1811	0.025389
<b>13</b>	1624	0.022767
<b>25</b>	1582	0.022178
<b>1</b>	1566	0.021954
<b>Less than 1 year</b>	1413	0.019809
<b>16</b>	1406	0.019711
<b>30</b>	1330	0.018645
<b>22</b>	1176	0.016487
<b>18</b>	1150	0.016122
<b>17</b>	1080	0.015141
<b>40</b>	777	0.010893
<b>24</b>	752	0.010542
<b>35</b>	742	0.010402
<b>23</b>	644	0.009028
<b>19</b>	563	0.007893
<b>26</b>	489	0.006855
<b>27</b>	485	0.006799
<b>21</b>	483	0.006771
<b>28</b>	425	0.005958
<b>32</b>	371	0.005201

YearsCode	count	pct
<b>38</b>	288	0.004038
<b>37</b>	277	0.003883
<b>34</b>	241	0.003379
<b>36</b>	236	0.003309
<b>33</b>	229	0.003210
<b>42</b>	216	0.003028
<b>29</b>	208	0.002916
<b>More than 50 years</b>	172	0.002411
<b>31</b>	170	0.002383
<b>45</b>	167	0.002341
<b>39</b>	159	0.002229
<b>41</b>	143	0.002005
<b>43</b>	115	0.001612
<b>44</b>	104	0.001458
<b>50</b>	68	0.000953
<b>46</b>	57	0.000799
<b>47</b>	42	0.000589
<b>48</b>	41	0.000575
<b>49</b>	35	0.000491

## Observations

- This is an object type column but should be numeric
- Contains some string values

## Steps

- Some values need to be replaced:
  - `More than 50 years` ---> 50
  - `Less than 1 year` ---> 1
- The type should be `int`

## YearsCodePro

In [69]: `df.YearsCodePro`

```
Out[69]: 0      NaN
          1      NaN
          2      5
          3     17
          4      3
          ...
          73263    5
          73264    5
          73265   33
          73266   31
          73267    5
Name: YearsCodePro, Length: 73268, dtype: object
```

```
In [70]: df.YearsCodePro.unique()
```

```
Out[70]: array([nan, '5', '17', '3', '6', '30', '2', '10', '15', '4', '22', '20',
       '40', '9', '14', '21', '7', '18', '25', '8', '12', '45', '1', '19',
       '28', '24', '11', '23', 'Less than 1 year', '32', '27', '16', '44',
       '26', '37', '46', '13', '31', '39', '34', '38', '35', '29', '42',
       '36', '33', '43', '41', '48', '50', 'More than 50 years', '47',
       '49'], dtype=object)
```

- Same steps as YearsCode

## DevType

```
In [71]: df.DevType
```

```
Out[71]: 0                      NaN
          1                      NaN
          2      Data scientist or machine learning specialist;...
          3      Developer, full-stack
          4      Developer, front-end;Developer, full-stack;Dev...
          ...
          73263      Developer, back-end
          73264      Data scientist or machine learning specialist
          73265      Developer, full-stack;Developer, desktop or en...
          73266      Developer, front-end;Developer, desktop or ent...
          73267      Developer, front-end;Engineer, data;Engineer, ...
Name: DevType, Length: 73268, dtype: object
```

```
In [72]: df.DevType.unique()
```

```
Out[72]: array([nan,
   'Data scientist or machine learning specialist;Developer, front-end;Engineer,
data;Engineer, site reliability',
   'Developer, full-stack', ...,
   'Data scientist or machine learning specialist;Developer, front-end;Developer,
full-stack;Developer, back-end;Developer, QA or test;Developer, mobile;Database admin
istrator;Cloud infrastructure engineer;Data or business analyst;Designer;Blockchain',
   'Developer, front-end;Developer, full-stack;Developer, back-end;Developer, des
ktop or enterprise applications;Developer, mobile;Educator;Developer, embedded applic
ations or devices',
   'Developer, front-end;Engineer, data;Engineer, site reliability;Developer, ful
l-stack;Developer, back-end;Developer, desktop or enterprise applications;Developer,
QA or test;Student;Developer, mobile;Academic researcher;DevOps specialist;Developer,
embedded applications or devices;Developer, game or graphics;Cloud infrastructure eng
ineer;Data or business analyst;Designer;Scientist;Product manager;System administrato
r'],
   dtype=object)
```

```
In [73]: col_value_counts("DevType").head(50)
```

Out[73]:

DevType	count	pct
<b>Developer, full-stack</b>	7142	0.116505
<b>Developer, back-end</b>	5301	0.086474
<b>Developer, front-end</b>	2385	0.038906
<b>Developer, front-end;Developer, full-stack;Developer, back-end</b>	1807	0.029477
<b>Developer, full-stack;Developer, back-end</b>	1535	0.025040
<b>Developer, mobile</b>	1492	0.024339
<b>Other (please specify):</b>	1229	0.020048
<b>Student</b>	976	0.015921
<b>Developer, front-end;Developer, full-stack</b>	941	0.015350
<b>Developer, desktop or enterprise applications</b>	758	0.012365
<b>Developer, back-end;Developer, desktop or enterprise applications</b>	666	0.010864
<b>Developer, embedded applications or devices</b>	579	0.009445
<b>Engineering manager</b>	528	0.008613
<b>Data scientist or machine learning specialist</b>	498	0.008124
<b>Developer, front-end;Developer, full-stack;Developer, back-end;Developer, desktop or enterprise applications</b>	472	0.007700
<b>Developer, full-stack;Student</b>	469	0.007651
<b>Developer, full-stack;Developer, desktop or enterprise applications</b>	364	0.005938
<b>Developer, full-stack;Developer, mobile</b>	361	0.005889
<b>Developer, back-end;Cloud infrastructure engineer</b>	350	0.005709
<b>Developer, full-stack;Developer, back-end;Developer, desktop or enterprise applications</b>	341	0.005563
<b>Developer, front-end;Developer, mobile</b>	335	0.005465
<b>Engineer, data</b>	334	0.005448
<b>Developer, front-end;Developer, full-stack;Developer, back-end;Developer, mobile</b>	321	0.005236
<b>Developer, QA or test</b>	316	0.005155
<b>Developer, back-end;Student</b>	306	0.004992
<b>Academic researcher</b>	294	0.004796
<b>Developer, back-end;DevOps specialist</b>	293	0.004780
<b>Developer, front-end;Developer, back-end</b>	293	0.004780
<b>Senior Executive (C-Suite, VP, etc.)</b>	284	0.004633
<b>Developer, full-stack;DevOps specialist</b>	258	0.004209
<b>Developer, full-stack;Engineering manager</b>	252	0.004111

		count	pct
DevType			
	<b>Data or business analyst</b>	244	0.003980
	<b>DevOps specialist</b>	243	0.003964
	<b>Engineer, data;Developer, back-end</b>	239	0.003899
	<b>Developer, front-end;Student</b>	188	0.003067
	<b>Developer, back-end;Engineering manager</b>	180	0.002936
	<b>Developer, game or graphics</b>	175	0.002855
	<b>Developer, front-end;Developer, full-stack;Developer, back-end;Student</b>	172	0.002806
	<b>System administrator</b>	169	0.002757
	<b>Developer, back-end;Developer, mobile</b>	168	0.002741
	<b>Developer, desktop or enterprise applications;Developer, embedded applications or devices</b>	160	0.002610
	<b>Developer, full-stack;Cloud infrastructure engineer</b>	149	0.002431
	<b>Developer, front-end;Designer</b>	148	0.002414
	<b>Academic researcher;Scientist</b>	145	0.002365
	<b>Developer, front-end;Developer, full-stack;Developer, mobile</b>	143	0.002333
	<b>Security professional</b>	135	0.002202
	<b>Developer, back-end;DevOps specialist;Cloud infrastructure engineer</b>	133	0.002170
	<b>Product manager</b>	132	0.002153
	<b>Developer, front-end;Developer, full-stack;Developer, back-end;DevOps specialist</b>	131	0.002137
	<b>Cloud infrastructure engineer</b>	130	0.002121

In [74]: `col_strip_nested("DevType")`

Unique Values:

```
Out[74]: [('Developer, full-stack', 28701),
('Developer, back-end', 26595),
('Developer, front-end', 15915),
('Developer, desktop or enterprise applications', 9546),
('Developer, mobile', 7634),
('DevOps specialist', 6170),
('Student', 5595),
('Cloud infrastructure engineer', 5283),
('Database administrator', 4934),
('System administrator', 4908),
('Developer, embedded applications or devices', 3923),
('Project manager', 3897),
('Designer', 3764),
('Engineer, data', 3600),
('Engineering manager', 3574),
('Data scientist or machine learning specialist', 3424),
('Data or business analyst', 3201),
('Developer, QA or test', 3096),
('Academic researcher', 2709),
('Other (please specify):', 2618),
('Product manager', 2514),
('Educator', 2090),
('Engineer, site reliability', 1947),
('Security professional', 1928),
('Developer, game or graphics', 1837),
('Senior Executive (C-Suite, VP, etc.)', 1805),
('Scientist', 1762),
('Blockchain', 1302),
('Marketing or sales professional', 518)]
```

```
In [75]: df
.DevType
.fillna("unanswered")
.str.replace("Other \Please specify\:", "other", regex=True)
.str.replace("Developer, embedded applications or devices", "embedded_app_dev")
.str.replace("Engineer, data", "data_engineer")
.str.replace("Developer, desktop or enterprise applications", "enterprise_app_dev")
.str.replace("Developer, full-stack", "full_stack_dev")
.str.replace("Developer, front-end", "front_end_dev")
.str.replace("Developer, back-end", "back_end_dev")
.str.replace("Developer, mobile", "mobile_dev")
.str.replace("Data scientist or machine learning specialist", "data_scientist")
.str.replace("Data or business analyst", "data_analyst")
.str.replace("Developer, QA or test", "testing_dev")
.str.replace("Engineer, site reliability", "site_reliability_engineer")
.str.replace("Developer, game or graphics", "game_dev")
.str.replace("Senior Executive \C-Suite, VP, etc.\", "senior_executive", regex=True)
.str.replace("Marketing or sales professional", "marketing_professional")
.lower()
.str.replace(" ", "_")
.value_counts()
```

```
Out[75]: unanswered
11966
full_stack_dev
7142
back_end_dev
5301
front_end_dev
2385
front_end_dev;full_stack_dev;back_end_dev
1807

...
site_reliability_engineer;testing_dev;devops_specialist;project_manager;system_administrator;security_professional
1
front_end_dev;full_stack_dev;back_end_dev;project_manager;product_manager;blockchain
1
front_end_dev;testing_dev;game_dev;system_administrator
1
database_administrator;academic_researcher;designer;blockchain
1
front_end_dev;data_engineer;site_reliability_engineer;full_stack_dev;back_end_dev;enterprise_app_dev;testing_dev;student;mobile_dev;academic_researcher;devops_specialist;embedded_app_dev;game_dev;cloud_infrastructure_engineer;data_analyst;designer;scientist;product_manager;system_administrator      1
Name: DevType, Length: 9985, dtype: int64
```

## Observations

- There're 9,984 unique entries
- Entries are nested, separated by a semicolon

## Steps

- The entries can be shortened
- This variable can be handled in various ways:
  - The rare categories (<1%) can be grouped together
  - The categories can be shortened based on presence of keywords
  - Multiple binary columns can be created
- These steps should be kept in mind for performing feature engineering during exploratory analysis

## OrgSize

```
In [76]: df.OrgSize
```

```
Out[76]: 0          NaN
         1          NaN
         2    20 to 99 employees
         3  100 to 499 employees
         4    20 to 99 employees
            ...
        73263  100 to 499 employees
        73264      I don't know
        73265    20 to 99 employees
        73266    10 to 19 employees
        73267          NaN
Name: OrgSize, Length: 73268, dtype: object
```

In [77]: `df.OrgSize.unique()`

```
Out[77]: array([nan, '20 to 99 employees', '100 to 499 employees', 'I don't know',
       'Just me - I am a freelancer, sole proprietor, etc.',
       '2 to 9 employees', '5,000 to 9,999 employees',
       '1,000 to 4,999 employees', '10,000 or more employees',
       '500 to 999 employees', '10 to 19 employees'], dtype=object)
```

In [78]: `col_value_counts("OrgSize")`

OrgSize	count	pct
<b>20 to 99 employees</b>	10343	0.202649
<b>100 to 499 employees</b>	9289	0.181998
<b>10,000 or more employees</b>	6922	0.135622
<b>1,000 to 4,999 employees</b>	5736	0.112385
<b>2 to 9 employees</b>	4887	0.095750
<b>10 to 19 employees</b>	4251	0.083289
<b>500 to 999 employees</b>	3645	0.071416
<b>Just me - I am a freelancer, sole proprietor, etc.</b>	2771	0.054292
<b>5,000 to 9,999 employees</b>	2189	0.042889
<b>I don't know</b>	1006	0.019710

In [79]: `df.OrgSize.isna().sum()`

Out[79]: 22229

```
In [80]: (df
      .OrgSize
      .str.replace(" employees", "")
      .replace("Just me - I am a freelancer, sole proprietor, etc.", "freelancer")
      .replace(["2 to 9",
                "10 to 19",
                "20 to 99",
                "100 to 499",
                "500 to 999",
                "1,000 to 4,999",
                "5,000 to 9,999",
                "10,000 or more employees"])
```

```

        "5,000 to 9,999",
        "10,000 or more",
        "I don't know"],
      ["small",
       "small",
       "small",
       "small",
       "medium",
       "medium",
       "large",
       "large",
       "unknown"])
.value_counts(dropna=False))

```

```

Out[80]: small      28770
          NaN       22229
          medium     9381
          large      9111
          freelancer 2771
          unknown    1006
Name: OrgSize, dtype: int64

```

## Observations

- There're 10 unique entries
- the entries are unnecessarily lengthy
- Less cardinality

## Steps

- The entries can be grouped together to make meaningful values
- The type can be made `category` for less memory usage

## PurchaseInfluence

```

In [81]: df.PurchaseInfluence

```

```

Out[81]: 0                  NaN
          1                  NaN
          2      I have some influence
          3      I have some influence
          4      I have some influence
          ...
          73263     I have some influence
          73264     I have little or no influence
          73265     I have a great deal of influence
          73266     I have a great deal of influence
          73267                  NaN
Name: PurchaseInfluence, Length: 73268, dtype: object

```

```
In [82]: df.PurchaseInfluence.unique()
```

```
Out[82]: array([nan, 'I have some influence', 'I have little or no influence',
   'I have a great deal of influence'], dtype=object)
```

```
In [83]: col_value_counts("PurchaseInfluence")
```

Out[83]:

	count	pct
<b>PurchaseInfluence</b>		
I have some influence	21991	0.431458
I have little or no influence	17345	0.340305
I have a great deal of influence	11633	0.228237

In [84]:

```
(df
    .PurchaseInfluence
    .replace(["I have some influence",
              "I have little or no influence",
              "I have a great deal of influence"],
              ["some",
               "little",
               "great"])
    .astype("category"))
```

Out[84]:

```
0      NaN
1      NaN
2    some
3    some
4    some
...
73263   some
73264   little
73265   great
73266   great
73267     NaN
Name: PurchaseInfluence, Length: 73268, dtype: category
Categories (3, object): ['great', 'little', 'some']
```

## Observations

- There're 3 unique entries
- Less cardinality

## Steps

- The entries can be shortened
- The type can be made `category` for less memory usage

## BuyNewTool

In [85]:

```
col_description("BuyNewTool")
```

### Description:

When buying a new tool or software, how do you discover and research available solutions? Select all that apply.

In [86]:

```
df.BuyNewTool
```

```
Out[86]: 0          NaN
         1          NaN
         2          NaN
         3          Other (please specify):
         4      Start a free trial;Visit developer communities...
                  ...
93263    Visit developer communities like Stack Overflow...
93264    Other (please specify);Ask developers I know/...
93265    Start a free trial;Ask developers I know/work ...
93266    Start a free trial;Visit developer communities...
93267    Start a free trial;Visit developer communities...
Name: BuyNewTool, Length: 93268, dtype: object
```

```
In [87]: # df.BuyNewTool.unique()
```

```
In [88]: col_value_counts("BuyNewTool").head(20)
```

Out[88]:

	count	pct
BuyNewTool		
<b>Start a free trial;Visit developer communities like Stack Overflow;Ask developers I know/work with</b>	11787	0.173433
<b>Start a free trial;Ask developers I know/work with</b>	7454	0.109677
<b>Start a free trial;Visit developer communities like Stack Overflow;Ask developers I know/work with;Read ratings or reviews on third party sites like G2Crowd</b>	5377	0.079117
<b>Start a free trial</b>	5210	0.076659
<b>Visit developer communities like Stack Overflow;Ask developers I know/work with</b>	4886	0.071892
<b>Start a free trial;Visit developer communities like Stack Overflow</b>	4192	0.061681
<b>Ask developers I know/work with</b>	2555	0.037594
<b>Visit developer communities like Stack Overflow</b>	2441	0.035917
<b>Start a free trial;Visit developer communities like Stack Overflow;Read ratings or reviews on third party sites like G2Crowd</b>	2215	0.032591
<b>Start a free trial;Ask developers I know/work with;Read ratings or reviews on third party sites like G2Crowd</b>	2023	0.029766
<b>Visit developer communities like Stack Overflow;Ask developers I know/work with;Read ratings or reviews on third party sites like G2Crowd</b>	1882	0.027692
<b>Other (please specify):</b>	1248	0.018363
<b>Start a free trial;Read ratings or reviews on third party sites like G2Crowd</b>	1219	0.017936
<b>Visit developer communities like Stack Overflow;Read ratings or reviews on third party sites like G2Crowd</b>	1159	0.017053
<b>Start a free trial;Visit developer communities like Stack Overflow;Ask developers I know/work with;Research companies that have advertised on sites I visit</b>	1145	0.016847
<b>Start a free trial;Visit developer communities like Stack Overflow;Ask developers I know/work with;Research companies that have advertised on sites I visit;Read ratings or reviews on third party sites like G2Crowd</b>	1058	0.015567
<b>Start a free trial;Visit developer communities like Stack Overflow;Ask developers I know/work with;Research companies that have advertised on sites I visit;Read ratings or reviews on third party sites like G2Crowd;Research companies that have emailed me</b>	968	0.014243
<b>Start a free trial;Ask developers I know/work with;Research companies that have advertised on sites I visit</b>	659	0.009696
<b>Ask developers I know/work with;Read ratings or reviews on third party sites like G2Crowd</b>	655	0.009638
<b>Start a free trial;Visit developer communities like Stack Overflow;Research companies that have advertised on sites I visit</b>	541	0.007960

In [89]: `col_strip_nested("BuyNewTool")`**Unique Values:**

```
Out[89]: [('Start a free trial', 48849),
           ('Ask developers I know/work with', 45588),
           ('Visit developer communities like Stack Overflow', 42762),
           ('Read ratings or reviews on third party sites like G2Crowd', 20235),
           ('Research companies that have advertised on sites I visit', 9136),
           ('Other (please specify):', 4438),
           ('Research companies that have emailed me', 3667)]
```

```
In [90]: df
    .BuyNewTool
    .fillna("unanswered")
    .str.replace("Other \pleas\especify\:", "other", regex=True)
    .str.replace("Start a free trial", "free_trial")
    .str.replace("Ask developers I know/work with", "ask_known_devs")
    .str.replace("Visit developer communities like Stack Overflow", "stack_overflow")
    .str.replace("Read ratings or reviews on third party sites like G2Crowd", "Read ratir")
    .str.replace("Research companies that have advertised on sites I visit", "Research co")
    .str.replace("Research companies that have emailed me", "Research companies")
    .str.lower()
    .str.replace(" ", "_")
    .value_counts()
    .index)
```

```
Out[90]: Index(['free_trial;stack_overflow;ask_known_devs', 'free_trial;ask_known_devs',
               'free_trial;stack_overflow;ask_known_devs;read_ratings', 'unanswered',
               'free_trial', 'stack_overflow;ask_known_devs',
               'free_trial;stack_overflow', 'ask_known_devs', 'stack_overflow',
               'free_trial;stack_overflow;read_ratings',
               ...
               'read_ratings;research_companies',
               'other;stack_overflow;research_companies;research_companies',
               'other;research_companies;research_companies',
               'other;stack_overflow;read_ratings;research_companies',
               'other;free_trial;read_ratings;research_companies',
               'other;free_trial;stack_overflow;research_companies;read_ratings;research_compa
               nies',
               'other;ask_known_devs;read_ratings;research_companies',
               'other;ask_known_devs;research_companies;research_companies',
               'other;free_trial;ask_known_devs;read_ratings;research_companies',
               'other;ask_known_devs;research_companies;read_ratings'],
              dtype='object', length=110)
```

## Observations

- There're 125 unique entries
- Many entries occur in less than 1% of the total observations
- Entries are nested, separated by a semicolon

## Steps

- The entries can be shortened
- This variable can be handled in various ways:
  - The rare categories (<1%) can be grouped together
  - The categories can be shortened based on presence of keywords
  - Multiple binary columns can be created

- These steps should be kept in mind for performing feature engineering during exploratory analysis

## Country

```
In [91]: df.Country
```

```
Out[91]: 0                      NaN
1                      Canada
2      United Kingdom of Great Britain and Northern I...
3                      Israel
4      United States of America
...
73263                  Nigeria
73264      United States of America
73265      United States of America
73266      United Kingdom of Great Britain and Northern I...
73267                  Canada
Name: Country, Length: 73268, dtype: object
```

```
In [92]: df.Country.unique()
```

```
Out[92]: array([nan, 'Canada',
   'United Kingdom of Great Britain and Northern Ireland', 'Israel',
   'United States of America', 'Germany', 'India', 'Netherlands',
   'Croatia', 'Australia', 'Russian Federation', 'Czech Republic',
   'Austria', 'Serbia', 'Italy', 'Ireland', 'Poland', 'Slovenia',
   'Iraq', 'Sweden', 'Madagascar', 'Norway', 'Taiwan',
   'Hong Kong (S.A.R.)', 'Mexico', 'France', 'Brazil', 'Lithuania',
   'Uruguay', 'Denmark', 'Spain', 'Egypt', 'Turkey', 'South Africa',
   'Ukraine', 'Finland', 'Romania', 'Portugal', 'Singapore', 'Oman',
   'Belgium', 'Chile', 'Bulgaria', 'Latvia', 'Philippines', 'Greece',
   'Belarus', 'Saudi Arabia', 'Kenya', 'Switzerland', 'Iceland',
   'Viet Nam', 'Thailand', 'China', 'Montenegro', 'Slovakia', 'Japan',
   'Luxembourg', 'Turkmenistan', 'Argentina', 'Hungary', 'Tunisia',
   'Bangladesh', 'Maldives', 'Dominican Republic', 'Jordan',
   'Pakistan', 'Nepal', 'Iran, Islamic Republic of...', 'Indonesia',
   'Ecuador', 'Bosnia and Herzegovina', 'Armenia', 'Colombia',
   'Kazakhstan', 'South Korea', 'Costa Rica', 'Honduras', 'Mauritius',
   'Estonia', 'Algeria', 'Trinidad and Tobago', 'Mali', 'Morocco',
   'Swaziland', 'New Zealand',
   'The former Yugoslav Republic of Macedonia', 'Afghanistan',
   'Cyprus', 'United Arab Emirates', 'Peru', 'Uzbekistan', 'Ethiopia',
   'Bahrain', 'Malta', 'Nicaragua', 'Andorra', 'Republic of Korea',
   'Lebanon', 'Belize', 'Zambia', 'Bolivia', 'Malaysia', 'Sri Lanka',
   "Lao People's Democratic Republic", 'Guatemala', 'Azerbaijan',
   'Suriname', 'El Salvador', 'Syrian Arab Republic', 'Qatar',
   'Nigeria', 'Kyrgyzstan', 'Zimbabwe', 'Rwanda', 'Georgia',
   'Cambodia', 'Malawi', 'Yemen', 'Fiji', 'Nomadic', 'Uganda',
   'Albania', 'Timor-Leste', 'Mongolia', 'Republic of Moldova',
   'Tajikistan', 'Ghana', 'United Republic of Tanzania', 'Myanmar',
   'Kuwait', 'Cameroon', 'Kosovo', 'Jamaica', 'Benin', 'Botswana',
   'Niger', 'Palestine', 'Cape Verde', 'Libyan Arab Jamahiriya',
   'Venezuela, Bolivarian Republic of...', 'Senegal', 'Cuba', 'Togo',
   'Angola', 'Isle of Man', 'Panama', 'Bahamas', 'Paraguay', 'Sudan',
   'Liberia', 'Bhutan', 'Democratic Republic of the Congo',
   "Côte d'Ivoire", 'Barbados', 'Congo, Republic of the...',
   'Namibia', 'Somalia', 'Sierra Leone', 'Mozambique', 'Lesotho',
   'Chad', 'North Korea', 'Antigua and Barbuda', 'Papua New Guinea',
   'Palau', 'Guinea', 'Haiti', 'Gabon', 'Mauritania', 'San Marino',
   'Guyana', 'Saint Lucia', 'Burkina Faso', 'Brunei Darussalam',
   'Gambia', 'Monaco', 'Djibouti', 'Seychelles', 'Solomon Islands',
   'Saint Kitts and Nevis'], dtype=object)
```

```
In [93]: with pd.option_context("display.max_rows", 180):
    display(col_value_counts("Country").head(180))
```

		count	pct
Country			
	<b>United States of America</b>	13543	0.188697
	<b>India</b>	6639	0.092503
	<b>Germany</b>	5395	0.075170
<b>United Kingdom of Great Britain and Northern Ireland</b>		4190	0.058380
	<b>Canada</b>	2490	0.034694
	<b>France</b>	2328	0.032436
	<b>Brazil</b>	2109	0.029385
	<b>Poland</b>	1732	0.024132
	<b>Netherlands</b>	1555	0.021666
	<b>Spain</b>	1521	0.021192
	<b>Italy</b>	1514	0.021095
	<b>Australia</b>	1462	0.020370
	<b>Russian Federation</b>	1229	0.017124
	<b>Turkey</b>	1020	0.014212
	<b>Sweden</b>	1020	0.014212
	<b>Switzerland</b>	932	0.012986
	<b>Austria</b>	833	0.011606
	<b>Israel</b>	795	0.011077
<b>Iran, Islamic Republic of...</b>		723	0.010074
	<b>Pakistan</b>	711	0.009907
	<b>Czech Republic</b>	692	0.009642
	<b>China</b>	632	0.008806
	<b>Belgium</b>	626	0.008722
	<b>Bangladesh</b>	613	0.008541
	<b>Ukraine</b>	593	0.008262
	<b>Romania</b>	581	0.008095
	<b>Mexico</b>	574	0.007998
	<b>Portugal</b>	563	0.007844
	<b>Greece</b>	548	0.007635
	<b>Denmark</b>	547	0.007621
	<b>Indonesia</b>	537	0.007482
	<b>Argentina</b>	521	0.007259
	<b>Nigeria</b>	517	0.007203

		count	pct
Country			
	<b>South Africa</b>	493	0.006869
	<b>Norway</b>	479	0.006674
	<b>Finland</b>	468	0.006521
	<b>Hungary</b>	466	0.006493
	<b>New Zealand</b>	440	0.006131
	<b>Egypt</b>	400	0.005573
	<b>Philippines</b>	366	0.005100
	<b>Bulgaria</b>	364	0.005072
	<b>Colombia</b>	341	0.004751
	<b>Japan</b>	333	0.004640
	<b>Ireland</b>	332	0.004626
	<b>Viet Nam</b>	317	0.004417
	<b>Sri Lanka</b>	296	0.004124
	<b>Slovenia</b>	286	0.003985
	<b>Serbia</b>	282	0.003929
	<b>Kenya</b>	270	0.003762
	<b>Taiwan</b>	233	0.003246
	<b>South Korea</b>	229	0.003191
	<b>Slovakia</b>	229	0.003191
	<b>Morocco</b>	226	0.003149
	<b>Singapore</b>	226	0.003149
	<b>Nepal</b>	222	0.003093
	<b>Lithuania</b>	212	0.002954
	<b>Malaysia</b>	206	0.002870
	<b>Chile</b>	203	0.002828
	<b>Croatia</b>	192	0.002675
	<b>Hong Kong (S.A.R.)</b>	184	0.002564
	<b>Thailand</b>	175	0.002438
	<b>United Arab Emirates</b>	165	0.002299
	<b>Estonia</b>	141	0.001965
	<b>Peru</b>	136	0.001895
	<b>Belarus</b>	136	0.001895
	<b>Venezuela, Bolivarian Republic of...</b>	119	0.001658

		count	pct
Country			
	<b>Saudi Arabia</b>	116	0.001616
	<b>Uruguay</b>	106	0.001477
	<b>Georgia</b>	105	0.001463
	<b>Bosnia and Herzegovina</b>	103	0.001435
	<b>Tunisia</b>	101	0.001407
	<b>Latvia</b>	95	0.001324
	<b>Armenia</b>	93	0.001296
	<b>Ghana</b>	86	0.001198
	<b>Uzbekistan</b>	84	0.001170
	<b>Dominican Republic</b>	83	0.001156
	<b>Ecuador</b>	83	0.001156
	<b>Costa Rica</b>	76	0.001059
	<b>Ethiopia</b>	75	0.001045
	<b>Republic of Korea</b>	72	0.001003
	<b>Lebanon</b>	71	0.000989
	<b>Kazakhstan</b>	70	0.000975
	<b>Azerbaijan</b>	69	0.000961
	<b>Afghanistan</b>	69	0.000961
	<b>Jordan</b>	64	0.000892
	<b>Uganda</b>	63	0.000878
	<b>Cyprus</b>	61	0.000850
	<b>Albania</b>	59	0.000822
	<b>Luxembourg</b>	56	0.000780
	<b>Guatemala</b>	54	0.000752
	<b>Iraq</b>	54	0.000752
	<b>Myanmar</b>	53	0.000738
	<b>The former Yugoslav Republic of Macedonia</b>	51	0.000711
	<b>El Salvador</b>	45	0.000627
	<b>Bolivia</b>	44	0.000613
	<b>Paraguay</b>	42	0.000585
	<b>Malta</b>	39	0.000543
	<b>Nomadic</b>	37	0.000516
	<b>Republic of Moldova</b>	37	0.000516

		count	pct
Country			
	<b>Algeria</b>	36	0.000502
	<b>Syrian Arab Republic</b>	35	0.000488
	<b>Cambodia</b>	35	0.000488
	<b>Cameroon</b>	34	0.000474
	<b>Palestine</b>	34	0.000474
	<b>Sudan</b>	32	0.000446
	<b>Cuba</b>	31	0.000432
	<b>Zimbabwe</b>	30	0.000418
	<b>Kosovo</b>	29	0.000404
	<b>Honduras</b>	28	0.000390
	<b>Panama</b>	28	0.000390
	<b>United Republic of Tanzania</b>	28	0.000390
	<b>Iceland</b>	28	0.000390
	<b>Madagascar</b>	26	0.000362
	<b>Jamaica</b>	22	0.000307
	<b>Kyrgyzstan</b>	22	0.000307
	<b>Mauritius</b>	21	0.000293
	<b>Qatar</b>	20	0.000279
	<b>Zambia</b>	20	0.000279
	<b>Rwanda</b>	20	0.000279
	<b>Kuwait</b>	20	0.000279
	<b>Montenegro</b>	19	0.000265
	<b>Mongolia</b>	19	0.000265
	<b>Nicaragua</b>	18	0.000251
	<b>Turkmenistan</b>	16	0.000223
	<b>Andorra</b>	15	0.000209
	<b>Côte d'Ivoire</b>	14	0.000195
	<b>Malawi</b>	14	0.000195
	<b>Bahrain</b>	14	0.000195
	<b>Yemen</b>	14	0.000195
	<b>Oman</b>	13	0.000181
	<b>Trinidad and Tobago</b>	13	0.000181
	<b>Libyan Arab Jamahiriya</b>	13	0.000181

		count	pct
Country			
	<b>Senegal</b>	13	0.000181
	<b>Angola</b>	12	0.000167
	<b>Congo, Republic of the...</b>	12	0.000167
	<b>Somalia</b>	12	0.000167
	<b>Isle of Man</b>	12	0.000167
	<b>Maldives</b>	12	0.000167
	<b>Tajikistan</b>	11	0.000153
	<b>Benin</b>	11	0.000153
	<b>Mozambique</b>	11	0.000153
	<b>Mali</b>	10	0.000139
	<b>Togo</b>	9	0.000125
	<b>Swaziland</b>	7	0.000098
	<b>Democratic Republic of the Congo</b>	7	0.000098
	<b>Lao People's Democratic Republic</b>	7	0.000098
	<b>Barbados</b>	6	0.000084
	<b>Guyana</b>	6	0.000084
	<b>Botswana</b>	6	0.000084
	<b>Guinea</b>	5	0.000070
	<b>Namibia</b>	5	0.000070
	<b>Bhutan</b>	5	0.000070
	<b>Suriname</b>	4	0.000056
	<b>North Korea</b>	4	0.000056
	<b>Timor-Leste</b>	4	0.000056
	<b>Cape Verde</b>	4	0.000056
	<b>Bahamas</b>	4	0.000056
	<b>Niger</b>	3	0.000042
	<b>Belize</b>	3	0.000042
	<b>Saint Lucia</b>	3	0.000042
	<b>Chad</b>	3	0.000042
	<b>Haiti</b>	3	0.000042
	<b>Gambia</b>	2	0.000028
	<b>Mauritania</b>	2	0.000028
	<b>Liberia</b>	2	0.000028

	count	pct
Country		
<b>Gabon</b>	2	0.000028
<b>Palau</b>	2	0.000028
<b>Papua New Guinea</b>	2	0.000028
<b>Antigua and Barbuda</b>	2	0.000028
<b>Lesotho</b>	2	0.000028
<b>Fiji</b>	2	0.000028
<b>Sierra Leone</b>	2	0.000028
<b>San Marino</b>	1	0.000014
<b>Burkina Faso</b>	1	0.000014
<b>Brunei Darussalam</b>	1	0.000014
<b>Monaco</b>	1	0.000014
<b>Djibouti</b>	1	0.000014
<b>Seychelles</b>	1	0.000014
<b>Solomon Islands</b>	1	0.000014
<b>Saint Kitts and Nevis</b>	1	0.000014

In [94]:

```
(df
    .Country
    .replace([
        "Lao People's Democratic Republic",
        "Democratic Republic of the Congo",
        "Brunei Darussalam",
        "Swaziland",
        "Congo, Republic of the...",
        "Libyan Arab Jamahiriya",
        "United Republic of Tanzania",
        "Syrian Arab Republic",
        "Republic of Moldova",
        "The former Yugoslav Republic of Macedonia",
        "Republic of Korea",
        "Venezuela, Bolivarian Republic of...",
        "United Arab Emirates",
        "Hong Kong (S.A.R.)",
        "Viet Nam",
        "Iran, Islamic Republic of...",
        "Russian Federation",
        "United Kingdom of Great Britain and Northern Ireland",
        "United States of America"],
        [
            "Laos",
            "Congo",
            "Brunei",
            "Eswatini",
            "Congo",
            "Libya",
            "Tanzania",
            "Syria",
            "Angola",
            "Moldova",
            "Russia",
            "United Kingdom",
            "United States"])
```

```
"Moldova",
"North Macedonia",
"South Korea",
"Venezuela",
"UAE",
"Hong Kong",
"Vietnam",
"Iran",
"Russia",
"UK",
"USA"]))
```

```
Out[94]: 0      NaN
          1      Canada
          2      UK
          3      Israel
          4      USA
          ...
73263    Nigeria
73264      USA
73265      USA
73266      UK
73267    Canada
Name: Country, Length: 73268, dtype: object
```

```
In [95]: df.Country.dropna().loc[lambda ser: ser.str.contains("Congo")].iloc[1]
```

```
Out[95]: 'Congo, Republic of the...'
```

- Some of the entries were inconsistent and irregular
- These inconsistencies will be handled accordingly

## Currency

```
In [323...]: col_description("Currency")
```

### Description:

Which currency do you use day-to-day? If your answer is complicated, please pick the one you're most comfortable estimating in. \*

```
In [96]: df.Currency
```

```
Out[96]: 0      NaN
          1      CAD\tCanadian dollar
          2      GBP\tPound sterling
          3      ILS\tIsraeli new shekel
          4      USD\tUnited States dollar
          ...
73263    USD\tUnited States dollar
73264    USD\tUnited States dollar
73265    USD\tUnited States dollar
73266      GBP\tPound sterling
73267      NaN
Name: Currency, Length: 73268, dtype: object
```

In [333...]

```
(df
    .loc[df.Country.isna()]
    .Currency
    .unique())
```

Out[333]: array([nan], dtype=object)

In [339...]

```
(df
    .Country
    .fillna("unanswered")
    .replace([
        "Lao People's Democratic Republic", "Democratic Republic of the Congo",
        "Brunei Darussalam", "Swaziland", "Congo, Republic of the...", "Libyan Arab
        United Republic of Tanzania", "Syrian Arab Republic", "Republic of Moldova",
        "The former Yugoslav Republic of Macedonia", "Republic of Korea",
        "Venezuela, Bolivarian Republic of...", "United Arab Emirates",
        "Hong Kong (S.A.R.)", "Viet Nam", "Iran, Islamic Republic of...", "Russian
        United Kingdom of Great Britain and Northern Ireland", "United States of A
        ["Laos", "Congo", "Brunei", "Eswatini", "Congo", "Libya", "Tanzania", "Syria",
        "North Macedonia", "South Korea", "Venezuela", "UAE", "Hong Kong", "Vietnam",
        "Russia", "UK", "USA"]),
    .unique())
```

```
Out[339]: array(['unanswered', 'Canada', 'UK', 'Israel', 'USA', 'Germany', 'India',
    'Netherlands', 'Croatia', 'Australia', 'Russia', 'Czech Republic',
    'Austria', 'Serbia', 'Italy', 'Ireland', 'Poland', 'Slovenia',
    'Iraq', 'Sweden', 'Madagascar', 'Norway', 'Taiwan', 'Hong Kong',
    'Mexico', 'France', 'Brazil', 'Lithuania', 'Uruguay', 'Denmark',
    'Spain', 'Egypt', 'Turkey', 'South Africa', 'Ukraine', 'Finland',
    'Romania', 'Portugal', 'Singapore', 'Oman', 'Belgium', 'Chile',
    'Bulgaria', 'Latvia', 'Philippines', 'Greece', 'Belarus',
    'Saudi Arabia', 'Kenya', 'Switzerland', 'Iceland', 'Vietnam',
    'Thailand', 'China', 'Montenegro', 'Slovakia', 'Japan',
    'Luxembourg', 'Turkmenistan', 'Argentina', 'Hungary', 'Tunisia',
    'Bangladesh', 'Maldives', 'Dominican Republic', 'Jordan',
    'Pakistan', 'Nepal', 'Iran', 'Indonesia', 'Ecuador',
    'Bosnia and Herzegovina', 'Armenia', 'Colombia', 'Kazakhstan',
    'South Korea', 'Costa Rica', 'Honduras', 'Mauritius', 'Estonia',
    'Algeria', 'Trinidad and Tobago', 'Mali', 'Morocco', 'Eswatini',
    'New Zealand', 'North Macedonia', 'Afghanistan', 'Cyprus', 'UAE',
    'Peru', 'Uzbekistan', 'Ethiopia', 'Bahrain', 'Malta', 'Nicaragua',
    'Andorra', 'Lebanon', 'Belize', 'Zambia', 'Bolivia', 'Malaysia',
    'Sri Lanka', 'Laos', 'Guatemala', 'Azerbaijan', 'Suriname',
    'El Salvador', 'Syria', 'Qatar', 'Nigeria', 'Kyrgyzstan',
    'Zimbabwe', 'Rwanda', 'Georgia', 'Cambodia', 'Malawi', 'Yemen',
    'Fiji', 'Nomadic', 'Uganda', 'Albania', 'Timor-Leste', 'Mongolia',
    'Moldova', 'Tajikistan', 'Ghana', 'Tanzania', 'Myanmar', 'Kuwait',
    'Cameroon', 'Kosovo', 'Jamaica', 'Benin', 'Botswana', 'Niger',
    'Palestine', 'Cape Verde', 'Libya', 'Venezuela', 'Senegal', 'Cuba',
    'Togo', 'Angola', 'Isle of Man', 'Panama', 'Bahamas', 'Paraguay',
    'Sudan', 'Liberia', 'Bhutan', 'Congo', "Côte d'Ivoire", 'Barbados',
    'Namibia', 'Somalia', 'Sierra Leone', 'Mozambique', 'Lesotho',
    'Chad', 'North Korea', 'Antigua and Barbuda', 'Papua New Guinea',
    'Palau', 'Guinea', 'Haiti', 'Gabon', 'Mauritania', 'San Marino',
    'Guyana', 'Saint Lucia', 'Burkina Faso', 'Brunei', 'Gambia',
    'Monaco', 'Djibouti', 'Seychelles', 'Solomon Islands',
    'Saint Kitts and Nevis'], dtype=object)
```

In [ ]:

```
currency_map = {"unanswered": "unanswered", "Canada": "CAD", "UK": "GBP", "Israel": "]
    "USA": "USD", "Germany": "EUR", "India": "INR", "Netherlands": "EUR",
```

```

"Croatia": "HRK", "Australia": "AUD", "Russia": "RUB",
"Czech Republic": "CZK", "Austria": "EUR", "Serbia": "RSD", "Italy": "EUR",
"Ireland": "EUR", "Poland": "PLN", "Slovenia": "EUR", "Iraq": "IQD",
"Sweden": "SEK", "Madagascar": "MGA", "Norway": "NOK", "Taiwan": "TWD",
"Hong Kong": "HKD", "Mexico": "MXN", "France": "EUR", "Brazil": "BRL",
"Lithuania": "LTL", "Uruguay": "UYU", "Denmark": "DKK", "Spain": "EUR",
"Egypt": "EGP", "Turkey": "TRY", "South Africa": "ZAR", "Ukraine": "UAH",
"Finland": "EUR", "Romania": "RON", "Portugal": "EUR", "Singapore": "SGD",
"Oman": "OMR", "Belgium": "EUR", "Chile": "CLP", "Bulgaria": "BGN",
"Latvia": "LVL", "Philippines": "PHP", "Greece": "EUR", "Belarus": "BYN",
"Saudi Arabia": "SAR", "Kenya": "KES", "Switzerland": "CHF", "Iceland": "ISK",
"Vietnam": "VND", "Thailand": "THB", "China": "CNY", "Montenegro": "EUR",
"Slovakia": "EUR", "Japan": "JPY", "Luxembourg": "LUF", "Turkmenistan": "TMT",
"Argentina": "ARS", "Hungary": "HUF", "Tunisia": "TND", "Bangladesh": "BDT",
"Maldives": "MVR", "Dominican Republic": "DOP", "Jordan": "JOD", "Pakistan": "PKR",
"Nepal": "NPR", "Iran": "IRR", "Indonesia": "IDR", "Ecuador": "USD",
"Bosnia and Herzegovina": "BAM", "Armenia": "AMD",
"Colombia": "COP", "Kazakhstan": "KZT", "South Korea": "KRW",
"Costa Rica": "CRC", "Honduras": "HNL", "Mauritius": "MUR",
"Estonia": "EUR", "Algeria": "DZD", "Trinidad and Tobago": "TTD",
"Malta": "EUR", "Morocco": "MAD", "Eswatini": "SZL",
"New Zealand": "NZD", "North Macedonia": "MKD", "Afghanistan": "AFN",
"Cyprus": "CYP", "UAE": "AED", "Peru": "PEN",
"Uzbekistan": "UZS", "Ethiopia": "ETB", "Bahrain": "BHD",
"Malta": "MLT", "Nicaragua": "NIO", "Andorra": "ADP",
"Lebanon": "LBP", "Belize": "BZD", "Zambia": "ZMW",
"Bolivia": "BOB", "Malaysia": "MYR", "Sri Lanka": "LKR",
"Laos": "LAK", "Guatemala": "GTQ", "Azerbaijan": "AZN",
"Suriname": "SRD", "El Salvador": "USD", "Syria": "SYP",
"Qatar": "QAR", "Nigeria": "NGN", "Kyrgyzstan": "KGS",
"Zimbabwe": "ZWD", "Rwanda": "RWF", "Georgia": "GEL",
"Cambodia": "KHR", "Malawi": "MWK", "Yemen": "YER",
"Fiji": "FJD", "Nomadic": "unknown", "Uganda": "UGX",
"Albania": "ALL", "Timor-Leste": "USD", "Mongolia": "MNT",
"Moldova": "MDL", "Tajikistan": "TJS", "Ghana": "GHS",
"Tanzania": "TZS", "Myanmar": "MMK", "Kuwait": "KWD",
"Cameroon": "XAF", "Kosovo": "EUR", "Jamaica": "JMD",
"Benin": "XOF", "Botswana": "BWP", "Niger": "XOF",
"Palestine": "EGP", "Cape Verde": "CVE", "Libya": "LYD",
"Venezuela": "VES", "Senegal": "XOF", "Cuba": "CUP",
"Togo": "XOF", "Angola": "AOA", "Isle of Man": "IMP",
"Panama": "PAB", "Bahamas": "BSD", "Paraguay": "PYG",
"Sudan": "SDG", "Liberia": "LRD", "Bhutan": "BTN",
"Congo": "CDF", "Côte d'Ivoire": "XOF", "Barbados": "BBD",
"Namibia": "NAD", "Somalia": "SOS", "Sierra Leone": "SLL",
"Mozambique": "MZN", "Lesotho": "LSL", "Chad": "XAF",
"North Korea": "KPW", "Antigua and Barbuda": "XCD", "Papua New Guinea": "PGK",
"Palau": "USD", "Guinea": "GNF", "Haiti": "HTG",
"Gabon": "XAF", "Mauritania": "MRU", "San Marino": "EUR",
"Guyana": "GYD", "Saint Lucia": "XCD", "Burkina Faso": "XOF",
"Brunei": "BND", "Gambia": "GMD", "Monaco": "MCO",
"Djibouti": "DJF", "Seychelles": "SCR", "Solomon Islands": "SBD",
"Saint Kitts and Nevis": "KN"} 
```

## Observations:

- Many countries are missing values for currency
- Some of the currencies are incorrectly mentioned

- Whenever the country column is missing, so is the corresponding value of the currency column
- It doesn't make sense to have missing values for this column when the country is specified

## Steps:

- The accurate currencies will be mapped manually based on the values of the `Country` feature

## CompTotal

```
In [100...]: col_description("CompTotal")
```

### Description:

What is your current total compensation (salary, bonuses, and perks, before taxes and deductions)? Please enter a whole number in the box below, without any punctuation. If you are paid hourly, please estimate an equivalent weekly, monthly, or yearly salary. If you prefer not to answer, please leave the box empty.

```
In [101...]: df.CompTotal
```

```
Out[101]:
```

0	NaN
1	NaN
2	32000.0
3	60000.0
4	NaN
...	
73263	60000.0
73264	107000.0
73265	NaN
73266	58500.0
73267	NaN

Name: CompTotal, Length: 73268, dtype: float64

```
In [102...]: with pd.option_context("display.float_format", "{:.3f}".format):
    display(df.CompTotal.describe(percentiles=[0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95,
```

count	38,422.000
mean	23,424,340,221,747,959,297,504,942,608,856,691,...
std	4,591,478,187,815,219,735,465,859,539,580,935,5...
min	0.000
5%	2,200.000
10%	4,000.000
25%	30,000.000
50%	77,500.000
75%	154,000.000
90%	500,000.000
95%	1,800,000.000
99%	30,000,000.000
max	900,000,000,000,000,060,934,480,090,350,342,481...

Name: CompTotal, dtype: float64

```
In [103...]: (df
    .dropna(subset=["CompTotal"])
    .sort_values(by="CompTotal", ascending=False))
```



Out[103]:

	RespondId	MainBranch	Employment	RemoteWork	CodingActivities	EdLevel
35786	35787	I am a developer by profession	Employed, full-time; Employed, part-time	Hybrid (some remote, some in-person)	I don't code outside of work	Secondary school (e.g. American high school, G...
3068	3069	I am a developer by profession	Employed, full-time	Hybrid (some remote, some in-person)	Hobby;Contribute to open-source projects;Boots...	Bachelor's degree (B.A., B.S., B.Eng., etc.)
70597	70598	I am a developer by profession	Independent contractor, freelancer, or self-em...	Fully remote	Hobby;Contribute to open-source projects	Bachelor's degree (B.A., B.S., B.Eng., etc.)
17567	17568	I am a developer by profession	Employed, full-time	Hybrid (some remote, some in-person)	Hobby;Contribute to open-source projects;Freel...	Secondary school (e.g. American high school, G...
19244	19245	I am not primarily a developer, but I write co...	Independent contractor, freelancer, or self-em...	Fully remote	Hobby;Contribute to open-source projects;Freel...	Master's degree (M.A., M.S., M.Eng., MBA, etc.)
...	...	...	...	...	...	...
3891	3892	I am a developer by profession	Independent contractor, freelancer, or self-em...	Fully remote	Contribute to open-source projects;Freelance/c...	Master's degree (M.A., M.S., M.Eng., MBA, etc.)
8085	8086	I am a developer by profession	Independent contractor, freelancer, or self-em...	Hybrid (some remote, some in-person)	Hobby;Contribute to open-source projects;Boots...	Some college/university study without earning ...
47360	47361	I am a developer by profession	Independent contractor, freelancer, or self-em...	Fully remote	Bootstrapping a business	Bachelor's degree (B.A., B.S., B.Eng., etc.)
51860	51861	I am a developer by profession	Independent contractor, freelancer, or self-em...	Fully remote	Contribute to open-source projects;Freelance/c...	Some college/university study without earning ...

ResponseId	MainBranch	Employment	RemoteWork	CodingActivities	EdLevel
73012	73013	I am a developer by profession	Independent contractor, freelancer, or self-em...	Fully remote	Hobby;Contribute to open-source projects;Boots... Master's degree (M.A., M.S., M.Eng., MBA, etc.)

## Checking who are the people having 0 compensation

In [104...]

```
(df
 .query("CompTotal == 0")
 .Employment
 .pipe(lambda ser: col_strip_nested(ser.name)))
```

### Unique Values:

Out[104]:

```
[('Employed, full-time', 49199),
 ('Student, full-time', 10932),
 ('Independent contractor, freelancer, or self-employed', 10721),
 ('Employed, part-time', 4154),
 ('Student, part-time', 3722),
 ('Not employed, but looking for work', 3381),
 ('Not employed, and not looking for work', 1244),
 ('I prefer not to say', 611),
 ('Retired', 396)]
```

In [105...]

```
(df
 .query("CompTotal == 0")
 .Employment
 .pipe(lambda ser: col_value_counts(ser.name)))
```

Out[105]:

		count	pct
Employment			
	<b>Employed, full-time</b>	42962	0.599116
	<b>Student, full-time</b>	6756	0.094214
	<b>Independent contractor, freelancer, or self-employed</b>	4978	0.069419
	<b>Employed, full-time;Independent contractor, freelancer, or self-employed</b>	3486	0.048613
	<b>Not employed, but looking for work</b>	1831	0.025534
	...	...	...
	<b>Student, part-time;Independent contractor, freelancer, or self-employed;Retired</b>	1	0.000014
	<b>Employed, full-time;Independent contractor, freelancer, or self-employed;Employed, part-time;Not employed, and not looking for work</b>	1	0.000014
	<b>Employed, full-time;Student, full-time;Retired</b>	1	0.000014
	<b>Employed, part-time;Not employed, and not looking for work</b>	1	0.000014
	<b>Student, part-time;Independent contractor, freelancer, or self-employed;Employed, part-time;Retired</b>	1	0.000014

103 rows × 2 columns

In [106]: `((df.Employment.str.contains("Employed, full-time")) & (df.CompTotal == 0)).sum()`

Out[106]: 27

In [107]: `df.CompTotal.isna().sum()`

Out[107]: 34846

In [108]: `(df
 .CompTotal
 .where(lambda ser: ~(df.Employment.str.contains("Employed, full-time") & (ser == 0)),
 .isna().sum())`

Out[108]: 34873

## Observations

- The values are of the right type
- Some values are extremely large and way out of range
  - These could be because of the varying compensation frequencies
  - Or due to different currencies
- Some entries are 0; these are people who are:
  - Students
  - Unemployed
  - Retired
  - Self-employed

- Part-time employed
- There're also some `Full-time Employed` people having 0 compensation:
  - Maybe they didn't want to share their income

## Steps

- The column `CompFreq` must be analyzed for inaccuracies
- For full-time employees having 0 compensation:
  - These values will be replaced with `np.nan`

## CompFreq

```
In [109]: col_description("CompFreq")
```

### Description:

Is that compensation weekly, monthly, or yearly?

```
In [110]: df.CompFreq
```

```
Out[110]: 0           NaN
          1           NaN
          2      Yearly
          3     Monthly
          4           NaN
          ...
          73263    Yearly
          73264    Yearly
          73265      NaN
          73266    Yearly
          73267      NaN
Name: CompFreq, Length: 73268, dtype: object
```

```
In [111]: df.CompFreq.unique()
```

```
Out[111]: array([nan, 'Yearly', 'Monthly', 'Weekly'], dtype=object)
```

```
In [112]: col_value_counts("CompFreq")
```

	count	pct
<b>CompFreq</b>		
<b>Yearly</b>	23267	0.523737
<b>Monthly</b>	19983	0.449814
<b>Weekly</b>	1175	0.026449

## Observations

- There're 3 unique entries
- Values are valid
- No cleaning required

## Steps

- The type can be made `category` for less memory usage

## LanguageHaveWorkedWith

In [113...]: `df.LanguageHaveWorkedWith`

```
Out[113]: 0                      NaN
1          JavaScript;TypeScript
2          C#;C++;HTML/CSS;JavaScript;Python
3          C#;JavaScript;SQL;TypeScript
4          C#;HTML/CSS;JavaScript;SQL;Swift;TypeScript
...
73263    Bash/Shell;Dart;JavaScript;PHP;Python;SQL;Type...
73264    Bash/Shell;HTML/CSS;JavaScript;Python;SQL
73265    HTML/CSS;JavaScript;PHP;Python;SQL
73266          C#;Delphi;VBA
73267    C#;JavaScript;Lua;PowerShell;SQL;TypeScript
Name: LanguageHaveWorkedWith, Length: 73268, dtype: object
```

In [114...]: `col_value_counts("LanguageHaveWorkedWith")`

		count
LanguageHaveWorkedWith		
HTML/CSS;JavaScript;TypeScript	1250	0.0170
Python	962	0.0139
HTML/CSS;JavaScript	914	0.0128
HTML/CSS;JavaScript;PHP;SQL	745	0.0104
C#;HTML/CSS;JavaScript;SQL;TypeScript	570	0.0080
...	...	...
C#;Java;JavaScript;PHP;PowerShell;SQL;TypeScript	1	0.0000
C++;HTML/CSS;JavaScript;Python;Solidity;SQL	1	0.0000
Dart;Haskell;HTML/CSS;Java;R;SQL;TypeScript	1	0.0000
Bash/Shell;C;C#;Go;HTML/CSS;Java;JavaScript;Lua;Perl;PHP;Python;Ruby;Rust;SQL;TypeScript	1	0.0000
C#;JavaScript;Lua;PowerShell;SQL;TypeScript	1	0.0000

25068 rows × 2 columns

In [115...]: `col_strip_nested("LanguageHaveWorkedWith")`

Unique Values:

```
Out[115]: [('JavaScript', 46443),
 ('HTML/CSS', 39142),
 ('SQL', 35127),
 ('Python', 34155),
 ('TypeScript', 24752),
 ('Java', 23644),
 ('Bash/Shell', 20656),
 ('C#', 19883),
 ('C++', 16024),
 ('PHP', 14827),
 ('C', 13692),
 ('PowerShell', 8575),
 ('Go', 7922),
 ('Rust', 6625),
 ('Kotlin', 6507),
 ('Dart', 4648),
 ('Ruby', 4299),
 ('Assembly', 3887),
 ('Swift', 3489),
 ('R', 3308),
 ('VBA', 3185),
 ('MATLAB', 2913),
 ('Lua', 2867),
 ('Groovy', 2357),
 ('Delphi', 2311),
 ('Scala', 1837),
 ('Objective-C', 1698),
 ('Perl', 1644),
 ('Haskell', 1577),
 ('Elixir', 1528),
 ('Julia', 1084),
 ('Clojure', 1070),
 ('Solidity', 1031),
 ('LISP', 932),
 ('F#', 730),
 ('Fortran', 646),
 ('Erlang', 641),
 ('APL', 504),
 ('COBOL', 464),
 ('SAS', 435),
 ('OCaml', 422),
 ('Crystal', 340)]
```

```
In [116...]  
df  
.LanguageHaveWorkedWith  
.apply(lambda val: len(val.split(";")) if isinstance(val, str) else np.nan))
```

```
Out[116]: 0      NaN  
1      2.0  
2      5.0  
3      4.0  
4      6.0  
...  
73263    7.0  
73264    5.0  
73265    5.0  
73266    3.0  
73267    6.0  
Name: LanguageHaveWorkedWith, Length: 73268, dtype: float64
```

## Observations:

- There're 25,068 unique entries
- Entries are nested, separated by a semicolon
- The entries are clean and valid

## Steps:

- This variable requires feature engineering
- Can create a column, `Num of languages known`
- These steps should be kept in mind for performing feature engineering during exploratory analysis

## LanguageWantToWorkWith

```
In [117...]: df.LanguageWantToWorkWith
```

```
Out[117]: 0                      NaN
1                  Rust;TypeScript
2          C#;C++;HTML/CSS;JavaScript;TypeScript
3                  C#;SQL;TypeScript
4          C#;Elixir;F#;Go;JavaScript;Rust;TypeScript
...
73263    Bash/Shell;Go;JavaScript;Python;SQL;TypeScript
73264                  HTML/CSS;JavaScript;Python
73265          C#;HTML/CSS;JavaScript;PHP;Python;SQL
73266                  Delphi
73267          PowerShell;Rust;TypeScript
Name: LanguageWantToWorkWith, Length: 73268, dtype: object
```

- Same steps as `LanguageHaveWorkedWith`

## DatabaseHaveWorkedWith

```
In [118...]: df.DatabaseHaveWorkedWith
```

```
Out[118]: 0                      NaN
1                      NaN
2          Microsoft SQL Server
3          Microsoft SQL Server
4          Cloud Firestore;Elasticsearch;Microsoft SQL Se...
...
73263    Elasticsearch;MySQL;PostgreSQL;Redis
73264          Elasticsearch;MongoDB;Oracle;SQLite
73265    MariaDB;Microsoft SQL Server;MySQL;PostgreSQL;...
73266          Microsoft SQL Server;MongoDB;Oracle
73267          Microsoft SQL Server;Neo4j;Redis
Name: DatabaseHaveWorkedWith, Length: 73268, dtype: object
```

```
In [119...]: col_value_counts("DatabaseHaveWorkedWith")
```

Out[119]:

		count	pct
<b>DatabaseHaveWorkedWith</b>			
	<b>MySQL</b>	3563	0.059264
	<b>PostgreSQL</b>	3519	0.058532
	<b>Microsoft SQL Server</b>	3238	0.053858
	<b>SQLite</b>	2022	0.033632
	<b>MongoDB</b>	1655	0.027528
	...	...	...
	<b>Elasticsearch;IBM DB2;MongoDB;MySQL</b>	1	0.000017
	<b>Cloud Firestore;Microsoft SQL Server;MongoDB;MySQL;Neo4j</b>	1	0.000017
	<b>Cloud Firestore;MongoDB;Oracle;PostgreSQL;Redis;SQLite</b>	1	0.000017
	<b>MariaDB;Microsoft SQL Server;MySQL;Neo4j;Oracle;PostgreSQL;SQLite</b>	1	0.000017
	<b>Microsoft SQL Server;Neo4j;Redis</b>	1	0.000017

4223 rows × 2 columns

In [120...]

col\_strip\_nested("DatabaseHaveWorkedWith")

### Unique Values:

```
Out[120]: [('MySQL', 28520),
('PostgreSQL', 26538),
('SQLite', 19487),
('MongoDB', 17228),
('Microsoft SQL Server', 16355),
('Redis', 13471),
('MariaDB', 10912),
('Elasticsearch', 7430),
('Oracle', 6994),
('Firebase Realtime Database', 5309),
('DynamoDB', 5029),
('Cloud Firestore', 4535),
('Cassandra', 1617),
('Neo4j', 1291),
('IBM DB2', 1219),
('Couchbase', 807),
('CouchDB', 783)]
```

In [121...]

```
(df
 .DatabaseHaveWorkedWith
 .str.replace(" ", "_"))
```

```
Out[121]: 0                               NaN
           1                               NaN
           2             Microsoft_SQL_Server
           3             Microsoft_SQL_Server
           4 Cloud_Firebase;Elasticsearch;Microsoft_SQL_S...
              ...
           73263      Elasticsearch;MySQL;PostgreSQL;Redis
           73264      Elasticsearch;MongoDB;Oracle;SQLite
           73265 MariaDB;Microsoft_SQL_Server;MySQL;PostgreSQL;...
           73266      Microsoft_SQL_Server;MongoDB;Oracle
           73267      Microsoft_SQL_Server;Neo4j;Redis
Name: DatabaseHaveWorkedWith, Length: 73268, dtype: object
```

## DatabaseWantToWorkWith

```
In [122... col_strip_nested("DatabaseWantToWorkWith")]
```

### Unique Values:

```
Out[122]: [('PostgreSQL', 25212),
            ('MongoDB', 17297),
            ('MySQL', 16271),
            ('Redis', 16211),
            ('SQLite', 14085),
            ('Microsoft SQL Server', 9867),
            ('Elasticsearch', 8533),
            ('MariaDB', 7181),
            ('Firebase Realtime Database', 5312),
            ('DynamoDB', 5309),
            ('Cloud Firestore', 3873),
            ('Oracle', 3511),
            ('Cassandra', 3178),
            ('Neo4j', 2354),
            ('CouchDB', 1154),
            ('Couchbase', 955),
            ('IBM DB2', 519)]
```

## PlatformHaveWorkedWith

```
In [123... col_value_counts("PlatformHaveWorkedWith")]
```

Out[123]:

		count	pct
PlatformHaveWorkedWith			
	<b>AWS</b>	8719	0.174645
	<b>Microsoft Azure</b>	4989	0.099932
	<b>Google Cloud</b>	2306	0.046190
	<b>AWS;Microsoft Azure</b>	2039	0.040842
	<b>Heroku</b>	2003	0.040121
	...	...	...
<b>AWS;Colocation;Firebase;Google Cloud;IBM Cloud or Watson;OpenStack;VMware</b>	1	0.000020	
<b>Firebase;Google Cloud;Heroku;Linode;Managed Hosting;Microsoft Azure</b>	1	0.000020	
<b>AWS;DigitalOcean;Firebase;Google Cloud;Heroku;IBM Cloud or Watson;Managed Hosting;Microsoft Azure;Oracle Cloud Infrastructure;VMware</b>	1	0.000020	
<b>AWS;DigitalOcean;Firebase;Google Cloud;IBM Cloud or Watson;Oracle Cloud Infrastructure;OpenStack;VMware</b>	1	0.000020	
<b>Colocation;DigitalOcean;Heroku;Linode;Oracle Cloud Infrastructure;OVH</b>	1	0.000020	

1500 rows × 2 columns

In [124...]

col\_strip\_nested("PlatformHaveWorkedWith")

### Unique Values:

Out[124]:

```
[('AWS', 25939),
 ('Microsoft Azure', 14604),
 ('Google Cloud', 13634),
 ('Firebase', 10751),
 ('Heroku', 10160),
 ('DigitalOcean', 7953),
 ('VMware', 4429),
 ('Managed Hosting', 2927),
 ('Linode', 1994),
 ('OVH', 1913),
 ('Oracle Cloud Infrastructure', 1110),
 ('OpenStack', 1029),
 ('IBM Cloud or Watson', 853),
 ('Colocation', 642)]
```

In [125...]

```
(df
 .PlatformHaveWorkedWith
 .fillna("unanswered")
 .str.replace("IBM Cloud or Watson", "watson")
 .str.replace("OpenStack", "open_stack")
 .str.replace("DigitalOcean", "Digital Ocean")
 .str.lower()
 .str.replace(" ", "_")
 .unique())
```

```
Out[125]: array(['unanswered', 'firebase;microsoft_azure',
   'aws;google_cloud;heroku', ...,
   'aws;digital_ocean;google_cloud;heroku;open_stack',
   'watson;linode;vmware',
   'colocation;digital_ocean;heroku;linode;oracle_cloud_infrastructure;ovh'],
  dtype=object)
```

## PlatformWantToWorkWith

```
In [126... col_strip_nested("PlatformWantToWorkWith")]
```

### Unique Values:

```
Out[126]: [('AWS', 23701),
 ('Google Cloud', 13394),
 ('Microsoft Azure', 12785),
 ('Firebase', 8643),
 ('DigitalOcean', 6983),
 ('Heroku', 5951),
 ('VMware', 2495),
 ('Linode', 2259),
 ('Managed Hosting', 1887),
 ('OpenStack', 1385),
 ('OVH', 1233),
 ('Oracle Cloud Infrastructure', 1144),
 ('IBM Cloud or Watson', 888),
 ('Colocation', 537)]
```

## WebframeHaveWorkedWith

```
In [127... df.WebframeHaveWorkedWith]
```

```
Out[127]: 0                               NaN
1                               NaN
2                               Angular.js
3                               ASP.NET;ASP.NET Core
4      Angular;ASP.NET;ASP.NET Core ;jQuery;Node.js
...
73263                         Express;FastAPI;Node.js
73264                         FastAPI;Flask;React.js
73265                         ASP.NET;React.js
73266                         NaN
73267      ASP.NET Core ;Blazor;Node.js;React.js;Svelte
Name: WebframeHaveWorkedWith, Length: 73268, dtype: object
```

```
In [128... col_strip_nested("WebframeHaveWorkedWith")]
```

### Unique Values:

```
Out[128]: [('Node.js', 25733),
 ('React.js', 23277),
 ('jQuery', 15602),
 ('Express', 12557),
 ('Angular', 11138),
 ('Vue.js', 10278),
 ('ASP.NET Core ', 10155),
 ('ASP.NET', 8139),
 ('Django', 8002),
 ('Flask', 7994),
 ('Next.js', 7386),
 ('Laravel', 5159),
 ('Angular.js', 4912),
 ('FastAPI', 3289),
 ('Ruby on Rails', 3182),
 ('Svelte', 2500),
 ('Blazor', 2438),
 ('Nuxt.js', 2089),
 ('Symfony', 1955),
 ('Gatsby', 1889),
 ('Drupal', 1211),
 ('Phoenix', 1164),
 ('Fastify', 1008),
 ('Deno', 925),
 ('Play Framework', 450)]
```

In [129...]

```
(df
 .WebframeHaveWorkedWith
 .fillna("unanswered")
 .str.lower()
 .str.replace(" ", "_")
 .str.replace(".", "_", regex=True))
```

```
Out[129]: 0                      unanswered
1                      unanswered
2                      angular_js
3          asp_net;asp_net_core_
4      angular;asp_net;asp_net_core_;jquery;node_js
...
73263                  express;fastapi;node_js
73264                  fastapi;flask;react_js
73265                  asp_net;react_js
73266                  unanswered
73267  asp_net_core_;blazor;node_js;react_js;svelte
Name: WebframeHaveWorkedWith, Length: 73268, dtype: object
```

## WebframeWantToWorkWith

In [130...]

```
col_strip_nested("WebframeWantToWorkWith")
```

**Unique Values:**

```
Out[130]: [('React.js', 21910),
 ('Node.js', 20948),
 ('Vue.js', 12666),
 ('Next.js', 10348),
 ('Express', 9333),
 ('ASP.NET Core ', 8894),
 ('Angular', 8402),
 ('Django', 7200),
 ('Svelte', 6741),
 ('jQuery', 5937),
 ('Flask', 5437),
 ('FastAPI', 4571),
 ('Blazor', 3880),
 ('ASP.NET', 3848),
 ('Laravel', 3808),
 ('Nuxt.js', 3805),
 ('Deno', 3717),
 ('Ruby on Rails', 3272),
 ('Angular.js', 2773),
 ('Phoenix', 1862),
 ('Gatsby', 1569),
 ('Fastify', 1452),
 ('Symfony', 1433),
 ('Drupal', 600),
 ('Play Framework', 371)]
```

## MiscTechHaveWorkedWith

In [131... df.MiscTechHaveWorkedWith

```
Out[131]: 0                               NaN
1                               NaN
2                               Pandas
3                               .NET
4                               .NET
...
73263                         Flutter
73264 Keras;NumPy;Pandas;Scikit-learn;TensorFlow;Tor...
73265                               .NET;Pandas;React Native
73266                               NaN
73267 Apache Kafka;Apache Spark
Name: MiscTechHaveWorkedWith, Length: 73268, dtype: object
```

In [132... col\_strip\_nested("MiscTechHaveWorkedWith")

**Unique Values:**

```
Out[132]: [(''.NET', 15850),
 ('NumPy', 13144),
 ('Pandas', 11506),
 ('Spring', 7399),
 ('TensorFlow', 5942),
 ('Flutter', 5799),
 ('Scikit-learn', 5776),
 ('React Native', 5765),
 ('Apache Kafka', 4748),
 ('Electron', 4390),
 ('Torch/PyTorch', 3952),
 ('Qt', 3906),
 ('Keras', 3333),
 ('Ionic', 2417),
 ('Xamarin', 2388),
 ('Apache Spark', 2298),
 ('Cordova', 1903),
 ('Hadoop', 1581),
 ('GTK', 1380),
 ('Capacitor', 1142),
 ('Tidyverse', 996),
 ('Hugging Face Transformers', 925),
 ('Uno Platform', 334)]
```

```
In [133...]: df
    .MiscTechHaveWorkedWith
    .fillna("unanswered")
    .str.replace("Torch/PyTorch", "pytorch")
    .str.replace(".NET", "dot_net", regex=True)
    .str.lower()
    .str.replace(" ", "_")
    .str.replace("-", "_"))
```

```
Out[133]: 0                      unanswered
1                      unanswered
2                      pandas
3                      dot_net
4                      dot_net
...
73263                  flutter
73264      keras;numpy;pandas;scikit_learn;tensorflow;pyt...
73265                      dot_net;pandas;react_native
73266                      unanswered
73267                      apache_kafka;apache_spark
Name: MiscTechHaveWorkedWith, Length: 73268, dtype: object
```

## MiscTechWantToWorkWith

```
In [134...]: col_strip_nested("MiscTechWantToWorkWith")
```

**Unique Values:**

```
Out[134]: [('.NET', 11824),
 ('NumPy', 10248),
 ('Pandas', 9070),
 ('TensorFlow', 8836),
 ('Flutter', 7980),
 ('React Native', 6728),
 ('Apache Kafka', 6153),
 ('Torch/PyTorch', 6073),
 ('Spring', 5472),
 ('Scikit-learn', 5175),
 ('Electron', 4349),
 ('Keras', 3493),
 ('Apache Spark', 3155),
 ('Qt', 2895),
 ('Xamarin', 2259),
 ('Hadoop', 1993),
 ('Ionic', 1679),
 ('Hugging Face Transformers', 1462),
 ('GTK', 1314),
 ('Capacitor', 1073),
 ('Tidyverse', 828),
 ('Cordova', 745),
 ('Uno Platform', 578)]
```

## ToolsTechHaveWorkedWith

```
In [135...]: df.ToolsTechHaveWorkedWith
```

0	NaN
1	NaN
2	NaN
3	NaN
4	npm
...	
73263	Docker;Homebrew;Kubernetes;npm
73264	NaN
73265	npm
73266	NaN
73267	Docker;Kubernetes;npm;Pulumi;Terraform

Name: ToolsTechHaveWorkedWith, Length: 73268, dtype: object

```
In [136...]: col_strip_nested("ToolsTechHaveWorkedWith")
```

### Unique Values:

```
Out[136]: [('npm', 35778),
 ('Docker', 34981),
 ('Yarn', 15175),
 ('Homebrew', 14420),
 ('Kubernetes', 12624),
 ('Terraform', 6160),
 ('Unity 3D', 5840),
 ('Ansible', 5210),
 ('Unreal Engine', 2180),
 ('Puppet', 1025),
 ('Chef', 828),
 ('Pulumi', 461),
 ('Flow', 444)]
```

In [137...]

```
(df
    .ToolsTechHaveWorkedWith
    .fillna("unanswered")
    .str.lower()
    .str.replace(" ", "_"))
```

Out[137]:

0	unanswered
1	unanswered
2	unanswered
3	unanswered
4	npm
...	
73263	docker;homebrew;kubernetes;npm
73264	unanswered
73265	npm
73266	unanswered
73267	docker;kubernetes;npm;pulumi;terraform

Name: ToolsTechHaveWorkedWith, Length: 73268, dtype: object

## ToolsTechWantToWorkWith

In [138...]

```
col_strip_nested("ToolsTechWantToWorkWith")
```

### Unique Values:

Out[138]:

```
[('Docker', 32642),
 ('npm', 23147),
 ('Kubernetes', 18647),
 ('Yarn', 11154),
 ('Homebrew', 10424),
 ('Terraform', 8130),
 ('Unity 3D', 6293),
 ('Ansible', 5540),
 ('Unreal Engine', 5061),
 ('Puppet', 1056),
 ('Pulumi', 1050),
 ('Chef', 911),
 ('Flow', 453)]
```

## NEWCollabToolsHaveWorkedWith

In [139...]

```
df.NEWCollabToolsHaveWorkedWith
```

Out[139]:

0	NaN
1	NaN
2	Notepad++;Visual Studio
3	Notepad++;Visual Studio;Visual Studio Code
4	Notepad++;Visual Studio;Visual Studio Code;Xcode
...	
73263	IPython/Jupyter;Sublime Text;Vim;Visual Studio...
73264	IPython/Jupyter;Notepad++;Spyder;Vim;Visual St...
73265	Spyder;Visual Studio;Visual Studio Code
73266	RAD Studio (Delphi, C++ Builder);Visual Studio
73267	Visual Studio;Visual Studio Code

Name: NEWCollabToolsHaveWorkedWith, Length: 73268, dtype: object

In [140...]

```
col_strip_nested("NEWCollabToolsHaveWorkedWith")
```

## Unique Values:

```
Out[140]: [('Visual Studio Code', 52523),
 ('Visual Studio', 22673),
 ('IntelliJ', 19723),
 ('Notepad++', 19543),
 ('Vim', 16458),
 ('Android Studio', 13963),
 ('PyCharm', 12158),
 ('Sublime Text', 11698),
 ('Eclipse', 8866),
 ('IPython/Jupyter', 8188),
 ('Xcode', 7425),
 ('Atom', 6595),
 ('Nano', 6530),
 ('Webstorm', 5602),
 ('PhpStorm', 4790),
 ('Neovim', 4759),
 ('NetBeans', 3695),
 ('CLion', 3543),
 ('Rider', 3480),
 ('Emacs', 3178),
 ('RStudio', 2387),
 ('GoLand', 2345),
 ('RAD Studio (Delphi, C++ Builder)', 1894),
 ('Qt Creator', 1892),
 ('Spyder', 1637),
 ('RubyMine', 975),
 ('TextMate', 516)]
```

```
In [141...]: df
    .NEWCollabToolsHaveWorkedWith
    .fillna("unanswered")
    .str.replace("RAD Studio \Delphi, C\+\+ Builder\)", "RAD Studio", regex=True)
    .str.replace("IPython/Jupyter", "Jupyter", regex=True)
    .str.lower()
    .str.replace(" ", "_"))
```

```
Out[141]: 0                      unanswered
1                      unanswered
2                  notepad++;visual_studio
3          notepad++;visual_studio;visual_studio_code
4      notepad++;visual_studio;visual_studio_code;xcode
...
73263      jupyter;sublime_text;vim;visual_studio_code
73264      jupyter;notepad++;spyder;vim;visual_studio_code
73265          spyder;visual_studio;visual_studio_code
73266          rad_studio;visual_studio
73267          visual_studio;visual_studio_code
Name: NEWCollabToolsHaveWorkedWith, Length: 73268, dtype: object
```

## NEWCollabToolsWantToWorkWith

```
In [142...]: col_strip_nested("NEWCollabToolsWantToWorkWith")
```

## Unique Values:

```
Out[142]: [('Visual Studio Code', 44282),  
 ('IntelliJ', 15040),  
 ('Visual Studio', 14958),  
 ('Vim', 12862),  
 ('Notepad++', 11734),  
 ('Android Studio', 9096),  
 ('PyCharm', 8882),  
 ('Sublime Text', 6370),  
 ('IPython/Jupyter', 6224),  
 ('Xcode', 5780),  
 ('Neovim', 4977),  
 ('Webstorm', 4638),  
 ('Nano', 3897),  
 ('Rider', 3537),  
 ('PhpStorm', 3265),  
 ('Eclipse', 2942),  
 ('Emacs', 2885),  
 ('CLion', 2828),  
 ('GoLand', 2725),  
 ('Atom', 2508),  
 ('RAD Studio (Delphi, C++ Builder)', 1566),  
 ('RStudio', 1449),  
 ('Qt Creator', 1226),  
 ('NetBeans', 1080),  
 ('RubyMine', 850),  
 ('Spyder', 837),  
 ('TextMate', 299)]
```

- The following columns have the exact same workflow as LanguageHaveWorkedWith
  - LanguageWantToWorkWith
  - DatabaseHaveWorkedWith
  - DatabaseWantToWorkWith
  - PlatformHaveWorkedWith
  - PlatformWantToWorkWith
  - WebframeHaveWorkedWith
  - WebframeWantToWorkWith
  - MiscTechHaveWorkedWith
  - MiscTechWantToWorkWith
  - ToolsTechHaveWorkedWith
  - ToolsTechWantToWorkWith
  - NEWCollabToolsHaveWorkedWith
  - NEWCollabToolsWantToWorkWith

## OpSysProfessional use

```
In [143...]
```

```
df["OpSysProfessional use"]
```

```
Out[143]: 0          NaN
           1          macOS
           2        Windows
           3        Windows
           4        Windows
           ...
           73263      macOS
           73264  Linux-based;Windows
           73265      Windows
           73266      Windows
           73267  Linux-based;Windows
Name: OpSysProfessional use, Length: 73268, dtype: object
```

```
In [144... df["OpSysProfessional use"].unique()
```

```
Out[144]: array([nan, 'macOS', 'Windows', 'Linux-based;macOS',
   'Windows;Windows Subsystem for Linux (WSL)',
   'Linux-based;macOS;Windows', 'Linux-based;Windows',
   'Windows Subsystem for Linux (WSL)', 'macOS;Windows',
   'Linux-based', 'Linux-based;Windows Subsystem for Linux (WSL)',
   'Linux-based;Windows;Windows Subsystem for Linux (WSL)',
   'Linux-based;macOS;Windows;Windows Subsystem for Linux (WSL)',
   'Other (please specify):',
   'BSD;Linux-based;Windows;Windows Subsystem for Linux (WSL)',
   'macOS;Other (please specify):',
   'macOS;Windows;Windows Subsystem for Linux (WSL)',
   'macOS;Windows Subsystem for Linux (WSL)',
   'Windows;Other (please specify):', 'BSD;Linux-based;macOS;Windows',
   'BSD;Linux-based',
   'BSD;Linux-based;Windows;Other (please specify):',
   'Linux-based;macOS;Windows Subsystem for Linux (WSL)',
   'BSD;Linux-based;Windows Subsystem for Linux (WSL)',
   'BSD;Linux-based;macOS',
   'Linux-based;Windows;Windows Subsystem for Linux (WSL);Other (please specif
y):',
   'BSD', 'Linux-based;Other (please specify):', 'BSD;macOS',
   'BSD;Windows',
   'BSD;Linux-based;macOS;Windows;Windows Subsystem for Linux (WSL)',
   'BSD;macOS;Windows;Windows Subsystem for Linux (WSL)',
   'BSD;Linux-based;macOS;Windows Subsystem for Linux (WSL)',
   'BSD;Linux-based;Windows',
   'Linux-based;Windows;Other (please specify):',
   'BSD;Windows Subsystem for Linux (WSL)', 'BSD;macOS;Windows',
   'Linux-based;macOS;Windows;Other (please specify):',
   'BSD;Linux-based;Other (please specify):',
   'Linux-based;macOS;Windows;Windows Subsystem for Linux (WSL);Other (please spe
cify):',
   'Linux-based;macOS;Other (please specify):',
   'Linux-based;Windows Subsystem for Linux (WSL);Other (please specify):',
   'Windows Subsystem for Linux (WSL);Other (please specify):',
   'BSD;Linux-based;macOS;Windows;Windows Subsystem for Linux (WSL);Other (please
specify):',
   'Windows;Windows Subsystem for Linux (WSL);Other (please specify):',
   'BSD;Windows;Windows Subsystem for Linux (WSL)',
   'BSD;Linux-based;Windows;Windows Subsystem for Linux (WSL);Other (please speci
fy):',
   'Linux-based;macOS;Windows Subsystem for Linux (WSL);Other (please specify):',
   'macOS;Windows;Other (please specify):',
   'BSD;Other (please specify):',
   'BSD;macOS;Windows Subsystem for Linux (WSL)',
   'BSD;Linux-based;macOS;Windows;Other (please specify):',
   'macOS;Windows;Windows Subsystem for Linux (WSL);Other (please specify):',
   'BSD;Linux-based;macOS;Other (please specify):',
   'BSD;Windows;Other (please specify):',
   'BSD;Linux-based;macOS;Windows Subsystem for Linux (WSL);Other (please specif
y):'],
  dtype=object)
```

In [145...]: col\_value\_counts("OpSysProfessional use")#.shape

Out[145]:

		count	pct
OpSysProfessional use			
	<b>Windows</b>	16645	0.254110
	<b>macOS</b>	12541	0.191457
	<b>Linux-based</b>	10934	0.166924
	<b>Linux-based;Windows</b>	5637	0.086057
	<b>Linux-based;macOS</b>	4906	0.074897
	<b>Windows;Windows Subsystem for Linux (WSL)</b>	3800	0.058013
	<b>Linux-based;Windows;Windows Subsystem for Linux (WSL)</b>	3003	0.045845
	<b>macOS;Windows</b>	2327	0.035525
	<b>Linux-based;macOS;Windows</b>	1551	0.023678
	<b>Linux-based;macOS;Windows;Windows Subsystem for Linux (WSL)</b>	897	0.013694
	<b>Linux-based;Windows Subsystem for Linux (WSL)</b>	712	0.010870
	<b>Windows Subsystem for Linux (WSL)</b>	586	0.008946
	<b>macOS;Windows;Windows Subsystem for Linux (WSL)</b>	557	0.008503
	<b>macOS;Windows Subsystem for Linux (WSL)</b>	219	0.003343
	<b>Linux-based;macOS;Windows Subsystem for Linux (WSL)</b>	206	0.003145
	<b>BSD;Linux-based</b>	113	0.001725
	<b>BSD;Linux-based;macOS</b>	102	0.001557
	<b>BSD</b>	92	0.001405
	<b>BSD;Linux-based;Windows</b>	85	0.001298
	<b>BSD;Linux-based;macOS;Windows</b>	67	0.001023
	<b>BSD;Linux-based;macOS;Windows;Other (please specify):</b>	55	0.000840
	<b>Linux-based;Other (please specify):</b>	52	0.000794
	<b>BSD;Linux-based;Windows;Windows Subsystem for Linux (WSL)</b>	51	0.000779
	<b>BSD;Linux-based;Windows;Windows Subsystem for Linux (WSL);Other (please specify):</b>	46	0.000702
	<b>Windows;Other (please specify):</b>	42	0.000641
	<b>Windows;Windows;Other (please specify):</b>	35	0.000534
	<b>BSD;macOS</b>	35	0.000534
	<b>BSD;Windows</b>	33	0.000504
	<b>BSD;Linux-based;macOS;Windows;Windows Subsystem for Linux (WSL);Other (please specify):</b>	18	0.000275
	<b>BSD;Linux-based;Windows;Windows Subsystem for Linux (WSL);Other (please specify):</b>	17	0.000260
	<b>BSD;Windows;Windows Subsystem for Linux (WSL)</b>	15	0.000229
	<b>BSD;Linux-based;macOS;Windows Subsystem for Linux (WSL)</b>	15	0.000229

		count	pct
OpSysProfessional use			
<b>BSD;Linux-based;Windows Subsystem for Linux (WSL)</b>	13	0.000198	
<b>Linux-based;macOS;Windows;Other (please specify):</b>	11	0.000168	
<b>BSD;Windows Subsystem for Linux (WSL)</b>	10	0.000153	
<b>Windows;Windows Subsystem for Linux (WSL);Other (please specify):</b>	8	0.000122	
<b>BSD;Linux-based;Other (please specify):</b>	8	0.000122	
<b>Linux-based;macOS;Windows;Windows Subsystem for Linux (WSL);Other (please specify):</b>	7	0.000107	
<b>Linux-based;macOS;Other (please specify):</b>	7	0.000107	
<b>macOS;Windows;Other (please specify):</b>	6	0.000092	
<b>BSD;macOS;Windows</b>	6	0.000092	
<b>BSD;macOS;Windows;Windows Subsystem for Linux (WSL)</b>	5	0.000076	
<b>Linux-based;Windows Subsystem for Linux (WSL);Other (please specify):</b>	3	0.000046	
<b>BSD;Linux-based;Windows;Other (please specify):</b>	3	0.000046	
<b>BSD;Linux-based;macOS;Windows;Other (please specify):</b>	3	0.000046	
<b>BSD;Other (please specify):</b>	2	0.000031	
<b>Windows Subsystem for Linux (WSL);Other (please specify):</b>	2	0.000031	
<b>BSD;Linux-based;Windows;Windows Subsystem for Linux (WSL);Other (please specify):</b>	2	0.000031	
<b>Linux-based;macOS;Windows Subsystem for Linux (WSL);Other (please specify):</b>	1	0.000015	
<b>BSD;macOS;Windows Subsystem for Linux (WSL)</b>	1	0.000015	
<b>macOS;Windows;Windows Subsystem for Linux (WSL);Other (please specify):</b>	1	0.000015	
<b>BSD;Linux-based;macOS;Other (please specify):</b>	1	0.000015	
<b>BSD;Windows;Other (please specify):</b>	1	0.000015	
<b>BSD;Linux-based;macOS;Windows Subsystem for Linux (WSL);Other (please specify):</b>	1	0.000015	

In [146...]: col\_strip\_nested("OpSysProfessional use")

## Unique Values:

Out[146]:

```
[('Windows', 34905),
 ('Linux-based', 28523),
 ('macOS', 23578),
 ('Windows Subsystem for Linux (WSL)', 10252),
 ('BSD', 737),
 ('Other (please specify):', 284)]
```

In [147...]: df["OpSysProfessional use"]
 .str.replace("Linux-based", "Linux", regex=True)

```
.str.replace("Windows Subsystem for Linux \(\WSL\)", "WSL", regex=True)
.str.replace("Other \(please specify\)\:", "Other", regex=True)
.str.lower()
.str.replace("macos", "mac_os"))
```

Out[147]:

0	NaN
1	mac_os
2	windows
3	windows
4	windows
...	
73263	mac_os
73264	linux;windows
73265	windows
73266	windows
73267	linux;windows

Name: OpSysProfessional use, Length: 73268, dtype: object

## Observations:

- There're 55 unique entries
- Entries are nested, separated by a semicolon

## Steps:

- Multiple binary columns can be created based on the languages known
- Can create a column, Num of operating systems used
- These steps should be kept in mind for performing feature engineering during exploratory analysis

## OpSysPersonal use

In [148...]

```
df["OpSysPersonal use"]
```

Out[148]:

0	NaN
1	Windows Subsystem for Linux (\WSL)
2	Windows
3	Windows
4	macOS;Windows
...	
73263	Linux-based;macOS
73264	Linux-based;Windows
73265	Windows
73266	Windows
73267	Linux-based;Windows

Name: OpSysPersonal use, Length: 73268, dtype: object

In [149...]

```
col_strip_nested("OpSysPersonal use")
```

## Unique Values:

```
Out[149]: [('Windows', 44567),
 ('Linux-based', 28765),
 ('macOS', 22217),
 ('Windows Subsystem for Linux (WSL)', 10724),
 ('BSD', 1054),
 ('Other (please specify):', 349)]
```

- Same steps as OpSysProfessional use

## VersionControlSystem

```
In [150... col_description("VersionControlSystem")
```

### Description:

What are the primary **version control systems** you use? Select all that apply.

```
In [151... df.VersionControlSystem
```

```
Out[151]: 0           NaN
 1           Git
 2           Git
 3           Git
 4       Git;Other (please specify):
 ...
 73263        Git
 73264        Git
 73265        Git
 73266        SVN
 73267        Git
Name: VersionControlSystem, Length: 73268, dtype: object
```

```
In [152... df.VersionControlSystem.unique()
```

```
Out[152]: array([nan, 'Git', 'Git;Other (please specify):', 'Mercurial;SVN',
   "I don't use one", 'Git;SVN', 'SVN', 'Other (please specify):',
   'Git;Mercurial', 'Git;Other (please specify):;SVN', 'Mercurial',
   'Git;Other (please specify):;Mercurial', 'Git;Mercurial;SVN',
   'Other (please specify):;SVN',
   'Git;Other (please specify):;Mercurial;SVN',
   'Other (please specify):;Mercurial',
   'Other (please specify):;Mercurial;SVN'], dtype=object)
```

```
In [153... col_value_counts("VersionControlSystem").shape
```

Out[153]:

		count	pct
<b>VersionControlSystem</b>			
	<b>Git</b>	62055	0.869373
	<b>I don't use one</b>	3080	0.043150
	<b>Git;SVN</b>	2858	0.040040
	<b>Git;Other (please specify):</b>	1356	0.018997
	<b>SVN</b>	590	0.008266
	<b>Other (please specify):</b>	523	0.007327
	<b>Git;Mercurial</b>	498	0.006977
	<b>Mercurial</b>	134	0.001877
	<b>Git;Mercurial;SVN</b>	104	0.001457
	<b>Git;Other (please specify)::SVN</b>	83	0.001163
	<b>Git;Other (please specify)::Mercurial</b>	28	0.000392
	<b>Other (please specify)::SVN</b>	26	0.000364
	<b>Git;Other (please specify)::Mercurial;SVN</b>	24	0.000336
	<b>Mercurial;SVN</b>	13	0.000182
	<b>Other (please specify)::Mercurial</b>	5	0.000070
	<b>Other (please specify)::Mercurial;SVN</b>	2	0.000028

In [154...]

col\_strip\_nested("VersionControlSystem")

## Unique Values:

Out[154]:

```
[('Git', 67006),
 ('SVN', 3700),
 ("I don't use one", 3080),
 ('Other (please specify):', 2047),
 ('Mercurial', 808)]
```

In [155...]

```
(df
 .VersionControlSystem
 .fillna("unanswered")
 .str.replace("Other \pleas e specify\:):", "other", regex=True)
 .str.replace("I don't use one", "none")
 .str.lower()
 .unique())
```

Out[155]:

```
array(['unanswered', 'git', 'git;other', 'mercurial;svn', 'none',
       'git;svn', 'svn', 'other', 'git;mercurial', 'git;other;svn',
       'mercurial', 'git;other;mercurial', 'git;mercurial;svn',
       'other;svn', 'git;other;mercurial;svn', 'other;mercurial',
       'other;mercurial;svn'], dtype=object)
```

## Observations

- There're 16 unique entries
- Git is most widely used version control system

## Steps

- Will create a binary column `use_git`

## VCInteraction

```
In [156...]: col_description("VCInteraction")
```

### Description:

How do you interact with your version control system? Select all that apply.

```
In [157...]: df.VCInteraction
```

```
Out[157]: 0                      NaN  
1                      NaN  
2                      Code editor  
3      Code editor;Command-line;Version control hosti...  
4                      Code editor  
...  
73263                  Code editor;Command-line  
73264                  Code editor;Command-line  
73265      Code editor;Command-line;Version control hosti...  
73266          Dedicated version control GUI application  
73267      Code editor;Command-line;Version control hosti...  
Name: VCInteraction, Length: 73268, dtype: object
```

```
In [158...]: col_value_counts("VCInteraction")
```

Out[158]:

VCInteraction	count	pct
<b>Command-line</b>	17602	0.258260
<b>Code editor;Command-line</b>	16502	0.242121
<b>Code editor;Command-line;Version control hosting service web GUI</b>	7288	0.106931
<b>Code editor;Command-line;Version control hosting service web GUI;Dedicated version control GUI application</b>	4420	0.064851
<b>Command-line;Dedicated version control GUI application</b>	3841	0.056356
<b>Code editor</b>	3762	0.055197
<b>Dedicated version control GUI application</b>	3373	0.049489
<b>Command-line;Version control hosting service web GUI</b>	3363	0.049343
<b>Code editor;Command-line;Dedicated version control GUI application</b>	2665	0.039101
<b>Command-line;Version control hosting service web GUI;Dedicated version control GUI application</b>	1280	0.018780
<b>Code editor;Dedicated version control GUI application</b>	1029	0.015098
<b>Code editor;Version control hosting service web GUI</b>	929	0.013630
<b>Version control hosting service web GUI;Dedicated version control GUI application</b>	826	0.012119
<b>Version control hosting service web GUI</b>	734	0.010769
<b>Code editor;Version control hosting service web GUI;Dedicated version control GUI application</b>	542	0.007952

In [159...]

col\_strip\_nested("VCInteraction")

### Unique Values:

Out[159]:

```
[('Command-line', 56961),
 ('Code editor', 37137),
 ('Version control hosting service web GUI', 19382),
 ('Dedicated version control GUI application', 17976)]
```

In [160...]

```
(df
 .VCInteraction
 .fillna("unanswered")
 .str.replace("Version control hosting service web GUI", "vc hosting service web gui")
 .str.replace("Dedicated version control GUI application", "ded vc gui app")
 .str.lower()
 .str.replace(" ", "_")
 .str.replace("-", "_")
 .unique())
```

```
Out[160]: array(['unanswered', 'code_editor',
       'code_editor;command_line;vc_hosting_service_web_gui;ded_vc_gui_app',
       'command_line;vc_hosting_service_web_gui;ded_vc_gui_app',
       'code_editor;command_line', 'command_line',
       'command_line;ded_vc_gui_app',
       'vc_hosting_service_web_gui;ded_vc_gui_app',
       'code_editor;ded_vc_gui_app',
       'code_editor;command_line;ded_vc_gui_app',
       'command_line;vc_hosting_service_web_gui',
       'code_editor;command_line;vc_hosting_service_web_gui',
       'code_editor;vc_hosting_service_web_gui',
       'vc_hosting_service_web_gui', 'ded_vc_gui_app',
       'code_editor;vc_hosting_service_web_gui;ded_vc_gui_app'],
      dtype=object)
```

## VCHostingPersonal use

```
In [161... df["VCHostingPersonal use"].isna().all()
```

```
Out[161]: True
```

## VCHostingProfessional use

```
In [162... df["VCHostingProfessional use"].isna().all()
```

```
Out[162]: True
```

- All entries in VCHostingPersonal use and VCHostingProfessional use are nan
- These columns will be dropped

## OfficeStackAsyncHaveWorkedWith

```
In [163... df.OfficeStackAsyncHaveWorkedWith
```

```
Out[163]: 0                               NaN
           1                               NaN
           2                               NaN
           3           Jira Work Management;Trello
           4                               NaN
...
           73263          Jira Work Management
           73264                               NaN
           73265          Microsoft Lists
           73266                               NaN
           73267  Confluence;Jira Work Management;Trello;Wrike
Name: OfficeStackAsyncHaveWorkedWith, Length: 73268, dtype: object
```

```
In [164... col_value_counts("OfficeStackAsyncHaveWorkedWith")
```

Out[164]:

		count	pct
OfficeStackAsyncHaveWorkedWith			
<b>Confluence;Jira Work Management</b>	7084	0.153257	
<b>Jira Work Management</b>	6059	0.131082	
<b>Trello</b>	4953	0.107154	
<b>Confluence</b>	3791	0.082015	
<b>Notion</b>	2577	0.055751	
...	...	...	...
<b>Adobe Workfront;Airtable;Asana;Confluence;Jira Work Management;Microsoft Planner;Smartsheet;Trello</b>	1	0.000022	
<b>Asana;DingTalk (Teambition);Microsoft Planner;monday.com</b>	1	0.000022	
<b>ClickUp;monday.com;Notion;Stack Overflow for Teams</b>	1	0.000022	
<b>Airtable;Asana;Confluence;Jira Work Management;Microsoft Lists;monday.com;Trello</b>	1	0.000022	
<b>Airtable;Microsoft Planner;Microsoft Lists;Trello</b>	1	0.000022	

1085 rows × 2 columns

In [165...]

col\_strip\_nested("OfficeStackAsyncHaveWorkedWith")

### Unique Values:

Out[165]:

```
[('Jira Work Management', 24234),
 ('Confluence', 19496),
 ('Trello', 16324),
 ('Notion', 9711),
 ('Asana', 3874),
 ('ClickUp', 2704),
 ('Microsoft Planner', 2282),
 ('Stack Overflow for Teams', 1804),
 ('monday.com', 1639),
 ('Airtable', 1438),
 ('Microsoft Lists', 973),
 ('Smartsheet', 654),
 ('Wrike', 417),
 ('Adobe Workfront', 368),
 ('DingTalk (Teambition)', 227),
 ('Swit', 131),
 ('Workzone', 112),
 ('Planview Projectplace or Clarizen', 84),
 ('Cerri', 61),
 ('Wimi', 61),
 ('Leankor', 57)]
```

In [166...]

```
(df
 .OfficeStackAsyncHaveWorkedWith
 .fillna("unanswered")
 .str.replace("Stack Overflow for Teams", "Stack Overflow Teams")
 .str.replace("monday\.com", "monday_com", regex=True)
 .str.replace("DingTalk \(\Teambition\)", "DingTalk", regex=True))
```

```
.str.replace("Planview Projectplace or Clarizen", "Clarizen")
.str.lower()
.str.replace(" ", "_")
.value_counts()
```

Out[166]:

```
unanswered
27045
confluence;jira_work_management
7084
jira_work_management
6059
trello
4953
confluence
3791

...
adobe_workfront;airtable;asana;confluence;jira_work_management;microsoft_planner;smar
tsheet;trello      1
asana;dingtalk;microsoft_planner;monday_com
1
clickup;monday_com;notion;stack_overflow_teams
1
airtable;asana;confluence;jira_work_management;microsoft_lists;monday_com;trello
1
airtable;microsoft_planner;microsoft_lists;trello
1
Name: OfficeStackAsyncHaveWorkedWith, Length: 1086, dtype: int64
```

## OfficeStackAsyncWantToWorkWith

In [167...]

```
col_strip_nested("OfficeStackAsyncWantToWorkWith")
```

### Unique Values:

Out[167]:

```
[('Jira Work Management', 14695),
 ('Confluence', 10568),
 ('Trello', 8968),
 ('Notion', 7923),
 ('Stack Overflow for Teams', 2440),
 ('Asana', 1973),
 ('ClickUp', 1808),
 ('Microsoft Planner', 1379),
 ('monday.com', 1024),
 ('Airtable', 997),
 ('Microsoft Lists', 690),
 ('Adobe Workfront', 314),
 ('Smartsheet', 307),
 ('Wrike', 211),
 ('DingTalk (Teambition)', 156),
 ('Workzone', 136),
 ('Swit', 134),
 ('Planview Projectplace or Clarizen', 81),
 ('Leankor', 77),
 ('Cerri', 75),
 ('Wimi', 73)]
```

## OfficeStackSyncHaveWorkedWith

```
In [168]: df.OfficeStackSyncHaveWorkedWith
```

```
Out[168]: 0                      NaN
1                      NaN
2          Microsoft Teams
3          Slack;Zoom
4  Microsoft Teams;Zoom
...
73263          Slack;Zoom
73264          Rocketchat
73265  Microsoft Teams;Zoom
73266          Zoom
73267  Microsoft Teams;Slack;Zoom
Name: OfficeStackSyncHaveWorkedWith, Length: 73268, dtype: object
```

```
In [169]: col_strip_nested("OfficeStackSyncHaveWorkedWith")
```

## Unique Values:

```
Out[169]: [('Zoom', 36153),
('Microsoft Teams', 36097),
('Slack', 34440),
('Google Chat', 13019),
('Cisco Webex Teams', 6238),
('Mattermost', 2603),
('Rocketchat', 1438),
('RingCentral', 560),
('Symphony', 359),
('Wire', 283),
('Wickr', 189),
('Unify Circuit', 123),
('Coolfire Core', 89)]
```

```
In [170]: (df
    .OfficeStackSyncHaveWorkedWith
    .fillna("unanswered")
    .str.replace("Rocketchat", "Rocket chat")
    .str.replace("RingCentral", "Ring Central")
    .str.lower()
    .str.replace(" ", "_")
    .value_counts())
```

```
Out[170]: unanswered
11140
microsoft_teams
9257
slack;zoom
7722
microsoft_teams;slack;zoom
6068
slack
5334

...
cisco_webex_teams;coolfire_core;google_chat;ring_central;rocket_chat;slack;symphony;wickr;zoom      1
coolfire_core;google_chat;microsoft_teams;slack;zoom
1
google_chat;microsoft_teams;ring_central;rocket_chat;slack;zoom
1
coolfire_core;google_chat;mattermost;microsoft_teams
1
cisco_webex_teams;google_chat;microsoft_teams;slack;unify_circuit;wire;zoom
1
Name: OfficeStackSyncHaveWorkedWith, Length: 398, dtype: int64
```

## OfficeStackSyncWantToWorkWith

```
In [171...]: col_strip_nested("OfficeStackSyncWantToWorkWith")
```

### Unique Values:

```
Out[171]: [('Slack', 27844),
('Microsoft Teams', 19140),
('Zoom', 18042),
('Google Chat', 7655),
('Mattermost', 1972),
('Cisco Webex Teams', 1944),
('Rocketchat', 973),
('Wire', 301),
('Symphony', 268),
('RingCentral', 215),
('Wickr', 183),
('Unify Circuit', 118),
('Coolfire Core', 102)]
```

- The following columns have the exact same workflow as LanguageHaveWorkedWith
  - OfficeStackAsyncHaveWorkedWith
  - OfficeStackAsyncWantToWorkWith
  - OfficeStackSyncHaveWorkedWith
  - OfficeStackSyncWantToWorkWith

## Blockchain

```
In [172...]: col_description("Blockchain")
```

## Description:

How favorable are you about blockchain, crypto, and decentralization?

In [173...]: df.Blockchain

```
Out[173]: 0           NaN
1      Very unfavorable
2      Very unfavorable
3      Very unfavorable
4      Unfavorable
...
73263    Very favorable
73264        Unsure
73265    Very unfavorable
73266        Indifferent
73267        Favorable
Name: Blockchain, Length: 73268, dtype: object
```

In [174...]: df.Blockchain.unique()

```
Out[174]: array([nan, 'Very unfavorable', 'Unfavorable', 'Favorable',
   'Very favorable', 'Indifferent', 'Unsure'], dtype=object)
```

In [175...]: col\_value\_counts("Blockchain")

	count	pct
<b>Blockchain</b>		
<b>Indifferent</b>	18331	0.257925
<b>Favorable</b>	14629	0.205836
<b>Very unfavorable</b>	11625	0.163569
<b>Unfavorable</b>	10549	0.148429
<b>Unsure</b>	8128	0.114365
<b>Very favorable</b>	7809	0.109876

## Observations

- There're 6 unique entries
- The entries are clean and valid

## Steps

- No cleaning required
- The type can be made `category` for less memory usage

## NEWSOSites

In [176...]: col\_description("NEWSOSites")

## Description:

Which of the following Stack Overflow sites have you visited? Select all that apply.

```
In [177...]: df.NEWSOSites
```

```
Out[177]: 0                               NaN
1      Collectives on Stack Overflow;Stack Overflow f...
2      Collectives on Stack Overflow;Stack Overflow;S...
3      Collectives on Stack Overflow;Stack Overflow f...
4      Collectives on Stack Overflow;Stack Overflow f...
...
73263          Stack Overflow;Stack Exchange
73264          Stack Overflow
73265          Stack Overflow;Stack Exchange
73266          Stack Overflow
73267  Collectives on Stack Overflow;Stack Overflow;S...
Name: NEWSOSites, Length: 73268, dtype: object
```

```
In [178...]: col_value_counts("NEWSOSites").shape
```

Out[178]:

		count	pct
NEWSOSites			
	<b>Stack Overflow;Stack Exchange</b>	41859	0.586548
	<b>Stack Overflow</b>	19229	0.269446
	<b>Collectives on Stack Overflow;Stack Overflow;Stack Exchange</b>	3814	0.053444
	<b>Stack Overflow for Teams (private knowledge sharing &amp; collaboration platform for companies);Stack Overflow;Stack Exchange</b>	2100	0.029426
	<b>Collectives on Stack Overflow;Stack Overflow for Teams (private knowledge sharing &amp; collaboration platform for companies);Stack Overflow;Stack Exchange</b>	1124	0.015750
	<b>Collectives on Stack Overflow;Stack Overflow</b>	992	0.013900
	<b>Stack Overflow for Teams (private knowledge sharing &amp; collaboration platform for companies);Stack Overflow</b>	589	0.008253
	<b>Collectives on Stack Overflow</b>	542	0.007595
	<b>I have never visited Stack Overflow or the Stack Exchange network</b>	461	0.006460
	<b>Stack Exchange</b>	219	0.003069
	<b>Collectives on Stack Overflow;Stack Overflow for Teams (private knowledge sharing &amp; collaboration platform for companies);Stack Overflow</b>	172	0.002410
	<b>Stack Overflow for Teams (private knowledge sharing &amp; collaboration platform for companies)</b>	110	0.001541
	<b>Collectives on Stack Overflow;Stack Exchange</b>	71	0.000995
	<b>Collectives on Stack Overflow;Stack Overflow for Teams (private knowledge sharing &amp; collaboration platform for companies)</b>	54	0.000757
	<b>Stack Overflow for Teams (private knowledge sharing &amp; collaboration platform for companies);Stack Exchange</b>	15	0.000210
	<b>Collectives on Stack Overflow;Stack Overflow for Teams (private knowledge sharing &amp; collaboration platform for companies);Stack Exchange</b>	14	0.000196

In [179...]

col\_strip\_nested("NEWSOSites")

## Unique Values:

Out[179]:

```
[('Stack Overflow', 69879),
 ('Stack Exchange', 49216),
 ('Collectives on Stack Overflow', 6783),
 ('Stack Overflow for Teams (private knowledge sharing & collaboration platform for companies)',
  4178),
 ('I have never visited Stack Overflow or the Stack Exchange network', 461)]
```

In [180...]

```
(df
 .NEWSOSites
 .str.replace("Collectives on Stack Overflow", "collectives")
 .str.replace("Stack Overflow for Teams \private knowledge sharing & collaboration pl
 .str.replace("I have never visited Stack Overflow or the Stack Exchange network", "no
 .str.lower()
```

```
.str.replace(" ", "_")
.value_counts()
```

```
Out[180]: stack_overflow;stack_exchange          41859
stack_overflow                           19229
collectives;stack_overflow;stack_exchange      3814
stack_overflow_teams;stack_overflow;stack_exchange 2100
collectives;stack_overflow_teams;stack_overflow;stack_exchange 1124
collectives;stack_overflow                  992
stack_overflow_teams;stack_overflow           589
collectives                               542
none                                      461
stack_exchange                            219
collectives;stack_overflow_teams;stack_overflow 172
stack_overflow_teams                      110
collectives;stack_exchange                71
collectives;stack_overflow_teams           54
stack_overflow_teams;stack_exchange         15
collectives;stack_overflow_teams;stack_exchange 14
Name: NEWSOSites, dtype: int64
```

## Observations

- There're 16 unique entries
- Most of the entries entries occur in less than 1% of the total observations
- Entries are nested, separated by a semicolon

## Steps

- The entries can be shortened
- Can create a column based on no. of sites visited
- Multiple binary columns can be created
- These steps should be kept in mind for performing feature engineering during exploratory analysis

## SOVisitFreq

```
In [181... col_description("SOVisitFreq")]
```

### Description:

How frequently would you say you visit Stack Overflow?

```
In [182... df.SOVisitFreq]
```

```
Out[182]: 0           NaN
1       Daily or almost daily
2      Multiple times per day
3       Daily or almost daily
4      Multiple times per day
...
73263   Daily or almost daily
73264   Daily or almost daily
73265   Multiple times per day
73266   Daily or almost daily
73267   Daily or almost daily
Name: SOVisitFreq, Length: 73268, dtype: object
```

In [183...]: `col_value_counts("SOVisitFreq")`

SOVisitFreq	count	pct
<b>Daily or almost daily</b>	21712	0.305971
<b>A few times per week</b>	19770	0.278604
<b>Multiple times per day</b>	15965	0.224983
<b>A few times per month or weekly</b>	11185	0.157622
<b>Less than once per month or monthly</b>	2329	0.032821

```
(df
    .SOVisitFreq
    .dropna()
    .replace(["Daily or almost daily",
              "Multiple times per day",
              "A few times per week",
              "A few times per month or weekly",
              "Less than once per month or monthly"],
              ["daily",
               "daily",
               "weekly",
               "weekly",
               "monthly"])
    .astype("category")
    .value_counts())
```

```
Out[184]: daily      37677
weekly     30955
monthly    2329
Name: SOVisitFreq, dtype: int64
```

## Observations

- There're 5 unique entries
- the entries are unnecessarily lengthy

## Steps

- The entries can be shortened
- The type can be made `category` for less memory usage

## SOAccount

In [185...]: `col_description("SOAccount")`

### Description:

Do you have a Stack Overflow account?

In [186...]: `df.SOAccount`

```
Out[186]: 0           NaN
          1           Yes
          2           Yes
          3           Yes
          4           Yes
          ...
          73263        Yes
          73264    Not sure/can't remember
          73265        Yes
          73266        Yes
          73267        Yes
Name: SOAccount, Length: 73268, dtype: object
```

In [187...]: `col_value_counts("SOAccount")`

	count	pct
<b>SOAccount</b>		
<b>Yes</b>	58519	0.817624
<b>No</b>	8951	0.125063
<b>Not sure/can't remember</b>	4102	0.057313

In [188...]: `(df
 .SOAccount
 .replace("Not sure/can't remember", "unsure"))`

```
Out[188]: 0           NaN
          1           Yes
          2           Yes
          3           Yes
          4           Yes
          ...
          73263        Yes
          73264    unsure
          73265        Yes
          73266        Yes
          73267        Yes
Name: SOAccount, Length: 73268, dtype: object
```

### Observations

- There're 3 unique entries
- The entries are clean and valid

## Steps

- No cleaning required
- Just one category will be renamed for clarity
- The type can be made `category` for less memory usage

## SOPartFreq

```
In [189]: col_description("SOPartFreq")
```

### Description:

How frequently would you say you participate in Q&A on Stack Overflow? By participate we mean ask, answer, vote for, or comment on questions.

```
In [190]: df.SOPartFreq
```

```
Out[190]: 0                               NaN
1                               Daily or almost daily
2                           Multiple times per day
3                           A few times per week
4                           Daily or almost daily
...
73263                     A few times per month or weekly
73264                               NaN
73265             Less than once per month or monthly
73266 I have never participated in Q&A on Stack Over...
73267             Less than once per month or monthly
Name: SOPartFreq, Length: 73268, dtype: object
```

```
In [191]: col_value_counts("SOPartFreq")
```

	count	pct
<b>SOPartFreq</b>		
<b>Less than once per month or monthly</b>	26846	0.461042
<b>I have never participated in Q&amp;A on Stack Overflow</b>	13498	0.231809
<b>A few times per month or weekly</b>	10559	0.181336
<b>A few times per week</b>	4433	0.076130
<b>Daily or almost daily</b>	1881	0.032303
<b>Multiple times per day</b>	1012	0.017380

```
In [192]:
```

```
(df
 .SOPartFreq
 .replace(["Daily or almost daily",
           "Multiple times per day",
           "A few times per week",
           "A few times per month or weekly",
           "Less than once per month or monthly",
           "I have never participated in Q&A on Stack Overflow"],
          ["daily",
           "multiple",
           "several",
           "monthly",
           "quarterly",
           "never"])
```

```

    "daily",
    "weekly",
    "weekly",
    "monthly",
    "never"])
.astype("category")
.value_counts()

```

Out[192]:

SOPartFreq	Count
monthly	26846
weekly	14992
never	13498
daily	2893

Name: SOPartFreq, dtype: int64

### Observations:

- There're 6 unique entries
- The entries are clean and valid

### Steps:

- Some categories will be renamed for clarity
- The type can be made `category` for less memory usage

## SOComm

In [193...]

```
col_description("SOComm")
```

### Description:

Do you consider yourself a member of the Stack Overflow community?

In [194...]

```
df.SOComm
```

Out[194]:

Index	Description
0	NaN
1	Not sure
2	Neutral
3	Yes, definitely
4	Yes, definitely
...	
73263	Yes, definitely
73264	Neutral
73265	Yes, somewhat
73266	No, not at all
73267	Yes, somewhat

Name: SOComm, Length: 73268, dtype: object

In [195...]

```
col_value_counts("SOComm")
```

Out[195]:

	count	pct
<b>SOComm</b>		
<b>Yes, somewhat</b>	19674	0.275515
<b>No, not really</b>	18728	0.262268
<b>Neutral</b>	14929	0.209066
<b>Yes, definitely</b>	10381	0.145376
<b>No, not at all</b>	6456	0.090410
<b>Not sure</b>	1240	0.017365

In [196...]

```
(df
    .SOComm
    .replace(["Yes, somewhat",
              "Yes, definitely",
              "No, not at all",
              "No, not really",
              "Not sure"],
              ["yes",
               "yes",
               "no",
               "no",
               "unsure"])
    .str.lower()
    .astype("category")
    .value_counts())
```

Out[196]:

yes	30055
no	25184
neutral	14929
unsure	1240
Name:	SOComm, dtype: int64

### Observations:

- There're 6 unique entries
- The entries are clean and valid

### Steps:

- Some categories will be renamed for clarity
- The type can be made `category` for less memory usage

## Age

In [197...]

```
df.Age
```

```
Out[197]: 0           NaN
           1           NaN
           2    25-34 years old
           3    35-44 years old
           4    25-34 years old
           ...
          73263   25-34 years old
          73264   25-34 years old
          73265   55-64 years old
          73266   55-64 years old
          73267   25-34 years old
Name: Age, Length: 73268, dtype: object
```

In [198... df.Age.unique()

```
Out[198]: array([nan, '25-34 years old', '35-44 years old', 'Under 18 years old',
       '18-24 years old', '45-54 years old', '55-64 years old',
       '65 years or older', 'Prefer not to say'], dtype=object)
```

In [199... col\_value\_counts("Age")

Age	count	pct
<b>25-34 years old</b>	28112	0.396245
<b>18-24 years old</b>	16646	0.234629
<b>35-44 years old</b>	13988	0.197164
<b>45-54 years old</b>	5281	0.074437
<b>Under 18 years old</b>	3866	0.054492
<b>55-64 years old</b>	1978	0.027880
<b>65 years or older</b>	554	0.007809
<b>Prefer not to say</b>	521	0.007344

In [200... (df
 .Age
 .rename("age\_group")
 .replace({"Under 18 years old": "minor",
 "18-24 years old": "young\_adult",
 "25-34 years old": "young\_adult",
 "35-44 years old": "middle\_aged",
 "45-54 years old": "middle\_aged",
 "55-64 years old": "senior",
 "65 years or older": "senior",
 "Prefer not to say": "unanswered"})
 .astype("category"))

```
Out[200]: 0      NaN
           1      NaN
           2    young_adult
           3  middle_aged
           4    young_adult
           ...
          73263  young_adult
          73264  young_adult
          73265    senior
          73266    senior
          73267  young_adult
Name: age_group, Length: 73268, dtype: category
Categories (5, object): ['middle_aged', 'minor', 'senior', 'unanswered', 'young_adult']
```

## Observations:

- There're 8 unique entries
- The entries are clean and valid

## Steps:

- The categories will be renamed for clarity
- The type can be made `category` for less memory usage

## Gender

```
In [201... df.Gender
Out[201]: 0      NaN
           1      NaN
           2      Man
           3      Man
           4      NaN
           ...
          73263    Man
          73264    Man
          73265    Man
          73266    Man
          73267    Man
Name: Gender, Length: 73268, dtype: object
```

```
In [202... df.Gender.unique()
```

```
Out[202]: array([nan, 'Man', 'Or, in your own words:', 'Woman',
       'Non-binary, genderqueer, or gender non-conforming',
       'Prefer not to say',
       'Man;Non-binary, genderqueer, or gender non-conforming',
       'Or, in your own words:;Non-binary, genderqueer, or gender non-conforming',
       'Woman;Non-binary, genderqueer, or gender non-conforming',
       'Man;Woman', 'Man;Or, in your own words:',
       'Or, in your own words:;Woman;Non-binary, genderqueer, or gender non-conformin
g',
       'Man;Woman;Non-binary, genderqueer, or gender non-conforming',
       'Or, in your own words:;Woman',
       'Man;Or, in your own words:;Woman;Non-binary, genderqueer, or gender non-confo
rming',
       'Man;Or, in your own words:;Non-binary, genderqueer, or gender non-conformin
g',
       'Man;Or, in your own words:;Woman'], dtype=object)
```

In [203... col\_value\_counts("Gender")#.shape

Out[203]:

		count	pct
Gender			
	<b>Man</b>	64607	0.911846
	<b>Woman</b>	3399	0.047973
	<b>Prefer not to say</b>	1172	0.016541
	<b>Non-binary, genderqueer, or gender non-conforming</b>	704	0.009936
	<b>Or, in your own words:</b>	279	0.003938
	<b>Man;Non-binary, genderqueer, or gender non-conforming</b>	235	0.003317
	<b>Man;Or, in your own words:</b>	171	0.002413
	<b>Woman;Non-binary, genderqueer, or gender non-conforming</b>	160	0.002258
	<b>Man;Woman;Non-binary, genderqueer, or gender non-conforming</b>	31	0.000438
	<b>Man;Woman</b>	24	0.000339
	<b>Man;Or, in your own words:;Woman;Non-binary, genderqueer, or gender non- conforming</b>	18	0.000254
	<b>Or, in your own words:;Woman;Non-binary, genderqueer, or gender non-conforming</b>	15	0.000212
	<b>Or, in your own words:;Non-binary, genderqueer, or gender non-conforming</b>	14	0.000198
	<b>Or, in your own words:;Woman</b>	13	0.000183
	<b>Man;Or, in your own words:;Non-binary, genderqueer, or gender non-conforming</b>	9	0.000127
	<b>Man;Or, in your own words:;Woman</b>	2	0.000028

In [204... col\_strip\_nested("Gender")

## Unique Values:

```
Out[204]: [('Man', 65097),
 ('Woman', 3662),
 ('Non-binary, genderqueer, or gender non-conforming', 1186),
 ('Prefer not to say', 1172),
 ('Or, in your own words:', 521)]
```

In [205...]

```
(df
 .Gender
 .fillna("unanswered")
 .str.replace("Non-binary, genderqueer, or gender non-conforming", "non_binary")
 .str.replace("Or, in your own words:", "other")
 .str.replace("Prefer not to say", "unanswered")
 .str.replace("Man", "male")
 .str.replace("Woman", "female")
 .value_counts())
```

```
Out[205]: male           64607
unanswered        3587
female            3399
non_binary        704
other             279
male;non_binary   235
male;other         171
female;non_binary 160
male;female;non_binary 31
male;female        24
male;other;female;non_binary 18
other;female;non_binary 15
other;non_binary   14
other;female       13
male;other;non_binary 9
male;other;female  2
Name: Gender, dtype: int64
```

## Observations:

- There're 16 unique entries
- The entries are nested

## Steps:

- The categories will be renamed for clarity
- The type can be made `category` for less memory usage

## Trans

In [206...]

```
col_description("Trans")
```

## Description:

Do you identify as transgender?

In [207...]

```
df.Trans
```

```
Out[207]: 0      NaN
           1      NaN
           2      No
           3      No
           4      NaN
           ...
          73263    No
          73264    No
          73265    No
          73266    No
          73267    No
Name: Trans, Length: 73268, dtype: object
```

In [208... df.Trans.unique()

```
Out[208]: array([nan, 'No', 'Or, in your own words:', 'Yes', 'Prefer not to say'],
                 dtype=object)
```

In [209... col\_value\_counts("Trans")

	count	pct
<b>Trans</b>		
<b>No</b>	67392	0.958430
<b>Prefer not to say</b>	1379	0.019612
<b>Yes</b>	1064	0.015132
<b>Or, in your own words:</b>	480	0.006826

In [210... (df
 .Trans
 .str.lower()
 .replace({"prefer not to say": "unanswered",
 "or, in your own words": "yes"})
 .astype("category"))

```
Out[210]: 0      NaN
           1      NaN
           2      no
           3      no
           4      NaN
           ...
          73263    no
          73264    no
          73265    no
          73266    no
          73267    no
Name: Trans, Length: 73268, dtype: category
Categories (3, object): ['no', 'unanswered', 'yes']
```

## Observations:

- There're 4 unique entries
- The entries are clean and valid
- No cleaning required

**Steps:**

- The type can be made `category` for less memory usage

## Sexuality

```
In [211...]: col_description("Sexuality")
```

### Description:

Which of the following describe you, if any? Please check all that apply.

```
In [212...]: df.Sexuality
```

```
Out[212]: 0           NaN
1           NaN
2           Bisexual
3      Straight / Heterosexual
4           NaN
...
73263    Straight / Heterosexual
73264    Straight / Heterosexual
73265    Straight / Heterosexual
73266    Straight / Heterosexual
73267    Straight / Heterosexual
Name: Sexuality, Length: 73268, dtype: object
```

```
In [213...]: col_value_counts("Sexuality")#.shape
```

Out[213]:

		count	pct
Sexuality			
	<b>Straight / Heterosexual</b>	55238	0.829835
	<b>Prefer not to say</b>	4350	0.065350
	<b>Bisexual</b>	2700	0.040562
	<b>Gay or Lesbian</b>	1382	0.020762
	<b>Prefer to self-describe:</b>	1079	0.016210
	<b>Queer</b>	394	0.005919
	<b>Bisexual;Straight / Heterosexual</b>	354	0.005318
	<b>Bisexual;Queer</b>	282	0.004236
	<b>Straight / Heterosexual;Prefer to self-describe:</b>	169	0.002539
	<b>Gay or Lesbian;Queer</b>	150	0.002253
	<b>Straight / Heterosexual;Queer</b>	79	0.001187
	<b>Bisexual;Gay or Lesbian</b>	73	0.001097
	<b>Prefer to self-describe;;Queer</b>	51	0.000766
	<b>Bisexual;Gay or Lesbian;Queer</b>	44	0.000661
	<b>Bisexual;Straight / Heterosexual;Gay or Lesbian;Queer</b>	36	0.000541
	<b>Bisexual;Prefer to self-describe:</b>	30	0.000451
	<b>Bisexual;Prefer to self-describe;;Queer</b>	26	0.000391
	<b>Bisexual;Straight / Heterosexual;Prefer to self-describe;;Gay or Lesbian;Queer</b>	25	0.000376
	<b>Bisexual;Straight / Heterosexual;Gay or Lesbian</b>	21	0.000315
	<b>Bisexual;Straight / Heterosexual;Queer</b>	19	0.000285
	<b>Prefer to self-describe;;Gay or Lesbian</b>	12	0.000180
	<b>Prefer to self-describe;;Gay or Lesbian;Queer</b>	12	0.000180
	<b>Straight / Heterosexual;Gay or Lesbian</b>	12	0.000180
	<b>Bisexual;Straight / Heterosexual;Prefer to self-describe:</b>	8	0.000120
	<b>Straight / Heterosexual;Prefer to self-describe;;Queer</b>	7	0.000105
	<b>Bisexual;Prefer to self-describe;;Gay or Lesbian</b>	3	0.000045
	<b>Straight / Heterosexual;Gay or Lesbian;Queer</b>	2	0.000030
	<b>Bisexual;Straight / Heterosexual;Prefer to self-describe;;Gay or Lesbian</b>	2	0.000030
	<b>Bisexual;Prefer to self-describe;;Gay or Lesbian;Queer</b>	2	0.000030
	<b>Bisexual;Straight / Heterosexual;Prefer to self-describe;;Queer</b>	1	0.000015
	<b>Straight / Heterosexual;Prefer to self-describe;;Gay or Lesbian;Queer</b>	1	0.000015
	<b>Straight / Heterosexual;Prefer to self-describe;;Gay or Lesbian</b>	1	0.000015

In [214]: `col_strip_nested("Sexuality")`

### Unique Values:

Out[214]:

```
[('Straight / Heterosexual', 55975),
 ('Prefer not to say', 4350),
 ('Bisexual', 3626),
 ('Gay or Lesbian', 1778),
 ('Prefer to self-describe:', 1429),
 ('Queer', 1131)]
```

In [215...]

```
(df
 .Sexuality
 .fillna("unanswered")
 .str.replace("Straight / Heterosexual", "Heterosexual")
 .str.replace("Prefer not to say", "unanswered")
 .str.replace("Gay or Lesbian", "Homosexual")
 .str.replace("Prefer to self-describe:", "other")
 .str.lower()
 .value_counts())
```

Out[215]:

heterosexual	55238
unanswered	11053
bisexual	2700
homosexual	1382
other	1079
queer	394
bisexual;heterosexual	354
bisexual;queer	282
heterosexual;other	169
homosexual;queer	150
heterosexual;queer	79
bisexual;homosexual	73
other;queer	51
bisexual;homosexual;queer	44
bisexual;heterosexual;homosexual;queer	36
bisexual;other	30
bisexual;other;queer	26
bisexual;heterosexual;other;homosexual;queer	25
bisexual;heterosexual;homosexual	21
bisexual;heterosexual;queer	19
other;homosexual	12
other;homosexual;queer	12
heterosexual;homosexual	12
bisexual;heterosexual;other	8
heterosexual;other;queer	7
bisexual;other;homosexual	3
heterosexual;homosexual;queer	2
bisexual;heterosexual;other;homosexual	2
bisexual;other;homosexual;queer	2
bisexual;heterosexual;other;queer	1
heterosexual;other;homosexual;queer	1
heterosexual;other;homosexual	1

Name: Sexuality, dtype: int64

### Observations:

- There're 32 unique entries

- The entries are nested

### Steps:

- The categories will be renamed for clarity
- The entries having more than 1 value will be renamed `lgbqt`
- The entries having `Prefer to self-describe` will be renamed `lgbqt`
- The type can be made `category` for less memory usage

## Ethnicity

```
In [216...]: col_description("Ethnicity")
```

### Description:

Which of the following describe you, if any? Please check all that apply.

```
In [217...]: df.Ethnicity
```

```
Out[217]: 0           NaN
1           NaN
2           White
3           White
4           NaN
...
73263      African
73264      White
73265      Multiracial
73266      European
73267      Or, in your own words:
Name: Ethnicity, Length: 73268, dtype: object
```

```
In [218...]: (df
    .Ethnicity
    .unique())
```

```
Out[218]: array(['nan', 'White', 'Or, in your own words:', ...,
   'White;European;North American;Middle Eastern;Asian;Multiracial',
   'White;Middle Eastern;Central American;Hispanic or Latino/a',
   'White;European;North African;Hispanic or Latino/a'], dtype=object)
```

```
In [219...]: col_value_counts("Ethnicity")
```

Out[219]:

		count	pct
Ethnicity			
<b>European</b>	14612	0.210323	
<b>White</b>	13633	0.196232	
<b>White;European</b>	8694	0.125140	
<b>Indian</b>	5240	0.075424	
<b>Asian</b>	3346	0.048162	
...	...	...	...
<b>Indian;European;African;Multiracial</b>	1	0.000014	
<b>Asian;East Asian;Southeast Asian;Pacific Islander</b>	1	0.000014	
<b>Middle Eastern;Asian;Multiracial;Biracial</b>	1	0.000014	
<b>Or, in your own words:;European;Asian</b>	1	0.000014	
<b>White;European;North African;Hispanic or Latino/a</b>	1	0.000014	

1055 rows × 2 columns

In [220...]

col\_strip\_nested("Ethnicity")

## Unique Values:

```
Out[220]: [('White', 27360),
 ('European', 25877),
 ('Indian', 6739),
 ('Asian', 6586),
 ('Hispanic or Latino/a', 3967),
 ('Middle Eastern', 2850),
 ('South American', 2624),
 ('North American', 2331),
 ('African', 2294),
 ('South Asian', 1797),
 ('Prefer not to say', 1732),
 ('Southeast Asian', 1618),
 ('Or, in your own words:', 1524),
 ('Multiracial', 1222),
 ('East Asian', 1214),
 ('Black', 1028),
 ('Biracial', 798),
 ("I don't know", 701),
 ('North African', 611),
 ('Caribbean', 460),
 ('Central American', 416),
 ('Central Asian', 397),
 ('Ethnoreligious group', 348),
 ('Indigenous (such as Native American or Indigenous Australian)', 330),
 ('Pacific Islander', 147)]
```

In [221...]

```
(df
 .Ethnicity
 .fillna("unanswered"))
```

```
.str.replace("Hispanic or Latino/a", "Hispanic")
.str.replace("Prefer not to say", "unanswered")
.str.replace("Or, in your own words:", "other")
.str.replace("I don't know", "unsure")
.str.replace("Indigenous \s(such as Native American or Indigenous Australian\s)", "Indigenous")
.str.lower()
.str.replace(" ", "_")
.unique()
.tolist()
```

```
Out[221]: ['unanswered',
 'white',
 'other',
 'indian',
 'european',
 'white;european',
 'white;north_american',
 'white;european;middle_eastern;ethnoreligious_group',
 'middle_eastern',
 'european;ethnoreligious_group',
 'north_american',
 'african',
 'white;european;asian;east_asian',
 'white;middle_eastern',
 'black;caribbean',
 'asian;east_asian;southeast_asian',
 'asian',
 'central_american',
 'white;european;north_african',
 'hispanic',
 'european;african',
 'european;southeast_asian',
 'white;hispanic',
 'european;hispanic',
 'white;european;middle_eastern',
 'east_asian',
 'white;south_american',
 'african;black',
 'white;hispanic;south_american',
 'african;north_african',
 'indian;asian',
 'asian;south_asian',
 'southeast_asian',
 'unsure',
 'middle_eastern;ethnoreligious_group',
 'white;european;hispanic',
 'north_african',
 'european;north_american;hispanic;multiracial',
 'biracial',
 'white;other;european',
 'other;ethnoreligious_group',
 'white;european;north_american;multiracial;indigenous',
 'middle_eastern;hispanic;south_american',
 'european;hispanic;multiracial;indigenous',
 'indian;asian;south_asian',
 'european;north_american',
 'european;middle_eastern',
 'white;other',
 'asian;southeast_asian',
 'white;european;north_american;multiracial;biracial',
 'white;hispanic;south_american;multiracial',
 'white;ethnoreligious_group',
 'white;european;middle_eastern;asian;multiracial;biracial',
 'other;european',
 'white;hispanic;biracial',
 'multiracial',
 'european;east_asian',
 'white;european;north_american',
 'white;european;hispanic;biracial',
 'north_american;asian;hispanic',
```

```
'south_american',
'white;asian;multiracial',
'european;asian',
'white;asian;east_asian;biracial',
'white;european;african;black;caribbean;multiracial;biracial',
'north_american;asian',
'asian;central_american',
'middle_eastern;asian;east_asian',
'white;asian',
'white;biracial;pacific_islander',
'white;african',
'asian;southeast_asian;south_asian',
'european;indigenous',
'caribbean;central_american',
'asian;east_asian',
'african;black;caribbean',
'white;north_american;north_african;indigenous',
'white;other;north_american',
'ethnoreligious_group',
'white;hispanic;multiracial',
'white;european;indigenous',
'south_asian',
'asian;east_asian;black;southeast_asian;south_asian',
'asian;multiracial',
'white;other;middle_eastern;ethnoreligious_group',
'est_asian;south_asian',
'north_american;east_asian',
'south_american;multiracial',
'north_american;hispanic;multiracial;indigenous',
'indian;south_asian',
'white;european;south_american',
'ethnoreligious_group;multiracial',
'european;black;biracial',
'white;european;middle_eastern;biracial',
'middle_eastern;asian',
'hispanic;south_american',
'middle_eastern;african',
'white;other;european;north_american',
'hispanic;south_american;multiracial',
'indian;southeast_asian',
'north_american;asian;east_asian;biracial',
'black',
'white;middle_eastern;hispanic',
'white;asian;multiracial;biracial',
'white;european;east_asian;multiracial',
'white;asian;biracial',
'indian;north_american',
'white;southeast_asian;multiracial',
'hispanic;biracial',
'white;east_asian',
'hispanic;indigenous',
'european;caribbean;multiracial',
'central_american;hispanic',
'white;south_asian;multiracial',
'white;middle_eastern;ethnoreligious_group',
'white;southeast_asian;biracial',
'white;biracial',
'European;southeast_asian;biracial',
'European;black',
'indian;hispanic;multiracial',
```

'north\_american;asian;biracial',  
'white;north\_american;biracial;indigenous',  
'indigenous',  
'white;ethnoreligious\_group;east\_asian;biracial',  
'indian;asian;southeast\_asian;south\_asian',  
'white;european;hispanic;south\_american;biracial',  
'indian;african;black;biracial',  
'north\_american;multiracial;biracial',  
'east\_asian;hispanic',  
'white;asian;southeast\_asian;biracial',  
'black;central\_american;hispanic',  
'central\_asian',  
'white;multiracial',  
'white;indian;indigenous',  
'hispanic;multiracial',  
'other;indian;central\_asian',  
'white;other;european;ethnoreligious\_group',  
'white;east\_asian;biracial',  
'african;asian;multiracial',  
'white;european;hispanic;multiracial',  
'caribbean',  
'african;caribbean',  
'middle\_eastern;multiracial',  
'white;european;central\_asian',  
'east\_asian;south\_american',  
'white;european;asian;southeast\_asian;biracial',  
'white;north\_american;south\_american',  
'white;european;hispanic;south\_american',  
'white;european;african',  
'white;north\_american;hispanic;indigenous',  
'white;southeast\_asian',  
'european;multiracial',  
'white;north\_american;asian;multiracial',  
'indian;european;asian;multiracial',  
'white;indigenous',  
'north\_american;indigenous',  
'north\_american;asian;multiracial',  
'white;middle\_eastern;biracial',  
'european;middle\_eastern;biracial',  
'north\_american;asian;east\_asian',  
'white;european;north\_american;middle\_eastern;multiracial',  
'european;north\_african;biracial',  
'white;east\_asian;multiracial',  
'black;central\_american',  
'european;asian;biracial',  
'white;black;multiracial;biracial',  
'european;hispanic;south\_american;multiracial',  
'white;european;asian;east\_asian;biracial',  
'black;south\_american',  
'european;biracial;pacific\_islander',  
'caribbean;hispanic',  
'white;north\_american;asian;east\_asian;multiracial',  
'white;indian;european;biracial',  
'indian;middle\_eastern;south\_asian',  
'white;north\_american;asian;east\_asian;biracial',  
'north\_american;asian;southeast\_asian',  
'white;black;caribbean',  
'white;indian;multiracial;biracial',  
'european;biracial',  
'white;other;ethnoreligious\_group',

```
'north_american;multiracial',
'east_asian;southeast_asian',
'asian;central_asian',
'white;other;black;biracial',
'caribbean;hispanic;multiracial',
'european;multiracial;pacific_islander',
'white;european;south_american;multiracial;biracial',
'european;north_american;central_american',
'black;hispanic;multiracial',
'north_american;middle_eastern',
'white;pacific_islander',
'white;black',
'other;middle_eastern;african',
'white;indian;north_american',
'indian;middle_eastern;asian',
'white;asian;south_asian;biracial',
'asian;east_asian;southeast_asian;south_asian',
'white;north_american;hispanic;south_american',
'asian;southeast_asian;pacific_islander',
'middle_eastern;asian;multiracial;biracial',
'european;asian;south_asian;central_asian',
'white;south_american;biracial',
'other;asian',
'indian;east_asian',
'other;asian;central_asian',
'white;north_american;african',
'east_asian;southeast_asian;south_asian',
'asian;east_asian;hispanic;south_american',
'european;african;biracial',
'indian;north_american;asian;south_asian',
'white;north_american;hispanic;multiracial;indigenous',
'southeast_asian;south_asian',
'est_asian;central_asian',
'white;indian;south_american;multiracial',
'other;asian;southeast_asian',
'other;middle_eastern',
'asian;hispanic',
'white;asian;east_asian',
'white;european;asian;southeast_asian;multiracial',
'indian;european;north_american;middle_eastern;ethnoreligious_group;african;asian;ea
st_asian;caribbean;southeast_asian;central_american;north_african;south_american;sout
h_asian;multiracial;indigenous;pacific_islander;central_asian',
'asian;black;south_asian',
'white;biracial;central_asian',
'white;central_asian',
'middle_eastern;north_african',
'indian;african;asian;black;multiracial',
'indian;southeast_asian;south_asian',
'indian;middle_eastern',
'multiracial;biracial;indigenous',
'indian;european;middle_eastern;asian',
'white;european;ethnoreligious_group',
'white;middle_eastern;asian;east_asian;central_asian',
'other;indian;asian',
'white;other;middle_eastern',
'indian;asian;southeast_asian',
'black;biracial',
'indian;south_american',
'est_asian;biracial',
'other;hispanic;south_american',
```

'white;indian;european;north\_american;middle\_eastern;ethnoreligious\_group;african;asian;east\_asian;black;caribbean;southeast\_asian;central\_american;north\_african;hispanic;south\_american;south\_asian;multiracial;biracial;indigenous;pacific\_islander;central\_asian',  
'white;european;asian;multiracial',  
'indian;north\_american;asian;biracial',  
'european;central\_asian',  
'white;indian;biracial',  
'asian;southeast\_asian;hispanic',  
'southeast\_asian;pacific\_islander',  
'european;hispanic;south\_american',  
'white;european;multiracial',  
'black;caribbean;multiracial',  
'middle\_eastern;asian;central\_asian',  
'white;european;asian',  
'white;european;african;biracial',  
'other;southeast\_asian',  
'north\_american;black;multiracial;biracial',  
'asian;east\_asian;central\_asian',  
'other;south\_asian',  
'white;north\_african',  
'asian;east\_asian;multiracial',  
'white;ethnoreligious\_group;south\_american',  
'white;european;north\_african;biracial',  
'white;african;north\_african',  
'african;black;multiracial;biracial',  
'european;north\_american;indigenous',  
'european;middle\_eastern;asian',  
'white;middle\_eastern;asian',  
'other;asian;east\_asian',  
'white;indian;asian;east\_asian;south\_asian;multiracial;biracial',  
'middle\_eastern;african;multiracial',  
'white;european;middle\_eastern;north\_african;multiracial',  
'european;asian;central\_american',  
'indian;black',  
'european;southeast\_asian;multiracial;indigenous',  
'indian;european',  
'white;indian;european;african;asian;black;multiracial',  
'middle\_eastern;south\_asian',  
'middle\_eastern;central\_asian',  
'white;middle\_eastern;north\_african',  
'european;middle\_eastern;central\_asian',  
'white;european;south\_asian;biracial',  
'african;multiracial',  
'white;north\_american;ethnoreligious\_group',  
'middle\_eastern;african;north\_african',  
'white;middle\_eastern;asian;central\_asian',  
'other;indian;south\_asian',  
'white;european;asian;biracial',  
'white;middle\_eastern;central\_asian',  
'white;asian;hispanic;multiracial',  
'other;asian;south\_asian',  
'european;asian;multiracial',  
'pacific\_islander',  
'middle\_eastern;south\_american',  
'white;middle\_eastern;multiracial',  
'european;south\_american;multiracial',  
'white;hispanic;south\_american;multiracial;indigenous',  
'white;north\_american;asian',  
'other;european;middle\_eastern',

```
'european;south_american',
'white;european;middle_eastern;asian',
'white;north_american;east_asian;biracial',
'middle_eastern;biracial',
'white;asian;central_asian',
'white;middle_eastern;african',
'african;biracial',
'european;asian;central_asian',
'white;indian',
'north_american;hispanic',
'white;european;middle_eastern;asian;central_asian',
'white;middle_eastern;african;north_african',
'black;caribbean;hispanic;south_american;multiracial;biracial;indigenous',
'white;north_american;middle_eastern;north_african',
'white;asian;east_asian;multiracial;biracial',
'european;middle_eastern;multiracial',
'european;african;multiracial;biracial',
'north_american;middle_eastern;hispanic;multiracial;indigenous',
'black;caribbean;hispanic;south_american;biracial;indigenous',
'middle_eastern;african;north_african;multiracial',
'white;south_american;multiracial',
'middle_eastern;multiracial;biracial',
'other;middle_eastern;multiracial',
'asian;hispanic;multiracial',
'white;other;european;south_american',
'indian;middle_eastern;asian;south_asian',
'hispanic;multiracial;indigenous',
'other;indian;asian;south_asian',
'asian;hispanic',
'white;caribbean;hispanic',
'black;hispanic;biracial',
'indian;caribbean;south_american;south_asian',
'asian;south_american',
'white;caribbean;biracial',
'white;other;indian;european;north_american;middle_eastern;ethnoreligious_group;afri
can;asian;east_asian;black;caribbean;southeast_asian;central_american;north_african;h
ispanic;south_american;south_asian;multiracial;biracial;indigenous;pacific_islander;c
entral_asian',
'black;hispanic',
'white;other;hispanic;multiracial',
'european;african;hispanic',
'indian;north_american;south_asian',
'asian;asian;east_asian',
'middle_eastern;asian;southeast_asian',
'white;european;north_american;middle_eastern',
'white;european;middle_eastern;east_asian;multiracial',
'north_american;hispanic;south_american',
'black;hispanic;south_american;indigenous',
'white;european;central_american;hispanic',
'white;hispanic;south_american;biracial',
'african;north_african;multiracial',
'north_american;black;hispanic;multiracial',
'white;other;north_american;middle_eastern',
'european;central_american',
'white;asian;south_american;biracial',
'white;other;european;middle_eastern;ethnoreligious_group',
'white;asian;hispanic;multiracial;indigenous',
'white;north_american;asian;southeast_asian;multiracial;indigenous',
'european;north_american;multiracial',
'other;east_asian',
```

'european;north\_american;asian',  
'hispanic;south\_american;multiracial;biracial',  
'white;european;biracial',  
'white;other;european;ethnoreligious\_group;central\_asian',  
'other;multiracial',  
'hispanic;south\_american;multiracial;indigenous',  
'white;european;biracial;indigenous',  
'white;european;north\_american;african;black;multiracial;biracial;indigenous',  
'indian;south\_asian;multiracial',  
'european;middle\_eastern;african;multiracial',  
'indian;asian;east\_asian;southeast\_asian;south\_asian',  
'white;ethnoreligious\_group;indigenous',  
'white;multiracial;pacific\_islander',  
'asian;black;biracial',  
'white;caribbean;hispanic;multiracial;indigenous',  
'european;asian;multiracial;biracial',  
'european;asian;east\_asian',  
'asian;southeast\_asian;multiracial',  
'caribbean;south\_american',  
'white;multiracial;biracial;indigenous',  
'european;african;black;multiracial;biracial',  
'white;south\_asian',  
'asian;hispanic;biracial',  
'other;biracial',  
'european;middle\_eastern;hispanic;multiracial',  
'north\_american;central\_american',  
'asian;east\_asian;hispanic;biracial',  
'white;black;biracial',  
'white;north\_american;hispanic',  
'white;european;caribbean;central\_american;hispanic',  
'white;central\_american;hispanic',  
'middle\_eastern;hispanic;multiracial;pacific\_islander',  
'north\_american;asian;south\_american',  
'other;black;caribbean',  
'white;north\_american;black;multiracial;biracial',  
'north\_american;southeast\_asian',  
'north\_american;central\_american;hispanic;south\_american',  
'white;indigenous;pacific\_islander',  
'white;caribbean;central\_american;hispanic',  
'north\_american;middle\_eastern;hispanic',  
'white;african;multiracial;indigenous',  
'white;european;north\_american;indigenous',  
'caribbean;hispanic;south\_american',  
'white;north\_american;asian;east\_asian;multiracial;biracial',  
'ethnoreligious\_group;south\_american',  
'other;african',  
'indian;north\_american;southeast\_asian',  
'white;european;middle\_eastern;multiracial;biracial;central\_asian',  
'black;south\_american;multiracial',  
'indian;middle\_eastern;ethnoreligious\_group;african;asian;east\_asian;black;caribbea  
n;southeast\_asian;north\_african;hispanic;south\_american;south\_asian;multiracial;birac  
ial;indigenous;pacific\_islander;central\_asian',  
'european;ethnoreligious\_group;biracial;central\_asian',  
'north\_american;central\_american;hispanic',  
'indian;asian;black',  
'indian;european;asian',  
'white;asian;multiracial;pacific\_islander',  
'asian;black;southeast\_asian;biracial',  
'north\_american;asian;south\_asian',  
'white;european;south\_american;multiracial',

```
'white;north_american;caribbean',
'white;indian;european',
'white;north_american;ethnoreligious_group;hispanic;biracial',
'white;north_american;central_american',
'white;other;east_asian;biracial',
'middle_eastern;asian;central_american',
'north_american;middle_eastern;north_african',
'european;north_american;middle_eastern',
'white;north_american;middle_eastern',
'african;black;north_african',
'asian;pacific_islander',
'white;hispanic;multiracial;indigenous',
'white;other;european;middle_eastern;ethnoreligious_group;multiracial',
'white;caribbean',
'white;asian;caribbean',
'multiracial;biracial',
'white;european;middle_eastern;african',
'white;european;asian;east_asian;southeast_asian;south_asian;multiracial;biracial',
'white;other;european;hispanic;south_american',
'white;european;north_american;asian;east_asian;multiracial;biracial',
'indian;caribbean',
'white;north_american;hispanic;biracial',
'white;north_american;middle_eastern;black;caribbean;indigenous;central_asian',
'other;central_asian',
'african;south_american',
'indian;east_asian;south_asian',
'euopean;hispanic;south_american;multiracial;biracial',
'white;european;asian;east_asian;southeast_asian;south_asian',
'african;asian',
'black;north_african',
'white;east_asian;multiracial;biracial',
'north_american;black;hispanic;south_american',
'white;east_asian;multiracial;pacific_islander',
'black;multiracial',
'biracial;pacific_islander',
'white;european;east_asian;biracial',
'white;european;hispanic;south_american;multiracial;biracial',
'other;north_american',
'asian;caribbean;south_asian',
'white;asian;southeast_asian;hispanic',
'north_american;black',
'white;indian;north_american;asian;east_asian;southeast_asian;south_asian;multiracia
l',
'white;multiracial;indigenous',
'white;european;ethnoreligious_group;african;hispanic;multiracial;biracial;indigenou
s',
'white;north_african;hispanic',
'other;south_american',
'south_american;biracial',
'white;european;middle_eastern;multiracial',
'north_american;african;black;south_american;biracial',
'white;european;black',
'north_african;multiracial',
'white;other;european;north_american;central_american;north_african',
'white;other;south_american',
'european;middle_eastern;ethnoreligious_group;biracial',
'hispanic;south_american;indigenous',
'middle_eastern;southeast_asian;biracial',
'indian;asian;east_asian;south_asian',
'white;north_american;indigenous',
```

```
'other;indian;african',
'european;african;black;multiracial',
'white;middle_eastern;asian;multiracial',
'white;indian;multiracial',
'caribbean;hispanic;biracial',
'indian;asian;hispanic',
'south_american;multiracial;biracial',
'african;hispanic;biracial',
'indian;north_american;south_american',
'indian;asian;east_asian',
'white;european;ethnoreligious_group;multiracial',
'european;east_asian;biracial',
'middle_eastern;hispanic;biracial',
'other;european;hispanic',
'white;middle_eastern;black;biracial',
'middle_eastern;southeast_asian',
'white;european;caribbean;hispanic',
'north_american;north_african;pacific_islander',
'caribbean;central_american;hispanic;south_american',
'indian;asian;southeast_asian;multiracial',
'asian;multiracial;biracial',
'other;european;east_asian;biracial',
'white;european;middle_eastern;central_asian',
'european;north_african',
'east_asian;pacific_islander',
'white;east_asian;hispanic;multiracial',
'white;european;east_asian;multiracial;biracial',
'other;north_african',
'european;south_asian',
'other;black',
'european;multiracial;central_asian',
'north_american;south_american;central_asian',
'white;african;black;biracial',
'indian;european;biracial',
'indian;asian;east_asian;biracial',
'white;european;hispanic;indigenous',
'white;asian;multiracial;central_asian',
'white;caribbean;multiracial;biracial;indigenous',
'european;middle_eastern;ethnoreligious_group;multiracial;indigenous',
'multiracial;central_asian',
'indian;black;caribbean;south_asian;multiracial;biracial',
'other;ethnoreligious_group;biracial',
'white;european;north_american;multiracial;biracial;indigenous',
'white;ethnoreligious_group;asian;multiracial;pacific_islander',
'european;asian;hispanic',
'white;asian;southeast_asian;multiracial;biracial',
'white;other;hispanic;south_american',
'caribbean;biracial',
'white;north_american;hispanic;multiracial',
'white;other;african;hispanic;south_american;multiracial;indigenous',
'white;european;central_american;biracial',
'white;biracial;indigenous',
'european;middle_eastern;north_african',
'north_american;southeast_asian;south_asian;multiracial',
'white;multiracial;biracial',
'ethnoreligious_group;african;black;caribbean;multiracial;biracial',
'central_american;hispanic;indigenous',
'indian;north_american;black;hispanic;multiracial;indigenous',
'other;european;multiracial',
'caribbean;central_american;hispanic',
```

```
'white;european;multiracial;central_asian',
'other;north_american;asian;east_asian;southeast_asian;south_asian',
'indian;central_asian',
'white;north_american;southeast_asian;multiracial',
'white;asian;biracial;pacific_islander',
'european;north_american;hispanic',
'black;hispanic;south_american',
'white;north_american;caribbean;hispanic',
'european;ethnoreligious_group;biracial',
'est_asian;multiracial',
'indian;european;north_american;east_asian;north_african;hispanic;south_american;sou
th_asian;pacific_islander;central_asian',
'white;european;north_american;ethnoreligious_group;indigenous',
'south_american;multiracial;indigenous',
'ethnoreligious_group;hispanic',
'white;european;african;north_african;biracial',
'black;south_american;multiracial;biracial',
'other;european;indigenous',
'indian;east_asian;southeast_asian;multiracial',
'white;middle_eastern;ethnoreligious_group;african;north_african',
'european;pacific_islander',
'asian;east_asian;south_american;multiracial',
'european;asian;hispanic;multiracial',
'european;asian;south_american',
'indian;asian;central_asian',
'other;indian',
'african;black;hispanic;south_american;multiracial;indigenous',
'ethnoreligious_group;asian;south_asian',
'middle_eastern;asian;multiracial',
'white;european;asian;east_asian;multiracial;biracial',
'european;hispanic;multiracial;biracial',
'asian;east_asian;biracial',
'white;middle_eastern;asian;biracial',
'white;european;multiracial;indigenous',
'european;hispanic;biracial',
'other;middle_eastern;asian',
'white;other;european;middle_eastern',
'white;asian;southeast_asian',
'north_american;middle_eastern;asian;east_asian;southeast_asian;multiracial',
'black;hispanic;south_american;multiracial;biracial',
'middle_eastern;south_asian;central_asian',
'european;black;caribbean',
'southeast_asian;multiracial',
'other;middle_eastern;ethnoreligious_group',
'asian;black',
'north_african;biracial',
'european;east_asian;south_american;biracial',
'white;middle_eastern;east_asian',
'indian;african',
'other;north_american;hispanic',
'southeast_asian;multiracial;biracial',
'indian;african;asian',
'european;african;central_american;central_asian',
'white;indian;european;north_american;middle_eastern;ethnoreligious_group;asian;cari
bbean;southeast_asian;central_american;north_african;hispanic;south_american;south_as
ian;indigenous;pacific_islander;central_asian',
'white;european;african;caribbean;hispanic;multiracial',
'other;pacific_islander',
'white;european;asian;central_asian',
'asian;east_asian;southeast_asian;south_asian;multiracial;central_asian',
```

```
'asian;caribbean;multiracial',
'european;hispanic;multiracial',
'indian;asian;south_asian;central_asian',
'asian;east_asian;south_asian',
'white;european;asian;multiracial;central_asian',
'european;asian;caribbean;multiracial',
'white;middle_eastern;north_african;multiracial;biracial',
'other;caribbean;hispanic;south_american;multiracial',
'european;north_american;hispanic;south_american;multiracial',
'indian;european;multiracial',
'african;north_african;biracial',
'white;european;asian;east_asian;multiracial',
'european;east_asian;hispanic;south_american;multiracial',
'white;european;african;indigenous',
'white;other;european;north_american;asian;east_asian;hispanic;south_american;multiracial',
'white;indian;european;asian;south_asian;multiracial;biracial',
'middle_eastern;african;black',
'white;ethnoreligious_group;central_american',
'european;caribbean',
'white;other;european;african;black;multiracial;biracial',
'white;black;hispanic;multiracial',
'euopean;asian;southeast_asian;multiracial',
'white;european;middle_eastern;asian;multiracial',
'white;other;pacific_islander',
'white;middle_eastern;caribbean;multiracial',
'white;other;asian;biracial',
'black;caribbean;hispanic',
'asian;southeast_asian;biracial',
'euopean;black;multiracial;biracial',
'euopean;middle_eastern;african',
'euopean;middle_eastern;hispanic',
'indian;european;north_american;middle_eastern;ethnoreligious_group;asian;east_asia
n;south_american;multiracial',
'asian;east_asian;southeast_asian;biracial',
'indian;european;caribbean;multiracial',
'white;hispanic;multiracial;biracial',
'euopean;asian;east_asian;southeast_asian;multiracial;biracial',
'indian;european;african;multiracial',
'middle_eastern;asian;biracial',
'asian;east_asian;southeast_asian;pacific_islander',
'white;european;hispanic;south_american;multiracial',
'other;european;asian',
'other;middle_eastern;north_african',
'euopean;african;black;caribbean;biracial',
'middle_eastern;hispanic',
'middle_eastern;east_asian;biracial',
'euopean;north_american;multiracial;indigenous',
'other;african;black;biracial',
'white;middle_eastern;ethnoreligious_group;north_african',
'white;hispanic;south_american;multiracial;biracial;indigenous',
'african;hispanic;south_american;multiracial;biracial',
'african;asian;east_asian;black;caribbean;central_american;central_asian',
'other;north_american;asian;south_american;multiracial;biracial',
'euopean;central_american;hispanic',
'white;european;asian;east_asian;southeast_asian;hispanic',
'indian;middle_eastern;asian;east_asian;black;southeast_asian;south_asian;multiracia
l;central_asian',
'indian;asian;multiracial',
'white;asian;east_asian;multiracial',
```

'white;indian;european;north\_american;middle\_eastern;ethnoreligious\_group;african;asian;east\_asian;black;caribbean;southeast\_asian;central\_american;north\_african;hispanic;south\_american;south\_asian;biracial;indigenous;pacific\_islander;central\_asian',  
'asian;south\_asian;biracial',  
'white;asian;southeast\_asian;multiracial;biracial;pacific\_islander',  
'indian;ethnoreligious\_group;asian;southeast\_asian;south\_asian;multiracial;central\_asian',  
'white;north\_american;asian;biracial',  
'european;north\_american;south\_american',  
'white;european;black;hispanic;south\_american;multiracial;indigenous',  
'white;north\_american;multiracial;indigenous',  
'white;indian;north\_american;multiracial;biracial',  
'white;indian;middle\_eastern;multiracial',  
'white;north\_american;biracial',  
'white;north\_american;asian;multiracial;biracial',  
'multiracial;indigenous;pacific\_islander',  
'white;south\_asian;biracial',  
'east\_asian;black;central\_asian',  
'asian;east\_asian;indigenous',  
'white;indian;european;ethnoreligious\_group;asian;black',  
'european;north\_american;asian;central\_american;south\_american',  
'white;european;north\_american;middle\_eastern;north\_african;multiracial',  
'white;european;african;asian;east\_asian;black;caribbean;southeast\_asian;central\_american;north\_african;south\_american;south\_asian;biracial',  
'white;north\_african;south\_asian',  
'white;black;multiracial',  
'white;other;indigenous',  
'indian;european;north\_american;middle\_eastern;asian;southeast\_asian;north\_african;south\_american;south\_asian;pacific\_islander',  
'white;other;black;multiracial;indigenous',  
'south\_asian;central\_asian',  
'african;southeast\_asian;biracial',  
'white;indian;middle\_eastern;central\_asian',  
'white;indian;north\_american;south\_asian',  
'other;european;middle\_eastern;central\_asian',  
'white;european;north\_american;north\_african',  
'white;asian;black',  
'white;indian;european;middle\_eastern;multiracial',  
'european;middle\_eastern;ethnoreligious\_group',  
'indian;european;central\_american;south\_american',  
'asian;east\_asian;hispanic;pacific\_islander',  
'other;north\_american;southeast\_asian;south\_american;south\_asian',  
'white;central\_american',  
'european;asian;east\_asian;multiracial',  
'asian;hispanic;south\_american;multiracial',  
'white;indian;asian',  
'indian;multiracial',  
'white;other;european;biracial',  
'asian;indigenous',  
'european;asian;black',  
'european;black;multiracial',  
'white;other;european;east\_asian;biracial',  
'european;middle\_eastern;african;east\_asian',  
'middle\_eastern;east\_asian',  
'white;other;asian;southeast\_asian',  
'white;other;european;multiracial',  
'white;other;european;middle\_eastern;multiracial',  
'white;european;black;caribbean;biracial',  
'european;black;hispanic',  
'european;asian;southeast\_asian',

```
'white;ethnoreligious_group;southeast_asian;south_american;south_asian;multiracial;b  
iracial;pacific_islander',  
'white;african;black;caribbean;multiracial;biracial',  
'white;other;middle_eastern;hispanic',  
'white;asian;hispanic',  
'black;caribbean;biracial',  
'ethnoreligious_group;african;black',  
'black;indigenous',  
'white;ethnoreligious_group;hispanic',  
'white;european;ethnoreligious_group;hispanic;south_american',  
'white;asian;south_asian',  
'black;hispanic;south_american;multiracial',  
'middle_eastern;southeast_asian;south_asian',  
'african;black;biracial',  
'white;european;caribbean',  
'indian;african;black;multiracial;biracial',  
'white;european;hispanic;south_american;multiracial;indigenous',  
'white;north_american;asian;multiracial;pacific_islander',  
'indian;asian;black;hispanic;multiracial;indigenous',  
'white;caribbean;hispanic;south_american',  
'european;middle_eastern;south_american',  
'caribbean;hispanic;south_american;multiracial',  
'white;north_american;east_asian;black;multiracial',  
'white;european;caribbean;multiracial;indigenous',  
'white;other;indian;european;north_american;ethnoreligious_group;central_american;hi  
spanic;multiracial;biracial;indigenous',  
'white;asian;black;hispanic',  
'indian;north_american;african;north_african;indigenous;pacific_islander',  
'central_american;indigenous',  
'white;european;north_american;middle_eastern;ethnoreligious_group;african;asian;nor  
th_african;hispanic',  
'white;asian;black;multiracial;biracial',  
'white;indian;european;north_american;middle_eastern;ethnoreligious_group;african;as  
ian;east_asian;black;caribbean;southeast_asian;central_american;north_african;hispan  
ic;south_american;multiracial;biracial;indigenous;pacific_islander;central_asian',  
'central_american;hispanic;multiracial',  
'white;other;hispanic',  
'white;european;north_american;central_american;multiracial',  
'north_american;caribbean',  
'african;black;multiracial',  
'white;european;north_american;middle_eastern;african;north_african;multiracial;bira  
cial',  
'white;ethnoreligious_group;hispanic;multiracial',  
'north_american;african;black',  
'other;hispanic',  
'hispanic;south_american;biracial',  
'white;middle_eastern;indigenous',  
'white;indian;ethnoreligious_group;african;asian;east_asian;black;caribbean;southeas  
t_asian;central_american;north_african;hispanic;south_american;multiracial;biracial;p  
acific_islander;central_asian',  
'white;north_african;hispanic;multiracial;biracial',  
'black;south_american;biracial',  
'other;european;biracial',  
'european;caribbean;hispanic',  
'white;asian;black;hispanic;multiracial',  
'white;european;middle_eastern;multiracial;biracial',  
'north_american;asian;east_asian;south_asian',  
'other;indigenous',  
'white;ethnoreligious_group;hispanic;south_american',  
'white;caribbean;hispanic;multiracial',
```

```
'european;north_american;hispanic;biracial',
'middle_eastern;north_african;multiracial',
'indian;european;southeast_asian;hispanic;multiracial',
'white;other;north_american;hispanic',
'white;black;caribbean;south_asian;biracial',
'asian;south_american;multiracial',
'white;asian;south_american',
'indigenous;pacific_islander',
'north_american;ethnoreligious_group',
'african;caribbean;multiracial',
'white;east_asian;biracial;pacific_islander',
'white;central_american;multiracial',
'white;european;caribbean;hispanic;south_american;multiracial',
'white;north_american;pacific_islander',
'white;indian;north_american;ethnoreligious_group;asian;south_asian;multiracial;bira
cial',
'african;caribbean;hispanic',
'north_american;african',
'white;north_american;southeast_asian;biracial',
'central_american;hispanic;south_american',
'white;indian;european;north_american;east_asian;multiracial;indigenous',
'white;indian;european;south_asian;multiracial;biracial',
'white;indian;european;north_american;indigenous',
'black;hispanic;south_american;multiracial;indigenous',
'asian;southeast_asian;hispanic;multiracial;biracial;pacific_islander',
'indian;north_american;central_american',
'indian;hispanic;biracial',
'southeast_asian;biracial',
'white;asian;southeast_asian;multiracial',
'other;biracial;pacific_islander',
'indian;asian;black;south_asian',
'white;european;central_american',
'european;asian;southeast_asian;biracial',
'european;african;black',
'other;african;indigenous',
'european;north_african;multiracial',
'indian;middle_eastern;asian;black;south_asian',
'indian;european;multiracial;biracial',
'european;african;black;biracial',
'white;european;caribbean;hispanic;multiracial',
'white;caribbean;north_african',
'white;black;hispanic;south_american;multiracial',
'asian;central_american;hispanic',
'european;east_asian;hispanic',
'asian;east_asian;black;caribbean;multiracial',
'white;hispanic;multiracial;biracial;indigenous',
'white;asian;east_asian;hispanic;multiracial',
'white;north_american;black;multiracial;indigenous',
'white;european;north_american;multiracial',
'white;european;north_american;ethnoreligious_group',
'hispanic;south_american;biracial;indigenous',
'white;indian;north_american;biracial',
'other;european;north_american',
'white;european;black;caribbean;multiracial',
'white;hispanic;pacific_islander',
'ethnoreligious_group;hispanic;biracial',
'black;hispanic;multiracial;indigenous',
'white;hispanic;south_american;multiracial;biracial',
'indian;european;north_american;middle_eastern;ethnoreligious_group;african;asian;ea
st_asian;black;caribbean;southeast_asian;central_american;north_african;hispanic;sout
```

```
h_american;south_asian;multiracial;biracial;indigenous;pacific_islander;central_asia
n',
'indian;european;asian;south_asian',
'white;caribbean;hispanic;multiracial;biracial;indigenous',
'european;east_asian;multiracial',
'asian;biracial;central_asian',
'north_american;hispanic;multiracial',
'african;south_american;multiracial',
'african;southeast_asian;hispanic',
'white;north_american;asian;southeast_asian;multiracial;biracial;pacific_islander',
'european;caribbean;hispanic;biracial',
'white;other;european;hispanic',
'european;central_american;hispanic;multiracial',
'caribbean;multiracial',
'south_american;south_asian',
'white;middle_eastern;african;black;north_african;biracial',
'southeast_asian;central_asian',
'white;north_american;middle_eastern;african;black;multiracial',
'white;european;north_american;middle_eastern;asian;east_asian;caribbean;central_ame
rican;hispanic;multiracial;indigenous;pacific_islander',
'black;central_american;hispanic;multiracial',
'north_american;black;indigenous',
'north_american;black;multiracial',
'european;middle_eastern;asian;multiracial',
'other;asian;pacific_islander',
'asian;black;southeast_asian',
'black;caribbean;central_american',
'white;european;southeast_asian',
'indian;east_asian;indigenous',
'indian;ethnoreligious_group;asian;southeast_asian;south_asian;indigenous',
'white;central_american;hispanic;multiracial',
'european;middle_eastern;multiracial;biracial',
'other;african;north_african',
'south_american;indigenous',
'middle_eastern;african;biracial',
'white;south_american;multiracial;indigenous',
'other;european;caribbean;biracial',
'indian;north_american;middle_eastern;asian;east_asian;southeast_asian;central_ameri
can;south_american;south_asian;indigenous;central_asian',
'other;north_african;indigenous',
'european;middle_eastern;north_african;biracial',
'white;indian;north_american;indigenous',
'white;african;multiracial',
'white;ethnoreligious_group;multiracial',
'indian;north_american;asian;south_american;south_asian',
'white;other;european;north_american;hispanic',
'white;central_american;hispanic;south_american',
'black;caribbean;central_american;hispanic',
'indian;north_american;multiracial',
'white;european;north_american;african;black;central_american;hispanic;south_america
n;multiracial;biracial;indigenous',
'white;asian;east_asian;biracial;pacific_islander',
'white;north_american;asian;hispanic;multiracial',
'white;european;south_asian;multiracial;biracial',
'white;european;asian;multiracial;biracial',
'white;european;north_american;asian;hispanic;multiracial;indigenous',
'european;african;black;indigenous',
'asian;east_asian;central_american',
'white;indian;european;multiracial',
'european;african;multiracial',
```

```
'caribbean;central_american;hispanic;multiracial',
'other;hispanic;south_american;multiracial',
'asian;black;caribbean;central_american;hispanic;multiracial',
'white;indian;european;north_american;middle_eastern;ethnoreligious_group;african;ea
st_asian;caribbean;southeast_asian;central_american;north_african;hispanic;south_amer
ican;south_asian;pacific_islander;central_asian',
'african;black;south_american',
'middle_eastern;south_american;multiracial',
'white;middle_eastern;multiracial;biracial',
'white;other;african',
'white;african;multiracial;biracial',
'white;european;hispanic;biracial;indigenous',
'european;caribbean;central_american;hispanic;multiracial',
'indian;pacific_islander',
'european;asian;south_asian;multiracial',
'african;hispanic;multiracial',
'african;asian;biracial;central_asian',
'white;other;black;indigenous',
'north_american;biracial',
'european;black;biracial;indigenous',
'white;north_american;ethnoreligious_group;multiracial',
'white;north_american;multiracial;biracial;indigenous',
'white;east_asian;pacific_islander',
'white;north_american;black;biracial',
'indian;north_american;asian',
'ethnoreligious_group;black;multiracial;indigenous',
'white;european;south_american;biracial',
'indian;middle_eastern;african;asian;east_asian;black;caribbean;southeast_asian;cent
ral_american;north_african;south_american;south_asian;multiracial;biracial;pacific_is
lander;central_asian',
'white;north_american;multiracial',
'white;other;north_american;hispanic;multiracial',
'european;middle_eastern;ethnoreligious_group;multiracial;biracial',
'hispanic;biracial;indigenous',
'white;asian;black;south_asian;multiracial',
'african;black;central_american',
'caribbean;central_american;multiracial',
'white;black;hispanic;multiracial;biracial',
'white;east_asian;indigenous',
'white;european;north_american;african;caribbean;multiracial',
'south_asian;biracial',
'indian;middle_eastern;south_asian;multiracial',
'white;other;european;north_american;multiracial;indigenous',
'african;asian;black;southeast_asian;pacific_islander',
'white;middle_eastern;asian;south_asian',
'white;other;european;ethnoreligious_group;asian;east_asian;hispanic;south_asian;mul
tiracial;biracial',
'white;european;middle_eastern;indigenous',
'asian;south_asian;central_asian',
'other;african;black',
'other;european;middle_eastern;ethnoreligious_group;hispanic',
'north_american;black;caribbean',
'middle_eastern;ethnoreligious_group;biracial;central_asian',
'asian;biracial',
'white;other;european;north_american;african;multiracial;indigenous',
'white;european;asian;multiracial;biracial;central_asian',
'european;south_american;biracial',
'other;asian;southeast_asian;south_asian',
'white;indian;middle_eastern;asian;multiracial',
'african;asian;black;multiracial',
```

```
'north_american;african;central_american;north_african;south_american',
'asian;biracial;pacific_islander',
'white;black;multiracial;pacific_islander',
'african;caribbean;hispanic;south_american',
'middle_eastern;ethnoreligious_group;african',
'other;african;caribbean',
'indian;asian;hispanic;south_american;south_asian;multiracial',
'white;middle_eastern;ethnoreligious_group;african;asian;black;southeast_asian;south
_amERICAN;south_asian;multiracial;biracial;pacific_islander',
'european;north_american;african;black;biracial',
'asian;central_american;hispanic;multiracial',
'caribbean;central_american;south_american',
'white;asian;southeast_asian;south_asian;multiracial;biracial',
'est_asian;southeast_asian;multiracial',
'white;other;european;asian',
'asian;east_asian;pacific_islander',
'white;european;ethnoreligious_group;central_asian',
'other;indian;ethnoreligious_group;asian',
'european;caribbean;hispanic;south_american',
'indian;asian;black;south_american',
'central_american;south_american',
'white;european;hispanic;multiracial;biracial',
'european;north_american;south_american;multiracial',
'european;biracial;central_asian',
'hispanic;multiracial;biracial',
'white;indian;european;asian;black;caribbean;south_asian;multiracial',
'white;other;european;north_american;ethnoreligious_group',
'white;south_american;multiracial;biracial',
'white;middle_eastern;ethnoreligious_group;hispanic;south_american;multiracial',
'white;north_american;east_asian;multiracial',
'white;european;asian;east_asian;southeast_asian;multiracial;biracial',
'north_american;south_asian',
'white;north_american;asian;southeast_asian;multiracial',
'asian;east_asian;southeast_asian;south_asian;biracial',
'african;south_asian',
'indian;african;east_asian',
'european;caribbean;biracial',
'indian;asian;multiracial;central_asian',
'european;middle_eastern;ethnoreligious_group;multiracial',
'est_asian;south_american;multiracial',
'indian;european;asian;east_asian;southeast_asian;south_asian;central_asian',
'south_asian;multiracial;biracial',
'european;middle_eastern;african;multiracial;biracial',
'indian;middle_eastern;multiracial',
'white;other;asian;east_asian;multiracial;biracial',
'white;middle_eastern;asian;east_asian;southeast_asian',
'european;african;asian',
'other;middle_eastern;caribbean;hispanic;central_asian',
'white;indian;north_american;asian;multiracial;biracial',
'european;north_american;asian;multiracial',
'other;ethnoreligious_group;african;black;multiracial;indigenous',
'white;indian;asian;south_asian',
'caribbean;central_american;hispanic;biracial',
'white;european;ethnoreligious_group;multiracial;indigenous',
'white;indian;pacific_islander',
'north_american;central_american;south_american;indigenous',
'white;other;indian;european;north_american;middle_eastern;ethnoreligious_group;afri
can;asian;east_asian;black;caribbean;southeast_asian;central_american;north_african;h
ispanic;south_american;south_asian;biracial;indigenous;pacific_islander;central_asia
n',
```

```
'indian;asian;southeast_asian;south_asian;multiracial',
'black;biracial;pacific_islander',
'european;asian;south_asian;biracial',
'white;european;african;black;caribbean;multiracial',
'indian;middle_eastern;asian;southeast_asian',
'european;north_american;east_asian',
'white;african;hispanic;multiracial',
'multiracial;indigenous',
'white;european;african;black;north_african;multiracial;biracial',
'white;european;north_american;asian;central_american;south_american;south_asian',
'white;north_american;asian;southeast_asian;multiracial;biracial',
'white;other;north_african',
'european;southeast_asian;multiracial',
'european;african;asian;black;multiracial',
'white;other;european;black',
'white;european;ethnoreligious_group;asian',
'white;other;north_american;ethnoreligious_group',
'black;hispanic;south_american;biracial',
'indian;african;black',
'indian;european;african;black;multiracial',
'white;european;north_american;hispanic',
'european;north_american;biracial',
'est_asian;hispanic;south_american;biracial',
'white;indian;european;north_american;asian;est_asian;multiracial;biracial',
'european;middle_eastern;caribbean',
'est_asian;hispanic;indigenous',
'european;asian;est_asian;biracial',
...]
```

## Observations:

- There're 1055 unique entries
- The entries are nested

## Steps:

- The categories will be renamed for clarity
- I will create a sum column and binary indicators
- The type can be made `category` for less memory usage

## Accessibility

In [222...]: `col_description("Accessibility")`

### Description:

Which of the following describe you, if any? Please check all that apply.

In [223...]: `df.Accessibility`

```
Out[223]: 0           NaN  
1           NaN  
2    None of the above  
3    None of the above  
4           NaN  
      ...  
73263  None of the above  
73264  None of the above  
73265  None of the above  
73266  None of the above  
73267  None of the above  
Name: Accessibility, Length: 73268, dtype: object
```

```
In [224... df.Accessibility.unique()
```

```
Out[224]: array([nan, 'None of the above', 'Or, in your own words:',  
   'I am deaf / hard of hearing', 'Prefer not to say',  
   'I am blind / have difficulty seeing',  
   'I am unable to / find it difficult to type',  
   'I am unable to / find it difficult to walk or stand without assistance',  
   'Or, in your own words:;I am blind / have difficulty seeing',  
   'I am unable to / find it difficult to type;I am unable to / find it difficult  
to walk or stand without assistance',  
   'I am deaf / hard of hearing;I am unable to / find it difficult to walk or sta  
nd without assistance',  
   'I am deaf / hard of hearing;I am blind / have difficulty seeing',  
   'I am deaf / hard of hearing;I am blind / have difficulty seeing;I am unable t  
o / find it difficult to type;I am unable to / find it difficult to walk or stand wit  
hout assistance',  
   'Or, in your own words:;I am deaf / hard of hearing;I am blind / have difficul  
ty seeing',  
   'I am blind / have difficulty seeing;I am unable to / find it difficult to wal  
k or stand without assistance',  
   'I am deaf / hard of hearing;I am unable to / find it difficult to type',  
   'Or, in your own words:;I am unable to / find it difficult to walk or stand wi  
thout assistance',  
   'I am blind / have difficulty seeing;I am unable to / find it difficult to typ  
e',  
   'Or, in your own words:;I am deaf / hard of hearing',  
   'Or, in your own words:;I am deaf / hard of hearing;I am blind / have difficul  
ty seeing;I am unable to / find it difficult to type;I am unable to / find it difficu  
lt to walk or stand without assistance',  
   'I am deaf / hard of hearing;I am blind / have difficulty seeing;I am unable t  
o / find it difficult to walk or stand without assistance',  
   'Or, in your own words:;I am blind / have difficulty seeing;I am unable to / f  
ind it difficult to type;I am unable to / find it difficult to walk or stand without  
assistance',  
   'I am deaf / hard of hearing;I am blind / have difficulty seeing;I am unable t  
o / find it difficult to type',  
   'Or, in your own words:;I am deaf / hard of hearing;I am blind / have difficul  
ty seeing;I am unable to / find it difficult to type',  
   'Or, in your own words:;I am unable to / find it difficult to type',  
   'Or, in your own words:;I am blind / have difficulty seeing;I am unable to / f  
ind it difficult to type',  
   'I am deaf / hard of hearing;I am unable to / find it difficult to type;I am u  
nable to / find it difficult to walk or stand without assistance',  
   'Or, in your own words:;I am unable to / find it difficult to type;I am unable  
to / find it difficult to walk or stand without assistance'],  
   dtype=object)
```

In [225...]

```
col_value_counts("Accessibility")
```

Out[225]:

		count	pct
Accessibility			
	<b>None of the above</b>	63064	0.937838
	<b>Prefer not to say</b>	1633	0.024285
	<b>I am blind / have difficulty seeing</b>	981	0.014589
	<b>Or, in your own words:</b>	579	0.008610
	<b>I am deaf / hard of hearing</b>	436	0.006484
	<b>I am unable to / find it difficult to walk or stand without assistance</b>	186	0.002766
	<b>I am unable to / find it difficult to type</b>	131	0.001948
	<b>I am deaf / hard of hearing;I am blind / have difficulty seeing</b>	48	0.000714
	<b>I am deaf / hard of hearing;I am blind / have difficulty seeing;I am unable to / find it difficult to type;I am unable to / find it difficult to walk or stand without assistance</b>	31	0.000461
	<b>Or, in your own words;;I am blind / have difficulty seeing</b>	30	0.000446
	<b>I am unable to / find it difficult to type;I am unable to / find it difficult to walk or stand without assistance</b>	25	0.000372
	<b>I am deaf / hard of hearing;I am unable to / find it difficult to walk or stand without assistance</b>	16	0.000238
	<b>I am blind / have difficulty seeing;I am unable to / find it difficult to walk or stand without assistance</b>	15	0.000223
	<b>I am blind / have difficulty seeing;I am unable to / find it difficult to type</b>	11	0.000164
	<b>Or, in your own words;;I am deaf / hard of hearing</b>	9	0.000134
	<b>Or, in your own words;;I am deaf / hard of hearing;I am blind / have difficulty seeing;I am unable to / find it difficult to type;I am unable to / find it difficult to walk or stand without assistance</b>	9	0.000134
	<b>Or, in your own words;;I am unable to / find it difficult to type</b>	8	0.000119
	<b>Or, in your own words;;I am unable to / find it difficult to walk or stand without assistance</b>	7	0.000104
	<b>I am deaf / hard of hearing;I am unable to / find it difficult to type</b>	6	0.000089
	<b>I am deaf / hard of hearing;I am blind / have difficulty seeing;I am unable to / find it difficult to walk or stand without assistance</b>	5	0.000074
	<b>I am deaf / hard of hearing;I am blind / have difficulty seeing;I am unable to / find it difficult to type</b>	5	0.000074
	<b>Or, in your own words;;I am deaf / hard of hearing;I am blind / have difficulty seeing</b>	3	0.000045
	<b>Or, in your own words;;I am blind / have difficulty seeing;I am unable to / find it difficult to type;I am unable to / find it difficult to walk or stand without assistance</b>	2	0.000030
	<b>Or, in your own words;;I am deaf / hard of hearing;I am blind / have difficulty seeing;I am unable to / find it difficult to type</b>	1	0.000015
	<b>Or, in your own words;;I am blind / have difficulty seeing;I am unable to / find it difficult to type</b>	1	0.000015

	count	pct
Accessibility		
I am deaf / hard of hearing;I am unable to / find it difficult to type;I am unable to / find it difficult to walk or stand without assistance	1	0.000015
Or, in your own words;I am unable to / find it difficult to type;I am unable to / find it difficult to walk or stand without assistance	1	0.000015

In [226]: `col_strip_nested("Accessibility")`

### Unique Values:

Out[226]:

```
[('None of the above', 63064),
 ('Prefer not to say', 1633),
 ('I am blind / have difficulty seeing', 1142),
 ('Or, in your own words:', 650),
 ('I am deaf / hard of hearing', 570),
 ('I am unable to / find it difficult to walk or stand without assistance',
  298),
 ('I am unable to / find it difficult to type', 232)]
```

In [227]:

```
(df
 .Accessibility
 .fillna("unanswered")
 .replace({"Prefer not to say": "unanswered",
            "None of the above": "none_of_these"})
 .str.replace("none of the above", "none")
 .str.replace("I am blind / have difficulty seeing", "blind")
 .str.replace("Or, in your own words:", "other")
 .str.replace("I am deaf / hard of hearing", "deaf")
 .str.replace("I am unable to / find it difficult to walk or stand without assistance",
             "cant_type")
 .value_counts())
```

```
Out[227]:
```

none_of_these	63064
unanswered	7657
blind	981
other	579
deaf	436
cant_walk	186
cant_type	131
deaf;blind	48
deaf;blind;cant_type;cant_walk	31
other;blind	30
cant_type;cant_walk	25
deaf;cant_walk	16
blind;cant_walk	15
blind;cant_type	11
other;deaf	9
other;deaf;blind;cant_type;cant_walk	9
other;cant_type	8
other;cant_walk	7
deaf;cant_type	6
deaf;blind;cant_walk	5
deaf;blind;cant_type	5
other;deaf;blind	3
other;blind;cant_type;cant_walk	2
other;deaf;blind;cant_type	1
other;blind;cant_type	1
deaf;cant_type;cant_walk	1
other;cant_type;cant_walk	1

Name: Accessibility, dtype: int64

### Observations:

- There're 27 unique entries
- The entries are nested

### Steps:

- The categories will be renamed for clarity
- I will create a sum column and binary indicators
- The type can be made `category` for less memory usage

## MentalHealth

```
In [228...]: col_description("MentalHealth")
```

### Description:

Which of the following describe you, if any? Please check all that apply.

```
In [229...]: df.MentalHealth
```

```
Out[229]: 0          NaN
           1          NaN
           2    I have a mood or emotional disorder (e.g., dep...
           3          None of the above
           4          NaN
           ...
           73263      None of the above
           73264      None of the above
           73265      None of the above
           73266      None of the above
           73267      None of the above
Name: MentalHealth, Length: 73268, dtype: object
```

```
In [230... col_value_counts("MentalHealth")#.shape
```

Out[230]:

		count	pct
MentalHealth			
	<b>None of the above</b>	46849	0.705058
	<b>Prefer not to say</b>	3435	0.051695
	<b>I have a concentration and/or memory disorder (e.g., ADHD, etc.)</b>	2936	0.044186
	<b>I have an anxiety disorder</b>	2404	0.036179
	<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.)</b>	1914	0.028805
	<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder</b>	1495	0.022499
	<b>I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	1025	0.015426
	<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;I have a concentration and/or memory disorder (e.g., ADHD, etc.)</b>	995	0.014974
	<b>I have learning differences (e.g., Dyslexic, Dyslexia, etc.)</b>	833	0.012536
	<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have a concentration and/or memory disorder (e.g., ADHD, etc.)</b>	677	0.010189
	<b>I have an anxiety disorder;I have a concentration and/or memory disorder (e.g., ADHD, etc.)</b>	640	0.009632
	<b>Or, in your own words:</b>	599	0.009015
	<b>I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	427	0.006426
	<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	297	0.004470
	<b>I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.)</b>	228	0.003431
	<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	162	0.002438
	<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	160	0.002408
	<b>I have an anxiety disorder;I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	152	0.002288
	<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	142	0.002137
	<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.)</b>	135	0.002032
	<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	115	0.001731

	count	pct
MentalHealth		
I have an anxiety disorder;I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)	114	0.001716
I have an anxiety disorder;I have learning differences (e.g., Dyslexic, Dyslexia, etc.)	76	0.001144
I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.)	67	0.001008
I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.)	62	0.000933
I have an anxiety disorder;I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.)	61	0.000918
I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;I have learning differences (e.g., Dyslexic, Dyslexia, etc.)	50	0.000752
I have learning differences (e.g., Dyslexic, Dyslexia, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)	49	0.000737
I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)	41	0.000617
Or, in your own words:;I have a concentration and/or memory disorder (e.g., ADHD, etc.)	39	0.000587
I have an anxiety disorder;Or, in your own words:	28	0.000421
I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);Or, in your own words:	27	0.000406
I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)	26	0.000391
I have an anxiety disorder;I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)	20	0.000301
I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;Or, in your own words:;I have a concentration and/or memory disorder (e.g., ADHD, etc.)	19	0.000286
I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;Or, in your own words:;I have a concentration and/or memory disorder (e.g., ADHD, etc.)	19	0.000286
I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;I have learning differences (e.g., Dyslexic, Dyslexia, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)	18	0.000271
I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)	17	0.000256
I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;Or, in your own words:;I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)	12	0.000181

	count	pct
<b>MentalHealth</b>		
<b>Or, in your own words::I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	11	0.000166
<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;Or, in your own words::I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	10	0.000150
<b>I have an anxiety disorder;I have learning differences (e.g., Dyslexic, Dyslexia, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	10	0.000150
<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);Or, in your own words::I have a concentration and/or memory disorder (e.g., ADHD, etc.)</b>	7	0.000105
<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;Or, in your own words::I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.)</b>	6	0.000090
<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);Or, in your own words::I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	5	0.000075
<b>Or, in your own words::I have learning differences (e.g., Dyslexic, Dyslexia, etc.)</b>	5	0.000075
<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);Or, in your own words::I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	5	0.000075
<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;Or, in your own words::I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	5	0.000075
<b>Or, in your own words::I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	4	0.000060
<b>Or, in your own words::I have learning differences (e.g., Dyslexic, Dyslexia, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	3	0.000045
<b>I have an anxiety disorder;Or, in your own words::I have a concentration and/or memory disorder (e.g., ADHD, etc.)</b>	3	0.000045
<b>I have an anxiety disorder;Or, in your own words::I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	2	0.000030
<b>Or, in your own words::I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.);I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)</b>	2	0.000030
<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);I have an anxiety disorder;Or, in your own words::I have learning differences (e.g., Dyslexic, Dyslexia, etc.)</b>	1	0.000015
<b>I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.);Or, in your own words::I have learning differences (e.g., Dyslexic, Dyslexia, etc.)</b>	1	0.000015
<b>I have an anxiety disorder;Or, in your own words::I have learning differences (e.g., Dyslexic, Dyslexia, etc.)</b>	1	0.000015
<b>Or, in your own words::I have a concentration and/or memory disorder (e.g., ADHD, etc.);I have learning differences (e.g., Dyslexic, Dyslexia, etc.)</b>	1	0.000015

```
In [231]: col_strip_nested("MentalHealth")
```

## Unique Values:

```
Out[231]: [('None of the above', 46849),  
 ('I have a concentration and/or memory disorder (e.g., ADHD, etc.)', 7026),  
 ('I have an anxiety disorder', 6848),  
 ('I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.)',  
 6449),  
 ('Prefer not to say', 3435),  
 ("I have autism / an autism spectrum disorder (e.g. Asperger's, etc.)", 2834),  
 ('I have learning differences (e.g., Dyslexic, Dyslexia, etc.)', 1840),  
 ('Or, in your own words:', 815)]
```

```
In [232]:
```

```
(df  
    .MentalHealth  
    .fillna("unanswered")  
    .replace({"None of the above": "none_of_these",  
              "Prefer not to say": "unanswered"})  
    .str.replace("Or, in your own words:", "other")  
    .str.replace("I have an anxiety disorder", "anxiety_disorder")  
    .str.replace("I have a mood or emotional disorder \n(e.g., depression, bipolar disorder)",  
    "mood_or_emotional_disorder")  
    .str.replace("I have a concentration and/or memory disorder \n(e.g., ADHD, etc.\n)", "concentration_and_memory_disorder")  
    .str.replace("I have autism / an autism spectrum disorder \n(e.g. Asperger's, etc.\n)", "autism_spectrum_disorder")  
    .str.replace("I have learning differences \n(e.g., Dyslexic, Dyslexia, etc.\n)", "learning_differences")  
    .unique()  
    .tolist())
```

```
Out[232]: ['unanswered',
 'emotional_disorder;anxiety_disorder',
 'none_of_these',
 'other',
 'emotional_disorder',
 'emotional_disorder;anxiety_disorder;memory_disorder',
 'memory_disorder;dyslexia',
 'anxiety_disorder',
 'autism',
 'dyslexia',
 'memory_disorder',
 'anxiety_disorder;memory_disorder',
 'memory_disorder;dyslexia;autism',
 'emotional_disorder;memory_disorder;autism',
 'anxiety_disorder;other',
 'emotional_disorder;anxiety_disorder;memory_disorder;autism',
 'emotional_disorder;autism',
 'memory_disorder;autism',
 'other;memory_disorder',
 'emotional_disorder;anxiety_disorder;memory_disorder;dyslexia',
 'emotional_disorder;anxiety_disorder;autism',
 'emotional_disorder;memory_disorder',
 'anxiety_disorder;memory_disorder;dyslexia;autism',
 'anxiety_disorder;memory_disorder;autism',
 'anxiety_disorder;autism',
 'anxiety_disorder;dyslexia',
 'emotional_disorder;anxiety_disorder;dyslexia',
 'other;autism',
 'emotional_disorder;dyslexia',
 'dyslexia;autism',
 'emotional_disorder;anxiety_disorder;memory_disorder;dyslexia;autism',
 'emotional_disorder;anxiety_disorder;other',
 'emotional_disorder;other',
 'emotional_disorder;dyslexia;autism',
 'emotional_disorder;memory_disorder;dyslexia',
 'anxiety_disorder;memory_disorder;dyslexia',
 'emotional_disorder;anxiety_disorder;dyslexia;autism',
 'emotional_disorder;anxiety_disorder;other;memory_disorder',
 'emotional_disorder;anxiety_disorder;other;memory_disorder;autism',
 'emotional_disorder;memory_disorder;dyslexia;autism',
 'emotional_disorder;other;autism',
 'emotional_disorder;other;memory_disorder',
 'emotional_disorder;anxiety_disorder;other;autism',
 'emotional_disorder;other;dyslexia',
 'anxiety_disorder;dyslexia;autism',
 'emotional_disorder;anxiety_disorder;other;memory_disorder;dyslexia;autism',
 'emotional_disorder;other;memory_disorder;autism',
 'other;dyslexia;autism',
 'anxiety_disorder;other;memory_disorder;autism',
 'emotional_disorder;anxiety_disorder;other;dyslexia',
 'other;dyslexia',
 'other;memory_disorder;autism',
 'anxiety_disorder;other;dyslexia',
 'emotional_disorder;anxiety_disorder;other;memory_disorder;dyslexia',
 'other;memory_disorder;dyslexia;autism',
 'anxiety_disorder;other;memory_disorder',
 'other;memory_disorder;dyslexia']
```

In [233...]

```
(df
 .MentalHealth
```

```
.str.lower()
.replace({np.nan: "unanswered"})
.loc[lambda ser: ser.str.contains("none")]
.value_counts()
```

Out[233]: none of the above 46849  
Name: MentalHealth, dtype: int64

### Observations:

- There're 57 unique entries
- The entries are nested

### Steps:

- The categories will be renamed for clarity

## TBranch

In [234... col\_description("TBranch")

### Description:

<span style="font-size:16px;">Would you like to participate in the Professional Development Series?</span>

In [235... df.TBranch

Out[235]:

0	NaN
1	No
2	No
3	No
4	No
...	
73263	Yes
73264	Yes
73265	Yes
73266	No
73267	NaN

Name: TBranch, Length: 73268, dtype: object

In [236... df.TBranch.unique()

Out[236]: array([nan, 'No', 'Yes'], dtype=object)

In [237... col\_value\_counts("TBranch")

Out[237]:

	count	pct
<b>TBranch</b>		
<b>Yes</b>	37200	0.706284
<b>No</b>	15470	0.293716

### Observations:

- There're 2 unique entries
- The entries are clean and valid

### Steps:

- No cleaning required
- The type can be made `category` for less memory usage

## ICorPM

In [238]: `col_description("ICorPM")`

### Description:

Are you an independent contributor or people manager?

In [239]: `df.ICorPM`

```
Out[239]: 0           NaN
           1           NaN
           2           NaN
           3           NaN
           4           NaN
...
73263  Independent contributor
73264  Independent contributor
73265  Independent contributor
73266           NaN
73267           NaN
Name: ICorPM, Length: 73268, dtype: object
```

In [240...]: `df.ICorPM.unique()`

```
Out[240]: array([nan, 'Independent contributor', 'People manager'], dtype=object)
```

In [241...]: `col_value_counts("ICorPM")`

	count	pct
<b>ICorPM</b>		
<b>Independent contributor</b>	30592	0.84315
<b>People manager</b>	5691	0.15685

### Observations:

- There're 2 unique entries
- The entries are clean and valid

### Steps:

- No cleaning required

- The type can be made `category` for less memory usage

## WorkExp

In [242...]: `col_description("WorkExp")`

### Description:

How many years of working experience do you have?

In [243...]: `df.WorkExp`

```
Out[243]: 0      NaN
           1      NaN
           2      NaN
           3      NaN
           4      NaN
           ...
          73263    5.0
          73264    6.0
          73265   42.0
          73266    NaN
          73267    NaN
Name: WorkExp, Length: 73268, dtype: float64
```

In [244...]: `df.WorkExp.describe()`

```
Out[244]: count    36769.000000
           mean     10.242378
           std      8.706850
           min      0.000000
           25%     4.000000
           50%     8.000000
           75%    15.000000
           max     50.000000
Name: WorkExp, dtype: float64
```

In [245...]: `(df
 .query("WorkExp == 0")
 .Age
 .unique())`

```
Out[245]: array(['18-24 years old', '25-34 years old', '35-44 years old',
       '65 years or older', nan, 'Under 18 years old',
       'Prefer not to say', '45-54 years old'], dtype=object)
```

### Observations:

- The entries are clean and valid

### Steps:

- No cleaning required

## Knowledge\_1

```
In [246...]: col_description("Knowledge_1")
```

### Description:

I have interactions with people outside of my immediate team.

```
In [247...]: df.Knowledge_1
```

```
Out[247]: 0      NaN  
1      NaN  
2      NaN  
3      NaN  
4      NaN  
      ...  
73263    Agree  
73264    Agree  
73265  Disagree  
73266      NaN  
73267      NaN  
Name: Knowledge_1, Length: 73268, dtype: object
```

```
In [248...]: df.Knowledge_1.unique()
```

```
Out[248]: array([nan, 'Agree', 'Strongly agree', 'Neither agree nor disagree',  
   'Strongly disagree', 'Disagree'], dtype=object)
```

### Observations:

- There're 5 unique entries
- The entries are clean and valid

### Steps:

- No cleaning required
- The type can be made `category` for less memory usage

## Knowledge\_2

```
In [249...]: col_description("Knowledge_2")
```

### Description:

Knowledge silos prevent me from getting ideas across the organization (i.e., one individual or team has information that isn't shared with others)

```
In [250...]: df.Knowledge_2
```

```
Out[250]: 0          NaN
           1          NaN
           2          NaN
           3          NaN
           4          NaN
           ...
           73263      Disagree
           73264      Agree
           73265  Neither agree nor disagree
           73266      NaN
           73267      NaN
Name: Knowledge_2, Length: 73268, dtype: object
```

In [251... df.Knowledge\_2.unique()

```
Out[251]: array([nan, 'Disagree', 'Agree', 'Neither agree nor disagree',
   'Strongly agree', 'Strongly disagree'], dtype=object)
```

### Observations:

- There're 5 unique entries
- The entries are clean and valid

### Steps:

- No cleaning required
- The type can be made `category` for less memory usage

## Knowledge\_3

In [252... col\_description("Knowledge\_3")

### Description:

I can find up-to-date information within my organization to help me do my job.

In [253... df.Knowledge\_3.unique()

```
Out[253]: array([nan, 'Agree', 'Disagree', 'Strongly agree',
   'Neither agree nor disagree', 'Strongly disagree'], dtype=object)
```

### Observations:

- There're 5 unique entries
- The entries are clean and valid

### Steps:

- No cleaning required
- The type can be made `category` for less memory usage

## Knowledge\_4

```
In [254... col_description("Knowledge_4")]
```

### Description:

I am able to quickly find answers to my questions with existing tools and resources.

```
In [255... df.Knowledge_4.unique())
```

```
Out[255]: array([nan, 'Agree', 'Strongly agree', 'Neither agree nor disagree',  
   'Disagree', 'Strongly disagree'], dtype=object)
```

### Observations:

- There're 5 unique entries
- The entries are clean and valid

### Steps:

- No cleaning required
- The type can be made `category` for less memory usage

## Knowledge\_5

```
In [256... col_description("Knowledge_5")]
```

### Description:

I know which system or resource to use to find information and answers to questions I have.

```
In [257... df.Knowledge_5.unique())
```

```
Out[257]: array([nan, 'Agree', 'Strongly agree', 'Neither agree nor disagree',  
   'Disagree', 'Strongly disagree'], dtype=object)
```

### Observations:

- There're 5 unique entries
- The entries are clean and valid

### Steps:

- No cleaning required
- The type can be made `category` for less memory usage

## Knowledge\_6

```
In [258... col_description("Knowledge_6")]
```

### Description:

I often find myself answering questions that I've already answered before.

```
In [259... df.Knowledge_6.unique()
```

```
Out[259]: array([nan, 'Agree', 'Neither agree nor disagree', 'Strongly agree', 'Disagree', 'Strongly disagree'], dtype=object)
```

### Observations:

- There're 5 unique entries
- The entries are clean and valid

### Steps:

- No cleaning required
- The type can be made `category` for less memory usage

## Knowledge\_7

```
In [260... col_description("Knowledge_7")
```

### Description:

Waiting on answers to questions often causes interruptions and disrupts my workflow.

```
In [261... df.Knowledge_7.unique()
```

```
Out[261]: array([nan, 'Disagree', 'Agree', 'Neither agree nor disagree', 'Strongly agree', 'Strongly disagree'], dtype=object)
```

### Observations:

- There're 5 unique entries
- The entries are clean and valid

### Steps:

- No cleaning required
- The type can be made `category` for less memory usage

## Frequency\_1

```
In [262... col_description("Frequency_1")
```

### Description:

Needing help from people outside of your immediate team?

```
In [263... df.Frequency_1.unique()
```

```
Out[263]: array([nan, '3-5 times a week', '10+ times a week', 'Never', '1-2 times a week', '6-10 times a week'], dtype=object)
```

In [264]: `col_value_counts("Frequency_1")`

	count	pct
<b>Frequency_1</b>		
<b>1-2 times a week</b>	21689	0.613186
<b>Never</b>	8754	0.247491
<b>3-5 times a week</b>	3489	0.098640
<b>6-10 times a week</b>	796	0.022504
<b>10+ times a week</b>	643	0.018179

In [265]:

```
(df
    .Frequency_1
    .str.lower()
    .fillna("unanswered")
    .replace({"1-2 times a week": "rarely",
              "3-5 times a week": "mildly",
              "6-10 times a week": "frequently",
              "10+ times a week": "frequently"})
    .astype("category"))
```

Out[265]:

0	unanswered
1	unanswered
2	unanswered
3	unanswered
4	unanswered
...	
73263	never
73264	rarely
73265	never
73266	unanswered
73267	unanswered

Name: Frequency\_1, Length: 73268, dtype: category  
Categories (5, object): ['frequently', 'mildly', 'never', 'rarely', 'unanswered']

## Observations:

- There're 5 unique entries
- The entries are clean and valid

## Steps:

- The categories will be renamed for clarity
- Type will be made `category`

## Frequency\_2

In [266]: `col_description("Frequency_2")`

## Description:

Interacting with people outside of your immediate team?

In [267... df.Frequency\_2.unique()

Out[267]: array([nan, '3-5 times a week', '10+ times a week', 'Never', '6-10 times a week', '1-2 times a week'], dtype=object)

### Observations:

- There're 5 unique entries
- The entries are clean and valid

### Steps:

- The categories will be renamed for clarity
- Type will be made category

## Frequency\_3

In [268... col\_description("Frequency\_3")

### Description:

Encountering knowledge silos (where one individual or team has information that's not shared or distributed with other individuals or teams) at work?

In [269... df.Frequency\_3.unique()

Out[269]: array([nan, 'Never', '3-5 times a week', '1-2 times a week', '6-10 times a week', '10+ times a week'], dtype=object)

### Observations:

- There're 5 unique entries
- The entries are clean and valid

### Steps:

- The categories will be renamed for clarity
- Type will be made category

## TimeSearching

In [270... col\_description("TimeSearching")

### Description:

On an average day, how much time do you typically spend searching for answers or solutions to problems you encounter at work? (This includes time spent searching on your own, asking a colleague, and waiting for a response).

In [271... df.TimeSearching

```
Out[271]: 0          NaN
           1          NaN
           2          NaN
           3          NaN
           4          NaN
           ...
           73263    30-60 minutes a day
           73264    15-30 minutes a day
           73265    30-60 minutes a day
           73266          NaN
           73267          NaN
Name: TimeSearching, Length: 73268, dtype: object
```

In [272... df.TimeSearching.unique()

```
Out[272]: array([nan, '15-30 minutes a day', '30-60 minutes a day',
       '60-120 minutes a day', 'Less than 15 minutes a day',
       'Over 120 minutes a day'], dtype=object)
```

In [273... col\_value\_counts("TimeSearching")

	count	pct
<b>TimeSearching</b>		
<b>30-60 minutes a day</b>	13652	0.377148
<b>15-30 minutes a day</b>	10122	0.279629
<b>60-120 minutes a day</b>	6371	0.176004
<b>Less than 15 minutes a day</b>	3528	0.097464
<b>Over 120 minutes a day</b>	2525	0.069755

In [274... df
 .TimeSearching
 .fillna("unanswered")
 .str.lower()
 .str.replace(" minutes a day", "")
 .replace({"less than 15": "quarter",
 "15-30": "half",
 "30-60": "one",
 "60-120": "two",
 "over\_120": "above\_two"})
 .astype("category"))

```
Out[274]: 0      unanswered
           1      unanswered
           2      unanswered
           3      unanswered
           4      unanswered
           ...
           73263    one
           73264    half
           73265    one
           73266  unanswered
           73267  unanswered
Name: TimeSearching, Length: 73268, dtype: category
Categories (6, object): ['half', 'one', 'over 120', 'quarter', 'two', 'unanswered']
```

## Observations:

- There're 5 unique entries
- The entries are clean and valid

## Steps:

- The categories will be renamed for clarity
- Type will be made `category`

## TimeAnswering

```
In [275...]: col_description("TimeAnswering")
```

### Description:

On an average day, how much time do you typically spend answering questions you get asked at work?

```
In [276...]: df.TimeAnswering
```

```
Out[276]: 0           NaN
1           NaN
2           NaN
3           NaN
4           NaN
...
73263     Less than 15 minutes a day
73264     60-120 minutes a day
73265     60-120 minutes a day
73266           NaN
73267           NaN
Name: TimeAnswering, Length: 73268, dtype: object
```

```
In [277...]: df.TimeAnswering.unique()
```

```
Out[277]: array([nan, 'Over 120 minutes a day', '60-120 minutes a day',
   'Less than 15 minutes a day', '30-60 minutes a day',
   '15-30 minutes a day'], dtype=object)
```

## Observations:

- There're 5 unique entries
- The entries are clean and valid

## Steps:

- The categories will be renamed for clarity
- Type will be made `category`

## Onboarding

In [278]: `col_description("Onboarding")`

### Description:

The time it takes to onboard new hires at my company is:

In [279]: `df.Onboarding`

```
Out[279]: 0           NaN
1           NaN
2           NaN
3           NaN
4           NaN
...
73263    Just right
73264    Very long
73265    Just right
73266        NaN
73267        NaN
Name: Onboarding, Length: 73268, dtype: object
```

In [280]: `df.Onboarding.unique()`

```
Out[280]: array([nan, 'Somewhat long', 'Just right', 'Somewhat short', 'Very short',
   'Very long'], dtype=object)
```

In [281]: `col_value_counts("Onboarding")`

	count	pct
<b>Onboarding</b>		
<b>Somewhat long</b>	12961	0.363267
<b>Just right</b>	12526	0.351075
<b>Somewhat short</b>	4434	0.124275
<b>Very long</b>	4352	0.121977
<b>Very short</b>	1406	0.039407

In [282]:

```
(df
 .Onboarding
 .fillna("unanswered")
 .str.lower()
 .str.replace(" ", "_")
 .str.replace("somewhat_", "")
 .astype("category")
 .value_counts())
```

```
Out[282]: unanswered    37589
long          12961
just_right    12526
short          4434
very_long      4352
very_short     1406
Name: Onboarding, dtype: int64
```

### Observations:

- There're 5 unique entries
- The entries are clean and valid

## Steps:

- The categories will be renamed for clarity
- Type will be made `category`

## ProfessionalTech

```
In [283...]: col_description("ProfessionalTech")
```

### Description:

My company has:

```
In [284...]: df.ProfessionalTech
```

```
Out[284]: 0                               NaN  
1                               NaN  
2                               NaN  
3                               NaN  
4                               NaN  
...  
73263  DevOps function;Microservices;Developer portal...  
73264                      None of these  
73265                      None of these  
73266                           NaN  
73267                           NaN  
Name: ProfessionalTech, Length: 73268, dtype: object
```

```
In [285...]: (df  
    .ProfessionalTech  
    .unique()  
    .tolist())
```

Out[285]: [nan,  
 'Innersource initiative;DevOps function;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
 'Innersource initiative;DevOps function;Microservices;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
 'DevOps function;Microservices',  
 'Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
 'DevOps function;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
 'None of these',  
 'DevOps function;Microservices;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
 'Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
 'DevOps function;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
 'DevOps function',  
 'DevOps function;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
 'Microservices',  
 'DevOps function;Microservices;Developer portal or other central places to find tools/services;Automated testing;Observability tools',  
 'Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
 'DevOps function;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
 'DevOps function;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
 'Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
 'Microservices;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
 'Microservices;Continuous integration (CI) and (more often) continuous delivery',  
 'DevOps function;Microservices;Continuous integration (CI) and (more often) continuous delivery',  
 'Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
 'Automated testing',  
 'DevOps function;Microservices;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
 'DevOps function;Developer portal or other central places to find tools/services;Observability tools',  
 'Innersource initiative;DevOps function;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
 'DevOps function;Continuous integration (CI) and (more often) continuous delivery',  
 'Innersource initiative;DevOps function;Developer portal or other central places to find tools/services',  
 'Microservices;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
 'DevOps function;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery',  
 'Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
 'DevOps function;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing']

testing',  
'Microservices;Automated testing;Observability tools',  
'DevOps function;Microservices;Automated testing',  
'Continuous integration (CI) and (more often) continuous delivery',  
'DevOps function;Microservices;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
'DevOps function;Microservices;Automated testing;Observability tools',  
'Innersource initiative;DevOps function;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
'Developer portal or other central places to find tools/services',  
'Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery',  
'Innersource initiative;DevOps function;Microservices;Automated testing',  
'Developer portal or other central places to find tools/services;Automated testing;Observability tools',  
'Innersource initiative;DevOps function;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
'DevOps function;Automated testing',  
'DevOps function;Developer portal or other central places to find tools/services',  
'Automated testing;None of these',  
'DevOps function;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Microservices;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Innersource initiative;DevOps function;Microservices;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
'DevOps function;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Innersource initiative;DevOps function;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
'DevOps function;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Innersource initiative;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
'DevOps function;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery',  
'Developer portal or other central places to find tools/services;Observability tools',  
'Microservices;Observability tools',  
'Innersource initiative;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
'Innersource initiative;DevOps function',  
'Microservices;Developer portal or other central places to find tools/services;Automated testing',  
'Microservices;Automated testing',  
'Developer portal or other central places to find tools/services;Automated testing',  
'Innersource initiative;DevOps function;Continuous integration (CI) and (more often) continuous delivery',  
'Innersource initiative;DevOps function;Microservices;Observability tools',  
'Microservices;Developer portal or other central places to find tools/services',  
'Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery',

'DevOps function;Microservices;Developer portal or other central places to find tools/services',  
'Observability tools',  
'DevOps function;Observability tools',  
'Innersource initiative;DevOps function;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Innersource initiative;Microservices;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
'Innersource initiative',  
'Innersource initiative;Microservices',  
'DevOps function;Microservices;Developer portal or other central places to find tools/services;Automated testing',  
'Innersource initiative;DevOps function;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
'Innersource initiative;DevOps function;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery',  
'Innersource initiative;DevOps function;Microservices;Automated testing;Observability tools',  
'Innersource initiative;DevOps function;Automated testing',  
'DevOps function;Developer portal or other central places to find tools/services;Automated testing',  
'Innersource initiative;DevOps function;Microservices;Developer portal or other central places to find tools/services',  
'DevOps function;Developer portal or other central places to find tools/services;Automated testing;Observability tools',  
'DevOps function;Automated testing;Observability tools',  
'Innersource initiative;Automated testing',  
'Innersource initiative;DevOps function;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery',  
'Innersource initiative;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
'Innersource initiative;Microservices;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Innersource initiative;DevOps function;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Innersource initiative;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
'Innersource initiative;DevOps function;Microservices;Continuous integration (CI) and (more often) continuous delivery',  
'Innersource initiative;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
'Innersource initiative;DevOps function;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Innersource initiative;DevOps function;Microservices',  
'Innersource initiative;Developer portal or other central places to find tools/services',  
'DevOps function;Microservices;Observability tools',  
'Innersource initiative;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools',  
'Microservices;Developer portal or other central places to find tools/services;Automated testing;Observability tools',  
'Automated testing;Observability tools',

'Microservices;Developer portal or other central places to find tools/services;Observability tools',  
'Innersource initiative;DevOps function;Developer portal or other central places to find tools/services;Automated testing',  
'Innersource initiative;Microservices;Developer portal or other central places to find tools/services;Automated testing',  
'Innersource initiative;Microservices;Continuous integration (CI) and (more often) continuous delivery',  
'Innersource initiative;Microservices;Observability tools',  
'Innersource initiative;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery',  
'Innersource initiative;DevOps function;Microservices;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Innersource initiative;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery',  
'Innersource initiative;Continuous integration (CI) and (more often) continuous delivery',  
'Innersource initiative;DevOps function;Microservices;Developer portal or other central places to find tools/services;Automated testing;Observability tools',  
'Innersource initiative;DevOps function;Automated testing;Observability tools',  
'Innersource initiative;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Continuous integration (CI) and (more often) continuous delivery;None of these',  
'DevOps function;Microservices;Developer portal or other central places to find tools/services;Observability tools',  
'Innersource initiative;Observability tools',  
'Innersource initiative;Microservices;Automated testing',  
'Innersource initiative;DevOps function;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools;None of these',  
'Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;None of these',  
'Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Innersource initiative;DevOps function;Developer portal or other central places to find tools/services;Observability tools',  
'Innersource initiative;Microservices;Developer portal or other central places to find tools/services',  
'Innersource initiative;Continuous integration (CI) and (more often) continuous delivery;Observability tools',  
'Innersource initiative;Microservices;Developer portal or other central places to find tools/services;Observability tools',  
'Developer portal or other central places to find tools/services;None of these',  
'Innersource initiative;Microservices;Continuous integration (CI) and (more often) continuous delivery;Automated testing',  
'Microservices;None of these',  
'Innersource initiative;Developer portal or other central places to find tools/services;Observability tools',  
'Observability tools;None of these',  
'Innersource initiative;DevOps function;Microservices;Developer portal or other central places to find tools/services;Observability tools',  
'Innersource initiative;DevOps function;Microservices;Developer portal or other central places to find tools/services;Automated testing',  
'Innersource initiative;Automated testing;Observability tools',  
'DevOps function;None of these',  
'Innersource initiative;DevOps function;Observability tools',  
'Innersource initiative;Developer portal or other central places to find tools/services;Automated testing;Observability tools',  
'Microservices;Continuous integration (CI) and (more often) continuous delivery;None of these'

```
of these',
 'Innersource initiative;Developer portal or other central places to find tools/services;Automated testing',
 'Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;None of these',
 'Innersource initiative;DevOps function;Developer portal or other central places to find tools/services;Automated testing;Observability tools',
 'Innersource initiative;Microservices;Developer portal or other central places to find tools/services;Automated testing;Observability tools',
 'Microservices;Automated testing;None of these',
 'Innersource initiative;Developer portal or other central places to find tools/services;Automated testing;Observability tools;None of these',
 'Innersource initiative;None of these',
 'DevOps function;Developer portal or other central places to find tools/services;None of these',
 'Innersource initiative;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Observability tools',
 'Continuous integration (CI) and (more often) continuous delivery;Automated testing;None of these',
 'Innersource initiative;Microservices;Developer portal or other central places to find tools/services;None of these',
 'DevOps function;Observability tools;None of these',
 'DevOps function;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools;None of these',
 'Innersource initiative;Microservices;Automated testing;Observability tools',
 'Innersource initiative;DevOps function;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;None of these',
 'DevOps function;Continuous integration (CI) and (more often) continuous delivery;Automated testing;None of these',
 'DevOps function;Microservices;Automated testing;None of these',
 'Innersource initiative;Microservices;Automated testing;Observability tools;None of these',
 'Innersource initiative;Microservices;Continuous integration (CI) and (more often) continuous delivery;None of these',
 'DevOps function;Continuous integration (CI) and (more often) continuous delivery;None of these',
 'Innersource initiative;DevOps function;None of these',
 'DevOps function;Microservices;None of these',
 'Innersource initiative;DevOps function;Microservices;Continuous integration (CI) and (more often) continuous delivery;None of these']
```

In [286...]

col\_value\_counts("ProfessionalTech")

Out[286]:

		count	pct
ProfessionalTech			
	<b>None of these</b>	4658	0.133444
<b>DevOps function;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools</b>		2421	0.069358
<b>Innersource initiative;DevOps function;Microservices;Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools</b>		2414	0.069157
<b>DevOps function;Microservices;Continuous integration (CI) and (more often) continuous delivery;Automated testing;Observability tools</b>		1915	0.054862
<b>DevOps function;Microservices;Continuous integration (CI) and (more often) continuous delivery;Automated testing</b>		1575	0.045121
	...	...	...
<b>Microservices;Continuous integration (CI) and (more often) continuous delivery;None of these</b>		1	0.000029
<b>Innersource initiative;Developer portal or other central places to find tools/services;Automated testing;Observability tools;None of these</b>		1	0.000029
<b>Microservices;Automated testing;None of these</b>		1	0.000029
<b>Developer portal or other central places to find tools/services;Continuous integration (CI) and (more often) continuous delivery;None of these</b>		1	0.000029
<b>Innersource initiative;DevOps function;Microservices;Continuous integration (CI) and (more often) continuous delivery;None of these</b>		1	0.000029

155 rows × 2 columns

In [287...]

col\_strip\_nested("ProfessionalTech")

### Unique Values:

Out[287]:

```
[('Continuous integration (CI) and (more often) continuous delivery', 24361),
 ('DevOps function', 20716),
 ('Automated testing', 20278),
 ('Microservices', 17094),
 ('Developer portal or other central places to find tools/services', 13327),
 ('Observability tools', 12941),
 ('Innersource initiative', 5692),
 ('None of these', 4757)]
```

In [288...]

```
(df
 .ProfessionalTech
 .fillna("unanswered")
 .str.replace(";None of these", "")
 .str.replace("Continuous integration \b(CI\b) and \b(more often\b) continuous delivery",
 .str.replace("DevOps function", "devops")
 .str.replace("Developer portal or other central places to find tools/services", "dev_
 .str.replace(" ", "_")
 .str.lower()
 .value_counts().to_frame())
```

Out[288]:

Professi

	unanswered
	none_of_these
innersource_initiative;devops;microservices;dev_portal;ci_cd;automated_testing;observability_tools	
devops;microservices;dev_portal;ci_cd;automated_testing;observability_tools	
devops;microservices;ci_cd;automated_testing;observability_tools	
	...
innersource_initiative;devops;observability_tools	
innersource_initiative;devops;dev_portal;automated_testing;observability_tools	
innersource_initiative;microservices;dev_portal;observability_tools	
innersource_initiative;microservices;dev_portal;automated_testing;observability_tools	
innersource_initiative;microservices;automated_testing;observability_tools	

129 rows × 1 columns



### Observations:

- There're 155 unique entries
- Entries are nested, separated by a semicolon
- Some of the entries contain `none of these` along with other values. which doesn't make much sense

### Steps:

- The entries can be shortened
- This variable can be handled in various ways:
  - The rare categories (<1%) can be grouped together
  - The categories can be shortened based on presence of keywords
  - Multiple binary columns can be created
- These steps should be kept in mind for performing feature engineering during exploratory analysis
- The entries in which there's `none of these` along with other values, `none of these` will be removed

## TrueFalse\_1

In [289...]

col\_description("TrueFalse\_1")

### Description:

Are you involved in supporting new hires during their onboarding?

```
In [290... df.TrueFalse_1
```

```
Out[290]: 0      NaN  
1      NaN  
2      NaN  
3      NaN  
4      NaN  
...  
73263  Yes  
73264  No  
73265  No  
73266  NaN  
73267  NaN  
Name: TrueFalse_1, Length: 73268, dtype: object
```

```
In [291... df.TrueFalse_1.unique()
```

```
Out[291]: array([nan, 'Yes', 'No'], dtype=object)
```

### Observations:

- There're 2 unique entries
- The entries are clean and valid

### Steps:

- The categories will be renamed for clarity
- Type will be made `category`

## TrueFalse\_2

```
In [292... col_description("TrueFalse_2")
```

### Description:

Do you use learning resources provided by your employer?

```
In [293... df.TrueFalse_2
```

```
Out[293]: 0      NaN  
1      NaN  
2      NaN  
3      NaN  
4      NaN  
...  
73263  Yes  
73264  Yes  
73265  No  
73266  NaN  
73267  NaN  
Name: TrueFalse_2, Length: 73268, dtype: object
```

```
In [294... df.TrueFalse_2.unique()
```

```
Out[294]: array([nan, 'Yes', 'No'], dtype=object)
```

- Same as `TrueFalse_1`

## TrueFalse\_3

```
In [295...]: col_description("TrueFalse_3")
```

### Description:

Does your employer give you time to learn new skills?

```
In [296...]: df.TrueFalse_3
```

```
Out[296]: 0      NaN  
1      NaN  
2      NaN  
3      NaN  
4      NaN  
...  
73263  Yes  
73264  Yes  
73265  No  
73266  NaN  
73267  NaN  
Name: TrueFalse_3, Length: 73268, dtype: object
```

```
In [297...]: df.TrueFalse_3.unique()
```

```
Out[297]: array([nan, 'Yes', 'No'], dtype=object)
```

- Same as `TrueFalse_1`

## SurveyLength

```
In [298...]: col_description("SurveyLength")
```

### Description:

How do you feel about the length of the survey this year?

```
In [299...]: df.SurveyLength
```

```
Out[299]: 0                  NaN  
1                  Too long  
2      Appropriate in length  
3      Appropriate in length  
4                  Too long  
...  
73263              Too long  
73264              Too long  
73265  Appropriate in length  
73266  Appropriate in length  
73267  Appropriate in length  
Name: SurveyLength, Length: 73268, dtype: object
```

In [300...]: df.SurveyLength.unique()

Out[300]: array([nan, 'Too long', 'Appropriate in length', 'Too short'],  
dtype=object)

In [301...]:  
df  
.SurveyLength  
.fillna("unanswered")  
.str.lower()  
.str.replace(" in length", "")  
.str.replace(" ", "\_")  
.astype("category")  
.value\_counts()

Out[301]: appropriate 53883  
too\_long 14491  
unanswered 2824  
too\_short 2070  
Name: SurveyLength, dtype: int64

### Observations:

- There're 3 unique entries
- The entries are clean and valid

### Steps:

- The categories will be renamed for clarity
- Type will be made category

## SurveyEase

In [302...]: col\_description("SurveyEase")

### Description:

How easy or difficult was this survey to complete?

In [303...]: df.SurveyEase

Out[303]: 0 NaN  
1 Difficult  
2 Neither easy nor difficult  
3 Easy  
4 Easy  
...  
73263 Easy  
73264 Easy  
73265 Easy  
73266 Easy  
73267 Easy  
Name: SurveyEase, Length: 73268, dtype: object

In [304...]: df.SurveyEase.unique()

```
In [304]: array([nan, 'Difficult', 'Neither easy nor difficult', 'Easy'],
      dtype=object)
```

```
In [305...]: (df
    .SurveyEase
    .fillna("unanswered")
    .str.lower()
    .replace("neither easy nor difficult", "neutral")
    .astype("category")
    .value_counts())
```

```
Out[305]: easy        47886
neutral      21627
unanswered     2760
difficult       995
Name: SurveyEase, dtype: int64
```

### Observations:

- There're 3 unique entries
- The entries are clean and valid

### Steps:

- The categories will be renamed for clarity
- Type will be made category

## ConvertedCompYearly

```
In [306...]: df.ConvertedCompYearly
```

```
Out[306]: 0          NaN
1          NaN
2        40205.0
3      215232.0
4          NaN
...
73263      NaN
73264      NaN
73265      NaN
73266      NaN
73267      NaN
Name: ConvertedCompYearly, Length: 73268, dtype: float64
```

```
In [307...]: df.ConvertedCompYearly.describe()
```

```
Out[307]: count    3.807100e+04
mean     1.707613e+05
std      7.814132e+05
min      1.000000e+00
25%     3.583200e+04
50%     6.784500e+04
75%     1.200000e+05
max      5.000000e+07
Name: ConvertedCompYearly, dtype: float64
```

# Data Cleaning Function

In [340...]

```
def clean_data(data):
    """
    This function accepts a raw dataset and returns the cleaned version

    Parameters:
    -----
    data: pd.DataFrame
        The raw dataset to clean and transform
    """

    return (data
            .drop(columns=["ResponseId",
                           "VCHostingPersonal use",
                           "VCHostingProfessional use"])
            .rename(columns=str.lower)
            .rename(columns={"mainbranch": "coding_proficiency",
                            "remotework": "work_type",
                            "codingactivities": "coding_activity",
                            "edlevel": "education_level",
                            "learncode": "learnt_coding",
                            "learncodeonline": "learnt_coding_online",
                            "learncodecoursescert": "learnt_coding_courses",
                            "yearscode": "coding_years",
                            "yearscodepro": "coding_pro_years",
                            "devtype": "profession",
                            "orgsize": "org_size",
                            "purchaseinfluence": "purchase_influence_level",
                            "buynewtool": "learn_new_tool",
                            "comptotal": "comp_total",
                            "compfreq": "comp_freq",
                            "languagehaveworkedwith": "lang_worked_with",
                            "languagewanttoworkwith": "lang_want_work_with",
                            "databasehaveworkedwith": "db_worked_with",
                            "databasewanttoworkwith": "db_want_work_with",
                            "platformhaveworkedwith": "platform_worked_with",
                            "platformwanttoworkwith": "platform_want_work_with",
                            "webframehaveworkedwith": "web_frame_worked_with",
                            "webframewanttoworkwith": "web_frame_want_work_with",
                            "misctechhaveworkedwith": "misc_tech_worked_with",
                            "misctechwanttoworkwith": "misc_tech_want_work_with",
                            "toolsTechhaveworkedwith": "tools_tech_worked_with",
                            "toolsTechwanttoworkwith": "tools_tech_want_work_with",
                            "newcollabtoolshaveworkedwith": "new_collab_tools_worked_",
                            "newcollabtoolswanttoworkwith": "new_collab_tools_want_w",
                            "opsysprofessional use": "op_sys_pro_use",
                            "opsyspersonal use": "op_sys_personal_use",
                            "versioncontrolsystem": "ver_control_sys",
                            "vcinteraction": "vc_interaction",
                            "officestackasynchaveworkedwith": "office_stack_async_wor",
                            "officestackasyncwanttoworkwith": "office_stack_async_wan",
                            "officestacksynchaveworkedwith": "office_stack_sync_work",
                            "officestacksyncwanttoworkwith": "office_stack_sync_want_",
                            "newsosites": "new_sites_visited",
                            "sovisitfreq": "sites_visit_freq",
                            "soaccount": "have_account",
```

```

"sopartfreq": "participate",
"socomm": "consider_self_member",
"age": "age_group",
"mentalhealth": "mental_health",
"tbranch": "participate_dev_series",
"icorpm": "ind_cont_ppl_manager",
"workexp": "work_exp",
"knowledge_1": "interact_ppl_out_team",
"knowledge_2": "info_not_shared_team",
"knowledge_3": "can_find_info_org",
"knowledge_4": "useful_resources",
"knowledge_5": "know_resources",
"knowledge_6": "ans_questions_repeated",
"knowledge_7": "interrupted_waiting",
"frequency_1": "get_help_out_team",
"frequency_2": "interact_out_team",
"frequency_3": "meet_silos_work",
"timesearching": "hours_spent_searching",
"timeanswering": "hours_spent_answering",
"onboarding": "onboarding_time",
"professionaltech": "company_tech",
"truefalse_1": "support_new_emp",
"truefalse_2": "use_resources",
"truefalse_3": "given_time_learning",
"surveylength": "survey_length",
"surveyease": "survey_difficulty",
"convertedcompyearly": "conv_yearly_comp")}

.apply(lambda ser: ser.str.strip().fillna("unanswered") if ser.dtype == "O"
.assign(coding_proficiency=lambda df_: df_
                           .coding_proficiency
                           .replace([
                               "I am a developer by pr
                               "I am learning to code"
                               "I am not primarily a c
                               "I code primarily as a
                               "None of these",
                               "I used to be a develop
                               ["developer",
                               "learning",
                               "work_partly",
                               "hobby",
                               "other",
                               "former_developer"])
                           .astype("category"),
employment=lambda df_: df_
                           .employment
                           .str.lower()
                           .replace({"i prefer not to say": "unanswe
                           .str.replace("independent contractor, fre
                           .str.replace("not employed, but looking f
                           .str.replace("not employed, and not looki
                           .str.replace(" ", "_")
                           .str.replace("-", "_")),
work_type=lambda df_: df_
                           .work_type
                           .replace([
                               "Fully remote",
                               "Hybrid (some remote, some in-per
                               "Full in-person"],
                               ["remote",
                               "hybrid",
                               "in_person"]))

```

```
        .astype("category"),
coding_activity=lambda df_: df_
    .coding_activity
    .str.lower()
    .str.replace("contribute to open-sou
    .str.replace("freelance/contract wor
    .str.replace("school or academic wor
    .str.replace("bootstrapping a busine
    .str.replace("i don't code outside o
    .str.replace("other \please specify
education_level=lambda df_: df_
    .education_level
    .replace([
        "Bachelor's degree (B.A., B.S.)",
        "Master's degree (M.A., M.S.)",
        "Some college/university student",
        "Secondary school (e.g. American
        "Associate degree (A.A., A.S.)",
        "Other doctoral degree (Ph.D., D
        "Primary/elementary school",
        "Something else",
        "Professional degree (JD, MD, PhD
        ["bachelors",
        "masters",
        "college",
        "secondary_school",
        "associate",
        "doctorate",
        "primary_school",
        "other",
        "professional"])
    ].astype("category"),
learnt_coding=lambda df_: df_
    .learnt_coding
    .str.replace("Other online resources \n
    .str.replace("School \n(i.e., University
    .str.replace("Books / Physical media",
    .str.replace("Online Courses or Certific
    .str.replace("On the job training", "Work
    .str.replace("Friend or family member" \n
    .str.replace("Hackathons \n(virtual or in
    .str.replace("Other \n(please specify)
learnt_coding_online=lambda df_: df_
    .learnt_coding_online
    .str.replace("Other \n(Please specify)", "O
    .str.replace("Technical documents", "Tech
    .str.replace("How-to videos", "Tutorials")
    .str.replace("Video-based Online courses", "V
    .str.replace("Online books", "Books"),
    .str.replace("Online forum", "Forums"),
    .str.replace("Written-based Online courses", "W
    .str.replace("Coding sessions \n", "Coding"),
    .str.replace("Written Tutorials", "Written Tuto
    .str.replace("Interactive tutorials", "Interactive
    .str.replace("Certification video", "Certificati
    .str.replace("Auditory material", "Auditory"),
    .str.replace("Online challenges", "Challenges"),
    .str.lower(),
    .str.replace(" ", "_"),
learnt_coding_courses=lambda df_: df_
    .learnt_coding_courses
```

```

        .str.lower(),
coding_years=lambda df_: df_
    .coding_years
    .replace(["Less than 1 year",
              "More than 50 years",
              "unanswered"],
              ["1",
               "50",
               np.nan])
    .pipe(lambda ser: pd.to_numeric(ser)),
coding_pro_years=lambda df_: df_
    .coding_pro_years
    .replace(["Less than 1 year",
              "More than 50 years",
              "unanswered"],
              ["1",
               "50",
               np.nan])
    .pipe(lambda ser: pd.to_numeric(ser))
profession=lambda df_: df_
    .profession
    .str.replace("Other \Please specify\:", "")
    .str.replace("Developer, embedded applica",
    .str.replace("Engineer, data", "data_engi",
    .str.replace("Developer, desktop or enter",
    .str.replace("Developer, full-stack", "fu",
    .str.replace("Developer, front-end", "fro",
    .str.replace("Developer, back-end", "back",
    .str.replace("Developer, mobile", "mobile",
    .str.replace("Data scientist or machine l",
    .str.replace("Data or business analyst",
    .str.replace("Developer, QA or test", "te",
    .str.replace("Engineer, site reliability",
    .str.replace("Developer, game or graphics",
    .str.replace("Senior Executive \C-Suite",
    .str.replace("Marketing or sales professi",
    .str.lower()
    .str.replace(" ", "_"),
org_size=lambda df_: df_
    .org_size
    .str.replace(" employees", "")
    .replace("Just me - I am a freelancer, sole",
    .replace(["2 to 9",
              "10 to 19",
              "20 to 99",
              "100 to 499",
              "500 to 999",
              "1,000 to 4,999",
              "5,000 to 9,999",
              "10,000 or more",
              "I don't know"],
              ["small",
               "small",
               "small",
               "small",
               "medium",
               "medium",
               "large",
               "large",
               "unknown"]))

```

```
    .astype("category"),
purchase_influence_level=lambda df_: df_
                                .purchase_influence_level
                                .replace(["I have some infl"
                                         "I have little or"
                                         "I have a great c"
                                         ["some",
                                         "little",
                                         "great"]])
                                .astype("category"),
learn_new_tool=lambda df_: df_
                            .learn_new_tool
                            .str.replace("Other \please specify\"
                            .str.replace("Start a free trial", "f
                            .str.replace("Ask developers I know/w
                            .str.replace("Visit developer commun
                            .str.replace("Read ratings or reviews
                            .str.replace("Research companies that
                            .str.replace("Research companies that
                            .str.lower()
                            .str.replace(" ", "_"),
country=lambda df_: df_
                    .country
                    .replace(["Lao People's Democratic Republic"
                             "Democratic Republic of the Congo"
                             "Brunei Darussalam",
                             "Swaziland",
                             "Congo, Republic of the...",
                             "Libyan Arab Jamahiriya",
                             "United Republic of Tanzania",
                             "Syrian Arab Republic",
                             "Republic of Moldova",
                             "The former Yugoslav Republic of M
                             "Republic of Korea",
                             "Venezuela, Bolivarian Republic of
                             "United Arab Emirates",
                             "Hong Kong (S.A.R.)",
                             "Viet Nam",
                             "Iran, Islamic Republic of...",
                             "Russian Federation",
                             "United Kingdom of Great Britain a
                             "United States of America"],
                             ["Laos",
                             "Congo",
                             "Brunei",
                             "Eswatini",
                             "Congo",
                             "Libya",
                             "Tanzania",
                             "Syria",
                             "Moldova",
                             "North Macedonia",
                             "South Korea",
                             "Venezuela",
                             "UAE",
                             "Hong Kong",
                             "Vietnam",
                             "Iran",
                             "Russia",
                             "UK"])
```

```
        "USA"]),
currency=lambda df_: df_
    .country
    .map({ "unanswered": "unanswered", "Canada": "CA",
           "UK": "GBP", "Israel": "ILS", "USA": "USD",
           "Germany": "EUR", "India": "INR", "Netherlands": "EUR",
           "Croatia": "HRK", "Australia": "AUD", "Austria": "EUR",
           "Czech Republic": "CZK", "Italy": "EUR", "Ireland": "EUR", "Poland": "PLN",
           "Slovenia": "EUR", "Iraq": "IQD", "Sweden": "SEK",
           "Madagascar": "MGA", "Norway": "NOK", "Hong Kong": "HKD", "Mexico": "MXN",
           "Brazil": "BRL", "Lithuania": "LTL", "Denmark": "DKK", "Spain": "EUR", "Ecuador": "EUR",
           "South Africa": "ZAR", "Ukraine": "UAH", "Romania": "RON", "Portugal": "EUR",
           "Oman": "OMR", "Belgium": "EUR", "China": "CNY", "Latvia": "LVL", "Philippines": "PHP",
           "Saudi Arabia": "SAR", "Kenya": "KES", "Iceland": "ISK", "Vietnam": "VND", "Montenegro": "EUR",
           "Slovakia": "EUR", "Luxembourg": "LUF", "Turkmenistan": "TMT", "Hungary": "HUF", "Tunisia": "TND",
           "Dominican Republic": "DOP", "Jordan": "JOD", "Nepal": "NPR", "Iran": "IRR", "Indonesia": "IDR",
           "Bosnia and Herzegovina": "BAM", "Armenia": "AMD", "Colombia": "COP", "Kazakhstan": "KZT",
           "Costa Rica": "CRC", "Honduras": "HNL", "Estonia": "EUR", "Algeria": "DZD", "Mali": "XOF",
           "Morocco": "MAD", "Eswatini": "SZL", "New Zealand": "NZD", "North Macedonia": "MKD",
           "Cyprus": "CYP", "UAE": "AED", "Peru": "PEN", "Uzbekistan": "UZS", "Ethiopia": "ETB", "Malta": "MLT",
           "Nicaragua": "NIO", "Lebanon": "LBP", "Belize": "BZD", "Zambia": "ZMW", "Bolivia": "BOB",
           "Malaysia": "MYR", "Laos": "LAK", "Guatemala": "GTQ", "Angola": "AOA", "Suriname": "SRD",
           "El Salvador": "USD", "Qatar": "QAR", "Nigeria": "NGN", "Kyrgyzstan": "KGS", "Zimbabwe": "ZWD",
           "Rwanda": "RWF", "Cambodia": "KHR", "Malawi": "MWK", "Fiji": "FJD", "Nomadic": "unknown",
           "Albania": "ALL", "Timor-Leste": "USD", "Moldova": "MDL", "Tajikistan": "TJS", "Tanzania": "TZS",
           "Myanmar": "MMK", "Cameroon": "XAF", "Kosovo": "EUR", "Benin": "XOF", "Botswana": "BWP",
           "Namibia": "NAD", "Somalia": "SOS", "Mozambique": "MZN", "Lesotho": "LSL", "North Korea": "KPW",
           "Antigua and Barbuda": "XBT", "Palau": "USD", "Guinea": "GNF", "Haiti": "HTG", "Gabon": "XAF",
           "Mauritania": "MRU", "Guyana": "GYD", "Saint Lucia": "XCD", "Brunei": "BND", "Gambia": "GMD", "Micronesia": "MVR"}))
```

```

    "Djibouti": "DJF", "Seychelles": "SCR",
    "Saint Kitts and Nevis": "KN"}),
comp_total=lambda df_: df_
    .comp_total
    .where(lambda ser: ~(data.Employment.str.contains("Total")))
    .pipe(lambda ser: pd.to_numeric(ser)),
comp_freq=lambda df_: df_
    .comp_freq
    .str.lower()
    .astype("category"),
db_worked_with=lambda df_: df_
    .db_worked_with
    .str.replace(" ", "_"),
db_want_work_with=lambda df_: df_
    .db_want_work_with
    .str.replace(" ", "_"),
platform_worked_with=lambda df_: df_
    .platform_worked_with
    .str.replace("IBM Cloud or Watsons", "IBM Cloud or Watsons")
    .str.replace("OpenStack", "openstack")
    .str.replace("DigitalOcean", "DigitalOcean")
    .str.lower()
    .str.replace(" ", "_"),
platform_want_work_with=lambda df_: df_
    .platform_want_work_with
    .str.replace("IBM Cloud or Watsons", "IBM Cloud or Watsons")
    .str.replace("OpenStack", "openstack")
    .str.replace("DigitalOcean", "DigitalOcean")
    .str.lower()
    .str.replace(" ", "_"),
web_frame_worked_with=lambda df_: df_
    .web_frame_worked_with
    .str.lower()
    .str.replace(" ", "_")
    .str.replace(".", "_", regex=True),
web_frame_want_work_with=lambda df_: df_
    .web_frame_want_work_with
    .str.lower()
    .str.replace(" ", "_")
    .str.replace(".", "_", regex=True),
misc_tech_worked_with=lambda df_: df_
    .misc_tech_worked_with
    .str.replace("Torch/PyTorch", "torch/pytorch")
    .str.replace(".NET", "dot_net")
    .str.lower()
    .str.replace(" ", "_")
    .str.replace("-", "_"),
misc_tech_want_work_with=lambda df_: df_
    .misc_tech_want_work_with
    .str.replace("Torch/PyTorch", "torch/pytorch")
    .str.replace(".NET", "dot_net")
    .str.lower()
    .str.replace(" ", "_")
    .str.replace("-", "_"),
tools_tech_worked_with=lambda df_: df_
    .tools_tech_worked_with
    .str.lower()
    .str.replace(" ", "_"),
tools_tech_want_work_with=lambda df_: df_
    .tools_tech_want_work_with

```

```
.str.lower()
.str.replace(" ", "_"),
new_collab_tools_worked_with=lambda df_: df_
    .new_collab_tools_worked_
    .str.replace("RAD Studio")
    .str.replace("IPython/Jupyter")
    .str.lower()
    .str.replace(" ", "_"),
new_collab_tools_want_work_with=lambda df_: df_
    .new_collab_tools_want_
    .str.replace("RAD Studio")
    .str.replace("IPython/Jupyter")
    .str.lower()
    .str.replace(" ", "_"),
op_sys_pro_use=lambda df_: df_
    .op_sys_pro_use
    .str.replace("Linux-based", "Linux"),
    .str.replace("Windows Subsystem for Linux", "WSL"),
    .str.replace("Other \please specify"),
    .str.lower()
    .str.replace("macos", "mac_os"),
op_sys_personal_use=lambda df_: df_
    .op_sys_personal_use
    .str.replace("Linux-based", "Linux"),
    .str.replace("Windows Subsystem for Linux", "WSL"),
    .str.replace("Other \please specify"),
    .str.lower()
    .str.replace("macos", "mac_os"),
ver_control_sys=lambda df_: df_
    .ver_control_sys
    .str.replace("Other \please specify"),
    .str.replace("I don't use one", "none"),
    .str.lower(),
vc_interaction=lambda df_: df_
    .vc_interaction
    .str.replace("Version control hosting service")
    .str.replace("Dedicated version control system")
    .str.lower()
    .str.replace(" ", "_")
    .str.replace("-", "_"),
office_stack_async_worked_with=lambda df_: df_
    .office_stack_async_worked_
    .str.replace("Stack Overflow")
    .str.replace("monday")
    .str.replace("DingTalk")
    .str.replace("Planview")
    .str.lower()
    .str.replace(" ", "_"),
office_stack_async_want_work_with=lambda df_: df_
    .office_stack_async_want_
    .str.replace("Stack Overflow")
    .str.replace("monday")
    .str.replace("DingTalk")
    .str.replace("Planview")
    .str.lower()
    .str.replace(" ", "_"),
office_stack_sync_worked_with=lambda df_: df_
    .office_stack_sync_worked_
    .str.replace("RocketChat")
    .str.replace("RingCentral")
```

```
        .str.lower()
        .str.replace(" ", "_")
    office_stack_sync_want_work_with=lambda df_: df_
        .office_stack_sync_
        .str.replace("Rocke"
        .str.replace("RingO"
        .str.lower()
        .str.replace(" ", "
    blockchain=lambda df_: df_
        .blockchain
        .str.lower()
        .astype("category"),
    new_sites_visited=lambda df_: df_
        .new_sites_visited
        .str.replace("Collectives on Stack"
        .str.replace("Stack Overflow for T
        .str.replace("I have never visited
        .str.lower()
        .str.replace(" ", "_"),
    sites_visit_freq=lambda df_: df_
        .sites_visit_freq
        .replace(["Daily or almost daily",
                  "Multiple times per day",
                  "A few times per week",
                  "A few times per month o
                  "Less than once per mont
                  ["daily",
                   "daily",
                   "weekly",
                   "weekly",
                   "monthly"])
        .astype("category"),
    have_account=lambda df_: df_
        .have_account
        .str.lower()
        .replace("Not sure/can't remember", "un
        .astype("category"),
    participate=lambda df_: df_
        .participate
        .replace(["Daily or almost daily",
                  "Multiple times per day",
                  "A few times per week",
                  "A few times per month or week
                  "Less than once per month or w
                  "I have never participated in
                  ["daily",
                   "daily",
                   "weekly",
                   "weekly",
                   "monthly",
                   "never"])
        .astype("category"),
    consider_self_member=lambda df_: df_
        .consider_self_member
        .replace(["Yes, somewhat",
                  "Yes, definitely",
                  "No, not at all",
                  "No, not really",
                  "Not sure"],
                  ["yes",
```

```

                                "yes",
                                "no",
                                "no",
                                "unsure"])
.str.lower()
.astype("category"),

age_group=lambda df_: df_
    .age_group
    .replace({"Under 18 years old": "minor",
              "18-24 years old": "young_adult",
              "25-34 years old": "young_adult",
              "35-44 years old": "middle_aged",
              "45-54 years old": "middle_aged",
              "55-64 years old": "senior",
              "65 years or older": "senior",
              "Prefer not to say": "unanswered"})
.astype("category"),

gender=lambda df_: df_
    .gender
    .str.replace("Non-binary, genderqueer, or ger",
                 "other")
    .str.replace("Or, in your own words:", "other")
    .str.replace("Prefer not to say", "unanswered")
    .str.replace("Man", "male")
    .str.replace("Woman", "female"),

trans=lambda df_: df_
    .trans
    .str.lower()
    .replace({"prefer not to say": "unanswered",
              "or, in your own words": "yes"})
    .astype("category"),

sexuality=lambda df_: df_
    .sexuality
    .str.replace("Straight / Heterosexual", "h")
    .str.replace("Prefer not to say", "unanswe")
    .str.replace("Gay or Lesbian", "Homosexua")
    .str.replace("Prefer to self-describe:", "t")
    .str.lower(),
.ethnicity=df_
    .ethnicity
    .replace({"Prefer not to say": "unanswered",
              "I don't know": "unsure"})
    .str.replace("Hispanic or Latino/a", "Hispi")
    .str.replace("Or, in your own words:", "oth")
    .str.replace("Indigenous \such as Native")
    .str.lower()
    .str.replace(" ", "_"),

accessibility=lambda df_: df_
    .accessibility
    .replace({"Prefer not to say": "unansw",
              "None of the above": "none_c"})
    .str.replace("none of the above", "nor")
    .str.replace("I am blind / have diffic")
    .str.replace("Or, in your own words:", "ot")
    .str.replace("I am deaf / hard of hear")
    .str.replace("I am unable to / find it")
    .str.replace("I am unable to / find it"),

mental_health=lambda df_: df_
    .mental_health
    .replace({"None of the above": "none_c",
              "Prefer not to say": "unansw"
}

```

```

    .str.replace("Or, in your own words:", " ")
    .str.replace("I have an anxiety disorder", " ")
    .str.replace("I have a mood or emotion disorder", " ")
    .str.replace("I have a concentration issue", " ")
    .str.replace("I have autism / an autism spectrum disorder", " ")
    .str.replace("I have learning differences", " ")

participate_dev_series=lambda df_: df_
    .participate_dev_series
    .str.lower()
    .astype("category"),

ind_cont_ppl_manager=lambda df_: df_
    .ind_cont_ppl_manager
    .str.lower()
    .str.replace(" ", "_")
    .astype("category"),

interact_ppl_out_team=lambda df_: df_
    .interact_ppl_out_team
    .str.lower()
    .str.replace(" ", "_")
    .astype("category"),

info_not_shared_team=lambda df_: df_
    .info_not_shared_team
    .str.lower()
    .str.replace(" ", "_")
    .astype("category"),

can_find_info_org=lambda df_: df_
    .can_find_info_org
    .str.lower()
    .str.replace(" ", "_")
    .astype("category"),

useful_resources=lambda df_: df_
    .useful_resources
    .str.lower()
    .str.replace(" ", "_")
    .astype("category"),

know_resources=lambda df_: df_
    .know_resources
    .str.lower()
    .str.replace(" ", "_")
    .astype("category"),

ans_questions_repeated=lambda df_: df_
    .ans_questions_repeated
    .str.lower()
    .str.replace(" ", "_")
    .astype("category"),

interrupted_waiting=lambda df_: df_
    .interrupted_waiting
    .str.lower()
    .str.replace(" ", "_")
    .astype("category"),

get_help_out_team=lambda df_: df_
    .get_help_out_team
    .str.lower()
    .replace({"1-2 times a week": "rarely", "3-5 times a week": "mil", "6-10 times a week": "frequent", "10+ times a week": "frequently"}, " ")
    .astype("category"),

interact_out_team=lambda df_: df_
    .interact_out_team

```

```

        .str.lower()
        .replace({"1-2 times a week": "rarely",
                  "3-5 times a week": "mildly",
                  "6-10 times a week": "frequently",
                  "10+ times a week": "frequently"})
        .astype("category"),
meet_silos_work=lambda df_: df_
        .meet_silos_work
        .str.lower()
        .replace({"1-2 times a week": "rarely",
                  "3-5 times a week": "mildly",
                  "6-10 times a week": "frequently",
                  "10+ times a week": "frequently"})
        .astype("category"),
hours_spent_searching=lambda df_: df_
        .hours_spent_searching
        .str.lower()
        .str.replace(" minutes a day",
                     " less than 15": "quarterly",
                     "15-30": "half",
                     "30-60": "one",
                     "60-120": "two",
                     "over 120": "above two")
        .astype("category"),
hours_spent_answering=lambda df_: df_
        .hours_spent_answering
        .str.lower()
        .str.replace(" minutes a day",
                     " less than 15": "quarterly",
                     "15-30": "half",
                     "30-60": "one",
                     "60-120": "two",
                     "over 120": "above two")
        .astype("category"),
onboarding_time=lambda df_: df_
        .onboarding_time
        .str.lower()
        .str.replace(" ", "_")
        .str.replace("somewhat_", "")
        .astype("category"),
company_tech=lambda df_: df_
        .company_tech
        .str.replace(";None of these", "")
        .str.replace("Continuous integration \\", "CI\\")
        .str.replace("DevOps function", "devops")
        .str.replace("Developer portal or other", "devportal")
        .str.lower()
        .str.replace(" ", "_"),
support_new_emp=lambda df_: df_
        .support_new_emp
        .str.lower()
        .astype("category"),
use_resources=lambda df_: df_
        .use_resources
        .str.lower()
        .astype("category"),
given_time_learning=lambda df_: df_
        .given_time_learning
        .str.lower()
        .astype("category"),

```

```

survey_length=lambda df_: df_
    .survey_length
    .str.lower()
    .str.replace(" in length", "")
    .str.replace(" ", "_")
    .astype("category"),
survey_difficulty=lambda df_: df_
    .survey_difficulty
    .str.lower()
    .replace("neither easy nor difficult", "neutral")
    .astype("category"))

```

In [341]:

```
df_cleaned = clean_data(df)
df_cleaned.shape
```

Out[341]:

```
(73268, 76)
```

## Data Validation

### 1. Checking if the total no. of years coded is less than no. of years coded professionally:

In [311]:

```
col_description("YearsCode")
```

#### Description:

Including any education, how many years have you been coding in total?

In [312]:

```
col_description("YearsCodePro")
```

#### Description:

NOT including education, how many years have you coded professionally (as a part of your work)?

In [313]:

```
(clean_data(df)
    .dropna(subset=["coding_years", "coding_pro_years"], how="any")
    .pipe(lambda df_: df_.coding_pro_years.gt(df_.coding_years).sum()))
```

Out[313]:

```
536
```

In [314]:

```

def validate_coding_years(data):
    result = np.full(data.shape[0], True)
    for i, (ix, row) in enumerate(data.iterrows()):
        try:
            years_coded = int(row.coding_years)
        except:
            continue

        try:
            years_coded_pro = int(row.coding_pro_years)
        except:
            continue

```

```

if years_coded_pro > years_coded:
    result[i] = False

return data[result]

```

## Observations:

- There're 536 entries where `coding_pro_years` is greater than `coding_years`
  - This is not possible and it doesn't make sense
  - Might be better to delete these observations

## Steps:

- The function `validate_coding_years` will handle this error and delete the erranous observations

## 2. Checking the currency values when country isn't mentioned

```
In [345...]: (df_cleaned
    .query("country == 'unanswered'")
    .currency
    .unique())
```

```
Out[345]: array(['unanswered'], dtype=object)
```

## Observations:

- Whenever country is not mentioned, the corresponding currency value is also missing, as expected

## Steps:

- No cleaning steps required

## Cleaned Data

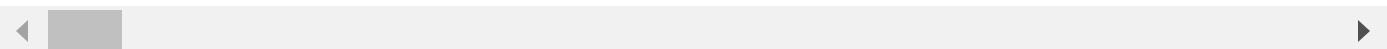
```
In [346...]: df_cleaned = (df
    .pipe(clean_data)
    .pipe(validate_coding_years)
    .reset_index(drop=True))

df_cleaned
```

Out[346]:

	<b>coding_proficiency</b>	<b>employment</b>	<b>work_type</b>	<b>coding_activity</b>	<b>educa</b>
<b>0</b>	other	unanswered	unanswered	unanswered	ui
<b>1</b>	developer	employed_full_time	remote	hobby;open_source_contribution	ui
<b>2</b>	work_partly	employed_full_time	hybrid	hobby	
<b>3</b>	developer	employed_full_time	remote	only_work	
<b>4</b>	developer	employed_full_time	hybrid	hobby	
...	...	...	...	...	...
<b>72727</b>	developer	employed_full_time	remote	freelancing	
<b>72728</b>	developer	employed_full_time	in_person	hobby	
<b>72729</b>	work_partly	employed_full_time	hybrid	hobby;academics	
<b>72730</b>	developer	employed_full_time	hybrid	hobby	
<b>72731</b>	former_developer	freelancer	remote	hobby;open_source_contribution;startup	

72732 rows × 76 columns



## Memory Usage

In [347...]

`df.memory_usage(deep=True).sum()`

Out[347]:

375776556

In [348...]

`df_cleaned.memory_usage(deep=True).sum()`

Out[348]:

216710996

In [349...]

375776346 / 218260617

Out[349]:

1.7216864460710288

## Meta-data

In [319...]

`df_cleaned.info(memory_usage="deep")`

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 72732 entries, 0 to 72731
Data columns (total 76 columns):
 #   Column           Non-Null Count Dtype
 ---  -- 
 0   coding_proficiency    72732 non-null category
 1   employment          72732 non-null object
 2   work_type            72732 non-null category
 3   coding_activity       72732 non-null object
 4   education_level      72732 non-null category
 5   learnt_coding         72732 non-null object
 6   learnt_coding_online  72732 non-null object
 7   learnt_coding_courses 72732 non-null object
 8   coding_years          70795 non-null float64
 9   coding_pro_years     51297 non-null float64
 10  profession           72732 non-null object
 11  org_size              72732 non-null category
 12  purchase_influence_level 72732 non-null category
 13  learn_new_tool        72732 non-null object
 14  country               72732 non-null object
 15  currency              72732 non-null object
 16  comp_total             38091 non-null float64
 17  comp_freq              72732 non-null category
 18  lang_worked_with      72732 non-null object
 19  lang_want_work_with   72732 non-null object
 20  db_worked_with        72732 non-null object
 21  db_want_work_with     72732 non-null object
 22  platform_worked_with  72732 non-null object
 23  platform_want_work_with 72732 non-null object
 24  web_frame_worked_with 72732 non-null object
 25  web_frame_want_work_with 72732 non-null object
 26  misc_tech_worked_with 72732 non-null object
 27  misc_tech_want_work_with 72732 non-null object
 28  tools_tech_worked_with 72732 non-null object
 29  tools_tech_want_work_with 72732 non-null object
 30  new_collab_tools_worked_with 72732 non-null object
 31  new_collab_tools_want_work_with 72732 non-null object
 32  op_sys_pro_use        72732 non-null object
 33  op_sys_personal_use   72732 non-null object
 34  ver_control_sys       72732 non-null object
 35  vc_interaction         72732 non-null object
 36  office_stack_async_worked_with 72732 non-null object
 37  office_stack_async_want_work_with 72732 non-null object
 38  office_stack_sync_worked_with 72732 non-null object
 39  office_stack_sync_want_work_with 72732 non-null object
 40  blockchain             72732 non-null category
 41  new_sites_visited      72732 non-null object
 42  sites_visit_freq       72732 non-null category
 43  have_account           72732 non-null category
 44  participate             72732 non-null category
 45  consider_self_member   72732 non-null category
 46  age_group               72732 non-null category
 47  gender                  72732 non-null object
 48  trans                   72732 non-null category
 49  sexuality                72732 non-null object
 50  ethnicity                72732 non-null object
 51  accessibility             72732 non-null object
 52  mental_health            72732 non-null object
 53  participate_dev_series  72732 non-null category
 54  ind_cont_ppl_manager    72732 non-null category

```

```

      data_cleaning_notebook
55   work_exp           36414 non-null float64
56   interact_ppl_out_team 72732 non-null category
57   info_not_shared_team 72732 non-null category
58   can_find_info_org    72732 non-null category
59   useful_resources     72732 non-null category
60   know_resources       72732 non-null category
61   ans_questions_repeated 72732 non-null category
62   interrupted_waiting 72732 non-null category
63   get_help_out_team   72732 non-null category
64   interact_out_team   72732 non-null category
65   meet_silos_work     72732 non-null category
66   hours_spent_searching 72732 non-null category
67   hours_spent_answering 72732 non-null category
68   onboarding_time     72732 non-null category
69   company_tech        72732 non-null object
70   support_new_emp    72732 non-null category
71   use_resources       72732 non-null category
72   given_time_learning 72732 non-null category
73   survey_length       72732 non-null category
74   survey_difficulty  72732 non-null category
75   conv_yearly_comp   37770 non-null float64
dtypes: category(33), float64(5), object(38)
memory usage: 207.6 MB

```

## Final Results:

- The raw dataset had 73,268 rows and 79 columns
- The cleaned dataset has 72,732 rows and 76 columns
- The 76 columns in the cleaned dataset are of the types:
  - `float` - 5
  - `category` - 33
  - `object` - 38
- The cleaned dataset utilizes nearly 2 times less memory compared to the raw dataset

## Feature Engineering (for the nested columns)

### Convenience Functions

In [320...]

```

def get_top_categories(data, name, n=5):
    """
    This function will return the top categories of the specified column
    from the dataset

    Parameters:
    -----
    data: pd.DataFrame
        The dataframe to work with

    name: str
        The column name whose top categories to return

    n: int
    """

    # Your code here
    pass

```

```

    No. of top categories to return
    """

    result_set = dict()
    for i, entry in data[name].items():
        if isinstance(entry, str):
            for value in entry.split(","):
                result_set[value] = result_set.get(value, 0) + 1

    return sorted(result_set,
                  key=lambda val: result_set[val],
                  reverse=True)[:n]

```

In [321...]

```

def clean_nested_column(data,
                       name,
                       replace_values=False,
                       to_replace=None,
                       replace_with=None,
                       default=None,
                       sum_values=False,
                       suffix=None,
                       terms=[],
                       add_indicators=False,
                       n=None):
    """
    Description:
    -----

```

This function will perform cleaning operations based on the user input on the nested columns and return the newly transformed variables concatenated with the provided dataset

Operations performed:

-----

1. Replace nested entries with values provided, according to a mapping
2. Add the no. of values present in the nested entries
3. Create binary indicator features for values in the nested entries

Parameters:

-----

`data: pd.DataFrame`  
The dataframe to work with

`name: str`  
The column name whose top categories to return

`replace_values: bool`  
Whether the nested entries should be replaced with valid values

`to_replace: list`  
A list of values to replace in the nested entries  
Used only if 'replace\_values' is True

`replace_with: list`  
A list of valid values to replace the nested entries

```

        Used only if 'replace_values' is True

default: str
    Default value to replace the nested entries
    Used only if 'replace_values' is True

sum_values: bool
    Whether to add the no. of values in the nested entries

suffix: str
    To append to column new column name
    Used only if 'sum_values' is True

terms: list
    Terms to avoid while calculating the number of values in the nested entries

add_indicators: bool
    Whether to create binary indicators for values in the nested entries

n: int
    The no. of values in the nested entries to create the binary indicators for
    Used only if 'add_indicators' is True
    """
    """

temp = pd.DataFrame()

# replacing nested entries with valid values; flattening the nested feature
if replace_values:
    temp[f"{name}_flattened"] = np.select([data[name].str.contains(value)
                                             for value in to_replace],
                                            replace_with,
                                            default)

    temp[f"{name}_flattened"] = temp[f"{name}_flattened"].astype("category")

# for accurate implementation of the following 2 operations
data = data.assign(**{f"{name}": data[name].replace("unanswered", np.nan)})

# summing the no. of values in the nested entries
if sum_values:
    avoid_terms = ["none", "unanswered"]
    avoid_terms.extend(terms)
    result = np.empty(data.shape[0])
    for i, entry in data[name].items():
        if isinstance(entry, str):
            total = 0
            values = set(entry.split(";"))
            for value in values:
                if value in avoid_terms:
                    total += 0
                else:
                    total += 1
            result[i] = total
        else:
            result[i] = np.nan
    temp[f"{name}_num_{suffix}"] = result

# adding indicators for values in the nested entries
if add_indicators:
    data = (data)
    top_categories = get_top_categories(data, name, n)

```

```

for category in top_categories:
    temp[f"{name}_{category}"] = (data
        .loc[:, name]
        .str.contains(category)
        .astype("float"))

return pd.concat([data, temp], axis=1)

```

- The above functions can be used for extracting various new features from the nested columns present in the dataset
- The function `clean_nested_column` performs 3 feature engineering operations:
  - Flattening the nested column
  - Creating a new column by calculating the total no. of values in the nested entries
  - Creating new binary features for the most frequent values in the nested entries
 ▪ All these operations can be controlled by specifying the corresponding parameters
- This step should actually be done during EDA
- By varying the specific parameters of this function, the usefulness of different features will be analyzed
  - This is an iterative step

## Example

In [322...]

```

clean_nested_column(data=df_cleaned,
                    name="lang_worked_with",
                    replace_values=True,
                    to_replace=[ "JavaScript",
                                "HTML/CSS",
                                "Python",
                                "SQL",
                                "C++"],
                    replace_with=[ "java_script",
                                "html_css",
                                "python",
                                "sql",
                                "c_plus_plus"],
                    default="other",
                    sum_values=True,
                    suffix="used",
                    terms=[],
                    add_indicators=True,
                    n=4) \
.head(50)

```

id	lang_worked_with_JavaScript	lang_worked_with_HTML/CSS	lang_worked_with_SQL	lang_worked_with_Python
aN	NaN	NaN	NaN	NaN
2.0	1.0	0.0	0.0	0.0
5.0	1.0	1.0	0.0	1.0
4.0	1.0	0.0	1.0	0.0
6.0	1.0	1.0	1.0	0.0
2.0	0.0	0.0	0.0	0.0
6.0	1.0	1.0	0.0	1.0
6.0	1.0	1.0	1.0	0.0
aN	NaN	NaN	NaN	NaN
3.0	0.0	0.0	0.0	0.0
6.0	1.0	1.0	1.0	0.0
7.0	1.0	1.0	1.0	1.0
6.0	0.0	1.0	1.0	0.0
7.0	1.0	1.0	1.0	0.0
7.0	1.0	1.0	0.0	1.0
5.0	1.0	1.0	1.0	0.0
5.0	0.0	0.0	0.0	1.0
2.0	0.0	0.0	1.0	1.0
3.0	0.0	0.0	1.0	0.0
5.0	1.0	1.0	0.0	1.0
8.0	1.0	1.0	1.0	0.0
3.0	0.0	0.0	1.0	1.0

id	lang_worked_with_JavaScript	lang_worked_with_HTML/CSS	lang_worked_with_SQL	lang_worked_with_Python
3.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	0.0
5.0	1.0	0.0	0.0	0.0
6.0	0.0	0.0	0.0	0.0
6.0	1.0	1.0	1.0	0.0
5.0	1.0	1.0	0.0	0.0
2.0	0.0	0.0	0.0	0.0
4.0	1.0	0.0	1.0	0.0
aN	NaN	NaN	NaN	NaN
4.0	1.0	1.0	1.0	0.0
2.0	0.0	0.0	0.0	1.0
3.0	0.0	0.0	0.0	1.0
5.0	1.0	1.0	1.0	1.0
2.0	0.0	0.0	0.0	1.0
8.0	1.0	1.0	1.0	1.0
5.0	1.0	1.0	0.0	1.0
4.0	1.0	1.0	1.0	1.0
9.0	1.0	1.0	1.0	0.0
7.0	1.0	1.0	1.0	0.0
4.0	0.0	1.0	1.0	1.0
6.0	1.0	1.0	1.0	0.0
aN	NaN	NaN	NaN	NaN
8.0	1.0	1.0	1.0	1.0

ed	lang_worked_with_JavaScript	lang_worked_with_HTML/CSS	lang_worked_with_SQL	lang_worked_with_Python
2.0	1.0	0.0	0.0	0.0
6.0	0.0	0.0	0.0	1.0
5.0	1.0	0.0	1.0	0.0
3.0	0.0	0.0	0.0	0.0
5.0	1.0	1.0	1.0	0.0

- This example demonstrates engineering new features for the column `lang_worked_with`
- In the resulting dataset, the last 6 columns depict the newly engineered features
- This is a very convenient function and provides various options to create new features from any of the nested columns present in the dataset