```
!pip install opendatasets --upgrade --quiet
```

```
import opendatasets as od
```

```
od.download('stackoverflow-developer-survey-2020')
```

```
Downloading https://raw.githubusercontent.com/JovianML/opendatasets/master/data/stackoverflow-developer-survey-2020/survey_results_public.csv to ./stackoverflow-developer-survey-2020/survey_results_public.csv

94609408it [00:05, 18321324.67it/s]

Downloading https://raw.githubusercontent.com/JovianML/opendatasets/master/data/stackoverflow-developer-survey-2020/survey_results_schema.csv to ./stackoverflow-developer-survey-2020/survey_results_schema.csv

16384it [00:00, 46231.97it/s]

Downloading https://raw.githubusercontent.com/JovianML/opendatasets/master/data/stackoverflow-developer-survey-2020/README.txt to ./stackoverflow-developer-survey-2020/README.txt

8192it [00:00, 21086.50it/s]
```

```
import os
```

```
os.listdir('stackoverflow-developer-survey-2020')
```

```
['survey_results_public.csv', 'survey_results_schema.csv', 'README.txt']
```

```
import pandas as pd
```

```
survey_raw_df = pd.read_csv('stackoverflow-developer-survey-2020/survey_results_public.
```

```
survey_raw_df
```

	Respondent	MainBranch	Hobbyist	Age	Age1stCode	CompFreq	CompTotal	ConvertedComp	Country	C
0	1	I am a developer by profession	Yes	NaN	13	Monthly	NaN	NaN	Germany	
1	2	I am a developer by profession	No	NaN	19	NaN	NaN	NaN	United Kingdom	

	Respondent	MainBranch	Hobbyist	Age	Age1stCode	CompFreq	CompTotal	ConvertedComp	Country	C
2	3	I code primarily as a hobby	Yes	NaN	15	NaN	NaN	NaN	Russian Federation	
3	4	I am a developer by profession	Yes	25.0	18	NaN	NaN	NaN	Albania	
4	5	I used to be a developer by profession, but no...	Yes	31.0	16	NaN	NaN	NaN	United States	
...	...	...	...	...	...	...	...	...	...	
64456	64858	NaN	Yes	NaN	16	NaN	NaN	NaN	United States	
64457	64867	NaN	Yes	NaN	NaN	NaN	NaN	NaN	Morocco	
64458	64898	NaN	Yes	NaN	NaN	NaN	NaN	NaN	Viet Nam	
64459	64925	NaN	Yes	NaN	NaN	NaN	NaN	NaN	Poland	
64460	65112	NaN	Yes	NaN	NaN	NaN	NaN	NaN	Spain	

64461 rows × 61 columns

```
survey_raw_df.columns
```

```
Index(['Respondent', 'MainBranch', 'Hobbyist', 'Age', 'Age1stCode', 'CompFreq',
       'CompTotal', 'ConvertedComp', 'Country', 'CurrencyDesc',
       'CurrencySymbol', 'DatabaseDesireNextYear', 'DatabaseWorkedWith',
       'DevType', 'EdLevel', 'Employment', 'Ethnicity', 'Gender', 'JobFactors',
       'JobSat', 'JobSeek', 'LanguageDesireNextYear', 'LanguageWorkedWith',
       'MiscTechDesireNextYear', 'MiscTechWorkedWith',
       'NEWCollabToolsDesireNextYear', 'NEWCollabToolsWorkedWith', 'NEWDevOps',
       'NEWDevOpsImpt', 'NEWEdImpt', 'NEWJobHunt', 'NEWJobHuntResearch',
       'NEWLearn', 'NEWOffTopic', 'NEWOnboardGood', 'NEWOtherComms',
       'NEWOvertime', 'NEWPurchaseResearch', 'NEWPurpleLink', 'NEWSOSites',
       'NEWStuck', 'OpSys', 'OrgSize', 'PlatformDesireNextYear',
       'PlatformWorkedWith', 'PurchaseWhat', 'Sexuality', 'SOAccount',
       'SOComm', 'SOPartFreq', 'SOVisitFreq', 'SurveyEase', 'SurveyLength',
       'Trans', 'UndergradMajor', 'WebframeDesireNextYear',
       'WebframeWorkedWith', 'WelcomeChange', 'WorkWeekHrs', 'YearsCode',
       'YearsCodePro'],
      dtype='object')
```

```
schema_fname = ('stackoverflow-developer-survey-2020/survey_results_schema.csv')
```

```
pd.read_csv(schema_fname, index_col = 'Column').QuestionText
```

```
Column
Respondent             Randomized respondent ID number (not in order ...
MainBranch             Which of the following options best describes ...
Hobbyist                                        Do you code as a hobby?
Age                    What is your age (in years)? If you prefer not...
Age1stCode             At what age did you write your first line of c...
                                                     ...
WebframeWorkedWith     Which web frameworks have you done extensive d...
WelcomeChange          Compared to last year, how welcome do you feel...
WorkWeekHrs            On average, how many hours per week do you wor...
YearsCode              Including any education, how many years have y...
YearsCodePro           NOT including education, how many years have y...
Name: QuestionText, Length: 61, dtype: object
```

```
schema_raw = pd.read_csv(schema_fname, index_col = 'Column').QuestionText
```

```
schema_raw
```

```
Column
Respondent             Randomized respondent ID number (not in order ...
MainBranch             Which of the following options best describes ...
Hobbyist                                        Do you code as a hobby?
Age                    What is your age (in years)? If you prefer not...
Age1stCode             At what age did you write your first line of c...
                                                     ...
WebframeWorkedWith     Which web frameworks have you done extensive d...
WelcomeChange          Compared to last year, how welcome do you feel...
WorkWeekHrs            On average, how many hours per week do you wor...
YearsCode              Including any education, how many years have y...
YearsCodePro           NOT including education, how many years have y...
Name: QuestionText, Length: 61, dtype: object
```

```
schema_raw['YearsCodePro']
```

```
'NOT including education, how many years have you coded professionally (as a part of your work)?'
```

```
project = 'python-survey'
```

```
!pip install jovian --upgrade --quiet
```

```
import jovian
```

```
jovian.commit(project=project)
```

```
[jovian] Updating notebook "ranjithamal231/python-survey" on https://jovian.com
[jovian] Committed successfully! https://jovian.com/ranjithamal231/python-survey
```

'https://jovian.com/ranjithamal231/python-survey'

```
selected_columns = [
   'Country',
    'Age',
    'Gender',
    'EdLevel',
    'UndergradMajor',
    # Programming experience
    'Hobbyist',
    'Age1stCode',
    'YearsCode',
    'YearsCodePro',
    'LanguageWorkedWith',
    'LanguageDesireNextYear',
    'NEWLearn',
    'NEWStuck',
    # Employment
    'Employment',
    'DevType',
    'WorkWeekHrs',
    'JobSat',
    'JobFactors',
    'NEWOvertime',
    'NEWEdImpt'
]
```

```
len(selected_columns)
```

20

```
survey_df = survey_raw_df[selected_columns].copy()
```

```
schema = schema_raw[selected_columns]
```

```
survey_df
```

	Country	Age	Gender	EdLevel	UndergradMajor	Hobbyist	Age1stCode	YearsCode	YearsCodeP
0	Germany	NaN	Man	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	Yes	13	36	
1	United Kingdom	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Computer science, computer engineering, or sof...	No	19	7	

	Country	Age	Gender	EdLevel	UndergradMajor	Hobbyist	Age1stCode	YearsCode	YearsCodeP
2	Russian Federation	NaN	NaN	NaN	NaN	Yes	15	4	N
3	Albania	25.0	Man	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	Yes	18	7	
4	United States	31.0	Man	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Computer science, computer engineering, or sof...	Yes	16	15	
...	...	...	...	...	...	...	...	...	
64456	United States	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	Yes	16	10	Less thar y
64457	Morocco	NaN	NaN	NaN	NaN	Yes	NaN	NaN	N
64458	Viet Nam	NaN	NaN	Primary/elementary school	NaN	Yes	NaN	NaN	N
64459	Poland	NaN	NaN	NaN	NaN	Yes	NaN	NaN	N
64460	Spain	NaN	NaN	Other doctoral degree (Ph.D., Ed.D., etc.)	Computer science, computer engineering, or sof...	Yes	NaN	NaN	N

64461 rows × 20 columns

schema

Column	
Country	Where do you live?
Age	What is your age (in years)? If you prefer not...
Gender	Which of the following describe you, if any? P...
EdLevel	Which of the following best describes the high...
UndergradMajor	What was your primary field of study?
Hobbyist	Do you code as a hobby?
Age1stCode	At what age did you write your first line of c...
YearsCode	Including any education, how many years have y...
YearsCodePro	NOT including education, how many years have y...
LanguageWorkedWith	Which programming, scripting, and markup langu...
LanguageDesireNextYear	Which programming, scripting, and markup langu...
NEWLearn	How frequently do you learn a new language or ...
NEWStuck	What do you do when you get stuck on a problem...
Employment	Which of the following best describes your cur...
DevType	Which of the following describe you? Please se...
WorkWeekHrs	On average, how many hours per week do you wor...

```
JobSat                    How satisfied are you with your current job? (...
JobFactors                Imagine that you are deciding between two job ...
NEWOvertime               How often do you work overtime or beyond the f...
NEWEdImpt                 How important is a formal education, such as a...
Name: QuestionText, dtype: object
```

```
survey_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 64461 entries, 0 to 64460
Data columns (total 20 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   Country                 64072 non-null  object 
 1   Age                     45446 non-null  float64
 2   Gender                  50557 non-null  object 
 3   EdLevel                 57431 non-null  object 
 4   UndergradMajor          50995 non-null  object 
 5   Hobbyist                64416 non-null  object 
 6   Age1stCode              57900 non-null  object 
 7   YearsCode               57684 non-null  object 
 8   YearsCodePro            46349 non-null  object 
 9   LanguageWorkedWith      57378 non-null  object 
 10  LanguageDesireNextYear  54113 non-null  object 
 11  NEWLearn                56156 non-null  object 
 12  NEWStuck                54983 non-null  object 
 13  Employment              63854 non-null  object 
 14  DevType                 49370 non-null  object 
 15  WorkWeekHrs             41151 non-null  float64
 16  JobSat                  45194 non-null  object 
 17  JobFactors              49349 non-null  object 
 18  NEWOvertime             43231 non-null  object 
 19  NEWEdImpt               48465 non-null  object 
dtypes: float64(2), object(18)
memory usage: 9.8+ MB
```

```
survey_df.Age1stCode.unique()
```

```
array(['13', '19', '15', '18', '16', '14', '12', '20', '42', '8', '25',
       '22', '30', '17', '21', '10', '46', '9', '7', '11', '6', nan, '31',
       '29', '5', 'Younger than 5 years', '28', '38', '23', '27', '41',
       '24', '53', '26', '35', '32', '40', '33', '36', '54', '48', '56',
       '45', '44', '34', 'Older than 85', '39', '51', '68', '50', '37',
       '47', '43', '52', '85', '64', '55', '58', '49', '76', '72', '73',
       '83', '63'], dtype=object)
```

```
survey_df['Age1stCode'] = pd.to_numeric(survey_df.Age1stCode, errors='coerce')
survey_df['YearsCode'] = pd.to_numeric(survey_df.YearsCode, errors='coerce')
survey_df['YearsCodePro'] = pd.to_numeric(survey_df.YearsCodePro, errors='coerce')
```

```
survey_df.drop(survey_df[survey_df.Age < 10].index, inplace=True)
survey_df.drop(survey_df[survey_df.Age > 100].index, inplace=True)
```

```
survey_df.drop(survey_df[survey_df.WorkWeekHrs > 140].index, inplace=True)
```

```
survey_df.describe()
```

	Age	Age1stCode	YearsCode	YearsCodePro	WorkWeekHrs
count	45319.000000	57326.000000	56636.000000	43993.000000	41002.000000
mean	30.832322	15.475317	12.783883	8.873003	40.024395
std	9.505965	5.114952	9.494519	7.762089	10.630010
min	10.000000	5.000000	1.000000	1.000000	1.000000
25%	24.000000	12.000000	6.000000	3.000000	40.000000
50%	29.000000	15.000000	10.000000	6.000000	40.000000
75%	35.000000	18.000000	17.000000	12.000000	43.000000
max	99.000000	85.000000	50.000000	50.000000	140.000000

```
import numpy as np
```

```
survey_df.where(~(survey_df.Gender.str.contains(';', na=False)),np.nan, inplace=True)
```

```
survey_df['Gender'].value_counts()
```

```
Man                                                 45895
Woman                                                3835
Non-binary, genderqueer, or gender non-conforming     385
Name: Gender, dtype: int64
```

```
survey_df.sample(10)
```

	Country	Age	Gender	EdLevel	UndergradMajor	Hobbyist	Age1stCode	YearsCode	YearsCodePr
40986	Russian Federation	42.0	Man	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Mathematics or statistics	Yes	13.0	30.0	8.
23915	Poland	20.0	Man	Secondary school (e.g. American high school, G...	NaN	Yes	17.0	4.0	1.

	Country	Age	Gender	EdLevel	UndergradMajor	Hobbyist	Age1stCode	YearsCode	YearsCodePr
2722	Russian Federation	40.0	Man	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	Yes	14.0	25.0	21.
9610	Turkey	NaN	Man	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	No	18.0	16.0	12.
4750	Mexico	23.0	Woman	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Computer science, computer engineering, or sof...	Yes	18.0	5.0	4.
3199	United States	31.0	Man	Bachelor's degree (B.A., B.S., B.Eng., etc.)	A business discipline (such as accounting, fin...	No	25.0	5.0	Nal
24426	Italy	26.0	Man	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Computer science, computer engineering, or sof...	Yes	19.0	6.0	4.
48869	Uruguay	NaN	NaN	Some college/university study without earning ...	Computer science, computer engineering, or sof...	No	22.0	9.0	6.
19655	United States	24.0	Man	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Computer science, computer engineering, or sof...	Yes	18.0	6.0	2.
55352	Ukraine	NaN	NaN	NaN	NaN	Yes	13.0	NaN	Nal

```
import jovian
```

```
jovian.commit()
```

```
[jovian] Updating notebook "ranjithamal231/python-survey" on https://jovian.com
[jovian] Committed successfully! https://jovian.com/ranjithamal231/python-survey
```

```
'https://jovian.com/ranjithamal231/python-survey'
```

```
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline

sns.set_style('darkgrid')
matplotlib.rcParams['font.size'] = 14
matplotlib.rcParams['figure.figsize'] = (9, 5)
matplotlib.rcParams['figure.facecolor'] = '#00000000'
```

```
schema.Country
```

```
'Where do you live?'
```

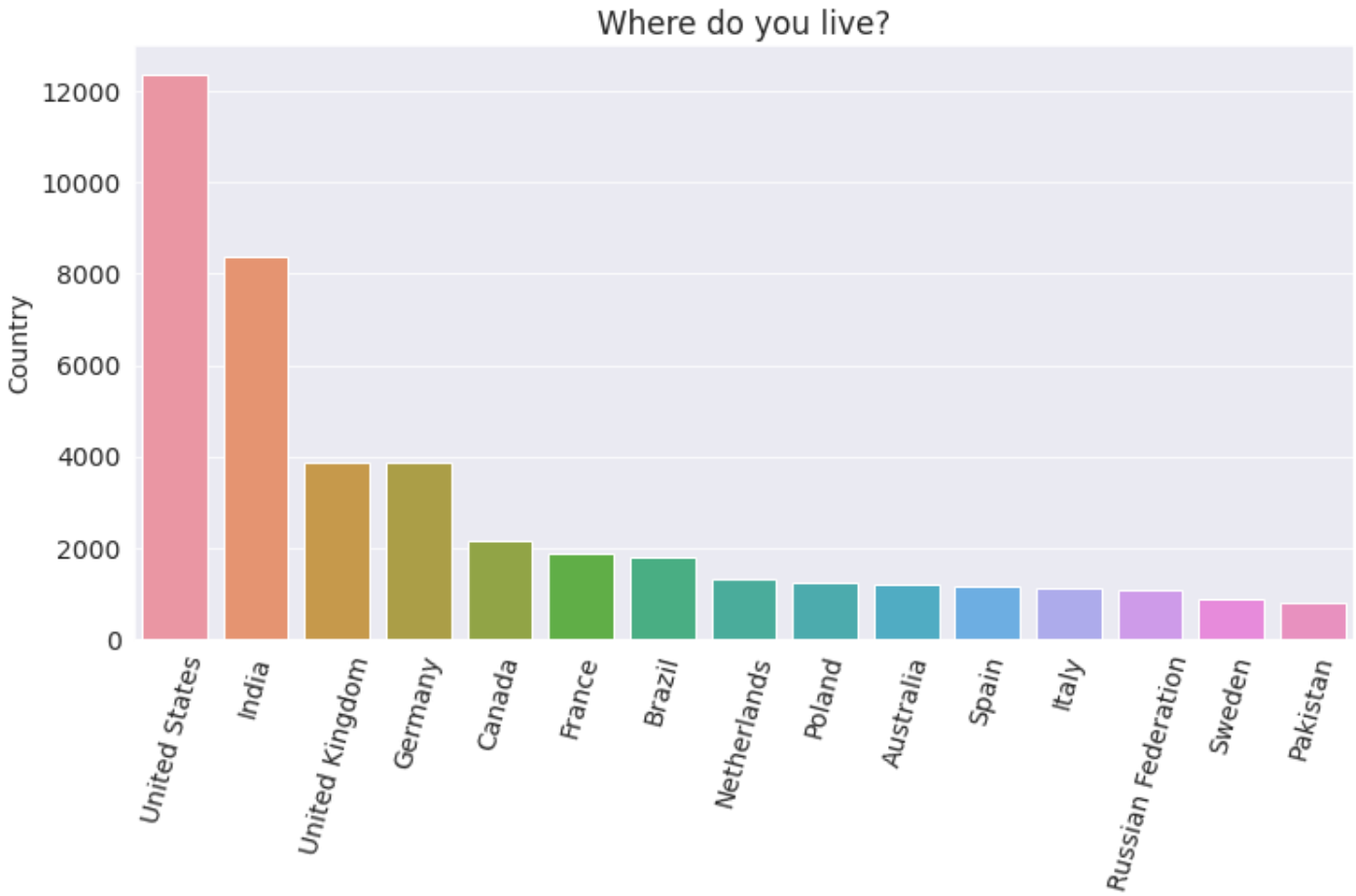
```
survey_df.Country.nunique()
```

```
183
```

```
top_countries = survey_df.Country.value_counts().head(15)
top_countries
```

```
United States         12371
India                  8364
United Kingdom         3881
Germany                3864
Canada                 2175
France                 1884
Brazil                 1804
Netherlands            1332
Poland                 1259
Australia              1199
Spain                  1157
Italy                  1115
Russian Federation     1085
Sweden                  879
Pakistan                802
Name: Country, dtype: int64
```

```
plt.figure(figsize=(12,6))
plt.xticks(rotation=75)
plt.title(schema.Country)
sns.barplot(x=top_countries.index, y=top_countries);
```




```
countries_languages_df = pd.read_csv('stackoverflow-developer-survey-2020/countries_lan
```

```
countries_languages_df
```

	Country	Languages Spoken
0	Afghanistan	Dari Persian, Pashtu (both official), other Tu...
1	Albania	Albanian (Tosk is the official dialect), Greek
2	Algeria	Arabic (official), French, Berber dialects
3	Andorra	Catalán (official), French, Castilian, Portuguese
4	Angola	Portuguese (official), Bantu and other African...
...	...	...
193	Vietnam	Vietnamese (official); English (increasingly f...
194	Western Sahara (proposed state)	Hassaniya Arabic, Moroccan Arabic
195	Yemen	Arabic
196	Zambia	English (official); major vernaculars: Bemba, ...
197	Zimbabwe	English (official), Shona, Ndebele (Sindebele)...

198 rows × 2 columns

```
countries_languages_df.columns
```

```
Index(['Country', 'Languages Spoken'], dtype='object')
```

```
countries_languages_df.Country.nunique()
```

```
198
```

```
def split_multicolumn(col_series):
    result_df = col_series.to_frame()
    options = []
    # Iterate over the column
    for idx, value  in col_series[col_series.notnull()].iteritems():
        # Break each value into list of options
        for option in value.split(','):
            # Add the option as a column to result
            if not option in result_df.columns:
                options.append(option)
                result_df[option] = False
            # Mark the value in the option column as True
            result_df.at[idx, option] = True
    return result_df[options]
```

```
de_type_df = split_multicolumn(merged_df.Languagessp)
de_type_df
```

```
/tmp/ipykernel_64/1262042527.py:11: PerformanceWarning: DataFrame is highly fragmented.
This is usually the result of calling `frame.insert` many times, which has poor
performance.  Consider joining all columns at once using pd.concat(axis=1) instead.  To
get a de-fragmented frame, use `newframe = frame.copy()`
  result_df[option] = False
```

	German	English	Welsh	Scots Gaelic	Albanian (Tosk is the official dialect)	Greek	English 82%	Spanish 11% (2000)	Hindi 30%	English	...	120 indigenous languages	Portugues
0	True	False	False	False	False	False	False	False	False	False	...	False	Fals
1	True	False	False	False	False	False	False	False	False	False	...	False	Fals
2	True	False	False	False	False	False	False	False	False	False	...	False	Fals
3	True	False	False	False	False	False	False	False	False	False	...	False	Fals
4	True	False	False	False	False	False	False	False	False	False	...	False	Fals
...	...	...	...	...	...	...	...	...	...	...	...	...	.
61479	False	False	False	False	False	False	False	False	False	False	...	True	Fals
61480	False	False	False	False	False	False	False	False	False	False	...	False	Tru
61481	False	False	False	False	False	False	False	False	False	False	...	False	Tru
61482	False	False	False	False	False	False	False	False	False	False	...	False	Fals

	German	English	Welsh	Scots Gaelic	Albanian (Tosk is the official dialect)	Greek	English 82%	Spanish 11% (2000)	Hindi 30%	English	...	120 indigenous languages	Portugues
61483	False	False	False	False	False	False	False	False	False	False	...	False	Fals

61484 rows × 447 columns

```
countries_languages_df[countries_languages_df.Country == "India"]
```

	Country	Languages Spoken
76	India	Hindi 30%, English, Bengali, Gujarati, Kashmir...

```
merged_df = survey_df.merge(countries_languages_df, on="Country")
```

```
merged_df
```

	Country	Age	Gender	EdLevel	UndergradMajor	Hobbyist	Age1stCode	YearsCode	YearsCodePro
0	Germany	NaN	Man	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	Yes	13.0	36.0	27.0
1	Germany	NaN	Man	Secondary school (e.g. American high school, G...	NaN	No	14.0	6.0	4.0
2	Germany	27.0	Man	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Another engineering discipline (such as civil,...	Yes	14.0	8.0	3.0
3	Germany	45.0	Man	Professional degree (JD, MD, etc.)	Another engineering discipline (such as civil,...	Yes	14.0	30.0	20.0
4	Germany	29.0	Man	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	Yes	15.0	10.0	NaN
...	...	...	...	...	...	...	...	...	...
61479	Solomon Islands	25.0	Man	Secondary school (e.g. American high school, G...	NaN	Yes	12.0	NaN	NaN
61480	Cape Verde	NaN	NaN	Some college/university study without earning ...	Information systems, information technology, o...	Yes	22.0	7.0	NaN

	Country	Age	Gender	EdLevel	UndergradMajor	Hobbyist	Age1stCode	YearsCode	YearsCodePro
61481	Cape Verde	35.0	Man	Some college/university study without earning ...	Information systems, information technology, o...	Yes	21.0	5.0	NaN
61482	Gabon	NaN	NaN	NaN	NaN	Yes	NaN	NaN	NaN
61483	Marshall Islands	NaN	NaN	Some college/university study without earning ...	A health science (such as nursing, pharmacy, r...	Yes	20.0	20.0	NaN

61484 rows × 21 columns

```
merged_df.rename(columns={'Languages Spoken': 'Languagessp'}, inplace=True)
```

```
merged_df
```

	Country	Age	Gender	EdLevel	UndergradMajor	Hobbyist	Age1stCode	YearsCode	YearsCodePro
0	Germany	NaN	Man	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	Yes	13.0	36.0	27.0
1	Germany	NaN	Man	Secondary school (e.g. American high school, G...	NaN	No	14.0	6.0	4.0
2	Germany	27.0	Man	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Another engineering discipline (such as civil,...	Yes	14.0	8.0	3.0
3	Germany	45.0	Man	Professional degree (JD, MD, etc.)	Another engineering discipline (such as civil,...	Yes	14.0	30.0	20.0
4	Germany	29.0	Man	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	Yes	15.0	10.0	NaN
...	...	...	...	...	...	...	...	...	...
61479	Solomon Islands	25.0	Man	Secondary school (e.g. American high school, G...	NaN	Yes	12.0	NaN	NaN
61480	Cape Verde	NaN	NaN	Some college/university study without earning ...	Information systems, information technology, o...	Yes	22.0	7.0	NaN

	Country	Age	Gender	EdLevel	UndergradMajor	Hobbyist	Age1stCode	YearsCode	YearsCodePro
61481	Cape Verde	35.0	Man	Some college/university study without earning ...	Information systems, information technology, o...	Yes	21.0	5.0	NaN
61482	Gabon	NaN	NaN	NaN	NaN	Yes	NaN	NaN	NaN
61483	Marshall Islands	NaN	NaN	Some college/university study without earning ...	A health science (such as nursing, pharmacy, r...	Yes	20.0	20.0	NaN

61484 rows × 21 columns

```
merged_df = ['LanguageDesireNextYear']
```

```
interested_df = merged_df['LanguageDesireNextYear']
interested_df
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
/tmp/ipykernel_64/2204227420.py in <module>
----> 1 interested_df = merged_df['LanguageDesireNextYear']
      2 interested_df

TypeError: list indices must be integers or slices, not str
```

```
for col in merged_df.columns:
    print(col)
```

```
Country
Age
Gender
EdLevel
UndergradMajor
Hobbyist
Age1stCode
YearsCode
YearsCodePro
LanguageWorkedWith
LanguageDesireNextYear
NEWLearn
NEWStuck
Employment
DevType
WorkWeekHrs
```

JobSat

JobFactors

NEWOvertime

NEWEdImpt

Languagessp

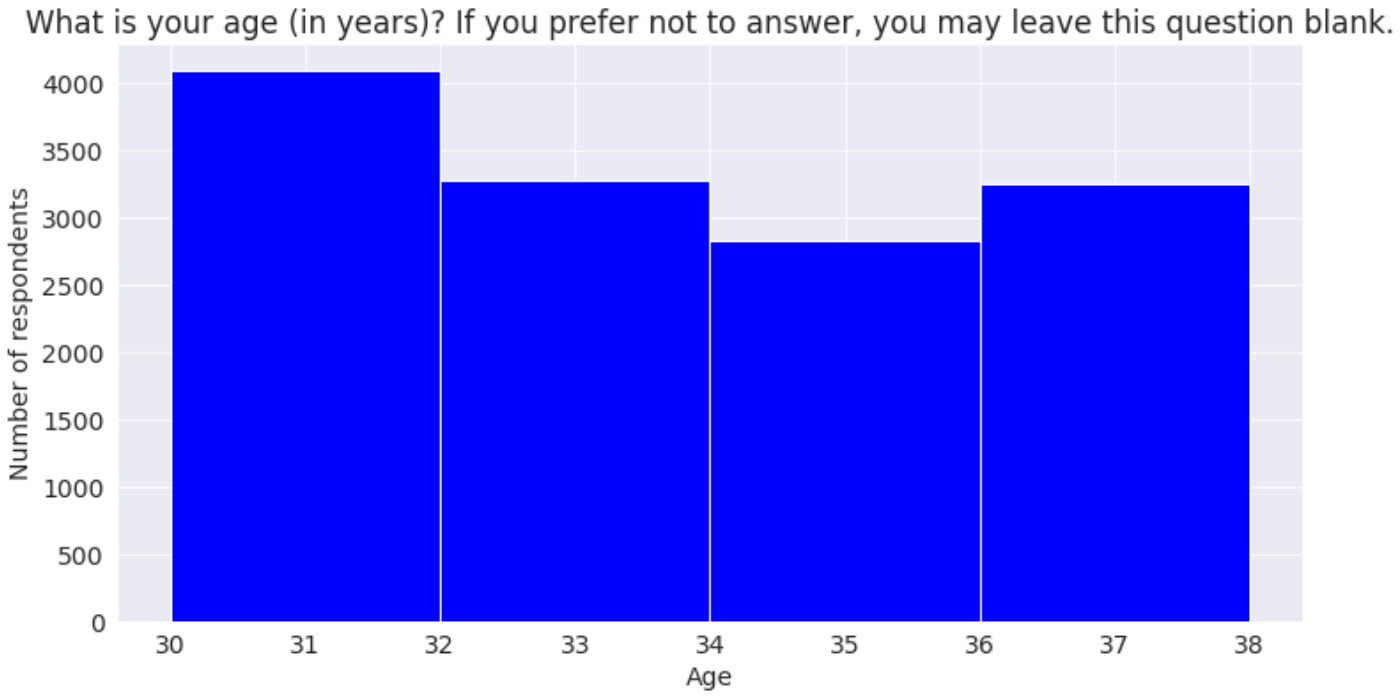
```
merged_df.Languagessp.nunique()
```

156

```
schema.Age
```

'What is your age (in years)? If you prefer not to answer, you may leave this question blank.'

```
plt.figure(figsize=(12, 6))
plt.title(schema.Age)
plt.xlabel('Age')
plt.ylabel('Number of respondents')

plt.hist(survey_df.Age, bins=np.arange(30,40,2), color='Blue');
```




```
schema.Gender
```

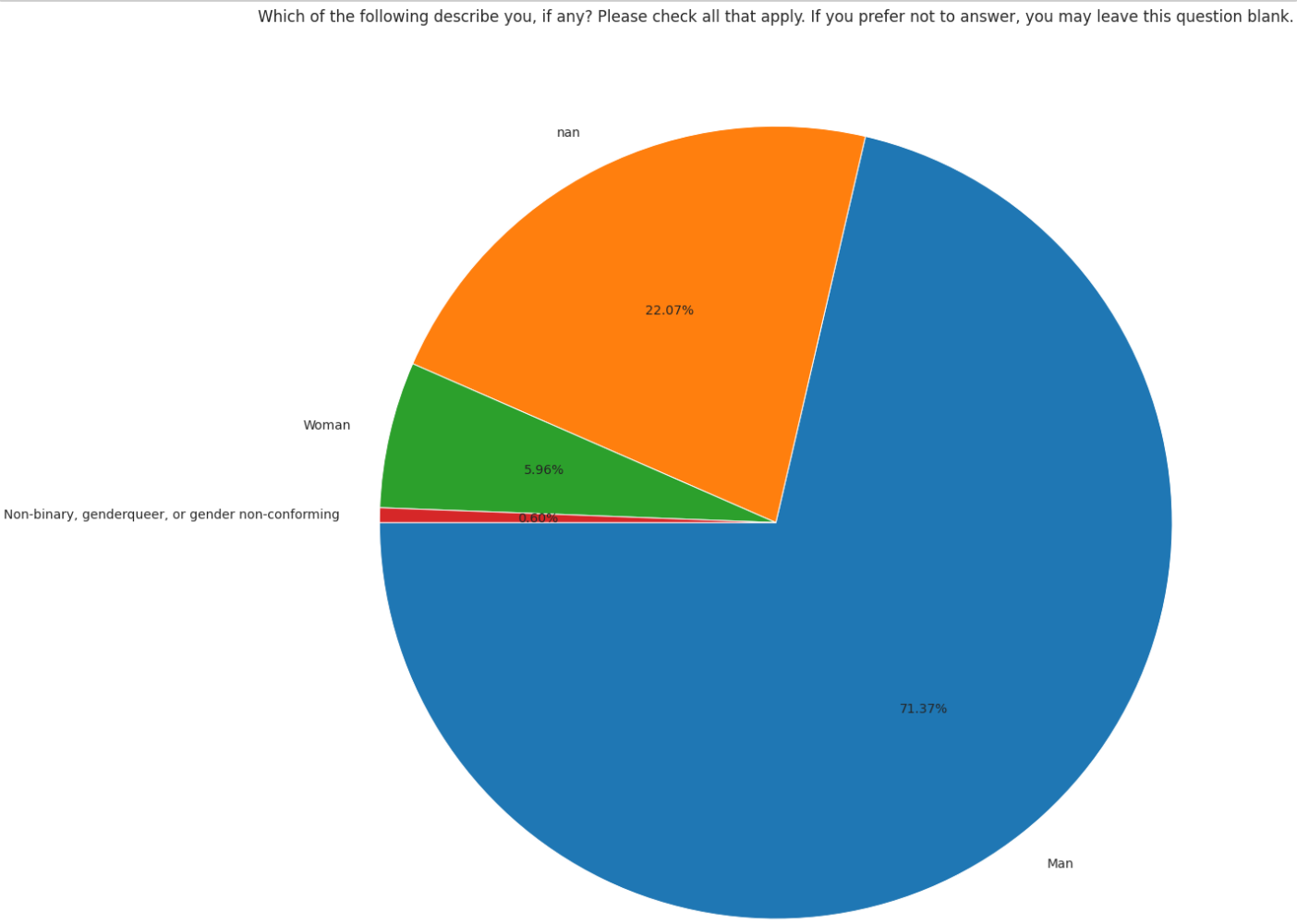
'Which of the following describe you, if any? Please check all that apply. If you prefer not to answer, you may leave this question blank.'

```
gender_counts = survey_df.Gender.value_counts(dropna=False)
gender_counts
```

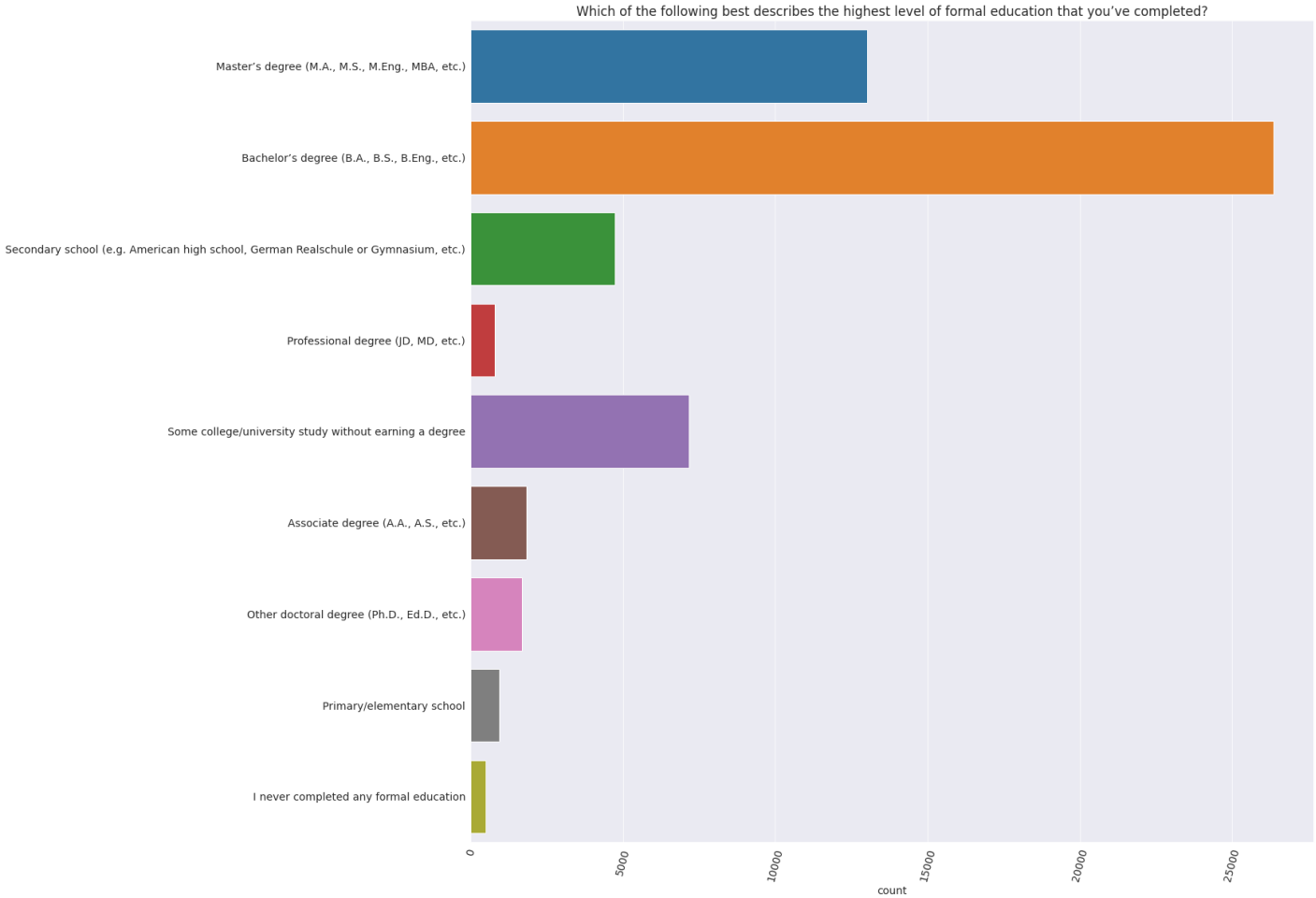
Man	45895
NaN	14191

```
Woman                                                   3835
Non-binary, genderqueer, or gender non-conforming        385
Name: Gender, dtype: int64
```

```
plt.figure(figsize=(20,20))
plt.title(schema.Gender)
plt.pie(gender_counts, labels=gender_counts.index,autopct='%2.2f%%', startangle=180);
```




```
plt.figure(figsize=(20,20))
sns.countplot(y=survey_df.EdLevel)
plt.xticks(rotation=75);
plt.title(schema['EdLevel'])
plt.ylabel(None);
```

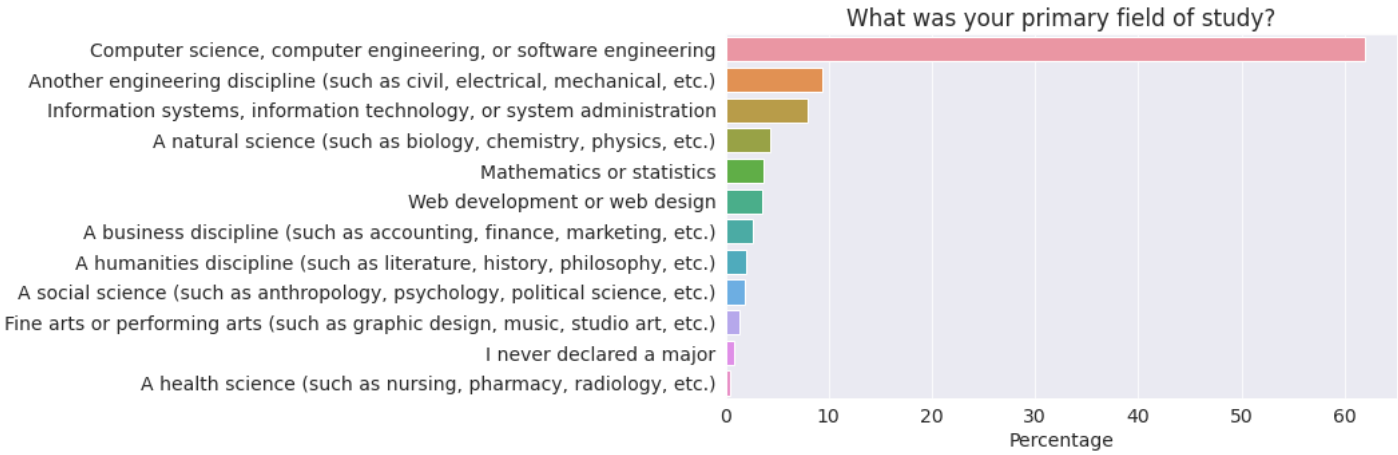



```
undergrad_pct = survey_df.UndergradMajor.value_counts() * 100 / survey_df.UndergradMajc
undergrad_pct
```

```
Computer science, computer engineering, or software engineering
61.939193
Another engineering discipline (such as civil, electrical, mechanical, etc.)
9.354195
Information systems, information technology, or system administration
7.983168
A natural science (such as biology, chemistry, physics, etc.)
4.316561
Mathematics or statistics
3.627097
Web development or web design
3.502637
A business discipline (such as accounting, finance, marketing, etc.)
2.700567
A humanities discipline (such as literature, history, philosophy, etc.)
1.969616
A social science (such as anthropology, psychology, political science, etc.)
1.821450
Fine arts or performing arts (such as graphic design, music, studio art, etc.)
1.412513
I never declared a major
0.885043
A health science (such as nursing, pharmacy, radiology, etc.)
```

```
0.487959
Name: UndergradMajor, dtype: float64
```

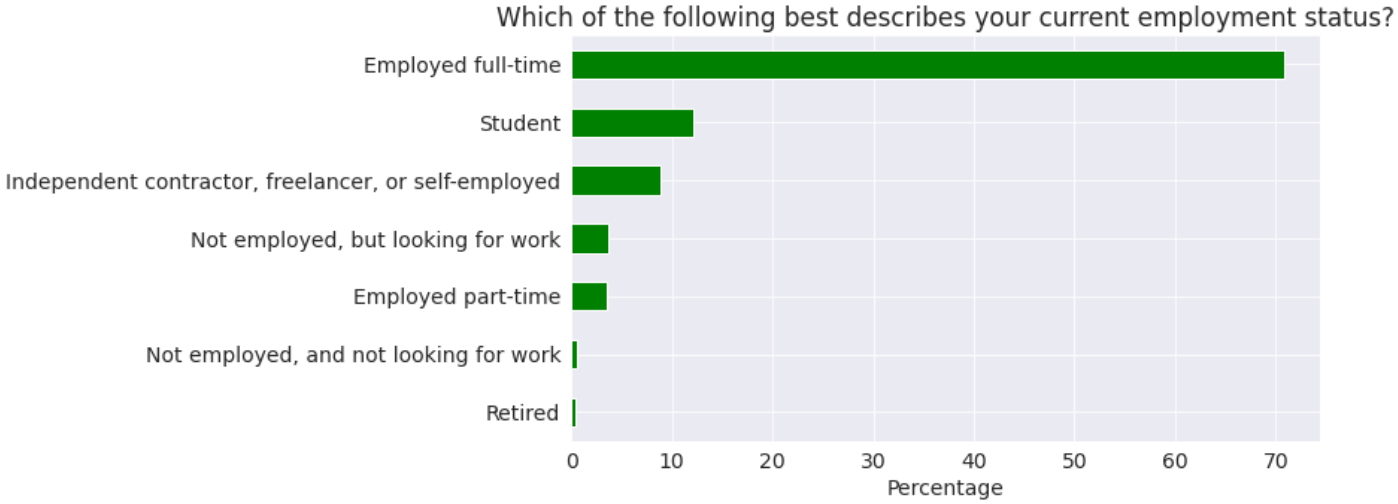
```
undergrad_pct = survey_df.UndergradMajor.value_counts() * 100 / survey_df.UndergradMajc

sns.barplot(x=undergrad_pct, y=undergrad_pct.index)

plt.title(schema.UndergradMajor)
plt.ylabel(None);
plt.xlabel('Percentage');
```




```
schema.Employment
```

```
'Which of the following best describes your current employment status?'
```

```
(survey_df.Employment.value_counts(normalize=True, ascending=True)*100).plot(kind='barh
plt.title(schema.Employment)
plt.xlabel('Percentage');
```




```
schema.DevType
```

```
'Which of the following describe you? Please select all that apply.'
```

```
survey_df.DevType
```

```
0        Developer, desktop or enterprise applications;...
1                  Developer, full-stack;Developer, mobile
2                                                      NaN
3                                                      NaN
4                                                      NaN
                              ...
64456                                  Senior executive/VP
64457                                                  NaN
64458                                                  NaN
64459                                                  NaN
64460                                                  NaN
Name: DevType, Length: 64306, dtype: object
```

```
def split_multicolumn(col_series):
    result_df = col_series.to_frame()
    options = []
    # Iterate over the column
    for idx, value  in col_series[col_series.notnull()].iteritems():
        # Break each value into list of options
        for option in value.split(';'):
            # Add the option as a column to result
            if not option in result_df.columns:
                options.append(option)
                result_df[option] = False
            # Mark the value in the option column as True
            result_df.at[idx, option] = True
    return result_df[options]
```

```
dev_type_df = split_multicolumn(survey_df.DevType)
```

```
dev_type_df
```

	Developer, desktop or enterprise applications	Developer, full-stack	Developer, mobile	Designer	Developer, front-end	Developer, back-end	Developer, QA or test	DevOps specialist	Developer, game or graphics	Data administ
0	True	True	False	False	False	False	False	False	False	
1	False	True	True	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	
64456	False	False	False	False	False	False	False	False	False	
64457	False	False	False	False	False	False	False	False	False	
64458	False	False	False	False	False	False	False	False	False	

	Developer, desktop or enterprise applications	Developer, full-stack	Developer, mobile	Designer	Developer, front-end	Developer, back-end	Developer, QA or test	DevOps specialist	Developer, game or graphics	Data administ
64459	False	False	False	False	False	False	False	False	False	
64460	False	False	False	False	False	False	False	False	False	

64306 rows × 23 columns

```
dev_type_total = dev_type_df.sum().sort_values(ascending=False)
dev_type_total
```

```
Developer, back-end                             26996
Developer, full-stack                           26915
Developer, front-end                            18128
Developer, desktop or enterprise applications   11687
Developer, mobile                                9406
DevOps specialist                                5915
Database administrator                           5658
Designer                                         5262
System administrator                             5185
Developer, embedded applications or devices      4701
Data or business analyst                         3970
Data scientist or machine learning specialist    3939
Developer, QA or test                            3893
Engineer, data                                   3700
Academic researcher                              3502
Educator                                         2895
Developer, game or graphics                      2751
Engineering manager                              2699
Product manager                                  2471
Scientist                                        2060
Engineer, site reliability                       1921
Senior executive/VP                              1292
Marketing or sales professional                   625
dtype: int64
```

```
import jovian
```

```
jovian.commit()
```

```
[jovian] Updating notebook "ranjithamal231/python-survey" on https://jovian.com
[jovian] Committed successfully! https://jovian.com/ranjithamal231/python-survey
```

```
'https://jovian.com/ranjithamal231/python-survey'
```

```
languages_worked_df = merged_df['LanguageWorkedWith']
languages_worked_df
```

```
0                                    C#;HTML/CSS;JavaScript
1                                  HTML/CSS;Java;JavaScript
2                               Bash/Shell/PowerShell;C#;C++
3                  Bash/Shell/PowerShell;Java;Kotlin;PHP;SQL
4                    HTML/CSS;Java;JavaScript;SQL;TypeScript
                                    ...
61479                                               C++;Java
61480                                                    NaN
61481                                  C;Java;JavaScript;PHP;SQL
61482    Bash/Shell/PowerShell;C++;HTML/CSS;Java;JavaSc...
61483                                               Assembly
Name: LanguageWorkedWith, Length: 61484, dtype: object
```

```
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
/tmp/ipykernel_64/452334947.py in <module>
----> 1 languages_interested_df = merged_df.int(input('LanguageDesireNextYear'))
      2 languages_interested_df

AttributeError: 'list' object has no attribute 'int'
```

```
languages_loved_df = languages_worked_df & LanguageDesireNextYear
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipykernel_64/1832210226.py in <module>
----> 1 languages_loved_df = languages_worked_df & LanguageDesireNextYear

NameError: name 'LanguageDesireNextYear' is not defined
```

```
```