Data Wrangling

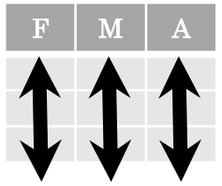
with pandas

Cheat Sheet

http://pandas.pydata.org

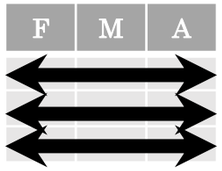
Tidy Data – A foundation for wrangling in pandas

In a tidy data set:



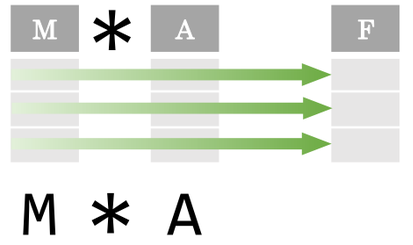
Each **variable** is saved in its own **column**

&



Each **observation** is saved in its own **row**

Tidy data complements pandas's **vectorized operations**. pandas will automatically preserve observations as you manipulate variables. No other format works as intuitively with pandas.



Syntax – Creating DataFrames

	a	b	c
1	4	7	10
2	5	8	11
3	6	9	12

```
df = pd.DataFrame(
        {"a" : [4 ,5, 6],
         "b" : [7, 8, 9],
         "c" : [10, 11, 12]},
       index = [1, 2, 3])
```

Specify values for each column.

```
df = pd.DataFrame(
    [[4, 7, 10],
     [5, 8, 11],
     [6, 9, 12]],
    index=[1, 2, 3],
    columns=['a', 'b', 'c'])
```

Specify values for each row.

		a	b	c
n	v			
d	1	4	7	10
	2	5	8	11
e	2	6	9	12

```
df = pd.DataFrame(
        {"a" : [4 ,5, 6],
         "b" : [7, 8, 9],
         "c" : [10, 11, 12]},
index = pd.MultiIndex.from_tuples(
        [('d',1),('d',2),('e',2)],
          names=['n','v']))
```

Create DataFrame with a MultiIndex

Method Chaining

Most pandas methods return a DataFrame so that another pandas method can be applied to the result. This improves readability of code.

```
df = (pd.melt(df)
        .rename(columns={
                'variable' : 'var',
                'value' : 'val'})
        .query('val >= 200')
    )
```

Reshaping Data – Change the layout of a data set

pd.melt(df)
Gather columns into rows.

df.pivot(columns='var', values='val')
Spread rows into columns.

pd.concat([df1,df2])
Append rows of DataFrames

pd.concat([df1,df2], axis=1)
Append columns of DataFrames

df.sort_values('mpg')
Order rows by values of a column (low to high).

df.sort_values('mpg',ascending=False)
Order rows by values of a column (high to low).

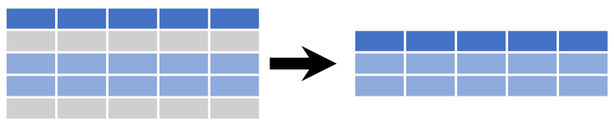
df.rename(columns = {'y':'year'})
Rename the columns of a DataFrame

df.sort_index()
Sort the index of a DataFrame

df.reset_index()
Reset index of DataFrame to row numbers, moving index to columns.

df.drop(columns=['Length','Height'])
Drop columns from DataFrame

Subset Observations (Rows)



df[df.Length > 7]
Extract rows that meet logical criteria.

df.drop_duplicates()
Remove duplicate rows (only considers columns).

df.head(n)
Select first n rows.

df.tail(n)
Select last n rows.

df.sample(frac=0.5)
Randomly select fraction of rows.

df.sample(n=10)
Randomly select n rows.

df.iloc[10:20]
Select rows by position.

df.nlargest(n, 'value')
Select and order top n entries.

df.nsmallest(n, 'value')
Select and order bottom n entries.

Logic in Python (and pandas)			
<	Less than	!=	Not equal to
>	Greater than	df.column.isin(*values*)	Group membership
==	Equals	pd.isnull(*obj*)	Is NaN
<=	Less than or equals	pd.notnull(*obj*)	Is not NaN
>=	Greater than or equals	&,|,~,^,df.any(),df.all()	Logical and, or, not, xor, any, all

Subset Variables (Columns)

df[['width','length','species']]
Select multiple columns with specific names.

df['width'] *or* **df.width**
Select single column with specific name.

df.filter(regex='*regex*')
Select columns whose name matches regular expression *regex*.

regex (Regular Expressions) Examples	
'\.'	Matches strings containing a period '.'
'Length$'	Matches strings ending with word 'Length'
'^Sepal'	Matches strings beginning with the word 'Sepal'
'^x[1-5]$'	Matches strings beginning with 'x' and ending with 1,2,3,4,5
'^(?!Species$).*'	Matches strings except the string 'Species'

df.loc[:,'x2':'x4']
Select all columns between x2 and x4 (inclusive).

df.iloc[:,[1,2,5]]
Select columns in positions 1, 2 and 5 (first column is 0).

df.loc[df['a'] > 10, ['a','c']]
Select rows meeting logical condition, and only the specific columns .

Summarize Data

`df['w'].value_counts()`
Count number of rows with each unique value of variable

`len(df)`
# of rows in DataFrame.

`df['w'].nunique()`
# of distinct values in a column.

`df.describe()`
Basic descriptive statistics for each column (or GroupBy)

pandas provides a large set of **summary functions** that operate on different kinds of pandas objects (DataFrame columns, Series, GroupBy, Expanding and Rolling (see below)) and produce single values for each of the groups. When applied to a DataFrame, the result is returned as a pandas Series for each column. Examples:

`sum()`
Sum values of each object.

`count()`
Count non-NA/null values of each object.

`median()`
Median value of each object.

`quantile([0.25,0.75])`
Quantiles of each object.

`apply(function)`
Apply function to each object.

`min()`
Minimum value in each object.

`max()`
Maximum value in each object.

`mean()`
Mean value of each object.

`var()`
Variance of each object.

`std()`
Standard deviation of each object.

Group Data

`df.groupby(by="col")`
Return a GroupBy object, grouped by values in column named "col".

`df.groupby(level="ind")`
Return a GroupBy object, grouped by values in index level named "ind".

All of the summary functions listed above can be applied to a group. Additional GroupBy functions:

`size()`
Size of each group.

`agg(function)`
Aggregate group using function.

The examples below can also be applied to groups. In this case, the function is applied on a per-group basis, and the returned vectors are of the length of the original DataFrame.

`shift(1)`
Copy with values shifted by 1.

`rank(method='dense')`
Ranks with no gaps.

`rank(method='min')`
Ranks. Ties get min rank.

`rank(pct=True)`
Ranks rescaled to interval [0, 1].

`rank(method='first')`
Ranks. Ties go to first value.

`shift(-1)`
Copy with values lagged by 1.

`cumsum()`
Cumulative sum.

`cummax()`
Cumulative max.

`cummin()`
Cumulative min.

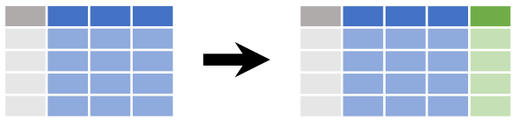
`cumprod()`
Cumulative product.

Handling Missing Data

`df.dropna()`
Drop rows with any column having NA/null data.

`df.fillna(value)`
Replace all NA/null data with value.

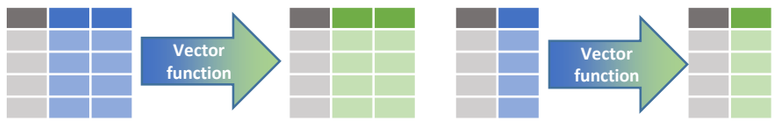
Make New Columns



`df.assign(Area=lambda df: df.Length*df.Height)`
Compute and append one or more new columns.

`df['Volume'] = df.Length*df.Height*df.Depth`
Add single column.

`pd.qcut(df.col, n, labels=False)`
Bin column into n buckets.



pandas provides a large set of **vector functions** that operate on all columns of a DataFrame or a single selected column (a pandas Series). These functions produce vectors of values for each of the columns, or a single Series for the individual Series. Examples:

`max(axis=1)`
Element-wise max.

`clip(lower=-10,upper=10)`
Trim values at input thresholds

`min(axis=1)`
Element-wise min.

`abs()`
Absolute value.

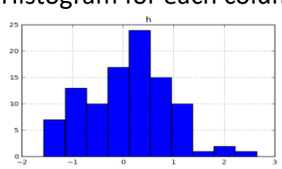
Windows

`df.expanding()`
Return an Expanding object allowing summary functions to be applied cumulatively.

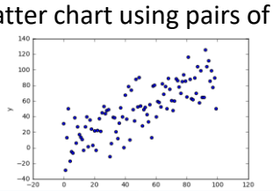
`df.rolling(n)`
Return a Rolling object allowing summary functions to be applied to windows of length n.

Plotting

`df.plot.hist()`
Histogram for each column



`df.plot.scatter(x='w',y='h')`
Scatter chart using pairs of points



Combine Data Sets

adf

x1	x2
A	1
B	2
C	3

+

bdf

x1	x3
A	T
B	F
D	T

=

Standard Joins

x1	x2	x3
A	1	T
B	2	F
C	3	NaN

`pd.merge(adf, bdf, how='left', on='x1')`
Join matching rows from bdf to adf.

x1	x2	x3
A	1.0	T
B	2.0	F
D	NaN	T

`pd.merge(adf, bdf, how='right', on='x1')`
Join matching rows from adf to bdf.

x1	x2	x3
A	1	T
B	2	F

`pd.merge(adf, bdf, how='inner', on='x1')`
Join data. Retain only rows in both sets.

x1	x2	x3
A	1	T
B	2	F
C	3	NaN
D	NaN	T

`pd.merge(adf, bdf, how='outer', on='x1')`
Join data. Retain all values, all rows.

Filtering Joins

x1	x2
A	1
B	2

`adf[adf.x1.isin(bdf.x1)]`
All rows in adf that have a match in bdf.

x1	x2
C	3

`adf[~adf.x1.isin(bdf.x1)]`
All rows in adf that do not have a match in bdf.

ydf

x1	x2
A	1
B	2
C	3

+

zdf

x1	x2
B	2
C	3
D	4

=

Set-like Operations

x1	x2
B	2
C	3

`pd.merge(ydf, zdf)`
Rows that appear in both ydf and zdf (Intersection).

x1	x2
A	1
B	2
C	3
D	4

`pd.merge(ydf, zdf, how='outer')`
Rows that appear in either or both ydf and zdf (Union).

x1	x2
A	1

```
pd.merge(ydf, zdf, how='outer',
         indicator=True)
.query('_merge == "left_only"')
.drop(columns=['_merge'])
```
Rows that appear in ydf but not zdf (Setdiff).

In [1]:

```
import pandas as pd
import numpy as np
pd.set_option("display.max_columns",None)
```

In [2]:

```
df = pd.DataFrame({
    "a":[4,5,6,7,8,9],
    "b":[7,8,9,10,11,12],
    "c":[10,12,14,15,16,18]},
    index = [1,2,3,4,5,6]
)
```

In [3]:

```
df
```

Out[3]:

	a	b	c
1	4	7	10
2	5	8	12
3	6	9	14
4	7	10	15
5	8	11	16
6	9	12	18

In [4]:

```
df = pd.DataFrame(
[[4, 7, 10],
[5, 8, 11],
[6, 9, 12]],
index = pd.MultiIndex.from_tuples(
[('d',1),('d',2),('e',2)],
names=['n','v']))
```

In [5]:

```
df
```

Out[5]:

		0	1	2
n	v			
d	1	4	7	10
	2	5	8	11
e	2	6	9	12

In [6]:

```
df
```

Out[6]:

		0	1	2
n	v			

		a	b	c
d	1	4	7	10
	2	5	8	11
e	2	6	9	12

In [7]:

```
df = pd.DataFrame(
{"a" : [4 ,5, 6],
"b" : [7, 8, 9],
"c" : [10, 11, 12]},
index = pd.MultiIndex.from_tuples(
[('d',1),('d',2),('e',2)],
names=['n','v']))
```

In [8]:

```
df
```

Out[8]:

		a	b	c
n	v			
d	1	4	7	10
	2	5	8	11
e	2	6	9	12

In [9]:

```
df = (pd.melt(df)
.rename(columns={
'variable' : 'var',
'value' : 'val'})
.query('val >= 200')
)
```

In [10]:

```
df
```

Out[10]:

var	val

In [11]:

```
df1 = pd.read_csv("/Users/narenderbeniwal/Downloads/stack-overflow-developer-survey-2021/survey_results_public.csv")
pd.set_option("display.max_columns", None)
```

In [12]:

```
df1
```

Out[12]:

	ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnC
0	1	I am a developer by profession	Independent contractor, freelancer, or self-em...	Slovakia	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Coc Bootcamp;Ot online resour (ex:
1	2	I am a student who is learning to	Student, full-time	Netherlands	NaN	NaN	Bachelor's degree (B.A., B.S.,	11 - 17 years	Other on resources videos, blogs, e

	ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnC
		learning to code					B.Eng., etc.)		videos, blogs, e
2	3	I am not primarily a developer, but I write co...	Student, full-time	Russian Federation	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other on resources videos, blogs, e
3	4	I am a developer by profession	Employed full-time	Austria	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	11 - 17 years	N
4	5	I am a developer by profession	Independent contractor, freelancer, or self-em...	United Kingdom of Great Britain and Northern I...	NaN	England	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	5 - 10 years	Friend or far mem
...	...	...	...	...	...	...	...	...	
83434	83435	I am a developer by profession	Employed full-time	United States of America	Texas	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other on resources videos, blogs, e
83435	83436	I am a developer by profession	Independent contractor, freelancer, or self-em...	Benin	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other on resources videos, blogs, e
83436	83437	I am a developer by profession	Employed full-time	United States of America	New Jersey	NaN	Secondary school (e.g. American high school, G...	11 - 17 years	Sch
83437	83438	I am a developer by profession	Employed full-time	Canada	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Online Course Certification;Bo / Phys
83438	83439	I am a developer by profession	Employed full-time	Brazil	NaN	NaN	Professional degree (JD, MD, etc.)	11 - 17 years	Sch

83439 rows × 48 columns

In [13]:

```
df.shape
```

Out[13]:

```
(0, 2)
```

In [14]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 0 entries
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   var     0 non-null      object
 1   val     0 non-null      int64 
dtypes: int64(1), object(1)
memory usage: 0.0+ bytes
```

In [15]:

```
schema_df = pd.read_csv("/Users/narenderbeniwal/Downloads/stack-overflow-developer-survey-2021/survey_results_schema.csv")
```

In [16]:

```
schema_df
```

Out[16]:

	qid	qname	question	force_resp	type	selector
0	QID16	S0	Hel...	False	DB	TB
1	QID12	MetaInfo	Browser Meta Info	False	Meta	Browser
2	QID1	S1	<span style="font-size:22px; font-family: aria...	False	DB	TB
3	QID2	MainBranch	Which of the following options best describes ...	True	MC	SAVR
4	QID24	Employment	Which of the following best describes your cur...	False	MC	MAVR
5	QID6	Country	Where do you live? <span style="font-weight: b...	True	MC	DL
6	QID7	US_State	<p>In which state or territory of the USA do y...	False	MC	DL
7	QID9	UK_Country	In which part of the United Kingdom do you liv...	False	MC	DL
8	QID190	S2	<span style="font-size:22px; font-family: aria...	False	DB	TB
9	QID25	EdLevel	Which of the following best describes the high...	False	MC	SAVR
10	QID149	Age1stCode	At what age did you write your first line of c...	False	MC	MAVR
11	QID276	LearnCode	How did you learn to code? Select all that apply.	False	MC	MAVR
12	QID32	YearsCode	Including any education, how many years have y...	False	MC	DL
13	QID34	YearsCodePro	NOT including education, how many years have y...	False	MC	DL
14	QID31	DevType	Which of the following describes your current ...	False	MC	MAVR
15	QID29	OrgSize	Approximately how many people are employed by ...	False	MC	MAVR
16	QID50	Currency	Which currency do you use day-to-day? If your ...	True	MC	SB
17	QID51	CompTotal	What is your current total compensation (salar...	False	TE	SL
18	QID52	CompFreq	Is that compensation weekly, monthly, or yearly?	False	MC	MAVR
19	QID61	S3	<span style="font-size:22px; font-family: aria...	False	DB	TB
20	QID233	Language	Which programming, scripting, and markup la...	False	Matrix	Likert
21	QID262	Database	Which database environments have you do...	False	Matrix	Likert
22	QID263	Platform	Which cloud platforms have you done ext...	False	Matrix	Likert
23	QID264	Webframe	Which web frameworks <span st...	False	Matrix	Likert
24	QID265	MiscTech	Which other frameworks and libraries ha...	False	Matrix	Likert
25	QID275	ToolsTech	Which tools have you done ext...	False	Matrix	Likert
26	QID274	NEWCollabTools	Which development environments</strong...	False	Matrix	Likert
27	QID71	OpSys	What is the primary operating system in which ...	False	MC	SAVR
28	QID243	NEWStuck	What do you do when you get stuck on a problem...	False	MC	MAVR
29	QID91	S4	<span style="font-size:22px; font-family: aria...	False	DB	TB
30	QID266	NEWSOSites	Which of the following Stack Overflow sites ha...	False	MC	MAVR
31	QID100	SOVisitFreq	How frequently would you say you visit Stack O...	False	MC	MAVR
32	QID101	SOAccount	Do you have a Stack Overflow account?	False	MC	MAVR
33	QID102	SOPartFreq	How frequently would you say you participate i...	False	MC	MAVR
34	QID106	SOComm	Do you consider yourself a member of the Stack...	False	MC	MAVR
35	QID267	NEWOtherComms	Are you a member of any other online developer...	False	MC	MAVR
36	QID268	NEWOtherCommsNames	Please name up to 5 other online developer com...	False	Matrix	Likert

	qid	qname	question	force_resp	type	selector
36	QID268	NEWOtherCommsNames	Please name up to 5 other online developer com...	False	Matrix	Likert
37	QID121	S5	<span style="font-size:22px; font-family: aria...	False	DB	TB
38	QID127	Age	What is your age?	False	MC	MAVR
39	QID122	Gender	Which of the following describe you, if any? P...	False	MC	MAVR
40	QID153	Trans	Do you identify as transgender?	False	MC	MAVR
41	QID136	Sexuality	Which of the following describe you, if any? P...	False	MC	MAVR
42	QID126	Ethnicity	Which of the following describe you, if any? P...	False	MC	MAVR
43	QID124	Accessibility	Which of the following describe you, if any? P...	False	MC	MAVR
44	QID125	MentalHealth	Which of the following describe you, if any? P...	False	MC	MAVR
45	QID131	S6	Final Qu...	False	DB	TB
46	QID132	SurveyLength	How do you feel about the length of the survey...	False	MC	MAVR
47	QID133	SurveyEase	How easy or difficult was this survey to compl...	False	MC	MAVR

In [17]:

```
schema_df.tail(5)
```

Out[17]:

	qid	qname	question	force_resp	type	selector
43	QID124	Accessibility	Which of the following describe you, if any? P...	False	MC	MAVR
44	QID125	MentalHealth	Which of the following describe you, if any? P...	False	MC	MAVR
45	QID131	S6	Final Qu...	False	DB	TB
46	QID132	SurveyLength	How do you feel about the length of the survey...	False	MC	MAVR
47	QID133	SurveyEase	How easy or difficult was this survey to compl...	False	MC	MAVR

DataFrame and Series Basics - Selecting Rows and Columns

In [18]:

```
person = {
    "first": "Corey",
    "last": "Schafer",
    "email": "CoreyMSchafer@gmail.com"
}
```

In [19]:

```
people = {
    "first": ["Corey"],
    "last": ["Schafer"],
    "email": ["CoreyMSchafer@gmail.com"]
}
```

In [20]:

```
people = {
    "first": ["Corey", 'Jane', 'John'],
    "last": ["Schafer", 'Doe', 'Doe'],
    "email": ["CoreyMSchafer@gmail.com", 'JaneDoe@email.com', 'JohnDoe@email.com']
}
```

In [21]:

```
people['email']
```

Out[21]:

```
['CoreyMSchafer@gmail.com', 'JaneDoe@email.com', 'JohnDoe@email.com']
```

In [22]:

```
people['first']
```

Out[22]:

```
['Corey', 'Jane', 'John']
```

In [23]:

```
df = pd.DataFrame(people)
```

In [24]:

```
df
```

Out[24]:

	first	last	email
0	Corey	Schafer	CoreyMSchafer@gmail.com
1	Jane	Doe	JaneDoe@email.com
2	John	Doe	JohnDoe@email.com

In [25]:

```
df['email']
```

Out[25]:

```
0    CoreyMSchafer@gmail.com
1          JaneDoe@email.com
2          JohnDoe@email.com
Name: email, dtype: object
```

In [26]:

```
type(df['email'])
```

Out[26]:

```
pandas.core.series.Series
```

In [27]:

```
df.email
```

Out[27]:

```
0    CoreyMSchafer@gmail.com
1          JaneDoe@email.com
2          JohnDoe@email.com
Name: email, dtype: object
```

In [28]:

```
df[['last','email']]
```

Out[28]:

	last	email
0	Schafer	CoreyMSchafer@gmail.com
1	Doe	JaneDoe@email.com
2	Doe	JohnDoe@email.com

In [29]:

```
df.columns
```

Out[29]:

```
Index(['first', 'last', 'email'], dtype='object')
```

In [30]:

```
df.iloc[[0,1],[2,1]]
```

Out[30]:

	email	last
0	CoreyMSchafer@gmail.com	Schafer
1	JaneDoe@email.com	Doe

In [31]:

```
df.loc[[0,1], ['email','last']]
```

Out[31]:

	email	last
0	CoreyMSchafer@gmail.com	Schafer
1	JaneDoe@email.com	Doe

Now working the stack overflow dataset

In [32]:

```
df1.shape
```

Out[32]:

```
(83439, 48)
```

In [33]:

```
df1.columns
```

Out[33]:

```
Index(['ResponseId', 'MainBranch', 'Employment', 'Country', 'US_State',
       'UK_Country', 'EdLevel', 'Age1stCode', 'LearnCode', 'YearsCode',
       'YearsCodePro', 'DevType', 'OrgSize', 'Currency', 'CompTotal',
       'CompFreq', 'LanguageHaveWorkedWith', 'LanguageWantToWorkWith',
       'DatabaseHaveWorkedWith', 'DatabaseWantToWorkWith',
       'PlatformHaveWorkedWith', 'PlatformWantToWorkWith',
       'WebframeHaveWorkedWith', 'WebframeWantToWorkWith',
       'MiscTechHaveWorkedWith', 'MiscTechWantToWorkWith',
       'ToolsTechHaveWorkedWith', 'ToolsTechWantToWorkWith',
       'NEWCollabToolsHaveWorkedWith', 'NEWCollabToolsWantToWorkWith', 'OpSys',
       'NEWStuck', 'NEWSOSites', 'SOVisitFreq', 'SOAccount', 'SOPartFreq',
       'SOComm', 'NEWOtherComms', 'Age', 'Gender', 'Trans', 'Sexuality',
       'Ethnicity', 'Accessibility', 'MentalHealth', 'SurveyLength',
       'SurveyEase', 'ConvertedCompYearly'],
      dtype='object')
```

In [34]:

```
df1['NEWSOSites'].value_counts()
```

Out[34]:

```
Stack Overflow;Stack Exchange
54658
Stack Overflow
23473
```

```
Stack Overflow;Stack Exchange;Stack Overflow for Teams (private knowledge sharing & colla
boration platform for companies)     3099
I have never visited Stack Overflow or the Stack Exchange network
750
Stack Overflow;Stack Overflow for Teams (private knowledge sharing & collaboration platfo
rm for companies)                    671
Stack Exchange
257
Stack Overflow for Teams (private knowledge sharing & collaboration platform for companie
s)                                   246
Stack Exchange;Stack Overflow for Teams (private knowledge sharing & collaboration platfo
rm for companies)                     17
Name: NEWSOSites, dtype: int64
```

In [35]:

```
df1['NEWSOSites']
```

Out[35]:

```
0                                        Stack Overflow
1                                        Stack Overflow
2                           Stack Overflow;Stack Exchange
3                                        Stack Overflow
4                           Stack Overflow;Stack Exchange
                              ...
83434                       Stack Overflow;Stack Exchange
83435                       Stack Overflow;Stack Exchange
83436                       Stack Overflow;Stack Exchange
83437                                    Stack Overflow
83438    Stack Overflow;Stack Exchange;Stack Overflow f...
Name: NEWSOSites, Length: 83439, dtype: object
```

In [36]:

```
#df1.loc[[0,1,2,67],'NEWSOSites']
```

In [37]:

```
df1.loc[0:9,'NEWSOSites':'Gender']
```

Out[37]:

	NEWSOSites	SOVisitFreq	SOAccount	SOPartFreq	SOComm	NEWOtherComms	Age	Gender
0	**Stack Overflow**	**Multiple times per day**	**Yes**	**A few times per month or weekly**	**Yes, definitely**	**No**	**25-34 years old**	**Man**
1	**Stack Overflow**	**Daily or almost daily**	**Yes**	**Daily or almost daily**	**Yes, definitely**	**No**	**18-24 years old**	**Man**
2	**Stack Overflow;Stack Exchange**	**Multiple times per day**	**Yes**	**Multiple times per day**	**Yes, definitely**	**Yes**	**18-24 years old**	**Man**
3	**Stack Overflow**	**Daily or almost daily**	**Yes**	**Daily or almost daily**	**Neutral**	**No**	**35-44 years old**	**Man**
4	**Stack Overflow;Stack Exchange**	**Daily or almost daily**	**Yes**	**A few times per week**	**Yes, somewhat**	**No**	**25-34 years old**	**Man**
5	**Stack Overflow;Stack Exchange**	**Multiple times per day**	**Yes**	**I have never participated in Q&A on Stack Over...**	**Yes, somewhat**	**No**	**18-24 years old**	**Prefer not to say**
6	**Stack Overflow**	**Daily or almost daily**	**Yes**	**A few times per week**	**Yes, somewhat**	**No**	**Prefer not to say**	**Prefer not to say**
7	**Stack Overflow;Stack Exchange**	**Multiple times per day**	**Yes**	**I have never participated in Q&A on Stack Over...**	**No, not at all**	**Yes**	**18-24 years old**	**Woman**

	NEWSOSites	SOVisitFreq	SOAccount	SOPartFreq	SOComm	NEWOtherComms	Age	Gender
8	Stack Overflow;Stack Exchange	A few times per week	Yes	Less than once per month or monthly	Yes, definitely	No	25-34 years old	Man
9	Stack Overflow;Stack Exchange	Multiple times per day	Yes	Daily or almost daily	Yes, somewhat	No	25-34 years old	Man

Indexes - How to Set, Reset, and Use Indexes

In [38]:

```
df
```

Out[38]:

	first	last	email
0	Corey	Schafer	CoreyMSchafer@gmail.com
1	Jane	Doe	JaneDoe@email.com
2	John	Doe	JohnDoe@email.com

In [39]:

```
df.set_index('email',inplace=True)
```

In [40]:

```
df
```

Out[40]:

email	first	last
CoreyMSchafer@gmail.com	Corey	Schafer
JaneDoe@email.com	Jane	Doe
JohnDoe@email.com	John	Doe

In [41]:

```
df.index
```

Out[41]:

```
Index(['CoreyMSchafer@gmail.com', 'JaneDoe@email.com', 'JohnDoe@email.com'], dtype='objec
t', name='email')
```

In [42]:

```
df.loc['JaneDoe@email.com']
```

Out[42]:

```
first    Jane
last      Doe
Name: JaneDoe@email.com, dtype: object
```

In [43]:

```
df.loc['JohnDoe@email.com']
```

Out[43]:

```
first    John
last      Doe
Name: JohnDoe@email.com, dtype: object
```

In [44]:

```
df.reset_index(inplace=True)
```

In [45]:

```
df
```

Out[45]:

	email	first	last
0	CoreyMSchafer@gmail.com	Corey	Schafer
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	John	Doe

Now working the stack overflow dataset

In [46]:

```
df1.head()
```

Out[46]:

	ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Yea
0	1	I am a developer by profession	Independent contractor, freelancer, or self-em...	Slovakia	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Coding Bootcamp;Other online resources (ex: vi...	
1	2	I am a student who is learning to code	Student, full-time	Netherlands	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
2	3	I am not primarily a developer, but I write co...	Student, full-time	Russian Federation	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
3	4	I am a developer by profession	Employed full-time	Austria	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	11 - 17 years	NaN	
4	5	I am a developer by profession	Independent contractor, freelancer, or self-em...	United Kingdom of Great Britain and Northern I...	NaN	England	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	5 - 10 years	Friend or family member	

In [47]:

```
df1 = pd.read_csv("/Users/narenderbeniwal/Downloads/stack-overflow-developer-survey-2021/survey_results_public.csv",index_col = 'ResponseId')
```

In [48]:

```
df1
```

Out[48]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Ye
1	I am a developer by profession	Independent contractor, freelancer, or self-em...	Slovakia	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Coding Bootcamp;Other online resources (ex: vi...	
2	I am a student who is learning to code	Student, full-time	Netherlands	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
3	I am not primarily a developer, but I write co...	Student, full-time	Russian Federation	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
4	I am a developer by profession	Employed full-time	Austria	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	11 - 17 years	NaN	
5	I am a developer by profession	Independent contractor, freelancer, or self-em...	United Kingdom of Great Britain and Northern I...	NaN	England	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	5 - 10 years	Friend or family member	
...	...	...	...	...	...	...	...	...	
83435	I am a developer by profession	Employed full-time	United States of America	Texas	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
83436	I am a developer by profession	Independent contractor, freelancer, or self-em...	Benin	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
83437	I am a developer by profession	Employed full-time	United States of America	New Jersey	NaN	Secondary school (e.g. American high school, G...	11 - 17 years	School	
83438	I am a developer by profession	Employed full-time	Canada	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Online Courses or Certification;Books / Physic...	
83439	I am a developer by profession	Employed full-time	Brazil	NaN	NaN	Professional degree (JD, MD, etc.)	11 - 17 years	School	

83439 rows × 47 columns

In [49]:

```
df1.loc[[1,4]]
```

Out[49]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	YearsCode
1	I am a developer by profession	Independent contractor, freelancer, or self-em...	Slovakia	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Coding Bootcamp;Other online resources (ex: vi...	NaN
4	I am a developer by profession	Employed full-time	Austria	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	11 - 17 years	NaN	NaN

In [50]:

```
schema_df
```

Out[50]:

	qid	qname	question	force_resp	type	selector
0	QID16	S0	Hel...	False	DB	TB
1	QID12	MetaInfo	Browser Meta Info	False	Meta	Browser
2	QID1	S1	<span style="font-size:22px; font-family: aria...	False	DB	TB
3	QID2	MainBranch	Which of the following options best describes ...	True	MC	SAVR
4	QID24	Employment	Which of the following best describes your cur...	False	MC	MAVR
5	QID6	Country	Where do you live? <span style="font-weight: b...	True	MC	DL
6	QID7	US_State	<p>In which state or territory of the USA do y...	False	MC	DL
7	QID9	UK_Country	In which part of the United Kingdom do you liv...	False	MC	DL
8	QID190	S2	<span style="font-size:22px; font-family: aria...	False	DB	TB
9	QID25	EdLevel	Which of the following best describes the high...	False	MC	SAVR
10	QID149	Age1stCode	At what age did you write your first line of c...	False	MC	MAVR
11	QID276	LearnCode	How did you learn to code? Select all that apply.	False	MC	MAVR
12	QID32	YearsCode	Including any education, how many years have y...	False	MC	DL
13	QID34	YearsCodePro	NOT including education, how many years have y...	False	MC	DL
14	QID31	DevType	Which of the following describes your current ...	False	MC	MAVR
15	QID29	OrgSize	Approximately how many people are employed by ...	False	MC	MAVR
16	QID50	Currency	Which currency do you use day-to-day? If your ...	True	MC	SB
17	QID51	CompTotal	What is your current total compensation (salar...	False	TE	SL
18	QID52	CompFreq	Is that compensation weekly, monthly, or yearly?	False	MC	MAVR
19	QID61	S3	<span style="font-size:22px; font-family: aria...	False	DB	TB
20	QID233	Language	Which programming, scripting, and markup la...	False	Matrix	Likert
21	QID262	Database	Which database environments have you do...	False	Matrix	Likert
22	QID263	Platform	Which cloud platforms have you done ext...	False	Matrix	Likert
23	QID264	Webframe	Which web frameworks <span st...	False	Matrix	Likert
24	QID265	MiscTech	Which other frameworks and libraries ha...	False	Matrix	Likert
25	QID275	ToolsTech	Which tools have you done ext...	False	Matrix	Likert
26	QID274	NEWCollabTools	Which development environments</strong...	False	Matrix	Likert
27	QID71	OpSys	What is the primary operating system in which ...	False	MC	SAVR

	qid	qname	question	force_resp	type	selector
28	QID243	NEWStuck	What do you do when you get stuck on a problem...	False	MC	MAVR
29	QID91	S4	<span style="font-size:22px; font-family: aria...	False	DB	TB
30	QID266	NEWSOSites	Which of the following Stack Overflow sites ha...	False	MC	MAVR
31	QID100	SOVisitFreq	How frequently would you say you visit Stack O...	False	MC	MAVR
32	QID101	SOAccount	Do you have a Stack Overflow account?	False	MC	MAVR
33	QID102	SOPartFreq	How frequently would you say you participate i...	False	MC	MAVR
34	QID106	SOComm	Do you consider yourself a member of the Stack...	False	MC	MAVR
35	QID267	NEWOtherComms	Are you a member of any other online developer...	False	MC	MAVR
36	QID268	NEWOtherCommsNames	Please name up to 5 other online developer com...	False	Matrix	Likert
37	QID121	S5	<span style="font-size:22px; font-family: aria...	False	DB	TB
38	QID127	Age	What is your age?	False	MC	MAVR
39	QID122	Gender	Which of the following describe you, if any? P...	False	MC	MAVR
40	QID153	Trans	Do you identify as transgender?	False	MC	MAVR
41	QID136	Sexuality	Which of the following describe you, if any? P...	False	MC	MAVR
42	QID126	Ethnicity	Which of the following describe you, if any? P...	False	MC	MAVR
43	QID124	Accessibility	Which of the following describe you, if any? P...	False	MC	MAVR
44	QID125	MentalHealth	Which of the following describe you, if any? P...	False	MC	MAVR
45	QID131	S6	Final Qu...	False	DB	TB
46	QID132	SurveyLength	How do you feel about the length of the survey...	False	MC	MAVR
47	QID133	SurveyEase	How easy or difficult was this survey to compl...	False	MC	MAVR

In [51]:

```
schema_df['question']
```

Out[51]:

```
0     <div><span style="font-size:19px;"><strong>Hel...
1                                  Browser Meta Info
2     <span style="font-size:22px; font-family: aria...
3     Which of the following options best describes ...
4     Which of the following best describes your cur...
5     Where do you live? <span style="font-weight: b...
6     <p>In which state or territory of the USA do y...
7     In which part of the United Kingdom do you liv...
8     <span style="font-size:22px; font-family: aria...
9     Which of the following best describes the high...
10    At what age did you write your first line of c...
11    How did you learn to code? Select all that apply.
12    Including any education, how many years have y...
13    NOT including education, how many years have y...
14    Which of the following describes your current ...
15    Approximately how many people are employed by ...
16    Which currency do you use day-to-day? If your ...
17    What is your current total compensation (salar...
18     Is that compensation weekly, monthly, or yearly?
19    <span style="font-size:22px; font-family: aria...
20    Which <b>programming, scripting, and markup la...
21    Which <b>database environments </b>have you do...
22    Which <b>cloud platforms</b> have you done ext...
23    Which <strong>web frameworks </strong><span st...
24    Which <b>other frameworks and libraries</b> ha...
25    Which <strong>tools</strong> have you done ext...
26    Which <strong>development environments</strong...
27    What is the primary operating system in which ...
28    What do you do when you get stuck on a problem...
29    <span style="font-size:22px; font-family: aria...
30    Which of the following Stack Overflow sites ha...
31    How frequently would you say you visit Stack O...
32                Do you have a Stack Overflow account?
33    How frequently would you say you participate i...
```

```
34    Do you consider yourself a member of the Stack...
35    Are you a member of any other online developer...
36    Please name up to 5 other online developer com...
37    <span style="font-size:22px; font-family: aria...
38                                  What is your age?
39    Which of the following describe you, if any? P...
40                     Do you identify as transgender?
41    Which of the following describe you, if any? P...
42    Which of the following describe you, if any? P...
43    Which of the following describe you, if any? P...
44    Which of the following describe you, if any? P...
45    <span style="font-size:22px;"><strong>Final Qu...
46    How do you feel about the length of the survey...
47    How easy or difficult was this survey to compl...
Name: question, dtype: object
```

In [52]:

```
schema_df = pd.read_csv("/Users/narenderbeniwal/Downloads/stack-overflow-developer-survey-2021/survey_results_schema.csv", index_col = "qname")
```

In [53]:

```
schema_df
```

Out[53]:

qname	qid	question	force_resp	type	selector
S0	**QID16**	** Hel...**	**False**	**DB**	**TB**
MetaInfo	**QID12**	**Browser Meta Info**	**False**	**Meta**	**Browser**
S1	**QID1**	**<span style="font-size:22px; font-family: aria...**	**False**	**DB**	**TB**
MainBranch	**QID2**	**Which of the following options best describes ...**	**True**	**MC**	**SAVR**
Employment	**QID24**	**Which of the following best describes your cur...**	**False**	**MC**	**MAVR**
Country	**QID6**	**Where do you live? <span style="font-weight: b...**	**True**	**MC**	**DL**
US_State	**QID7**	**<p>In which state or territory of the USA do y...**	**False**	**MC**	**DL**
UK_Country	**QID9**	**In which part of the United Kingdom do you liv...**	**False**	**MC**	**DL**
S2	**QID190**	**<span style="font-size:22px; font-family: aria...**	**False**	**DB**	**TB**
EdLevel	**QID25**	**Which of the following best describes the high...**	**False**	**MC**	**SAVR**
Age1stCode	**QID149**	**At what age did you write your first line of c...**	**False**	**MC**	**MAVR**
LearnCode	**QID276**	**How did you learn to code? Select all that apply.**	**False**	**MC**	**MAVR**
YearsCode	**QID32**	**Including any education, how many years have y...**	**False**	**MC**	**DL**
YearsCodePro	**QID34**	**NOT including education, how many years have y...**	**False**	**MC**	**DL**
DevType	**QID31**	**Which of the following describes your current ...**	**False**	**MC**	**MAVR**
OrgSize	**QID29**	**Approximately how many people are employed by ...**	**False**	**MC**	**MAVR**
Currency	**QID50**	**Which currency do you use day-to-day? If your ...**	**True**	**MC**	**SB**
CompTotal	**QID51**	**What is your current total compensation (salar...**	**False**	**TE**	**SL**
CompFreq	**QID52**	**Is that compensation weekly, monthly, or yearly?**	**False**	**MC**	**MAVR**
S3	**QID61**	**<span style="font-size:22px; font-family: aria...**	**False**	**DB**	**TB**
Language	**QID233**	**Which programming, scripting, and markup la...**	**False**	**Matrix**	**Likert**
Database	**QID262**	**Which database environments have you do...**	**False**	**Matrix**	**Likert**
Platform	**QID263**	**Which cloud platforms have you done ext...**	**False**	**Matrix**	**Likert**
Webframe	**QID264**	**Which web frameworks <span st...**	**False**	**Matrix**	**Likert**
MiscTech	**QID265**	**Which other frameworks and libraries ha...**	**False**	**Matrix**	**Likert**
ToolsTech	**QID275**	**Which tools have you done ext...**	**False**	**Matrix**	**Likert**

qname	qid	question	force_resp	type	selector
NEWCollabTools	QID274	Which development environments...	False	Matrix	Likert
OpSys	QID71	What is the primary operating system in which ...	False	MC	SAVR
NEWStuck	QID243	What do you do when you get stuck on a problem...	False	MC	MAVR
S4	QID91	<span style="font-size:22px; font-family: aria...	False	DB	TB
NEWSOSites	QID266	Which of the following Stack Overflow sites ha...	False	MC	MAVR
SOVisitFreq	QID100	How frequently would you say you visit Stack O...	False	MC	MAVR
SOAccount	QID101	Do you have a Stack Overflow account?	False	MC	MAVR
SOPartFreq	QID102	How frequently would you say you participate i...	False	MC	MAVR
SOComm	QID106	Do you consider yourself a member of the Stack...	False	MC	MAVR
NEWOtherComms	QID267	Are you a member of any other online developer...	False	MC	MAVR
NEWOtherCommsNames	QID268	Please name up to 5 other online developer com...	False	Matrix	Likert
S5	QID121	<span style="font-size:22px; font-family: aria...	False	DB	TB
Age	QID127	What is your age?	False	MC	MAVR
Gender	QID122	Which of the following describe you, if any? P...	False	MC	MAVR
Trans	QID153	Do you identify as transgender?	False	MC	MAVR
Sexuality	QID136	Which of the following describe you, if any? P...	False	MC	MAVR
Ethnicity	QID126	Which of the following describe you, if any? P...	False	MC	MAVR
Accessibility	QID124	Which of the following describe you, if any? P...	False	MC	MAVR
MentalHealth	QID125	Which of the following describe you, if any? P...	False	MC	MAVR
S6	QID131	Final Qu...	False	DB	TB
SurveyLength	QID132	How do you feel about the length of the survey...	False	MC	MAVR
SurveyEase	QID133	How easy or difficult was this survey to compl...	False	MC	MAVR

In [54]:

```
schema_df.loc['S3','question']
```

Out[54]:

```
'<span style="font-size:22px; font-family: arial,helvetica,sans-serif; font-weight: 700;"
>Tech and tech culture</span><br>\n<br>\n\n<p><span style="font-size:16px; font-family:ar
ial,helvetica,sans-serif;">The next set of questions will focus on technology and tech cu
lture.<br>\n<br>\nMost questions in this section are optional. Required questions are not
ed with *.</span></p>'
```

In [55]:

```
schema_df.sort_index(ascending=True)
```

Out[55]:

qname	qid	question	force_resp	type	selector
Accessibility	QID124	Which of the following describe you, if any? P...	False	MC	MAVR
Age	QID127	What is your age?	False	MC	MAVR
Age1stCode	QID149	At what age did you write your first line of c...	False	MC	MAVR
CompFreq	QID52	Is that compensation weekly, monthly, or yearly?	False	MC	MAVR
CompTotal	QID51	What is your current total compensation (salar...	False	TE	SL
Country	QID6	Where do you live? <span style="font-weight: b...	True	MC	DL
Currency	QID50	Which currency do you use day-to-day? If your ...	True	MC	SB
Database	QID262	Which database environments have you do...	False	Matrix	Likert
DevType	QID31	Which of the following describes your current ...	False	MC	MAVR

qname	qid	question	force_resp	type	selector
EdLevel	QID20	Which of the following best describes the high...	False	MC	MAVR
Employment	QID24	Which of the following best describes your cur...	False	MC	MAVR
Ethnicity	QID126	Which of the following describe you, if any? P...	False	MC	MAVR
Gender	QID122	Which of the following describe you, if any? P...	False	MC	MAVR
Language	QID233	Which programming, scripting, and markup la...	False	Matrix	Likert
LearnCode	QID276	How did you learn to code? Select all that apply.	False	MC	MAVR
MainBranch	QID2	Which of the following options best describes ...	True	MC	SAVR
MentalHealth	QID125	Which of the following describe you, if any? P...	False	MC	MAVR
MetaInfo	QID12	Browser Meta Info	False	Meta	Browser
MiscTech	QID265	Which other frameworks and libraries ha...	False	Matrix	Likert
NEWCollabTools	QID274	Which development environments</strong...	False	Matrix	Likert
NEWOtherComms	QID267	Are you a member of any other online developer...	False	MC	MAVR
NEWOtherCommsNames	QID268	Please name up to 5 other online developer com...	False	Matrix	Likert
NEWSOSites	QID266	Which of the following Stack Overflow sites ha...	False	MC	MAVR
NEWStuck	QID243	What do you do when you get stuck on a problem...	False	MC	MAVR
OpSys	QID71	What is the primary operating system in which ...	False	MC	SAVR
OrgSize	QID29	Approximately how many people are employed by ...	False	MC	MAVR
Platform	QID263	Which cloud platforms have you done ext...	False	Matrix	Likert
S0	QID16	Hel...	False	DB	TB
S1	QID1	<span style="font-size:22px; font-family: aria...	False	DB	TB
S2	QID190	<span style="font-size:22px; font-family: aria...	False	DB	TB
S3	QID61	<span style="font-size:22px; font-family: aria...	False	DB	TB
S4	QID91	<span style="font-size:22px; font-family: aria...	False	DB	TB
S5	QID121	<span style="font-size:22px; font-family: aria...	False	DB	TB
S6	QID131	Final Qu...	False	DB	TB
SOAccount	QID101	Do you have a Stack Overflow account?	False	MC	MAVR
SOComm	QID106	Do you consider yourself a member of the Stack...	False	MC	MAVR
SOPartFreq	QID102	How frequently would you say you participate i...	False	MC	MAVR
SOVisitFreq	QID100	How frequently would you say you visit Stack O...	False	MC	MAVR
Sexuality	QID136	Which of the following describe you, if any? P...	False	MC	MAVR
SurveyEase	QID133	How easy or difficult was this survey to compl...	False	MC	MAVR
SurveyLength	QID132	How do you feel about the length of the survey...	False	MC	MAVR
ToolsTech	QID275	Which tools have you done ext...	False	Matrix	Likert
Trans	QID153	Do you identify as transgender?	False	MC	MAVR
UK_Country	QID9	In which part of the United Kingdom do you liv...	False	MC	DL
US_State	QID7	<p>In which state or territory of the USA do y...	False	MC	DL
Webframe	QID264	Which web frameworks <span st...	False	Matrix	Likert
YearsCode	QID32	Including any education, how many years have y...	False	MC	DL
YearsCodePro	QID34	NOT including education, how many years have y...	False	MC	DL

Filtering - Using Conditionals to Filter Rows and Columns

In [56]:

```
df
```

Out[56]:

	email	first	last
0	CoreyMSchafer@gmail.com	Corey	Schafer
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	John	Doe

In [57]:

```
df['last'] == 'Doe'
```

Out[57]:

```
0    False
1     True
2     True
Name: last, dtype: bool
```

In [58]:

```
filt = (df['last'] == 'Doe')
```

In [59]:

```
filt
```

Out[59]:

```
0    False
1     True
2     True
Name: last, dtype: bool
```

In [60]:

```
df.loc[filt, 'email']
```

Out[60]:

```
1    JaneDoe@email.com
2    JohnDoe@email.com
Name: email, dtype: object
```

In [61]:

```
filt = (df['last']=='Doe')& (df['first']=='John')
```

In [62]:

```
df.loc[filt]
```

Out[62]:

	email	first	last
2	JohnDoe@email.com	John	Doe

In [63]:

```
filt = (df['last']=='Schafer')| (df['first']=='John')
```

In [64]:

```
df.loc[~filt, 'email']
```

Out[64]:

```
1    JaneDoe@email.com
Name: email, dtype: object
```

Now working the stack overflow dataset

In [65]:

```
df1
```

Out[65]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Ye
1	I am a developer by profession	Independent contractor, freelancer, or self-em...	Slovakia	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Coding Bootcamp;Other online resources (ex: vi...	
2	I am a student who is learning to code	Student, full-time	Netherlands	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
3	I am not primarily a developer, but I write co...	Student, full-time	Russian Federation	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
4	I am a developer by profession	Employed full-time	Austria	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	11 - 17 years	NaN	
5	I am a developer by profession	Independent contractor, freelancer, or self-em...	United Kingdom of Great Britain and Northern I...	NaN	England	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	5 - 10 years	Friend or family member	
...	...	...	...	...	...	...	...	...	
83435	I am a developer by profession	Employed full-time	United States of America	Texas	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
83436	I am a developer by profession	Independent contractor, freelancer, or self-em...	Benin	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
83437	I am a developer by profession	Employed full-time	United States of America	New Jersey	NaN	Secondary school (e.g. American high school, G...	11 - 17 years	School	
83438	I am a developer by profession	Employed full-time	Canada	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Online Courses or Certification;Books / Physic...	
83439	I am a developer by profession	Employed full-time	Brazil	NaN	NaN	Professional degree (JD, MD, etc.)	11 - 17 years	School	

83439 rows × 47 columns

In [66]:

```
df1.columns
```

Out[66]:

```
Index(['MainBranch', 'Employment', 'Country', 'US_State', 'UK_Country',
       'EdLevel', 'Age1stCode', 'LearnCode', 'YearsCode', 'YearsCodePro',
       'DevType', 'OrgSize', 'Currency', 'CompTotal', 'CompFreq',
       'LanguageHaveWorkedWith', 'LanguageWantToWorkWith',
       'DatabaseHaveWorkedWith', 'DatabaseWantToWorkWith',
       'PlatformHaveWorkedWith', 'PlatformWantToWorkWith',
       'WebframeHaveWorkedWith', 'WebframeWantToWorkWith',
       'MiscTechHaveWorkedWith', 'MiscTechWantToWorkWith',
       'ToolsTechHaveWorkedWith', 'ToolsTechWantToWorkWith',
       'NEWCollabToolsHaveWorkedWith', 'NEWCollabToolsWantToWorkWith', 'OpSys',
       'NEWStuck', 'NEWSOSites', 'SOVisitFreq', 'SOAccount', 'SOPartFreq',
       'SOComm', 'NEWOtherComms', 'Age', 'Gender', 'Trans', 'Sexuality',
       'Ethnicity', 'Accessibility', 'MentalHealth', 'SurveyLength',
       'SurveyEase', 'ConvertedCompYearly'],
      dtype='object')
```

In [ ]:

```

```

In [67]:

```
high_salary = (df1['ConvertedCompYearly']>70000)
```

In [68]:

```
high_salary
```

Out[68]:

```
ResponseId
1         False
2         False
3         False
4         False
5         False
          ...
83435      True
83436     False
83437      True
83438      True
83439     False
Name: ConvertedCompYearly, Length: 83439, dtype: bool
```

In [69]:

```
df1.loc[high_salary, ['Employment','Country','ConvertedCompYearly']]
```

Out[69]:

ResponseId	Employment	Country	ConvertedCompYearly
13	Employed full-time	Germany	77290.0
19	I prefer not to say	Singapore	160932.0
25	Employed full-time	Germany	77831.0
27	Employed full-time	Switzerland	81319.0
32	Employed full-time	Israel	122580.0
...	...	...	...
83431	Employed full-time	United States of America	125000.0
83433	Employed full-time	Canada	80169.0

ResponseId	Employment	Country	ConvertedCompYearly
83435	Employed full-time	United States of America	160500.0
83437	Employed full-time	United States of America	90000.0
83438	Employed full-time	Canada	816816.0

18793 rows × 3 columns

In [70]:

```
countries = ['Singapore','United States of America','Germany','Canada', 'India']
filt = df1['Country'].isin(countries)
```

In [71]:

```
df1.loc[filt,['Employment','Country','ConvertedCompYearly']]
```

Out[71]:

ResponseId	Employment	Country	ConvertedCompYearly
6	Student, part-time	United States of America	NaN
7	I prefer not to say	United States of America	NaN
9	Employed part-time	India	NaN
13	Employed full-time	Germany	77290.0
16	Student, full-time	United States of America	NaN
...	...	...	...
83431	Employed full-time	United States of America	125000.0
83433	Employed full-time	Canada	80169.0
83435	Employed full-time	United States of America	160500.0
83437	Employed full-time	United States of America	90000.0
83438	Employed full-time	Canada	816816.0

34775 rows × 3 columns

In [72]:

```
filt = df1['LanguageHaveWorkedWith'].str.contains('Python', na=False)
```

In [73]:

```
df1['LanguageHaveWorkedWith']
```

Out[73]:

```
ResponseId
1        C++;HTML/CSS;JavaScript;Objective-C;PHP;Swift
2                                  JavaScript;Python
3                           Assembly;C;Python;R;Rust
4                              JavaScript;TypeScript
5                    Bash/Shell;HTML/CSS;Python;SQL
                              ...
83435                             Clojure;Kotlin;SQL
83436                                            NaN
83437                             Groovy;Java;Python
83438          Bash/Shell;JavaScript;Node.js;Python
83439       Delphi;Elixir;HTML/CSS;Java;JavaScript
Name: LanguageHaveWorkedWith, Length: 83439, dtype: object
```

In [ ]:

In [74]:

```
df1
```

Out[74]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Ye
1	I am a developer by profession	Independent contractor, freelancer, or self-em...	Slovakia	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Coding Bootcamp;Other online resources (ex: vi...	
2	I am a student who is learning to code	Student, full-time	Netherlands	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
3	I am not primarily a developer, but I write co...	Student, full-time	Russian Federation	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
4	I am a developer by profession	Employed full-time	Austria	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	11 - 17 years	NaN	
5	I am a developer by profession	Independent contractor, freelancer, or self-em...	United Kingdom of Great Britain and Northern I...	NaN	England	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	5 - 10 years	Friend or family member	
...	...	...	...	...	...	...	...	...	
83435	I am a developer by profession	Employed full-time	United States of America	Texas	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
83436	I am a developer by profession	Independent contractor, freelancer, or self-em...	Benin	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
83437	I am a developer by profession	Employed full-time	United States of America	New Jersey	NaN	Secondary school (e.g. American high school, G...	11 - 17 years	School	
83438	I am a developer by profession	Employed full-time	Canada	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Online Courses or Certification;Books / Physic...	
83439	I am a developer by profession	Employed full-time	Brazil	NaN	NaN	Professional degree (JD, MD, etc.)	11 - 17 years	School	

83439 rows × 47 columns

Updating Rows and Columns - Modifying Data Within DataFrames

In [75]:

```
df
```

Out[75]:

	email	first	last
0	CoreyMSchafer@gmail.com	Corey	Schafer
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	John	Doe

In [76]:

```
df.columns
```

Out[76]:

```
Index(['email', 'first', 'last'], dtype='object')
```

In [77]:

```
df.columns = ['email','first_name','last_name']
```

In [78]:

```
df
```

Out[78]:

	email	first_name	last_name
0	CoreyMSchafer@gmail.com	Corey	Schafer
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	John	Doe

In [79]:

```
df.columns = [x.upper() for x in df.columns]
```

In [80]:

```
df
```

Out[80]:

	EMAIL	FIRST_NAME	LAST_NAME
0	CoreyMSchafer@gmail.com	Corey	Schafer
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	John	Doe

In [81]:

```
df.columns = df.columns.str.replace('_','')
```

In [82]:

```
df
```

Out[82]:

	EMAIL	FIRSTNAME	LASTNAME
0	CoreyMSchafer@gmail.com	Corey	Schafer

	EMAIL	FIRSTNAME	LASTNAME
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	John	Doe

In [83]:

```
df.columns = [x.lower() for x in df.columns]
df
```

Out[83]:

	email	firstname	lastname
0	CoreyMSchafer@gmail.com	Corey	Schafer
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	John	Doe

In [84]:

```
df.columns = df.columns.str.replace('_','')
```

In [85]:

```
df
```

Out[85]:

	email	firstname	lastname
0	CoreyMSchafer@gmail.com	Corey	Schafer
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	John	Doe

In [86]:

```
df.rename(columns={'firstname':'first', 'lastname':'last'}, inplace=True)
```

In [87]:

```
df
```

Out[87]:

	email	first	last
0	CoreyMSchafer@gmail.com	Corey	Schafer
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	John	Doe

In [88]:

```
df.loc[2] = ['John','Smith','JohnDoe@email.com']
```

In [ ]:

```

```

In [89]:

```
df
```

Out[89]:

	email	first	last
0	CoreyMSchafer@gmail.com	Corey	Schafer

1	JaneDoe@email.com	Jane	Doe
2	John	Smith	JohnDoe@email.com

In [90]:

```
df.loc[2,['last','email']] = ['Doe','JohnDoe@email.com']
```

In [91]:

```
df
```

Out[91]:

	email	first	last
0	CoreyMSchafer@gmail.com	Corey	Schafer
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	Smith	Doe

In [92]:

```
df.loc[2, 'last'] = 'Smith'
```

In [93]:

```
df
```

Out[93]:

	email	first	last
0	CoreyMSchafer@gmail.com	Corey	Schafer
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	Smith	Smith

In [94]:

```
df.at[2,'last'] = 'Doe'
```

In [95]:

```
df
```

Out[95]:

	email	first	last
0	CoreyMSchafer@gmail.com	Corey	Schafer
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	Smith	Doe

In [96]:

```
filt = (df['email']=='JohnDoe@email.com')
df[filt]['last'] = 'Smith'
```

```
/var/folders/12/9_hj00mj65v4tgx35t4t6c280000gn/T/ipykernel_1909/1777826079.py:2: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_g
uide/indexing.html#returning-a-view-versus-a-copy
  df[filt]['last'] = 'Smith'
```

In [97]:

```
df
```

Out[97]:

	email	first	last
0	CoreyMSchafer@gmail.com	Corey	Schafer
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	Smith	Doe

In [98]:

```
filt = (df['email']=='JohnDoe@email.com')
df.loc[filt,'last'] = 'Smith'
```

In [99]:

```
df
```

Out[99]:

	email	first	last
0	CoreyMSchafer@gmail.com	Corey	Schafer
1	JaneDoe@email.com	Jane	Doe
2	JohnDoe@email.com	Smith	Smith

In [100]:

```
df['email'] = df['email'].str.lower()
```

In [101]:

```
df
```

Out[101]:

	email	first	last
0	coreymschafer@gmail.com	Corey	Schafer
1	janedoe@email.com	Jane	Doe
2	johndoe@email.com	Smith	Smith

Four Importent Methods

1. Apply

2. map

3. Applymap

4. replace

In [102]:

```
df['email'].apply(len)
```

Out[102]:

```
0    23
1    17
```

```
2    17
Name: email, dtype: int64
```

In [103]:

```
def update_email(email):
    return email.upper()
```

In [104]:

```
df['email'].apply(update_email)
```

Out[104]:

```
0    COREYMSCHAFER@GMAIL.COM
1          JANEDOE@EMAIL.COM
2          JOHNDOE@EMAIL.COM
Name: email, dtype: object
```

In [105]:

```
df['email'] = df['email'].apply(update_email)
```

In [106]:

```
df
```

Out[106]:

	email	first	last
0	COREYMSCHAFER@GMAIL.COM	Corey	Schafer
1	JANEDOE@EMAIL.COM	Jane	Doe
2	JOHNDOE@EMAIL.COM	Smith	Smith

In [107]:

```
df['email'] = df['email'].apply(lambda x: x.lower())
```

In [108]:

```
df
```

Out[108]:

	email	first	last
0	coreymschafer@gmail.com	Corey	Schafer
1	janedoe@email.com	Jane	Doe
2	johndoe@email.com	Smith	Smith

In [109]:

```
df.apply(len)
```

Out[109]:

```
email    3
first    3
last     3
dtype: int64
```

In [110]:

```
len(df['email'])
```

Out[110]:

```
3
```

In [111]:

```
df.apply(len, axis = 'columns')
```

Out[111]:

```
0    3
1    3
2    3
dtype: int64
```

In [112]:

```
df.apply(pd.Series.min)
```

Out[112]:

```
email    coreymschafer@gmail.com
first                      Corey
last                         Doe
dtype: object
```

In [113]:

```
df.apply(lambda x:x.min())
```

Out[113]:

```
email    coreymschafer@gmail.com
first                      Corey
last                         Doe
dtype: object
```

In [114]:

```
df.applymap(len)
```

Out[114]:

	email	first	last
0	23	5	7
1	17	4	3
2	17	5	5

In [115]:

```
df.applymap(str.lower)
```

Out[115]:

	email	first	last
0	coreymschafer@gmail.com	corey	schafer
1	janedoe@email.com	jane	doe
2	johndoe@email.com	smith	smith

In [116]:

```
df['first'].map({'Corey':'Chris', 'jane':'mary'})
```

Out[116]:

```
0    Chris
1      NaN
2      NaN
Name: first, dtype: object
```

In [117]:

In [117]:

```
df['first'] = df['first'].replace({'Corey':'Chris', 'jane':'mary'})
```

In [118]:

```
df['first']
```

Out[118]:

```
0    Chris
1     Jane
2    Smith
Name: first, dtype: object
```

Now working the stack overflow dataset

In [119]:

```
df1.columns
```

Out[119]:

```
Index(['MainBranch', 'Employment', 'Country', 'US_State', 'UK_Country',
       'EdLevel', 'Age1stCode', 'LearnCode', 'YearsCode', 'YearsCodePro',
       'DevType', 'OrgSize', 'Currency', 'CompTotal', 'CompFreq',
       'LanguageHaveWorkedWith', 'LanguageWantToWorkWith',
       'DatabaseHaveWorkedWith', 'DatabaseWantToWorkWith',
       'PlatformHaveWorkedWith', 'PlatformWantToWorkWith',
       'WebframeHaveWorkedWith', 'WebframeWantToWorkWith',
       'MiscTechHaveWorkedWith', 'MiscTechWantToWorkWith',
       'ToolsTechHaveWorkedWith', 'ToolsTechWantToWorkWith',
       'NEWCollabToolsHaveWorkedWith', 'NEWCollabToolsWantToWorkWith', 'OpSys',
       'NEWStuck', 'NEWSOSites', 'SOVisitFreq', 'SOAccount', 'SOPartFreq',
       'SOComm', 'NEWOtherComms', 'Age', 'Gender', 'Trans', 'Sexuality',
       'Ethnicity', 'Accessibility', 'MentalHealth', 'SurveyLength',
       'SurveyEase', 'ConvertedCompYearly'],
      dtype='object')
```

In [120]:

```
df1.rename(columns = {'ConvertedCompYearly': 'SalaryUSD'}, inplace=True)
```

In [121]:

```
df1
```

Out[121]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Ye
1	I am a developer by profession	Independent contractor, freelancer, or self-em...	Slovakia	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Coding Bootcamp;Other online resources (ex: vi...	
2	I am a student who is learning to code	Student, full-time	Netherlands	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
3	I am not primarily a developer, but I write co...	Student, full-time	Russian Federation	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
4	I am a developer	Employed	Austria	NaN	NaN	Master's degree (M.A., M.S.,	11 - 17	NaN	

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Ye
5	I am a developer by profession	Independent contractor, freelancer, or self-em...	United Kingdom of Great Britain and Northern I...	NaN	England	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	5 - 10 years	Friend or family member	
...	...	...	...	...	...	...	...	...	
83435	I am a developer by profession	Employed full-time	United States of America	Texas	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
83436	I am a developer by profession	Independent contractor, freelancer, or self-em...	Benin	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
83437	I am a developer by profession	Employed full-time	United States of America	New Jersey	NaN	Secondary school (e.g. American high school, G...	11 - 17 years	School	
83438	I am a developer by profession	Employed full-time	Canada	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Online Courses or Certification;Books / Physic...	
83439	I am a developer by profession	Employed full-time	Brazil	NaN	NaN	Professional degree (JD, MD, etc.)	11 - 17 years	School	

83439 rows × 47 columns

In [122]:

```
df1['SalaryUSD']
```

Out[122]:

```
ResponseId
1          62268.0
2              NaN
3              NaN
4              NaN
5              NaN
           ...
83435     160500.0
83436       3960.0
83437      90000.0
83438     816816.0
83439      21168.0
Name: SalaryUSD, Length: 83439, dtype: float64
```

In [123]:

```
df1['OpSys']
```

Out[123]:

```
ResponseId
1                MacOS
2              Windows
3                MacOS
4              Windows
5          Linux-based
```

```
          ...
83435          MacOS
83436    Linux-based
83437        Windows
83438          MacOS
83439    Linux-based
Name: OpSys, Length: 83439, dtype: object
```

In [124]:

```
df1['OpSys'].map({'MacOS':True, 'Windows':False,'Linux-based':'Good'})
```

Out[124]:

```
ResponseId
1          True
2         False
3          True
4         False
5          Good
          ...
83435      True
83436      Good
83437     False
83438      True
83439      Good
Name: OpSys, Length: 83439, dtype: object
```

In [125]:

```
df1['OpSys'] = df1['OpSys'].map({'MacOS':True, 'Windows':False,'Linux-based':'Good'})
```

In [126]:

```
df1
```

Out[126]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Ye
1	I am a developer by profession	Independent contractor, freelancer, or self-em...	Slovakia	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Coding Bootcamp;Other online resources (ex: vi...	
2	I am a student who is learning to code	Student, full-time	Netherlands	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
3	I am not primarily a developer, but I write co...	Student, full-time	Russian Federation	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
4	I am a developer by profession	Employed full-time	Austria	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	11 - 17 years	NaN	
5	I am a developer by profession	Independent contractor, freelancer, or self-em...	United Kingdom of Great Britain and Northern I...	NaN	England	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	5 - 10 years	Friend or family member	
...	...	...	...	...	...	...	...	...	

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Ye
83435	I am a developer by profession	Employed full-time	United States of America	Texas	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
83436	I am a developer by profession	Independent contractor, freelancer, or self-em...	Benin	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
83437	I am a developer by profession	Employed full-time	United States of America	New Jersey	NaN	Secondary school (e.g. American high school, G...	11 - 17 years	School	
83438	I am a developer by profession	Employed full-time	Canada	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Online Courses or Certification;Books / Physic...	
83439	I am a developer by profession	Employed full-time	Brazil	NaN	NaN	Professional degree (JD, MD, etc.)	11 - 17 years	School	

83439 rows × 47 columns

In [127]:

```
df
```

Out[127]:

	email	first	last
0	coreymschafer@gmail.com	Chris	Schafer
1	janedoe@email.com	Jane	Doe
2	johndoe@email.com	Smith	Smith

Add/Remove Rows and Columns From DataFrames

In [128]:

```
df['first'] + ' '+ df['last']
```

Out[128]:

```
0    Chris Schafer
1         Jane Doe
2      Smith Smith
dtype: object
```

In [129]:

```
df['full_name'] = df['first'] + ' '+ df['last']
```

In [130]:

```
df
```

Out[130]:

	email	first	last	full_name
0	coreymschafer@gmail.com	Chris	Schafer	Chris Schafer
1	janedoe@email.com	Jane	Doe	Jane Doe

	email	first	last	full_name
2	johndoe@email.com	Smith	Smith	Smith Smith

In [131]:

```
df.drop(columns = ['first', 'last'], inplace=True)
df
```

Out[131]:

	email	full_name
0	coreymschafer@gmail.com	Chris Schafer
1	janedoe@email.com	Jane Doe
2	johndoe@email.com	Smith Smith

In [132]:

```
df['full_name'].str.split(' ', expand=True)
```

Out[132]:

	0	1
0	Chris	Schafer
1	Jane	Doe
2	Smith	Smith

In [133]:

```
df[['first','last']] = df['full_name'].str.split(' ', expand=True)
```

In [134]:

```
df
```

Out[134]:

	email	full_name	first	last
0	coreymschafer@gmail.com	Chris Schafer	Chris	Schafer
1	janedoe@email.com	Jane Doe	Jane	Doe
2	johndoe@email.com	Smith Smith	Smith	Smith

In [135]:

```
df.append({'firsst':'Tony'}, ignore_index=True)
```

Out[135]:

	email	full_name	first	last	firsst
0	coreymschafer@gmail.com	Chris Schafer	Chris	Schafer	NaN
1	janedoe@email.com	Jane Doe	Jane	Doe	NaN
2	johndoe@email.com	Smith Smith	Smith	Smith	NaN
3	NaN	NaN	NaN	NaN	Tony

In [136]:

```
people = {
    "first": ["Tony", "Steve"],
    "last": ["Stark", "Rogers"],
    "email": ["Ironman@gmail.com","Cap@avenge.com"]
}
```

```
df2 = pd.DataFrame(people)
```

In [137]:

```
df2
```

Out[137]:

	first	last	email
0	Tony	Stark	Ironman@gmail.com
1	Steve	Rogers	Cap@avenge.com

In [138]:

```
df.append(df2, ignore_index=True, sort = False)
```

Out[138]:

	email	full_name	first	last
0	coreymschafer@gmail.com	Chris Schafer	Chris	Schafer
1	janedoe@email.com	Jane Doe	Jane	Doe
2	johndoe@email.com	Smith Smith	Smith	Smith
3	Ironman@gmail.com	NaN	Tony	Stark
4	Cap@avenge.com	NaN	Steve	Rogers

In [139]:

```
df =df.append(df2, ignore_index=True, sort = False)
```

In [140]:

```
df
```

Out[140]:

	email	full_name	first	last
0	coreymschafer@gmail.com	Chris Schafer	Chris	Schafer
1	janedoe@email.com	Jane Doe	Jane	Doe
2	johndoe@email.com	Smith Smith	Smith	Smith
3	Ironman@gmail.com	NaN	Tony	Stark
4	Cap@avenge.com	NaN	Steve	Rogers

In [141]:

```
df.drop(index=4)
```

Out[141]:

	email	full_name	first	last
0	coreymschafer@gmail.com	Chris Schafer	Chris	Schafer
1	janedoe@email.com	Jane Doe	Jane	Doe
2	johndoe@email.com	Smith Smith	Smith	Smith
3	Ironman@gmail.com	NaN	Tony	Stark

In [142]:

```
filt = df['last'] == 'Doe'
df.drop(index=df[filt].index)
```

Out[142]:

	email	full_name	first	last
0	coreymschafer@gmail.com	Chris Schafer	Chris	Schafer
2	johndoe@email.com	Smith Smith	Smith	Smith
3	Ironman@gmail.com	NaN	Tony	Stark
4	Cap@avenge.com	NaN	Steve	Rogers

In [143]:

```
df
```

Out[143]:

	email	full_name	first	last
0	coreymschafer@gmail.com	Chris Schafer	Chris	Schafer
1	janedoe@email.com	Jane Doe	Jane	Doe
2	johndoe@email.com	Smith Smith	Smith	Smith
3	Ironman@gmail.com	NaN	Tony	Stark
4	Cap@avenge.com	NaN	Steve	Rogers

In [144]:

```
df.sort_values(by='last', ascending=False)
```

Out[144]:

	email	full_name	first	last
3	Ironman@gmail.com	NaN	Tony	Stark
2	johndoe@email.com	Smith Smith	Smith	Smith
0	coreymschafer@gmail.com	Chris Schafer	Chris	Schafer
4	Cap@avenge.com	NaN	Steve	Rogers
1	janedoe@email.com	Jane Doe	Jane	Doe

In [145]:

```
df.sort_values(by=['last','first'], ascending=False)
```

Out[145]:

	email	full_name	first	last
3	Ironman@gmail.com	NaN	Tony	Stark
2	johndoe@email.com	Smith Smith	Smith	Smith
0	coreymschafer@gmail.com	Chris Schafer	Chris	Schafer
4	Cap@avenge.com	NaN	Steve	Rogers
1	janedoe@email.com	Jane Doe	Jane	Doe

In [146]:

```
df.sort_values(by=['last','first'], ascending=[False, True],inplace=True)
```

In [147]:

```
df
```

Out[147]:

	email	full_name	first	last
3	Ironman@gmail.com	NaN	Tony	Stark
2	johndoe@email.com	Smith Smith	Smith	Smith
0	coreymschafer@gmail.com	Chris Schafer	Chris	Schafer
4	Cap@avenge.com	NaN	Steve	Rogers
1	janedoe@email.com	Jane Doe	Jane	Doe

In [148]:

```
df['last'].sort_values()
```

Out[148]:

```
1         Doe
4     Rogers
0    Schafer
2      Smith
3      Stark
Name: last, dtype: object
```

Now working the stack overflow dataset

In [149]:

```
df1
```

Out[149]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Ye
1	I am a developer by profession	Independent contractor, freelancer, or self-em...	Slovakia	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Coding Bootcamp;Other online resources (ex: vi...	
2	I am a student who is learning to code	Student, full-time	Netherlands	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
3	I am not primarily a developer, but I write co...	Student, full-time	Russian Federation	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
4	I am a developer by profession	Employed full-time	Austria	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	11 - 17 years	NaN	
5	I am a developer by profession	Independent contractor, freelancer, or self-em...	United Kingdom of Great Britain and Northern I...	NaN	England	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	5 - 10 years	Friend or family member	
...	...	...	...	...	...	...	...	...	
83435	I am a developer by profession	Employed full-time	United States of America	Texas	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Ye
83436	I am a developer by profession	Independent contractor, freelancer, or self-em...	Benin	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
83437	I am a developer by profession	Employed full-time	United States of America	New Jersey	NaN	Secondary school (e.g. American high school, G...	11 - 17 years	School	
83438	I am a developer by profession	Employed full-time	Canada	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Online Courses or Certification;Books / Physic...	
83439	I am a developer by profession	Employed full-time	Brazil	NaN	NaN	Professional degree (JD, MD, etc.)	11 - 17 years	School	

83439 rows × 47 columns

In [150]:

```
df1.sort_values(by=['Country','SalaryUSD'],ascending=[True, False], inplace=True)
```

In [151]:

```
df1
```

Out[151]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Y
65400	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	25 - 34 years	Online Courses or Certification	
22667	I am a developer by profession	Independent contractor, freelancer, or self-em...	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	45 - 54 years	School	
27199	I am a developer by profession	Independent contractor, freelancer, or self-em...	Afghanistan	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	5 - 10 years	Online Courses or Certification	
31087	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
44641	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Online Courses or Certification	
...	...	...	...	...	...	...	...	...	
75694	I am a student who is learning to code	Student, full-time	Zimbabwe	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Other online resources (ex: videos, blogs, etc...	
	I am a	Not				Bachelor's degree		Other online	

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode
76083	developer by profession	employed, but looking for work	Zimbabwe	NaN	NaN	(B.A., B.S., B.Eng., etc.)	18 - 24 years	resources (ex: videos, blogs, etc...
79302	I am a student who is learning to code	Student, part-time	Zimbabwe	NaN	NaN	Secondary school (e.g. American high school, G...	35 - 44 years	Other online resources (ex: videos, blogs, etc)
80997	I code primarily as a hobby	Not employed, and not looking for work	Zimbabwe	NaN	NaN	Primary/elementary school	11 - 17 years	Other online resources (ex: videos, blogs, etc...
82407	I am a student who is learning to code	Employed part-time	Zimbabwe	NaN	NaN	Some college/university study without earning ...	18 - 24 years	School;Books / Physical media

83439 rows × 47 columns

In [152]:

```
df1[['Country','SalaryUSD']].head(500)
```

Out[152]:

ResponseId	Country	SalaryUSD
65400	Afghanistan	30468516.0
22667	Afghanistan	155496.0
27199	Afghanistan	51804.0
31087	Afghanistan	23964.0
44641	Afghanistan	15132.0
...	...	...
60929	Argentina	25296.0
73682	Argentina	25296.0
43210	Argentina	25200.0
20319	Argentina	25000.0
56837	Argentina	24328.0

500 rows × 2 columns

In [153]:

```
df1['SalaryUSD'].nlargest(10)
```

Out[153]:

```
ResponseId
66911    45241312.0
65400    30468516.0
40587    21822250.0
28792    20000000.0
12701    19200000.0
9609     17500000.0
5306     15000000.0
12904    14411628.0
66489    12750000.0
7206     12500000.0
Name: SalaryUSD, dtype: float64
```

In [154]:

```
df1.nsmallest(10,'SalaryUSD')
```

Out[154]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCod
1925	I am a developer by profession	Employed full-time	United States of America	Pennsylvania	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	11 - 17 years	School;Othe (please specify)
24286	I am a developer by profession	Independent contractor, freelancer, or self-em...	United States of America	California	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Books / Physica medi
46465	I am not primarily a developer, but I write co...	Employed full-time	United States of America	Washington	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Codin Bootcamp;Othe onlin resources (ex vi..
15164	I am a developer by profession	Employed full-time	China	NaN	NaN	Primary/elementary school	Younger than 5 years	Books / Physica medi
56791	I am a developer by profession	Employed full-time	Taiwan	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	11 - 17 years	Schoc
22033	I am a developer by profession	Employed full-time	Republic of Korea	NaN	NaN	Associate degree (A.A., A.S., etc.)	18 - 24 years	Codin Bootcamp;Othe onlin resources (ex vi..
59044	I am a developer by profession	Employed full-time	Taiwan	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	18 - 24 years	Other onlin resources (ex videos, blogs etc..
22454	I am a developer by profession	Employed full-time	Japan	NaN	NaN	Secondary school (e.g. American high school, G...	5 - 10 years	Other onlin resources (ex videos, blogs etc
45142	I am not primarily a developer, but I write co...	Employed full-time	China	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	11 - 17 years	Other onlin resources (ex videos, blogs etc..
53307	I am a developer by profession	Employed full-time	Japan	NaN	NaN	Some college/university study without earning ...	11 - 17 years	Other onlin resources (ex videos, blogs etc..

In [155]:

```
df1
```

Out[155]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Y
65400	I am a developer	Employed	Afghanistan	NaN	NaN	Master's degree (M.A., M.S., M.Eng	25 - 34	Online Courses or	

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode
22667	I am a developer by profession	Independent contractor, freelancer, or self-em...	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	45 - 54 years	School
27199	I am a developer by profession	Independent contractor, freelancer, or self-em...	Afghanistan	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	5 - 10 years	Online Courses or Certification
31087	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...
44641	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Online Courses or Certification
...	...	...	...	...	...	...	...	...
75694	I am a student who is learning to code	Student, full-time	Zimbabwe	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Other online resources (ex: videos, blogs, etc...
76083	I am a developer by profession	Not employed, but looking for work	Zimbabwe	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	18 - 24 years	Other online resources (ex: videos, blogs, etc...
79302	I am a student who is learning to code	Student, part-time	Zimbabwe	NaN	NaN	Secondary school (e.g. American high school, G...	35 - 44 years	Other online resources (ex: videos, blogs, etc)
80997	I code primarily as a hobby	Not employed, and not looking for work	Zimbabwe	NaN	NaN	Primary/elementary school	11 - 17 years	Other online resources (ex: videos, blogs, etc...
82407	I am a student who is learning to code	Employed part-time	Zimbabwe	NaN	NaN	Some college/university study without earning ...	18 - 24 years	School;Books / Physical media

83439 rows × 47 columns

Grouping and Aggregating - Analyzing and Exploring Your Data

In [156]:

```
df1['SalaryUSD'].head(16)
```

Out[156]:

```
ResponseId
65400    30468516.0
22667      155496.0
27199       51804.0
31087       23964.0
44641       15132.0
11378        9792.0
17487        5448.0
51429        4668.0
```

```
14883         4200.0
17121         3108.0
55151          100.0
2070             NaN
4545             NaN
5101             NaN
9565             NaN
10946            NaN
Name: SalaryUSD, dtype: float64
```

In [157]:

```
df1['SalaryUSD'].median()
```

Out[157]:

```
56211.0
```

In [158]:

```
df1.median()
```

```
/var/folders/12/9_hj00mj65v4tgx35t4t6c280000gn/T/ipykernel_1909/2279417019.py:1: FutureWa
rning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is
deprecated; in a future version this will raise TypeError.  Select only valid columns bef
ore calling the reduction.
  df1.median()
```

Out[158]:

```
CompTotal    67000.0
SalaryUSD    56211.0
dtype: float64
```

In [159]:

```
df1.describe()
```

Out[159]:

	CompTotal	SalaryUSD
count	4.718300e+04	4.684400e+04
mean	2.119407e+69	1.184262e+05
std	4.603702e+71	5.272944e+05
min	0.000000e+00	1.000000e+00
25%	1.600000e+04	2.702500e+04
50%	6.700000e+04	5.621100e+04
75%	1.400000e+05	1.000000e+05
max	1.000000e+74	4.524131e+07

In [160]:

```
df1['SalaryUSD'].count()
```

Out[160]:

```
46844
```

In [161]:

```
df1['Accessibility'].value_counts()
```

Out[161]:

```
None of the above
72725
Prefer not to say
1918
```

I am blind / have difficulty seeing
1030
Or, in your own words:
842
I am deaf / hard of hearing
442
I am unable to / find it difficult to walk or stand without assistance
188
I am unable to / find it difficult to type
160
I am blind / have difficulty seeing;Or, in your own words:
54
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
49
I am deaf / hard of hearing;I am blind / have difficulty seeing
45
I am unable to / find it difficult to type;I am unable to / find it difficult to walk or stand without assistance
31
I am unable to / find it difficult to type;Or, in your own words:
16
I am blind / have difficulty seeing;I am unable to / find it difficult to type
15
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;Or, in your own words:
14
I am deaf / hard of hearing;I am unable to / find it difficult to walk or stand without assistance
10
I am deaf / hard of hearing;Or, in your own words:
10
I am unable to / find it difficult to walk or stand without assistance;Or, in your own words:
9
I am deaf / hard of hearing;I am unable to / find it difficult to type
7
I am deaf / hard of hearing;I am blind / have difficulty seeing;I am unable to / find it difficult to type
6
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
6
I am blind / have difficulty seeing;I am unable to / find it difficult to walk or stand without assistance
5
I am deaf / hard of hearing;I am blind / have difficulty seeing;Or, in your own words:
4
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:
4
I am blind / have difficulty seeing;I am unable to / find it difficult to type;Or, in your own words:
3
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;Or, in your own words:
2
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
2
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
2
I am deaf / hard of hearing;I am unable to / find it difficult to walk or stand without assistance;Or, in your own words:
1
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:
1
I am deaf / hard of hearing;I am blind / have difficulty seeing;I am unable to / find it difficult to type;Or, in your own words:
1

```
I am blind / have difficulty seeing;I am unable to / find it difficult to walk or stand w
ithout assistance;Or, in your own words:
1
Name: Accessibility, dtype: int64
```

In [162]:

```
df1['SurveyEase']
```

Out[162]:

```
ResponseId
65400                       Difficult
22667                            Easy
27199                       Difficult
31087    Neither easy nor difficult
44641    Neither easy nor difficult
                   ...
75694    Neither easy nor difficult
76083                            Easy
79302                             NaN
80997                            Easy
82407    Neither easy nor difficult
Name: SurveyEase, Length: 83439, dtype: object
```

In [163]:

```
schema_df.loc['SurveyEase']
```

Out[163]:

```
qid                                                        QID133
question      How easy or difficult was this survey to compl...
force_resp                                                 False
type                                                          MC
selector                                                    MAVR
Name: SurveyEase, dtype: object
```

In [164]:

```
df1['SurveyEase'].value_counts()
```

Out[164]:

```
Easy                         63087
Neither easy nor difficult   18091
Difficult                      770
Name: SurveyEase, dtype: int64
```

In [165]:

```
df1['SurveyEase'].value_counts(normalize=True)
```

Out[165]:

```
Easy                         0.769842
Neither easy nor difficult   0.220762
Difficult                    0.009396
Name: SurveyEase, dtype: float64
```

In [166]:

```
df1['Country']
```

Out[166]:

```
ResponseId
65400    Afghanistan
22667    Afghanistan
27199    Afghanistan
31087    Afghanistan
44641    Afghanistan
            ...
75694       Zimbabwe
```

```
76083        Zimbabwe
79302        Zimbabwe
80997        Zimbabwe
82407        Zimbabwe
Name: Country, Length: 83439, dtype: object
```

In [167]:

```
df1['Country'].value_counts()
```

Out[167]:

```
United States of America                                   15288
India                                                      10511
Germany                                                     5625
United Kingdom of Great Britain and Northern Ireland        4475
Canada                                                      3012
                                                           ...
Liechtenstein                                                  1
Papua New Guinea                                               1
Dominica                                                       1
Saint Kitts and Nevis                                          1
Saint Vincent and the Grenadines                               1
Name: Country, Length: 181, dtype: int64
```

In [168]:

```
country_grp = df1.groupby(['Country'])
```

In [169]:

```
country_grp.get_group('United States of America')
```

Out[169]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode
40587	I am a developer by profession	Employed full-time	United States of America	California	NaN	Some college/university study without earning ...	Younger than 5 years	Other online resources (ex: videos, blogs, etc...
9609	I am not primarily a developer, but I write co...	Employed full-time	United States of America	New York	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	18 - 24 years	School
5306	I am a developer by profession	Independent contractor, freelancer, or self-em...	United States of America	New Jersey	NaN	Professional degree (JD, MD, etc.)	11 - 17 years	Friend or family member;Books / Physical media
66489	I am a developer by profession	Employed full-time	United States of America	Washington	NaN	Some college/university study without earning ...	18 - 24 years	School;Online Courses or Certification;Books /...
7206	I am a developer by profession	Independent contractor, freelancer, or self-em...	United States of America	South Carolina	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	5 - 10 years	Other online resources (ex: videos, blogs, etc...
...	...	...	...	...	...	...	...	...
83401	I am a developer	Independent contractor,	United States	Indiana	NaN	Bachelor's degree (B.A.,	11 - 17	Other online resources (ex:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode
83413	I am a developer by profession	Independent contractor, freelancer, or self-em...	United States of America	Oregon	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...
83415	I code primarily as a hobby	Student, full-time	United States of America	Maryland	NaN	Secondary school (e.g. American high school, G...	5 - 10 years	Coding Bootcamp;Other online resources (ex: vi...
83427	I am a developer by profession	Employed full-time	United States of America	Nebraska	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...
83430	I code primarily as a hobby	Not employed, but looking for work	United States of America	Washington	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	18 - 24 years	Other online resources (ex: videos, blogs, etc)

15288 rows × 47 columns

In [170]:

```
country_grp.get_group('India')
```

Out[170]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	YearsCode
28792	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	35 - 44 years	Coding Bootcamp;Other online resources (ex: vi...	11
12701	I am a developer by profession	Employed full-time	India	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	25 - 34 years	Online Forum	5
64388	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	10
12794	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	School	18
3641	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other (please specify):	16
...	...	...	...	...	...	...	...	...	...
83387	I am a developer by profession	Employed full-time	India	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	18 - 24 years	NaN	4
	I used to be					Master's			

	ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	YearsCode
	83398	a developer by profession, but no...	I prefer not to say	India	NaN	NaN	degree (M.A., M.S., M.Eng., MBA, etc.)	18 - 24 years	Other (please specify):	14
	83407	I am a student who is learning to code	Student, full-time	India	NaN	NaN	Secondary school (e.g. American high school, G...	11 - 17 years	Other online resources (ex: videos, blogs, etc...	2
	83411	I am a student who is learning to code	Student, part-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	18 - 24 years	Other online resources (ex: videos, blogs, etc)	1
	83418	I am a developer by profession	Student, full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	School;Friend or family member;Online Courses ...	3

10511 rows × 47 columns

In [171]:

```
filt = df1['Country'] == 'United States of America'
df1.loc[filt]['Age1stCode'].value_counts()
```

Out[171]:

```
11 - 17 years           7995
5 - 10 years            3169
18 - 24 years           2776
25 - 34 years            739
35 - 44 years            223
Younger than 5 years     156
45 - 54 years            107
55 - 64 years             67
Older than 64 years       18
Name: Age1stCode, dtype: int64
```

In [172]:

```
country_grp['MainBranch'].value_counts().head(50
                                               )
```

Out[172]:

```
Country      MainBranch

Afghanistan  I am a developer by profession
27
             I am a student who is learning to code
15
             None of these
11
             I code primarily as a hobby
9
             I am not primarily a developer, but I write code sometimes as part of my wor
k       3
Albania      I am a developer by profession
52
             I am a student who is learning to code
17
             I am not primarily a developer, but I write code sometimes as part of my wor
k       4
Algeria      I am a developer by profession
27
             I am a student who is learning to code
```

```
8
                I code primarily as a hobby
7
                I am not primarily a developer, but I write code sometimes as part of my wor
k       3
                None of these
1
Andorra         I am a developer by profession
6
                I code primarily as a hobby
3
                I am a student who is learning to code
2
                None of these
1
Angola          I am a developer by profession
16
                I am a student who is learning to code
5
                None of these
3
                I code primarily as a hobby
2
                I used to be a developer by profession, but no longer am
1
Argentina       I am a developer by profession
449
                I am a student who is learning to code
55
                I am not primarily a developer, but I write code sometimes as part of my wor
k      52
                I code primarily as a hobby
23
                I used to be a developer by profession, but no longer am
5
                None of these
3
Armenia         I am a developer by profession
76
                I am a student who is learning to code
5
                I code primarily as a hobby
4
                None of these
4
                I am not primarily a developer, but I write code sometimes as part of my wor
k       3
Australia       I am a developer by profession
1091
                I am not primarily a developer, but I write code sometimes as part of my wor
k     198
                I am a student who is learning to code
163
                I code primarily as a hobby
149
                I used to be a developer by profession, but no longer am
39
                None of these
6
Austria         I am a developer by profession
570
                I am a student who is learning to code
110
                I am not primarily a developer, but I write code sometimes as part of my wor
k      78
                I code primarily as a hobby
40
                I used to be a developer by profession, but no longer am
7
                None of these
3
Azerbaijan      I am a developer by profession
```

```
42
             I am a student who is learning to code
18
             I am not primarily a developer, but I write code sometimes as part of my wor
k       3
             I code primarily as a hobby
1
Bahamas      I am a student who is learning to code
5
Name: MainBranch, dtype: int64
```

In [173]:

```
country_grp['MainBranch'].value_counts().loc['China']
```

Out[173]:

```
MainBranch
I am a developer by profession                                          638
I am a student who is learning to code                                  252
I am not primarily a developer, but I write code sometimes as part of my work     87
I code primarily as a hobby                                              54
I used to be a developer by profession, but no longer am                 14
None of these                                                            10
Name: MainBranch, dtype: int64
```

In [174]:

```
country_grp['SalaryUSD'].median().loc['India']
```

Out[174]:

```
14748.0
```

In [175]:

```
country_grp['SalaryUSD'].agg(['median','mean'])
```

Out[175]:

Country	median	mean
Afghanistan	9792.0	2.794748e+06
Albania	15900.0	4.499814e+04
Algeria	9875.0	1.446114e+04
Andorra	94045.5	8.928200e+04
Angola	9750.0	2.155680e+04
...	...	...
Venezuela, Bolivarian Republic of...	12000.0	2.246505e+04
Viet Nam	12678.0	1.995289e+04
Yemen	3954.0	5.628667e+03
Zambia	9816.0	1.991491e+04
Zimbabwe	7200.0	1.141477e+04

181 rows × 2 columns

In [176]:

```
country_grp['SalaryUSD'].agg(['median','mean']).loc['India']
```

Out[176]:

```
median    14748.000000
mean      42522.583464
Name: India, dtype: float64
```

In [177]:

```
filt = df1['Country'] == 'United States of America'
df1.loc[filt]['SurveyEase'].str.contains("Easy").sum()
```

Out[177]:

```
12218
```

In [178]:

```
country_grp['SurveyEase'].apply(lambda x: x.str.contains("Easy").sum())
```

Out[178]:

```
Country
Afghanistan                                 24
Albania                                     48
Algeria                                     31
Andorra                                      5
Angola                                      11
                                           ...
Venezuela, Bolivarian Republic of...        86
Viet Nam                                   219
Yemen                                       13
Zambia                                      16
Zimbabwe                                    20
Name: SurveyEase, Length: 181, dtype: int64
```

In [179]:

```
country_respomdents = df1['Country'].value_counts()
country_respomdents
```

Out[179]:

```
United States of America                                15288
India                                                   10511
Germany                                                  5625
United Kingdom of Great Britain and Northern Ireland     4475
Canada                                                   3012
                                                        ...
Liechtenstein                                               1
Papua New Guinea                                            1
Dominica                                                    1
Saint Kitts and Nevis                                       1
Saint Vincent and the Grenadines                            1
Name: Country, Length: 181, dtype: int64
```

In [180]:

```
easy = country_grp['SurveyEase'].apply(lambda x: x.str.contains("Easy"))
```

In [181]:

```
easy
```

Out[181]:

```
ResponseId
65400    False
22667     True
27199    False
31087    False
44641    False
         ...
75694    False
76083     True
79302      NaN
80997     True
82407    False
Name: SurveyEase, Length: 83439, dtype: object
```

In [182]:

```
data_df = pd.concat([easy,country_respomdents],axis='columns',sort = False)
```

In [183]:

```
data_df
```

Out[183]:

	SurveyEase	Country
1	True	NaN
2	True	NaN
3	True	NaN
4	False	NaN
5	True	NaN
...	...	...
Venezuela, Bolivarian Republic of...	NaN	104.0
Viet Nam	NaN	386.0
Yemen	NaN	20.0
Zambia	NaN	22.0
Zimbabwe	NaN	36.0

83620 rows × 2 columns

In [184]:

```
data_df.rename(columns = {"country":"no of easy", "SurveyEase": "easy"},inplace=True)
```

In [185]:

```
data_df
```

Out[185]:

	easy	Country
1	True	NaN
2	True	NaN
3	True	NaN
4	False	NaN
5	True	NaN
...	...	...
Venezuela, Bolivarian Republic of...	NaN	104.0
Viet Nam	NaN	386.0
Yemen	NaN	20.0
Zambia	NaN	22.0
Zimbabwe	NaN	36.0

83620 rows × 2 columns

In [ ]:

```

```

Cleaning Data - Casting Datatypes and Handling Missing Values

In [186]:

```
people = {
    'first': ['Corey', 'Jane', 'John', 'Chris', np.nan, None, 'NA'],
    'last': ['Schafer', 'Doe', 'Doe', 'Schafer', np.nan, np.nan, 'Missing'],
    'email': ['CoreyMSchafer@gmail.com', 'JaneDoe@email.com', 'JohnDoe@email.com', None,
np.nan, 'Anonymous@email.com', 'NA'],
    'age': ['33', '55', '63', '36', None, None, 'Missing']
}
```

In [187]:

```
df = pd.DataFrame(people)

df.replace('NA', np.nan, inplace=True)
df.replace('Missing', np.nan, inplace=True)
```

In [188]:

```
df
```

Out[188]:

	first	last	email	age
0	Corey	Schafer	CoreyMSchafer@gmail.com	33
1	Jane	Doe	JaneDoe@email.com	55
2	John	Doe	JohnDoe@email.com	63
3	Chris	Schafer	None	36
4	NaN	NaN	NaN	None
5	None	NaN	Anonymous@email.com	None
6	NaN	NaN	NaN	NaN

In [189]:

```
df.dropna()
```

Out[189]:

	first	last	email	age
0	Corey	Schafer	CoreyMSchafer@gmail.com	33
1	Jane	Doe	JaneDoe@email.com	55
2	John	Doe	JohnDoe@email.com	63

In [190]:

```
df.dropna(axis='index',how='all')
```

Out[190]:

	first	last	email	age
0	Corey	Schafer	CoreyMSchafer@gmail.com	33
1	Jane	Doe	JaneDoe@email.com	55
2	John	Doe	JohnDoe@email.com	63
3	Chris	Schafer	None	36
5	None	NaN	Anonymous@email.com	None

In [191]:

```
df.dropna(axis='columns',how='all')
```

Out[191]:

	first	last	email	age
0	Corey	Schafer	CoreyMSchafer@gmail.com	33
1	Jane	Doe	JaneDoe@email.com	55
2	John	Doe	JohnDoe@email.com	63
3	Chris	Schafer	None	36
4	NaN	NaN	NaN	None
5	None	NaN	Anonymous@email.com	None
6	NaN	NaN	NaN	NaN

In [192]:

```
df.dropna(axis='columns',how='any')
```

Out[192]:

0
1
2
3
4
5
6

In [193]:

```
df.dropna(axis='index',how='all', subset=['email'])
```

Out[193]:

	first	last	email	age
0	Corey	Schafer	CoreyMSchafer@gmail.com	33
1	Jane	Doe	JaneDoe@email.com	55
2	John	Doe	JohnDoe@email.com	63
5	None	NaN	Anonymous@email.com	None

In [194]:

```
df.dropna(axis='index',how='all', subset=['last','email'])
```

Out[194]:

	first	last	email	age
0	Corey	Schafer	CoreyMSchafer@gmail.com	33
1	Jane	Doe	JaneDoe@email.com	55
2	John	Doe	JohnDoe@email.com	63
3	Chris	Schafer	None	36
5	None	NaN	Anonymous@email.com	None

In [195]:

```
df
```

Out[195]:

	first	last	email	age
0	Corey	Schafer	CoreyMSchafer@gmail.com	33
1	Jane	Doe	JaneDoe@email.com	55
2	John	Doe	JohnDoe@email.com	63
3	Chris	Schafer	None	36
4	NaN	NaN	NaN	None
5	None	NaN	Anonymous@email.com	None
6	NaN	NaN	NaN	NaN

In [196]:

```
df = pd.DataFrame(people)

df.replace('NA',np.nan,inplace=True)
df.replace('Missing',np.nan,inplace=True)
```

In [197]:

```
df
```

Out[197]:

	first	last	email	age
0	Corey	Schafer	CoreyMSchafer@gmail.com	33
1	Jane	Doe	JaneDoe@email.com	55
2	John	Doe	JohnDoe@email.com	63
3	Chris	Schafer	None	36
4	NaN	NaN	NaN	None
5	None	NaN	Anonymous@email.com	None
6	NaN	NaN	NaN	NaN

In [198]:

```
df.dropna(axis='index',how='all', subset=['last','email'])
```

Out[198]:

	first	last	email	age
0	Corey	Schafer	CoreyMSchafer@gmail.com	33
1	Jane	Doe	JaneDoe@email.com	55
2	John	Doe	JohnDoe@email.com	63
3	Chris	Schafer	None	36
5	None	NaN	Anonymous@email.com	None

In [199]:

```
df.isna()
```

Out[199]:

	first	last	email	age
0	False	False	False	False
1	False	False	False	False

	first	last	email	age
2	False	False	False	False
3	False	False	True	False
4	True	True	True	True
5	True	True	False	True
6	True	True	True	True

In [200]:

```
df.fillna('0')
```

Out[200]:

	first	last	email	age
0	Corey	Schafer	CoreyMSchafer@gmail.com	33
1	Jane	Doe	JaneDoe@email.com	55
2	John	Doe	JohnDoe@email.com	63
3	Chris	Schafer	0	36
4	0	0	0	0
5	0	0	Anonymous@email.com	0
6	0	0	0	0

In [201]:

```
df.dtypes
```

Out[201]:

```
first     object
last      object
email     object
age       object
dtype: object
```

In [203]:

```
#df['age'].mean()
```

In [204]:

```
type(np.nan)
```

Out[204]:

```
float
```

In [205]:

```
df['age'] =df['age'].astype(float)
```

In [206]:

```
df.dtypes
```

Out[206]:

```
first     object
last      object
email     object
age      float64
dtype: object
```

In [207]:

```
df['age'].mean()
```

Out[207]:

46.75

Now working the stack overflow dataset

In [208]:

```
df1
```

Out[208]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Y
65400	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	25 - 34 years	Online Courses or Certification	
22667	I am a developer by profession	Independent contractor, freelancer, or self-em...	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	45 - 54 years	School	
27199	I am a developer by profession	Independent contractor, freelancer, or self-em...	Afghanistan	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	5 - 10 years	Online Courses or Certification	
31087	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
44641	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Online Courses or Certification	
...	...	...	...	...	...	...	...	...	
75694	I am a student who is learning to code	Student, full-time	Zimbabwe	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Other online resources (ex: videos, blogs, etc...	
76083	I am a developer by profession	Not employed, but looking for work	Zimbabwe	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	18 - 24 years	Other online resources (ex: videos, blogs, etc...	
79302	I am a student who is learning to code	Student, part-time	Zimbabwe	NaN	NaN	Secondary school (e.g. American high school, G...	35 - 44 years	Other online resources (ex: videos, blogs, etc)	
80997	I code primarily as a hobby	Not employed, and not looking for work	Zimbabwe	NaN	NaN	Primary/elementary school	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
82407	I am a student who is learning to code	Employed part-time	Zimbabwe	NaN	NaN	Some college/university study without earning ...	18 - 24 years	School;Books / Physical media	

83439 rows × 47 columns

In [209]:

```
na_vals = ['NA', 'Missing']
```

In [210]:

```
df1
```

Out[210]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Y
65400	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	25 - 34 years	Online Courses or Certification	
22667	I am a developer by profession	Independent contractor, freelancer, or self-em...	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	45 - 54 years	School	
27199	I am a developer by profession	Independent contractor, freelancer, or self-em...	Afghanistan	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	5 - 10 years	Online Courses or Certification	
31087	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
44641	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Online Courses or Certification	
...	...	...	...	...	...	...	...	...	
75694	I am a student who is learning to code	Student, full-time	Zimbabwe	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Other online resources (ex: videos, blogs, etc...	
76083	I am a developer by profession	Not employed, but looking for work	Zimbabwe	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	18 - 24 years	Other online resources (ex: videos, blogs, etc...	
79302	I am a student who is learning to code	Student, part-time	Zimbabwe	NaN	NaN	Secondary school (e.g. American high school, G...	35 - 44 years	Other online resources (ex: videos, blogs, etc)	
80997	I code primarily as a hobby	Not employed, and not looking for work	Zimbabwe	NaN	NaN	Primary/elementary school	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
82407	I am a student who is learning to code	Employed part-time	Zimbabwe	NaN	NaN	Some college/university study without earning ...	18 - 24 years	School;Books / Physical media	

83439 rows × 47 columns

In [213]:

```
#df1['YearsCode'].mean()
```

In [214]:

```
df1['YearsCode'].unique()
```

Out[214]:

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

In [215]:

```
df1['YearsCode'].replace('Less than 1 year',0,inplace=True)
```

In [216]:

```
df1['YearsCode'].replace('More than 50 years',51,inplace=True)
```

In [217]:

```
df1['YearsCode'].unique()
```

Out[217]:

```
array(['7', '5', '3', '16', '8', 0, '6', '4', nan, '2', '10', '9', '11',
       51, '32', '1', '20', '23', '12', '13', '25', '30', '15', '18',
       '46', '14', '17', '37', '28', '40', '35', '26', '38', '19', '22',
       '21', '33', '31', '36', '24', '34', '27', '41', '43', '50', '29',
       '45', '42', '48', '39', '47', '49', '44'], dtype=object)
```

In [218]:

```
df1['YearsCode'] = df1['YearsCode'].astype(float)
```

In [219]:

```
df1['YearsCode'].mean()
```

Out[219]:

```
12.338200169032717
```

In [220]:

```
df1['YearsCode'].median()
```

Out[220]:

```
10.0
```

Working with Dates and Time Series Data

In [221]:

```
d_parser = lambda x: pd.datetime.strptime(x, '%Y-%m-%d %I-%p')

df2 = pd.read_csv("/Users/narenderbeniwal/Downloads/ETH_1H.csv", date_parser=d_parser)
pd.set_option("display.max_columns", None)
```

In [222]:

```
df2
```

Out[222]:

	Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume
0	1586995200000	2020-04-16 00:00:00	ETHUSD	152.94	152.94	150.39	150.39	650.188125
1	1586991600000	2020-04-15 23:00:00	ETHUSD	155.81	155.81	151.39	152.94	4277.567299
2	1586988000000	2020-04-15 22:00:00	ETHUSD	157.18	157.30	155.32	155.81	106.337279
3	1586984400000	2020-04-15 21:00:00	ETHUSD	158.04	158.31	157.16	157.18	55.244131
4	1586980800000	2020-04-15 20:00:00	ETHUSD	157.10	158.10	156.87	158.04	144.262622
...	...	...	...	...	...	...	...	...
34492	1462813200	2016-05-09 17:00:00	ETHUSD	9.83	9.83	9.48	9.49	329.553213
34493	1462809600	2016-05-09 16:00:00	ETHUSD	9.99	9.99	9.79	9.83	62.379450
34494	1462806000	2016-05-09 15:00:00	ETHUSD	10.00	10.00	9.99	9.99	10.973567
34495	1462802400	2016-05-09 14:00:00	ETHUSD	9.55	10.00	9.55	10.00	235.774075
34496	1462798800	2016-05-09 13:00:00	ETHUSD	0.00	12.00	0.00	9.55	432.562115

34497 rows × 8 columns

In [223]:

```
df2.shape
```

Out[223]:

```
(34497, 8)
```

In [224]:

```
df2.loc[0,'Date']
```

Out[224]:

```
'2020-04-16 00:00:00'
```

In [225]:

```
df2['Date'] = pd.to_datetime(df2['Date'])
```

In [226]:

```
df2
```

Out[226]:

	Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume
0	1586995200000	2020-04-16 00:00:00	ETHUSD	152.94	152.94	150.39	150.39	650.188125
1	1586991600000	2020-04-15 23:00:00	ETHUSD	155.81	155.81	151.39	152.94	4277.567299
2	1586988000000	2020-04-15 22:00:00	ETHUSD	157.18	157.30	155.32	155.81	106.337279
3	1586984400000	2020-04-15 21:00:00	ETHUSD	158.04	158.31	157.16	157.18	55.244131
4	1586980800000	2020-04-15 20:00:00	ETHUSD	157.10	158.10	156.87	158.04	144.262622
...	...	...	...	...	...	...	...	...
34492	1462813200	2016-05-09 17:00:00	ETHUSD	9.83	9.83	9.48	9.49	329.553213
34493	1462809600	2016-05-09 16:00:00	ETHUSD	9.99	9.99	9.79	9.83	62.379450
34494	1462806000	2016-05-09 15:00:00	ETHUSD	10.00	10.00	9.99	9.99	10.973567
34495	1462802400	2016-05-09 14:00:00	ETHUSD	9.55	10.00	9.55	10.00	235.774075
34496	1462798800	2016-05-09 13:00:00	ETHUSD	0.00	12.00	0.00	9.55	432.562115

34497 rows × 8 columns

In [227]:

```
df2['Date'] = pd.to_datetime(df2['Date'], format='%Y=%m-%d %I-%p')
```

In [228]:

```
df2['Date']
```

Out[228]:

```
0        2020-04-16 00:00:00
1        2020-04-15 23:00:00
2        2020-04-15 22:00:00
3        2020-04-15 21:00:00
4        2020-04-15 20:00:00
                ...
34492    2016-05-09 17:00:00
34493    2016-05-09 16:00:00
34494    2016-05-09 15:00:00
34495    2016-05-09 14:00:00
34496    2016-05-09 13:00:00
Name: Date, Length: 34497, dtype: datetime64[ns]
```

In [229]:

```
df2
```

Out[229]:

	Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume
0	1586995200000	2020-04-16 00:00:00	ETHUSD	152.94	152.94	150.39	150.39	650.188125
1	1586991600000	2020-04-15 23:00:00	ETHUSD	155.81	155.81	151.39	152.94	4277.567299
2	1586988000000	2020-04-15 22:00:00	ETHUSD	157.18	157.30	155.32	155.81	106.337279
3	1586984400000	2020-04-15 21:00:00	ETHUSD	158.04	158.31	157.16	157.18	55.244131
4	1586980800000	2020-04-15 20:00:00	ETHUSD	157.10	158.10	156.87	158.04	144.262622
...	...	...	...	...	...	...	...	...
34492	1462813200	2016-05-09 17:00:00	ETHUSD	9.83	9.83	9.48	9.49	329.553213
34493	1462809600	2016-05-09 16:00:00	ETHUSD	9.99	9.99	9.79	9.83	62.379450
34494	1462806000	2016-05-09 15:00:00	ETHUSD	10.00	10.00	9.99	9.99	10.973567
34495	1462802400	2016-05-09 14:00:00	ETHUSD	9.55	10.00	9.55	10.00	235.774075
34496	1462798800	2016-05-09 13:00:00	ETHUSD	0.00	12.00	0.00	9.55	432.562115

34497 rows × 8 columns

In [230]:

```
df2.loc[0,'Date'].day_name()
```

Out[230]:

```
'Thursday'
```

In [231]:

```
df2['Date'].dt.day_name()
```

Out[231]:

```
0          Thursday
1         Wednesday
2         Wednesday
3         Wednesday
4         Wednesday
              ...
34492        Monday
34493        Monday
```

```
34494        Monday
34495        Monday
34496        Monday
Name: Date, Length: 34497, dtype: object
```

In [232]:

```
df2['Day of week'] = df2['Date'].dt.day_name()
```

In [233]:

```
df2
```

Out[233]:

	Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume	Day of week
0	1586995200000	2020-04-16 00:00:00	ETHUSD	152.94	152.94	150.39	150.39	650.188125	Thursday
1	1586991600000	2020-04-15 23:00:00	ETHUSD	155.81	155.81	151.39	152.94	4277.567299	Wednesday
2	1586988000000	2020-04-15 22:00:00	ETHUSD	157.18	157.30	155.32	155.81	106.337279	Wednesday
3	1586984400000	2020-04-15 21:00:00	ETHUSD	158.04	158.31	157.16	157.18	55.244131	Wednesday
4	1586980800000	2020-04-15 20:00:00	ETHUSD	157.10	158.10	156.87	158.04	144.262622	Wednesday
...	...	...	...	...	...	...	...	...	...
34492	1462813200	2016-05-09 17:00:00	ETHUSD	9.83	9.83	9.48	9.49	329.553213	Monday
34493	1462809600	2016-05-09 16:00:00	ETHUSD	9.99	9.99	9.79	9.83	62.379450	Monday
34494	1462806000	2016-05-09 15:00:00	ETHUSD	10.00	10.00	9.99	9.99	10.973567	Monday
34495	1462802400	2016-05-09 14:00:00	ETHUSD	9.55	10.00	9.55	10.00	235.774075	Monday
34496	1462798800	2016-05-09 13:00:00	ETHUSD	0.00	12.00	0.00	9.55	432.562115	Monday

34497 rows × 9 columns

In [234]:

```
df2['Date'].min()
```

Out[234]:

```
Timestamp('2016-05-09 13:00:00')
```

In [235]:

```
df2['Date'].max()
```

Out[235]:

```
Timestamp('2020-04-16 00:00:00')
```

In [236]:

```
df2['Date'].max() - df2['Date'].min()
```

Out[236]:

```
Timedelta('1437 days 11:00:00')
```

In [237]:

```
filt = (df2['Date']<'2020')
```

In [238]:

```
df2.loc[filt]
```

Out[238]:

Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume	Day of week

	Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume	Day of week
2545	1577833200000	2019-12-31 23:00:00	ETHUSD	128.68	128.85	127.82	128.84	857.650259	Tuesday
2546	1577829600000	2019-12-31 22:00:00	ETHUSD	128.40	128.93	127.77	128.68	3050.507350	Tuesday
2547	1577826000000	2019-12-31 21:00:00	ETHUSD	127.87	128.41	127.81	128.40	447.680372	Tuesday
2548	1577822400000	2019-12-31 20:00:00	ETHUSD	127.86	128.30	127.86	127.87	151.711128	Tuesday
2549	1577818800000	2019-12-31 19:00:00	ETHUSD	128.78	128.78	127.86	127.86	2450.933248	Tuesday
...	...	...	...	...	...	...	...	...	...
34492	1462813200	2016-05-09 17:00:00	ETHUSD	9.83	9.83	9.48	9.49	329.553213	Monday
34493	1462809600	2016-05-09 16:00:00	ETHUSD	9.99	9.99	9.79	9.83	62.379450	Monday
34494	1462806000	2016-05-09 15:00:00	ETHUSD	10.00	10.00	9.99	9.99	10.973567	Monday
34495	1462802400	2016-05-09 14:00:00	ETHUSD	9.55	10.00	9.55	10.00	235.774075	Monday
34496	1462798800	2016-05-09 13:00:00	ETHUSD	0.00	12.00	0.00	9.55	432.562115	Monday

31952 rows × 9 columns

In [239]:

```
filt1 = (df2['Date']>=pd.to_datetime('2019-01-01')) & (df2['Date']<pd.to_datetime('2020-01-01'))
```

In [240]:

```
df2.loc[filt1]
```

Out[240]:

	Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume	Day of week
2545	1577833200000	2019-12-31 23:00:00	ETHUSD	128.68	128.85	127.82	128.84	857.650259	Tuesday
2546	1577829600000	2019-12-31 22:00:00	ETHUSD	128.40	128.93	127.77	128.68	3050.507350	Tuesday
2547	1577826000000	2019-12-31 21:00:00	ETHUSD	127.87	128.41	127.81	128.40	447.680372	Tuesday
2548	1577822400000	2019-12-31 20:00:00	ETHUSD	127.86	128.30	127.86	127.87	151.711128	Tuesday
2549	1577818800000	2019-12-31 19:00:00	ETHUSD	128.78	128.78	127.86	127.86	2450.933248	Tuesday
...	...	...	...	...	...	...	...	...	...
11297	1546315200000	2019-01-01 04:00:00	ETHUSD	130.83	133.75	130.83	132.09	1035.840465	Tuesday
11298	1546311600000	2019-01-01 03:00:00	ETHUSD	129.79	131.00	129.79	130.83	1307.299291	Tuesday
11299	1546308000000	2019-01-01 02:00:00	ETHUSD	130.98	130.98	129.25	129.79	837.808380	Tuesday
11300	1546304400000	2019-01-01 01:00:00	ETHUSD	131.10	131.10	128.72	130.98	965.092541	Tuesday
11301	1546300800000	2019-01-01 00:00:00	ETHUSD	130.80	131.70	130.00	131.10	1288.434493	Tuesday

8757 rows × 9 columns

In [241]:

```
df2.set_index('Date', inplace=True)
```

In [242]:

```
df2
```

Out[242]:

Date	Unix Timestamp	Symbol	Open	High	Low	Close	Volume	Day of week
2020-04-16 00:00:00	1586995200000	ETHUSD	152.94	152.94	150.39	150.39	650.188125	Thursday
2020-04-15 23:00:00	1586991600000	ETHUSD	155.81	155.81	151.39	152.94	4277.567299	Wednesday

Date	Unix Timestamp	Symbol	Open	High	Low	Close	Volume	Day of week
2020-04-15 22:00:00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2020-04-15 21:00:00	1586984400000	ETHUSD	158.04	158.31	157.16	157.18	55.244131	Wednesday
2020-04-15 20:00:00	1586980800000	ETHUSD	157.10	158.10	156.87	158.04	144.262622	Wednesday
...	...	...	...	...	...	...	...	...
2016-05-09 17:00:00	1462813200	ETHUSD	9.83	9.83	9.48	9.49	329.553213	Monday
2016-05-09 16:00:00	1462809600	ETHUSD	9.99	9.99	9.79	9.83	62.379450	Monday
2016-05-09 15:00:00	1462806000	ETHUSD	10.00	10.00	9.99	9.99	10.973567	Monday
2016-05-09 14:00:00	1462802400	ETHUSD	9.55	10.00	9.55	10.00	235.774075	Monday
2016-05-09 13:00:00	1462798800	ETHUSD	0.00	12.00	0.00	9.55	432.562115	Monday

34497 rows × 8 columns

In [243]:

```
df2['2019']
```

```
/var/folders/12/9_hj00mj65v4tgx35t4t6c280000gn/T/ipykernel_1909/735323592.py:1: FutureWarning: Indexing a DataFrame with a datetimelike index using a single string to slice the rows, like `frame[string]`, is deprecated and will be removed in a future version. Use `frame.loc[string]` instead.
  df2['2019']
```

Out[243]:

Date	Unix Timestamp	Symbol	Open	High	Low	Close	Volume	Day of week
2019-12-31 23:00:00	1577833200000	ETHUSD	128.68	128.85	127.82	128.84	857.650259	Tuesday
2019-12-31 22:00:00	1577829600000	ETHUSD	128.40	128.93	127.77	128.68	3050.507350	Tuesday
2019-12-31 21:00:00	1577826000000	ETHUSD	127.87	128.41	127.81	128.40	447.680372	Tuesday
2019-12-31 20:00:00	1577822400000	ETHUSD	127.86	128.30	127.86	127.87	151.711128	Tuesday
2019-12-31 19:00:00	1577818800000	ETHUSD	128.78	128.78	127.86	127.86	2450.933248	Tuesday
...	...	...	...	...	...	...	...	...
2019-01-01 04:00:00	1546315200000	ETHUSD	130.83	133.75	130.83	132.09	1035.840465	Tuesday
2019-01-01 03:00:00	1546311600000	ETHUSD	129.79	131.00	129.79	130.83	1307.299291	Tuesday
2019-01-01 02:00:00	1546308000000	ETHUSD	130.98	130.98	129.25	129.79	837.808380	Tuesday
2019-01-01 01:00:00	1546304400000	ETHUSD	131.10	131.10	128.72	130.98	965.092541	Tuesday
2019-01-01 00:00:00	1546300800000	ETHUSD	130.80	131.70	130.00	131.10	1288.434493	Tuesday

8757 rows × 8 columns

In [244]:

```
df2['2020-01':'2020-03']
```

Out[244]:

Date	Unix Timestamp	Symbol	Open	High	Low	Close	Volume	Day of week
2020-03-01 00:00:00	1583020800000	ETHUSD	217.35	221.00	216.62	218.09	965.771245	Sunday
2020-02-29 23:00:00	1583017200000	ETHUSD	222.93	223.96	217.26	217.35	1247.560462	Saturday
2020-02-29 22:00:00	1583013600000	ETHUSD	223.59	223.76	222.32	222.93	118.446008	Saturday
2020-02-29 21:00:00	1583010000000	ETHUSD	224.81	225.08	223.05	223.59	191.602533	Saturday
2020-02-29 20:00:00	1583006400000	ETHUSD	225.56	225.56	223.55	224.81	158.232905	Saturday
...	...	...	...	...	...	...	...	...

Date	Unix Timestamp	Symbol	Open	High	Low	Close	Volume	Day of week
2020-02-01 04:00:00	1580529600000	ETHUSD	182.65	183.42	181.86	182.03	61.277147	Saturday
2020-02-01 03:00:00	1580526000000	ETHUSD	183.20	183.51	182.63	182.65	35.431896	Saturday
2020-02-01 02:00:00	1580522400000	ETHUSD	183.21	184.05	182.98	183.20	285.663975	Saturday
2020-02-01 01:00:00	1580518800000	ETHUSD	181.20	183.38	181.08	183.21	189.571066	Saturday
2020-02-01 00:00:00	1580515200000	ETHUSD	179.75	181.20	179.15	181.20	153.640006	Saturday

697 rows × 8 columns

In [245]:

```
df2['2020-01':'2020-03']['Close'].mean()
```

Out[245]:

```
238.04850789096147
```

In [246]:

```
df2['2020-01':'2020-03'].head(20)
```

Out[246]:

Date	Unix Timestamp	Symbol	Open	High	Low	Close	Volume	Day of week
2020-03-01 00:00:00	1583020800000	ETHUSD	217.35	221.00	216.62	218.09	965.771245	Sunday
2020-02-29 23:00:00	1583017200000	ETHUSD	222.93	223.96	217.26	217.35	1247.560462	Saturday
2020-02-29 22:00:00	1583013600000	ETHUSD	223.59	223.76	222.32	222.93	118.446008	Saturday
2020-02-29 21:00:00	1583010000000	ETHUSD	224.81	225.08	223.05	223.59	191.602533	Saturday
2020-02-29 20:00:00	1583006400000	ETHUSD	225.56	225.56	223.55	224.81	158.232905	Saturday
2020-02-29 19:00:00	1583002800000	ETHUSD	225.05	225.92	224.06	225.56	174.783652	Saturday
2020-02-29 18:00:00	1582999200000	ETHUSD	225.37	225.99	224.65	225.05	388.282143	Saturday
2020-02-29 17:00:00	1582995600000	ETHUSD	222.94	225.39	222.27	225.37	198.134810	Saturday
2020-02-29 16:00:00	1582992000000	ETHUSD	225.43	226.06	222.43	222.94	79.720833	Saturday
2020-02-29 15:00:00	1582988400000	ETHUSD	223.31	226.04	221.66	225.43	41.941194	Saturday
2020-02-29 14:00:00	1582984800000	ETHUSD	224.32	224.32	221.17	223.31	201.109898	Saturday
2020-02-29 13:00:00	1582981200000	ETHUSD	225.66	225.66	223.07	224.32	253.084937	Saturday
2020-02-29 12:00:00	1582977600000	ETHUSD	226.43	226.52	224.26	225.66	21.756591	Saturday
2020-02-29 11:00:00	1582974000000	ETHUSD	223.11	226.67	222.84	226.43	1012.890661	Saturday
2020-02-29 10:00:00	1582970400000	ETHUSD	225.48	228.24	223.00	223.11	355.191491	Saturday
2020-02-29 09:00:00	1582966800000	ETHUSD	229.08	229.08	225.40	225.48	54.195815	Saturday
2020-02-29 08:00:00	1582963200000	ETHUSD	229.37	230.23	228.22	229.08	1.354596	Saturday
2020-02-29 07:00:00	1582959600000	ETHUSD	228.69	230.16	227.81	229.37	10.881507	Saturday
2020-02-29 06:00:00	1582956000000	ETHUSD	231.86	233.11	228.65	228.69	111.868107	Saturday
2020-02-29 05:00:00	1582952400000	ETHUSD	232.13	233.36	231.20	231.86	131.271104	Saturday

In [247]:

```
df2['2020-01':'2020-03'].max()
```

Out[247]:

```
Unix Timestamp    1583020800000
Symbol                   ETHUSD
Open                      286.48
High                      289.58
Low                       285.29
```

```
Close                    286.48
Volume              14699.656142
Day of week            Wednesday
dtype: object
```

In [248]:

```
df2['High']
```

Out[248]:

```
Date
2020-04-16 00:00:00    152.94
2020-04-15 23:00:00    155.81
2020-04-15 22:00:00    157.30
2020-04-15 21:00:00    158.31
2020-04-15 20:00:00    158.10
                         ...
2016-05-09 17:00:00      9.83
2016-05-09 16:00:00      9.99
2016-05-09 15:00:00     10.00
2016-05-09 14:00:00     10.00
2016-05-09 13:00:00     12.00
Name: High, Length: 34497, dtype: float64
```

In [249]:

```
df2['High'].resample('D').max()
```

Out[249]:

```
Date
2016-05-09     12.00
2016-05-10      9.96
2016-05-11     10.47
2016-05-12     12.00
2016-05-13     11.59
               ...
2020-04-12    165.37
2020-04-13    159.51
2020-04-14    162.15
2020-04-15    161.52
2020-04-16    152.94
Freq: D, Name: High, Length: 1439, dtype: float64
```

In [250]:

```
%matplotlib inline
```

In [251]:

```
df2['High']
```

Out[251]:

```
Date
2020-04-16 00:00:00    152.94
2020-04-15 23:00:00    155.81
2020-04-15 22:00:00    157.30
2020-04-15 21:00:00    158.31
2020-04-15 20:00:00    158.10
                         ...
2016-05-09 17:00:00      9.83
2016-05-09 16:00:00      9.99
2016-05-09 15:00:00     10.00
2016-05-09 14:00:00     10.00
2016-05-09 13:00:00     12.00
Name: High, Length: 34497, dtype: float64
```

In [252]:

```
highs =df2['High'].resample('D').max()
```

In [253]:

```
highs
```

Out[253]:

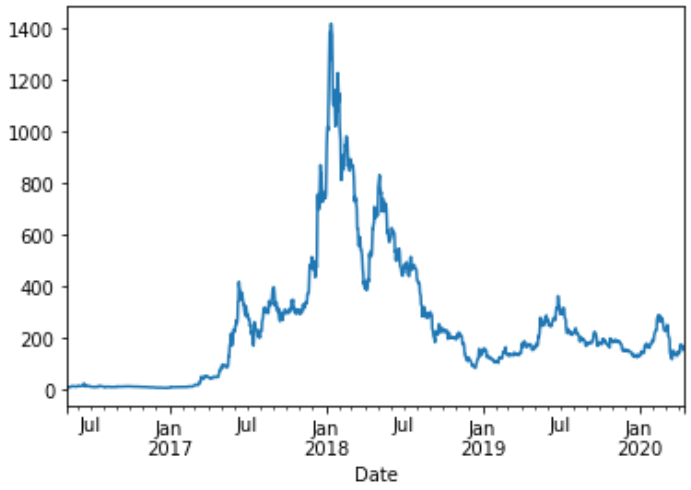
```
Date
2016-05-09     12.00
2016-05-10      9.96
2016-05-11     10.47
2016-05-12     12.00
2016-05-13     11.59
               ...
2020-04-12    165.37
2020-04-13    159.51
2020-04-14    162.15
2020-04-15    161.52
2020-04-16    152.94
Freq: D, Name: High, Length: 1439, dtype: float64
```

In [254]:

```
highs.plot()
```

Out[254]:

```
<AxesSubplot:xlabel='Date'>
```



In [255]:

```
df2.resample('W').mean()
```

Out[255]:

Date	Unix Timestamp	Open	High	Low	Close	Volume
2016-05-15	1.463076e+09	10.140387	10.310516	10.052387	10.205290	70.896402
2016-05-22	1.463657e+09	13.042262	13.144048	12.965179	13.066964	242.104139
2016-05-29	1.464262e+09	12.481012	12.555536	12.385357	12.471012	345.546483
2016-06-05	1.464867e+09	13.586369	13.651667	13.542202	13.594583	281.854432
2016-06-12	1.465472e+09	14.287500	14.326190	14.260179	14.297798	309.536737
...	...	...	...	...	...	...
2020-03-22	1.584617e+12	124.640952	126.387917	122.802262	124.636012	2618.930260
2020-03-29	1.585222e+12	133.274762	134.287857	132.255714	133.285893	1374.652289
2020-04-05	1.585827e+12	137.517083	138.314583	136.791310	137.627440	898.945625
2020-04-12	1.586432e+12	163.601548	164.731071	162.422560	163.693929	1218.547250
2020-04-19	1.586866e+12	156.890137	157.744384	155.952055	156.778219	695.573831

206 rows × 6 columns

In [256]:

```
df2.resample('W').agg({'Close':'mean','Low':'min','Volume':'sum'})
```

Out[256]:

Date	Close	Low	Volume
2016-05-15	10.205290	0.00	10988.942273
2016-05-22	13.066964	10.06	40673.495362
2016-05-29	12.471012	10.41	58051.809091
2016-06-05	13.594583	12.41	47351.544496
2016-06-12	14.297798	13.83	52002.171838
...	...	...	...
2020-03-22	124.636012	100.70	439980.283707
2020-03-29	133.285893	119.17	230941.584515
2020-04-05	137.627440	124.09	151022.864981
2020-04-12	163.693929	142.87	204715.937994
2020-04-19	156.778219	150.12	50776.889638

206 rows × 3 columns

Reading/Writing Data to Different Sources - Excel, JSON, SQL, Etc

In [257]:

```
df1
```

Out[257]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Y
65400	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	25 - 34 years	Online Courses or Certification	
22667	I am a developer by profession	Independent contractor, freelancer, or self-em...	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	45 - 54 years	School	
27199	I am a developer by profession	Independent contractor, freelancer, or self-em...	Afghanistan	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	5 - 10 years	Online Courses or Certification	
31087	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
44641	I am a developer by profession	Employed full-time	Afghanistan	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Online Courses or Certification	
...	...	...	...	...	...	...	...	...	

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	Y
75694	I am a student who is learning to code	Student, full-time	Zimbabwe	NaN	NaN	Secondary school (e.g. American high school, G...	18 - 24 years	Other online resources (ex: videos, blogs, etc...	
76083	I am a developer by profession	Not employed, but looking for work	Zimbabwe	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	18 - 24 years	Other online resources (ex: videos, blogs, etc...	
79302	I am a student who is learning to code	Student, part-time	Zimbabwe	NaN	NaN	Secondary school (e.g. American high school, G...	35 - 44 years	Other online resources (ex: videos, blogs, etc)	
80997	I code primarily as a hobby	Not employed, and not looking for work	Zimbabwe	NaN	NaN	Primary/elementary school	11 - 17 years	Other online resources (ex: videos, blogs, etc...	
82407	I am a student who is learning to code	Employed part-time	Zimbabwe	NaN	NaN	Some college/university study without earning ...	18 - 24 years	School;Books / Physical media	

83439 rows × 47 columns

In [258]:

```
filt = (df1['Country'] == 'India')
india_df = df1.loc[filt]
india_df.head()
```

Out[258]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	YearsCode
28792	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	35 - 44 years	Coding Bootcamp;Other online resources (ex: vi...	11.0
12701	I am a developer by profession	Employed full-time	India	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	25 - 34 years	Online Forum	5.0
64388	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	10.0
12794	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	School	18.0
3641	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other (please specify):	16.0

In [259]:

```
india_df.to_csv('modified.csv')
```

In [260]:

```
india_df.to_excel('modified.xlsx')
```

In [261]:

```
test = pd.read_excel('modified.xlsx', index_col='ResponseId')
```

In [262]:

```
test
```

Out[262]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	YearsCode
28792	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	35 - 44 years	Coding Bootcamp;Other online resources (ex: vi...	11.0
12701	I am a developer by profession	Employed full-time	India	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	25 - 34 years	Online Forum	5.0
64388	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	10.0
12794	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	School	18.0
3641	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other (please specify):	16.0
...	...	...	...	...	...	...	...	...	...
83387	I am a developer by profession	Employed full-time	India	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	18 - 24 years	NaN	4.0
83398	I used to be a developer by profession, but no...	I prefer not to say	India	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	18 - 24 years	Other (please specify):	14.0
83407	I am a student who is learning to code	Student, full-time	India	NaN	NaN	Secondary school (e.g. American high school, G...	11 - 17 years	Other online resources (ex: videos, blogs, etc...	2.0
	I am a					Bachelor's		Other online	

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	YearsCode
83411	student who is learning to code	Student, part-time	India	NaN	NaN	degree (B.A., B.S., B.Eng., etc.)	18 - 24 years	resources (ex: videos, blogs, etc)	1.0
83418	I am a developer by profession	Student, full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	School;Friend or family member;Online Courses ...	3.0

10511 rows × 47 columns

In [263]:

```
test.head()
```

Out[263]:

ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	YearsCode
28792	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	35 - 44 years	Coding Bootcamp;Other online resources (ex: vi...	11.0
12701	I am a developer by profession	Employed full-time	India	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	25 - 34 years	Online Forum	5.0
64388	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	10.0
12794	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	School	18.0
3641	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other (please specify):	16.0

In [264]:

```
india_df.to_json('modified.json', orient='records', lines=True)
```

In [265]:

```
test = pd.read_json('modified.json', orient='records', lines=True)
```

In [266]:

```
test.head()
```

Out[266]:

	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	YearsCode	YearsCo
0	I am a developer by	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S.,	35 - 44 years	Coding Bootcamp;Other online	11.0	

	MainBranch	Employment	Country	US_State	UK_Country	EdLevel	Age1stCode	LearnCode	YearsCode	YearsCo
1	I am a developer by profession	Employed full-time	India	NaN	NaN	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	25 - 34 years	Online Forum	5.0	
2	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other online resources (ex: videos, blogs, etc...	10.0	
3	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	School	18.0	
4	I am a developer by profession	Employed full-time	India	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)	11 - 17 years	Other (please specify):	16.0	

In [267]:

```
from sqlalchemy import create_engine
import psycopg2
```

In [268]:

```
conda install psycopg2
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /Users/narenderbeniwal/Applications/anaconda3

  added / updated specs:
    - psycopg2


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    certifi-2022.9.24          |   py39hecd8cb5_0         155 KB
    ------------------------------------------------------------
                                           Total:         155 KB

The following packages will be UPDATED:

  certifi                          2022.9.14-py39hecd8cb5_0 --> 2022.9.24-py39hecd8cb5_0
None



Downloading and Extracting Packages
certifi-2022.9.24    | 155 KB    | ##################################### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
Retrieving notices: ...working... done

Note: you may need to restart the kernel to use updated packages.
```

In [269]:

```
engine = create_engine('postgresql://dbuser:dbpass@localhost:5432/sample_db')
```

In [270]:

```
#india_df.to_sql('sample_table', engine, if_exists='replace')
```

In [271]:

```
#sql_df = pd.read_sql('sample_table', engine, index_col='Respondent')
```

In [272]:

```
#sql_df.head()
```

In [273]:

```
#sql_df = pd.read_sql_query('SELECT * FROM sample_table', engine, index_col='Respondent')
```

In [274]:

```
#sql_df.head()
```

In [275]:

```
#posts_df = pd.read_json('https://raw.githubusercontent.com/CoreyMSchafer/code_snippets/master/Python/Flask_Blog/snippets/posts.json')
```

In [276]:

```
#posts_df.head()
```

In [ ]:

```

```