In [1]:
```
import pandas as pd
```

In [2]:
```
#READING THE csv FILES, CHANGING THEIR IDS TO DESIRED COLUMN#

df = pd.read_csv('data/survey_results_public.csv', index_col='ResponseId')
schema_df = pd.read_csv('data/survey_results_schema.csv', index_col='qname')
```

In [29]:
```
#SETTING ALL COLUMNS VISIBLE#

pd.set_option('display.max_columns', 79)
pd.set_option('display.max_rows', 79)
```

In [4]:
```
#df JUST SHOW THE COMPLETE DATAFRAME, FIRST 5 ROWS AND LAST 5 ROWS BY DEFAULT#

df
```

Out[4]:

ResponseId	MainBranch	Employment	RemoteWork	CodingActivities	EdLevel	LearnCode
1	None of these	NaN	NaN	NaN	NaN	NaN
2	I am a developer by profession	Employed, full-time	Fully remote	Hobby;Contribute to open-source projects	NaN	NaN
3	I am not primarily a developer, but I write co...	Employed, full-time	Hybrid (some remote, some in-person)	Hobby	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Books / Physical media;Friend or family member... d
4	I am a developer by profession	Employed, full-time	Fully remote	I don't code outside of work	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Books / Physical media;School (i.e., Universit...
5	I am a developer by profession	Employed, full-time	Hybrid (some remote, some in-person)	Hobby	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Other online resources (e.g., videos, blogs, f...
...	...	...	...	...	...	...
73264	I am a developer by profession	Employed, full-time	Fully remote	Freelance/contract work	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Books / Physical media;Other online resources ...
73265	I am a developer by profession	Employed, full-time	Full in-person	Hobby	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Other online resources (e.g., videos, blogs, f...
73266	I am not primarily a developer, but I write co...	Employed, full-time	Hybrid (some remote, some in-person)	Hobby;School or academic work	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Books / Physical media;Other online resources ...
73267	I am a developer by profession	Employed, full-time	Hybrid (some remote, some in-person)	Hobby	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Books / Physical media;On the job training
73268	I used to be a developer by profession, but no...	Independent contractor, freelancer, or self-em...	Fully remote	Hobby;Contribute to open-source projects;Boots...	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Books / Physical media;Friend or family member... d

73268 rows × 78 columns

In [5]:
```
#FUNCTION TO FIND HOW MANY DEVELOPERS PER LANGUAGE IN BRAZIL#
#You can search by calling the function and putting the desired language in second
#You may also change the country by changing the cntry variable

cntry = 'Brazil'

def qtdePorLing(country, ling):
    num = df.loc[df['Country']==country]['LanguageHaveWorkedWith'].str.contains(lir
    if len(ling) > 5: #simple conditional just to print a better table#
        print(ling,':\t',num)
    else:
        print(ling,':\t\t',num)

qtdePorLing(cntry,'Java')
qtdePorLing(cntry,'Python')
qtdePorLing(cntry,'JavaScript')
qtdePorLing(cntry,'HTML')
qtdePorLing(cntry,'C#')
qtdePorLing(cntry,'C')
qtdePorLing(cntry,'CSS')
qtdePorLing(cntry,'Shell')
```

```
Java :          1566
Python :        873
JavaScript :    1394
HTML :          1102
C# :            462
C :             1483
CSS :           1102
Shell :         583
```

In [6]:
```
#TOP 10 COUNTRIES WITH MOST RESPONSES IN THE SURVEY, CONSEQUENTLY COUNTRIES WITH MC
df['Country'].value_counts().head(10)
```

Out[6]:
```
United States of America                                  13543
India                                                      6639
Germany                                                    5395
United Kingdom of Great Britain and Northern Ireland       4190
Canada                                                     2490
France                                                     2328
Brazil                                                     2109
Poland                                                     1732
Netherlands                                                1555
Spain                                                      1521
Name: Country, dtype: int64
```

In [7]:
```
#CREATING A GROUPBY (by countries) TO ENHANCE THE SEARCH FOR DATA AND VALUES#
country_grp = df.groupby(['Country'])
```

In [8]:
```
#THE 10 MOST COMMON SALARIES OF BRAZILIAN DEVELOPERS IN USD PER YEAR#

country_grp['CompTotal'].value_counts().loc['Brazil'].head(10)
```

Out[8]:
```
CompTotal
10000.0    60
8000.0     47
5000.0     46
6000.0     42
12000.0    42
15000.0    38
7000.0     37
11000.0    29
4000.0     27
20000.0    26
Name: CompTotal, dtype: int64
```

In [9]:
```
df['Gender'].value_counts()
```

Out[9]:
```
Man
64607
Woman
3399
Prefer not to say
1172
Non-binary, genderqueer, or gender non-conforming
704
Or, in your own words:
279
Man;Non-binary, genderqueer, or gender non-conforming
235
Man;Or, in your own words:
171
Woman;Non-binary, genderqueer, or gender non-conforming
160
Man;Woman;Non-binary, genderqueer, or gender non-conforming
31
Man;Woman
24
Man;Or, in your own words:;Woman;Non-binary, genderqueer, or gender non-conforming
18
Or, in your own words:;Woman;Non-binary, genderqueer, or gender non-conforming
15
Or, in your own words:;Non-binary, genderqueer, or gender non-conforming
14
Or, in your own words:;Woman
13
Man;Or, in your own words:;Non-binary, genderqueer, or gender non-conforming
9
Man;Or, in your own words:;Woman
2
Name: Gender, dtype: int64
```

In [10]:
```
country_grp['Gender'].value_counts().loc['Brazil'].head(5)
```

Out[10]:
```
Gender
Man                                                    1933
Woman                                                   106
Non-binary, genderqueer, or gender non-conforming        17
Prefer not to say                                        17
Man;Non-binary, genderqueer, or gender non-conforming    12
Name: Gender, dtype: int64
```

In [13]:
```
comp_grp = df.groupby(['CompTotal'])
```

In [28]:
```
#comp_grp['Country'].value_counts().loc[CompTotal>=70000]   #df['CompTotal']>70000]
```

In [ ]: