Skills
Network

Survey Dataset Exploration Lab

Estimated time needed: **30** minutes

Objectives

After completing this lab you will be able to:

- Load the dataset that will used thru the capstone project.
- Explore the dataset.
- Get familier with the data types.

Load the dataset

Import the required libraries.

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

The dataset is available on the IBM Cloud at the below url.

```
In [2]:  dataset_url = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/I
```

Load the data available at dataset_url into a dataframe.

```
In [3]:  df=pd.read_csv(dataset_url)
```

Explore the data set

It is a good idea to print the top 5 rows of the dataset to get a feel of how the dataset will look.

Display the top 5 rows and columns from your dataset.

```
In [10]:  print(df.head())
```

```
   Respondent                    MainBranch Hobbyist  \
0           4  I am a developer by profession       No
1           9  I am a developer by profession      Yes
2          13  I am a developer by profession      Yes
3          16  I am a developer by profession      Yes
4          17  I am a developer by profession      Yes


                                        OpenSourcer  \
0                                             Never
1                          Once a month or more often
2  Less than once a month but more than once per ...
3                                             Never
4  Less than once a month but more than once per ...


                                        OpenSource          Employment  \
0  The quality of OSS and closed source software ...  Employed full-time
1  The quality of OSS and closed source software ...  Employed full-time
2  OSS is, on average, of HIGHER quality than pro...  Employed full-time
3  The quality of OSS and closed source software ...  Employed full-time
4  The quality of OSS and closed source software ...  Employed full-time


          Country Student                                    EdLevel  \
0   United States      No          Bachelor's degree (BA, BS, B.Eng., etc.)
1     New Zealand      No  Some college/university study without earning ...
2   United States      No        Master's degree (MA, MS, M.Eng., MBA, etc.)
3  United Kingdom      No        Master's degree (MA, MS, M.Eng., MBA, etc.)
4       Australia      No          Bachelor's degree (BA, BS, B.Eng., etc.)


                                UndergradMajor  ...  \
0  Computer science, computer engineering, or sof...  ...
1  Computer science, computer engineering, or sof...  ...
2  Computer science, computer engineering, or sof...  ...
3                                            NaN  ...
4  Computer science, computer engineering, or sof...  ...


                              WelcomeChange  \
0   Just as welcome now as I felt last year
1   Just as welcome now as I felt last year
2  Somewhat more welcome now than last year
3   Just as welcome now as I felt last year
4   Just as welcome now as I felt last year


                                SONewContent   Age Gender Trans  \
0  Tech articles written by other developers;Indu...  22.0    Man    No
1                                             NaN  23.0    Man    No
2  Tech articles written by other developers;Cour...  28.0    Man    No
3  Tech articles written by other developers;Indu...  26.0    Man    No
4  Tech articles written by other developers;Indu...  29.0    Man    No


                 Sexuality                              Ethnicity Dependents  \
0  Straight / Heterosexual          White or of European descent         No
1                Bisexual          White or of European descent         No
2  Straight / Heterosexual          White or of European descent        Yes
3  Straight / Heterosexual          White or of European descent         No
4  Straight / Heterosexual  Hispanic or Latino/Latina;Multiracial         No
```

```
            SurveyLength                         SurveyEase
0  Appropriate in length                              Easy
1  Appropriate in length  Neither easy nor difficult
2  Appropriate in length                              Easy
3  Appropriate in length  Neither easy nor difficult
4  Appropriate in length                              Easy

[5 rows x 85 columns]
```

Find out the number of rows and columns

Start by exploring the numbers of rows and columns of data in the dataset.

Print the number of rows in the dataset.

```
In [5]: num_rows=df.shape[0]
        print(f"Number of rows: {num_rows}")
```

```
Number of rows: 11552
```

Print the number of columns in the dataset.

```
In [6]: num_columns=df.shape[1]
        print(f"Number of columns: {num_columns}")
```

```
Number of columns: 85
```

Identify the data types of each column

Explore the dataset and identify the data types of each column.

Print the datatype of all columns.

```
In [7]: print(df.dtypes)
```

```
Respondent       int64
MainBranch      object
Hobbyist        object
OpenSourcer     object
OpenSource      object
                 ...
Sexuality       object
Ethnicity       object
Dependents      object
SurveyLength    object
SurveyEase      object
Length: 85, dtype: object
```

Print the mean age of the survey participants.

```
In [8]: mean_age=df['Age'].mean()
        print(f"Mean age of survey participants: {mean_age}")
```

```
Mean age of survey participants: 30.77239449133718
```

The dataset is the result of a world wide survey. Print how many unique countries are there in the Country column.

```
In [9]:  unique_countries=df['Country'].nunique()
         print(f"Number of unique countries: {unique_countries}")
```

```
Number of unique countries: 135
```

Authors

Ramesh Sannareddy

Other Contributors

Rav Ahuja

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2020-10-17	0.1	Ramesh Sannareddy	Created initial version of the lab