



## Survey Dataset Exploration Lab

Estimated time needed: 30 minutes

### Objectives

After completing this lab you will be able to:

- Load the dataset that will be used thru the capstone project.
- Explore the dataset.
- Get familiar 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/IBM-DA0321EN-SkillsNetwork/LargeData/m1_survey_1.csv"
```

Load the data available at dataset\_url into a dataframe.

```
In [3]: # your code goes here
df = pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/LargeData/m1_survey_1.csv")
```

### 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 [4]: # your code goes here
df.head()
```

Out[4]:

	Respondent	MainBranch	Hobbyist	OpenSourcer	OpenSource	Employment	Country	Student	EdLevel	UndergradMajor	...	WelcomeChange	SO
0	4	I am a developer by profession	No	Never	The quality of OSS and closed source software ...	Employed full-time	United States	No	Bachelor's degree (BA, BS, B.Eng., etc.)	Computer science, computer engineering, or sof...	...	Just as welcome now as I felt last year	wr deve
1	9	I am a developer by profession	Yes	Once a month or more often	The quality of OSS and closed source software ...	Employed full-time	New Zealand	No	Some college/university study without earning ...	Computer science, computer engineering, or sof...	...	Just as welcome now as I felt last year	
2	13	I am a developer by profession	Yes	Less than once a month but more than once per ...	OSS is, on average, of HIGHER quality than pro...	Employed full-time	United States	No	Master's degree (MA, MS, M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	...	Somewhat more welcome now than last year	wr devel
3	16	I am a developer by profession	Yes	Never	The quality of OSS and closed source software ...	Employed full-time	United Kingdom	No	Master's degree (MA, MS, M.Eng., MBA, etc.)	NaN	...	Just as welcome now as I felt last year	wr deve
4	17	I am a developer by profession	Yes	Less than once a month but more than once per ...	The quality of OSS and closed source software ...	Employed full-time	Australia	No	Bachelor's degree (BA, BS, B.Eng., etc.)	Computer science, computer engineering, or sof...	...	Just as welcome now as I felt last year	wr deve

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]: # your code goes here
df.tail()
```

Out[5]:

	Respondent	MainBranch	Hobbyist	OpenSourcer	OpenSource	Employment	Country	Student	EdLevel	UndergradMajor	...	WelcomeChange	SC
11547	25136	I am a developer by profession	Yes	Never	OSS is, on average, of HIGHER quality than pro...	Employed full-time	United States	No	Master's degree (MA, MS, M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	...	Just as welcome now as I felt last year	w dev
11548	25137	I am a developer by profession	Yes	Never	The quality of OSS and closed source software ...	Employed full-time	Poland	No	Master's degree (MA, MS, M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	...	A lot more welcome now than last year	w dev
11549	25138	I am a developer by profession	Yes	Less than once per year	The quality of OSS and closed source software ...	Employed full-time	United States	No	Master's degree (MA, MS, M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	...	A lot more welcome now than last year	w dev
11550	25141	I am a developer by profession	Yes	Less than once a month but more than once per ...	OSS is, on average, of LOWER quality than prop...	Employed full-time	Switzerland	No	Secondary school (e.g. American high school, G...	NaN	...	Somewhat less welcome now than last year	
11551	25142	I am a developer by profession	Yes	Less than once a month but more than once per ...	OSS is, on average, of HIGHER quality than pro...	Employed full-time	United Kingdom	No	Other doctoral degree (Ph.D., Ed.D., etc.)	A natural science (ex. biology, chemistry, phy...	...	Just as welcome now as I felt last year	w dev

5 rows x 85 columns

Print the number of columns in the dataset.

```
In [6]: # your code goes here
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11552 entries, 0 to 11551
Data columns (total 85 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Respondent                           11552 non-null  int64
1   MainBranch                           11552 non-null  object
2   Hobbyist                             11552 non-null  object
3   OpenSourcer                          11552 non-null  object
4   OpenSource                           11471 non-null  object
5   Employment                           11552 non-null  object
6   Country                              11552 non-null  object
7   Student                              11499 non-null  object
8   EdLevel                              11436 non-null  object
9   UndergradMajor                       10812 non-null  object
10  EduOther                             11388 non-null  object
11  OrgSize                              11454 non-null  object
12  DevType                              11485 non-null  object
13  YearsCode                            11543 non-null  object
14  Age1stCode                           11539 non-null  object
15  YearsCodePro                         11536 non-null  object
16  CareerSat                            11552 non-null  object
17  JobSat                               11551 non-null  object
18  MgrIdiot                             11054 non-null  object
19  MgrMoney                             11050 non-null  object
20  MgrWant                              11054 non-null  object
21  JobSeek                              11552 non-null  object
22  LastHireDate                         11552 non-null  object
23  LastInt                              11129 non-null  object
24  FizzBuzz                             11515 non-null  object
25  JobFactors                           11549 non-null  object
26  ResumeUpdate                         11511 non-null  object
27  CurrencySymbol                       11552 non-null  object
28  CurrencyDesc                         11552 non-null  object
29  CompTotal                            10737 non-null  float64
30  CompFreq                             11346 non-null  object
31  ConvertedComp                        10730 non-null  float64
32  WorkWeekHrs                          11427 non-null  float64
33  WorkPlan                             11429 non-null  object
34  WorkChallenge                        11384 non-null  object
35  WorkRemote                           11544 non-null  object
36  WorkLoc                              11520 non-null  object
37  ImpSyn                               11547 non-null  object
38  CodeRev                              11551 non-null  object
39  CodeRevHrs                           9083 non-null   float64
40  UnitTests                            11523 non-null  object
41  PurchaseHow                           11354 non-null  object
42  PurchaseWhat                         11514 non-null  object
43  LanguageWorkedWith                  11541 non-null  object
44  LanguageDesireNextYear               11415 non-null  object
45  DatabaseWorkedWith                  11096 non-null  object
46  DatabaseDesireNextYear               10497 non-null  object
47  PlatformWorkedWith                  11130 non-null  object
48  PlatformDesireNextYear               10991 non-null  object
49  WebFrameWorkedWith                  10139 non-null  object
50  WebFrameDesireNextYear               9918 non-null   object
51  MiscTechWorkedWith                   9343 non-null   object
52  MiscTechDesireNextYear               10078 non-null  object
53  DevEnviron                           11523 non-null  object
54  OpSys                                11518 non-null  object
55  Containers                           11470 non-null  object
56  BlockchainOrg                        9198 non-null   object
--  --
```

```

57 Blockchainis      8915 non-null object
58 BetterLife       11452 non-null object
59 ITPerson          11517 non-null object
60 OffOn             11514 non-null object
61 SocialMedia       11251 non-null object
62 Extraversion      11532 non-null object
63 ScreenName        11039 non-null object
64 SOVisitist        11227 non-null object
65 SOVisitFreq       11547 non-null object
66 SOVisitTo         11551 non-null object
67 SOFindAnswer      11549 non-null object
68 SOTimeSaved        11501 non-null object
69 SOHowMuchTime     9616 non-null object
70 SOAccount         11551 non-null object
71 SOPartFreq        10404 non-null object
72 SOJobs            11546 non-null object
73 EntTeams          11547 non-null object
74 SOComm            11552 non-null object
75 WelcomeChange     11463 non-null object
76 SONewContent      9557 non-null object
77 Age               11255 non-null float64
78 Gender            11477 non-null object
79 Trans             11429 non-null object
80 Sexuality          11005 non-null object
81 Ethnicity         10869 non-null object
82 Dependents        11408 non-null object
83 SurveyLength      11533 non-null object
84 SurveyEase        11538 non-null object
dtypes: float64(5), int64(1), object(79)
memory usage: 7.5+ MB

```

## 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 [10]: *# your code goes here*

Print the mean age of the survey participants.

In [8]: *# your code goes here*

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

In [ ]: *# your code goes here*

## 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

Copyright © 2020 IBM Corporation. This notebook and its source code are released under the terms of the [MIT License](#).