

Hands-on Lab 5: Explore the Data Set

To launch the lab, check the box below indicating "I agree to use this app responsibly," and then click on the "Launch App" button. This will open up the lab environment in a new browser tab.

In case you need to download the lab notebook (.ipynb file) click [HERE](#) to download the lab notebook.

Once you **complete the lab** you can download the notebook as follows:

- Select the notebook which is displayed in the left **Navigation pane**.
- Right click on it and click on the **Download** option.
- Your notebook gets downloaded.

Later upload to GitHub by following the instructions in the reading [Getting Started with GitHub](#).

Further in the **MySubmission tab** of the assignment paste the GitHub link of the notebook in the **URL textbox** instead of the Watson studio link .

This course uses a third-party app, Hands-on Lab 5: Explore the Data Set, to enhance your learning experience. The app will reference basic information like your name, email, and Coursera ID.

About the dataset

Stack Overflow, a popular website for developers, conducted an online survey of software professionals across the world. The survey data was later open-sourced by Stack Overflow. The actual data set has around 90,000 responses.

The dataset I am going to use in this assignment comes from the following source: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/LargeData/m1_survey_data.csv.

You will be given a subset of the original data set in this capstone project. You will explore, analyze, and visualize this dataset and present your analysis.

Note: This randomized subset contains around 1/10th of the original data set. Any conclusions you draw after analyzing this subset may not reflect the real-world scenario.

The dataset is available as a .csv file [here](#).

The below table lists the questions asked in the survey and the column under which the response was collected.

Column Name	Question Text
Respondent	Randomized respondent ID number (not in order of survey response time)
MainBranch	Which of the following options best describes you today? Here, by “developer” we mean “someone who writes code.”
Hobbyist	Do you code as a hobby?
OpenSourcer	How often do you contribute to open source?
OpenSource	How do you feel about the quality of open source software (OSS)?
Employment	Which of the following best describes your current employment status?
Country	In which country do you currently reside?
Student	Are you currently enrolled in a formal, degree-granting college or university program?
EdLevel	Which of the following best describes the highest level of formal education that you’ ve completed?
UndergradMajor	What was your main or most important field of study?
EduOther	Which of the following types of non-degree education have you used or participated in? Please select all that apply.

Column Name	Question Text
OrgSize	Approximately how many people are employed by the company or organization you work for?
DevType	Which of the following describe you? Please select all that apply.
YearsCode	Including any education, how many years have you been coding?
Age1stCode	At what age did you write your first line of code or program? (E.g., webpage, Hello World, Scratch project)
YearsCodePro	How many years have you coded professionally (as a part of your work)?
CareerSat	Overall, how satisfied are you with your career thus far?
JobSat	How satisfied are you with your current job? (If you work multiple jobs, answer for the one you spend the most hours on.)
MgrIdiot	How confident are you that your manager knows what they 're doing?
MgrMoney	Do you believe that you need to be a manager to make more money?
MgrWant	Do you want to become a manager yourself in the future?
JobSeek	Which of the following best describes your current job-seeking status?
LastHireDate	When was the last time that you took a job with a new employer?

Column Name	Question Text
LastInt	In your most recent successful job interview (resulting in a job offer), you were asked to··· (check all that apply)
FizzBuzz	Have you ever been asked to solve FizzBuzz in an interview?
JobFactors	Imagine that you are deciding between two job offers with the same compensation, benefits, and location. Of the following factors, which 3 are MOST important to you?
ResumeUpdate	Think back to the last time you updated your resuméCV, or an online profile on a job site. What is the PRIMARY reason that you did so?
CurrencySymbol	Which currency do you use day-to-day? If your answer is complicated, please pick the one you’ re most comfortable estimating in.
CurrencyDesc	Which currency do you use day-to-day? If your answer is complicated, please pick the one you’ re most comfortable estimating in.
CompTotal	What is your current total compensation (salary, bonuses, and perks, before taxes and deductions), in CurrencySymbol? Please enter a whole number in the box below, without any punctuation. If you are paid hourly, please estimate an equivalent weekly, monthly, or yearly salary. If you prefer not to answer, please leave the box empty.
CompFreq	Is that compensation weekly, monthly, or yearly?
ConvertedComp	Salary converted to annual USD salaries using the exchange rate on 2019-02-01, assuming 12 working months and 50 working weeks.

Column Name	Question Text
WorkWeekHrs	On average, how many hours per week do you work?
WorkPlan	How structured or planned is your work?
WorkChallenge	Of these options, what are your greatest challenges to productivity as a developer? Select up to 3:
WorkRemote	How often do you work remotely?
WorkLoc	Where would you prefer to work?
ImpSyn	For the specific work you do, and the years of experience you have, how do you rate your own level of competence?
CodeRev	Do you review code as part of your work?
CodeRevHrs	On average, how many hours per week do you spend on code review?
UnitTests	Does your company regularly employ unit tests in the development of their products?
PurchaseHow	How does your company make decisions about purchasing new technology (cloud, AI, IoT, databases)?
PurchaseWhat	What level of influence do you, personally, have over new technology purchases at your organization?
LanguageWorkedWith	Which of the following programming, scripting, and markup languages have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the language and want to continue to do so, please check both boxes in that row.)

Column Name	Question Text
LanguageDesireNextYear	Which of the following programming, scripting, and markup languages have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the language and want to continue to do so, please check both boxes in that row.)
DatabaseWorkedWith	Which of the following database environments have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the database and want to continue to do so, please check both boxes in that row.)
DatabaseDesireNextYear	Which of the following database environments have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the database and want to continue to do so, please check both boxes in that row.)
PlatformWorkedWith	Which of the following platforms have you done extensive development work for over the past year? (If you both developed for the platform and want to continue to do so, please check both boxes in that row.)
PlatformDesireNextYear	Which of the following platforms have you done extensive development work for over the past year? (If you both developed for the platform and want to continue to do so, please check both boxes in that row.)
WebFrameWorkedWith	Which of the following web frameworks have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the framework and want to continue to do so, please check both boxes in that row.)

Column Name	Question Text
WebFrameDesireNextYear	Which of the following web frameworks have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the framework and want to continue to do so, please check both boxes in that row.)
MiscTechWorkedWith	Which of the following other frameworks, libraries, and tools have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the technology and want to continue to do so, please check both boxes in that row.)
MiscTechDesireNextYear	Which of the following other frameworks, libraries, and tools have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the technology and want to continue to do so, please check both boxes in that row.)
DevEnviron	Which development environment(s) do you use regularly? Please check all that apply.
OpSys	What is the primary operating system in which you work?
Containers	How do you use containers (Docker, Open Container Initiative (OCI), etc.)?
BlockchainOrg	How is your organization thinking about or implementing blockchain technology?
BlockchainIs	Blockchain / cryptocurrency technology is primarily:

Column Name	Question Text
BetterLife	Do you think people born today will have a better life than their parents?
ITperson	Are you the “IT support person” for your family?
OffOn	Have you tried turning it off and on again?
SocialMedia	What social media site do you use the most?
Extraversion	Do you prefer online chat or IRL conversations?
ScreenName	What do you call it?
SOVisit1st	To the best of your memory, when did you first visit Stack Overflow?
SOVisitFreq	How frequently would you say you visit Stack Overflow?
SOVisitTo	I visit Stack Overflow to... (check all that apply)
SOFindAnswer	On average, how many times a week do you find (and use) an answer on Stack Overflow?
SOTimeSaved	Think back to the last time you solved a coding problem using Stack Overflow, as well as the last time you solved a problem using a different resource. Which was faster?
SOHowMuchTime	About how much time did you save? If you’re not sure, please use your best estimate.
SOAccount	Do you have a Stack Overflow account?

Column Name	Question Text
SOPartFreq	How frequently would you say you participate in Q&A on Stack Overflow? By participate we mean ask, answer, vote for, or comment on questions.
SOJobs	Have you ever used or visited Stack Overflow Jobs?
EntTeams	Have you ever used Stack Overflow for Enterprise or Stack Overflow for Teams?
SOComm	Do you consider yourself a member of the Stack Overflow community?
WelcomeChange	Compared to last year, how welcome do you feel on Stack Overflow?
SONewContent	Would you like to see any of the following on Stack Overflow? Check all that apply.
Age	What is your age (in years)? If you prefer not to answer, you may leave this question blank.
Gender	Which of the following do you currently identify as? Please select all that apply. If you prefer not to answer, you may leave this question blank.
Trans	Do you identify as transgender?
Sexuality	Which of the following do you currently identify as? Please select all that apply. If you prefer not to answer, you may leave this question blank.
Ethnicity	Which of the following do you identify as? Please check all that apply. If you prefer not to answer, you may leave this question blank.

Column Name	Question Text
Dependents	Do you have any dependents (e.g., children, elders, or others) that you care for?
SurveyLength	How do you feel about the length of the survey this year?
SurveyEase	How easy or difficult was this survey to complete?

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.

```
import pandas as pd
```

```
# URL of the dataset
```

```
dataset_url = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/LargeData/m1_survey_data.csv"
```

```
# Load the data into a DataFrame
```

```
df = pd.read_csv(dataset_url)
```

```
# Display the first few rows of the DataFrame
```

```
print(df.head())
```

Explanation:

1. **Import pandas:** Ensure that the **pandas** library is imported.

2. **Define the URL:** Set the **URL** of the dataset.
3. **Load the Data:** Use **pd.read_csv()** to load the CSV file from the URL into a DataFrame.
4. **Display Data:** Use **df.head()** to display the first few rows of the DataFrame.

This code will download the CSV data from the provided URL and load it into a DataFrame, which you can then use for further analysis.

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.

```
df.head(5)
```

Explanation:

Display Data: Use **df.head(5)** to display the first 5 rows of the DataFrame.

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 & Columns in the dataset.

```
# Get the number of rows and columns

num_rows, num_columns = df.shape

# Print the number of rows and columns

print(f"Number of rows: {num_rows}")

print(f"Number of columns: {num_columns}")
```

Output: Number of rows: 11552

Number of columns: 85

Explanation:

1. **Get the Shape:** Use the **shape** attribute of the DataFrame to get the number of rows and columns.
2. **Print the Results:** Print the number of rows and columns.

Identify the data types of each column

Explore the dataset and identify the data types of each column. Print the datatype of all columns.

To identify and print the data types of each column in the dataset, I use the `dtypes` attribute of the pandas DataFrame. Here is the code to do that:

```
df.dtypes
```

Print Data Types: Use `df.dtypes` to print the data types of all columns in the DataFrame.

Print the mean age of the survey participants.

To calculate and print the **mean age** of the survey participants, I use the `mean()` function from the pandas library. Here is the complete code to do that:

```
# Calculate the mean age of the survey participants
mean_age = df['Age'].mean()

# Print the mean age
print(f"The mean age of the survey participants is: {mean_age}")
```

Explanation:

1. **Calculate Mean Age:** Use the `mean()` function on the 'Age' column to calculate the mean age.
2. **Print Mean Age:** Print the calculated mean age.

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

```
# Find the number of unique countries in the Country column
unique_countries = df['Country'].nunique()

# Print the number of unique countries
print(f"The number of unique countries in the survey is: {unique_countries}")
```

Output: The number of unique countries in the survey is: 135

