Summary IBM Data Analyst Capstone project_Jul 2023

Contents

What you'll learn

- Apply different techniques to collect and wrangle data
- Showcase your Data Analysis and Visualization skills
- Create a data analysis report and a compelling presentation
- Demonstrate proficiency with various Python Libraries

Run python code by **Visual Studio Code**

installed before: **Anaconda**

Introduction to Capstone Project

You have recently been hired as a Data Analyst by a global IT and business consulting services firm that is known for their expertise in IT solutions and their team of highly experienced IT consultants. In order to keep pace with changing technologies and remain competitive, your organization regularly analyzes data to help identify future skill requirements.

As a Data Analyst, you will be assisting with this initiative and have been tasked with collecting data from various sources and identifying trends for this year's report on emerging skills.

Your first task is to collect the top programming skills that are most in demand from various sources including:

- Job postings
- Training portals
- Surveys

Once you have collected enough data, you will begin analyzing the data and identify insights and trends that may include the following:

- What are the top programming languages in demand?
- What are the top database skills in demand?
- What are the popular IDEs?

You will begin by scraping internet web sites and accessing APIs to collect data in various formats like .csv files, excel sheets, and databases.

Once this is completed, you will make that data ready for analysis using data wrangling techniques.

When the data is ready you will then want to apply statistical techniques to analyze the data. Then bring all of your information together by using IBM Cognos Analytics to create your dashboard. And finally, show off your storytelling skills by sharing your findings in a presentation.

You will be evaluated using quizzes in each module as well as the final project presentation.

1/ Data Collection

Data Collection is the first step in solving any analysis problem and can be collected in many formats and from many sources. In the first module of the Capstone, we will collect data by scraping the internet and using web APIs.

- Collecting Data Using APIs
- Collecting Data Using Webscraping
- Exploring Data

Collecting Data Using APIs

(Optional) Hands-on Lab 1: Review of Accessing APIs

In the Python for Data Science course you learned to use REST APIs. However, if you require a refresher for invoking REST APIs from Python, please complete this lab before you move to the next one.

Right click and *copy this link* for the Jupyter notebook for this lab.

Note: Different browsers may have different options to copy the notebook.

If you are in Google Chrome, right click and "Copy link address".

In Mozilla Firefox, right click and choose "Copy Link Location".

Then as shown in the previous lab, create a new project in Watson Studio and add the Jupyter notebook referenced above to your notebooks and run through the instructions in the notebook to complete the lab.

Collecting Data Using Webscraping

Exploring Data

2/ Data Wrangling

In this module, you will be focusing on the cleaning of your dataset with various techniques. With these techniques you will be identifying duplicate rows, finding missing values, and normalizing the data.

- Finding Missing Values
- Determine Missing Values
- Finding Duplicates
- Removing Duplicates
- Normalizing Data

3/ Exploratory Data Analysis

In this module, begin working with the cleaned dataset from the previous module. You will now begin to analyze the dataset to find the distribution of data, presence of outliers and the correlation between different columns.

- Distribution
- Outliers
- Correlation

4/ Data Visualization

In module 4 of the Capstone, you will be required to create visualizations using the developer survey data. The visualizations you create should highlight the distribution of data, relationships between data, the composition of data, and comparison of data.

- Visualizing Distribution of Data
- Relationship
- Composition
- Comparison

5/ Building a Dashboard

In this module, you will create a dashboard using IBM Cognos Analytics. This platform will give you the ability to create various charts while assembling a dashboard that is appealing and easy to understand. Your dashboard will contain your data analysis, which should be intuitive and allow for the drill-down of data.

- Dashboards

6/ Final- assignment: Present Your Findings

You have analyzed the data in the previous modules, and now it is time to demonstrate your storytelling skills. In this module, you will create a compelling story that helps to clarify your analysis in an easy-to-understand presentation.

- Final Presentation