Course Two
Get Started with Python

Instructions

Use this PACE strategy document to record decisions and reflections as you work through this end-of-course project. You can use this document as a guide to consider your responses and reflections at different stages of the data analytical process. Additionally, the PACE strategy documents can be used as a resource when working on future projects.

Course Project Recap

Regardless of which track you have chosen to complete, your goals for this project are:

- ☐ Complete the questions in the Course 2 PACE strategy document
- ☐ Answer the questions in the Jupyter notebook project file
- ☐ Complete coding prep work on project's Jupyter notebook
- ☐ Summarize the column Dtypes
- ☐ Communicate important findings in the form of an executive summary

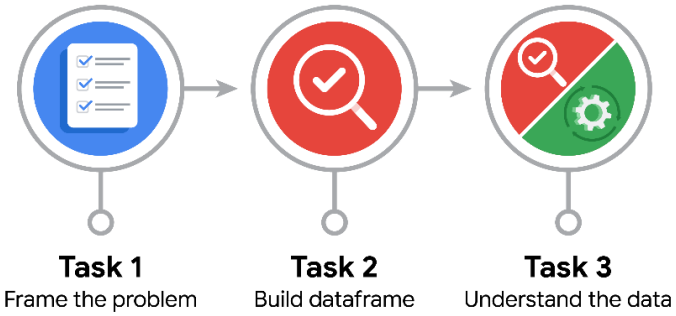
Relevant Interview Questions

Completing the end-of-course project will help you respond these types of questions that are often asked during the interview process:

- Describe the steps you would take to clean and transform an unstructured data set.
- What specific things might you look for as part of your cleaning process?
- What are some of the outliers, anomalies, or unusual things you might look for in the data cleaning process that might impact analyses or ability to create insights?

Reference Guide

This project has three tasks; the visual below identifies how the stages of PACE are incorporated across those tasks.



Data Project Questions & Considerations

PACE: Plan Stage

- How can you best prepare to understand and organize the provided information?

In this provided taxi cab information, I first started to familiarize every single variable from the dataset and understand its meaning. Then, perform the analysis on variables that might be useful to deeper understanding. It's essential to perform tasks such as checking data types, missing values, descriptive statistics, etc. Preparing and organizing these tasks will be beneficial for future activities about Exploratory Data Analysis, statistical testing, and so on.

- What follow-along and self-review codebooks will help you perform this work?

Jupyter Notebook to complete coding works.

- What are some additional activities a resourceful learner would perform before starting to code?

Practicing coding will be a great approach to enhancing coding skills. Besides the lessons and resources from the course work, there are sources from the internet, YouTube videos, books, etc. are the advantages to utilize on coding skills.

PACE: Analyze Stage

- Will the available information be sufficient to achieve the goal based on your intuition and the analysis of the variables?

The data might be sufficient to achieve the goal of the analysis in each variable. Further analysis is necessary to perform for deeper understanding.

- How would you build summary dataframe statistics and assess the min and max range of the data?

There's a method describe (), which is used to perform descriptive statistics for numerical variables. It shows metrics such as count, min, max, quartiles, etc.

- Do the averages of any of the data variables look unusual? Can you describe the interval data?

There's not much concern about the average for variables. Pickup time and dropoff time are two described for interval data.

PACE: Construct Stage

Note: The Construct stage does not apply to this workflow. The PACE framework can be adapted to fit the specific requirements of any project.

PACE: Execute Stage

- Given your current knowledge of the data, what would you initially recommend to your manager to investigate further prior to performing exploratory data analysis?

The first thing to investigate further is the appearance of negative values in the amount of tip, fare, and total in taxi rides. Also, the max values for these variables should be further investigated, which tend to be quite a lot higher than the normal range.

- What data initially presents as containing anomalies?

The negative values in fare, tip, etc. variables are the anomalies.

- What additional types of data could strengthen this dataset?
