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~~

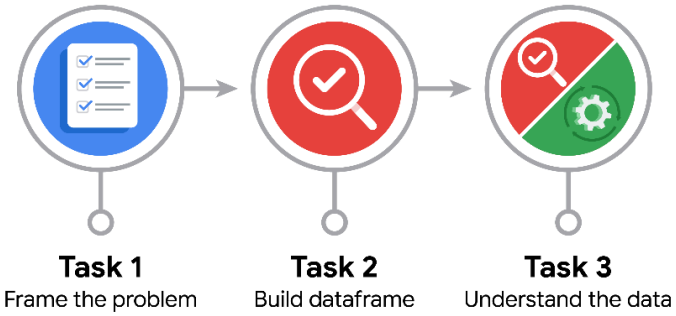
Relevant Interview Questions

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?

  To best prepare for understanding and organizing the provided taxi cab information, I read the emails with the provided information about what I have to do to accomplish the tasks and also reviewed the data dictionary, which contains information about the variables in the dataset.

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

  To perform this task, follow-along examples of data structures, conditional statements, loops, and strings will help me.

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

  Before starting to code, a resourceful learner would review past activities, research similar case studies, organize and manage their time, and focus on the provided data analysis task.

PACE: Analyze Stage

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

The sufficiency of available information for achieving the goal depends on specific details. More information about the goal and dataset is needed for a precise assessment.

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

I would use df.describe() function of pandas.

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

Some data look unusual, for example the ones related to the total_amount and trip_distance variables.

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?

I would recommend further investigation to determine if any null data is present but in a format different from NaN. Additionally, explore instances where values appear to be impossible, such as negative `total_amount`.

- What additional types of data could strengthen this dataset?

Geospatial coordinates, time-related details, weather, traffic, demographics, special events, payment processing time, and promotional data could enhance the dataset, offering a more comprehensive analysis of taxi ride patterns and user behavior.