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~~

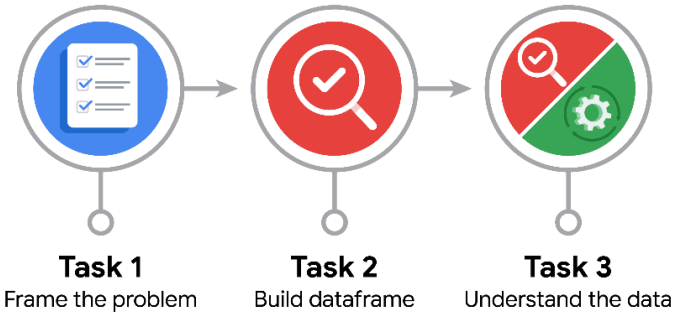
Relevant Interview Questions

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?

Best Preparation practice is to approach the information by referring to the data dictionary by reading the fields in it. Reviewing the fact sheet could provide helpful insights about the background of the data.

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

Python data science handbook, Data analysis with python and pandas by wes mckinney are few of the helpers

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

To name some- Previewing data, looking for missing values, identifying data types would be helpful along with defining goals and outlining the steps.

PACE: Analyze Stage

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

While it's unlikely that the available information will be enough, we require data visualization, exploration, feature selection, domain knowledge to name a few.

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

Describe method of pandas will be useful while summarizing the model and this methods generates various summary statistics of the numerical columns such as min and max range

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

Yes, Average trip measure is disrupted by a potential outlier as one trip is way higher compared to others and also average fare is affected by one unusual fare.

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?

What is the business model trying to achieve specifically? If there's any known domain specific insights or constraints? How the data was collected, is there any potential bias involved? the target variable the model is trying to predict.

- What data initially presents as containing anomalies?

Trip fare, Trip distance and Trip tip seems to contain anomalies

- What additional types of data could strengthen this dataset?

Data such as the amount of time long rides have been chosen, or any specific insights to those unusual data.