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~~

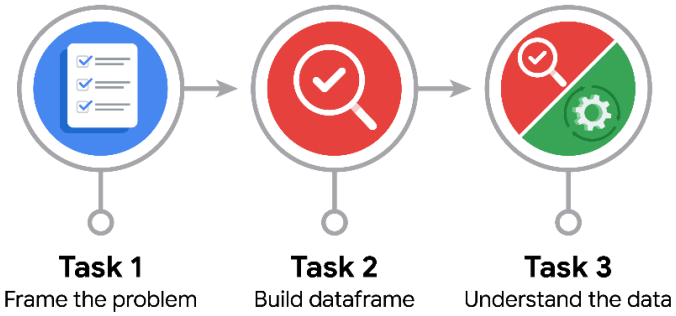
Relevant Interview Questions

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?

Our goal must be conspicuously defined. If our vision is clear then handling the task at hand will be a lot easier.

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

I'd take a through look at my previous notes.

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

Having python.org/ Stackoverflow opened on another device may come in handy.

PACE: Analyze Stage

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

Yes, definitely.

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

I'd first go through the the dataframe to check for any null values, once I'm done cleaning the data then I'd use the python skills to cut down the data frames into relevant and meaningful stacks. Using the .info() or describe() or head() functions are quite helpful, however as they are not going the help me right away so I have used nested queries to make find the unusual values.

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

Well the values given to described payment method were floats so when I averaged them, I got float values which clearly is pointless, for fares there were 4 values that were negative which made next to no sense.the intervals didn't have a pattern as there were no distinct relationship between the total fare and trip distance.

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 ask him to investigate the negative fare amounts as well as the 10 unusual fare amounts that were charged so that while building the regression model these unusual values don't cause a massive failure.

- What data initially presents as containing anomalies?

The fare amounts and the trip distances.

- What additional types of data could strengthen this dataset?

A trip_distance: total_fare_amount ratio that is relatively constant for all passengers.