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
- [x] ~~Answer the questions in the Jupyter notebook project file~~
- [x] ~~Complete coding prep work on project's Jupyter notebook~~
- [ ] Summarize the column Dtypes
- [ ] Communicate important findings in the form of an executive summary

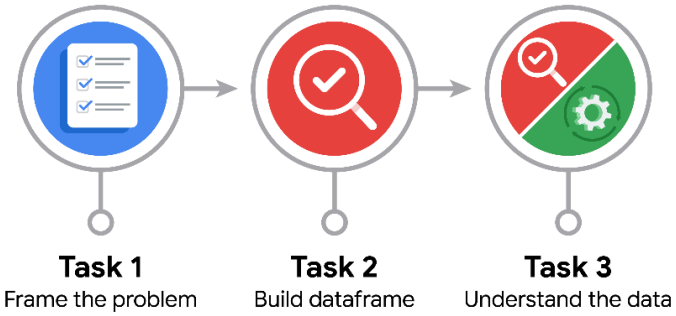
Relevant Interview Questions

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?

First, I will go through the provided data dictionary to gain a general idea of the dataset and the information provided by each variable.

After reading the data to the Jupyter Notebook environment, I intend to check the shape of the data (rows, columns), the metadata and the descriptive statistics of the numeric fields.

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

The course's annotated follow-along guide with be used to support my work. Additionally, I will be checking Python documentation to get information on more methods and attributes that have not been covered during the course.

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

Gain a thorough understanding of the project activities and deliverables.

PACE: Analyze Stage

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

Yes. Assuming that the dataset covers a representative sample of the user base.

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

The describe() function from the pandas library can be used to return summary statistics of a dataframe, including values such as min, max, mean, quartiles etc.

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

None of the averages of the data variables look unusual.

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 dataset also showed that the churned users drove a greater distance for each day that they drove, and also showed a higher drives per drive day. It would be worth exploring whether the users who have churned are 'superdrivers' who require different needs from a typical driver, and whether their reason for churning is due to the app not meeting their work needs.

- What data initially presents as containing anomalies?

total_navigations_fav1, total_navigations_fav2, sessions because of the gap between the median and mean values

- What additional types of data could strengthen this dataset?

User feedback, particularly from churned users would be enhance our understanding on what features either retain or churn a user.

More information about the users' activity such as why they use the app (for a commute, as a part of their occupation)