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~~

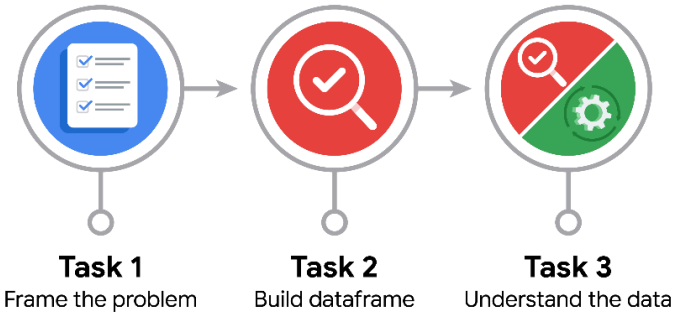
Relevant Interview Questions

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?

I can start by reviewing any materials or guidelines that was provided to me in this course and also from external sources like python.org or stackoverflow etc.

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

A few examples are like I mentioned earlier would be websites like github, stackoverflow, pythonschool, codecamp.org etc.

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

I can start by : 1.Research and Exploration 2.Reading and Documentation 3.Reviewing Examples 4.Problem Analysis 5.Planning, 6. Setting goals and finally. 7.Seeking Feedback

PACE: Analyze Stage

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

Based on intuition and analysis of the variables, it's likely that the available information, combined with additional research and problem-solving skills, will be sufficient to achieve the goal. However, the degree of sufficiency may vary depending on factors such as the complexity of the task, my prior knowledge(intermideate) and experience, and the resources available for further exploration.

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

This is how I will begin :

```
import pandas as pd

# Assuming df is my DataFrame

# Create summary statistics

summary_stats = df.describe()

# Assess the minimum and maximum range of the data

min_range = df.min()

max_range = df.max()
```

And the rest is the way I'd like it to show the data

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

There appear to be more of Android users than iPhone users; with about 65% of Android users and 35% of iOS users.

Users who churned averaged ~3 more drives in the last month than retained users, but retained users used the app on over twice as many days as churned users in the same time period. The median churned user drove ~200 more kilometers and 2.5 more hours during the last month than the median retained user.

Perhaps this is suggestive of a user profile.

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 before performing exploratory data analysis?
- **We recommend gathering more data on the super-drivers**. The reason they're driving so much may be also the reason why the Waze app does not meet their specific set of needs, which may differ from the typical driver.
- **The immediate next step is to conduct thorough EDA and develop data visualizations** to illustrate the narrative behind the data and guide future project decisions.

- What data initially presents as containing anomalies?

There were many null values in almost all the rows and columns, apart from that everything else was solid.

- What additional types of data could strengthen this dataset?

The strength of a dataset depends on its relevance to the problem at hand. Additional types of data that could strengthen a dataset **might include demographic information, geographic data, temporal data, sentiment analysis, or even data from related fields** that could provide context or additional insights. It ultimately depends on the specific goals and requirements of the analysis or project.