Reference Guide
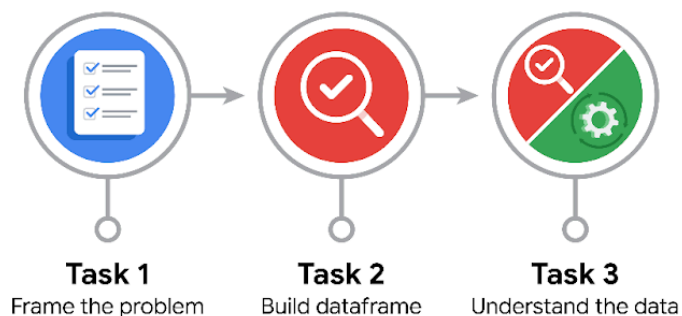
This project has three tasks; the visual below identifies how the stages of PACE are incorporated across those tasks.

Task 1
Frame the problem

Task 2
Build dataframe

Task 3
Understand the data

Data Project Questions & Considerations

PACE: Plan Stage

● How can you best prepare to understand and organize the provided information?

> I believe that the best approach is to have a plan with the data. I understand that I need to import and inspect the data, make sure that the data is clean and provide a quick EDA.

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

> I believe the materials that I have collected from my time in the course will help me with this project.

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

> I think that there are plenty of practice opportunities with the material provided, but also going the extra mile by looking at videos, and other resources online.

PACE: Analyze Stage

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

> I believe that the informations will be sufficient enough to achieve the goal of completing a quick EDA on the data.

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

> I will use the pandas library to analyze the data using masks, groupings, and aggregate functions.

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

> Because there are some outliers in the data, it is best to stay away from looking at mean data, instead, I used median data to analyze the data. I believe that by doing this, it eliminates the possibility of utilizing skewed data.

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 believe that is important to investigate whether the missing data is from a non-random sample. If it is, we might consider options to leave that sample data included, or eliminate it.

- What data initially presents as containing anomalies?

> The `label` column of the data had 700 rows of missing data.

- What additional types of data could strengthen this dataset?

> I believe that another column with data about the last time a user logged into the app would be useful, using it as a timestamp of their last activity. This can help us understand how long these breaks away from the app may be. Additionally, I would think it would be beneficial to see where exactly these people are traveling, if not city, as least given state names. Maybe there are issues inside of a particular city with navigation mishaps, making people feel obligated to choose another navigation service.