Course Two

Get Started with Python

Instructions

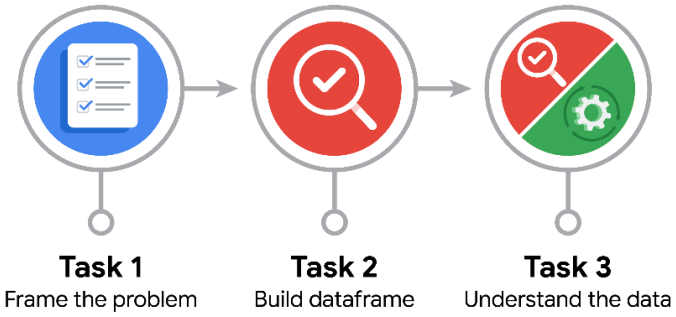
Use this PACE strategy document to record decisions and reflections as you work through this end-of-course project. You can use this document as a guide to consider your responses and reflections at different stages of the data analytical process. Additionally, the PACE strategy documents can be used as a resource when working on future projects.

Course Project Recap

- 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

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?

  By planning the steps ahead, then using python to get a summary of the given dataset and to check any outliers, null, duplicated values.

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

  Following the EDA 6 practices: Clean, Validate, Present, Visualize, Merge, Explore.

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

  Knowing the source of the dataset, knowing the estimated number of observations in it.
  Writing a PACE document to help planning and tracking the project.

PACE: Analyze Stage

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

  First, we need to perform some EDA in python and to get some insight, afterward we can conclude if we need some more information to merge it to our dataset in order to make more accurate prediction or not.

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

  Python have different useful packages that we can use to perform this type of tasks. First using pandas we will read our data and store it in a dataframe, then we can use the .describe() function that will return a descriptive summary for us.

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

  Some columns need to be checked to know what information they hold to analyze it right.

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?

  Is the churn of users occur because of some errors in the app itself? Is there specific month in the year most of the users churn in?

- What data initially presents as containing anomalies?

  NON

- What additional types of data could strengthen this dataset?

  Location can give an overview of most places people churn, because navigation apps not much needed in smaller places or cities.