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

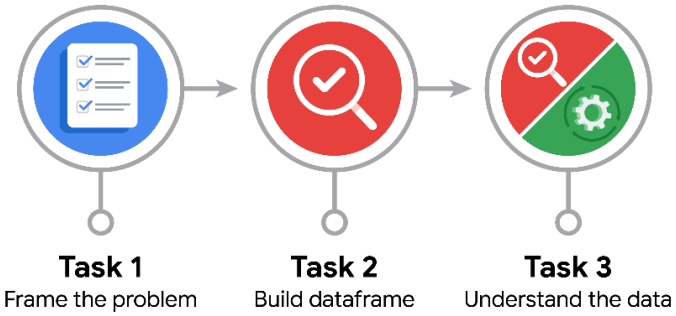
Relevant Interview Questions

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?

Read and understand the scenario and the project goals and look at what tools or escources will be required.

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

Jupyter notebooks

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

Review documentation on the language ,the project proposal , goals, and the scenario.

PACE: Analyze Stage

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

Yes ,the techniques I've learned about python and the tools and data at hand is enough to achieve the goal.

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

Using the .describe function/method of the pandas library.

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

The averages of some of the data variables are very low compared to their max values indicating outliers. They have very high standard deviation and maximum values that are very high compared to their quartile values.

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?

Engagement Level is highly correlated with claim status.This should be a focus for further inquiry.

- What data initially presents as containing anomalies?

The data for video duration and engagement metrics initially present anomalies due to extreme values and skewed distributions.

- What additional types of data could strengthen this dataset?

Including demographic data and engagement rates could further strengthen the dataset. I've already added few engagement rates like likes per view ,shares per view and comments per view.