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~~

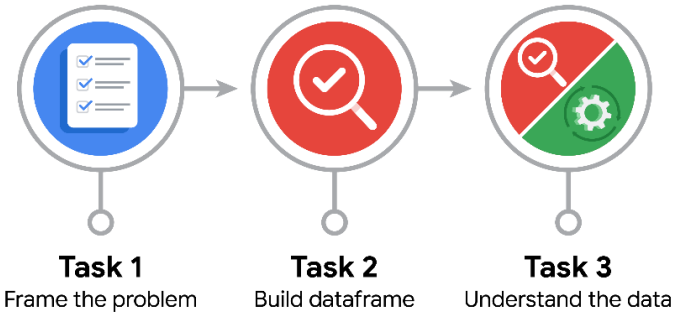
Relevant Interview Questions

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?

To understand and organize the provided information: I need to load the data, viewing some portion of it, doing some EDA for better understanding.

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

I have used the code example given by earlier in the course.

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

I think the best practices for learning code are:
1. Read documentation.
2. Try to write code.
3. Got error and try to solve them from internet.

PACE: Analyze Stage

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

I think the given information is good enough to create a model that can specify the claim and opinion videos. As the dataset has almost 12 columns, some of the variables have a much closer correlation to the outcome label.

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

I used the pandas build function .describe() to get the statistics of each numerical column.

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

The average of counts variable is something unusual as there are a large difference between the mean and median with a large number of std.

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?

  We can further study the relationship between the claim status and banned user type.

- What data initially presents as containing anomalies?

  The count's columns may have some anomalies as the difference between the mean and median is large and also the standard deviation is huge.

- What additional types of data could strengthen this dataset?

  If we add the number of reports or skip in the dataset for each video, I think it will be helpful to make better decisions.