Course Three

Go Beyond the Numbers: Translate Data into Insights

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 3 PACE strategy document
- ☐ Answer the questions in the Jupyter notebook project file
- ☐ Clean your data, perform exploratory data analysis (EDA)
- ☐ Create data visualizations
- ☐ Create an executive summary to share your results

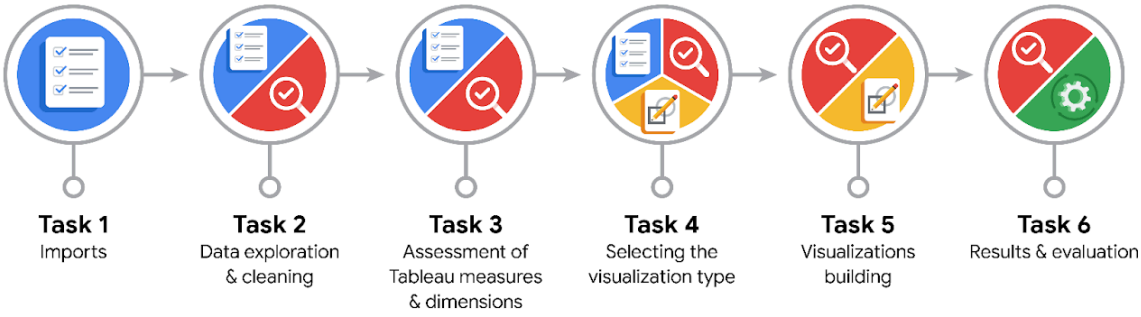
Relevant Interview Questions

Completing the end-of-course project will help you respond to these types of questions that are often asked during the interview process:

- How would you explain the difference between qualitative and quantitative data sources?
- Describe the difference between structured and unstructured data.
- Why is it important to do exploratory data analysis?
- How would you perform EDA on a given dataset?
- How do you create or alter a visualization based on different audiences?
- How do you avoid bias and ensure accessibility in a data visualization?
- How does data visualization inform your EDA?

Reference Guide

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



Data Project Questions & Considerations

PACE: Plan Stage

- What are the data columns and variables and which ones are most relevant to your deliverable?

Claim status,author ban status ,transcript and other engagement columns. The most relevant to the deliverable are claim status ,author ban status and engagement columns.

- What units are your variables in?

The variables are in ‘per video’ units.

- What are your initial presumptions about the data that can inform your EDA, knowing you will need to confirm or deny with your future findings?

There seems to be a correlation between engagement rates and claim status as well as author ban status.

- Is there any missing or incomplete data?

There seem to be missing for claim status and engagement rate columns for some videos

- Are all pieces of this dataset in the same format?

If each individual data point(row) is in the same format? partially true except the missing data for some columns.

- Which EDA practices will be required to begin this project?

discovering,structuring,validating,cleaning,presenting

PACE: Analyze Stage

- What steps need to be taken to perform EDA in the most effective way to achieve the project goal?

Discovering,structuring,cleaning,presenting with validating in between each

- Do you need to add more data using the EDA practice of joining? What type of structuring needs to be done to this dataset, such as filtering, sorting, etc.?

Currently no. As for filtering ,sorting can be done on a need to need basis such as for different engagement rates,author ban status and claim status.

- What initial assumptions do you have about the types of visualizations that might best be suited for the intended audience?

Pie charts,histograms,bar plots,box plots and scatter plots.

PACE: Construct Stage

- What data visualizations, machine learning algorithms, or other data outputs will need to be built in order to complete the project goals?

Pie charts,histograms,bar plots,box plots and scatter plots.I don't have much idea about the machine learning algorithms yet.

- What processes need to be performed in order to build the necessary data visualizations?

Structuring,cleaning and validating of the data and then use of tools like Tableau ,matplotlib and seaborn.

- Which variables are most applicable for the visualizations in this data project?

Claim status is applicable the msot.

- Going back to the Plan stage, how do you plan to deal with the missing data (if any)?

Ask the origin source of this data about it before making my final decision as to represent,categorize or delete it.

PACE: Execute Stage

- What key insights emerged from your EDA and visualizations(s)?

I analyzed the correlations between variables,particularly between claim status and others and recognized the presence of null values and skewed data.

- What business and/or organizational recommendations do you propose based on the visualization(s) built?

To account for the null values and skewed data.

- Given what you know about the data and the visualizations you were using, what other questions could you research for the team?

I could research about the distinctive characteristics that apply only to claims or only to opinions.

- How might you share these visualizations with different audiences?

By preparing and delivering a presentation.