

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

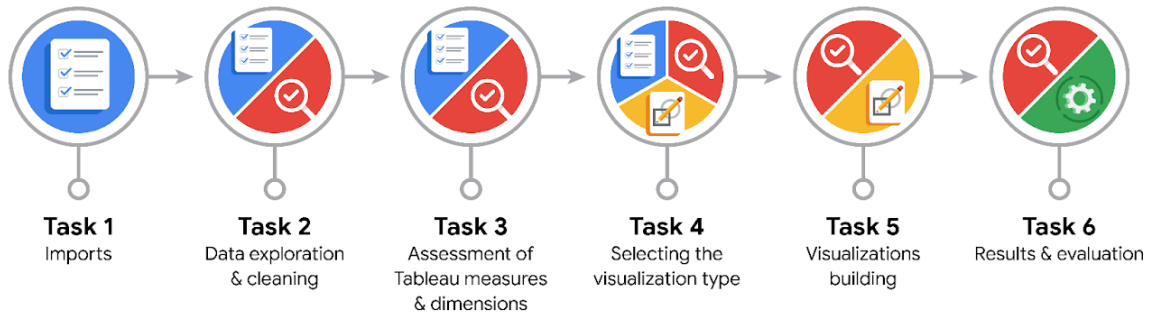
Completing the end-of-course project will help you respond these types of questions that are often asked during the interview process:

- 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?

video duration, video like count, video comment count, video view count, author ban status.

- What units are your variables in?

Int, float, obj.

- 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?

Find outliers of variables that will be deliverable.



- Is there any missing or incomplete data?

Below 1000 row or nothing.

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

There are Int, float, obj.

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

Discovering, Validating, Structuring, Cleaning, Presenting.



PACE: Analyze Stage

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

Validating.

- 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.?

Count and sorting.

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

Boxplot and histogram.



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?

Boxplot and histogram. Try most simple model first such as Linear Regression.

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

Cleaning first performed data visualizations.

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

video duration, video like count, video comment count, video view count, author ban status.

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

Start by discovering, using `.head()`, `.size`, and `.shape`.



PACE: Execute Stage

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

Understand the data.

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

Determine the variables that have the largest impact.

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

Understand the outliers help prepare for cleaning and modeling.

- How might you share these visualizations with different audiences?

Share all visualizations and explain what happening.