

PACE Strategy Document

Introduction

This PACE strategy document can be used 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.

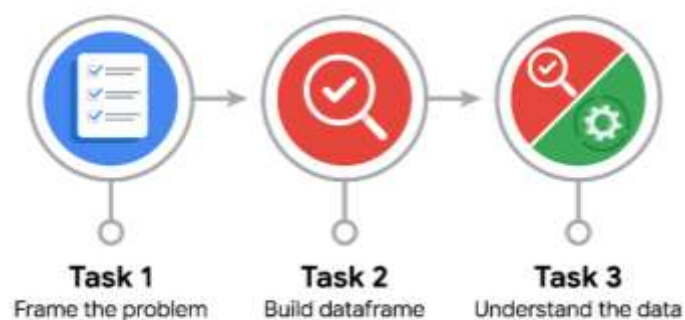
Relevant Interview Questions

Completing the end-of-course project will help respond to 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 it thoroughly and underline important phrases/tasks and start working on them.

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

Jupyter notebook embedded in code labs. But in my case notebook downloaded from Kaggle because I am auditing this course so don't have access to labs.

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

Make notes of the process and save templates so as to possess a roadmap for future projects.



PACE: Analyze Stage

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

Yes it would be sufficient.

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

Summary can be built based on output from `info()` and `describe()`. Max and min also appear in `describe()`.

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

Not really.



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?

Some of the data needs error correction. So, data cleaning needs to be performed before EDA.

- What data initially presents as containing anomalies?

The columns distance and total_amount contain anomalies.

- What additional types of data could strengthen this dataset?

Holiday and traffic data could help create an even more precise model for fares before ride.