

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

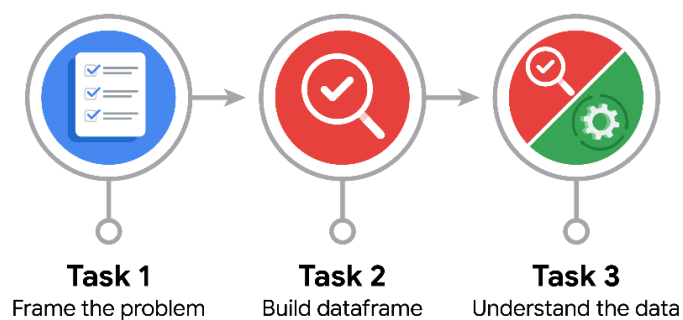
Relevant Interview Questions

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?

I would confirm the data sourcing and see how many different locations the information would be spread over.

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

Jupyter Notebooks will be crucial for this task.

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

Create an overall workflow and scope document to limit how much and what data will be included in the project's focus. This report must have approval from top management of the project.



PACE: Analyze Stage

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

I believe the data will provide enough information to draw insights into where to take the project in future stages.

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

I would use pandas documentation to confirm methods for info/describe summaries on the data as well as the min and max methods available.

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

The averages and the interval data are both standard with nothing unusual standing out.



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?

I would recommend looking into geographic locations of users to see if there is a connection between location and churning rates. I would also see if there is a bias towards Android users churning more than iPhone users.

- What data initially presents as containing anomalies?

The category for 'Labels on Churning', 'Retained' and nulls.

- What additional types of data could strengthen this dataset?

'Location', 'Population of city residence.'