Course One
Foundations of Data Science

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 PACE Strategy Document to plan your project while considering your audience members, teammates, key milestones, and overall project goal.
- ☐ Create a project proposal for the data team.

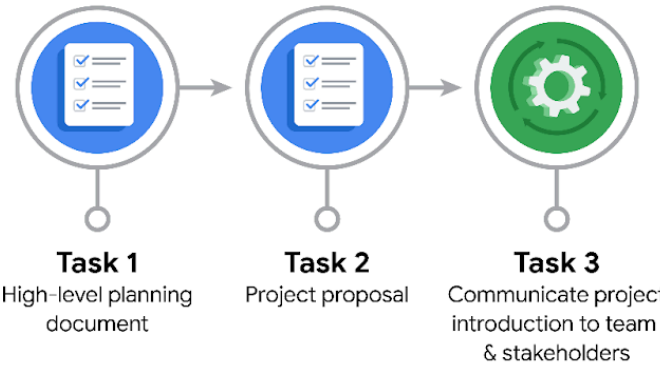
Relevant Interview Questions

Completing this end-of-course project will empower you to respond to the following interview topics:

- As a new member of a data analytics team, what steps could you take to get 'up to speed' with a current project? What steps would you take? Who would you like to meet with?
- How would you plan an analytics project?
- What steps would you take to translate a business question to an analytical solution?
- Why is actively managing data an important part of a data analytics team's responsibilities?
- What are some considerations you might need to be mindful of when reporting results?

Reference Guide

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



Data Project Questions & Considerations

PACE: Plan Stage

- Who is your audience for this project?

New York Taxi and Limousine Commission is the audience for this project.

- What are you trying to solve or accomplish? And, what do you anticipate the impact of this work will be on the larger needs of the client?

We are trying to build a model that estimates taxi and limo fares for riders before their ride.

- What questions need to be asked or answered?

What information could be relevant? What variables are more useful? What do the trends throughout the data mean? How can we reduce bias in analysis? Etc.

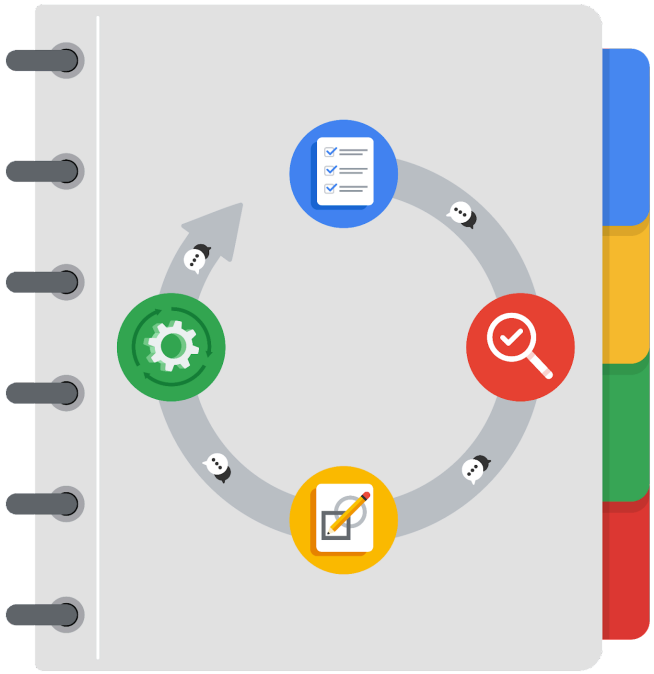
- What resources are required to complete this project?

Python notebook, the dataset, and feedback from stakeholders.

- What are the deliverables that will need to be created over the course of this project?

The prepared dataset that is ready for the analyses, EDA visualizations, statistical testing model, and the final regression model.

THE PACE WORKFLOW



[Alt-text: The PACE Workflow with the four stages in a circle: plan, analyze, construct, and execute.]

You have been asked to demonstrate for the company's data team how you would use the PACE workflow to organize and classify tasks for the upcoming project. Select a PACE stage from the dropdown buttons. A few tasks involve more than one stage of the PACE workflow. Additionally, not every workplace scenario will require every task. Refer back to the Course 1 end-of-course portfolio project overview reading if you need more information about the tasks within the project.

Project tasks

Following are a group of tasks your company's data team has determined need to be completed within this project. The data analysis manager has asked you to organize these tasks in preparation for the project proposal document. First, identify which stage of the PACE workflow each task would best fit under using the drop down menu. Next, give an explanation of why you selected the stage for each task. Review the following readings to help guide your selections and explanation: The PACE stages and Communicate objectives with a project proposal. You will later reorder these tasks within a project proposal.

1. **Evaluating the model:** Execute

   Why did you select this stage for this task?

Evaluating the final model comes after the construction of the final model. We execute this model to evaluate how accurate it is.

2. **Conduct hypothesis testing:** Analyze **and** Construct

   Why did you select these stages for this task?

Hypothesis testing is idealized in the analyze stage and executed in the construct stage.

3. **Begin exploring the data:** Analyze

   Why did you select this stage for this task?

After I plan out everything, collect the data, and scrub the data to prepare it for exploration, I will explore the data. However, this comes before the construction of any model, so it's in the analyze stage.

4. **Data exploration and cleaning:** Analyze **and** Plan

   Why did you select these stages for this task?

Initial exploration would be a part of the plan stage. Cleaning would be a part of the analyze stage.

5. **Establish structure for project workflow (PACE):** Plan

Why did you select this stage for this task?

Projects need to be developed in a planned and structured manner. During the planning stage, the structure of the project flow would be established. This needs to be done before doing anything else.

6. **Communicate final insights with stakeholders:** Execute

Why did you select this stage for this task?

This is in the execute stage because final insights would have already been generated and finalized and are ready to share with stakeholders in this stage.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

Descriptive statistics are done through exploratory analysis and need to be done before constructing the exploratory visualizations and the model.

8. **Visualization building:** Analyze **and** Construct

Why did you select these stages for this task?

The analysis leads to visualizations. Planning out how the visualizations would look like would be in the analysis stage, and actually building it would be in the construction stage.

9. **Write a project proposal:** Plan

Why did you select this stage for this task?

The project proposal is the initial document to list out the milestones and deliverables in each stage of the project.

10. Build a regression model: Analyze and Construct

Why did you select this stage for this task?

The model examination would be in the analysis stage, and the actual building will be in the construction stage.

11. Compile summary information about the data: Analyze

Why did you select this stage for this task?

Summary information is analyzed during exploratory analysis, before model building.

12. Build machine learning model: Construct

Why did you select this stage for this task?

Machine learning models are built in the construction stage, after regression model building.