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.

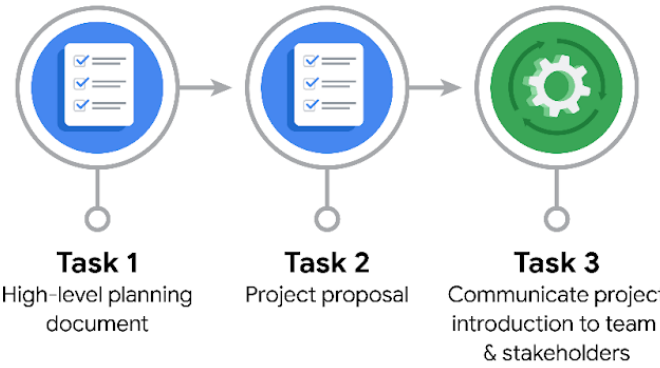
Relevant Interview Questions

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 City Taxi and Limousine Commission (TLC), an agency responsible for licensing and regulating New York City's taxi cabs and for-hire vehicles.

- 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 develop a regression model that helps estimate taxi fares before a ride. I believe this project may also give TLC a much deeper understanding of their customers to help them make more informed decisions in the future.

- What questions need to be asked or answered?

How was the data collected? What variables in the data will be the most useful? Are there any trends in the data that would be of interest to us? How can we reduce the impact of bias?

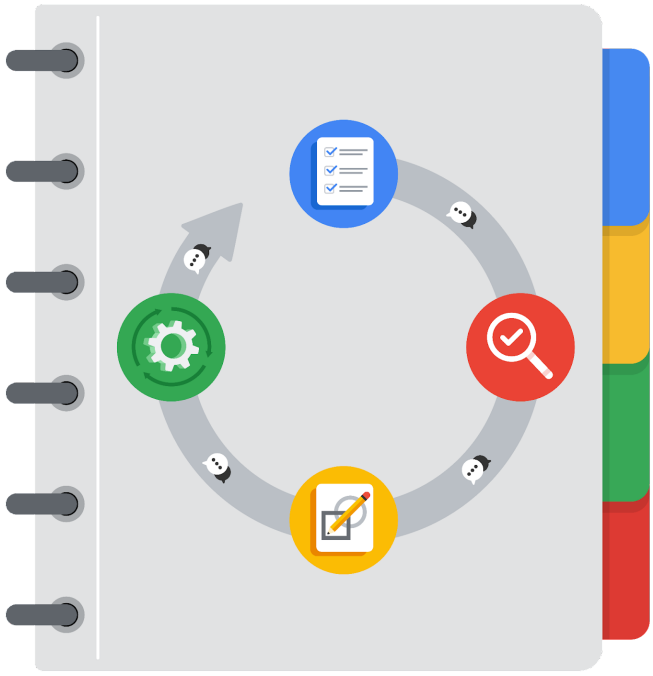
- What resources are required to complete this project?

The dataset, python, jupyter notebook and communication and input from all stakeholders.

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

The regression model, visualisations of EDA and a cleaned dataset.

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?

After the model has been executed it needs to be evaluated for accuracy.

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

   Why did you select these stages for this task?

During the analysis stage, the statistical test to be used will be decided. In the construct phase, the test will be carried out.

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

   Why did you select this stage for this task?

In this stage the data will be explored to develop a deeper understanding of the data and any trends it reveals.

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

   Why did you select these stages for this task?

The planning stage is for making the first choices on what to do with the data. The analysis stage allows us to explore trends and clean data for construction.

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

Why did you select this stage for this task?

So we can lay out a pathway for how we will complete the project.

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

Why did you select this stage for this task?

After constructing and finalizing the model we can then present to all stakeholders.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

This task involves finding statistics that can help us understand the data.

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

Why did you select these stages for this task?

Visualization will allow the analysts and stakeholders to understand the data and will help with the construction of the model.

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

Why did you select this stage for this task?

This is the start of the project.

10. **Build a regression model:** Analyze **and** Construct

Why did you select this stage for this task?

The analysis stage will help construct the model as the most important variables will be determined. The building of the regression model will take place in the construction phase.

11. **Compile summary information about the data:** Analyze

Why did you select this stage for this task?

Inspecting the data will happen in this stage.

12. **Build machine learning model:** Construct

Why did you select this stage for this task?

This is when the model will be constructed.