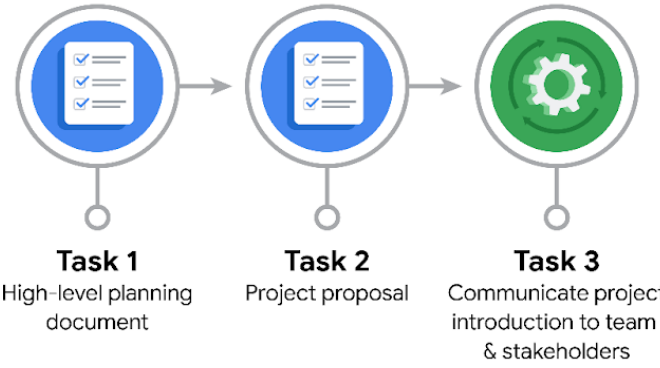
Course One
Foundations of Data Science

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.

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?

The audience for this project is New York City Taxi and Limousine Commission (TLC).

- 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?

The goal is to build a regression model that will predict the taxi fares before the ride. By developing a regression model for predicting taxi fares, New York City TLC can adopt a more robust pricing strategy. This strategy will enable them to compete effectively with their competitors in the market. Additionally, it is expected that accurately predicting fares will contribute to increasing customer satisfaction, as customers will have transparency and confidence in the pricing. This, in turn, is likely to drive higher ridership and ultimately increase the client's revenue.

- What questions need to be asked or answered?

The questions that need to be asked are: Is the data reliable and of good quality? Are there any biases in the data collected? If so, how can this be reduced? What variables are the most useful? Are there any trends in the data that uncover some insights?

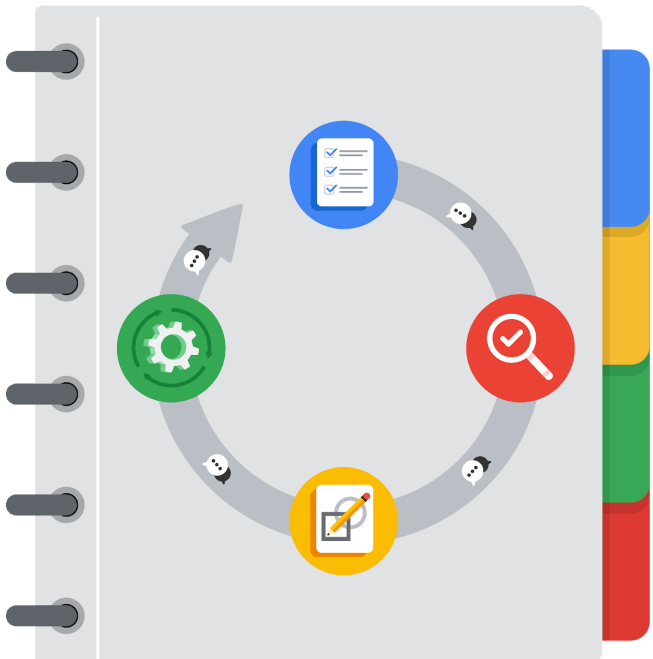
- What resources are required to complete this project?

Access to the internet, data availability, Jupyter NoteBook and feedback from the stakeholders.

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

Build a regression model that will predict the taxi fare before the ride, test accuracy of the model, retest the model after incorporating feedback from the stakeholders, present the findings through visualizations and an executive summary showcasing key insights, recommendations and next steps.

THE PACE WORKFLOW



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?

This is the stage where you determine if the constructed model aligns with the project's requirements and objectives.

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

   Why did you select these stages for this task?

At the Analyze stage, you identify the need for a statistical test, and in the construction stage, you perform the test.

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

   Why did you select this stage for this task?

This is the stage where you begin with data exploration and get a better understanding of the underlying data.

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

   Why did you select these stages for this task?

In the Planning stage, you decide about the different methods that are going to be used and in the Analyze phase, the exploratory analysis and cleaning are carried out.

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

Why did you select this stage for this task?

In the Planning stage, one of the key activities is establishing the milestones that are required to complete the project, thereby giving the project a structure and a project workflow.

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

Why did you select this stage for this task?

The final stage is the execution stage, where you share the insights with the stakeholders.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

It is in the Analyze stage where numerical measures and techniques are used to get insights about the central tendencies, variation and distribution.

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

Why did you select these stages for this task?

In the Analyze phase of the project, exploratory data analysis is done to get a better understanding of the data and in the Construction phase, the understanding gained from the exploratory data analysis is utilized to build meaningful visualizations.

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

Why did you select this stage for this task?

The Planning stage is where you conceptualize the scope of the project and develop a project proposal.

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

Why did you select this stage for this task?

In the Construct phase of the project, the modeling approach is selected and built. In the Execute phase, the findings are shared with the stakeholders. If necessary, adjustments or improvements may be made to the model based on the feedback.

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

Why did you select this stage for this task?

Compiling summary information about the data takes place in the Analyze phase.

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

Why did you select this stage for this task?

It is in the Construction phase where the building, interpreting and revising of the models takes place.