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.~~

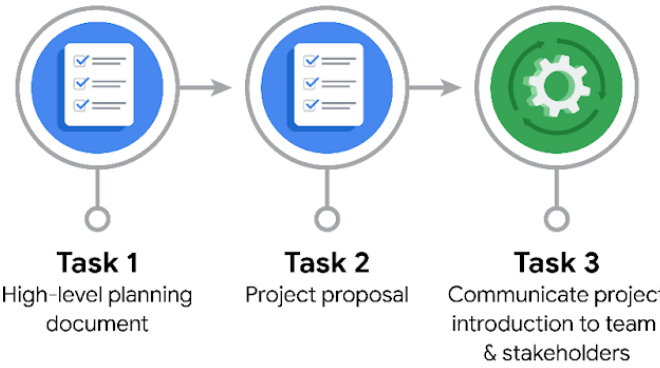
Relevant Interview Questions

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)

- 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 primary goal of this project is to develop a reliable and accurate taxi fare prediction model based on NYC taxi trip data. This model will be used by Automatidata to provide real-time fare estimates for taxi rides, which will benefit both riders and drivers.

- What questions need to be asked or answered?

1. Is the TLC trip data accessible and updated regularly?
2. How clean and consistent is the data?
3. What additional features can be added to improve the model?
4. How well does the model perform on unseen data?

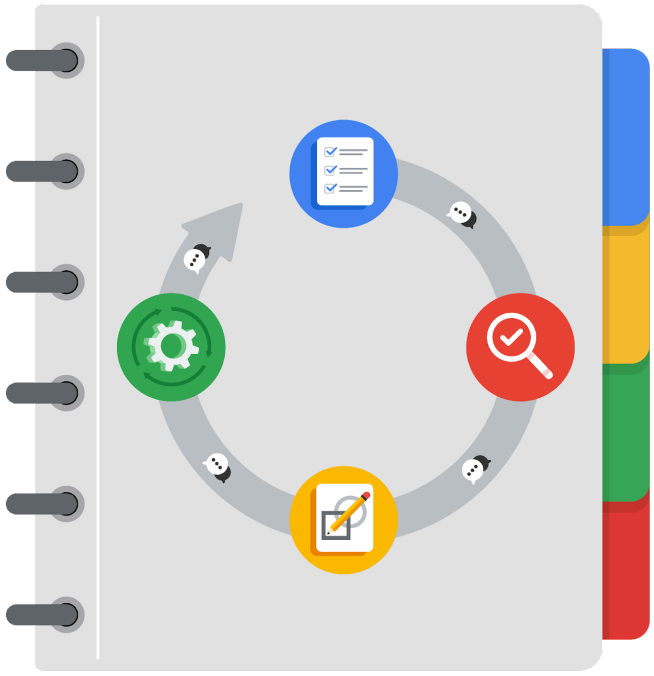
- What resources are required to complete this project?

• TLC trip data • Data Cleaning and Processing tools (pandas, NumPy) • Machine learning libraries

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

1. Project Report 2. Model Training Report 3. Performance Report

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:** Plan

   Why did you select this stage for this task?

The task of "Evaluating the model" was assigned to the Plan stage because evaluating the model is a critical step in the overall project lifecycle. It helps assess the model's performance, identify any potential issues, and refine the model before deployment.

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

   Why did you select these stages for this task?

Plan: To evaluate the model's performance and identify any potential issues before deployment. Analyze: To understand the project's context and requirements, identify hypotheses to be tested, and ensure that test results are interpreted correctly.

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

   Why did you select this stage for this task?

EDA is crucial for understanding the data, identifying patterns, trends, and relationships between variables.

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

Why did you select these stages for this task?

Plan: To ensure data is clean and ready for modeling. Analyze: To gain insights into the data and inform the modeling process.

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

   Why did you select this stage for this task?

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

   Why did you select this stage for this task?

To define the project's workflow and ensure all team members are on the same page.

7. **Compute descriptive statistics:** Construct

   Why did you select this stage for this task?

To summarize data and identify patterns.

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

   Why did you select these stages for this task?

Analyze: to understand the data and identify relevant patterns, Construct: to create visualizations that effectively communicate insights.

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

   Why did you select this stage for this task?

Final Part of the project

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

    Why did you select this stage for this task?

Construct: To build the regression model using the chosen algorithm and hyperparameters. Execute: To deploy the model into production and make it accessible for real-time usage.

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

    Why did you select this stage for this task?

To provide a concise and understandable overview of the data to users.

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

    Why did you select this stage for this task?

To build the machine learning model using the chosen algorithm, hyperparameters, and features.