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.

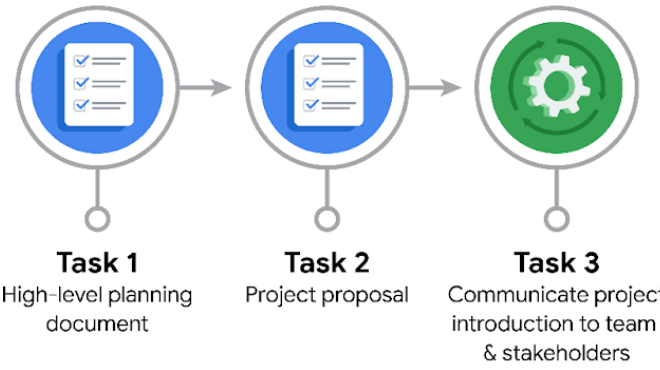
Relevant Interview Questions

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?

The audience for this project includes both internal team members at Automatidata and external stakeholders at the 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?

I am trying to build a regression model to predict taxi fares before each ride based on distance, time of day and other variables. The project will enhance customer satisfaction through accurate fare predictions, improve TLC's operational efficiency, and enable data-driven decision-making for better service and policy improvements.

- What questions need to be asked or answered?

What is the condition of the provided dataset? What variables will be the most useful? Are there trends within the data that can provide insight? What steps can I take to reduce the impact of bias?

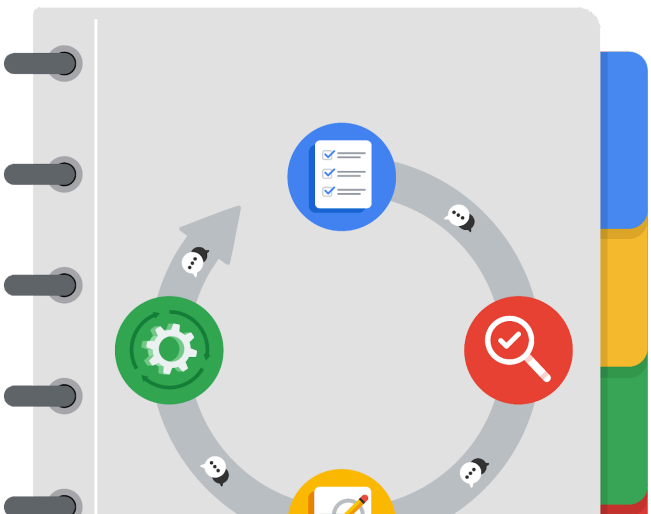
- What resources are required to complete this project?

1. Data that TLC has gathered. 2. Computational resources 3. Software tools 4. Whole team of data analysts, project manager and other technical support staff

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

1. Course 1 PACE Strategy Document to plan your project while considering your audience members, teammates, key milestones, and overall project goal. 2. Create a project proposal for the data team. 3. Dataset scrubbed for exploratory data analysis, visualizations, statistical model, regression analysis and/or machine learning 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 model falls under the Execute stage because it involves testing the final product to ensure it meets the project's requirements and performs accurately.

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

   Why did you select these stages for this task?

Hypothesis testing involves both analyzing the data to form hypotheses (Analyze) and constructing tests to validate these hypotheses (Construct).

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

   Why did you select this stage for this task?

Exploring the data is an initial step in understanding the dataset, identifying patterns, and forming initial insights, which is a key part of the Analyze stage.

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

   Why did you select these stages for this task?

Data exploration involves analyzing the data to identify issues, while data cleaning involves constructing methods to address these issues and prepare the dataset for further analysis.

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

   Why did you select this stage for this task?

Establishing the workflow structure is a planning activity, where tasks are organized, and milestones are set to guide the project.

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

   Why did you select this stage for this task?

Communicating final insights is part of the Execute stage, where the findings and results are shared with stakeholders after the analysis and model building are complete.

7. **Compute descriptive statistics:** Analyze

   Why did you select this stage for this task?

Computing descriptive statistics is part of the Analyze stage as it involves summarizing and describing the main features of the dataset.

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

   Why did you select these stages for this task?

   Visualization building involves analyzing the data to determine what to visualize (Analyze) and then constructing the visualizations (Construct) to effectively communicate insights.

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

   Why did you select this stage for this task?

Writing a project proposal is a planning activity that outlines the project goals, milestones, and tasks, setting the stage for the project's execution.

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

    Why did you select this stage for this task?

    Building a regression model involves analyzing the data to understand relationships (Analyze) and constructing the model to predict outcomes (Construct).

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

    Why did you select this stage for this task?

Compiling summary information involves analyzing the data to extract and summarize key points, providing an overview of the dataset.

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

    Why did you select this stage for this task?

Building a machine learning model is a constructive activity where algorithms are developed and trained to make predictions or classifications based on the data.