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.~~

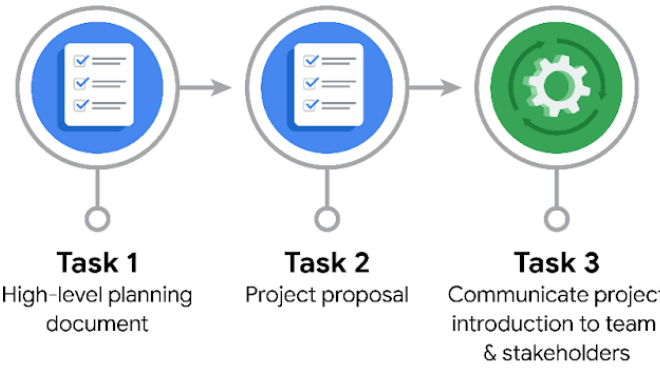
Relevant Interview Questions

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?

This project engages two main groups of stakeholders, each with distinct communication needs. The first group consists of the technical team at Automatidata including data analysts, data scientists, and project managers who require detailed, technical, and data-driven information to effectively develop, validate, and optimize the regression model.

The second group includes non-technical representatives from the New York City Taxi and Limousine Commission (TLC), such as Juliana Soto (Head of Finance and Administration) and Titus Nelson (Operations Manager). For this audience, communication must be simplified and focused on practical outcomes, operational impacts, and strategic decision-making.

Therefore, all project communications will be tailored accordingly: technical and data-rich for the Automatidata team, and strategic and visual for TLC stakeholders.

- 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 main objective of this project is to develop a regression model capable of predicting taxi fares in New York City before rides take place, using historical data provided by the Taxi and Limousine Commission (TLC). This solution aims to enhance transparency and predictability for both passengers and drivers, while streamlining operational and financial processes for the agency.

The expected outcomes include improved pricing efficiency, a reduction in disputes over fare amounts, and enhanced strategic planning for TLC such as fare policy adjustments and more effective resource allocation. For Automatidata, the project's success will further solidify its reputation as a consultancy that transforms data into practical solutions, paving the way for future partnerships and market opportunities.

- What questions need to be asked or answered?

- What is the condition of the given dataset?
- Which variables will be most useful?
- Are there trends in the data that can provide insights?
- What steps can I take to reduce the impact of bias?

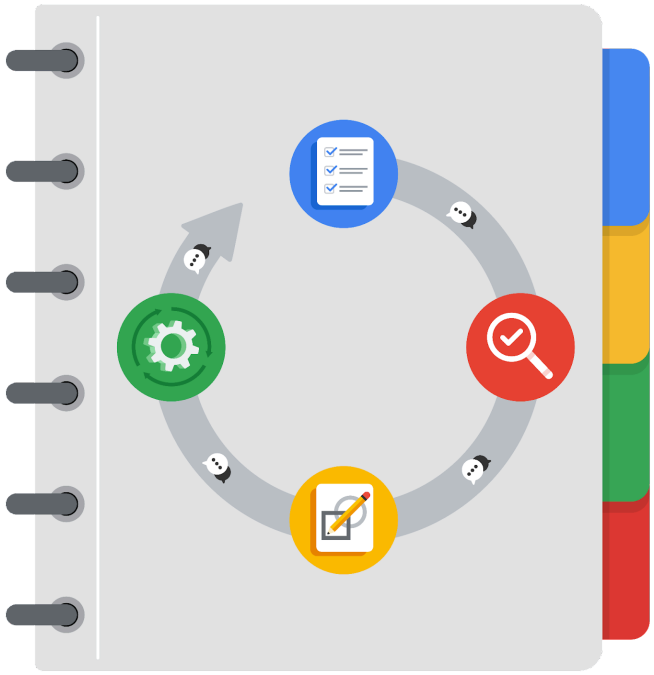
- What resources are required to complete this project?

- Project dataset
- Python
- Stakeholder input

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

Cleaned dataset 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?

   Once the model has been built, the data is processed to assess whether the outcomes align with the project's defined goals and expectations.

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

   Why did you select these stages for this task?

   These stages were selected because, during the analysis phase, the need for a statistical test is identified. In the construction phase, the test is then carried out accordingly.

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

   Why did you select this stage for this task?

   This stage was selected because, during the analysis phase, one can gain a deeper understanding of the dataset and the information it contains.

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

   Why did you select these stages for this task?

   These stages were selected because, during the planning phase, initial decisions are made about the required methods. The data cleaning process then takes place in the analysis phase.

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

Why did you select this stage for this task?

The planning phase was chosen for this task because creating the initial project PACE document helps structure the workflow and define the best approach to take.

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

Why did you select this stage for this task?

The execution phase was selected for this task because it is when the final key insights are communicated to stakeholders, making it a crucial moment to align expectations and consolidate the project's outcomes.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

The analysis phase was chosen for this task because it is during this stage that the statistical investigation of the data takes place, enabling the identification of meaningful patterns and trends.

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

Why did you select these stages for this task?

The analysis and construction phases were selected for this task because visualization begins with data assessment and is built during the construction stage.

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

Why did you select this stage for this task?

The planning phase was chosen for this task because the project proposal is the initial document that defines its objectives, scope, and guidelines, serving as the foundation for the entire development.

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

Why did you select this stage for this task?

During the analyzing stage, the model is examined in detail to be sure it will meet the needs of the task. 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?

The analysis phase was selected for this task because it involves inspecting the dataset and compiling summary information, allowing for a more structured understanding of the content.

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

Why did you select this stage for this task?

The construction phase was selected for this task because it is during this stage that the data model is developed a key step for consolidating the project's analytical structure.