

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

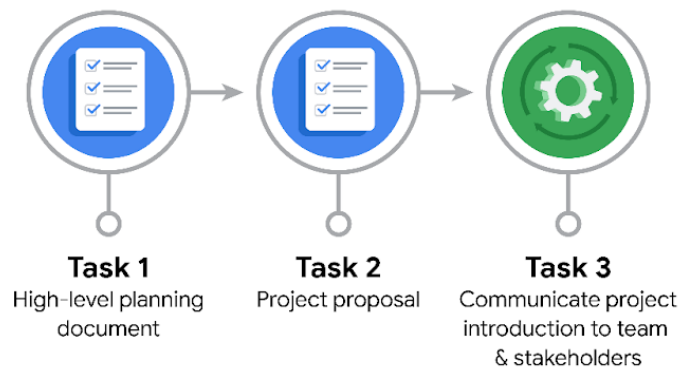
Completing this end-of-course project will empower you to respond to the following interview topics:

- 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) is the audience for this project.

- 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 working on the development of a predictive model that enables TLC riders to estimate the taxi fares in advance of their ride.

- What questions need to be asked or answered?

What are the **key variables** and how are they defined?

Is the data **complete** or are there missing values?

Are there any **biases** in how the data was collected?



What **patterns or trends** emerge from the data?

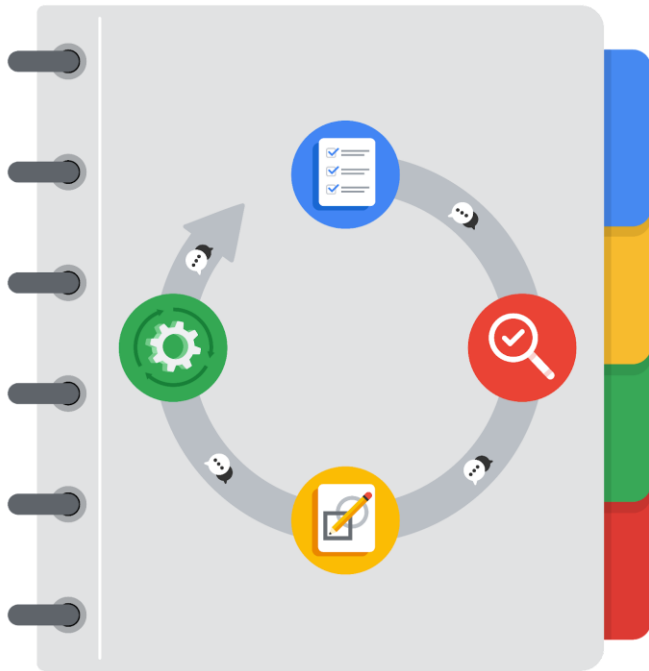
- What resources are required to complete this project?

We need the project dataset, Python notebook, and input from stakeholders.

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

Project deliverables consist of a cleaned dataset prepared for exploratory data analysis, accompanying visualizations, and the development of a statistical model, regression analysis, 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?



After building the model, it must be evaluated to ensure it aligns with the project's goals and requirements.

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

Why did you select these stages for this task?

Analyze because hypothesis testing requires statistical examination of data, and **Construct** because the results must be turned into conclusions, insights, or models that connect back to the project's objectives.

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

Why did you select this stage for this task?

Analyze stage was selected because exploring is about **examining and interpreting** it to build understanding, which is the essence of analysis.

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

Why did you select these stages for this task?

Data exploration in the plan phase helps to define the requirements and anticipate the challenges. Data cleaning is performed during the analysis phase to gain a clearer and more detailed understanding of the dataset.

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

Why did you select this stage for this task?

Establishing structure for the project workflow during the **plan phase** ensures clarity and organization.



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

Why did you select this stage for this task?

Communication with stakeholders happens during all phases however sharing the final insights happens during the final 'Execute phase'.

7. **Compute descriptive statistics:** **Analyze**

Why did you select this stage for this task?

Analysing the statistics of the data happens during Analyze phase.

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

Why did you select these stages for this task?

Visualization building fits into both the **Analyze** and **Construct** stages. In the **Analyze phase**, visualizations help uncover patterns, trends, and relationships in the data, while in the **Construct phase**, they are developed into clear, purposeful charts or dashboards that communicate insights effectively to stakeholders.

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

Why did you select this stage for this task?

The project proposal represents the first stage of the process and is required before any other phase begins.

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

Why did you select this stage for this task?



Building a regression model belongs to the Construct stage because we are creating the actual predictive model. During analysis, the model is examined to confirm it meets the project's needs.

11. Compile summary information about the data: Analyze

Why did you select this stage for this task?

Compiling the summary information is to get a clearer understanding of the structure of the data before building the model which would happen in the Analyze phase.

12. Build machine learning model: Construct

Why did you select this stage for this task?

Building the machine learning model/development would occur in the Construct phase.