

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 New York City Taxi and Limousine Commission

- 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?

a regression model that predicts taxi cab fares before each ride according to the related variables.

- What questions need to be asked or answered?

How can I use the data I receive most effectively?
Which variables may be important for the regression model?
What precautions can I take to avoid encountering an error?
How can I build a more effective system based on the data I have?

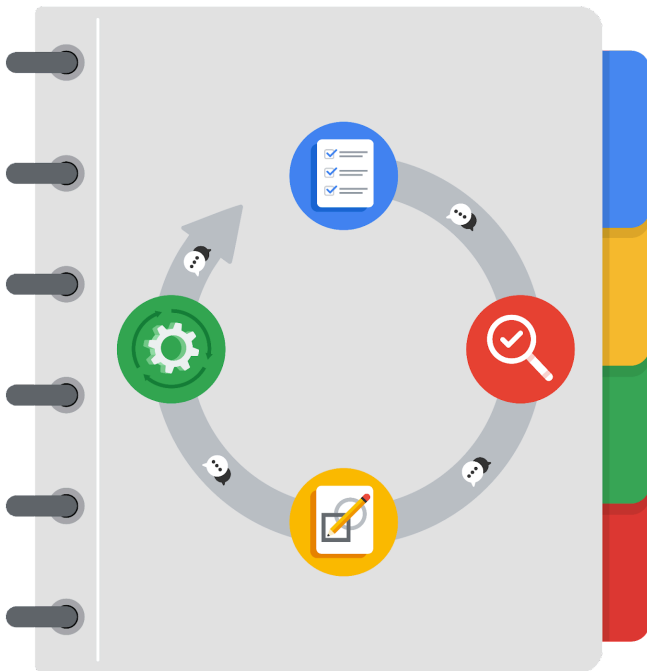
- What resources are required to complete this project?

A computer with Python installed and a csv file containing the data

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

Final version of regression model and test results that show how accurate the regression 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?

It is the execute phase because it is related to confirming the accuracy of the model after the construction phase.

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

Why did you select these stages for this task?

Construction and analyze phases were chosen because hypothesis testing is related to the phases to determine the test type to be used and to conduct the test.

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

Why did you select this stage for this task?

Exploring data is done to gain insight into what path to follow. Therefore, the analysis phase was chosen.

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

Why did you select these stages for this task?

While explore the data, we need to choose how to use the dataset (plan) and to convert the data into usable format (analyze).



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

Why did you select this stage for this task?

Before analyze phase, we need to develop a workflow. It is a part of plan phase.

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

Why did you select this stage for this task?

When we finish construction phase, we have final version of the regression model and we are ready to share our insights with stakeholders

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

Why did you select this stage for this task?

Statistics are done to determine the correct method before starting the construction phase. It is related to the analysis phase.

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

Why did you select these stages for this task?

While visualization is done in the analysis phase to choose the right method, visualization is done in the Construction phase to understand the results obtained from the study of the model.

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

Why did you select this stage for this task?

Before we do any analyze, we have to write a proposal in plan phase.



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

Why did you select this stage for this task?

While the regression model is built during the analysis phase to ensure that the correct method is applied, it is built to reach its final version during the construction phase.

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

Why did you select this stage for this task?

It is a part of analyze phase.

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

Why did you select this stage for this task?

Since model construction will begin after all analyzes are completed, this is done during the Construction phase.