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.

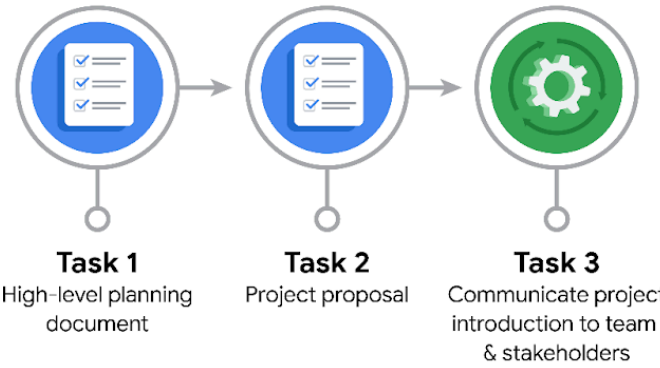
Relevant Interview Questions

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 Limo Commission (TLC) executives.

- 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?

What: to build a regression model that predicts taxi cab fares before each ride based on distance, time of day, and any additional variables we find necessary.

- What questions need to be asked or answered?

What TLC's dataset consists of.
What information the TLC data provides through exploratory data analysis (EDA).
Which model will deliver consistent results.

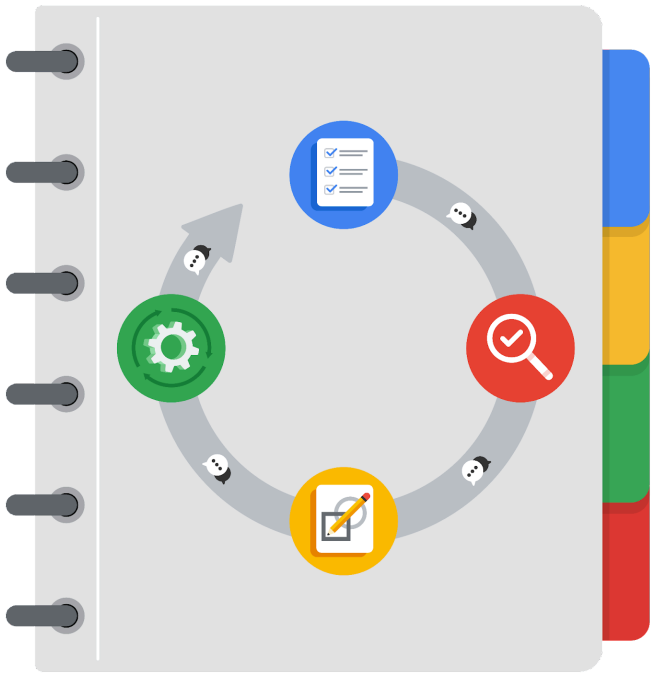
- What resources are required to complete this project?

Access to TLC data. Python programming language. Data wrangling. Exploratory Data Analysis (EDA). Regression models ML. Data Visualization. A/B Testing knowledge.

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

EDA information (Data team). Metrics and methodology for choosing a model (Data team). Model accuracy and score (Data Team). Visual presentation presenting results (Stakeholders). Written report presenting results (Stakeholders).

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?

The model is built, interpreted, and revised in the Construct stage. However, it is during the Execute stage that we put our analysis and construction into action

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

   Why did you select these stages for this task?

Because in the Analyze state, we prepare the data and perform EDA, whereas in the Construct stage we use the data to make an informed decision. We can test some hypotheses during the EDA (Analyze stage), and also while interpreting and revising the models from the Construct stage.

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

   Why did you select this stage for this task?

The analyzing stage is where one collects, prepares, and analyzes all of the data for the project to gain a deeper understanding of the dataset and the information within it.

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

   Why did you select these stages for this task?

   During the Plan stage one identifies the information needed and researches if the business data is congruent with the required data.

   The Analysis phase is where we interact with the data, by formatting, cleaning and converting it into usable formats.

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

   Why did you select this stage for this task?

   Because during the Plan phase we establish the foundations for succeeding in the project.

   This can be achieved by structuring the project workflow using PACE.

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

   Why did you select this stage for this task?

   During the Execute phase we put in practice our analysis and constructions into action, meaning, delivering our findings.

   Final insights are shared with stakeholders in this stage.

7. **Compute descriptive statistics:** Analyze

   Why did you select this stage for this task?

   Descriptive statistics are brief informational coefficients that summarize a given data set. This can be understood as EDA, where we are searching for information on the data that we have acquired from the business. EDA is usually performed in the Analyze stage.

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

Why did you select these stages for this task?

During EDA (part of the Analyze stage), we can make use of several graphs to understand the relationship between variables. When building and evaluating the models (Construct Stage), we also need to build other visualizations to present the findings.

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

Why did you select this stage for this task?

The project proposal helps to identify the information needed to understand the scope of the project. This is prepared during the Plan stage.

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

Why did you select this stage for this task?

Part of building a regression model is related to preparing the data, that's why we can select the Analyze stage as a required stage. During the Construct stage we will be building, interpreting, and revising models.

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

Why did you select this stage for this task?

Since the analyzing stage is where we engage in exploratory data analysis or EDA, compiling information about the data can be achieved in this stage, since the information on how data will influence the models can be gathered during this stage.

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

Why did you select this stage for this task?

The Construct stage is where we build, interpret, and revise models. Some projects will require machine learning algorithms to uncover correlations within the data.