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.

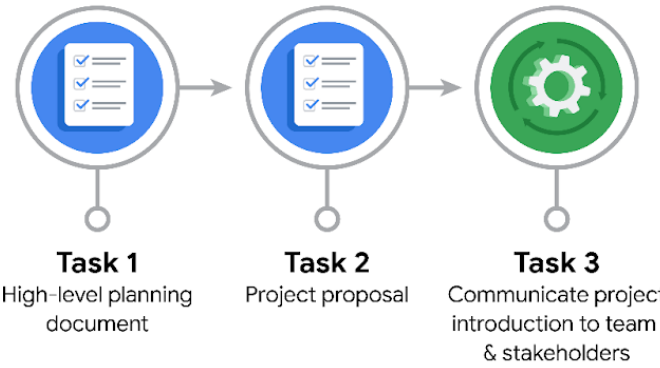
Relevant Interview Questions

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 of this project is the New York City Taxi and Limousine Commission (TLC) company.

- 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 trying to develop a regression model to predict the ride duration based on parameters we believe are important. This work will help the client make informed decisions about how they distribute their taxis and limousines in New York City.

- What questions need to be asked or answered?

What type of data is the New York City TLC providing us?
Would the models we currently use properly/accurately describe the data given to us?
Is there any bias in the data provided to us?

What are some important parameters that are directly related to the length of a ride?

- What resources are required to complete this project?

We will be using python to create and apply the regression model to our data

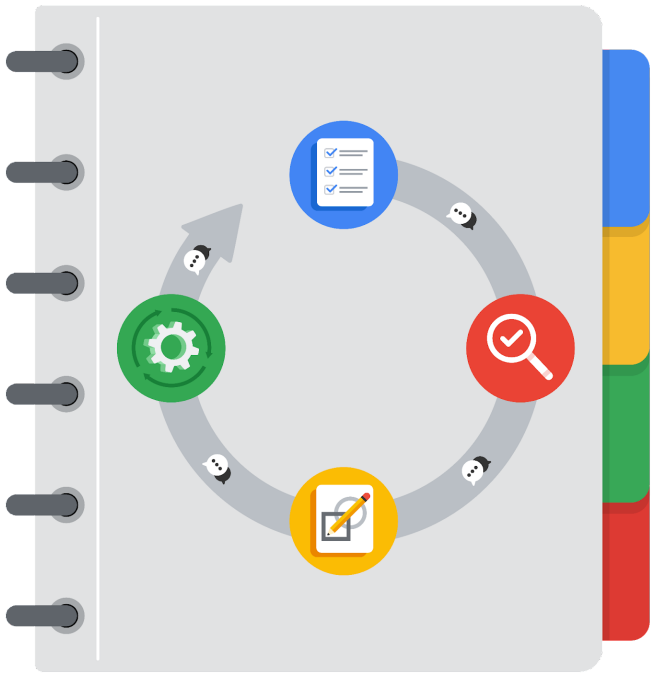
We will need a team of data professionals to work on this project

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

Exploratory data analysis (EDA) needs to be broken down to multiple parties in order to gather an unbiased understanding of the information within the data.

Drafting the regression model should also be considered a deliverable because we are yet to know what models can be applied to our data and if they are truly representative of what we are observing.

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:** Construct

   Why did you select this stage for this task?

During the stage of construction the team emphasizes their efforts in developing and evaluating models that will be used to study the data.

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

   Why did you select these stages for this task?

Hypothesis is an essential part of examining data which is the principal component of the analyze stage of the PACE workflow. Secondly during the construction phase you check if the model adapts well to the data by conducting hypothesis testing.

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

   Why did you select this stage for this task?

During the analysis phase we are tasked to start exploring the data, and perform reduction routines to clean and structure the data in a more readable form.

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

   Why did you select these stages for this task?

   The analyze stage focuses on working on the raw data, and investigating its trends. During the planning phase you can also outline how this data exploration can be carried out such that it is done in the most efficient way possible.

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

   Why did you select this stage for this task?

   During the planning stage is when you set up the foundations of the project such as the structure of the project workflow.

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

   Why did you select this stage for this task?

   During the execution stage data professionals present a compelling story of what the data is trying to say to the decision maker of the company.

7. **Compute descriptive statistics:** Analyze

   Why did you select this stage for this task?

   Computing descriptive statistics of a data set is a form of characterization of data which is part of the analysis stage.

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

   Why did you select these stages for this task?

In order to better understand data we construct clear visualizations which makes this action part of the analysis stage. Additionally, visualization techniques help showcase the adaptability of a machine learning model which is part of the construction stage.

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

Why did you select this stage for this task?

Writing a project proposal is a form of setting the aims and goals of a project which are part of the planning stage.

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

Why did you select this stage for this task?

Regression analysis is pivotal for both statistical analysis and machine learning models thus making this action part of both construct and analyze stages.

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

Why did you select this stage for this task?

During the execute stage we need to summarize the results of the data in an appropriate form in order to present it to the stakeholders.

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

Why did you select this stage for this task?

The construction stage focuses primarily on building machine learning model.