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.

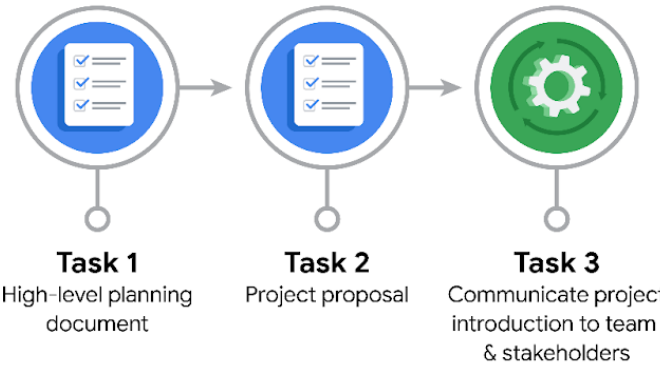
Relevant Interview Questions

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 is the data team members and cross-functional team members.

- 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?

Trying to automate the claim classification process. The impact of the project reflects both end users and the TikTok claim report unit. The end-users get a first response while the claim has been made, and the TikTok claim report unit gets fewer reports than usual, as most of them are classified automatically.

- What questions need to be asked or answered?

Q1: What is the condition of the dataset? Q2: What variables are most important? Q3: Are there trends within the data that can provide insights? Q4: What steps can be taken to reduce Bias?

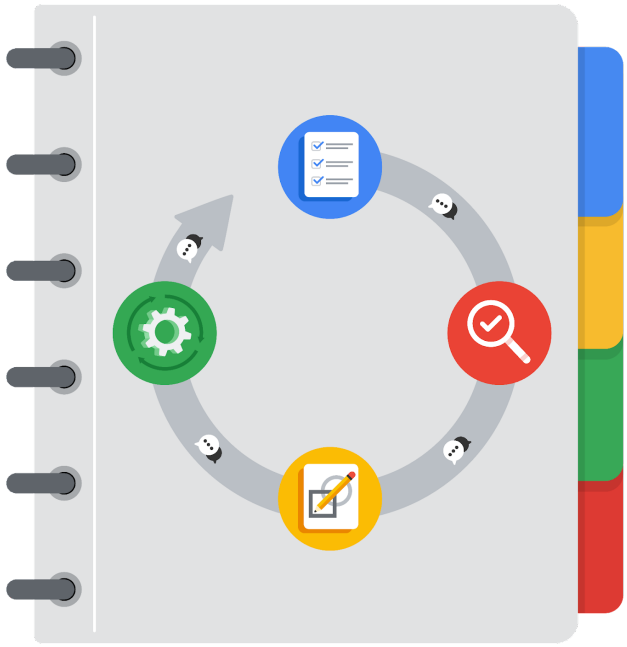
- What resources are required to complete this project?

Resources are- Dataset, Python notebook, and input from the expats and stakeholders.

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

The deliverables are:- a clean dataset, the EDA of each feature, the Statistical Significance of the features, and the Regression Analysis / 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 is 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?

Model evaluation is part of the “construct” phase.

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

   Why did you select these stages for this task?

In the “analyze” phase we need to conduct the hypothesis testing and after implementing the model in the “Construct” phase we also need to do the hypothesis testing.

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

   Why did you select this stage for this task?

In the “Analyze” step we need to do the EDA.

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

   Why did you select these stages for this task?

Planing takes place when we select the model using the data and data cleaning needs to be performed in the analysis phase.

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

   Why did you select this stage for this task?

In the “Plan” stage we structure our complete project.

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

   Why did you select this stage for this task?

In the ”Execute” phase we present the final insights to the stakeholders.

7. **Compute descriptive statistics:** Analyze

   Why did you select this stage for this task?

Computing the descriptive statistics is part of EDA, so we perform this task in the “Analyze” stage of PACE.

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

   Why did you select these stages for this task?

Visualization is the core part of the “Analyze” process. But in the “Construct” phase its also used while comparing different models.

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

   Why did you select this stage for this task?

Project prosal writing in the part of the “Plan” stage.

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

    Why did you select this stage for this task?

In the “Analyze” phase we can check if the data fits the model and in the “Construct” phase we implement the machine learning model such as regression.

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

    Why did you select this stage for this task?

To get the insights of the data we perform in the “Analyze” phase.

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

    Why did you select this stage for this task?

Building a machine learning model is the core portion of the “Construct” phase.