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.

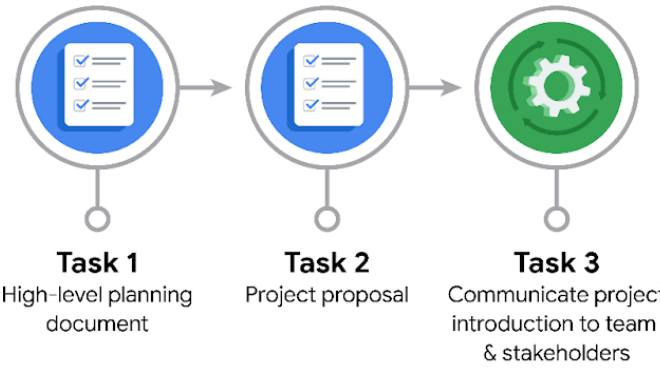
Relevant Interview Questions

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 data team, the cross-functional team members, our leadership team, and TikTok 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?

The goal is to build a machine learning model with the capability to reliably classify user interaction data as a claim or opinion.

- What questions need to be asked or answered?

Where will we obtain the data? What is the condition of the provided dataset? How will we handle data privacy and ethical considerations?What are the key characteristics of our data? Are there any patterns or anomalies that stand out? Are there trends within the data that can provide insight? What steps can we take to reduce the impact of bias? How can we visualize the data effectively?

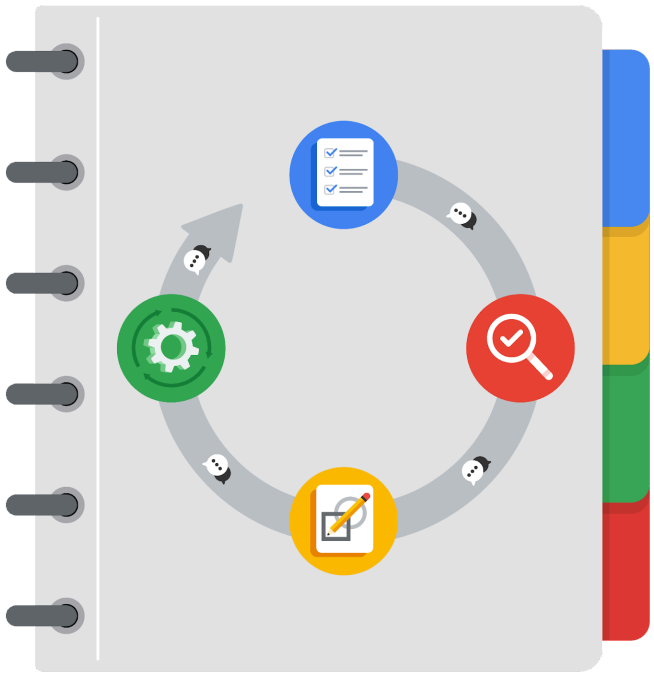
- What resources are required to complete this project?

Stakeholders input, Microsoft Teams, Project dataset, Python Jupyter notebook, and Tableau or Looker Studio.

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

Cleaned, converted, and formatted data for exploratory data analysis (EDA), statistical models for testing, regression analysis, a machine learning model, and visualizations.

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?

   In the Construct stage, we build and refine models. We build various models based on the problem and data, select the most promising models for further evaluation, evaluate model performance using appropriate metrics, and iterate on model development based on evaluation results.

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

   Why did you select these stages for this task?

   Hypothesis testing primarily falls within the Analyze phase of the PACE workflow. During the Analyze phase, we formulate hypotheses about the data based on the problem we are trying to solve and conduct hypothesis tests to determine the validity of our hypotheses. While hypothesis testing is primarily an Analyze phase activity, it is important to note that it can also be used in the Construct phase to evaluate model performance and assumptions.

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

   Why did you select this stage for this task?

   In the Analyze stage, we conduct exploratory data analysis (EDA) to identify potential relationships and patterns.

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

Why did you select these stages for this task?

Data exploration and cleaning primarily occur in the Analyze phase of the PACE workflow. Although we identify data sources in the Plan phase of the PACE workflow, we explore and prepare data for modeling. After we explore the data to understand its structure, distribution, and potential issues, we clean the data by handling missing values, outliers, inconsistencies, and errors.

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

Why did you select this stage for this task?

The Plan phase of the PACE workflow is where we establish the structure for our project. We define the problem, set goals, identify data sources, and establish the project workflow.

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

Why did you select this stage for this task?

Although communication is necessary at various points throughout a project, final insights are shared with stakeholders in the execute phase of the data project workflow.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

Descriptive statistics are primarily computed during the Analyze phase of the PACE workflow. During this phase, we compute descriptive statistics (mean, median, mode, standard deviation, etc.) to summarize data characteristics.

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

Why did you select these stages for this task?

Visualization building primarily occurs in the Analyze phase of the PACE workflow. Since this is where we explore and prepare our data for modeling, visualizations are a crucial component of this process. By creating visualizations during the Analyze phase, we can understand data distribution, explore relationships and, in the Construct phase, we communicate findings, effectively conveying insights to stakeholders.

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

Why did you select this stage for this task?

The project proposal is typically written during the Plan phase of the PACE workflow. This is when we outline the project's objectives, scope, methodology, timeline, and resource requirements.

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

Why did you select this stage for this task?

Building a regression model primarily occurs in the Construct phase of the PACE workflow. We build and refine models, including regression models. During the Construct phase, we select appropriate algorithms and train models, choosing regression techniques based on data characteristics and building regression models using the prepared data. We evaluate its performance, assessing model accuracy and reliability. Finally, we iterate and improve, refining the model based on evaluation results.

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

Why did you select this stage for this task?

Compiling summary information about the data typically occurs in the Analyze phase of the PACE workflow. This is when we delve into the data to understand its characteristics and prepare it for modeling. We gain valuable insights into the data and inform subsequent analysis steps.

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

Why did you select this stage for this task?

Building a machine learning model primarily occurs in the Construct phase of the PACE workflow. This is where we apply our data analysis insights to create predictive or prescriptive models.