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.

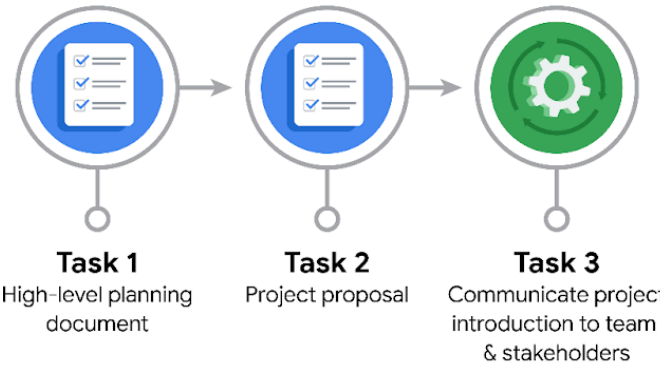
Relevant Interview Questions

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?

- Primarily the TikTok data team (Willow, Rosie, Orion) - Secondarily, cross-functional team members (Mary, Margery, Maika) and eventually 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?

Develop a machine learning model to classify user content (videos/comments) as either a "claim" or an "opinion." **Impact:** Reduce the backlog of user reports, improve efficiency of content moderation, potentially reduce the spread of misinformation.

- What questions need to be asked or answered?

What data is available for this project (e.g., user reports, video transcripts, user demographics)?
What are the ethical considerations of using this data?
What are the defining characteristics of a "claim" vs. an "opinion" in this context?
What metrics will be used to evaluate the model's performance (accuracy, precision, recall)?
What are the potential biases in the data and how can they be mitigated?

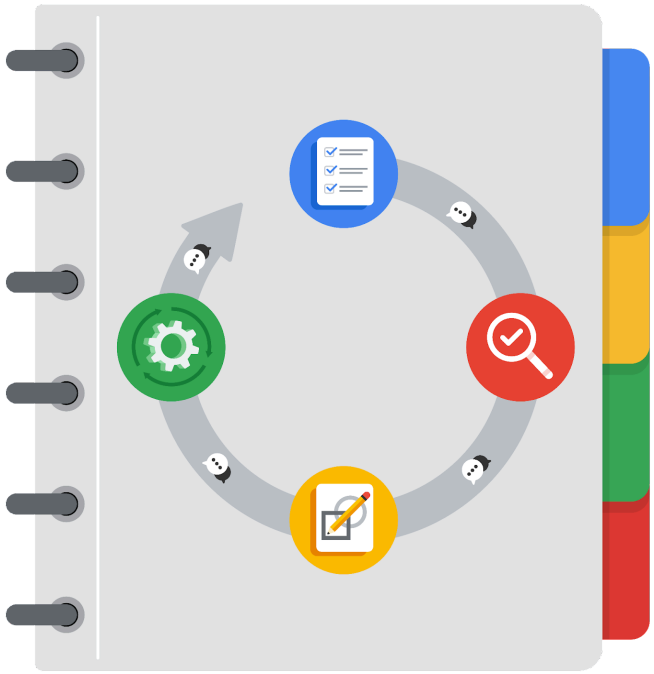
- What resources are required to complete this project?

Access to TikTok's user interaction and content data
Computing resources for model training (potentially cloud-based)
Software and libraries (Python as suggested by Rosie)
Expertise in machine learning, natural language processing, and data visualization

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

Project proposal
Data exploration and analysis reports
Model development and evaluation reports
Visualizations for stakeholders
Final presentation to leadership with actionable insights

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?

This final stage involves rigorously testing the constructed model against defined performance metrics to ensure it effectively fulfills the project's objectives. It confirms whether the model meets the established success criteria.

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

   Why did you select these stages for this task?

During **Analyze**, we determine the appropriate statistical tests needed to validate or refute our assumptions about the data. In **Construct**, we implement these tests, applying the chosen methods to the data to derive meaningful conclusions.

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

   Why did you select this stage for this task?

This foundational analysis phase involves initial data exploration to gain a comprehensive understanding of the dataset's structure, content, and potential relationships, laying the groundwork for subsequent analysis.

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

   Why did you select these stages for this task?

In **Plan**, we strategize and select the appropriate data cleaning methodologies based on initial data assessments. During **Analyze**, we execute these methods to rectify inconsistencies, handle missing values, and prepare the data for further analysis.

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

   Why did you select this stage for this task?

The **Plan** stage is crucial for establishing a structured framework using the PACE methodology. This ensures a systematic approach to the project, outlining clear milestones and task assignments for efficient execution

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

   Why did you select this stage for this task?

In the **Execute** stage, we effectively communicate the project's key findings and actionable insights to stakeholders, ensuring they understand the results and their implications for decision-making.

7. **Compute descriptive statistics:** Analyze

   Why did you select this stage for this task?

This analysis phase involves calculating summary statistics to provide a concise overview of the data's central tendencies, distributions, and variability, facilitating a deeper understanding of its characteristics.

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

   Why did you select these stages for this task?

During **Analyze**, we assess the data to determine the most effective visualization techniques. In **Construct**, we create these visualizations to visually represent data patterns, trends, and relationships, enhancing comprehension and communication.

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

Why did you select this stage for this task?

The **Plan** stage necessitates the creation of a comprehensive project proposal, outlining the project's objectives, scope, methodology, and deliverables. This document serves as the project's foundational blueprint.

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

Why did you select this stage for this task?

In **Analyze**, we evaluate the data and determine the suitability and specifications of the regression model. During **Construct**, we develop and train the model, applying the chosen algorithms to the data to establish predictive capabilities.

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

Why did you select this stage for this task?

This analysis phase involves systematically reviewing the data to extract and compile key summary information, providing a consolidated overview of the dataset's essential attributes.

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

Why did you select this stage for this task?

The **Construct** stage focuses on the development and implementation of the machine learning model, utilizing algorithms and techniques to train the model and achieve the desired predictive or classification outcomes.