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.

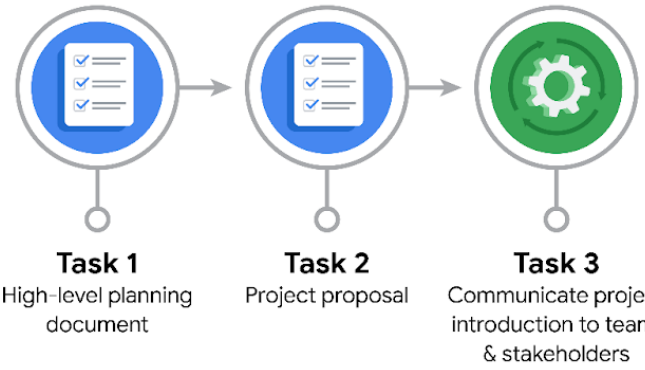
Relevant Interview Questions

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?

TikTok's Trust and Safety Team
They're responsible for keeping the platform safe, moderating content, and ensuring community guidelines are followed.

- 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?

Automatically classify user-submitted claims to assist moderators in detecting misleading or sensitive content. It will speed up moderation, reduce manual workload, and help TikTok maintain a safe and trustworthy platform.

- What questions need to be asked or answered?

What are the key indicators of harmful or misleading content?
How accurate is the classification model, and how can it be improved?

Can the system handle multilingual or evolving slang/language trends?

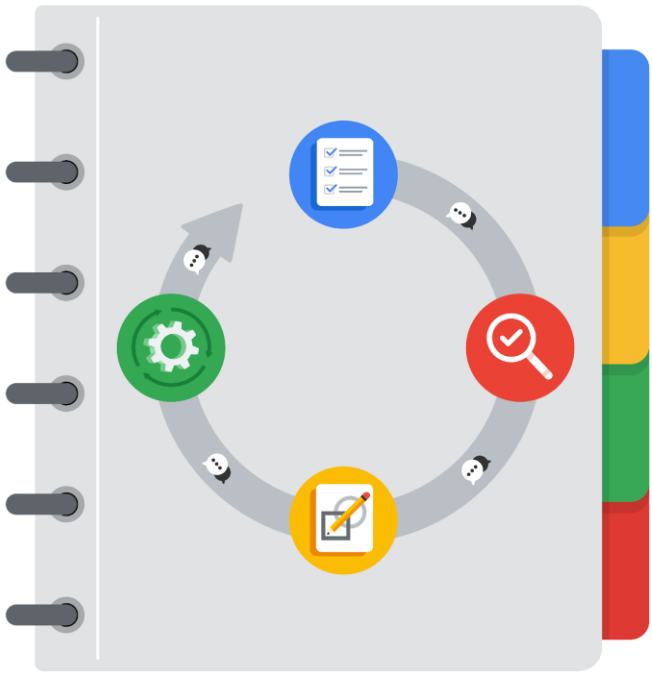
- What resources are required to complete this project?

TikTok content dataset with labels (true, false, misleading)
Python
Jupyter Notebook or Colab

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

Cleaned & Labeled Dataset – Preprocessed TikTok content with truth/misinformation labels.
EDA Report – Visualizations and insights from the dataset.
Evaluation Metrics – Accuracy, precision, recall, F1-score, confusion matrix.

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, explain 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: Select PACE stage**

   Why did you select this stage for this task?

Construct. You're **applying a machine learning model**, interpreting its results, and making data-driven decisions. The *Construct* stage focuses on executing solutions and evaluating their effectiveness, which aligns with model testing and performance evaluation.

2. **Conduct hypothesis testing: Select PACE stage and Select PACE stage**

   Why did you select these stages for this task?

Analyze: hypothesis testing involves exploring data, identifying variables, and determining relationships or differences worth testing. Construct: you design and perform the actual hypothesis tests (e.g., t-tests, chi-square) as part of building analytical outputs.

3. **Begin exploring the data: Select PACE stage**

   Why did you select this stage for this task?

Analyze and Construct I selected **Analyze** because hypothesis testing begins with examining the data and identifying patterns, distributions, and relationships that may warrant further investigation. I selected **Construct** because this stage involves designing and performing the statistical tests, interpreting results, and preparing insights based on the hypothesis outcomes.

4. **Data exploration and cleaning: Select PACE stage and Select PACE stage**

   Why did you select these stages for this task?

Analyze: Initial data exploration happens here to understand the structure, patterns, and issues in the dataset.

Construct: Data cleaning is performed in this stage to prepare the dataset for modeling and visualization, ensuring accuracy and reliability.

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

   Why did you select this stage for this task?

Why this stage?
This task involves setting up the project's roadmap, timeline, and tools—which is the core focus of the **Plan** stage. It ensures clarity, organization, and alignment before moving into analysis or modeling.

6. **Communicate final insights with stakeholders: Select PACE stage**

   Why did you select this stage for this task?

The **Execute** stage focuses on delivering outcomes. Communicating final insights to stakeholders is a key deliverable, where results are presented clearly and effectively to drive decisions based on the analysis.

7. **Compute descriptive statistics: Select PACE stage**

   Why did you select this stage for this task?

Descriptive statistics help summarize and understand the main features of the data. This aligns with the **Analyze** stage, which focuses on examining the data to uncover patterns, trends, and insights.

8. **Visualization building: Select PACE stage and Select PACE stage**

   Why did you select these stages for this task?

Construct: Visualizations are created during this stage as tools to explore data, identify patterns, and communicate findings.

Execute: Final visualizations are refined and presented to stakeholders during execution, helping drive decisions based on the analysis.

9. **Write a project proposal: Select PACE stage**

Why did you select this stage for this task?

Writing a project proposal is part of the **planning** phase, where you define project goals, scope, and strategies. It sets the foundation for analysis and development by clearly outlining the purpose and direction of the work.

10. **Build a regression model: Select PACE stage and Select PACE stage**

Why did you select this stage for this task?

Construct: Building a regression model involves designing and developing the solution based on analytical insights—key elements of the Construct stage.

Execute: After building, the model must be tested and applied to real or validation data, which falls under the implementation focus of the Execute stage.

11. **Compile summary information about the data: Select PACE stage**

Why did you select this stage for this task?

Why this stage?
I selected **Analyze** because summarizing information about the data—such as distributions, trends, and key metrics—is part of exploring and interpreting the dataset, which aligns with the analytical goals of this stage.

12. **Build machine learning model: Select PACE stage**

Why did you select this stage for this task?

I selected Construct because building a machine learning model involves designing, training, and fine-tuning algorithms—core activities of the Construct phase where solutions are developed based on analytical findings.