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.~~

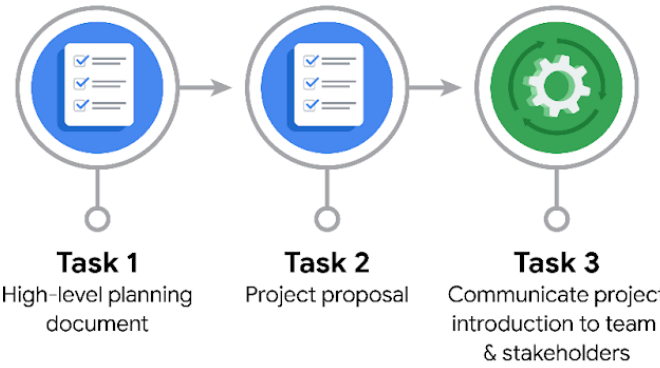
Relevant Interview Questions

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 for this project includes both the TikTok data team—who require concise, technical details on analysis, modeling, and metrics—and cross-functional managers, who need clear, high-level summaries, visuals, and impact-focused updates without technical jargon.

- 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 develop a reliable predictive model that classifies TikTok content as either a claim or an opinion, enabling the moderation team to prioritize reports more efficiently and reduce backlog. This will streamline the review process, improve response times, and help TikTok maintain a safe, engaging platform aligned with its mission to inspire creativity and bring joy.

- What questions need to be asked or answered?

  - What does the dataset contain, and are there missing or inconsistent values?
  - Which features are most useful for distinguishing claims from opinions?
  - What type of statistical testing is most appropriate for this dataset?
  - Which regression or classification model will yield the most accurate results?
  - How will model performance be measured and validated?

- What assumptions need to be checked for the chosen model?
- How should results be communicated to technical and non-technical stakeholders?

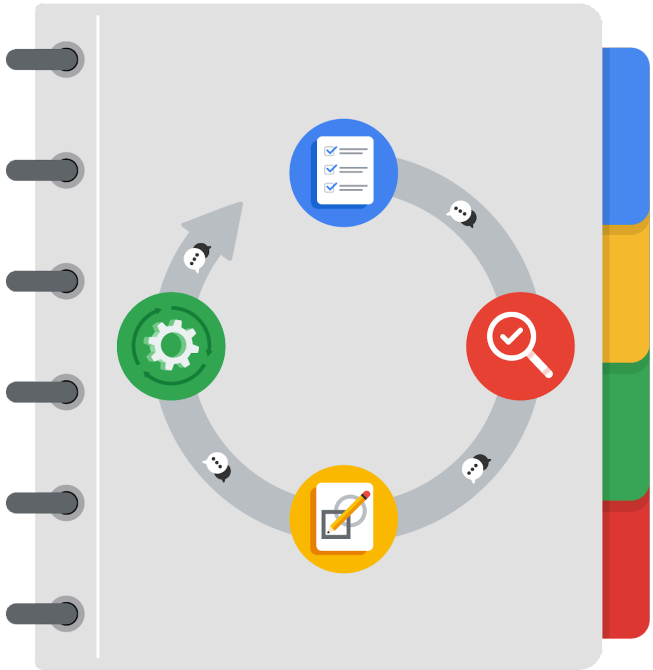
- What resources are required to complete this project?

The project will require the TikTok claims dataset, Python programming environment with relevant libraries (e.g., pandas, NumPy, scikit-learn, matplotlib, seaborn), data storage and processing tools, collaboration platforms for team communication, and access to subject matter expertise from the data science team and cross-functional managers for context and feedback.

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

Deliverables include a project proposal outlining milestones and stakeholders, a PACE strategy document, cleaned and preprocessed dataset, exploratory data analysis (EDA) report, statistical testing results, a trained and validated predictive model with performance metrics, clear visuals summarizing findings, executive-friendly talking points, and final project documentation for both technical and non-technical audiences.

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?

Model evaluation happens after the model has been built but before it is finalized and deployed. It's part of the Construct stage because this phase focuses on developing, testing, and refining the solution. Evaluating the model—by measuring performance metrics like accuracy, precision, recall, or F1 score—ensures that the constructed solution meets project goals and is ready for execution.

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

   Why did you select these stages for this task?

Hypothesis testing begins in the Analyze stage because it requires examining the dataset, understanding patterns, and determining which statistical tests are appropriate. It also extends into Construct because you apply the chosen tests to validate assumptions or relationships, producing results that guide model development and decision-making. This task bridges understanding the data and building actionable insights.

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

   Why did you select this stage for this task?

Exploring the data is part of the Analyze stage because it involves examining the dataset to understand its structure, identify patterns, detect anomalies, and determine which features are most relevant. This foundational analysis informs later modeling and decision-making steps.

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

Why did you select these stages for this task?

Data exploration falls under Analyze because it involves inspecting the dataset to understand its structure, detect missing values, and identify patterns or outliers. Data cleaning is part of Construct because it requires actively transforming and preparing the dataset—handling missing values, correcting errors, and formatting features—so it is ready for modeling and further analysis.

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

Why did you select this stage for this task?

Setting up the project workflow is part of the Plan stage because it involves organizing tasks, defining the sequence of work, and determining responsibilities before analysis or construction begins. This ensures the team has a clear roadmap to follow throughout the project.

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

Why did you select this stage for this task?

Sharing final insights occurs in the Execute stage because it involves delivering the completed analysis and model results to stakeholders. This includes creating visuals, summaries, and presentations that translate technical findings into actionable information for both technical and non-technical audiences.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

Computing descriptive statistics is part of the Analyze stage because it involves summarizing and understanding the dataset's characteristics, such as mean, median, standard deviation, and distributions. These insights help guide further analysis, feature selection, and model development.

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

Why did you select these stages for this task?

Visualization begins in the **Analyze** stage because creating graphs, charts, and plots helps the team explore patterns, trends, and relationships in the data. It continues into the **Execute** stage when visuals are refined and presented to stakeholders to communicate findings clearly and support decision-making.

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

Why did you select this stage for this task?

Writing the project proposal is part of the Plan stage because it involves defining the project's goals, milestones, tasks, and stakeholders before any analysis or modeling begins. This sets a clear roadmap for the team and ensures alignment on objectives and expectations.

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

Why did you select this stage for this task?

Building a regression model is part of the Construct stage because it involves actively creating and refining the predictive solution using the prepared dataset. This step focuses on model development, testing assumptions, and ensuring it meets project goals.

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

Why did you select this stage for this task?

Compiling summary information falls under Analyze because it involves reviewing descriptive statistics, distributions, and key characteristics of the dataset. This understanding guides feature selection and subsequent modeling decisions.

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

Why did you select this stage for this task?

Building a machine learning model is in the Construct stage because it requires implementing algorithms, training the model, and iteratively testing and refining it to produce accurate predictions. This is the core development phase of the project.