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.

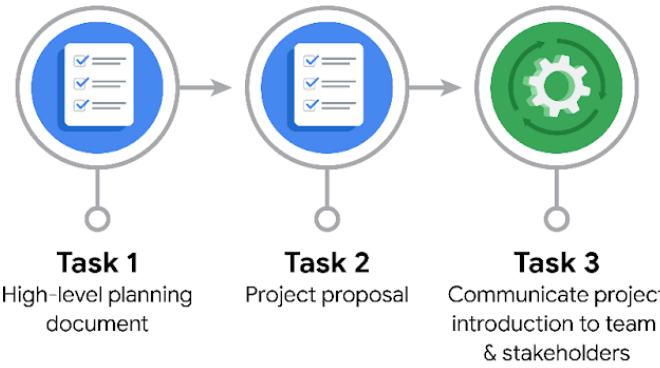
Relevant Interview Questions

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?

Data team (Technical Audience)

1)William Jaffey- Data Science Lead

2) Rosie mae bradshaw - Data Science manager

3) Orion - Data Scientist

Cross-Functional team member (Less Technical Audience)

1) Mary joanna Rodgers -Project management officer
2) Margery Adebowale - Financial lead , Americas
3) Maika Abadi -Operations Lead

- 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 objective is to develop a machine learning model for claims classification on TikTok, aiming to automate the identification of claims and opinions in user-generated content. This will enhance content moderation efficiency, improve the user experience, optimize resource allocation, and strengthen

platform trust by ensuring a safer and more positive environment.

- What questions need to be asked or answered?

1. How can we define and operationalize the distinction between claims and opinions in TikTok content for effective machine learning classification?
2. What preprocessing steps are necessary to curate a high-quality training dataset, considering the unique characteristics of TikTok user interactions?
3. Which machine learning algorithms align with the complexity of the claims classification task, taking into account factors like interpretability and scalability?
4. What key features or attributes should be prioritized in the model development to enhance its accuracy and generalization to diverse content?
5. How do we plan to address ethical considerations and potential biases in the claims classification process to ensure fairness and user trust in TikTok's content moderation?

- What resources are required to complete this project?

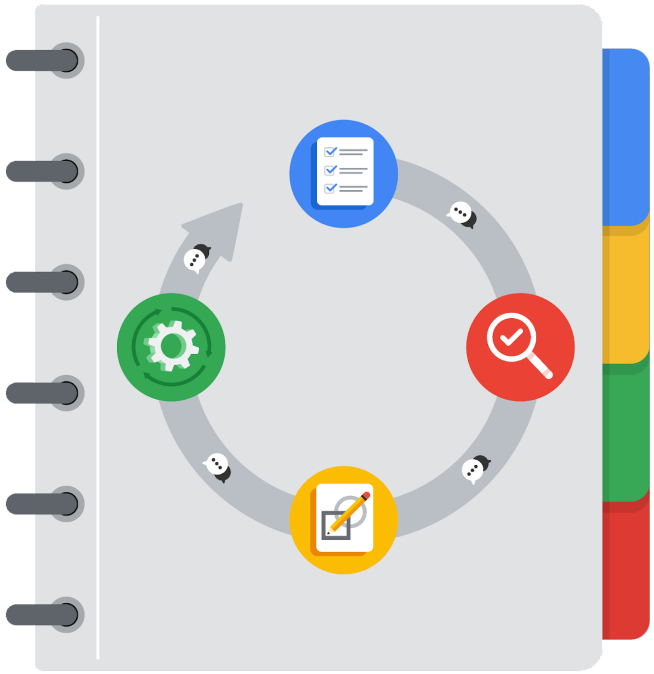
1. **TikTok User Interaction Data:** Access a comprehensive dataset containing user interactions, comments, and reported content.
2. **Programming Tools (Python):** Utilizing Python for its versatility in data analysis, machine learning, and efficient coding practices.
3. **Machine Learning Libraries:** Using libraries like Scikit-learn and TensorFlow to build and train the predictive model.
4. **Natural Language Processing Expertise:** Leveraging NLP techniques to enhance the model's ability to distinguish between claims and opinions in user-generated content.
5. **Project Management Tools:** Implementing tools like Jira or Asana for effective task tracking, collaboration, and milestone management.
6. **Stakeholder Engagement:** Ensuring active communication and collaboration with cross-functional teams, including project managers, finance leads, and operations leads.

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

With this end-of-course project, you will gain valuable practice and apply your new skills as you complete the following deliverables:

- Course 1 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.

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?

I selected the "Execute" stage for the task "Evaluating the model" because this is the phase where the focus is on putting the analysis and construction into action. During this phase, the data professional assesses the model's performance, accuracy, and alignment with project objectives. It involves the practical application of the model, results validation, and necessary refinements based on the evaluation. The "Execute" stage is the appropriate phase for evaluating the model as it represents the active implementation and validation of the constructed model.

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

Why did you select these stages for this task?

I selected the Analyze stage for "Conduct hypothesis testing" because hypothesis testing is essential to data analysis. During the Analyze stage, data professionals interact with the data, conduct exploratory data analysis (EDA), and perform statistical testing, which includes hypothesis testing. This stage involves assessing the dataset, identifying patterns, and gaining insights to inform subsequent steps.

I selected the Construct stage because hypothesis testing often involves constructing statistical models to test specific hypotheses. In this stage, data professionals build, interpret, and revise models. Forming a hypothesis testing framework, defining the variables, and setting up the statistical model to validate or invalidate hypotheses align with the tasks typically performed in the Construct stage.

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

Why did you select this stage for this task?

I selected the Analyze stage for "Begin exploring the data" because exploring the data is a fundamental step in data analysis. During the analysis stage, data professionals acquire, clean, and prepare the data for further investigation. Exploratory Data Analysis (EDA) is a vital component of this stage, involving tasks such as data visualization, summary statistics, and identifying patterns or trends in the data. Exploring the data is crucial to understanding its characteristics before moving on to more advanced analyses.

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

   Why did you select these stages for this task?

Data exploration and cleaning are strategically paired with the "Plan" stage to outline a comprehensive approach to data analysis. Planning ensures a structured strategy for compelling data exploration. Additionally, it is integrated into the "Analyze" stage, where professionals interact with the data for the first time, conducting exploratory data analysis and ensuring data quality. This dual pairing optimizes the early identification and resolution of data issues, setting the groundwork for more informed and successful research.

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

   Why did you select this stage for this task?

I selected the Plan stage for "Establish a structure for project workflow (PACE)" because it is essential to set up a solid foundation for success at the beginning of a project. During the Plan stage, professionals define the project's scope, identify informational needs, and develop a workflow. Establishing the structure for the project workflow is a foundational step in the planning process, ensuring a clear and organized path to guide the project from start to finish.

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

   Why did you select this stage for this task?

I selected the Execute stage for "Communicate final insights with stakeholders" because this task involves presenting and delivering the final insights to internal and external stakeholders. In the execution stage, the focus is on putting the analysis and construction into action. In this phase, the data professional shares the results, receives feedback and makes necessary revisions based on stakeholder input. Communicating final insights is crucial to achieving the project's goals and ensuring that the findings are effectively conveyed to those needing the information.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

I selected the Analyze stage for "Compute descriptive statistics" because this task involves interacting with the data for the first time, acquiring necessary information, and preparing the data for analysis. In the Analyze stage, data professionals collect, clean, and organize the data, and computing descriptive statistics falls within the realm of understanding the dataset and its characteristics. This task is foundational for subsequent analysis and model construction, making it fitting for the Analyze stage of the PACE workflow.

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

Why did you select these stages for this task?

I selected the Analyze and Construct stages for "Visualization building" because this task involves two key aspects. Professionals gather and understand the data in the analysis stage, identifying patterns and insights that can be visualized. Once the analysis is complete, visualizations are constructed, aligning with the construct stage. Visualization building is a bridge between analyzing the data to derive meaningful insights and producing a visual representation to communicate those insights effectively. Therefore, it spans both the Analyze and Construct stages in the PACE workflow.

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

Why did you select this stage for this task?

I selected the Plan stage for "Write a project proposal" because this task involves establishing the groundwork for the entire project. During planning, you define the scope, identify stakeholders, and outline the key milestones. Crafting a project proposal is integral to setting the direction for the project, aligning with the broader objectives, and ensuring that the team is on the same page before moving into the subsequent stages of the PACE workflow.

10. Build a regression model: Construct and Analyze

Why did you select this stage for this task?

Building a regression model involves tasks that span both the "Construct" and "Analyze" stages. In the "Construct" stage, you actively develop and create the model, focusing on building, interpreting, and refining the model. Meanwhile, in the "Analyze" stage, you interact with the data, perform exploratory data analysis (EDA), and conduct statistical testing, which includes tasks related to understanding and analyzing the dataset before and after constructing the regression model. Therefore, building a regression model encompasses activities in the data workflow's "Construct" and "Analyze" stages.

11. Compile summary information about the data: Analyze

Why did you select this stage for this task?

I selected the Analyze stage for "Compile summary information about the data" because this task involves interacting with the data for the first time. In the Analyze stage, data is collected, prepared, and analyzed. Compiling summary information requires an initial examination of the dataset, including data cleaning, organization, and fundamental statistical analysis to understand the key characteristics and trends. This stage sets the foundation for further exploration and modeling in the subsequent phases of the data workflow.

12. Build machine learning model: Construct

Why did you select this stage for this task?

I selected the "Construct" stage for the task "Build machine learning model" because, during this phase, I actively engaged in developing and creating models. The construction stage is designed for building, interpreting, and refining models. It is the most suitable phase for implementing machine learning algorithms to achieve the project's predictive or classification objectives. This aligns with the hands-on construction and development aspects required for building a machine-learning model based on the selected approach and methodology.