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.

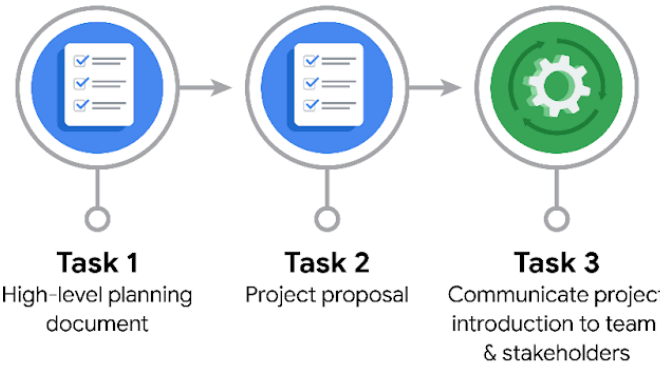
Relevant Interview Questions

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 members (Willow, Rosie, and Orion) and Cross-Functional team members (Mary, Margerie, and Maika). The Data Team members have a technical background, so be specific and concise. The Cross-functional team members oversee operations and are less technical, so communication should be in laymans terms and laid out in a straightforward fashion.

- 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?

Developing a machine learning model to classify TikTok videos as claims or opinions will streamline the review process for reported content. Currently, users having reported violations on their account face delays in content creation while awaiting moderator decisions, which can take weeks and impact livelihoods. This model will reduce backlog, allowing moderators to focus on genuine claims, improve user experience, and minimize disruption for content creators.

- What questions need to be asked or answered?

How long is the turnaround time for moderator review? How many reports are claims vs. opinions after having been reported? How many reports are made monthly, annually, etc? What type of content is reported more avidly? What variables are the most important? Are there any initial trends in the data that can provide insight?

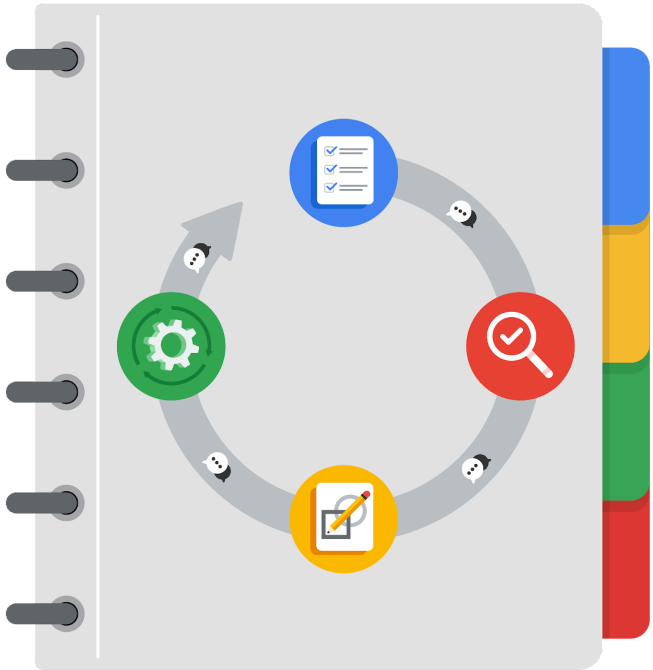
- What resources are required to complete this project?

Data professionals, python, machine learning model

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

A dataset scrubbed for exploratory data analysis, project proposal, statistical model, regression model, machine learning model, and any requested visualizations and reports.

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?

   Because testing the model happens after data analysis. This is where I am implementing the solution I curated during the planning and analyzing stage.

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

   Why did you select these stages for this task?

   Construct is when you are developing the model, Execute is when you are deploying the model and reporting findings and possibly enhancing the model with further training.

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

   Why did you select this stage for this task?

   This is where I'm focused on finding patterns, relationships, and anomalies. I'm also cleaning the data looking for duplicates and missing values and understanding the data collection in general.

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

Why did you select these stages for this task?

Cleaning and processing happens in exploratory data analysis. I include planning because this might mean understanding the tools and resources I need to carry out said stages.

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

Why did you select this stage for this task?

Beginning stages of planning out the workflow and milestones for this project.

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

Why did you select this stage for this task?

This is where I've deployed the model and have several iterations of results perhaps that I can share findings and visualizations to stakeholders with.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

Statistical Analysis and analyzing relationships and distributions is done here.

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

Why did you select these stages for this task?

I have the first results from training a ML model and first few iterations of results I can set visuals with.

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

Why did you select this stage for this task?

Writing a project proposal helps you **formalize** the scope, milestones, deliverables, and timelines for each part of a project."

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

Why did you select this stage for this task?

Building and then refining the model happens during these two stages.

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

Why did you select this stage for this task?

Getting data distributions and identifying patterns, relationships, correlations, is done during this step.

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

Why did you select this stage for this task?

Here we are taking the curated data we processed and cleaned to feed into a model for prediction or classification purposes.