

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

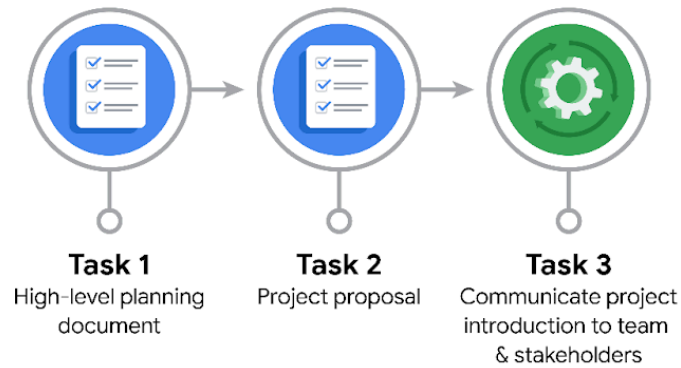
Completing this end-of-course project will empower you to respond to the following interview topics:

- 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 TikTok Executives and the TikTok application moderators

- 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 problem is with the number of submissions to Tiktok, it is difficult for moderators to inspect every video, comment, post, etc. in a timely manner. We are trying to solve this issue through developing a classification based machine learning model to sort each submission as either 'claim' or 'opinion'. Therefore, greatly increasing the efficiency of moderators by heavily reducing the number of submissions to review.

- What questions need to be asked or answered?

- 1) What is the data that the data team will be interacting with?
- 2) What type of regression model will be used?
- 3) What is the best method of hypothesis testing the model?
- 4) Are there any assumptions being made by the regression model?
- 5) What is the project workflow?
- 6) What graphics will best describe the efficacy of the final project for the executives to review?

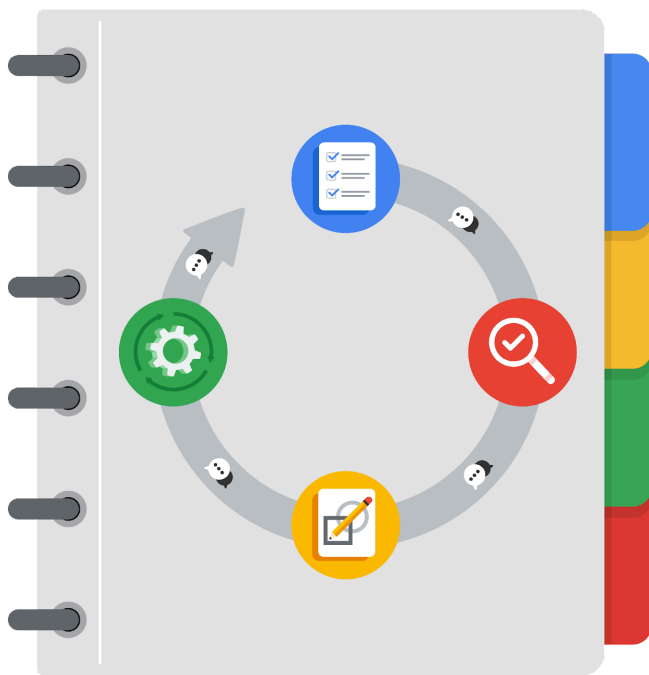
- What resources are required to complete this project?

- 1) High quality data and data storage
- 2) Hardware: computers, cloud storage and computing
- 3) The data team
- 4) Software: machine learning libraries, version control system (git), development environment
- 5) Network and security measures
- 6) Budget and budget plan

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

A efficient and dependable classification regression machine learning model for categorizing submissions as 'claim' or 'opinion'.

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?

During the construction phase, it is the time to build and test models for accuracy.

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

Why did you select these stages for this task?

We must test the model against the test data while we are in the construction phase. While in the analyzing stage, we must determine the statistical tests that we will use in the construction stage.

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

Why did you select this stage for this task?

During the analysis phase, we take the data and begin inspecting the data for trends with exploratory data analysis (EDA).

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

Why did you select these stages for this task?

Of course, during the analysis phase we take the raw data, remove missing and incomplete data and start the EDA. During the planning phase when the data is sourced, measures can be taken to collect better and cleaner sources of data.



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

Why did you select this stage for this task?

At the beginning of the project, we can lay out a solid groundwork for the structure of workflow, setting us up to stay on time and on budget.

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

Why did you select this stage for this task?

Once a final product is produced, we can bring the product and our insights to the stakeholders during the execute phase.

7. **Compute descriptive statistics:** Analyze ▾

Why did you select this stage for this task?

While in the analysis stage, we explore the data with statistical tests.

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

Why did you select these stages for this task?

During the analysis phase, we will be doing EDA, which will be producing graphics that are statistically and observationally driven, then in the construction phase there will be graphics produced to better understand the efficacy and accuracy of the machine learning model chosen.



9. Write a project proposal: Plan ▾

Why did you select this stage for this task?

The Writing of the project proposal happens during the planning stage, setting the team up for a successful and on time project.

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

Why did you select this stage for this task?

The initial regression model is built during the analysis stage, making sure it fits the data. The then building of the model takes place in construction stage.

11. Compile summary information about the data: Analyze ▾

Why did you select this stage for this task?

During the analysis stage, a good understanding of the data must be established. This includes the meaning of every column and row, which then must be recorded as the summary of information that way there will be less confusion with selection of data during construction.

12. Build machine learning model: Construct ▾

Why did you select this stage for this task?

During the construction phase, the cleaned, organized data will be used to construct a machine learning model to address the issue of concern.

