

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 data team:

Willow Jaffey- Data Science Lead

Rosie Mae Bradshaw- Data Science Manager

Orion Rainier- Data Scientist

TikTok executives:

Mary Joanna Rodgers- Project Management Officer

Margery Adebawale- Finance Lead, Americas

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?

Develop a predictive model that can determine whether a video contains a claim or offers opinion. By using it, TikTok can reduce the backlog of user reports and prioritize them more efficiently.



☐ What questions need to be asked or answered?

- ☐ What kind of data do we need?
- ☐ What is the condition of the provided dataset?
- ☐ Have we identified key words, phrases and expression, and even topics and intentions that can help us classify an user interaction data as a claim or opinion?
- ☐ What variables will be the most useful?
- ☐ Are there trends within the data that can provide insight?
- ☐ Have we checked any of the assumptions made by any regression modeling developed for this project?
- ☐ What steps can I take to reduce the impact of bias?

☐ What resources are required to complete this project?

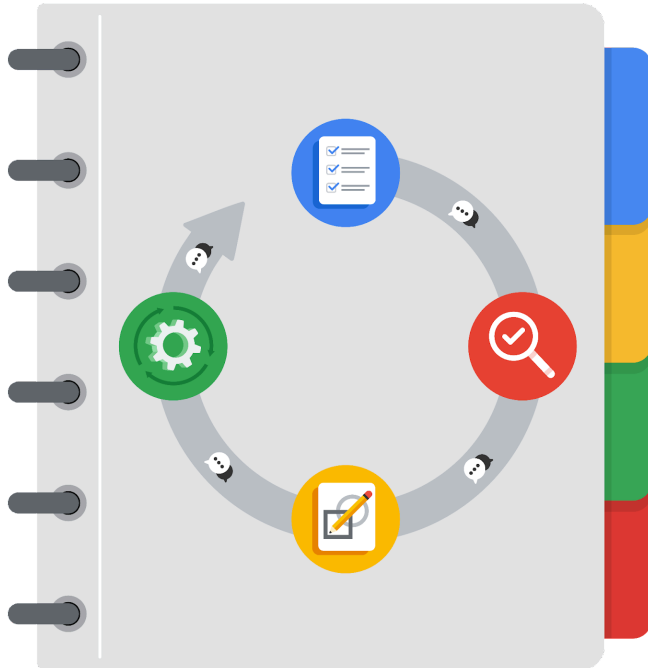
- ☐ To have a thorough understanding of TikTok's [community guidelines](#), including [safety measures and how users report](#) videos and comments containing claims made by other users within the application or website. Additionally, to stay updated on special or stricter protection policies regarding [war content](#).
- ☐ Project dataset
- ☐ Python as the programming language
- ☐ Input from stakeholders

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

- ☐ Project proposal for the data team including: milestones, tasks, deliverables, and estimated time.
- ☐ Dataset to do EDA
- ☐ Visualisations from the statistical models, regression analysis, hypothesis testing
- ☐ Machine learning model
- ☐ Generate visuals of the final results for TikTok executives
- ☐ Constantly updating stakeholders
- ☐ Executive summary



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?

Evaluating is done after constructing the regression model at Construction stage

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

Why did you select these stages for this task?

During the analyzing stage, it is determined that a statistical test will be used. During the construction phase, the test is carried out.

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

Why did you select this stage for this task?

Exploring data is part of the EDA process, where you get an understanding of the kind of data you have, like their features, quantity and quality.

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

Why did you select these stages for this task?

Planning takes place when you first make choices about the methods needed. The cleaning process then takes place in the analyzing stage.

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

Why did you select this stage for this task?



This is the first part of a project, where you establish those four different stages according to the project, and you set milestones, tasks, deliverables and estimated times.

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

Why did you select this stage for this task?

After build and testing the machine learning model in the Construct stage, the next is to show the final results of that work to the stakeholders by making an executive summary and showing some visualisations.

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

Why did you select this stage for this task?

Describing data with statistics is part of Exploratory Data Analysis stage which makes part of analysing the databases available for the project.

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

Why did you select these stages for this task?

Visualization begins with data assessment and is created during the construction stage.

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

Why did you select this stage for this task?

As it is a proposal, it must be part of planning the project, defining milestones, tasks, estimated times and deliverables beforehand.

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

Why did you select this stage for this task?

During the analyzing stage, the model is examined in detail to be sure it will meet the needs of the task. The building of the regression model will take place in the construction phase.



11. Compile summary information about the data: Analyze ▾

Why did you select this stage for this task?

Writing a detailed description of the data available for the project is the first part of the EDA.

12. Build machine learning model: Construct ▾

Why did you select this stage for this task?

The building of a data model would take place in the construct stage.