

# 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 Data Analytics and Moderation teams and stakeholders.

- 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?

We are trying to create a model that will distinguish real content claims from pure opinions to assist the moderation team. This will reduce the amount of time that is needed to moderate content by only requiring manual review for legitimate claims.

- What questions need to be asked or answered?

What sort of content is classed as opinion vs claim?  
Which machine learning models will we use?  
What accuracy are we aiming for?

Are we more concerned about precision or recall?  
Who will be responsible for what features?  
What is the scope of things we will be looking for?

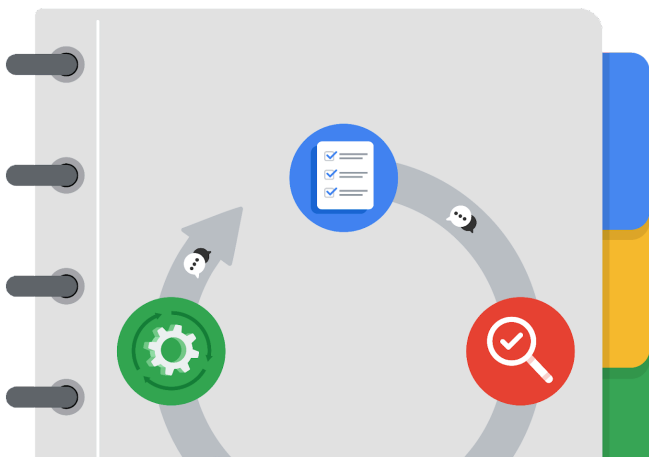
- What resources are required to complete this project?

Data engineers build pipelines to give useful clean data to the model.  
Data scientists create the model.  
Data set of labeled data for training the model.  
Infrastructure to run and store the model.

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

Stakeholder presentations/documentation during the planning and analysing stage.  
Responsibility list and project specification for each role.  
Possible solution presentation after analysis stage.  
Machine learning model that classifies correctly for a certain accuracy.  
Visualisations to explain results.  
Final presentation on final model and its efficacy.

## 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?

Overall evaluation of the model will occur at the end and will be performed when execution is occurring with the final presentation.

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

Why did you select these stages for this task?

Analyze stage for choosing which model to actually build and introduce hypothesis to test for and construct to actually test them.



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

Why did you select this stage for this task?

This will occur after planning and before constructing the model begins so we are aware of the data that can be used in the model.

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

Why did you select these stages for this task?

Both while analysing the data and cleaning during the construct phase.

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

Why did you select this stage for this task?

Establish everything at the start as a plan for the entire project.

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

Why did you select this stage for this task?

Happens at the very end after everything has been constructed.

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

Why did you select this stage for this task?

These will happen during analysis and will be used during the construction phase.



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

Why did you select these stages for this task?

Visualisations will be made for both statistics found in analyse stage and the model performance in construct stage

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

Why did you select this stage for this task?

Happens at the beginning and dictates the rest.

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

Why did you select this stage for this task?

The model is built during the construct stage while being planned and certain parts tested in the analyze stage.

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

Why did you select this stage for this task?

Summary information is done at the end of analysis to easily describe the data in the dataset.

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

Why did you select this stage for this task?

Constructs model based on previous analysis