**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 audience includes TikTok’s leadership team, the data science team, and cross-functional team members who will be involved in the project, such as the Project Management Officer, Finance Lead, and 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 goal is to develop a machine learning model that can accurately classify user-reported claims in TikTok videos. The impact will be significant as it will help TikTok prioritize and address user reports more efficiently, reducing the backlog and ensuring timely moderation.

* What questions need to be asked or answered?

 What type of regression model is most suitable for this task?

 How accurate is the model in classifying claims?

 What are the key variables that influence the classification?

 How will the model's performance be validated?

* What resources are required to complete this project?

A dataset of user-reported claims, Python for model development, tools for data exploration and cleaning, and collaboration platforms for team communication.

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

Project proposal, exploratory data analysis (EDA) report, a regression model, a final machine learning model, visualizations, and a final presentation for stakeholders.

## 

## **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](https://www.coursera.org/learn/foundations-of-data-science/supplement/4OtHr/the-pace-stages) and [Communicate objectives with a project proposal](https://www.coursera.org/learn/foundations-of-data-science/supplement/79Ysh/communicate-objectives-with-a-project-proposal). You will later reorder these tasks within a project proposal.

1. **Evaluating the model: Select PACE stage**

Why did you select this stage for this task?

**Evaluating the model:**

* **PACE Stage:** **Analyze**
* **Reason:** Evaluation of the model's performance fits within the Analyze stage, as it involves interpreting data to assess whether the model meets the desired criteria.

1. **Conduct hypothesis testing: Select PACE stage** **and** **Select PACE stage**

Why did you select these stages for this task?

**Conduct hypothesis testing:**

* **PACE Stage:** **Analyze and Construct**
* **Reason:** Hypothesis testing is part of both analyzing the data to understand patterns and constructing the model to ensure it aligns with theoretical expectations.

1. **Begin exploring the data: Select PACE stage**

Why did you select this stage for this task?

**Begin exploring the data:**

* **PACE Stage:** **Analyze**
* **Reason:** Data exploration is the first step in the analysis phase, where the team familiarizes itself with the dataset and identifies any issues or patterns.

1. **Data exploration and cleaning: Select PACE stage** **and Select PACE stage**

Why did you select these stages for this task?

**Data exploration and cleaning:**

* **PACE Stage:** **Analyze and Construct**
* **Reason:** While data cleaning falls under analyzing to ensure data quality, it also partially involves constructing the dataset in a usable format for modeling.

1. **Establish structure for project workflow (PACE): Select PACE stage**

Why did you select this stage for this task?

**Establish structure for project workflow (PACE):**

* **PACE Stage:** **Plan**
* **Reason:** Establishing the workflow is part of the planning stage to ensure that the project follows a structured approach.

1. **Communicate final insights with stakeholders: Select PACE stage**

Why did you select this stage for this task?

**Communicate final insights with stakeholders:**

* **PACE Stage:** **Execute**
* **Reason:** The final insights and results are communicated in the execution stage to provide actionable recommendations.

1. **Compute descriptive statistics: Select PACE stage**

Why did you select this stage for this task?

**Compute descriptive statistics:**

* **PACE Stage:** **Analyze**
* **Reason:** Descriptive statistics help summarize the main features of the data, making them part of the analysis phase.

1. **Visualization building: Select PACE stage and Select PACE stage**

Why did you select these stages for this task?

**Visualization building:**

* **PACE Stage:** **Analyze and Execute**
* **Reason:** Visualizations are used in both analyzing data and communicating findings to stakeholders.

1. **Write a project proposal: Select PACE stage**

Why did you select this stage for this task?

**Write a project proposal:**

* **PACE Stage:** **Plan**
* **Reason:** The project proposal is an essential part of the planning stage, outlining the scope, objectives, and tasks.

1. **Build a regression model: Select PACE stage and Select PACE stage**

Why did you select this stage for this task?

**Build a regression model:**

* **PACE Stage: Construct and Analyze**
* **Reason: Building the model involves constructing the algorithm and analyzing its fit to the data.**

1. **Compile summary information about the data: Select PACE stage**

Why did you select this stage for this task?

**Compile summary information about the data:**

* **PACE Stage:** **Analyze**
* **Reason:** Summarizing data is part of the analysis process to provide a comprehensive overview of the dataset.

1. **Build machine learning model: Select PACE stage**

Why did you select this stage for this task?

**Build machine learning model:**

* **PACE Stage: Construct**
* **Reason: Constructing the machine learning model is a key task in the construction phase.**