**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?

Audience: Tiktok’s leadership team, the data science manager, the project Management Officer, the data scientist, and the data science 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 project proposal will show milestones and tasks within the comment classification project. This serves as a plan of action where important questions are addressed before the project starts. The project aims to build a machine learning model to accurately predict whether ‘opinions’ or ‘claims’ are been made in TikTok videos.

* What questions need to be asked or answered?

The roles and responsibilities of the individuals: how the RACI table looks like (responsible, accountable, consulted, informed). Also, we outline the project workflow, gather data (videos), and identify hardware/software requirements as well as coding languages. When building the machine learning model, we need to know the desired accuracy of the model (90% or more?). We also need to know the estimated project’s completion to meet the client's needs.

* What resources are required to complete this project?

Clean data (no duplicates, outliers, or missing data) for exploratory data exploration (EDA). For the constructing stage of the project, we need an accurate model that classifies the comments into ‘opinion’ or ‘claim’.

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

The deliverables/outcomes of the project are: data ready for modelling and stakeholders updated about our progress.

## 

## **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?

Tasks in the Construct phase are: Finalize modeling strategies, build a few ML models, test or evaluate models with the desired accuracy. The outcome is the champion model which we will use to classify TikTok videos. In the Execute phase, after the model has been constructed, the data is given as input to the model to evaluate whether it meets the expectations and goals.

1. **Conduct hypothesis testing: Analyze** **and** **Construct**

Why did you select these stages for this task?

We formulate a hypothesis in the Analyze stage of the project after cleaning the data and EDA. It is determined that a statistical test will be used. We identify the champion model and use our input data to test the hypothesis in the Construct phase. The test is carried out.

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

Why did you select this stage for this task?

The EDA takes place after the data is clean in the Analyze stage. During this phase, we gain a deeper understanding of the dataset and the information within.

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

Why did you select these stages for this task?

In the Plan stage, we first make choices about the required methods. One has to gather the data (Plan stage) and then remove possible duplicates, outliers or missing data (cleaning the data), which takes place in the Analyze stage.

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

Why did you select this stage for this task?

This is the very first task in the Plan stage, even before we start collecting data and identifying software/hardware needs. Creating an initial project PACE document outlines the workflow and helps plan how to best approach a project.

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

Why did you select this stage for this task?

After finalizing the results, we can update the stakeholders and share our insights with them. Communication is necessary at various stages throughout the project. Final insights are shared with the stakeholders and actively listen to their comments/feedback. After that, we implement the feedback from the stakeholders and identify future steps. The main outcomes are an executive summary prepared by the data team, visualizations and results that aid the decision-making process.

1. **Compute descriptive statistics: Analyze**

Why did you select this stage for this task?

This is the first task after cleaning the data. Descriptive statistics shows how the frequency of the data, the central tendency of the data, and the variability. This is done during EDA to understand how the distribution of the data.

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

Why did you select these stages for this task?

Visualization begins with data assessment. We make plots and figures during EDA to understand the data’s distribution. The visualizations are created during the construction phase, once we have identified the champion model, we produce visualizations (Tableau, dashboards) for the stakeholders to share insights and facilitate the decision-making.

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

Why did you select this stage for this task?

This is where the goal and general considerations are included to ensure the project is clearly defined. Project workflow and needs are outlined in this document.

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

Why did you select this stage for this task?

We build a regression model to understand if the change in a dependent variable is due to the change of one or more independent variables. We need to understand this relationship in the Analyze phase where the model is examined in detail to be sure that it will meet the needs of the task. In the Construct phase, the regression model is implemented.

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

Why did you select this stage for this task?

We describe how the data is organized and some of the general features of the data. We address whether there are missing data, outliers, or repeated entries. General inspection of the data set to compile information takes place in this phase.

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

Why did you select this stage for this task?

After finalizing the machine learning strategies, we build a few ML models and test them to see which one performs better. We pick the one that reaches the desired accuracy described in the Plan stage (project proposal).