

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 science leadership

- 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?

Address a large volume of reports on the app by a developing machine learning model. It will turn the app more family friendly and content moderation

- What questions need to be asked or answered?

What features or patterns in user interactions are most predictive?
What is going to be banned or restricted?

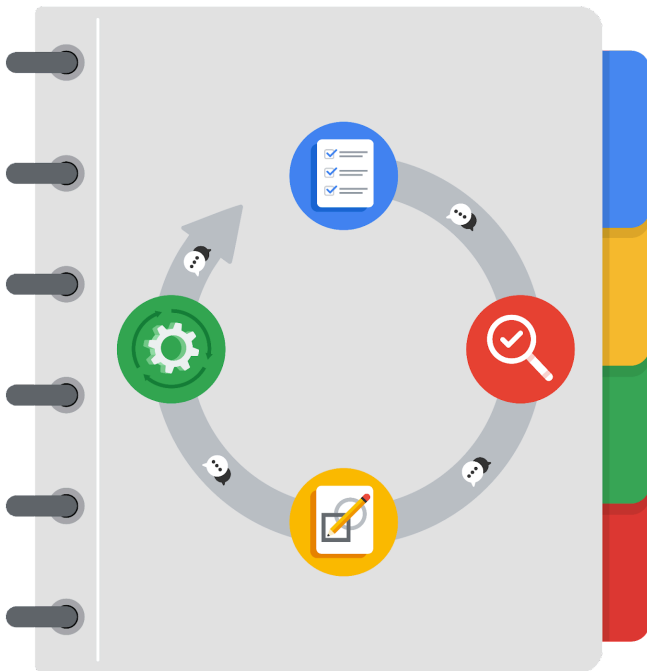
- What resources are required to complete this project?

Data, Machine Learning model, Reports, Tools for visualization (Tableau, Looker etc) and Programming (e.g Python)

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

A trained machine learning classification model

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: Construct ▾

Why did you select this stage for this task?

I selected the Construct stage because evaluating the model is part of building and refining it. This step ensures that the machine learning model performs well before it is deployed or used to generate final insights.

2. Conduct hypothesis testing: Construct ▾ and Execute ▾

Why did you select these stages for this task?

I selected Construct because hypothesis testing involves building the statistical foundation for analyzing the data. I also selected Execute because it requires applying the test results to inform decisions and share findings with stakeholders.

3. Begin exploring the data: Analyze ▾

Why did you select this stage for this task?

It's the stage where you begin exploring the dataset to understand patterns, trends, and potential issues before building any model

4. Data exploration and cleaning: Analyze ▾ and Construct ▾

Why did you select these stages for this task?

I selected Analyze because it involves understanding the structure and quality of the dataset. I also included Construct because data cleaning prepares the dataset for building models and conducting deeper analysis.



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

Why did you select this stage for this task?

It's a initial stage, where the data team will define the structure they will follow throughout the entire project

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

Why did you select this stage for this task?

It's final details about the project. This stage is where you show your results of ML model.

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

Why did you select this stage for this task?

I selected the Construct stage because descriptive statistics help summarize and prepare the data, which is an essential step before building and training a machine learning model.

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

Why did you select these stages for this task?

I selected the Construct stage because building visualizations is part of developing the tools to communicate insights. I also selected the Execute stage because visualizations are used to present final results and support stakeholder decision-making

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

Why did you select this stage for this task?

I selected the Plan stage because writing a project proposal is one of the first steps in a project. This task also supports the overall project structure that was previously defined.

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

Why did you select this stage for this task?



I selected Construct because it is the stage where the data professionals will develop the regression model. I also selected Execute stage because the stakeholders expect the answer from this regression model

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

Why did you select this stage for this task?

I selected the Analyze stage because summarizing the data is part of understanding its structure, distributions, and potential issues before moving on to modeling or deeper analysis.

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

Why did you select this stage for this task?

I selected the Construct stage because building the machine learning model is part of developing the solution. This stage involves selecting algorithms, training the model, and preparing it for evaluation.