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.

Task 1
High-level planning document

Task 2
Project proposal

Task 3
Communicate project introduction to team & stakeholders

Data Project Questions & Considerations

PACE: **Plan Stage**

- Who is your audience for this project?

The tiktok data team and cross functional members.

- 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?

Create a model that can categorize claims vs opinions reports. This will help the client by reducing their time in sorting and can directly focus on the claim reports.

- What questions need to be asked or answered?

What is the state of the dataset?
Which indicators are more important in the dataset.
How can we reduce the impact of bias?

Any trends in the data?

- What resources are required to complete this project?

Past labelled data which contains properties/indicators for each video and the category of claim or opinion it belongs to.

For language we are using python.

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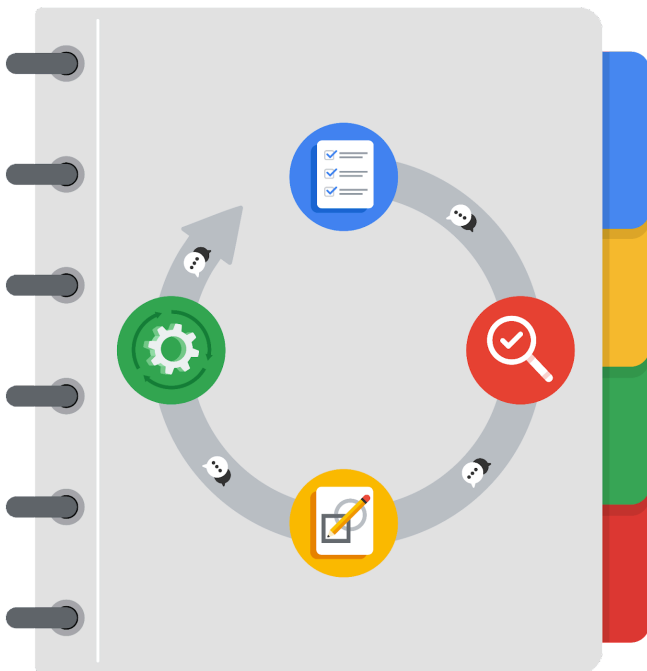
Trained model.

Vizualisations to show effectiveness of model with its predictions.

Executive summary

Cleaned dataset.

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?

   > After model creation, its tested by passing data through it to see if it meets goals.

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

   Why did you select these stages for this task?

   > In the analyze phase define the initial assumptions and which testing method to use. Define any initial parameters required before the construction of the model. After the model is made we test it and complete our hypothesis testing.

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

   Why did you select this stage for this task?

   > This phase will involving gathering data, exploring it, cleaning and getting it ready for model.

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

   Why did you select these stages for this task?

   > Planning stage is for choosing initial methods. analyze phase will involving gathering data, exploring it, cleaning and getting it ready for model.

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

Why did you select this stage for this task?

Before we begin working on the project the entire workflow is planned out to keep everyone on same page and removing any ambiguities regarding tasks and outcomes expected.

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

Why did you select this stage for this task?

This phase is focused on explaining the results to stakeholders using presentations and vizualisations.

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

Why did you select this stage for this task?

Statistics within the data happens during analyze

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

Why did you select these stages for this task?

Data assessment helps us understand the data and viz are created during construct phase.

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

Why did you select this stage for this task?

It needs to be made before the work on project is started.

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

Why did you select this stage for this task?

> Model selection happens during analyze phase and creation comes under construct after we have prepared the data.

11. Compile summary information about the data: Analyze ▾

Why did you select this stage for this task?

> After data exploration and cleaning a summary info can be made to give a quick glimpse of what the data is made of.

12. Build machine learning model: Construct ▾

Why did you select this stage for this task?

> Model is created during this phase.