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.

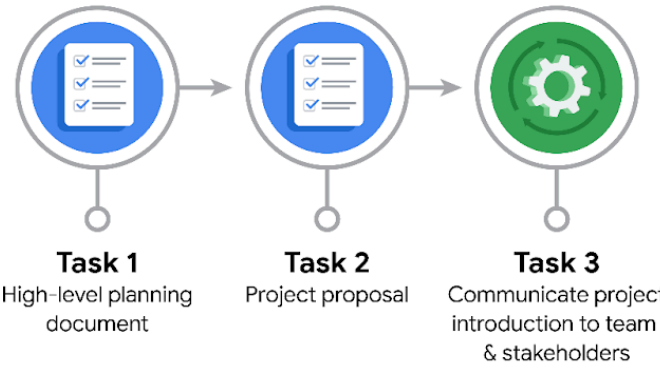
Relevant Interview Questions

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?

My audience is the data team from Tiktok

- 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?

I'm trying to predict if a submitted video contains a claim or is it an opinion, it will help classify the videos and handle them faster.

- What questions need to be asked or answered?

Does the dataset contain any type of key variables that are relevant to the model?
Is the data clean from the pipeline?
How can we overcome our bias?

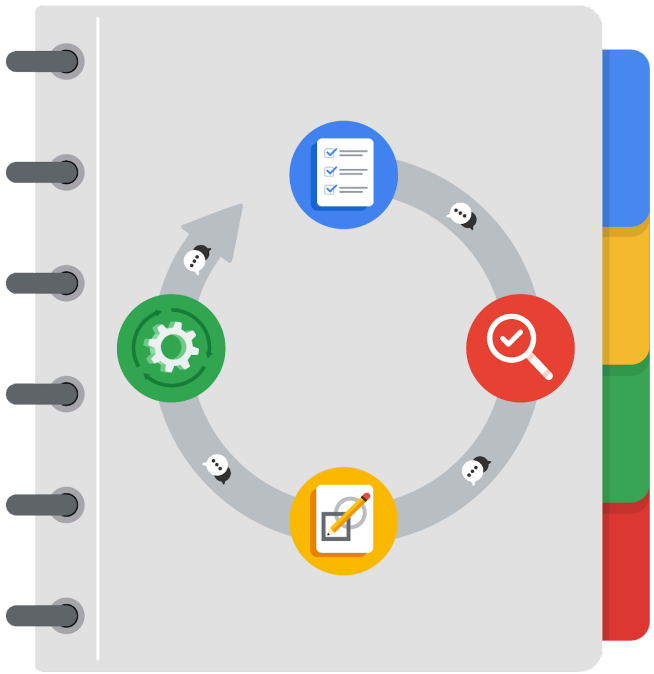
- What resources are required to complete this project?

We will need the dataset, Python and feedback from the stakeholders.

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

The project plan, the clean dataset, the visualization, the statistical regression and the 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:** Execute

   Why did you select this stage for this task?

Because once we have finished building the model, we need to evaluate its effectiveness in the real world right before deploying it.

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

   Why did you select these stages for this task?

During the analysis phase we need to choose the type of statistical test we will be carrying out based in the project requirements, then in the construct phase we will apply the statistical test.

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

   Why did you select this stage for this task?

We need to analyze the data and check the different kinds of variables and have a deep understanding of how they work.

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

   Why did you select these stages for this task?

Data exploration takes place after planning to have a more profound understanding of how data is structured,, and cleaning is after exploring the data and remove any null or invalid info

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

Why did you select this stage for this task?

The structure of the workflow first needs to be established at the beginning in order to have a clear understanding of the steps of the project.

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

Why did you select this stage for this task?

After deploying the statistical model we can learn and get key insights that need to be communicated in an effective and relatable way to the stakeholders.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

Delving into the statistics within data is an essential part of the Analyze phase.

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

Why did you select these stages for this task?

The visualization begins at early steep of the data analysis process and its formally applied at the construction stage.

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

Why did you select this stage for this task?

The project proposal is the earliest action done in a data analytics projects, because is the basis of all the project.

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

Why did you select this stage for this task?

The building of the regression model takes place in the construct phase, however it needs to be analyzed in detail to make sure that it fits within the project requirements.

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

Why did you select this stage for this task?

Its very important to have an index or summary of the data, this is done during the analysis phase.

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

Why did you select this stage for this task?

We build the machine learning model once we collected, cleaned and organized the data.