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.

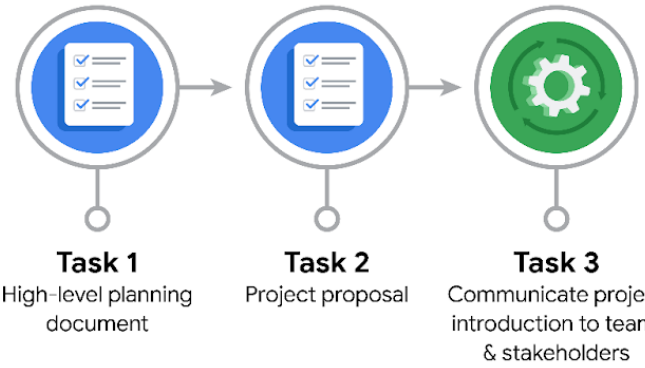
Relevant Interview Questions

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?

Waze Leadership team

- 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?

To prevent user churn by building a predictive model that can predict user churn. The impact of this work will be that the model will optimize Waze's user retention strategy, enhance user experience and make data-driven decisions about product development.

- What questions need to be asked or answered?

Who are the users most likely to churn? Why do users churn? When do users churn?

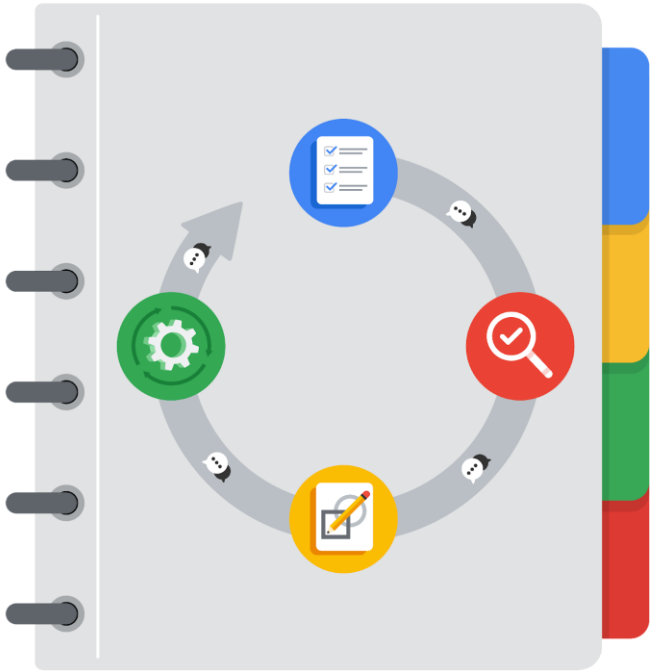
- What resources are required to complete this project?

Python

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

The deliverables will be a project proposal, Dashboard/Visualizations, EDA, Hypothesis Testing, Modelling and sharing the final insights with 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 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 chose this stage as this is where the model is being evaluated based on its metrics and so is in the later stage of the modelling selection process, and so goes in the construct stage.

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

   Why did you select these stages for this task?

   I selected these stages because here is where we are testing for any statistical significance or relationship between the two variables. This therefore fits into the analyze and construct stage.

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

   Why did you select this stage for this task?

   I selected the analyze stage because this is where data exploration begins and where the beginning of the EDA practices start.

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

   Why did you select these stages for this task?

   I selected these stages as some data may need to be consulted before removal such as outliers or duplicates which fits in with the plan stage and with the EDA itself it also fits in with the analyze stage.

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

   Why did you select this stage for this task?

   This is because it is the beginning of the project and where the important questions are going to be asked, and is part of the overall planning process for projects.

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

   Why did you select this stage for this task?

   This is because in the execute stage you communicate the final results/findings with stakeholders.

7. **Compute descriptive statistics:** Analyze

   Why did you select this stage for this task?

   This is part of both EDA (Discovering) and statistics and so is in the analyze stage.

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

   Why did you select these stages for this task?

   This is in both the analyze and construct stages as some visualizations can be useful in the regression process especially when checking for assumptions and so fits into both stages.

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

   Why did you select this stage for this task?

This is one of the first tasks to be done and is where milestones and deliverables are set and so fits in with the plan stage.

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

Why did you select this stage for this task?

It is in these two stages due to the fact that when evaluating the models performance it make both building and evaluating the model to be in both stages.

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

Why did you select this stage for this task?

This is where the EDA practice of discovering takes place and so is apart of the analyze stage.

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

Why did you select this stage for this task?

I chose both analyze and construct as the building of the model falls under the analyze stage and the evaluation and visualization of the models predictions falls into the construct stage.