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.

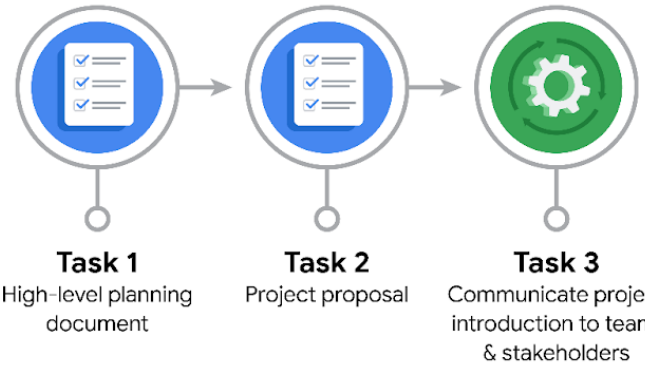
Relevant Interview Questions

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?

All stakeholders, both technical and non-technical, within and outside the data team. They include Harriet Hadzic, Director of Data Analysis, May Santner, Data Analysis Manager, Chidi Ga, Senior Data Analyst, and Sylvester Esperanza, Senior Project Manager on the Waze data team as well as non-technical users Emrick Larson, Finance and Administration Department Head and Ursula Sayo, Operations Manager outside the data 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?

We're trying to leverage data analysis and modeling to predict which segment(s) of users are at high risk of churning so that Waze could engage, understand, and retain them. This will help the client understand its customers and design effective business strategy to improve user retention and grow the user base.

- What questions need to be asked or answered?

1. What data do we need/are available for analysis and modeling?
2. What are the demographics of Waze users?
3. Which modeling / machine learning techniques are most appropriate?
4. What are the reasons that users are churning?

5. Where Waze could create value for users and raise satisfaction?

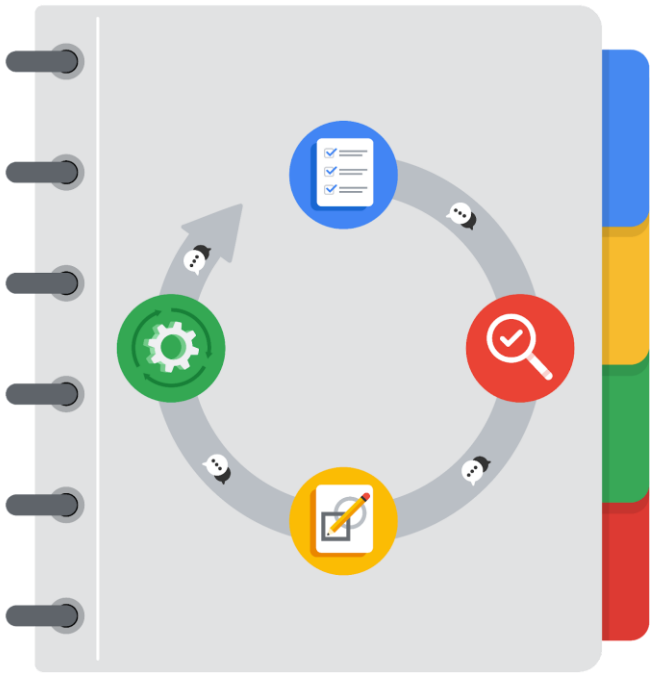
- What resources are required to complete this project?

1. Datasets that reveal user behavior patterns.
2. A computing environment where Python and packages are installed.
3. Compute power.

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

1. Dashboards and reports describing the Waze user base.
2. An analysis on why users are churning and staying.
3. Statistical analyses and machine learning models to predict churns based on patterns.
4. Presentation with recommended actions to retain customers.

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?

The Construct phase is where modeling happens.

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

   Why did you select these stages for this task?

Hypothesis testing is useful for understanding correlations between variables and feature engineering for machine learning. Hence it could appear in both Analyze and Construct where in Analyze we explore the data and in Construct we build models.

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

   Why did you select this stage for this task?

Analyze is the phase in which we make sense of the dataset.

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

   Why did you select these stages for this task?

Data exploration is at the core of Analyze. In order to build accurate models in Construct, we need to clean up the data. Exploration and cleaning support each other.

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

   Why did you select this stage for this task?

   Plan is where we structure and plan the entire project workflow before actual work begins.

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

   Why did you select this stage for this task?

   In Execute we communicate insights and formulate business strategy to put the findings into action.

7. **Compute descriptive statistics:** Analyze

   Why did you select this stage for this task?

   Descriptive statistics is useful for exploring data which happens in Analyze.

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

   Why did you select these stages for this task?

   Visualizations are useful for understanding the dataset in Analyze and presenting results to stakeholders in Execute.

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

   Why did you select this stage for this task?

   Project proposals are designed to communicate objectives to stakeholders so that collaboration is smooth. This should happen at the Plan stage.

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

    Why did you select this stage for this task?

Regression models are used for predicting outcomes. This happens at Construct. It could also be used for exploring relationships between variables in Analyze.

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

    Why did you select this stage for this task?

Making sense of data happens in Analyze.

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

    Why did you select this stage for this task?

Machine learning happens in Construct.