

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.

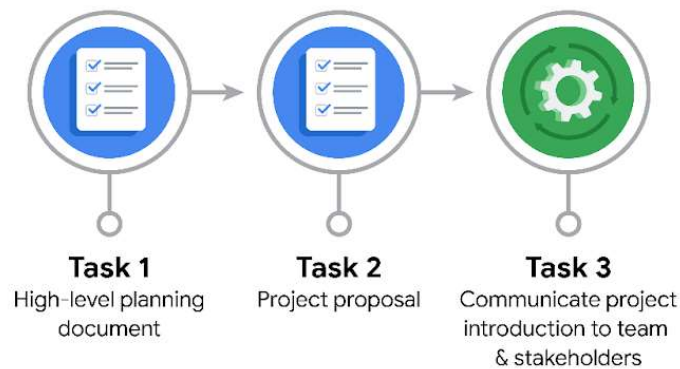
Relevant Interview Questions

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?

The direct audience for this project is the team members at Waze: Harriet Hadzic (Director of Data Analysis), May Santner (Data Analysis Manager), Chidi Ga (Senior Data Analyst), and Sylvester Esperanza (Senior Project Manager.) Indirectly, other members include: Emrick Larson (Finance and Administration Department Head) and Ursula Sayo (Operations Manager.)

- 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?

At this stage, it is necessary to create a project proposal that outlines the full scope of the project from beginning to end. This will ensure that there is a guideline to follow for all subsequent steps in the project, including the relevant PACE stages. Additionally, it will allow various team members collaborating on the project to quickly get an idea of the project's progress and its current stage. The overall project goal is to create a machine learning model that can predict Waze user churn. This will enable Waze to develop strategies to prevent customer churn and attract back users who have already left.

- What questions need to be asked or answered?

1. What is the data source?



2. Was the data acquired in a manner that is representative of the population at large?
3. How long is the project approximately going to take?
4. Are there any ethical concerns with any aspects of the model being developed?
5. What is the communication hierarchy going to look like? Who needs to be directly updated on the project's during each milestone?

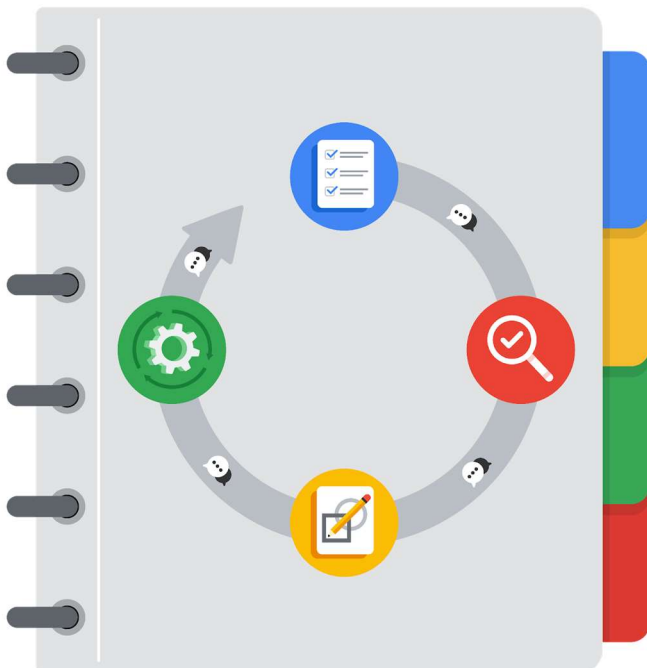
● What resources are required to complete this project?

1. Access to a sample of the Waze user data, and potentially more if deemed necessary.
2. Python notebook.
3. Input from stakeholders.

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

The Project Proposal, PACE Strategy Document and Executive Summaries for each phase of the project. Additionally, cleaned dataset for EDA, visualizations, outcomes of statistical tests, regression analysis and/or a machine learning 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?

It's at this stage that any models in development are built, evaluated and revised.

2. Conduct hypothesis testing: Analyze and Select PACE stage

Why did you select these stages for this task?

At this stage, data is examined for patterns, relationships, and correlations. Hypothesis testing is conducted to determine whether the original hypothesis is supported.

3. Begin exploring the data: Analyze

Why did you select this stage for this task?

Data exploration occurs during the EDA portion of the process, which falls under the Analyze step of the PACE framework.

4. Data exploration and cleaning: Analyze and Analyze

Why did you select these stages for this task?

Data exploration and cleaning both occur within the EDA process, which falls under the Analyze phase.

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



Why did you select this stage for this task?

The Plan section of the overall workflow outlines the project's structure. While adjustments can be made throughout the process, it's at this stage that the major milestones should be indicated.

6. Communicate final insights with stakeholders: Execute

Why did you select this stage for this task?

Communicating final insights to stakeholders occurs during the Execute stage, after the data has been cleaned, hypotheses have been tested, and models have been built. This is the portion of the PACE framework where the results of all these prior steps are communicated to the relevant audience.

7. Compute descriptive statistics: Analyze

Why did you select this stage for this task?

Descriptive statistics are generated in the Analyze phase, when the data is initially investigated. This allows individuals to identify key features of the dataset by providing information such as measures of central tendency (mean, median, mode), measures of variability (standard deviation, variance, range), frequency distributions (frequency tables, histograms), and data visualizations (bar charts, line charts, scatter plots).

8. Visualization building: Analyze and Execute

Why did you select these stages for this task?

Visualization building can occur in both the Analyze stage (when data is first investigated) and the Execute stage (when final insights are provided to stakeholders).

9. Write a project proposal: Plan

Why did you select this stage for this task?

A project proposal is written during the Plan stage, when the initial outline of the overall project is developed.

10. Build a regression model: Analyze and Construct

Why did you select this stage for this task?

The core development of a regression model typically occurs in the Analyze phase, where hypotheses about variable relationships are tested. However, it may also partially occur in the Construct phase, where other models, such as machine learning models, are developed.

11. Compile summary information about the data: Plan

Why did you select this stage for this task?

A summary of information about the data itself often occurs in the Plan phase, when data sources are identified, and a higher-level understanding of the data is acquired and summarized.

12. Build machine learning model: Construct

Why did you select this stage for this task?

Machine learning models are often built in the Construct phase, where they are evaluated, refined, and compared to other models to determine the highest-performing or "champion" model that can be tested on the final test set.