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.

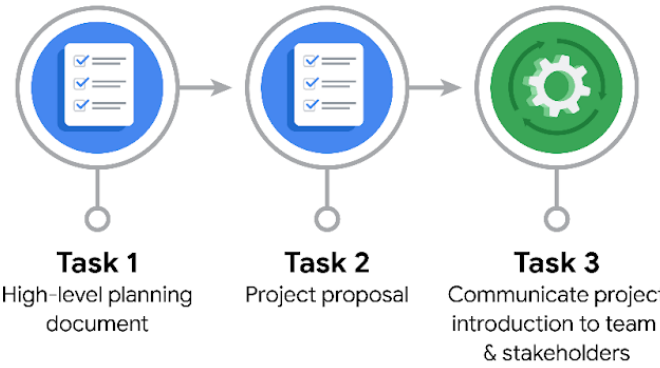
Relevant Interview Questions

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's 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?

The goal of this project is to develop a machine learning model to predict user churn. The impact of making this model is to let executives easily know and determine beforehand on how to prevent the user turn, imrpove the user rentention rate, and growth the business.

- What questions need to be asked or answered?

- Will we need to gather additional resources besides our interal data sources?
- What do we need to know before we start the Exploratory Data Analysis (EDA) process?
- What steps do we need to perform in the EDA process?

-
- Do we need to consider developing either one or multiple model to achieve maximum effectiveness?
-

- What resources are required to complete this project?

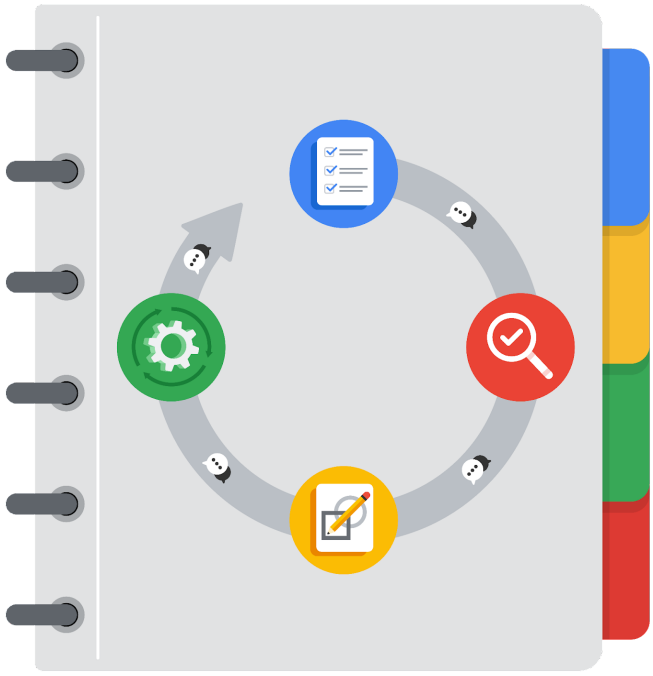
- Waze user data to develop the machine learning model.
- Jupyter Notebook to write Python scripts.

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

The deliverables are set specifically over the course of this project from the PACE workflow:

- With the **Planning** phase, we need to understand our project workflow.
- In the **Analyzing** process, we'll have the comprehensive and reliable data that we need and be ready to develop the model. Besides, we'll create visualizations which share the key insights and conduct descriptive statistics.
- In the **Construct** phase, we'll combine along with the **analyzing** phase to create regression model, develop the machine learning model, and start to deliver the results to the executives.
- Finally, in the **Execute** phase, we'll share the results to the executives and incorporate their feedback.

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?

   This stage is overall about developing the machine learning model. After getting things done such as training and testing the model, evaluating is the necessary step in order to get the best result that the model brings to.

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

   Why did you select these stages for this task?

   We might conduct the hypothesis testing to prove the given hypothesis is either true or false by making tests such as: t-test, ANNOVA, chi-squared test, etc.

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

   Why did you select this stage for this task?

   The analyzing stage is when we gather the necessary information we need, and then start to explore the data afterwards.

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

Why did you select these stages for this task?

In analyzing stage, data exploration and cleaning are included in EDA process. Cleaning data to have the reliable, accurate and consistent data. Exploring data is to obtain the information we truly need rather than working on a bunch of data that might include irrelevant information. Planning is to select the needed method/tools to analyze.

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

Why did you select this stage for this task?

In the plan stage, it's important to create a project workflow to let the whole team involved in the project having a clear understanding of what will be included in every stage of the project.

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

Why did you select this stage for this task?

After getting done with the analysis and finalizing the model, the execute stage is when to share the results to the executives. Then, making recommendations, getting the feedbacks, questions, or concerns from the stakeholders might have.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

In this stage, computing descriptive statistics is the most important step to let us understand the data better and perform complex calculations to have a bigger view of the picture from the data we're working with.

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

Why did you select these stages for this task?

In analyzing stage, visualization is essential to better understand the data. When visualizing data, we can easily catch and get the most valuable insights that the data is showing, rather than looking at the data without visualizations which might be a bit inefficient to thoroughly understand. Visualizations in the construct stage might use to refer in order to create model. Also the visualizations will share the key insights of the data to the executives.

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

Why did you select this stage for this task?

Writing a project proposal in the plan stage will let the whole team know about the overall of what will be included in the project.

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

Why did you select this stage for this task?

For instance, in analyzing stage, we can build simple linear or logistic regression model to visualize the relationship of specific variables within dataset. The same concept can be applied to the construct phase where we build the model to train and test with the purpose of predicting (for example: house prediction).

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

Why did you select this stage for this task?

In this phase, we majority work with the data. Create a summary information of data is useful in order to inform the whole team working on the project completely understanding on how to best approach with the data to get the most from it.

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

Why did you select this stage for this task?

We start to build the machine learning model after we get done from the analyzing stage, which the data is reliable to use.