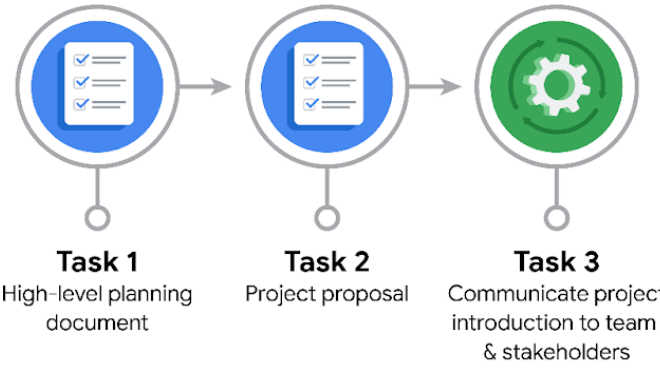
Course One
Foundations of Data Science

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 primary audience for this project includes:

1. Waze Data Team Members:
- Data Analysis Manager (May Santner)
- Senior Data Analyst (Chidi Ga)
- Director of Data Analysis (Harriet Hadzic)
- Senior Project Manager (Sylvester Esperanza)
2. Waze Co-workers Outside the Data Team:
- Emrick Larson (Finance and Administration Department Head)
- 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?

Problem to Solve:

- Address the challenge of reducing monthly user churn on the Waze app, focusing on understanding why users churn and predicting high-risk churn users.

Goals and Anticipated Impact:

1. Retain Users: Decrease the number of users uninstalling or discontinuing Waze each month.
2. Enhance Experience: Improve user satisfaction by addressing identified pain points and optimizing features.
3. Targeted Strategies: Develop personalized retention strategies based on predictive churn modeling. This may include personalized offers, notifications, or features aimed at retaining users identified as high-risk churn candidates.
4. Data-Driven Decisions: Promote a data-driven culture, providing actionable insights for informed decision-making.
5. Business Growth: Contribute to business growth by fostering high user retention rates.
6. Proactive Engagement: Engage with users at risk of churn proactively, preventing potential losses.

- What questions need to be asked or answered?

1. Who are the users most likely to churn?
- Identify user segments or characteristics associated with high churn rates.
2. Why do users churn?
- Uncover the reasons behind user churn, such as dissatisfaction with features, user experience issues, or external factors.
3. When do users churn?
- Determine patterns or specific points in the user journey when churn is more likely to occur.
4. What are the key factors influencing churn?
- Analyze which features, behaviors, or external factors contribute significantly to user churn.
5. How accurate is the churn prediction model?
- Evaluate the performance of the machine learning model in predicting user churn.
6. What insights can be derived from exploratory data analysis (EDA)?
- Understand user behavior patterns, correlations, and anomalies through EDA.
7. What are the main talking points for the presentation to the leadership team?
- Summarize the insights gained from the analysis and model development.
8. How can the model be tested to ensure consistent results?
- Establish methods to test the model's consistency across different subsets of data.
9. What actions can be taken to proactively engage users at high risk of churn?
- Develop strategies or offers to retain users identified as being at risk of churning.
10. How can the project contribute to overall business growth and user satisfaction?
- Link the project outcomes to broader business goals and strategies.

- What resources are required to complete this project?

1. Data

2. Computational Resources:
3. Software: Programming tools, particularly Python, for data analysis, machine learning model development, and visualization.
4. Machine Learning Libraries: Libraries such as scikit-learn, TensorFlow, or PyTorch for implementing machine learning algorithms.
5. Data Analysis Tools: Tools for exploratory data analysis (EDA), statistical analysis, and visualization (e.g., pandas, matplotlib, seaborn).
6. Documentation Tools: mTools for documenting insights, findings, and methodologies (e.g., Jupyter Notebooks, Markdown).
7. Collaboration Tools: e.g., Slack, Microsoft Teams.
8. Project Management Tools: e.g., Jira, Trello
9. Training Data
10. Testing Data
11. Presentation Tools: Tools for creating visualizations and presentations for communicating findings to stakeholders (e.g., Matplotlib, PowerPoint).

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

1. Project Proposal: Comprehensive document outlining project goals and milestones.
2. PACE Strategy Document: Strategy for project planning considering audience, team, and milestones.
3. Readme File
4. Data Exploration and Preparation Reports: Insights from data inspection, exploratory data analysis (EDA), and variable relationship analysis.
5. Model Development Documentation: Details on the chosen machine learning model, feature engineering, and model training.
6. Model Evaluation Reports: Performance evaluation of the churn prediction model and consistency testing results.
7. Presentation Preparation Documentation: Summaries of insights, presentation outline, and visuals for leadership.
8. Final Model and Implementation Plan: Documentation finalizing the churn prediction model and the plan for integration.
9. Comprehensive Project Documentation: Record of the entire project process for reference and knowledge sharing.
10. Regular Progress Updates: Ongoing communication of project progress through meetings and collaboration tools.
11. Insights Communication: Clear communication of derived insights to relevant stakeholders.
12. GitHub Repository
13. Training Materials: Materials for training and knowledge sharing within the team.
14. Feedback Integration Records: Documentation of feedback received and its integration into the project.
15. Project Closure Report: Summary of project outcomes, lessons learned, and recommendations for future projects.

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?

   - After the model has been constructed, data is run through to evaluate whether it meets the project's expectations and goals.

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

   Why did you select these stages for this task?

   - During the analyzing stage, it is determined that a statistical test will be used. During the construction phase, the test is carried out.

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

   Why did you select this stage for this task?

   - During the analysis phase, you will gain a deeper understanding of the dataset and the information within it.

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

   Why did you select these stages for this task?

- Planning takes place when you first make choices about the methods needed. The cleaning process then takes place in the analyzing stage.

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

Why did you select this stage for this task?

- Planning stage. Creating an initial project PACE document outlines the workflow and helps to plan how to best approach a project.

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

Why did you select this stage for this task?

- Communication is necessary at various points throughout a project. Final insights are shared with stakeholders in the execute phase of the data project workflow.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

- Investigating the statistics within data takes place during analysis.

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

Why did you select these stages for this task?

- Visualization begins with data assessment and is created during the construction stage.

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

Why did you select this stage for this task?

- Planning stage. A project proposal is the initial document used to define a project.

10. **Build a regression model:** Select PACE stage **and** Select PACE stage

Why did you select this stage for this task?

- During the analyzing stage, the model is examined in detail to be sure it will meet the needs of the task. The building of the regression model will take place in the construction phase.

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

Why did you select this stage for this task?

- Inspecting a dataset to compile information would take place in the analysis phase.

12. **Build machine learning model:** Select PACE stage

Why did you select this stage for this task?

- The building of a data model would take place in the construct stage.