**Waze Project**



**Data Project Questions & Considerations**

**PACE: Plan Stage**

* What are the data columns and variables and which ones are most relevant to your deliverable?

Since we are interested in user churn, the label column is essential. Besides label, variables that tie to user behaviors will be the most applicable. All variables tie to user behavior except ID.

* What units are your variables in?

The units of the variables are as follows: sessions and drives are counts, n\_days\_after\_onboarding is in days, total\_navigations\_fav1 and total\_navigations\_fav2 are counts, driven\_km\_drives is in kilometers, duration\_minutes\_drives is in minutes, and activity\_days and driving\_days are counts.

* What are your initial presumptions about the data that can inform your EDA, knowing you will need to confirm or deny with your future findings?

My initial presumption is that there will be a positive correlation between the number of sessions, total navigations, and driven kilometers, indicating higher user engagement leads to more driving activity.

* Is there any missing or incomplete data?

Yes, 700 rows have label missing. Other variables have no missing values.

* Are all pieces of this dataset in the same format?

Yes, all pieces of this dataset are in the same format, with each column representing a specific type of user activity or attribute.

* Which EDA practices will be required to begin this project?

The EDA practices required to begin this project include data cleaning, checking for missing values, summarizing statistics, visualizing distributions, exploring correlations, and identifying outliers.

**PACE: Analyze Stage**

* What steps need to be taken to perform EDA in the most effective way to achieve the project goal?

Define churn and retention metrics, clean the dataset, handle missing values, calculate descriptive statistics, visualize data distributions and relationships, perform time-series analysis, compute correlation coefficients, identify and analyze outliers, engineer relevant features, develop preliminary churn prediction models, analyze retained user patterns, summarize initial insights and hypotheses.

* Do you need to add more data using the EDA practice of joining? What type of structuring needs to be done to this dataset, such as filtering, sorting, etc.?

Determine if additional data is needed through joining, and structure the dataset by filtering out irrelevant records, sorting by key metrics, and aggregating data as necessary for analysis.

* What initial assumptions do you have about the types of visualizations that might best be suited for the intended audience?

Initial assumptions include using line charts for trends over time, bar charts for comparisons between user segments, and heatmaps for activity patterns, tailored to the audience's need for clear, actionable insights.

**PACE: Construct Stage**

* What data visualizations, machine learning algorithms, or other data outputs will need to be built in order to complete the project goals?

To complete the project goals, build visualizations such as histograms for trends, boxplot for segment comparisons; use machine learning algorithms like logistic regression or decision trees for churn prediction; and generate reports or dashboards to present actionable insights.

* What processes need to be performed in order to build the necessary data visualizations?

Prepare data, select visualization tools, define visual goals, create visualizations, refine designs, and validate insights.

* Which variables are most applicable for the visualizations in this data project?

Variables most applicable for visualizations include sessions, total\_sessions, total\_navigations\_fav1, total\_navigations\_fav2, driven\_km\_drives, duration\_minutes\_drives, activity\_days, and driving\_days.

* Going back to the Plan stage, how do you plan to deal with the missing data (if any)?

Identify missing data patterns, choose imputation or exclusion methods, and document the approach.

******PACE: Execute Stage**

* What key insights emerged from your EDA and visualizations(s)?

Key insights from the EDA and visualizations include patterns in user engagement, factors influencing churn, and trends in retention across different segments.

* What business and/or organizational recommendations do you propose based on the visualization(s) built?

Based on the visualizations, propose targeted retention strategies, personalized engagement initiatives, and adjustments to user onboarding processes to reduce churn and improve overall user satisfaction.

* Given what you know about the data and the visualizations you were using, what other questions could you research for the team?

Explore questions such as: What specific features correlate most with user churn? How do different user segments respond to engagement strategies? What are the long-term trends in user behavior post-onboarding?

* How might you share these visualizations with different audiences?

Share visualizations with different audiences by tailoring formats: use interactive dashboards for executives, detailed reports for analysts, and simplified charts for broader teams.