**Project Outline:**

**I. Introduction**

Briefly introduce the proposed project

Provide background information on the problem of rider cancellations in the ride-sharing industry

**II. Problem Statement**

Define the problem of rider cancellations and its impact on the ride-sharing industry

Explain why a rider-driven cancellation prediction model is necessary to address this problem

**III. Data Collection**

Identify the data sources that will be used to build the model

Discuss the types of data that will be collected, such as rider history, driver history, and weather data

**IV. Data Preprocessing**

Explain how the collected data will be preprocessed to ensure it's in a suitable format for analysis

Discuss common data preprocessing techniques, such as removing duplicates, handling missing values, and converting categorical variables into numerical variables

**V. Feature Engineering**

Discuss how new features will be created from the existing data to improve the model's accuracy

Provide examples of potential features, such as estimated time of arrival for the driver or the distance between the rider's pickup location and the driver's current location

**VI. Model Selection**

Explore different machine learning algorithms and select the best one for the problem

Compare the pros and cons of various algorithms, such as logistic regression, decision trees, and random forests

**VII. Model Training**

Train the selected model on the preprocessed data

Use cross-validation techniques to ensure the model generalizes well to new data

Tune the model's hyperparameters to optimize its performance

**VIII. Model Deployment**

Deploy the model so it can be used in real-time

Integrate it into the ride-sharing app and provide drivers with real-time notifications of potential cancellations

Monitor the model's performance and make updates as necessary

**IX. Conclusion**

Summarize the proposed project and its potential benefits to multiple stakeholders in the ride-sharing industry

Provide recommendations for future work in this area

Purpose of this proposed project (doc file):

The purpose of this proposed project is to address the problem of rider cancellations in the ride-sharing industry by developing a rider-driven cancellation prediction model. This project aims to collect and preprocess data from multiple sources, create new features from the existing data, select the best machine learning algorithm, train the model on the preprocessed data, and deploy it so it can be used in real-time. By providing drivers with real-time notifications of potential cancellations, this model can help reduce cancellations, increase revenue, and improve the experience for both riders and drivers. The proposed project outlines the steps necessary to build this model and provides a roadmap for future work in this area.