**Problem statement:**

In the ride-sharing industry, rider cancellations can lead to lost revenue and wasted time for drivers. To address this problem, we propose building a rider-driven cancellation prediction model that can accurately predict when a rider is likely to cancel their ride.

**Data collection:**

To build this model, we'll collect data from multiple sources, including rider history, driver history, and weather data. We'll also gather data on the time of day, day of the week, and location of the pickup and drop-off points.

**Data preprocessing:**

After collecting the data, we'll preprocess it to ensure it's in a suitable format for analysis. This may include removing duplicates, handling missing values, and converting categorical variables into numerical variables.

**Feature engineering:**

We'll create new features from the existing data to improve the model's accuracy. For example, we could calculate the estimated time of arrival for the driver based on their current location and the time of day.

**Model selection:**

We'll explore different machine learning algorithms and select the best one for this problem. We'll consider algorithms such as logistic regression, decision trees, and random forests.

**Model training:**

Once we've selected a model, we'll train it on our preprocessed data. We'll use cross-validation techniques to ensure the model generalizes well to new data. We'll also tune the model's hyperparameters to optimize its performance.

**Model deployment:**

Finally, we'll deploy the model so it can be used in real-time. We'll integrate it into the ride-sharing app and provide drivers with a notification when the model predicts a high likelihood of cancellation. We'll also monitor the model's performance and make updates as necessary.

Overall, we believe that a rider-driven cancellation prediction model can help reduce cancellations, increase revenue, and improve the experience for both riders and drivers.