Data Science Capstone

Understanding What Drives Ratings and Cancellations with Uber in India's Capital District

Prepared By:

Lee Manning

Jakob Theis

Michelle Flores Sanchez

Nicholas Robertson

DAT 490

Table of Contents

Abstract/Executive Summary

Our study analyzed the Uber Ride Analytics Dataset 2024 (“The Dataset”) with the aim of determining which factors influence both driver and passenger ratings, as well as the likely locations and predictors of ride cancellations. Cancellations create great inefficiencies in ride supply while also lowering driver morale and passenger satisfaction. We believe that reducing cancellations will lead to improved revenues. Since drivers are the primary face of the company for customers, driver ratings are indicative of how consumers feel about Uber. Finding correlations between driver ratings and operational factors would allow Uber an opportunity to address these pain points, thereby improving customer opinion of Uber.

The Dataset contained information for 150,000 unique Uber rides booked in India’s National Capital Region during 2024. The columns in The Dataset contain temporal and geographic details for each ride, including estimates of wait time, distance, and duration. It includes vehicle type, method of payment, and customer & driver ratings at the time of booking. The dataset also recorded the outcome of each booking, whether the ride was completed, cancelled, or marked as incomplete. Despite the large number of entries in The Dataset, many columns contained null entries.

The attempt to predict incomplete or completed rides was limited by a class imbalance within the booking status variable. The initial goal of this question was to predict all booking statuses, including cancellations. However, we ultimately found that a more comprehensive dataset is needed to achieve this. Our analysis of ratings revealed that ride distance, booking

value, and wait time were the factors most strongly associated with differences in driver ratings. Lower ratings were most often linked to long wait times paired with short and relatively costly trips, and vice versa for higher ratings. Our geographical analysis inspected pickup and drop-off locations for cancellation rates. We were able to identify cancellation hotspots where Uber could focus efforts for reducing cancellations. By parsing cancellations between pick-up and drop-off locations, we observed that reasons given for customer cancellations at pick-up points tended to be customer-based, whereas the trend for cancellations tied to drop-off locations showed elevated rates of bad driver behavior, like not moving towards the customer. While some of these causes are unavoidable, Uber could take action to ameliorate others we have highlighted in our analysis. Such actions would likely increase customer satisfaction and decrease cancellations, thereby boosting Uber's reputation and revenues.

Project Plan

Understanding What Drives Ratings and Cancellations with Uber in India's Capital District

Team Drop It Like It's Null Project Plan

Team Lead – Lee Manning
Analyst – Jakob Theis
Analyst – Michelle Flores Sanchez
Analyst – Nicholas Robertson

November 2, 2025

Profile of the organization and opportunity background

Company Details:

Founded – March 2009
Founders – Garrett Camp, Travis Kalanick
Headquarters – San Francisco, CA
IPO – Uber began public trading on May 10, 2019, after an IPO on May 9, 2019
Categories – Ride-hailing, food delivery, courier services, electric bicycle/scooter rental

Address:

1725 3rd Street
San Francisco, CA 94158
United States

Contact Information:

Website: www.uber.com
Phone: 415-612-8582

Business Description:

Uber is a technology company whose app allows consumer-end users to find rides and have food delivered to them. Uber also has business and freight offerings, but we will focus on their consumer services.

Uber connects people seeking rides with a driver or electric scooter/bicycle near their location. It also has a service called Uber Eats, which allows users to place orders with local restaurants. Uber Eats drivers then pick up the food from the restaurant and deliver it to the consumers. The ride-hailing service offers passengers an array of vehicle classes to suit their needs, from single-rider electric scooters and bicycles to sharing a ride with other passengers to guaranteeing a ride in a premium vehicle.

Financials:

Ticker Symbol:	UBER
Latest Financial Data:	August 2025
Revenue:	$12.7 billion
Assets:	$14.107 billion
Liabilities:	$32.352 billion
Return on Equity:	67.48%
Earnings:	33.5%, quarterly, year over year
Number of Employees:	31,100

Key Executives:

Dara Khosrowshahi:	Chief Executive Officer
Andrew Macdonald:	President and Chief Operating Officer
Prashanth Mahendra-Rajah:	Chief Financial Officer

Jill Hazelbaker:	Chief Marketing Officer
Nikki Krishnamurthy:	Senior Vice President, Chief People Officer
Tony West:	Senior Vice President, Chief Legal Officer

Major Competitors:

Lyft
Waymo
DoorDash
Ola
Rapido

Business Analysis Opportunity

This project leverages a comprehensive dataset that contains 150,000 Uber ride bookings from the National Capital Region (NCR) of India for 2024. This data provides an end-to-end view of the ride-hailing process by capturing everything from initial booking requests to ultimate outcomes whether a trip was completed or cancelled.

We'll leverage this data to go beyond basic reporting and advance into diagnostic and predictive analytics. We focus on addressing two key challenges the business may face: minimizing revenue loss from cancellations and improving customer satisfaction. Through predictive modeling and in-depth analysis, we can uncover the drivers behind these losses and enable Uber to apply targeted strategies such as reducing cancellations, refining driver and rider matching, and enhancing the overall user experience.

Research Questions

India is the most populous country in the world, its capital is the most populated city in the country, and one of the most populous metropolitan areas in the world. Consumers use

Uber's portfolio of offerings in this bustling metropolis as a way to get around in a country with low vehicle ownership per household. We will research the following questions in this project:

RQ1: Can cancellations be predicted based on factors like vehicle type, distances, cancellation history, and ratings?

This question is critical because each cancelled ride represents a direct loss of potential revenue for both Uber and its drivers while also damaging user trust and platform reliability. Understanding and predicting cancellations will allow for proactive interventions.

RQ2: Which driver and trip-related factors are most associated with consistently high or low ratings, respectively?

Ratings serve as a clear measure of both customer and driver satisfaction, which are key predictors of user retention and platform loyalty. For drivers, low ratings can limit their earnings and access to incentives. For customers, consistently poor ratings may expose deeper platform service issues. The insights we gather from this exploration can identify the underlying factors that could assist in creating a data-driven roadmap for platform improvement to elevate service quality.

RQ3: How do Uber ride cancellations vary by location, and what spatial patterns emerge when examining different cancellation reasons?

Location is a fundamental part of transportation. Moving people from Location A to Location B is what Uber does. If patterns emerge from this analysis that indicate cancellations occur with greater frequency in particular areas and for specific reasons, it would allow Uber to improve their ride completion rates.

Hypothesis

H1: Ride cancellations are primarily driven by service inefficiencies and economic misalignment

We hypothesize that cancellations follow predictable patterns rather than occurring by chance. Factors that contribute to inefficiency are expected to have a strong positive relationship with cancellation rates. Additionally, we think that certain vehicle types and pickup locations are likely to serve as strong predictors, as they influence both the economic incentive for drivers and the overall convenience for riders.

H2: Ride reliability and service smoothness are the foundational elements of high ratings

We hypothesize that ratings are heavily dependent on the consistency and quality of the core service. We expect that factors that introduce friction, such as longer pickup times, cancellations (from either party), and high fare values, will drive low ratings.

H3: Booking efficiency depends on where passengers are going to or coming from

Whether or not a booking is completed depends on geographical factors. Does the availability of vehicles in certain parts of the region affect whether rides are cancelled or completed? We intend to determine that through testing this hypothesis.

Data

Driver Perspective: Captures behavioral and performance data, including Driver Cancellation Reasons and operational metrics such as Avg VTAT.

Customer Perspective: Reflects user behavior through Cancellation flags and reasons, Payment Methods, and the Customer Ratings they assign.

Ratings: Both Driver and Customer Ratings act as key indicators of service quality and serve as primary target variables in our analysis.

The Cancellations data is particularly valuable as it includes not only binary indicators (e.g., cancelled by customer or driver) but also categorical variables like customer and driver cancellation reasons that add qualitative insights to our quantitative models.

Beyond the existing variables, we will engineer additional features to enrich our analysis. Time-related fields can be broken down into hour-of-day or day-of-week patterns, booking value and ride distance can be leveraged to derive efficiency metrics, and we can also add historical context by aggregating features such as the average cancellation rate for a given pickup location.

Measurements

Cancellation Probability: Represents the likelihood that a booking results in a cancellation. This can be modeled as either a binary outcome (Yes/No) or a multi-class outcome (Completed, Cancelled by Customer, Cancelled by Driver). It serves as our primary indicator of revenue loss.

Satisfaction Score: Captured through Driver and Customer Ratings on a 1–5 scale. We will analyze both the raw scores and a derived binary indicator (High vs. Low satisfaction) for classification purposes.

Service Efficiency: Measured using variables such as Avg VTAT (average vehicle turn-around time) and Avg CTAT (average customer trip actual time). We will examine the distributions and their relationship with our target outcomes.

Feature Importance: We will assess the predictive power and relative influence of independent variables (e.g., Vehicle Type, Time, Pickup/Drop Location) on our primary targets: cancellations and satisfaction. This analysis will highlight the key drivers shaping performance outcomes.

Methodology

For question 1, we want to predict an outcome, in this case, when a ride cancellation is likely to occur. To achieve this, we should use a variety of machine learning models such as random forest, regression, or any other classification methods. The importance of using multiple different machine learning models lies in the fact that we need to verify the accuracy of their predictions.

For question 2, we are trying to find which metrics are most influential to the ratings for both customers and drivers to see what both look for to be satisfied. Similarly to question 1, we should use a variety of methods to ensure accuracy and minimize bias. A useful feature to utilize for this question will be the feature importance scores, assuming there is a suitably accurate decision tree model for the data. Additionally, we should look at some type of clustering model, such as k-means clustering, to see if there are any groups that indicate a good driver and passenger rating combination.

For question 3, we want to see if there is a clear correlation between location and cancellations before investigating if there is a further correlation between cancellation reason and location. We should use multiple heat maps starting with location vs cancellation to see if there are any notable correlations. If there is, we should move on to location vs cancellation reasoning, verifying if any location-based reasons are increasing or decreasing cancellations.

This dataset has no unstructured data, such as written reviews, and is composed of structured data, including both categorical and numerical data. Prior to any machine learning models, we should do an exploratory data analysis to look at the metrics individually and ensure data usability. We should look at areas such as how many of the cancellations were from drivers vs. how many from customers, and potentially make a couple of visuals based on the findings. This could be displayed as graphs, dashboards, or other visualization techniques. After a broader look at the data, we should begin a more in-depth analysis where we try to find relationships between metrics and answer our research questions through the use of more advanced methods such as decision trees, regression, k-means clustering, and any other appropriate approaches.

Computational Methods and Outputs

When it comes to the methods, we will likely use a variety of computational methods depending on which research question we are analyzing. For question 1, we will be trying to predict when a rating is likely to happen, so multiple classification models will be used to ensure accuracy and reduce the chance of one of them overfitting. Question 2 will rely on slightly different computational methods since we are trying to see the impact of the metrics themselves. We will use decision trees along with their feature importance scores, followed by a clustering model to have an extra layer of confidence in the accuracy of our analysis. Question 3 will be working with visualizations and heat maps to understand the geographical importance of the data. We are not limited to the models and methods listed, as we may discover that other models are more accurate during testing.

The outputs will look quite different from each other for each of the questions due to the differing analysis methods. Question 1 will be a number that is the likelihood of a cancellation based on the metrics, along with performance values for the accuracy of the different machine learning models. The output for question 2 will be the results from the feature importance scores output by the decision tree, which will be compared to the clusters from the clustering model, in addition to the performance scores for the decision tree and clustering model. Question 3's output will be multiple heat maps of varying complexity, starting with a broad location vs cancellations. This will then be broken down into a more complex heat map visualizing location vs cancellation reason.

Output Summaries

RQ1: Can cancellations be predicted based on factors like vehicle type, distances, cancellation history, and ratings?

Our analysis will identify the factors that most often result in cancelled rides. Through analyzing historical cancellations, we will be able to determine what will cause future cancellations. Our output will include a pie chart showing the reasons given for cancelled rides. Other charts will consist of:

- Cancellation rates by vehicle type
- Cancellations by time of day
- Cancellations by ride distance

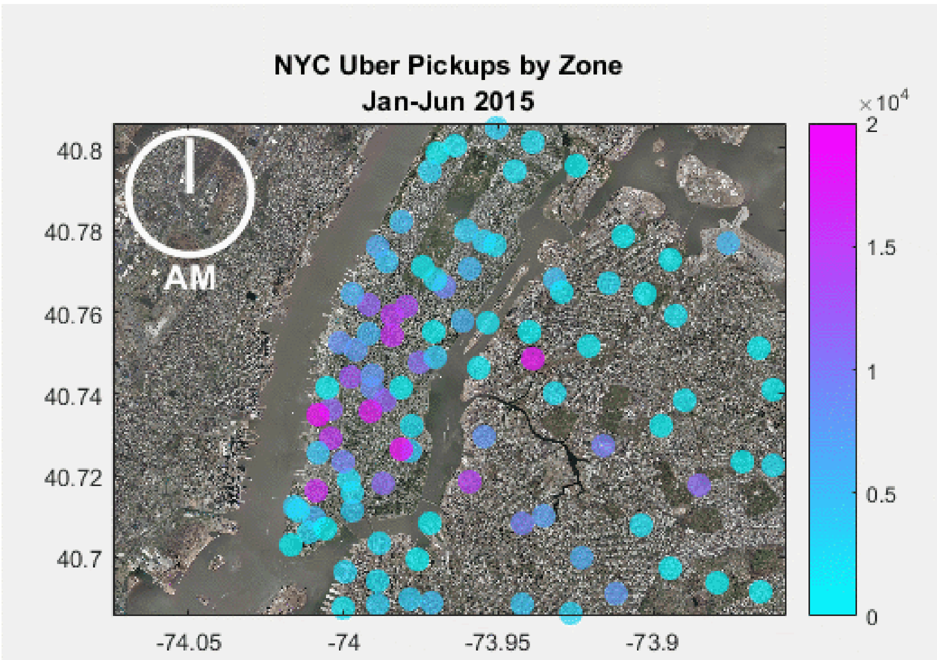
RQ2: Which driver and trip-related factors are most associated with consistently high or low ratings, respectively?

The analysis will identify which factors affect ratings. The ratings are based on a scale of 1-5. A bar chart similar to the example below will have ratings on one axis and variables on the other to show the correlation between ratings and the variables.

RQ3: How do Uber ride cancellations vary by location, and what spatial patterns emerge when examining different cancellation reasons?

The analysis will identify which locations are most likely to cancel rides or not. A heat map will show the rate of cancellations in the National Capital Region, similar to the map below of overall Uber pickups by time of day in New York City.

We will also explore what caused the cancellations and if there is a greater or lesser propensity of these reasons based on location. We will show such correlations with their own heat maps.



<u>Campaign Implementation</u>

Uber could use insight gained from each of these questions to implement changes to improve ratings, reduce cancellations, and, most importantly, increase profits. Cancellations mean lost money and, therefore, lower profit, so Uber could use question 1 to gain insight into what is causing the most cancellations. With this information, they could attempt to implement policies to minimize the factors causing the most cancellations. Additionally, more satisfied customers mean more return customers, and having satisfied workers is a great way to prevent an employee shortage. Using question 2, Uber could better understand what factors lead to high ratings and, therefore, satisfaction, and what factors lead to lower ratings. With this understanding, Uber could try to maximize the frequency of the positive factors while reducing the frequency of negative factors through policy change, training, or any other adjustments based on the results from question 2. Additionally, question 3 could help Uber determine which negative and positive factors are location-based. If there is a location with a high amount of cancellations for the same reason, Uber could tell drivers to avoid that area or make changes to how the service is operated in the area.

Literature Review

The transportation industry is continuously evolving with new technologies, methods, and approaches. In the modern world, rideshare organizations such as Uber, Lyft, and a variety of smaller companies have become massively popular. They allow people to easily order a ride from their phone and get to where they need to go with greater efficiency than traditional taxi or bus services. Uber is a ridesharing service that has expanded across all six non-Antarctic continents, making it widely available and well-known. When it comes to the success of rideshare organizations like Uber, ratings are crucial for both drivers and passengers. Not only do ratings allow users to assess the quality of their experience or decide whether to use a specific service, but they also allow the organization itself to evaluate the metrics that drive its success. This can be accomplished using a range of data exploration methods and analysis techniques.

A primary focus of this literature review is to clarify which factors most significantly affect customer satisfaction and cancellation within Uber's platform. The review identifies three main themes: (1) determinants of customer satisfaction in rideshare platforms, (2) how driving behavior and comfort directly influence satisfaction, and (3) cancellations as reflections of service quality and user trust. Exploring these, the literature seeks to uncover the underlying contributors to cancellations—such as location—and thereby enhance our broader understanding of rideshare data and its complexities.

Rideshare apps like Uber pride themselves on their customer satisfaction ratings and data collected that directly indicate user retention, driver incentives, and overall operational efficiency. Almaskati et al. (2025) explored this concept within a university rideshare system, identifying how trip characteristics such as ride distance, duration, month, and day of the week influence overall user ratings. This study was conducted using a Random Forest machine learn method, which demonstrated that longer rides tended to yield higher satisfaction ratings, possibly reflecting improved rapport between drivers and riders. While Almaskati et al. (2025) mainly focused on a university environment, their approach provides a framework for larger-scale applications such as Uber. The data-driven insights align with the broader view that satisfaction is multifactorial, dependent on the interplay of trip context, driver performance, and system-level dynamics (Almaskati et al. 2025).

Beyond trip logistics, rider satisfaction is strongly influenced by perceived comfort and driver behavior. Verma et al (2021) introduced *Ridergo*, an intelligent system that evaluates commuter comfort using smartphone sensor data during cab rides. It utilized continuous sensor data that would generate comfort scores on a five-point scale. The findings by Verma et al (2021) provide a novel perspective on Uber's feedback ecosystem. It also demonstrates that comfort is a highly personalized construct.

Cancellations by ridesharing can be attributed by an even specified range of variables such as weather, time waiting, and price mechanisms. Brodeur and Nield (2018) conducted an extensive empirical analysis encompassing taxi, Lyft, and Uber rides in New York City. Their focus was on examining how weather shocks influenced ride volume and service behavior. Their results identified an increase of approximately 22% during rainy periods, compared to only a 5% increase for taxis during the same weather period. Such external shocks can significantly impact the rider's experience. Brodeur and Nield's (2018) findings underscore how Uber's algorithmic pricing variants may inadvertently affect customer satisfaction when perceived as unfair or unpredictable. A key indicator of cancellations if acceptable external conditions negatively impact the consumer's experience.

Despite advances in understanding the determinants of satisfaction, the literature reveals several gaps relevant to Uber-specific research. Current studies often isolate a single dimension or focus solely on satisfaction measurements, which are frequently measured post hoc (via ratings) rather than dynamically predicted during rides. The opportunity presented is to build predictive models that aim to find correlations between the aforementioned factors and user cancellations. By mapping these correlations spatially, such as through heat maps, researchers can identify geographical patterns of dissatisfaction and cancellation, leading to targeted interventions for service improvement.

For a data science project, integrating these insights supports the development of predictive analytics that quantify and visualize customer experience. Together, these perspectives provide a comprehensive framework for exploring the relationships between satisfaction and cancellation within Uber's ecosystem. The resulting models not only enhance understanding of Uber's service but also provide actionable intelligence for improvement. If successful, reducing churn within Uber's rideshare market becomes a reality.

Final Research Questions

Our research followed the three core questions outlined in our project plan which we built to fully address the operational challenges of cancellations and user satisfaction on Uber's platform. This framework proved both solid and comprehensive, allowing us to move from predictive diagnostics to prescriptive spatial insights while examining how service efficiency, user experience, and geography interact.

Research Question 1: Can "Incomplete" rides be predicted based on factors like vehicle type, distances, cancellation history, and ratings?

Research Question 2: Which driver and trip-related factors are most associated with consistently high or low ratings, respectively?

Research Question 3: How do Uber ride cancellations vary by location, and what spatial patterns emerge when examining different cancellation reasons?

Exploratory Data Analysis

The data for this project is comprised of structured data from Kaggle.com and provides detailed information regarding the Uber rideshare service in the National Capital Region of India. The dataset is made up of 150,000 booking instances. Each booking represents a scheduled ride, with a majority being completed and a minority being cancelled. There are 21 total variables within the dataset, including time and date information for each instance. Three of these variables state whether or not a booking was cancelled or not completed, followed by another column containing the reason for the incomplete booking. We have removed these redundant variables for efficiency.

The variables we will be using are as follows:

- **Date:** Date of the booking (instance)
- **Time:** Time of the booking (instance)
- **Booking ID:** Identification number associated with the individual booking
- **Booking Status:** Tells if the booking is completed, cancelled by customer, or cancelled by driver
- **Customer ID:** Identification number associated with the individual customer (passenger)
- **Vehicle Type:** tells if the vehicle for the booking is a mini (compact car), sedan, auto (auto-rickshaw), eBike/Bike (back of a motorcycle), UberXL, or premier sedan
- **Pickup Location:** Location that the passenger was picked up by the driver
- **Drop Location:** Location that the passenger was dropped off by the driver
- **Avg VTAT:** Average time it takes the driver to reach the pickup location (in minutes)

- **Avg CTAT:** Average time it takes the driver to reach the dropoff location after picking up the passenger (in minutes)
- **Reason for cancelling by Customer:** Null, unless a ride is cancelled by the customer. Reasons are grouped into a small list of options
- **Driver Cancellation Reason:** Null, unless a ride is cancelled by the driver. Reasons are grouped into a small list of options
- **Incomplete Rides Reason:** If a ride is accepted by a driver, but not completed. Reason the ride was incomplete is selected from a small list of options
- **Booking Value:** Total cost of the ride in Indian Rupees
- **Ride Distance:** Distance from pickup to dropoff location (in km)
- **Driver Ratings:** Driver's averaged rating at the time of this booking (1 - 5 scale)
- **Customer Rating:** Customer's averaged rating at the time of this booking (1 - 5 scale)
- **Payment Method:** Tells if the rider paid with UPI, cash, credit card, Uber Wallet, or debit card

We will create the following variables:

- **Day of the Week:** Transform dates into days of the week
- **Weekend:** Indicate if the booking occurred on a weekend or weekday
- **Rush Hour:** For identifying peak hours of travel

For some basic exploration of the data, we checked to see if there was any direct relationship between the type of vehicle and the driver rating. Ratings are determined by averaging a driver's most recent rides. As such, we presume the rating listed with the booking was achieved using the vehicle type listed on the booking. This can help us see how much of an impact the type of vehicle will have on the rating of the driver and could lead to other potential discoveries such as how comfort or space impacts rating. Creating a quick table to show the average rating based on vehicle produced the results below:

Vehicle Type	Avg. Driver Rating (1-5)
Auto (Rickshaw)	4.232369
Bike (Back of motorcycle)	4.230056
Mini (Compact car)	4.227694
Sedan	4.231812
Premier Sedan	4.234865
Uber XL	4.238340
eBike	4.225614

Overall, the average ratings are very close between each class of vehicle only differing by 0.0127 between the highest average (Uber XL) and the lowest average (eBike).

Next, as shown in Appendix A, we built a sunburst chart illustrating the share of cancelled versus completed rides across the vehicle types. Despite differences in total ride volume among vehicle types, the ratio of cancellations to completed rides remains relatively consistent.

Appendix B shows us the distributions of booking value. It is heavily right-skewed, indicating that most rides are relatively low cost (~$11USD), with a gradual decline for booking values above 1000 rupee point.

In contrast, Appendix C shows that ride distances are more evenly distributed. Uber India rides cover a wide range of trip lengths. Rides less than 20 kilometers are more common, but not excessively so. As Uber uses a formula involving distance to calculate costs, and shown in the previous chart, the part of the formula that factors in vehicle type must vary substantially.

Both driver and customer ratings are concentrated between 4.0 and 4.8 ratings. In Appendix D, the distribution of driver ratings shows that the most common ratings for drivers are 4.2 and 4.3. Some cursory evaluation allows us to label the Top 10 percent of drivers as having a rating of 4.8 and above. The top quarter are 4.6+ and the top half are 4.3+. The bottom 10% of drivers are rated at 3.6 and lower, while the bottom quarter have 4.0 and lower ratings.

Appendix E, by contrast, shows higher ratings for Uber passengers in India, suggesting a difference in ratings based on perspective. While only 10% of drivers had a rating of 4.8+, 28% of passengers held this rating. Half of all riders carried at least a 4.5 rating. This carries through on the bottom of the spectrum, with the lowest 10% of riders being rated at 3.9 and below. The bottom 28% had ratings of 4.2 and lower.

In Appendices F and G, we've looked into temporal details of the bookings. Appendix F shows the average bookings for each day of the week. We can see that bookings occur rather evenly across the week. The variation between the busiest day(Monday) and the slowest day (Thursday) is only 2%.

Appendix G displays the Ride Status Distribution by Hour. We can see that the peak hours for Uber in the New Delhi area are 10am and 6pm. Booking reliably picks up every day at 5am, having a peak at 10 o'clock, before dipping and having an even greater peak during the 6pm hour. This visualization also shows that cancellations and incomplete bookings rise and ebb in a similar fashion.

We wanted to determine if payment methods had a correlation with customer ratings (i.e., how the driver rates the passenger), so we calculated the average customer rating for each payment method. The following table shows that the average customer rating is fairly even across payment methods, suggesting no preference for drivers. Appendix H shows a more detailed breakdown of distribution of customer ratings based on payment method. Again, the percentages are very evenly balanced across the different ratings which would suggest the payment method does not significantly impact how drivers rate customers.

Payment Method	Customer Rating
Cash	4.405369
Credit Card	4.408058
Debit Card	4.405753
UPI	4.402137
Uber Wallet	4.408424

Appendix I shows the top 10% of pickup locations with the highest number of customer cancellations. The most frequent reasons across these areas are "Wrong Address" and "Driver is not moving towards pickup location." This pattern may indicate that operational or geographic challenges, such as heavy traffic, are preventing drivers from reaching customers quickly. Additionally, the high number of cancellations due to “wrong address” could suggest that the platform’s location accuracy or address confirmation features are not prominent enough for users.

Finally, Appendix J shows the average monthly cancellations for both drivers and customers. Drivers cancel bookings at a rate of more than double that of customers. The respective rates do differ on a month-to-month basis, with some of the lowest average months of customer cancellations occurring during months where driver cancellations are among the highest of the year. While there are slight fluctuations, driver cancellations consistently occur more frequently than passenger cancellations.

Methodology

RQ1 – Can cancellations be predicted based on factors like vehicle type, distances, cancellation history, and ratings?

When trying to predict cancellations we are largely using multiple classification methods with the target (y) variable being the cancellation status. This status is already available in the dataset so we do not have to create any new variables for it. The classification can be shown two different ways using the data, with the simplest being either cancelled or completed, and the other being canceled by customer, canceled by driver, or completed. In preparation, we derive additional variables, Day_of_Week, Is_Weekend, Is_Rush_Hour from the original "Time" and "Date" variables. This allowed us to gain further insight into the specifics of when orders are cancelled.

In addition to the derived variables, there are variables "Reason for cancelling by Customer", "Driver Cancellation Reason", "Incomplete Rides Reason", "Vehicle Type", "Pickup Location", "Drop Location", "Avg VTAT", "Avg CTAT", "Booking Value", "Ride Distance", "Driver Ratings", and "Customer Rating" that help us to classify when a ride is likely to be cancelled (This is our X, also referred to as feature variables).

To prevent and check for overfitting, we use multiple different types of machine learning models allowing us to compare accuracy and verify that a model works on both a train and test set of data. In addition to the accuracy of the model, we plot the ROC curve and find the area underneath. This makes a comparable metric between the models and adds an extra layer of verification that a certain model is not overfitting.

<u>Modeling Techniques:</u>

- Modeling Technique 1: Logistic Regression

    - Works by using a logistic formula to predict the probability of a **binary** (0 or 1 / Yes or No) classification. We use it to establish a baseline understanding of how each predictor variable affects the possibility of a cancellation.
- Modeling Technique 2: Random Forest
    - Works by several decision trees in the training phase that are made of multiple random subsets of the training data and determines the most likely class. We use this to handle non-linear relationships and interactions between variables and to provide initial feature importance scores.
- Modeling Technique 3: Gradient Boosting
    - Works by making several models that are trained to correct the errors of previous models before producing a prediction for the class. We use this for high predictive accuracy and a different approach to capturing patterns.

RQ2 – Which driver and trip-related factors are most associated with consistently high or low ratings, respectively?

To identify factors that influence ratings, we are approaching the problem with both regression and classification methods. The target (y) variable for this is the driver rating value which our dataset already provides on a 1-5 scale. We plan to use this rating in two ways: first as a continuous variable for regression to predict a driver rating score for a booking, and second by creating categorical groups that driver ratings will fall into to predict the likelihood of a high rating.

The feature variables that we will be using for this analysis comes from two main categories: trip-related factors and efficiency metrics. Trip-related factors will include variables such as “Vehicle Type”, “Ride Distance”, “Fare”, “Pickup/Drop off locations” while efficiency metrics will consist of “Avg VTAT” and “Avg CTAT”.

To ensure we get a comprehensive and reliable answer, we will use two modeling techniques that complement each other. This will allow us to see the factors from both a supervised and an unsupervised learning perspective.

<u>Modeling Techniques:</u>

- Modeling Technique 1: Decision Tree / Random Forest
  - We use this to identify the most significant factors that split data into the rating buckets. This will be crucial for this research question as it directly provides feature importance scores which will tell a features influence on the rating outcome.
- Modeling Technique 2: K-Means Clustering
  - This unsupervised technique will group similar trips together based on their feature variables without using the rating data itself. We’ll use this method to identify natural groupings in the data (clusters) to best see ideal and problematic trips. Analyzing the profile of these clusters will help to validate the discoveries made by the Decision Tree and to understand the conditions of features that drive user satisfaction.

RQ3: How do Uber ride cancellations vary by location, and what spatial patterns emerge when examining different cancellation reasons?

For this analysis, our target (y) variables are “Booking Status” and “Reason for cancelling by Customer/Driver Cancellation Reason/Incomplete Rides Reason”. Our main features will be “Pickup Location” and “Drop Location”. Rather than any machine learning methods, we will use heatmaps. A series of visualizations will be created using the location names to uncover patterns in cancellation reasons.

Analytical Techniques:

- Analytical Technique 1: Cancellation Rate Heat Map
    - We will create a heat map that visualizes the cancellation rate across different locations. This will identify any location "hotspots" and "coldstops" across the region.
- Analytical Technique 2: Reason-Specific Heat Map
    - After identifying large cancellation "hotspots" from the first map, we will dig deeper into the reasoning for such cancellations by looking into the top 10% of locations with the most cancellations. This will allow us to go beyond just where cancellations are happening and reveal *why* they happen there.

Analysis

RQ1 – Can incomplete rides be predicted based on factors such as time of day, ride distance, and booking value?

Initially, we identified "Cancelled Rides by Customer" and "Cancelled Rides by Driver" as target variables that could be useful for Uber. However, due to insufficient data in rows with a cancelled booking status, we excluded them from subsets and our learning models. As a result, we instead focused on predicting Incomplete rides using operational and customer-related features such as time of day, ride distance, and booking value.

To explore this question, six classification models were tested using the Uber Ride Analytics 2024 dataset. The goal was to classify each booking as either "Completed" or "Incomplete" based on operational and customer-related variables.

Models Used:

1. Logistic Regression: A linear baseline model using the logistic function to predict "Incomplete" probability. This model helps interpret how each feature influences the likelihood of an incomplete ride.
2. Random Forest (Baseline): An ensemble of decision trees that captures non-linear relationships and interactions between variables.
3. Random Forest (Tuned): A refined version of the baseline forest using tuned hyperparameters (max_depth=10, min_samples_split=5, n_estimators=300) and class weighting (1:2.5) to address the imbalance of "Completed" and "Incomplete".
4. Random Forest + SMOTE: Applied Synthetic Minority Oversampling (SMOTE) to

increase the representation of incomplete rides in the training data.

5. Gradient Boosting: Sequentially fits trees that correct previous errors, producing strong performance with fewer overfitting issues.
6. XGBoost: An advanced, regularized form of gradient boosting that improves computational efficiency and generalization.

All models were trained on an 80/20 train-test split, with “Incomplete” as the target variable. They were evaluated using Accuracy, Precision, Recall, F1-score, and ROC-AUC. Class weighting and resampling methods better represented the minority “Incomplete” class.

Results

Method	Variable Selection Method	ROC/AUC	F1-score (Incomplete)	ROC/F1 Variance
Logistic Regression	Coefficient magnitude (absolute value)	0.8696	0.48	0.3896
Random Forest	Mean decrease in impurity (Gini)	0.8879	0.25	0.6379
Random Forest Tuned	Mean decrease in impurity (weighted 1:2.5)	0.8877	0.55	0.3377
Random Forest + SMOTE	Mean decrease in impurity (resampled data)	0.8815	0.52	0.3615
Gradient Boosting	Split gain importance	0.8872	0.13	0.7572
XGBoost	Gain-based feature importance	0.8865	0.3	0.5865

- Accuracy and ROC-AUC: All models achieved high ROC-AUC values (>0.87), indicating that the models can distinguish between Completed and Incomplete rides, though this performance is partly driven by the class imbalance. Boosting models slightly outperformed others in overall ranking ability.
- Recall (detecting “Incomplete”): Tuned Random Forest achieved the highest recall (≈0.96) while maintaining 0.80 accuracy, making it ideal for identifying likely incompletions before they occur.
- Precision: Gradient Boosting accurately predicts "Incomplete", but it only predicted "Incomplete" 100 times compared to the 10,000+ times it predicted "Completed". Such

an imbalance renders it unsuitable for our purposes, regardless of how strong its ROC scores are.

- Best balance: The RF + SMOTE and RF Tuned models offered the best compromise between recall and precision, with an F1-score of 0.55. The relatively low F1 score, compared to individual precision and recall values, reflects the continued impact of class imbalance. While some models captured a lot of incomplete rides, they also produced a high number of false positives.
- Feature importance: Across all models, Ride Distance and Average VTAT emerged as the only variables with any potential for predicting incomplete rides, as seen in Appendix X. Other features had minimal effect.

Conclusion

Among all models, the Tuned Random Forest (optimized parameters and 1:2.5 weighting) had the best overall trade-off between identifying "Incomplete" and minimizing false positives. Ride Distance and Avg VTAT were the only viable predictors of whether a ride was completed successfully or not. However, the imbalance in testing results, due to the difference in the number of completed and incomplete rides, makes it difficult to confidently say that we can predict "Incomplete" rides with significant accuracy.

Although ROC-AUC values were consistently high, the bar chart in Appendix Z shows comparatively low F1-scores across all models, indicating that we cannot reliably predict incomplete rides with the given data. This suggests that the dataset lacks the behavioral or contextual variables for accurate prediction.

RQ2 – Which driver and trip-related factors are most associated with consistently high or low ratings, respectively?

The goal is to see how the different features impacted the driver rating. The first step of the process was to make a regression decision tree and a classification decision tree that attempted to predict if a review is positive or negative depending on multiple metrics. The accuracy was measured using the mean squared error (MSE) which shows the average of squared distances and uses the formula below.

$$MSE = 1/n \cdot \Sigma (y_i - \hat{y}_i)^2$$

When running the classifier decision tree, and the regression decision tree, it was decided that accuracy metric would not be very reliable as the data is very heavily weighted towards positive reviews leading to the classification tree always picking positive for the review, heavily skewing the accuracy metric. This also led to the decision to use the regression decision tree over the classification tree for further analysis.

The regression models still struggled with the data so the decision was made to try out a gradient boosted regression model. This takes into account several weaker decision trees to build one stronger model. The result of this was an MSE of 0.440 which was decided to be acceptable for this dataset.

The next step was to get the feature importance which displays how much each feature impacted the model. The following visual in Appendix K displays the most important feature to rating at the top and the least important at the bottom.

After finding which features are important to the model, it is essential to know how these features impact the rating (i.e. positively or negatively). The SHapley Additive exPlanations (SHAP) library in Python added the ability to see how features contribute to the model in a more detailed manner than the feature importance metric. This library was used to create visualizations in Appendices L, M, and N.

In Appendix L, a red dot is a high value (for example: a red dot for Ride Distance would be a long distance covered during the trip) and blue is low value (for example: a blue dot for Ride Distance would be a short distance covered during the trip). Overall, a wider spread of the dots indicates a larger impact on the rating with more towards the negative meaning a negative impact and more towards the positive meaning a positive impact. This helps interpret in what way the important features are impacting the rating.

For the graph in Appendix M, when a point is in the negative x direction, this indicates a lower distance covered during the trip and vice versa for the positive direction. When a point is in the negative y direction, this indicates a negative impact to the drivers score with positive being a positive impact. When a point is red, it indicates a higher time for the driver to arrive at the pickup location (and longer wait time for the rider) with a blue point being a shorter amount of time.

Appendix M shows a lot more detail on how the features interact. There is a correlation when considering both ride distance and Avg VTAT (how long a rider had to wait for a driver to arrive). There is a larger concentration of red dots when the ride is shorter (negative x direction) and the Avg VTAT time is higher (The dot is more red than blue). This would suggest that there is a negative impact to the driver's rating (negative y direction) when a passenger has to wait a long time for a short ride.

The graph in Appendix N works the same way as the one in Appendix M, but to compare ride distance and booking value. A red dot would be a far distance covered during the ride with blue a short distance. A dot in the negative x direction would be a lower booking value with a dot in the positive being a higher booking value. A dot in the negative y direction would be a positive impact to rating with one in the negative being a negative impact to the rating.

Appendix N allows a more detailed examination of how the booking value and the ride distance work together to impact the rating. There is a trend that shows a massive drop for the location of the dots in the negative y direction and positive x direction (Negative rating impact and high booking value) plus more red dots concentrated in this area (shorter ride distance). This suggests that the driver's rating is negatively impacted when there is a ride that has a high booking value for a shorter distance covered.

After analysing the SHAP graphs and feature importance from the gradient boosted regression decision tree, we can gain some major insights into what factors are impacting driver's scores. The feature importance is displayed in the following table:

Feature / Metric	**Importance**
Ride Distance	0.180145
Booking Value	0.172292
Avg CTAT	0.158683
Drop Location	0.146429
Pickup Location	0.144665
Avg VTAT	0.140016
Vehicle Type	0.057771

The two most important features are ride distance and booking value but it is also important to know how these interact with the other important features and this is where the SHAP visuals become useful.

The SHAP library will automatically determine which of the other features is most useful to the feature that the user selects. In this case, the ride distance and booking value features were selected in separate code cells. The SHAP library determined that there was a relationship between ride distance, Avg VTAT (how long a rider had to wait for a driver to arrive), and rating plus a relationship between booking value, ride distance, and rating. This means that the Avg VTAT is a major contributor to the score received when also considering the distance covered during the ride. Additionally, the ride distance is a major contributor to the rating when considered with the booking value.

After looking at the features separately and then together, a more concise story can be put together that shows what the biggest impact is on a driver's ratings. The worst possible situation for a rating would be a long wait time for the driver to get to the location, a short distance covered during the ride, and an expensive booking cost. This would make sense logically as someone would not like paying a lot of money for a ride that was not long after waiting a long time for the driver to pick them up. The best situation for the driver's rating would be a short wait time, long distance covered, and a cheap booking value. That also makes sense logically as a customer would see the situation as getting more value out of their money if the ride is long, the wait time is short, and the cost is cheap.

Further analysis via a K-Means clustering plot groups trips based on Count of cancelled rides by Drivers vs. Riders over Average Ride Distance. Patterns such as longer trips did show a higher tendency to impact Ave CTAT on the back end. This represents rides with similar travel characteristics, helping to identify potential areas of issues consumers may experience. With proper analytics, it could help improve attributed to factors that lead to negative rider experiences. Appendix O illustrates the total variation within each cluster. This objective function helps reveal natural clusters in the data, such as users who cancel frequently with longer trips, users who cancel rarely, or intermediate groups with moderate behavior.

RQ3 – How do Uber ride cancellations vary by location, and what spatial patterns emerge when examining different cancellation reasons?

The cancellation rate heat maps in Appendix P revealed consistent patterns between pickup and drop locations, with both categories showing identical 25% average cancellation rates across all 176 locations analyzed. The top 10% of locations with the highest cancellation rates identified 18 locations in each category, with four locations (Chhatarpur, GTB Nagar, Karkarduma, and Keshav Puram) appearing as problematic for both pickups and drops. This shows that while cancellation issues affect both origin and destination points equally in aggregate, the specific location pain points still differ with 14 unique locations in each category.

There's only a slight 0.13% gap in the top 10% average cancellation rates between pickup (27.86%) and drop (27.99%) locations. This minimal difference suggests that cancellations aren't only driven by location-specific factors but rather by broader, platform-wide operations.

The reason-specific heat map analysis in Appendix R shows insights into why cancellations happen in the hotspots. By examining the top 10% of locations with the highest cancellation rates, we discovered that driver-initiated cancellations dominate both pickup (72.4%) and drop (70.9%) hotspots, with “Customer Related Issues” ranking high as a primary reason across locations.

The analysis also uncovers distinct patterns in cancellation reasoning between pickup and drop locations. While both share common top reasons like "Customer Related Issues" and "Too Many Passengers", unique patterns emerge: pickup hotspots show higher "Driver Personal/Vehicle Issues" (18.1%), whereas drop locations exhibit more "Customer Appeared Sick" cancellations (17.7%). The heat maps visually demonstrate that cancellation reasons maintain consistent distributions across different geographic hotspots, with driver behavior accounting for the majority of cancellations regardless of whether it’s a pickup or drop location.

PICKUP HOTSPOTS - CANCELLATION SOURCES	
Driver:	3044 (72.4%)
Customer:	1161 (27.6%)
DROP HOTSPOTS - CANCELLATION SOURCES	
Driver:	3045 (70.9%)
Customer:	1247 (29.1%)
DRIVER CANCELLATION REASON COMPARISON	
PICKUP - Top 3 Driver Reasons	
Customer Related Issues: 795 (34.5%)	
Driver Personal/Vehicle Issues: 760 (33.0%)	
Too Many Passengers: 748 (32.5%)	
DROP - Top 3 Driver Reasons	
Customer Related Issues: 790 (34.3%)	
Customer Appeared Sick: 758 (32.9%)	
Too Many Passengers: 753 (32.7%)	
CUSTOMER CANCELLATION REASON COMPARISON	
PICKUP - Top 3 Customer Reasons	
Wrong Address Provided: 268 (34.2%)	
Driver Requested Cancellation: 267 (34.1%)	
Customer Change of Plans: 248 (31.7%)	
DROP - Top 3 Customer Reasons	
Driver Requested Cancellation: 298 (34.8%)	
Driver Not Moving Towards Pickup: 280 (32.7%)	
Customer Change of Plans: 279 (32.6%)	

As we see in the Appendix R and the table above, customer cancellations mostly reflect reactions to service breakdowns rather than independent choices. The high frequency of "Driver Requested Cancellation" and "Driver Not Moving Towards Pickup" shows that customers are often canceling in response to driver unavailability or poor service. While "Wrong Address Provided" appears more frequently at pickups and "Customer Change of Plans" stays consistent across areas, the broader trend points to service reliability as the main driver of cancellations. The consistency of these patterns across locations reinforces that this isn’t a localized issue but is instead a network-wide operational problem that calls for improvement of driver behavior and overall reliability.

Ethical Recommendations

Our analysis of the Uber ride data provides valuable insights that can improve the platform, but it's important to consider the ethical implications of this work. The core goal is to utilize data to support drivers and customers, rather than creating unfair systems. Through proper analysis, we may provide insights into challenges that affect customer experience, market conduct, and corporate accountability.

The main issue surrounding the data, which doesn't account for instances where it relates to both driver and customer ratings, is the range of human error and bias. People can be inherently biased regardless of how hard they try not to be. Small things, such as a person's negative mood because they got up early or are getting off work, may lead to a lower rating for the driver due to a factor outside of their control. This relates to the dilemma of determining how much weight to give to driver ratings, as there's a chance that factors beyond a driver's control can negatively impact them. This can cause a driver to appear as a worse quality due to what is essentially a random chance, since the passenger's mood is difficult to factor into the analysis.

Additionally, the dataset contains information such as customer ID, pickup location, drop-off location, time, and date. If this information is not kept secure, it could be used incorrectly to track individuals' movement patterns. Not only could this be used to commit crimes, but it also could be used to identify a specific individual, which would essentially remove any anonymity associated with the information. The customer IDs don't contain names, but if someone is trying to identify a specific person, they can easily look at the pickup/drop-off location with the date and time to identify who a specific customer ID is associated with. Misuse of data privacy and algorithm ethics are on the minds of all consumers.

One consideration to be mindful of stems from ethical concerns surrounding aggressive underpricing strategies that may harm Uber’s drivers. Competitors may use the collected data to strategically lower their rates, potentially stealing their consumer base. The counterpart is how Uber offers transparency of internal problems(declining service quality, pricing, driver shortages), and if analysis plays into negative PR. Consumers want to know how their data is stored and used, and there is a risk of data leaks, misuse, or surveillance.

A positive impact of analysing this data is to improve both customer and driver experiences. The data can help identify and pinpoint specific issues that negatively impact both customer and driver experiences. In response to the data findings, Uber could implement changes to help both parties. Numbers can be analyzed in various ways. The impact on how to issue responses to improve customer satisfaction, driver treatment & labor practices, data safety, and then strengthen a better overall company culture and business practices is pivotal.

Challenges

A challenge we faced was an initial misunderstanding of the data. Due to the compressed timeframe for assembling a team and then finding a large dataset, we had to rely on the dashboard of the chosen dataset. However, we later found that the dataset's author incorrectly defined some of the variables. Where we believed we could assess a ride on the rating given at the end of it, turned out to be the ratings at the time of booking.

There were additionally several redundant columns for cancellations and incomplete rides. For these rides, there was also a dearth of data. They did not include quoted booking values, customer ratings, expected ride distance, or Average CTAT. We did not immediately notice this at first glance because the redundant columns were filled with null values. These limited the usefulness of our dataset. If we were to start over with the same project, finding a more complete dataset would be desirable.

RQ1 – Can cancellations be predicted based on factors like vehicle type, distances, cancellation history, and ratings?

The big challenge on this question turned out to be the lack of data when a booking was cancelled. The only data available for cancelled rides were temporal variables, vehicle type, and location. Through exploratory analysis, we found that vehicle types and time of day were very consistent across booking statuses. Location-based cancellations were handled in Research Question 3. We had to modify RQ1to "Can "Incomplete" rides be predicted based on factors like vehicle type, distances, cancellation history, and ratings?"

Overfitting was a difficult obstacle to overcome on this question. Due to the data mostly belonging to completed rides, it was impossible to ascertain if some variables correlated with cancellations. It also appears that the data was somewhat curated to contain balanced data of variables, rather than being a comprehensive or random collection of data. This deduction is based on the aforementioned “Vehicle Types” and temporal variables, and their consistency across booking statuses.

RQ2 – Which driver and trip-related factors are most associated with consistently high or low ratings, respectively?

The biggest challenge that we ran into when it came to the decision trees was the unbalanced data. A large amount of the reviews are clustered in the higher values of the 0 to 5 range. It would show high accuracy but the confusion matrix would tell a different story. The confusion matrix showed that the classification tree was picking all positive ratings but since there were not a lot of lower ratings, it would still think there was high accuracy. Using different parameters and features along with pruning the tree was attempted but still resulted in the same problem. The ultimate decision of this was to focus on the regression decision tree as it performed better and less biased.

RQ3 – How do Uber ride cancellations vary by location, and what spatial patterns emerge when examining different cancellation reasons?

While creating the heatmaps was technically straightforward, the main challenge was related to the lack of deeper context for the location data as well as vagueness of top cancellation reason codes such as “Customer Related Issues” which functions as a catch-all that obscures true and specific problems. We successfully identified where cancellations were clustered based on location name, but without data on local traffic patterns, infrastructure issues, or socioeconomic factors, our analysis isn't able to reach the depth that we'd prefer to fully explain the underlying "why".

Recommendations and Next Steps

- Obtain more balanced rating data
  - Find more comprehensive datasets that can explain a wider range of questions and account for additional context
- Try to find more specific factors that impact the trip
  - Factors such as weather, traffic conditions, driver/customer age, etc. could have a large impact but are not accounted for in this dataset.
- Additional questions we could ask.
  - Analyze repeat customers: If a customer ID pops up more than once in the data, are there consistent factors between them and other repeat customers?
  - Analyze customer or driver bias: Do some customers or drivers always rate low or high? If so, does this skew the data in any way?
  - Analyze if all ratings are fair: How do certain customer, regional, weather, traffic, etc. factors outside of the driver's control impact the score negatively?)
- Instead of just using the 1 to 5 rating, maybe a more sophisticated method to compare customer / driver satisfaction could be used.
  - Percentile rankings or rank a driver's rating compared to their daily average and look into days/metrics that don't align with the average. This could be done both when more positive or more negative than the average.

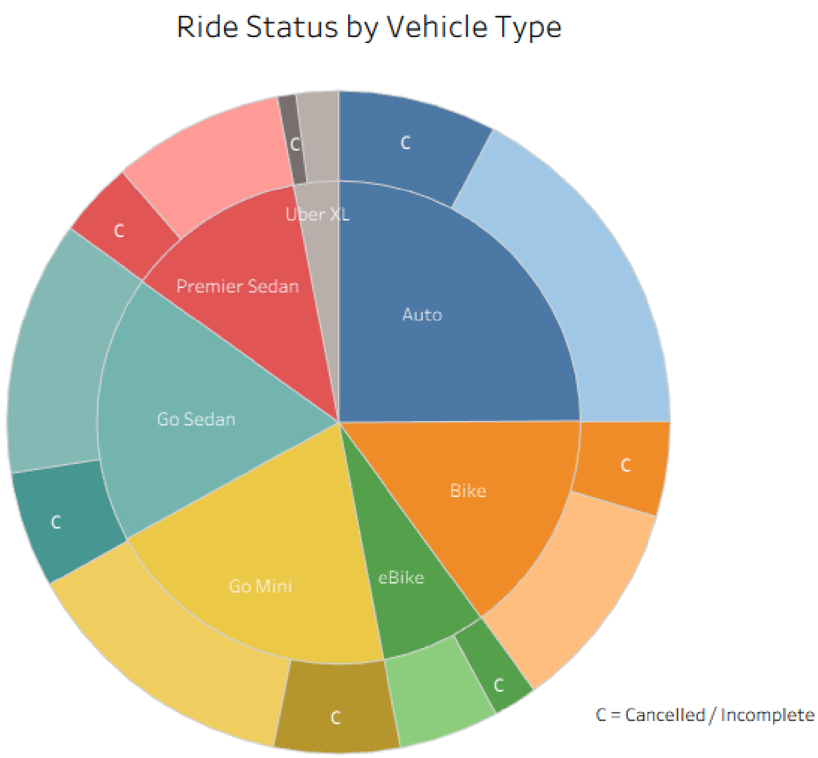
Source Code

The following Google Colab notebook was used as a collaborative file to do the machine learning and visualizations:

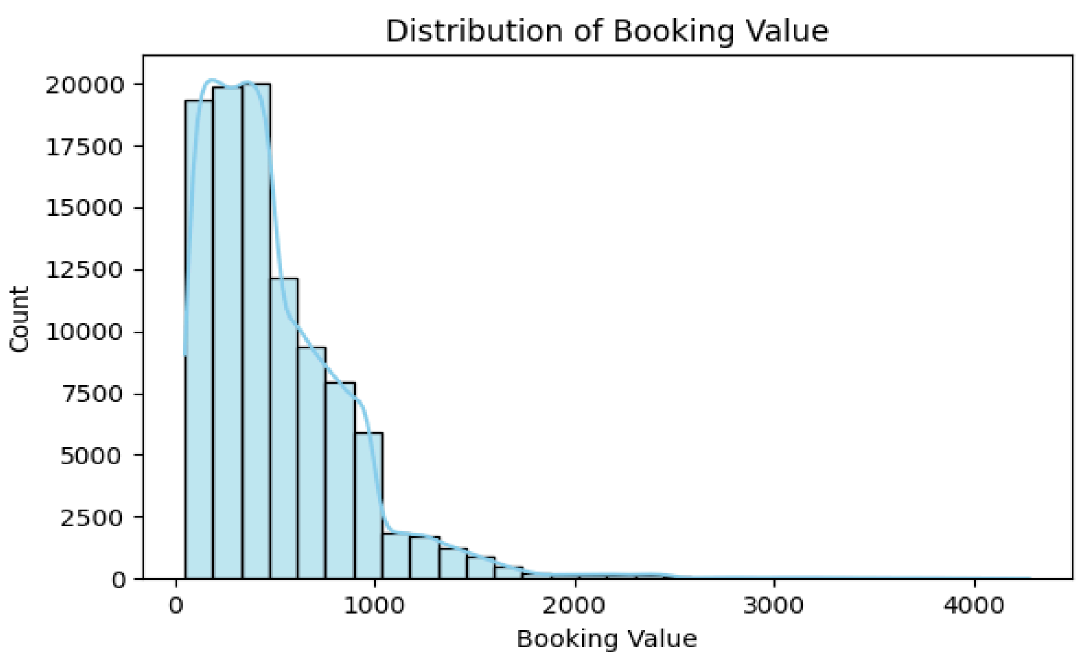
https://colab.research.google.com/drive/182rIBeUjhP4yJ5T-yZmdrj6UeFFm7Jjf?authuser=6#scrollTo=SCs4cHWF0b5A

Appendix

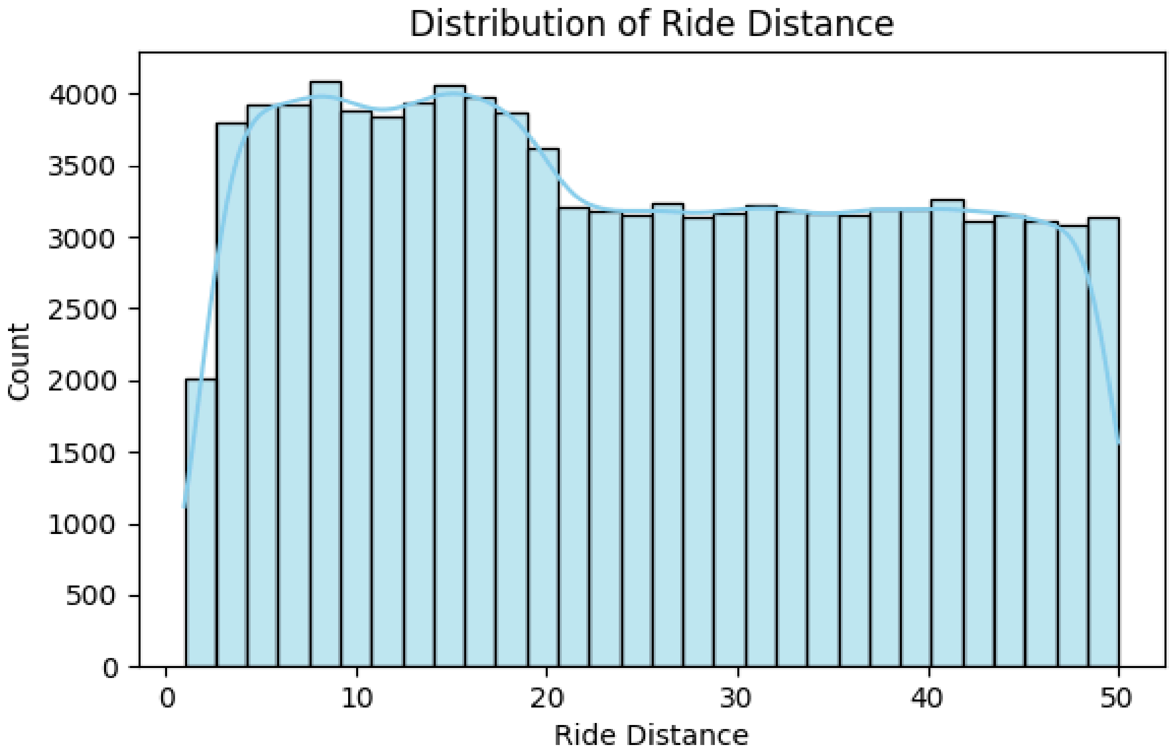
Appendix A: Cancellations/Incomplete Rides by Vehicle Type



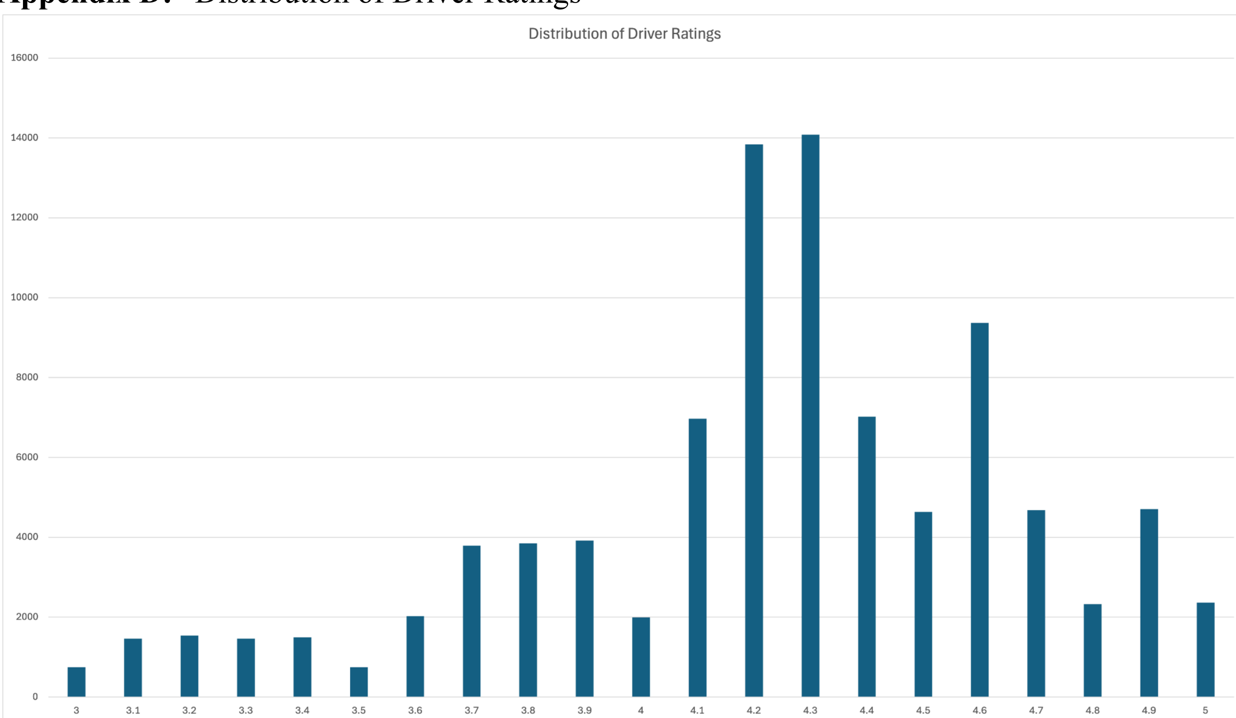
Appendix B: "Distribution of Booking Value"



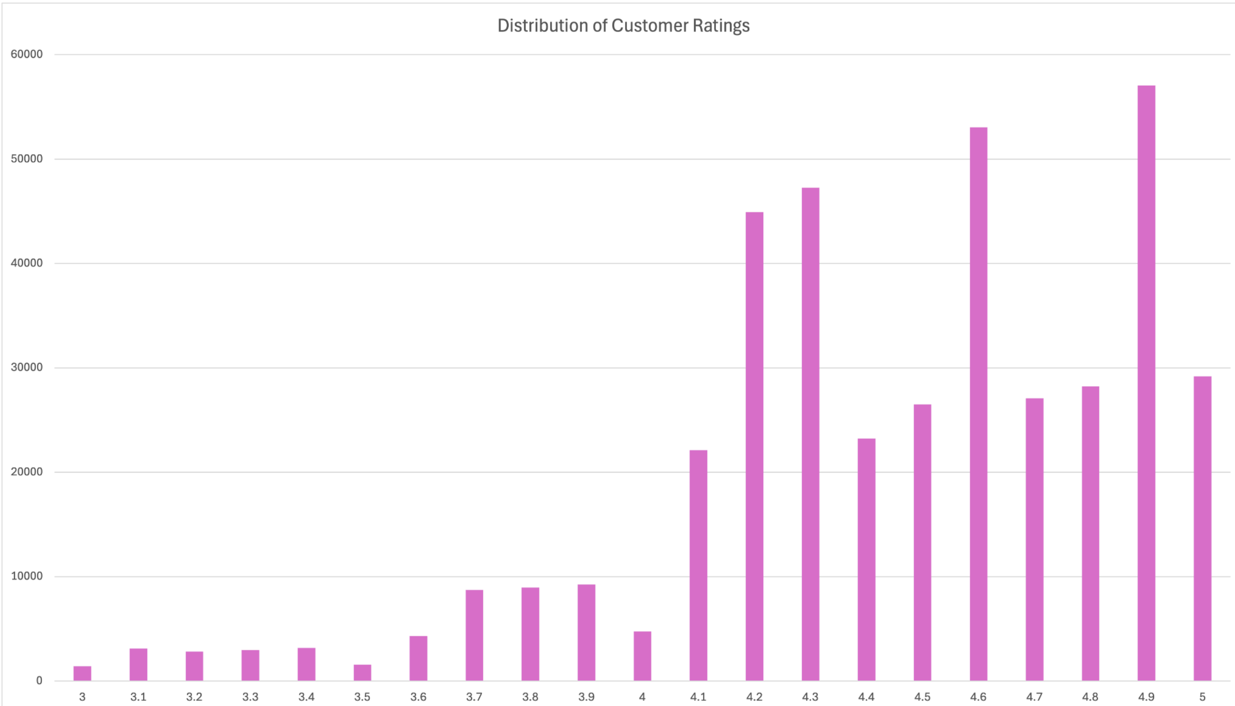
Appendix C: “Distribution of Ride Distance”



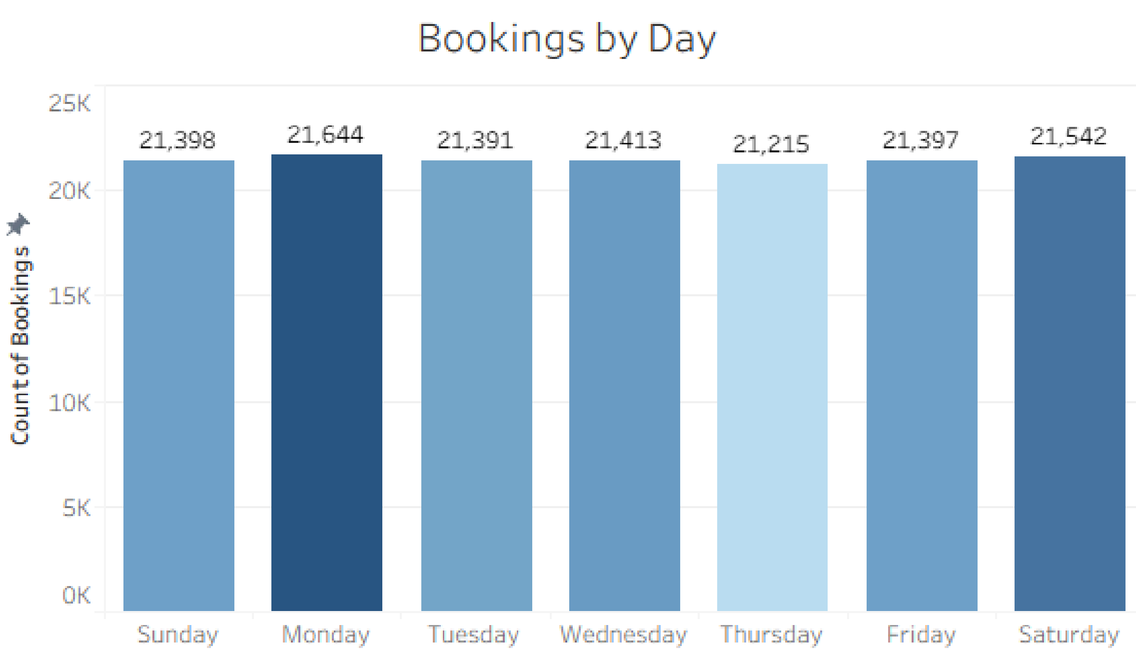
Appendix D: “Distribution of Driver Ratings”



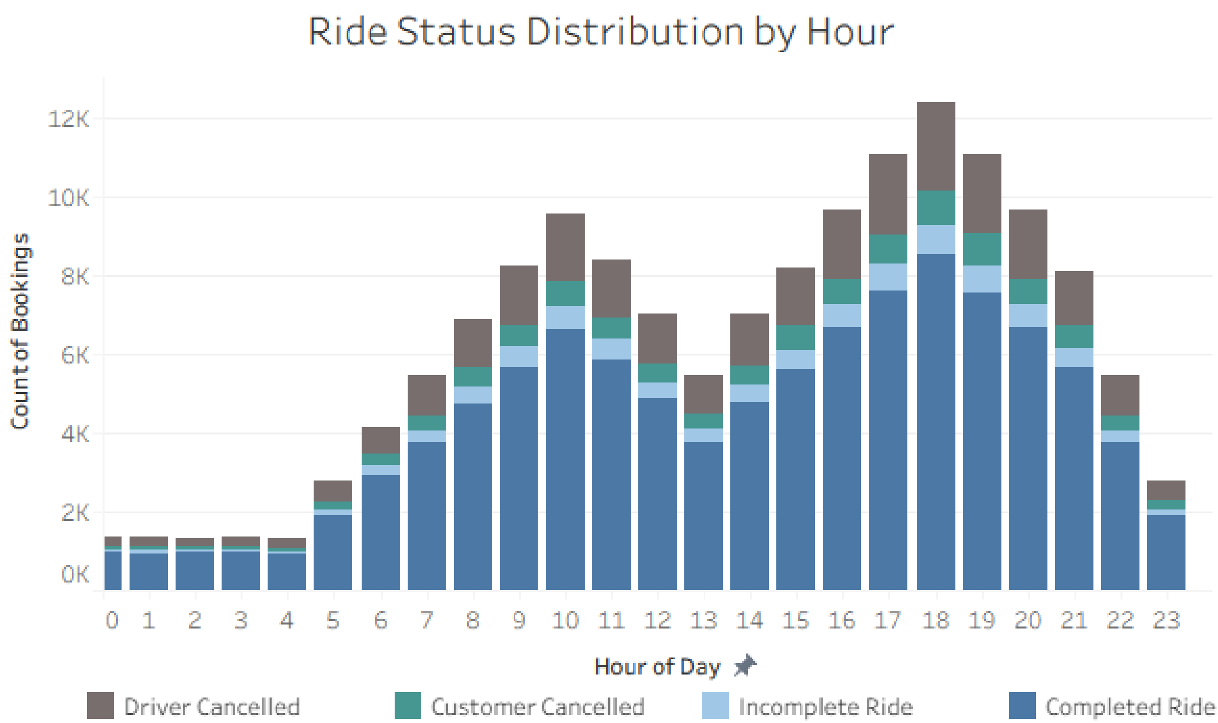
Appendix E: "Distribution of Customer/Passenger Ratings"



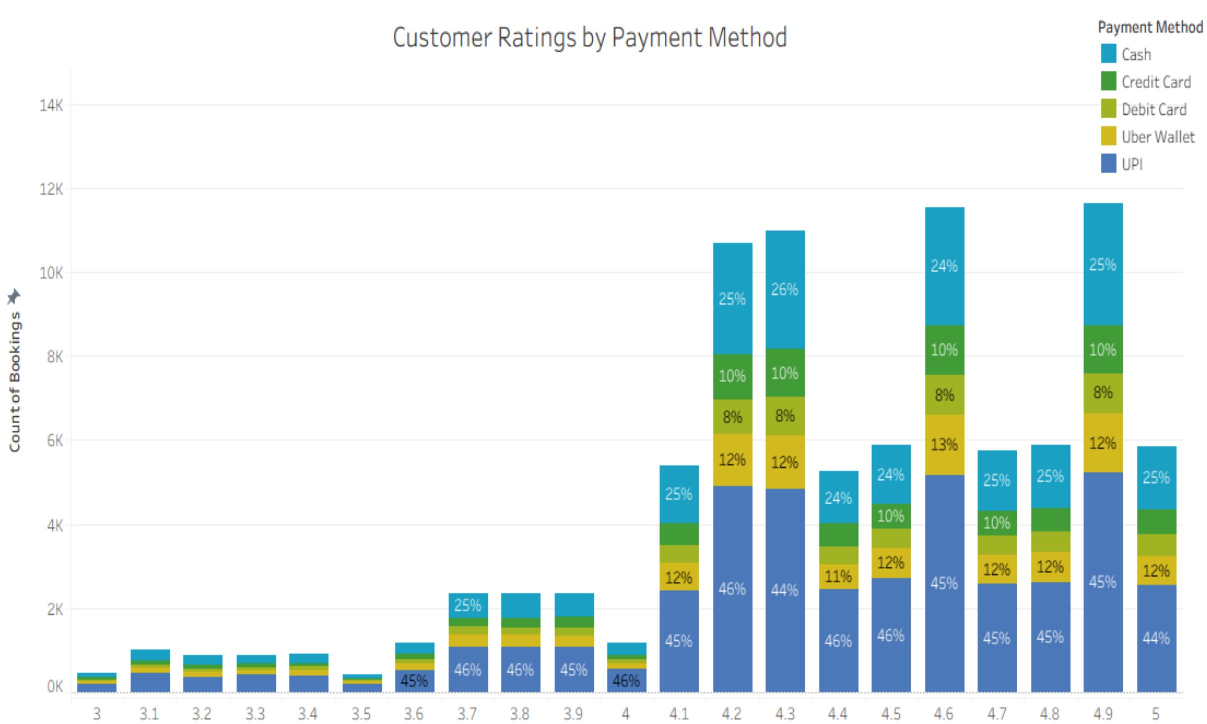
Appendix F: "Average Bookings per Day of the Week"



Appendix G: "Hourly Distribution of Completed and Incomplete Rides"



Appendix H: "Customer Rating by Payment Method"

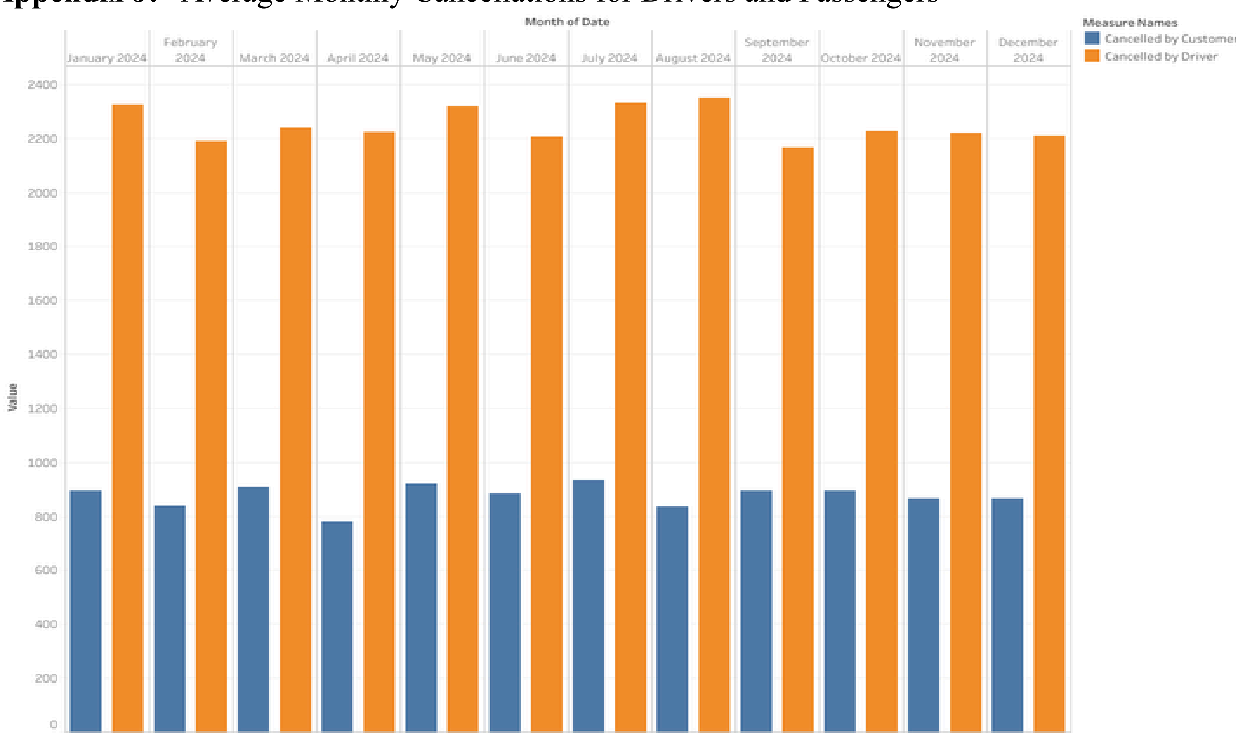


Appendix I: “Top 10% of Pickup Locations by Customer Cancellation Reason”

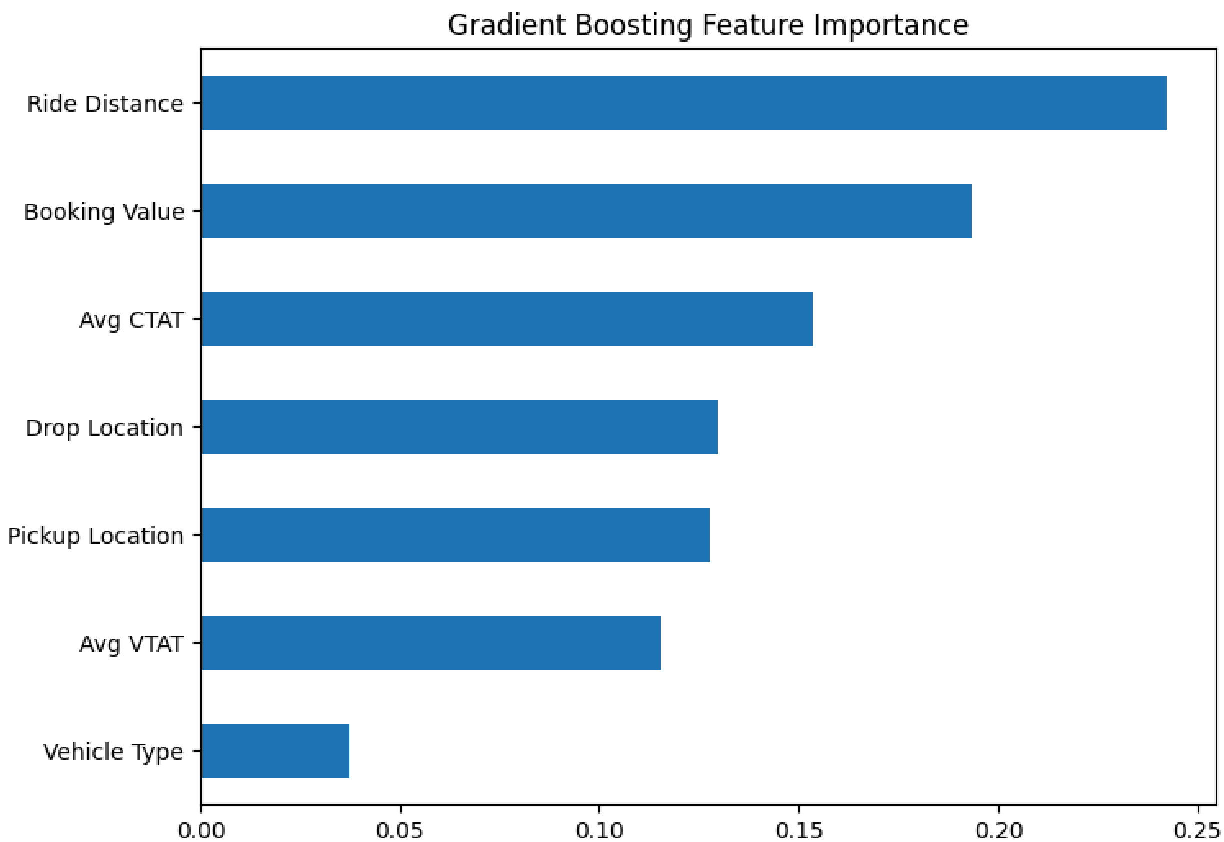
Top 10% Pickup Locations by Customer Cancellations

Pickup Location	Reason for cancelling by Customer: AC is not working	Change of plans	Driver asked to cancel	Driver is not moving towards pickup location	Wrong Address
Seelampur	8	20	14	18	21
Preet Vihar	12	17	19	14	18
Keshav Puram	7	13	19	21	19
Akshardham	7	18	17	18	18
Mansarovar Park	3	16	18	17	21
Arjangarh	11	13	20	17	13
Kashmere Gate ISBT	5	14	16	18	21
Greater Kailash	11	14	12	21	15
Badarpur	9	16	15	22	10
Greater Noida	8	18	10	16	20
Saket A Block	9	15	14	16	18
IGI Airport	6	17	15	19	14
Indraprastha	6	19	18	10	18
Badshahpur	9	19	16	11	15
Gurgaon Railway Station	7	18	17	14	14
Hauz Khas	12	12	21	14	11
Inderlok	9	11	16	15	19
Saket	7	16	15	20	12

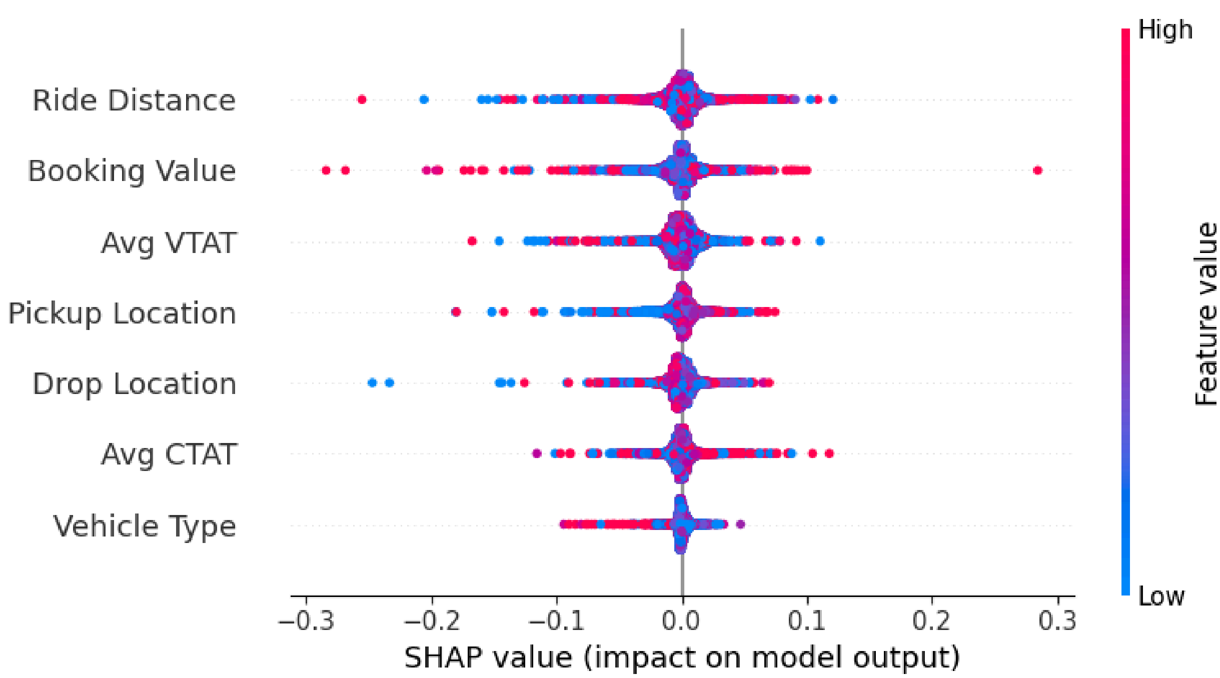
Appendix J: “Average Monthly Cancellations for Drivers and Passengers”



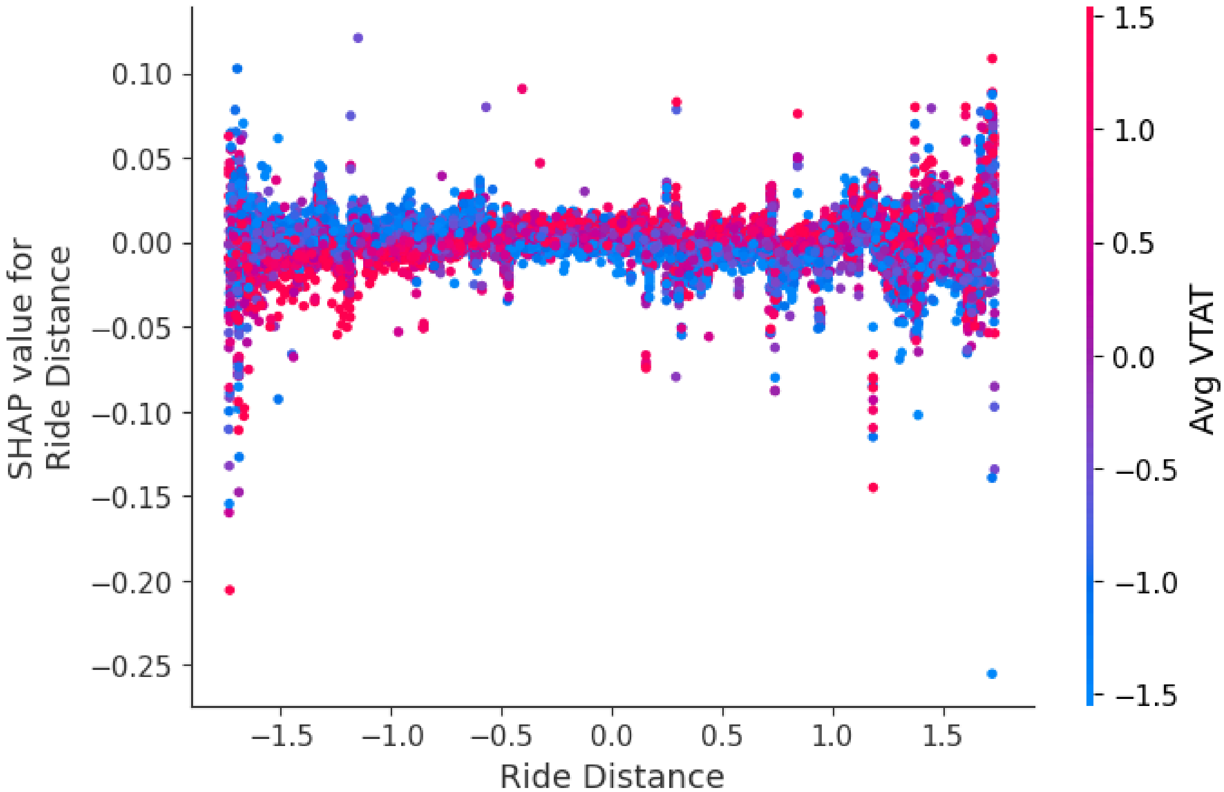
Appendix K: “Feature Importance from Gradient Boosted Regression Decision Tree”



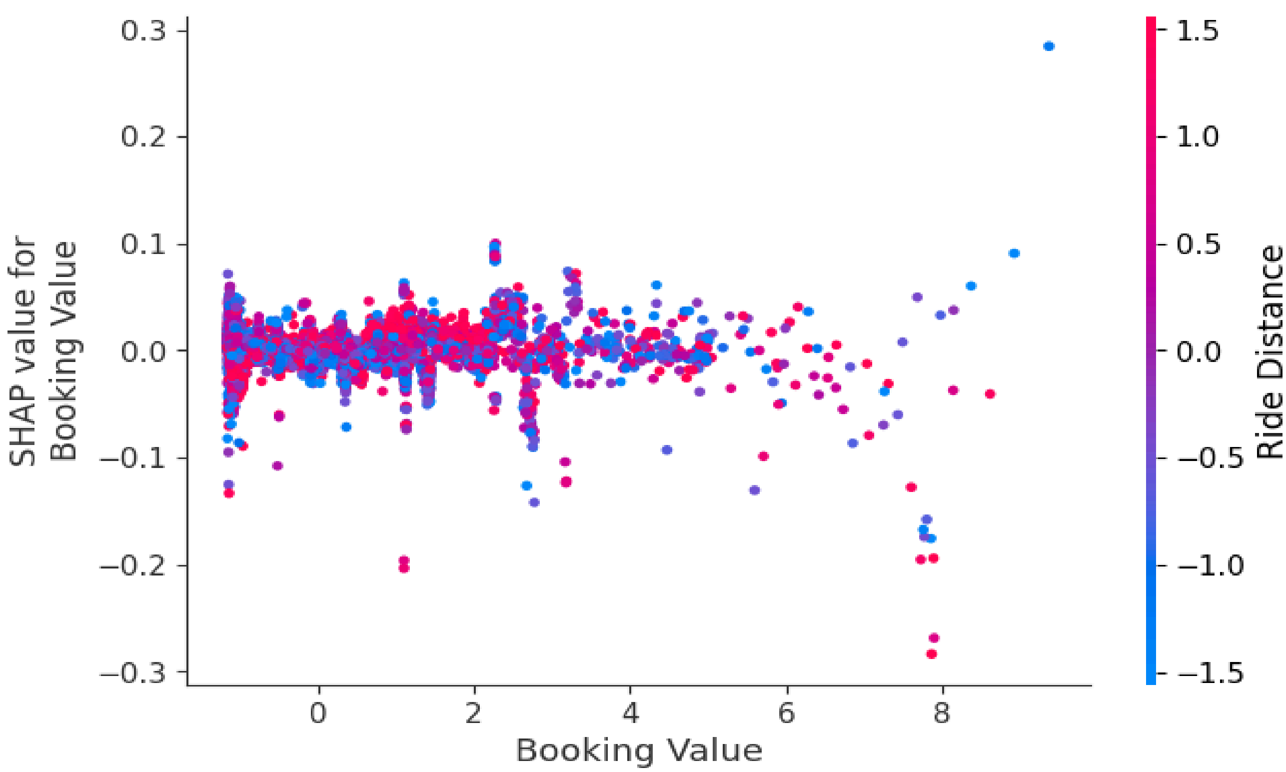
Appendix L: "Feature Impact on Final Model Based on SHAP Value"



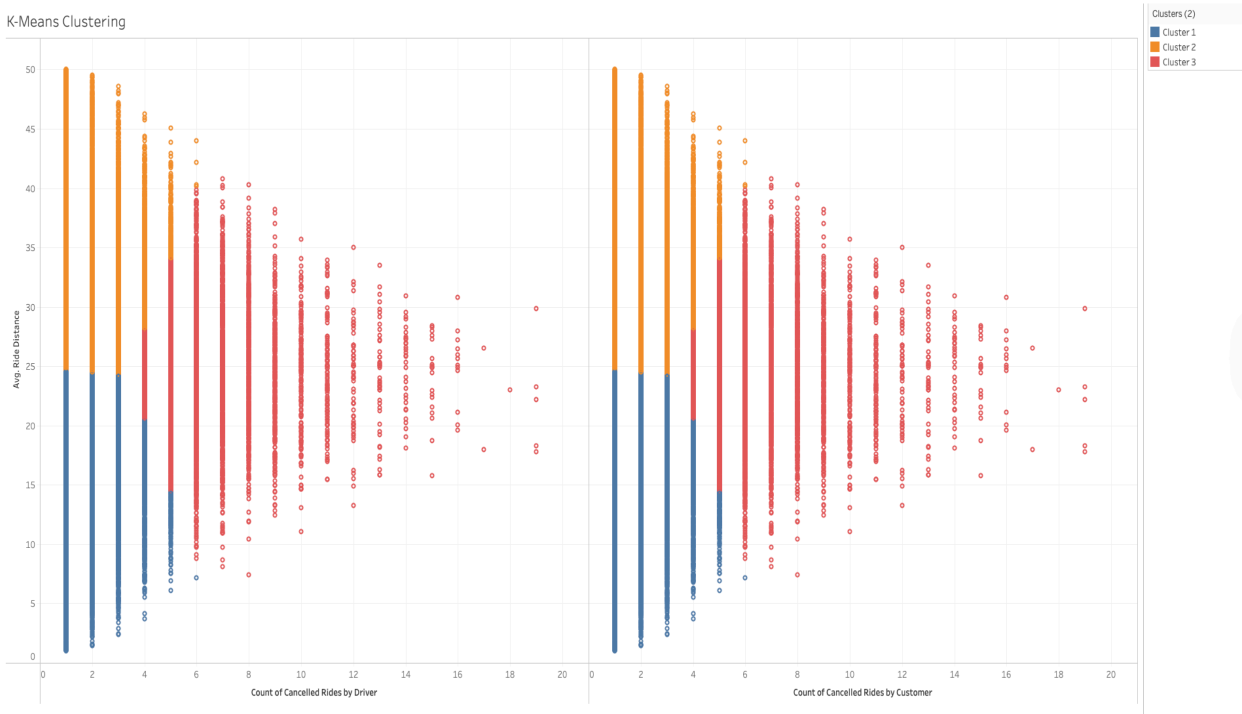
Appendix M: "Impact on Driver Rating Based on Ride Distance When Accounting For Avg VTAT"



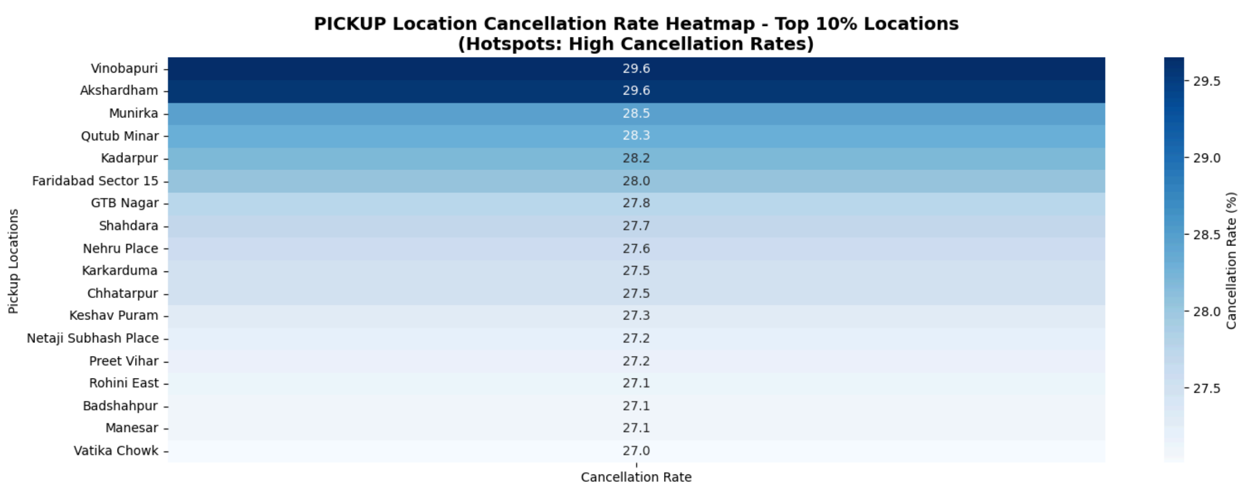
Appendix N: "Impact on Driver Rating Based on Booking Value When Accounting for Ride Distance"



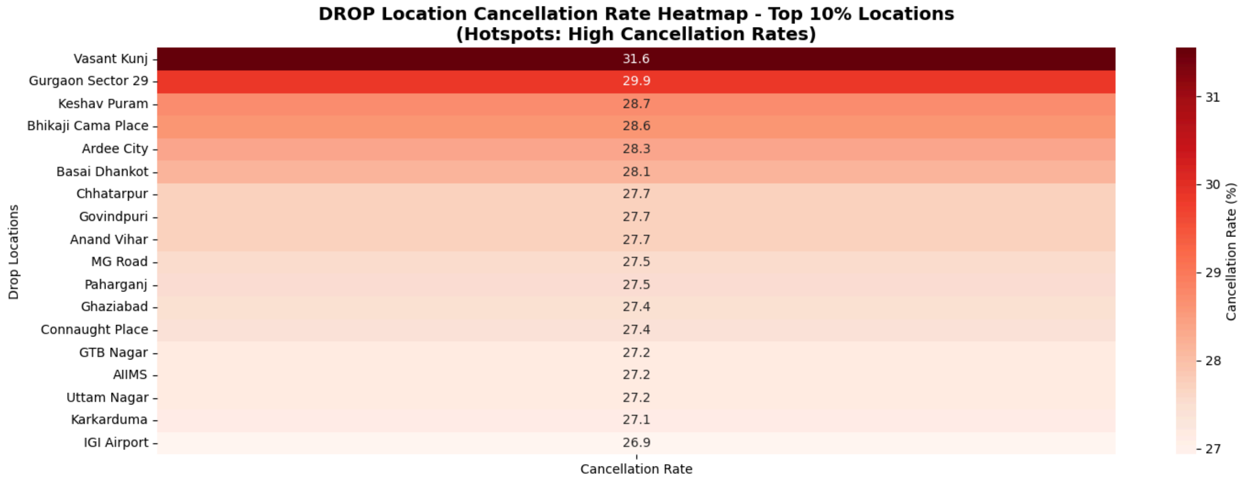
Appendix O: "K-means Clustering, Ride cancellation vs trip length and cancellation type"



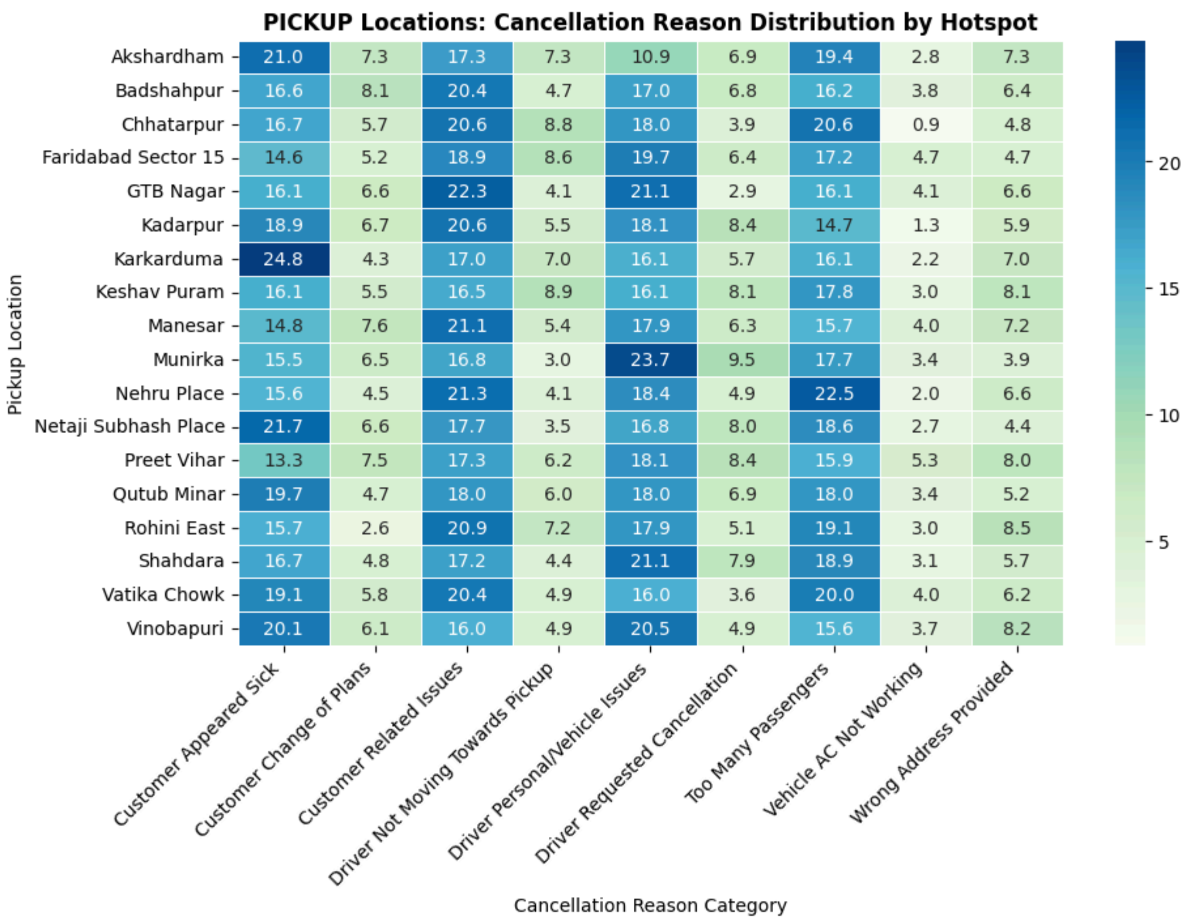
Appendix P: "Top 10% of Pickup Locations With The Highest Cancellation Rates"



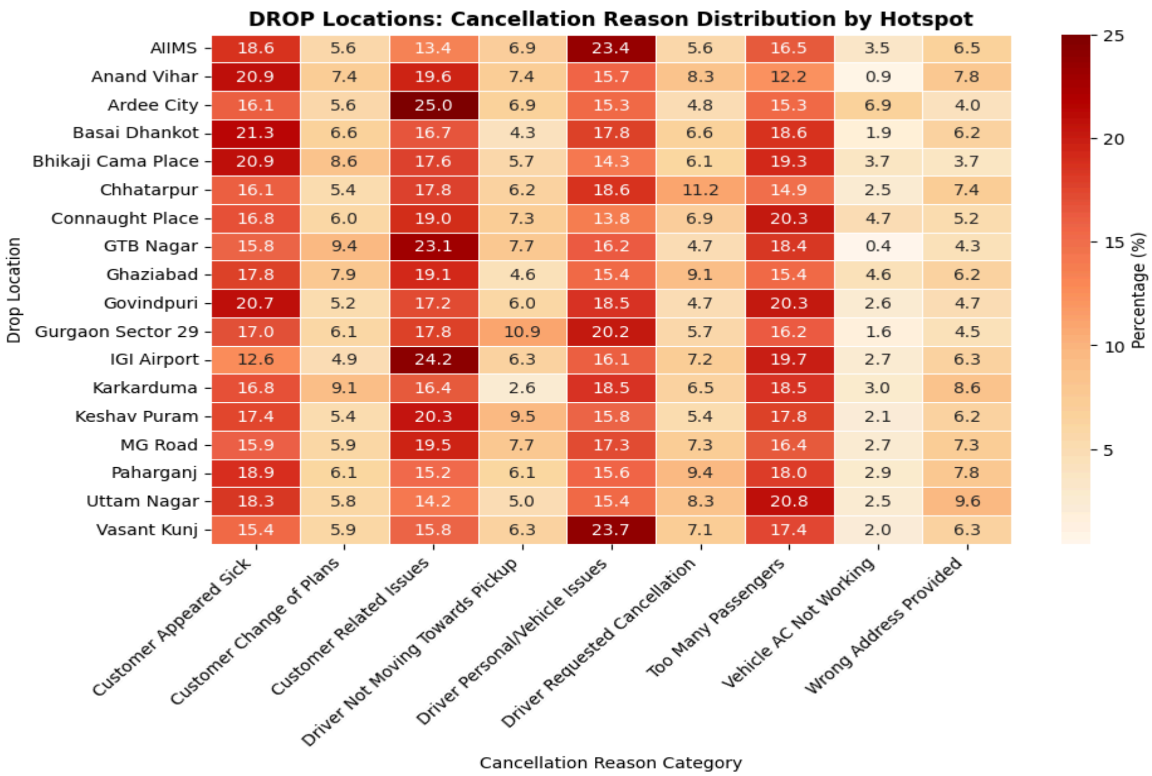
Appendix Q: "Top 10% of Drop Locations With The Highest Cancellation Rates"



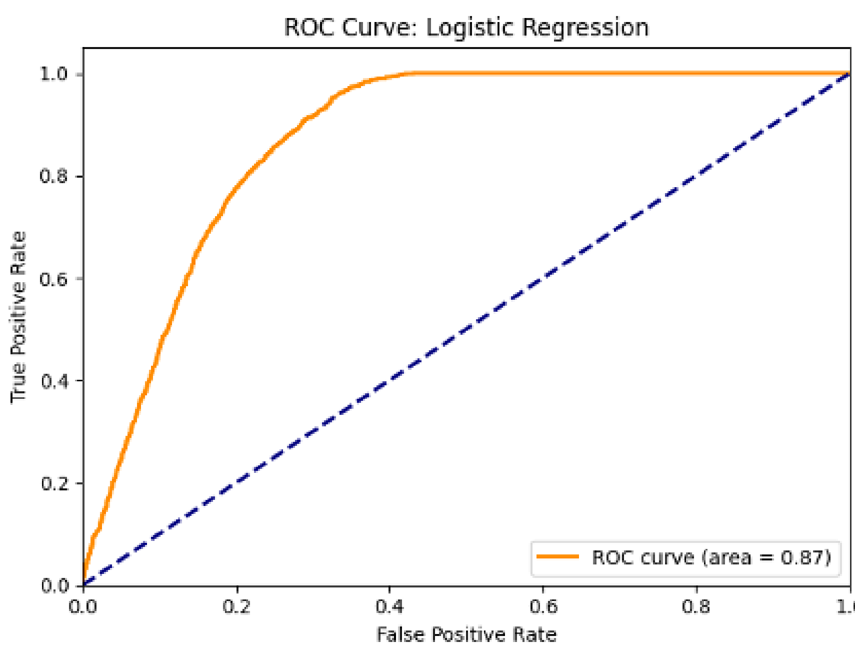
Appendix R: "Cancellation Reason Distribution in Pickup Location Hotspots"



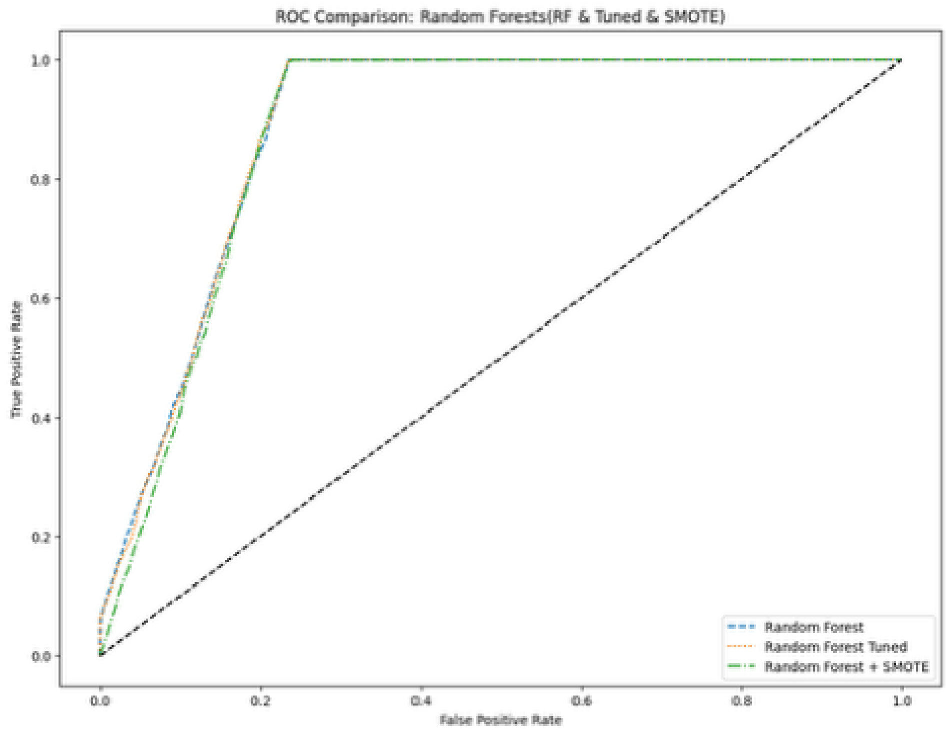
Appendix S: "Cancellation Reason Distribution in Drop Location Hotspots"



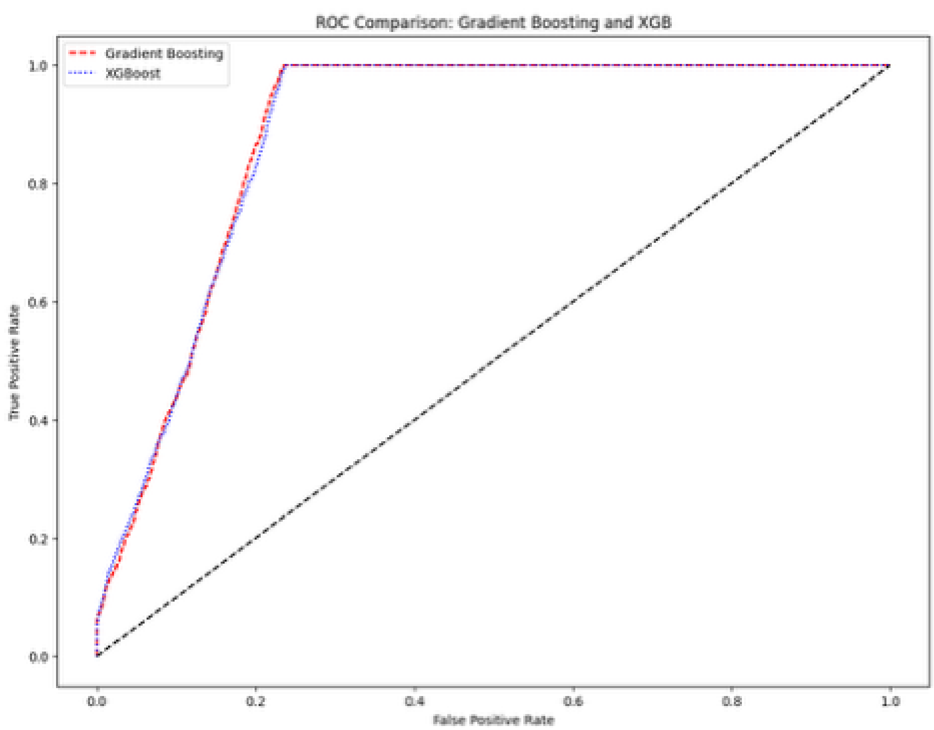
Appendix T: ROC Curve for Logistic Regression Modeling, Predicting Incomplete Rides



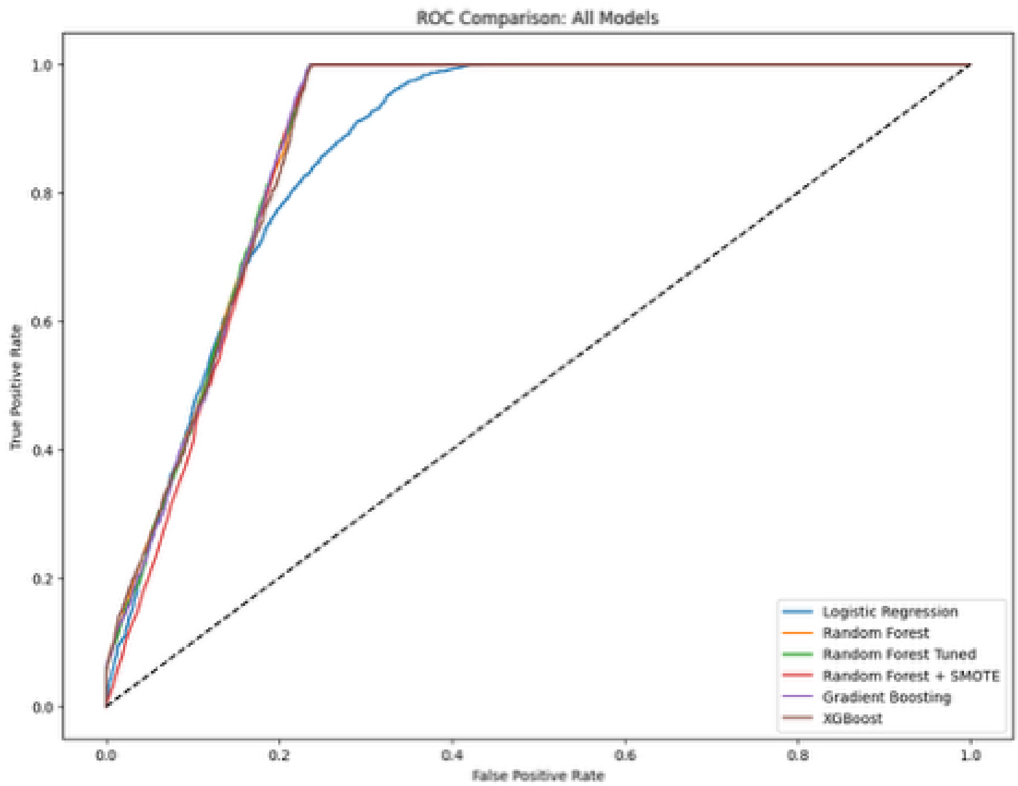
Appendix U: Composite ROC Curve for Random Forest Modeling (Base, Tuned, SMOTE) and Gradient Boosting, Predicting Incomplete Rides



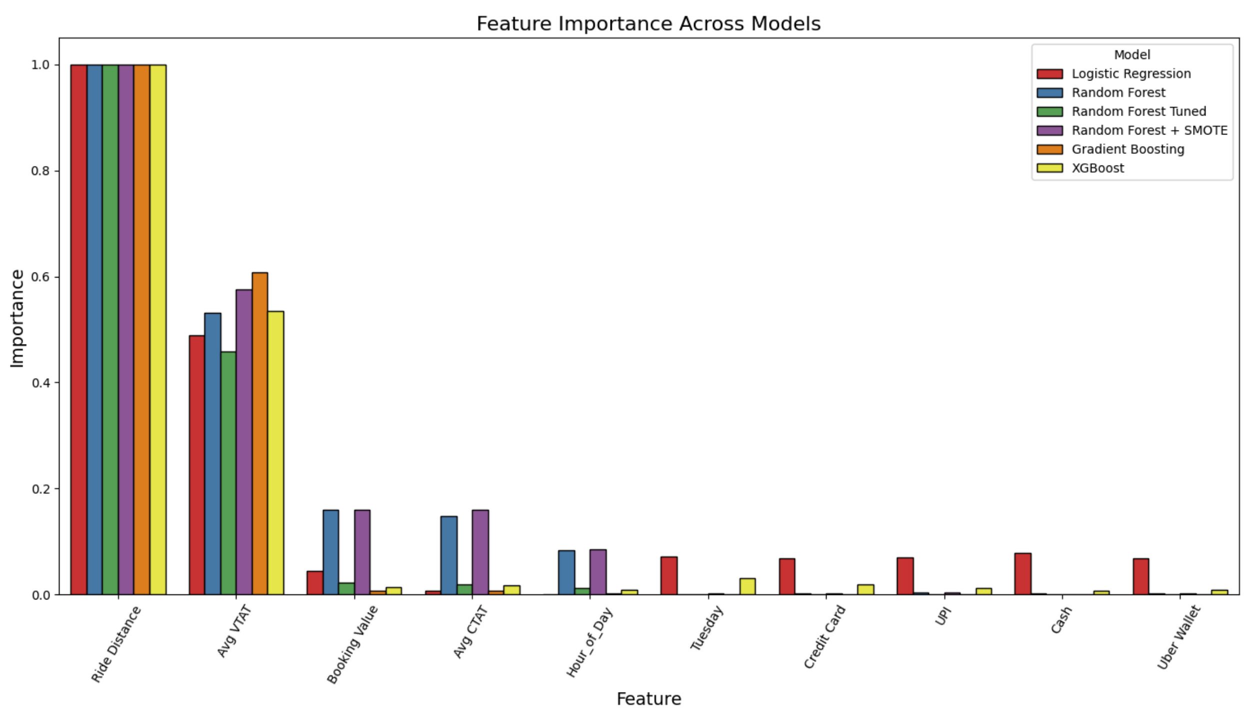
Appendix V: Composite ROC Curve, Gradient Boosting & XGBoost, Predicting “Incomplete”



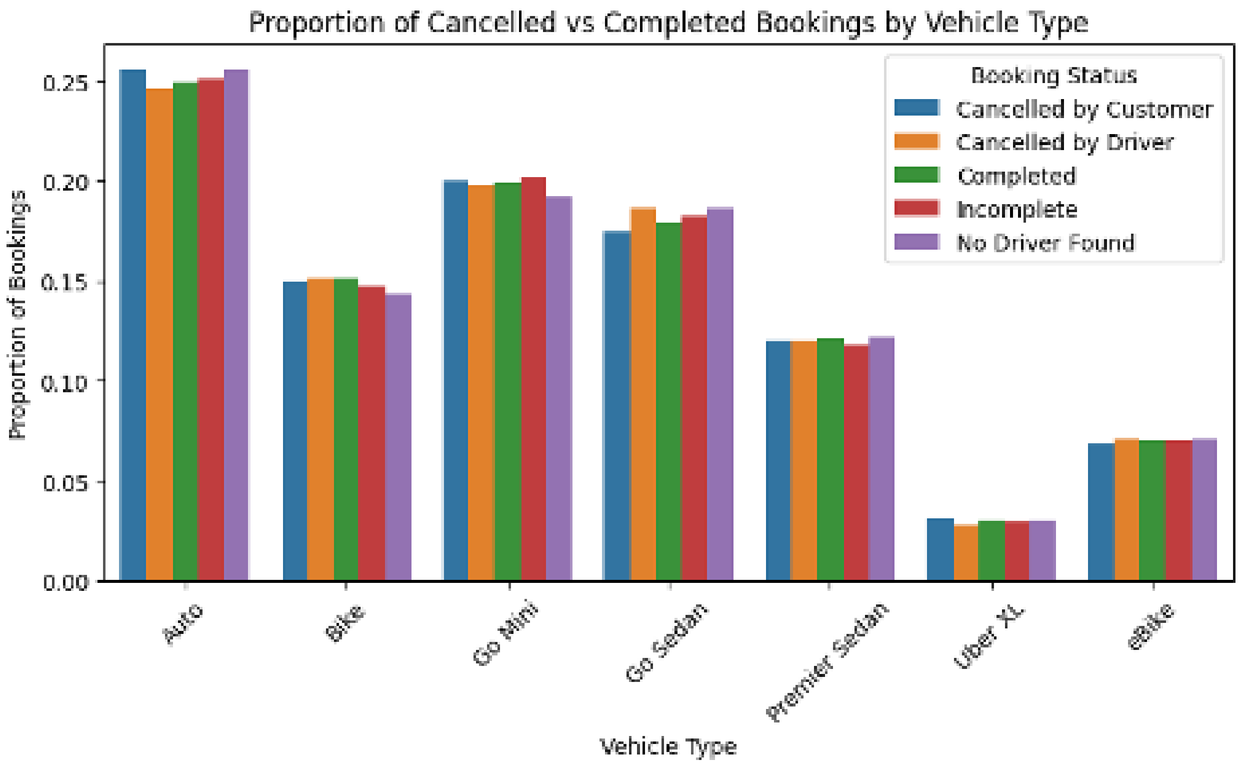
Appendix W: Composite ROC Curve for Logistic Regression, Random Forest Modeling (Base, Tuned, SMOTE), Gradient Boosting, and XGB, Predicting Incomplete Rides



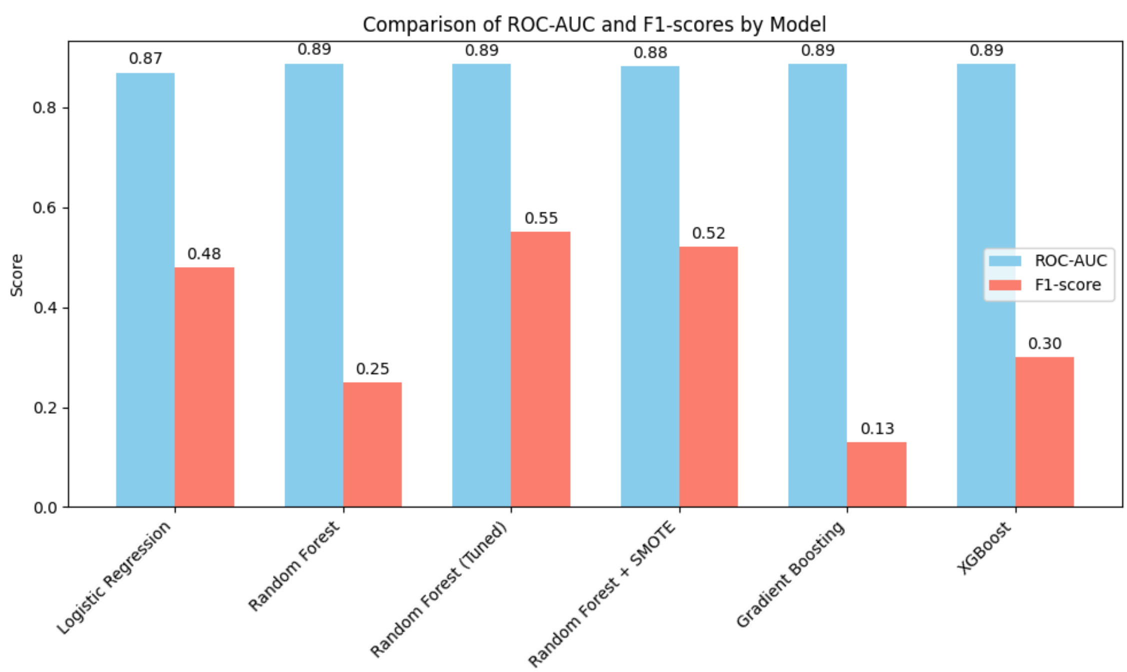
Appendix X: Composite Feature Importance for all Models, Predicting Incomplete Rides



Appendix Y: Share of Bookings for each Vehicle Type



Appendix Z: Differences between ROC and F1 Scores



References

Almaskati, D., Kermanshachi, S., Rosenberger, J. M., Pamidimukkala, A., Kan, C., & Foss, A. (2025). Factors influencing rideshare satisfaction in a university community.Transportation Research Interdisciplinary Perspectives, 32, 101547.https://doi.org/10.1016/j.trip.2025.101547

Bhargave, Naina, and Madhumita Sharma. "The Plight of Gig Workers in India." The Wire.

https://thewire.in/labour/the-plight-of-gig-workers-in-india

Brodeur, A., & Nield, K. (2018). An empirical analysis of taxi, Lyft, and Uber rides: Evidence from weather shocks in NYC. Journal of Economic Behavior & Organization, 152, 1–16.https://doi.org/10.1016/j.jebo.2018.06.004

"EDGAR Filing Documents for 0001543151-25-000008." Sec.gov, 2025, www.sec.gov/Archives/edgar/data/1543151/000154315125000008/0001543151-25-000008-index.htm. Accessed 19 Nov. 2025.

Farrell, Corrie Driebusch and Maureen. "Uber's High-Profile IPO Upsets with Weak Debut." Wall Street Journal, 10 May 2019, www.wsj.com/articles/uber-stumbles-in-trading-debut-11557503554.

Ngai Keung Chan. "The Rating Game: The Discipline of Uber's User-Generated Ratings."

ResearchGate, Queen's University Library, 31 Mar. 2019,

www.researchgate.net/publication/332114372_The_Rating_Game_The_Discipline_of_Uber.

"Uber Apps, Products, and Offerings | Uber." Uber, 2025, www.uber.com/us/en/about/uber-offerings/. Accessed 19 Nov. 2025.

"Uber Technologies, Inc. - FAQ and Resources." Uber.com, 2019, investor.uber.com/faq-resources/default.aspx. Accessed 19 Nov. 2025.

"Uber Technologies, Inc. (UBER) Valuation Measures & Financial Statistics." Yahoo Finance, 2025, finance.yahoo.com/quote/UBER/key-statistics/?guccounter=1. Accessed 19 Nov. 2025.

Verma, R., Pargal, S., Das, D., Parbat, T., Kambalapalli, S. S., Mitra, B., & Chakraborty, S. (2021). Impact of driving behavior on commuters' comfort during cab rides: Towards a new perspective of driver rating.arXiv preprint arXiv:2108.10944.http://arxiv.org/abs/2108.10944v1