A Mini Project Report on

UBER DATA ANALYSIS

T.E. - I.T Engineering

Submitted By

NAME OF STUDENT	MOODLE ID
Muskan Rao	20104096
Sanskruti Rawal	20104020
Divyal Chheda	21204005

Under The Guidance Of

Prof. Geetanjali Kalme

DEPARTMENT OF INFORMATION TECHNOLOGY

A.P.SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic year : 2022-23

CERTIFICATE

This to certify that the Mini Project report on **Uber Data Analysis** has been submitted by **Divyal Chheda (21204005), Muskan Rao (20104096), Sanskruti Rawal (20104020)** who are a Bonafede students of A. P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfillment of the requirement for the degree in Information Technology, during the academic year 2022-2023 in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Prof. Geetanjali Kalme
Guide

Dr. Kiran Deshpande   Dr. Uttam D.Kolekar

Head Department of Information Technology   Principal

External Examiner(s)

1.

2.

Place: A.P.Shah Institute of Technology, Thane

Date:

ACKNOWLEDGEMENT

ABSTRACT

Concept of organized rental cab was introduced to Indian consumers in 2004 when Meru cab service was launched in major metro cities of India and then almost after six years App based rental cab service was introduced in the Indian market. Among various transportation modes cab service gained popularity because of its advantage of door to door service and now because of technological advancement customers were able to book cabs at competitive price in just one click using their smartphones. These App based cab services were having tremendous potential for growth in densely populated countries like India where parking is a major problem because of space crunch as well as public transports are overcrowded during peak hours. Slowly this convenient mode of travel started gaining popularity and competition became dense after Uber's launch in 2013.As customers have become more demanding it's a challenging job for the rental cab industry to meet the customer's expectations.

In a data-driven world, our project navigates the fastest route to accelerate our business. Taking numbers, statistics, and digital metrics and transforming them into actionable insights, we utilize machine learning, natural language processing, and advanced statistical modelling to implement automations and algorithms that enhance safety, amplify performance, improve our customer experience across Uber, and make magic in the marketplace.

Two papers of related works are of interest to same projects – those related to case studies of uber data in different cities and ones related to urban transport development. It can greatly enhance the quality and aesthetics of your graphics, and will make you much more efficient in creating them. It is a system for declarative creating graphics, based on the grammar of graphics.

TABLE OF CONTENTS

Chapter 1

Introduction :

Uber is a transportation network company that provides ride-sharing and food delivery services worldwide. Since its launch in 2010, Uber has revolutionized the transportation industry, making it easier and more affordable for people to get around.

Uber collects massive amounts of data on its riders, drivers, and trips, and uses this data to improve its services and make more informed business decisions. The data includes information on trip duration, distance, location, fare, ratings, driver information, and more.

Analyzing this data can provide valuable insights into Uber's operations, such as identifying peak hours, popular routes, and driver performance. It can also help Uber optimize its pricing and allocation of resources to improve customer satisfaction and profitability.

Data analysis can be done using a variety of tools and techniques, including statistical analysis, data visualization, and machine learning. With the help of these techniques, Uber can gain a deeper understanding of its business and make data-driven decisions to improve its services.

Uber data analysis involves the process of examining and interpreting the large amounts of data generated by Uber rides. This includes analyzing data on rider behavior, driver behavior, trip details, and other variables that impact the overall performance of the Uber platform. The goal of this analysis is to gain insights into how the Uber platform is being used and how it can be optimized to provide better service to riders and drivers.

The data used in Uber analysis can come from a variety of sources, including user data, trip data, and location data. Machine learning algorithms are often used to identify patterns and trends in this data, and to make predictions about future trends and behaviors.

Uber data analysis can provide insights into a variety of topics, including rider demographics, driver earnings, popular pickup and drop-off locations, and the impact of external factors like weather and events on ride demand. This information can be used by Uber to improve its services and offerings, as well as to inform marketing and advertising campaigns.

Overall, Uber data analysis is an important tool for understanding the behavior of users and drivers on the platform, and it plays a critical role in helping Uber to optimize its services and achieve its business objectives.

1.1. Purpose :

The purpose of Uber data analysis is to gain insights into the operations of the taxi service and inform solutions to improve efficiency and customer satisfaction. By analyzing data on demand patterns, route optimization, surge pricing, driver incentives, and customer behavior, Uber can optimize its operations and provide a better experience for both passengers and drivers. The ultimate goal of Uber data analysis is to improve the quality of service provided by the taxi service, increase revenue and profitability, and maintain a competitive edge in the transportation industry.

The purpose of Uber data analysis is to improve the efficiency and effectiveness of the Uber ride-sharing service. By analyzing data on various aspects of the service, including demand patterns, driver performance, customer behavior, and route optimization, Uber can gain insights that inform decision-making and lead to improvements in operations, customer experience, and profitability. The goal of Uber data analysis is to provide a better experience for passengers and drivers while optimizing the use of resources and maximizing revenue.

1.2.   Problem Statement :

The problem statement of Uber data analysis can vary depending on the specific area of focus. However, some common problem statements include:

1. **Pricing optimization:** Uber's pricing strategy is crucial to its success. Data analysis can help Uber identify pricing patterns and trends to optimize fares and increase revenue.

2. **Rider satisfaction:** Uber's success is also heavily dependent on rider satisfaction. Data analysis can help identify areas where riders are dissatisfied and suggest improvements to enhance the overall customer experience.

3. **Driver performance:** Uber needs to ensure that its drivers are providing a high level of service to riders. Data analysis can help identify areas where drivers are underperforming and suggest training or other measures to improve driver performance.

4. **Safety:** Safety is a critical concern for Uber. Data analysis can help identify areas where safety may be a concern and suggest measures to improve safety, such as improving driver screening processes or implementing new safety features in the app.

In summary, the problem statement of Uber data analysis is to use data to improve various aspects of the business, including driver allocation, pricing, rider satisfaction, driver performance, and safety.

1.3. Objectives :

1. **Optimize driver allocation:** One objective of Uber data analysis may be to optimize the allocation of drivers to different areas and times of day. By analyzing data on demand patterns and driver availability, Uber can identify areas of high demand and ensure that there are enough drivers to meet that demand.

2. **Improve routing efficiency:** Another objective of Uber data analysis may be to improve the efficiency of routing algorithms. By analyzing real-time data on traffic patterns and road conditions, Uber can optimize routes to reduce travel time and improve the overall efficiency of the service.

3. **Enhance the user experience:** Uber may use data analysis to improve the user experience for passengers and drivers. By analyzing data on customer behavior and preferences, Uber can develop personalized marketing campaigns, tailor service offerings, and improve customer support.

4. **Increase driver satisfaction:** Uber may use data analysis to identify areas where driver satisfaction can be improved. By analyzing data on driver performance, feedback, and incentives, Uber can develop programs and policies that motivate drivers to provide high-quality service.

5. **Maximize revenue:** Finally, Uber may use data analysis to maximize revenue by identifying areas where pricing can be adjusted or by developing new service offerings that meet customer needs. By analyzing data on customer behavior and demand patterns, Uber can develop pricing strategies that balance supply and demand and optimize revenue.

Overall, the objectives of Uber data analysis are to improve the efficiency and effectiveness of the service, enhance the user experience, and maximize revenue while maintaining a high level of service quality.

1.4.Scope :

Rider behavior analysis involves analyzing data on rider demographics, ride frequency, payment methods, and preferred ride types. This analysis can help Uber understand rider preferences and tailor its services to meet their needs. Driver behavior analysis involves analyzing data on driver performance, ratings, and feedback from riders. This analysis can help Uber identify drivers who may need additional training or support to improve their performance. Route analysis involves analyzing data on popular routes, travel times, and traffic patterns. This analysis can help Uber optimize its routing algorithms to improve trip

efficiency and reduce wait times. Pricing analysis involves analyzing data on fares, surge pricing, and rider demand. This analysis can help Uber optimize pricing strategies to increase revenue and improve rider satisfaction. Safety analysis involves analyzing data on safety incidents, driver behavior, and rider feedback. This analysis can help Uber identify safety risks and implement measures to improve safety. Market analysis involves analyzing data on market trends, competitor activity, and new market opportunities. This analysis can help Uber make informed decisions about expanding its services to new markets or developing new products. Overall, the scope of Uber data analysis is vast and can provide valuable insights into various aspects of the business, enabling Uber to improve its services, increase efficiency, and enhance the customer experience.

Chapter 2

Literature Review :

Sr No.	Title	Author(s)	Year	Outcomes	Methodology	Result
1	UBER DATA ANALYSIS USING GGPLOT	Mrunal Patil, Vidya Kumari, Adarsh Patil, Laxmikant Ahire and Asst.Prof. Umakant Mandawkar from B.Tech, Computer Science and Engineering, Sandip University, Nashik, India	2021	Uber analyze historical data for say, last 3 or 4 weeks and identifies pockets within the city that witness extremely high demand	Estimated Time of Arrival can be reduced with increase in the number of Uber drivers which successively will make Uber more liked by the customers and hence, the company will get more revenue and drivers will also be profited.	At the end of all procedure we get to see different graphs giving us insights.
2	Uber Related Data Analysis using Machine Learning	Rishi Srinivas, B. Ankayarkanni, R. Sathya Bama Krishna	2021	The paper explains the working of an Uber dataset, which contains data produced by Uber for New York City. Uber is defined as a P2P platform.	Using the information, the paper explains the use of the k-means clustering algorithm on the set of data and classify the various parts of New York City. Since the industry is booming and expected to grow shortly.	ffective taxi dispatching will facilitate each driver and passenger to reduce the wait time to seek out one another. The model is employed to predict the demand on points of the city.

Table 2.1

Chapter 3

Proposed System :

We use the NYC dataset containing data on over 4.5 million trips in NYC from Apr. to Sept. 2014. The files are separated by month and the following columns:

•Date

•Time Stamp

•Destination

•Source

•Price

There are some of the questions this dataset can be used to answer, we'll choose the following:

- Uber trips and distribution
- Time when Uber trips occur regularly
- Price range of all the ride

3.1. Feature & Functionality:

The features of Uber data analysis include:

1. **Data-driven decision-making:** Uber data analysis is based on insights extracted from data, which provides a solid foundation for decision-making. This ensures that decisions made are evidence-based and less prone to bias or speculation.

2. **Real-time analytics:** Uber's business operates in real-time, and data analysis needs to keep up with the pace of the business. Real-time analytics provides up-to-date insights that enable Uber to respond quickly to changes in demand, rider behavior, or other factors.

3. **Predictive analytics:** Predictive analytics is an essential feature of Uber data analysis as it enables Uber to forecast future demand, pricing trends, and driver behavior. This provides Uber with an edge in the highly competitive ride-sharing industry.

4. **Personalization:** Uber collects data on riders and drivers, enabling it to provide a personalized experience for each user. Data analysis helps Uber to identify user preferences, enabling it to customize services and offerings to meet the needs of individual users.

5. **Visualization:** Data visualization is a crucial feature of Uber data analysis as it enables stakeholders to understand complex data quickly. Visualization tools such as charts,

graphs, and maps enable stakeholders to see trends, patterns, and insights that might not be immediately apparent from raw data.

6. **Integration with other systems:** Uber data analysis needs to integrate with other systems such as driver management, payment processing, and app development. This integration provides a more comprehensive view of the business, enabling Uber to make more informed decisions.

In summary, the features of Uber data analysis include data-driven decision-making, real-time analytics, predictive analytics, personalization, visualization, and integration with other systems. These features enable Uber to gain insights, improve operations, and provide a better experience for riders and drivers.

Chapter 4

Requirement Analysis :

● Performance Requirements

The load time for the user interface screen should take no longer than 5 seconds.

● Design Constraints

The application should be able to run on any Pc or Laptop.

● Availability

The application should be always available whenever the user wants to use it.

● Hardware requirements: - RAM

The application requires a device with a minimum of 512 MB RAM while running.

Processor speed

The application requires a device with a minimum processor speed of 1 GHz while running.

Software requirements: - Operating system

The application must run on any Operating System.

Chapter 5
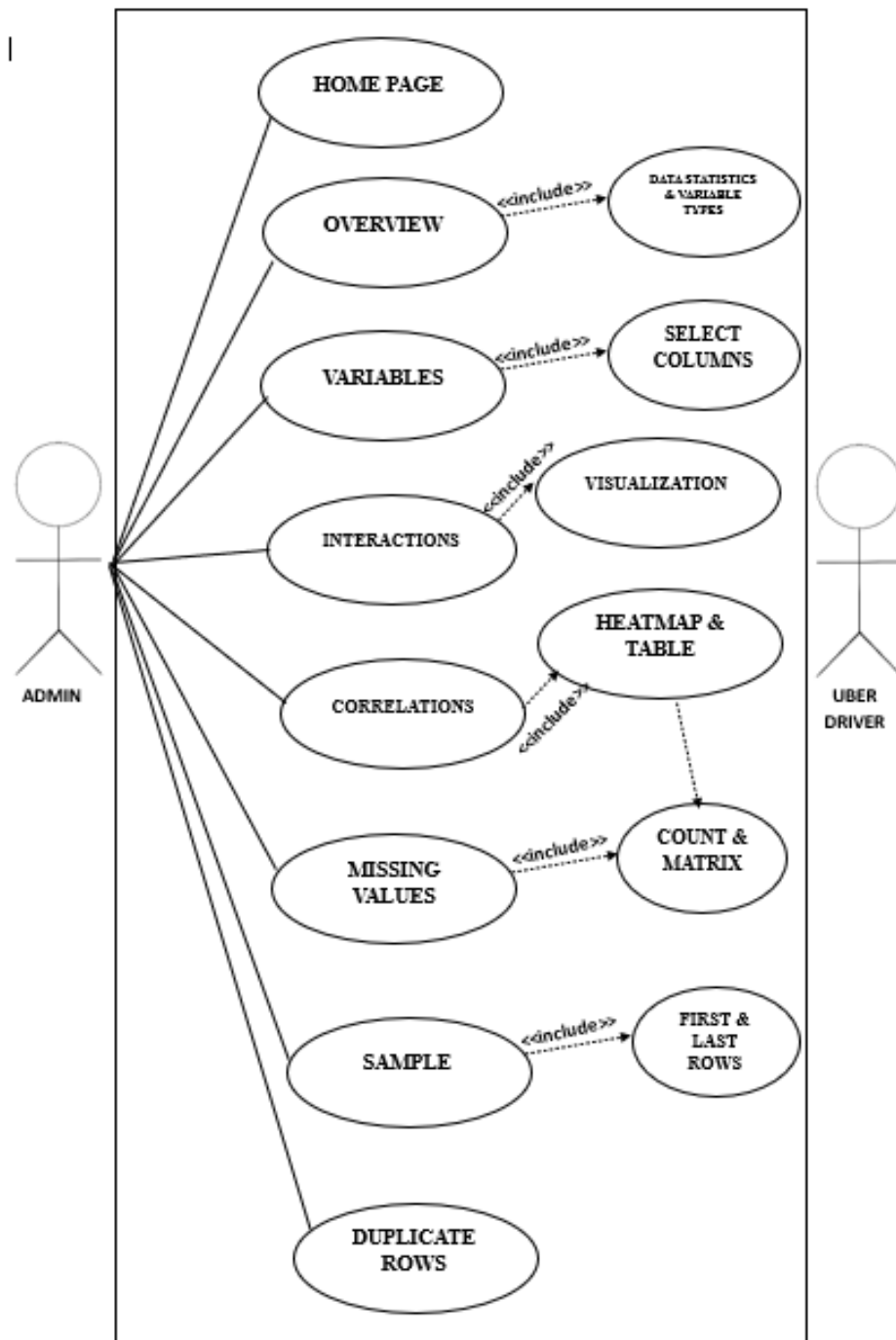
Project Design:

Use case diagram:

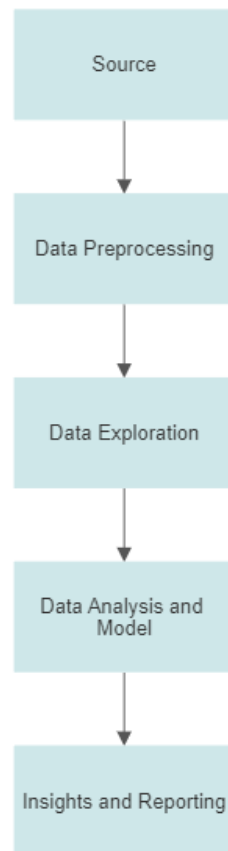Figure 5.1 Use Case Diagram

Data Flow diagram:

Figure 5.2 Level 0 DFD

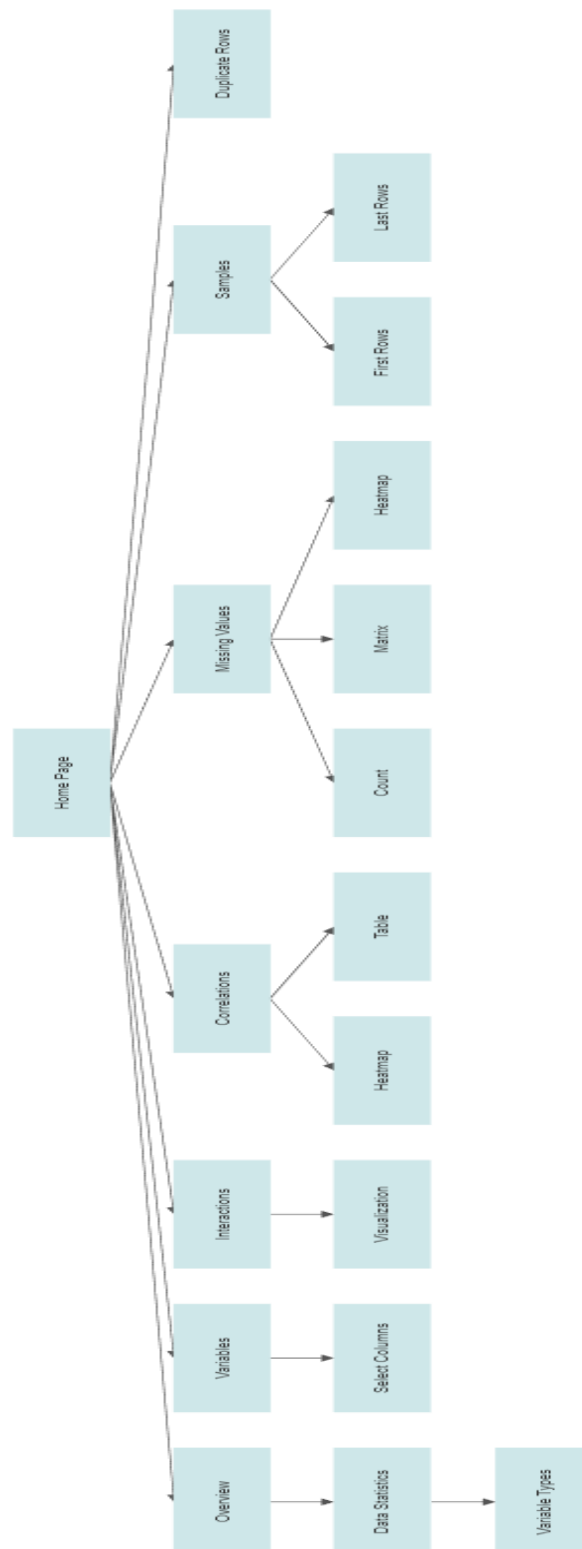**System Architecture :**

Figure 5.3 Flow Chart

Chapter 6

Technical Specifications :

FRONTEND :

Streamlit

BACKEND :

Python 3

PACKAGES USED :

numpy, pandas, sklearn, scipy, matplotlib, seaborn etc.

Chapter No: 7

Project Scheduling

Date	Weeks	Contents
14/01/2023 TO 18/01/2023	1	Group formation and Topic finalization. Identifying the scope and objectives of the Mini Project
24/01/2023 TO 30/01/2023	2	Identifying the functionalities of the Mini Project
31/01/2023 TO 4/02/2023	3	Discussing the ML Algorithm.
6/02/2023 TO 10/02/2023	4	Designing the Graphical User Interface (GUI)
17/02/2023 TO 19/02/2023	5	Review 1 Presentations
20/02/2023 TO 28/02/2022	6	Detail ML Algorithm implementation
03/03/2023 TO 10/03/2023	7	Integration of GUI with ML Algorithm code
14/03/2023 To 21/03/2023	8	Report Writing
20/04/2023 TO 20/04/2023	9	Review 2 Presentations

Table 7.1
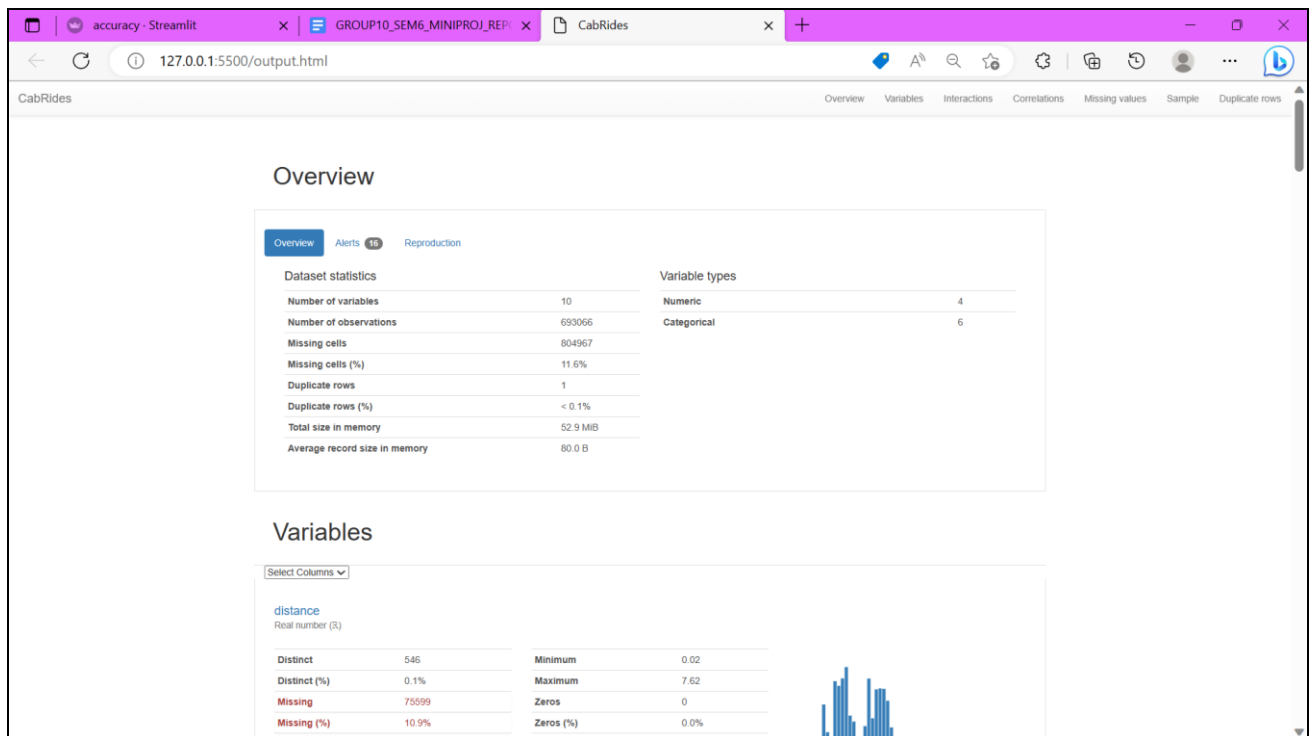
Chapter 8

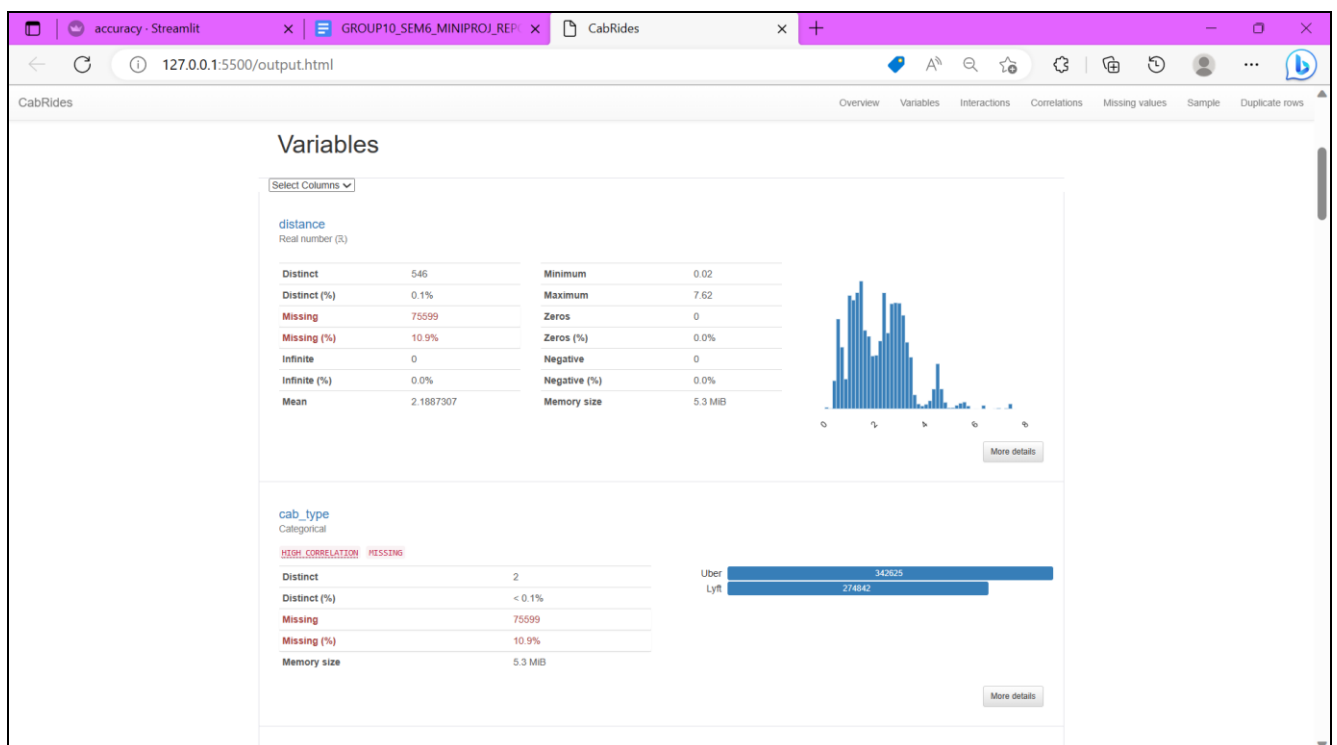## Implementation :

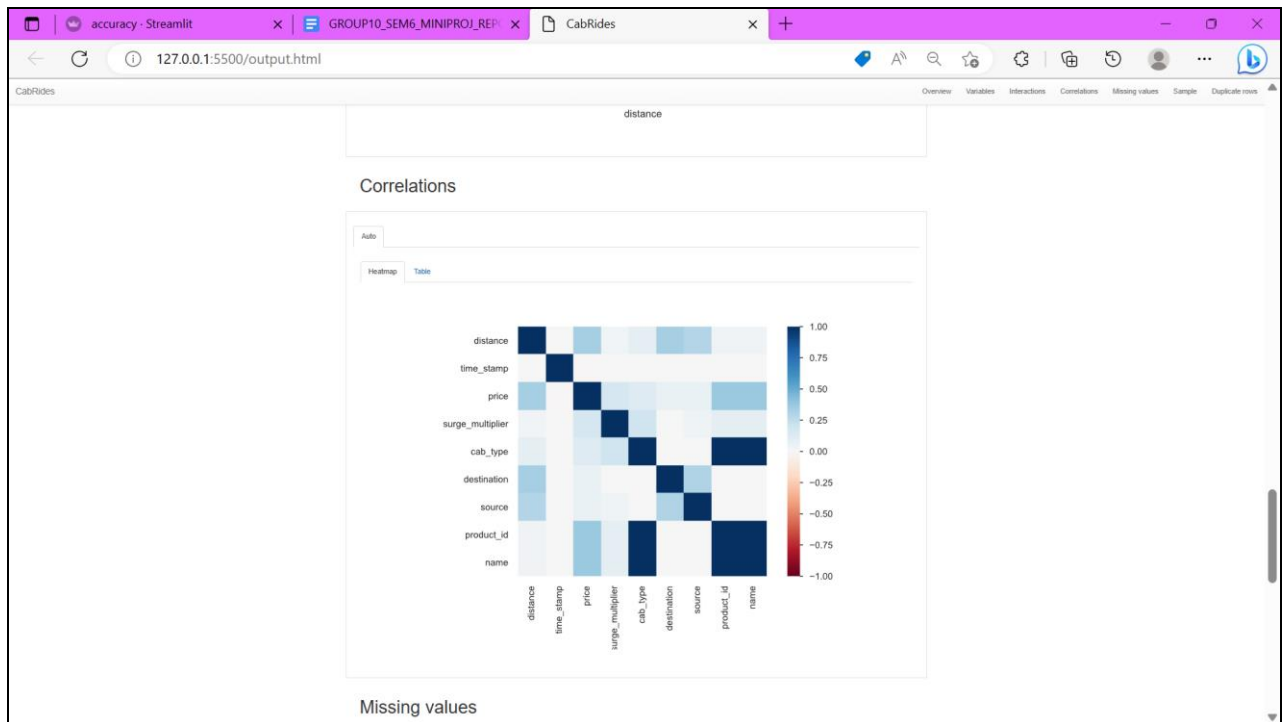Figure 8.1 Overview

Figure 8.2 Variables

Figure 8.3 Correlation

Chapter 9

Result & Future Scope :

The result of Uber data analysis can provide valuable insights into the operations of the taxi service and inform solutions to improve efficiency and customer satisfaction. Here are some possible results and discussions that may arise from Uber data analysis:

- Demand Patterns
- Route Optimization
- Surge pricing
- Driver Incentives
- Customer Behavior

Autonomous driving: As autonomous driving technology advances, Uber may leverage data analysis to optimize its self-driving fleet. This could involve analyzing data on traffic patterns, weather conditions, and road infrastructure to develop algorithms that enable autonomous vehicles to navigate safely and efficiently.

Predictive maintenance: By analyzing data on driver behavior, vehicle performance, and maintenance records, Uber could develop predictive maintenance models that identify potential issues before they occur. This could help reduce downtime and maintenance costs while improving the safety and reliability of the service.

Personalization: Uber could use data analysis to develop personalized service offerings that cater to the unique needs and preferences of individual passengers. This could involve analyzing data on previous trips, ratings, and reviews to develop tailored recommendations for routes, drivers, and other aspects of the service.

Integration with public transit: Uber could use data analysis to integrate its service with public transit systems, providing a seamless and convenient transportation experience for passengers. This could involve analyzing data on transit schedules, routes, and demand patterns to optimize service offerings and reduce congestion.

Environmental impact: Uber could use data analysis to reduce its environmental impact by optimizing routing and driver allocation to minimize emissions and fuel consumption. This could involve analyzing data on vehicle performance, traffic patterns, and fuel efficiency to develop strategies that reduce the carbon footprint of the service.

Overall, the future scope of Uber data analysis is vast and holds immense potential for improving the efficiency, effectiveness, and sustainability of the ride-sharing service.

Chapter 10

Conclusion :

In conclusion, data analysis is an essential tool for Uber to gain insights and make data-driven decisions that improve the overall performance of its platform. By analyzing data on rider behaviour, driver behaviour, trip details, and other variables, Uber can uncover patterns and insights that inform future decisions and strategies.

The objectives of Uber data analysis include improving the user experience, increasing driver earnings, enhancing safety, optimizing pricing, and expanding services. The scope of Uber exploratory data analysis can include an examination of data quality and completeness, rider and driver behavior, trip details, external factors, and visualization and communication. The outcome of Uber data analysis is actionable insights that can be used to improve the user experience, increase driver earnings, enhance safety, optimize pricing, and expand services. By leveraging the knowledge gained from data analysis, Uber can continue to grow its business and maintain its position as a leader in the transportation industry.

References

- https://www.analyticsvidhya.com/blog/2021/10/end-to-end-predictive-analysis-on-ubers-data/
- https://data-flair.training/blogs/r-data-science-project-uber-data-analysis/
- https://www.projectpro.io/article/uber-data-analysis-project-using-machine-learning-in-python/589