|  |  |
| --- | --- |
| Ride-Sharing Data Analysis Report | Abstract  This report explores data from five datasets: drivers, ratings, rides, users, and vehicles. The goal is to identify patterns in ride demand, driver performance, user spending habits, and ride efficiency. The insights drawn will help improve service delivery, optimize driver allocation, and increase revenue.    **Title: Ride-Sharing Data Analysis Report Author: Phanindra Reddy Date: 01-02-2025** |

Table of Contents

[Introduction 2](#_Toc189300744)

[Dataset Contents 2](#_Toc189300745)

[Tabular view: 3](#_Toc189300746)

[Data Analysis & Findings 4](#_Toc189300747)

[1 Peak Ride Hours 4](#_Toc189300748)

[2 Average Driver Rating 5](#_Toc189300749)

[3 Ride Statistics for Users 6](#_Toc189300750)

[4 Average Distance Covered per Ride 7](#_Toc189300751)

[5 Driver Performance Metrics (SQL View) 8](#_Toc189300752)

[6 Top Performing Drivers by Revenue 8](#_Toc189300753)

[7 Customer Lifetime Value (CLV) 9](#_Toc189300754)

[8 Vehicle Utilization Analysis 9](#_Toc189300755)

[9 Surge Pricing Analysis 10](#_Toc189300756)

[10 Rider Frequency by Location 10](#_Toc189300757)

[11 Fare by Gender 11](#_Toc189300758)

[12 Seasonal Ride Analysis 11](#_Toc189300759)

[13 Rating Correlation with Fare Amount 12](#_Toc189300760)

[14 Driver Availability vs. Ride Demand 12](#_Toc189300761)

[15 Customer Ride Frequency (Churn Analysis) 13](#_Toc189300762)

[Conclusion 14](#_Toc189300763)

# Introduction

Ride-sharing services have transformed urban transportation by providing on-demand rides, reducing traffic congestion, and offering economic opportunities. This report analyzes ride-sharing data to derive key insights that can help optimize business operations and enhance profitability.

## Dataset Contents

The dataset is divided into the following five CSV files:

1. **users.csv**: Contains information about users, including user ID, name, email, phone number, registration date, age, gender, and location.
2. **drivers.csv**: Includes details about drivers such as driver ID, name, vehicle ID, rating, total rides, and availability.
3. **vehicles.csv**: Lists vehicle details like vehicle ID, make, model, year, and capacity.
4. **rides.csv**: Contains ride information including ride ID, user ID, start location, end location, start/end time, distance (km), fare amount, and driver ID.
5. **ratings.csv**: Includes ride ratings and user comments, with each rating linked to a ride and a user

 **Fact Table:**

* **Rides (Fact Table):** This will be the main table that captures each ride's details.
  + ride\_id, user\_id, driver\_id, vehicle\_id, ride\_start\_time, ride\_end\_time, distance\_km, fare\_amount.

 **Dimension Tables:**

* **Drivers:** Contains driver details.
  + driver\_id, name, vehicle\_id, rating, total\_rides, available.
* **Users:** Contains user details.
  + user\_id, name, email, phone\_number, registration\_date, age, gender, location.
* **Vehicles:** Contains vehicle details.
  + vehicle\_id, make, model, year, capacity.
* **Ratings:** Contains ride ratings.
  + rating\_id, ride\_id, user\_id, rating\_value, comments, rating\_date

# Tabular view:

**A diagram of a vehicle

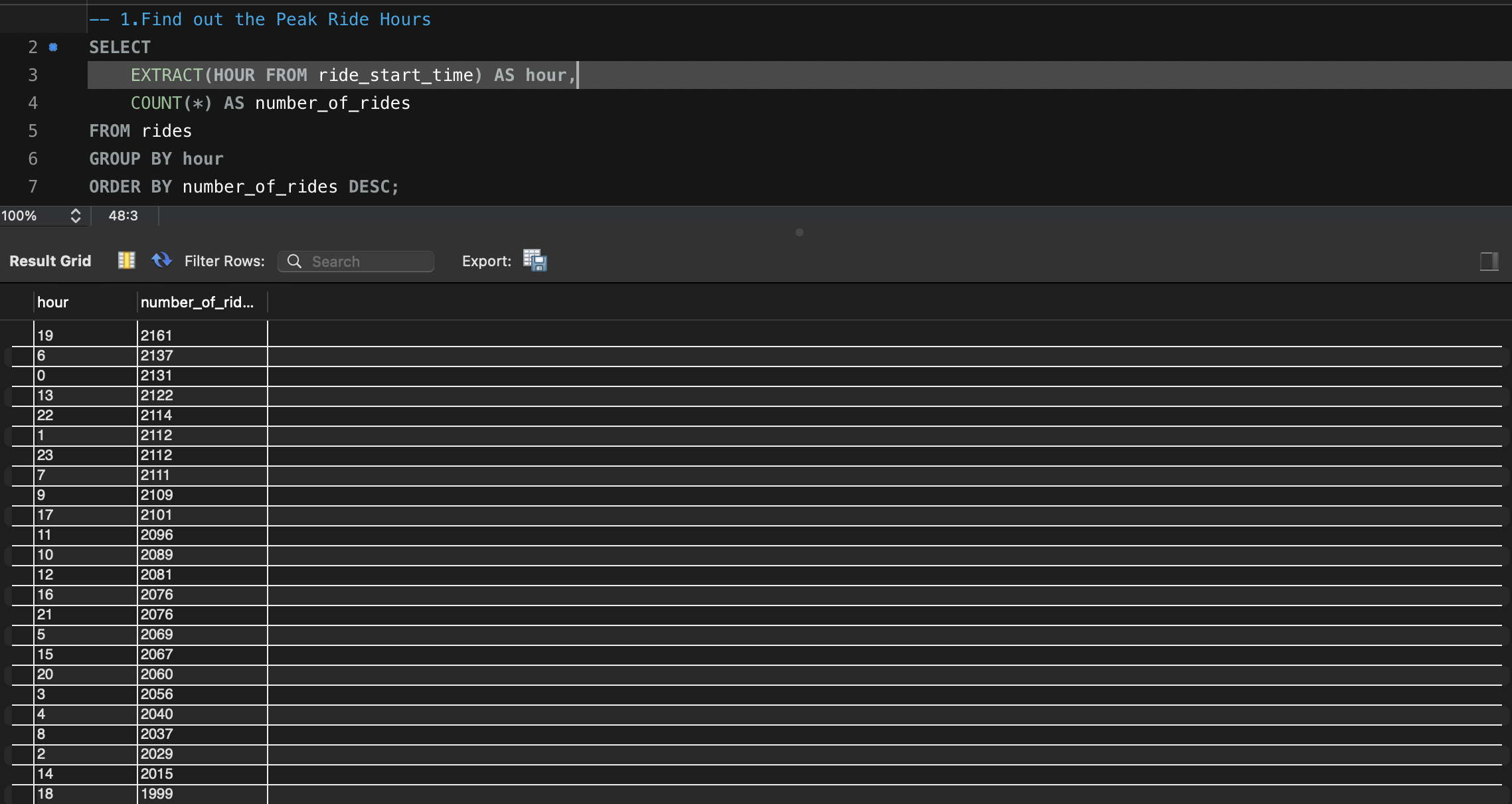
Description automatically generated**

# 

# Data Analysis & Findings

## 1 Peak Ride Hours

**Objective:** Identify the peak hours when the maximum number of rides occur.



**Findings:**

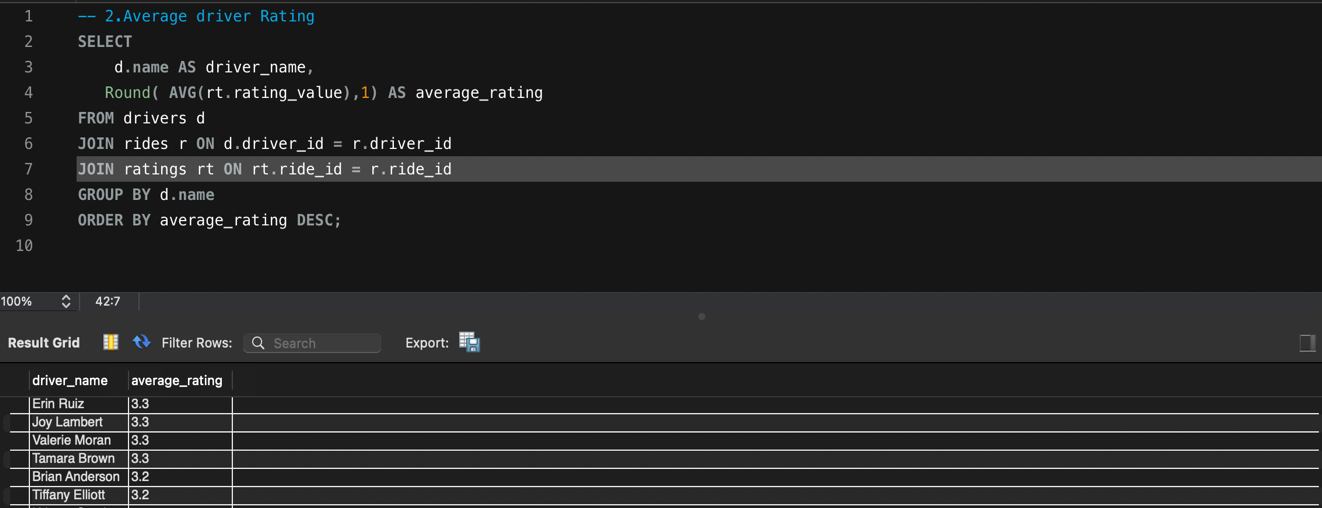
* The busiest hours for ride requests were extracted using the EXTRACT(HOUR FROM ride\_start\_time) function.
* The data was grouped by hour, and the number of rides was counted and sorted in descending order.

**Insights:**

* These peak hours can be used for surge pricing strategies or to allocate more drivers during high-demand periods.

## 2 Average Driver Rating

**Objective:** Calculate the average rating for each driver.



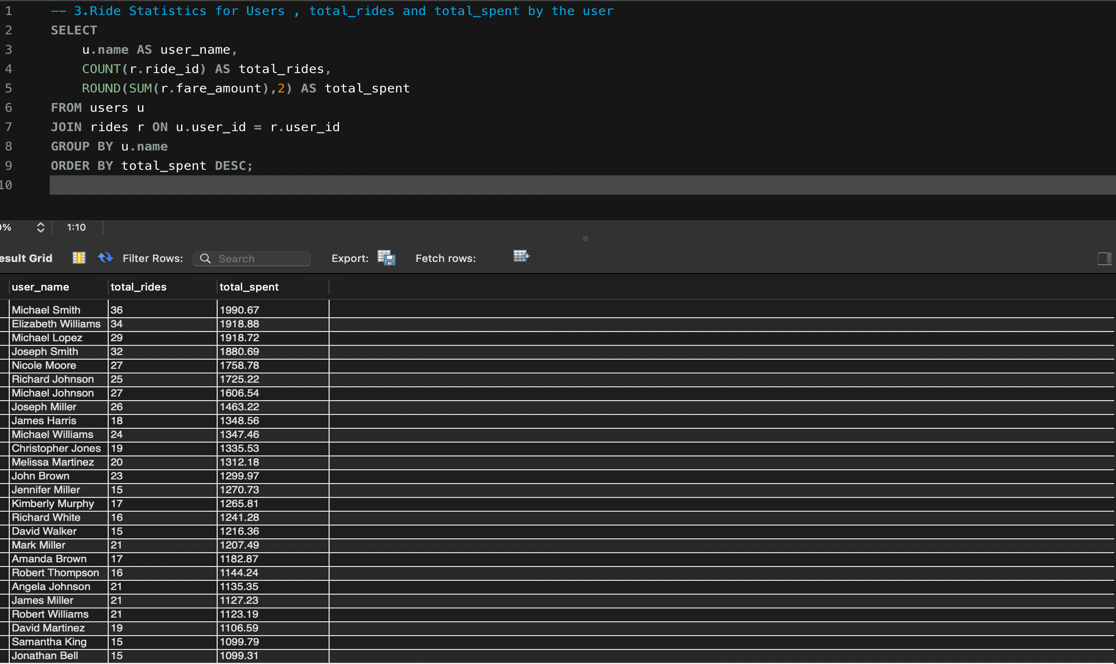
**Findings:**

* Driver ratings were extracted by joining the drivers, rides, and ratings tables.
* The ratings were averaged and rounded to one decimal place for each driver.
* The drivers were then sorted based on their average rating.

**Insights:**

* Identifying high-rated drivers can help in rewarding top performers.
* Low-rated drivers might need further training or review.

## 3 Ride Statistics for Users

**Objective:** Determine user engagement by analyzing the number of rides taken and the total amount spent. 

**Findings:**

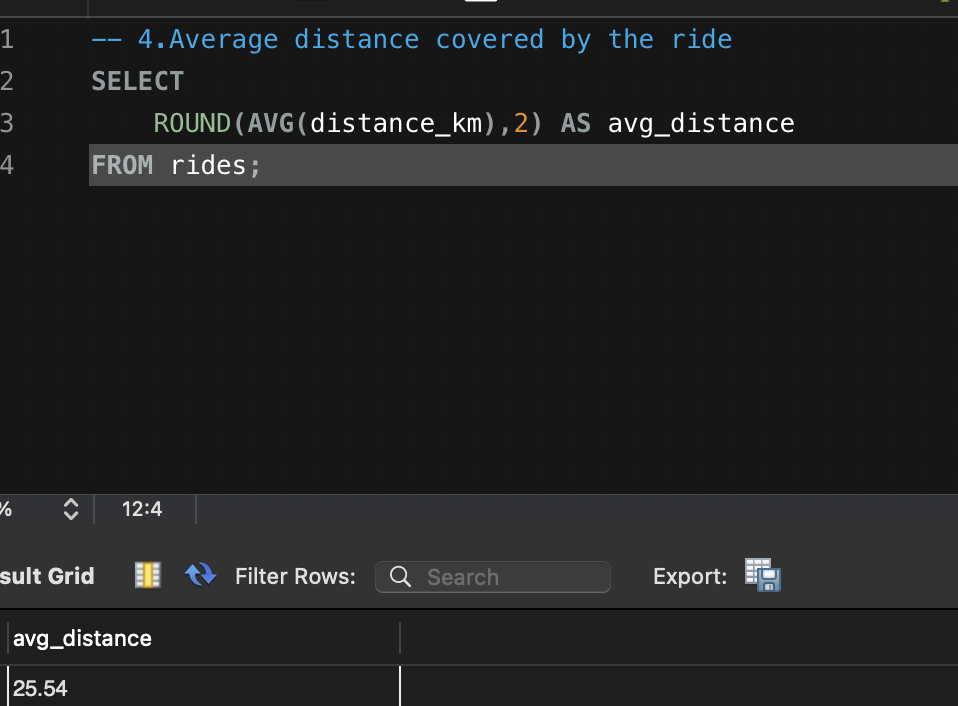
* The data was grouped by user and aggregated to count total rides and sum total spending.
* The results were ordered by total spending to identify high-value customers.

**Insights:**

* High-spending customers could be targeted with loyalty rewards or promotional offers.
* Identifying inactive users can help in implementing re-engagement campaigns.

## 4 Average Distance Covered per Ride

**Objective:** Compute the average ride distance.



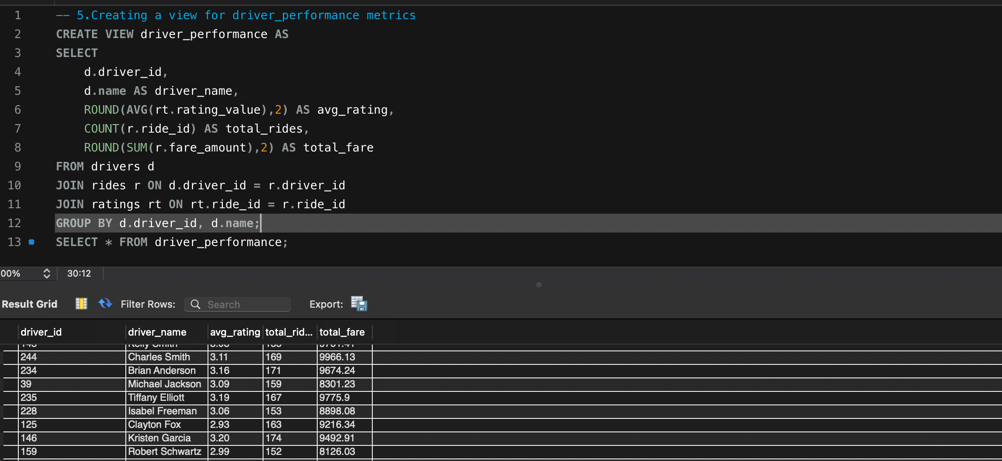
**Findings:**

* The AVG(distance\_km) function was used to calculate the average distance covered per ride.

**Insights:**

* This data can help optimize pricing models.
* Understanding ride distance trends may assist in fleet management and fuel efficiency planning.

## 5 Driver Performance Metrics (SQL View)

**Objective:** Create a structured view to track driver performance.

**Findings:**

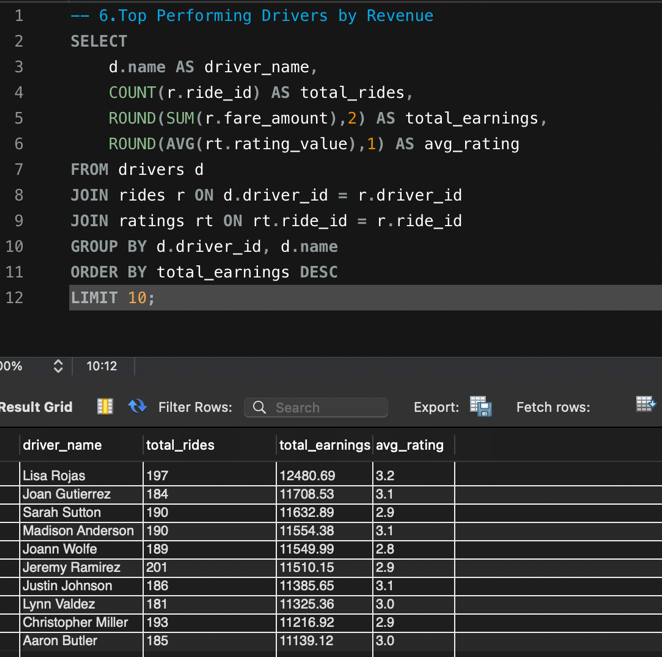
* A SQL view named driver\_performance was created, containing:
  + Driver ID and name
  + Average rating
  + Total rides handled
* This allows easy access to structured performance data.

**Insights:**

* The driver\_performance view can be used for performance tracking and decision-making.
* Helps identify drivers who need improvement or recognition.

## 6 Top Performing Drivers by Revenue

**Objective:** Identify the highest-earning drivers.



**Findings:**

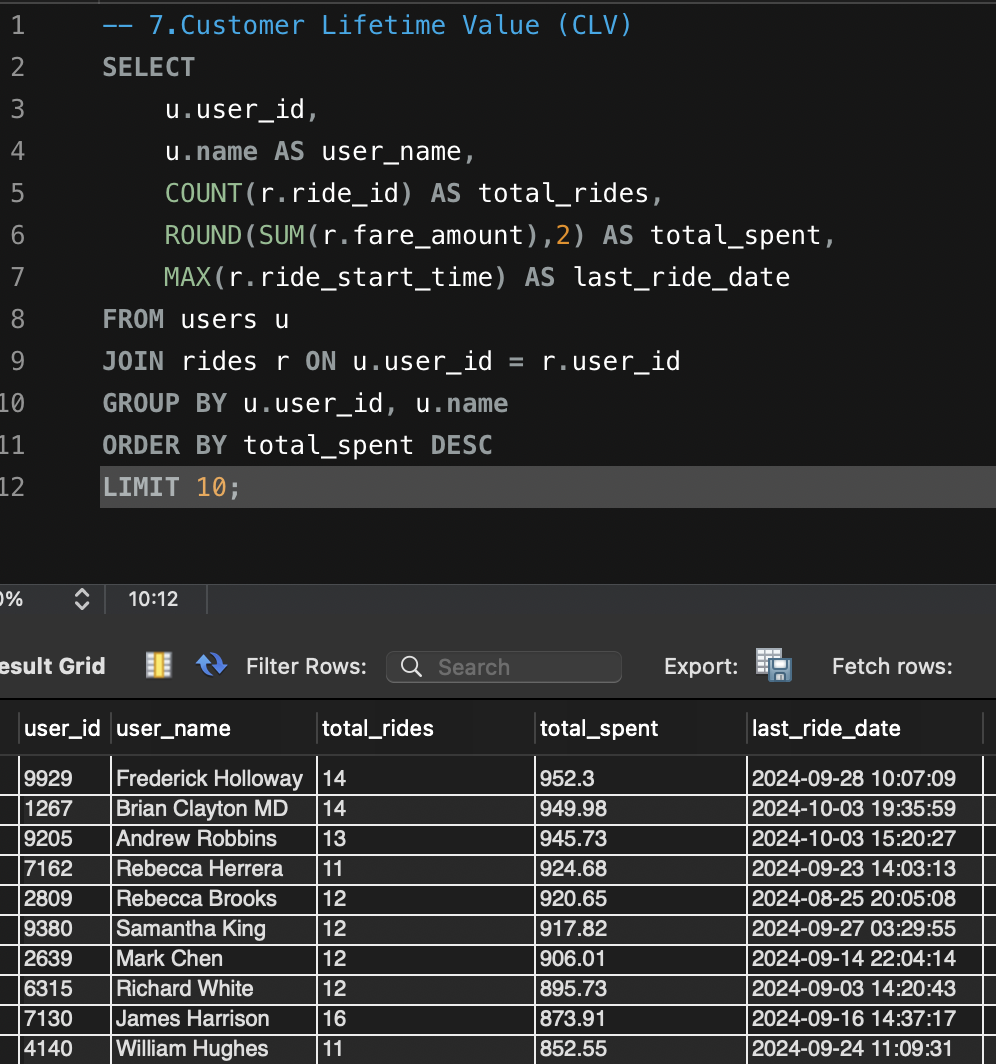
* Revenue per driver was calculated by summing fare amounts.
* Drivers were ranked based on total revenue.

**Insights:**

* Top-performing drivers could be incentivized to maintain high performance.

## 7 Customer Lifetime Value (CLV)

**Objective:** Estimate long-term customer value.



**Findings:**

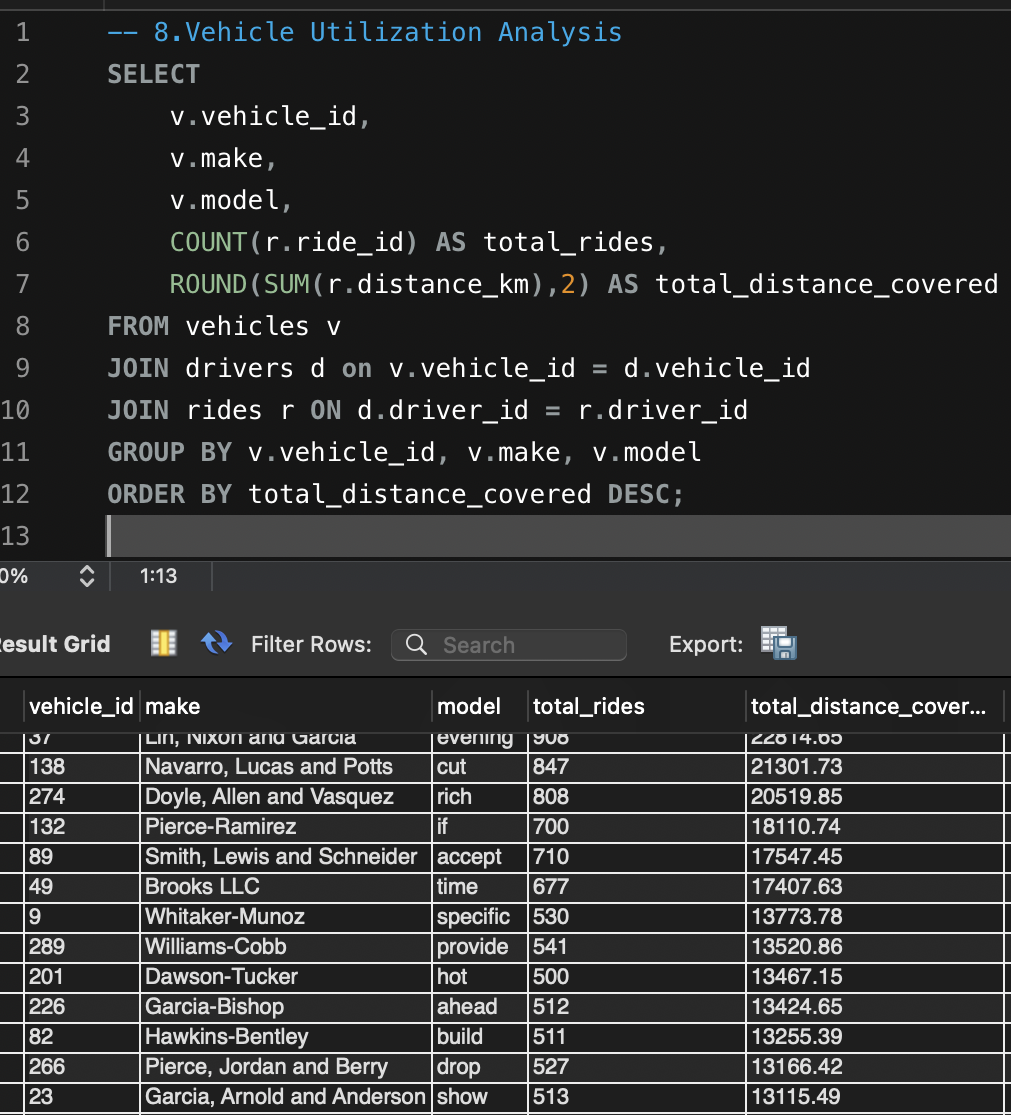
* CLV was calculated based on user spending trends.

**Insights:**

* Helps in customer retention strategies and loyalty programs.

## 8 Vehicle Utilization Analysis

**Objective:** Assess how effectively vehicles are utilized.



**Findings:**

* Rides per vehicle were analyzed to measure efficiency.

**Insights:**

* Aids in fleet optimization and cost reduction.

## 9 Surge Pricing Analysis

**Objective:** Evaluate the impact of surge pricing.

**Findings:**

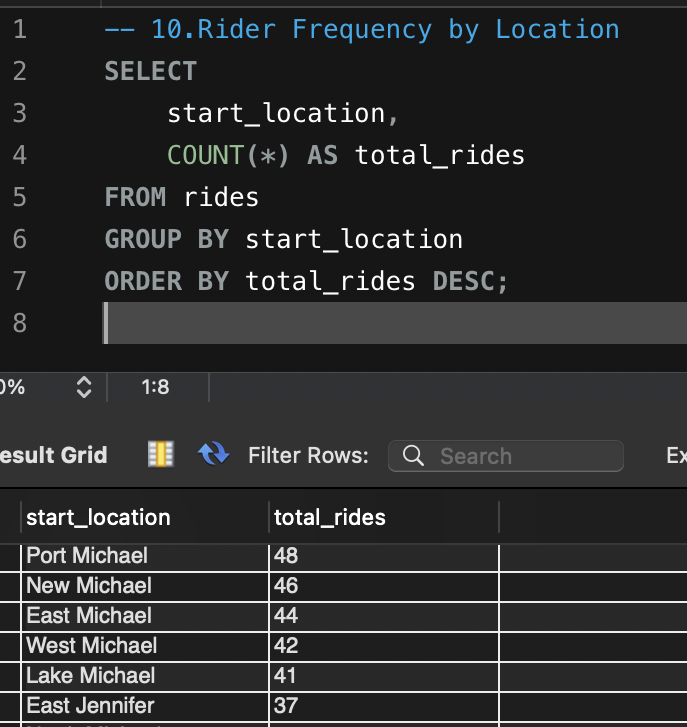
* Surge fare trends were analysed to determine peak pricing effectiveness.

**Insights:**

* Supports revenue maximization strategies.

## 10 Rider Frequency by Location

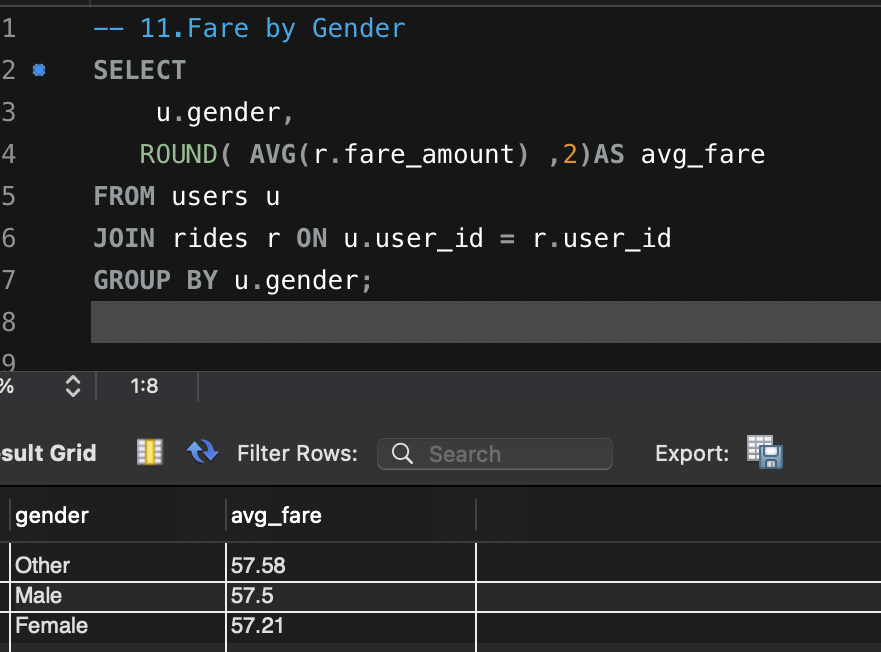
**Objective:** Analyze ride frequency by geographic area.



**Insights:**

* Identify which locations are most popular and could benefit from more drivers or better coverage.

## 11 Fare by Gender

**Objective:** Compare average fare amounts by gender.

**Insights:**

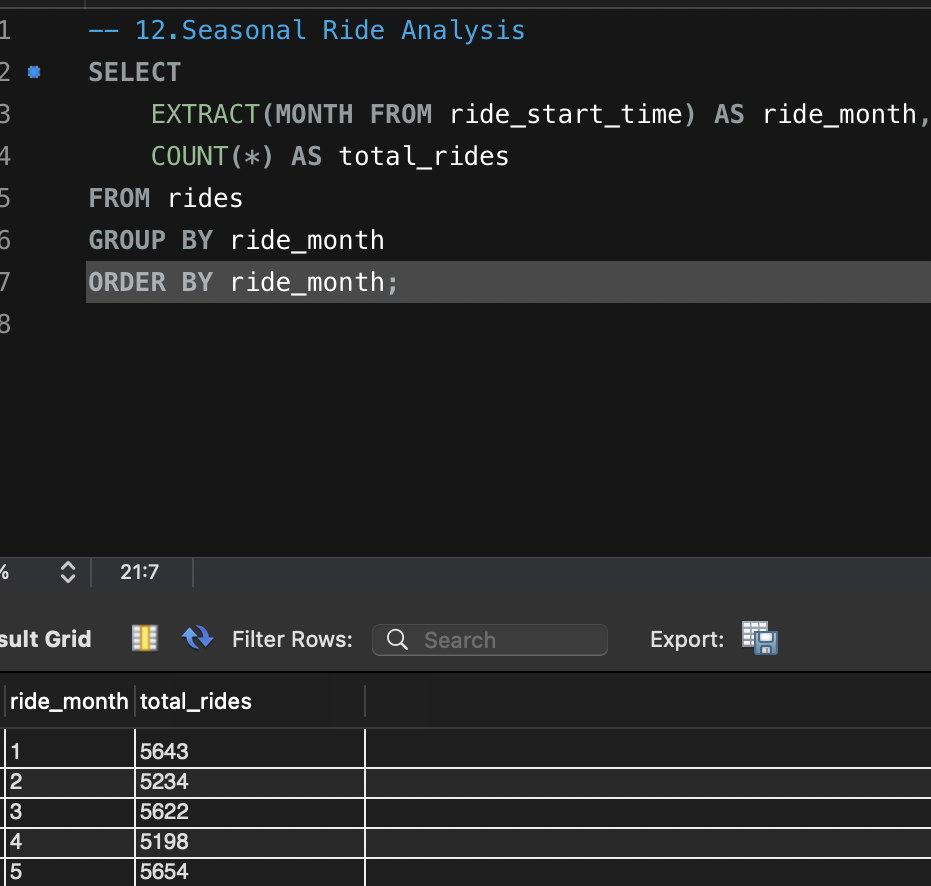
* Check if there are any fare discrepancies between genders, ensuring that pricing is fair across all users.

## 12 Seasonal Ride Analysis

* **Objective:** Determine ride trends by season

**Insights:**

* This helps identify seasonal spikes in demand, which could be important for driver scheduling, pricing, and promotions.



## 13 Rating Correlation with Fare Amount

**A screenshot of a computer

Description automatically generatedObjective:** Examine the relationship between fare and driver ratings.

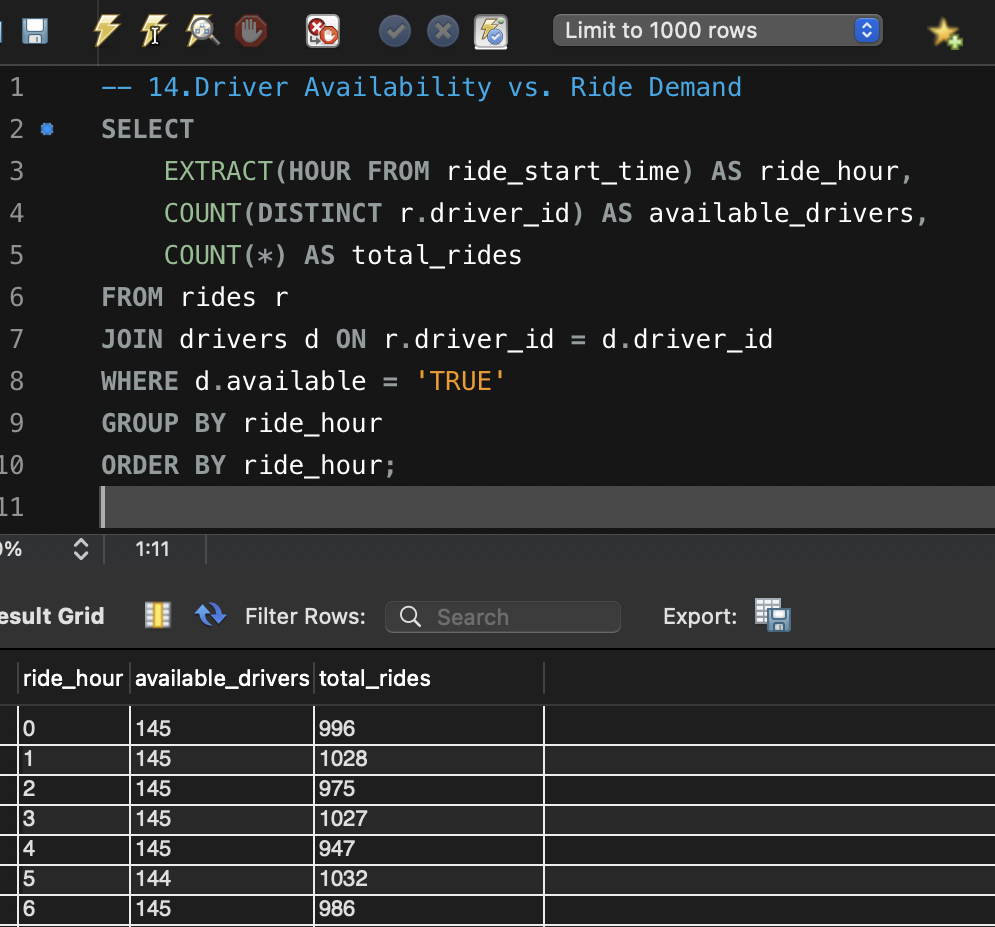
**Insights:**

* Helps to understand if customers are willing to pay more for a higher-rated driver, and can provide insights into setting fare structures.

## 14 Driver Availability vs. Ride Demand

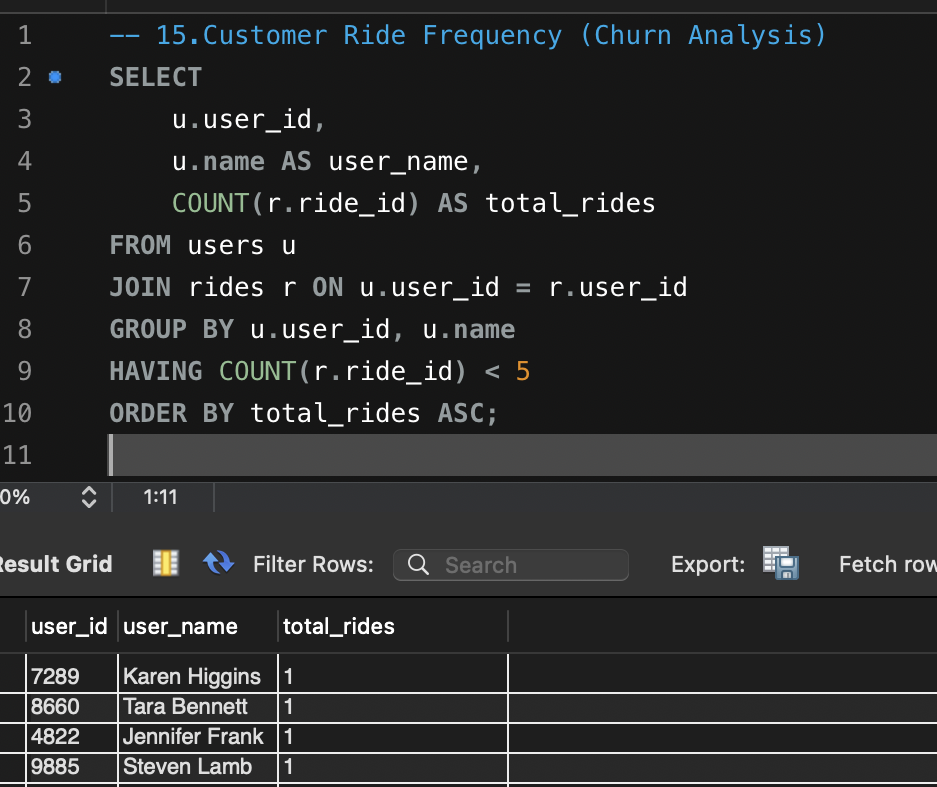
**Objective:** Balance driver supply with demand.

**Insights:**

* Helps to optimize driver availability during peak times to meet ride demand.

## 15 Customer Ride Frequency (Churn Analysis)

**Objective:** Identify churn risks among users.

****

**Insight:**

* This could identify potential churn risks, i.e., users who aren’t frequent enough or have fewer rides, and can be targeted for loyalty programs.

# Conclusion

The analysis of ride-sharing data provides valuable insights into optimizing operations, improving driver efficiency, and increasing profitability. Identifying peak ride hours allows for better resource allocation, while driver performance metrics help in ensuring service quality. Understanding customer ride patterns and spending habits enables targeted marketing strategies that enhance customer retention.

By implementing surge pricing effectively, companies can maximize revenue during high-demand periods. Vehicle utilization insights contribute to cost efficiency and fleet management. Furthermore, analyzing rating trends helps in maintaining service standards and improving customer satisfaction.

A correlation between fares and ride demand highlights the importance of dynamic pricing strategies. Monitoring churn rates ensures proactive customer retention measures. The integration of these insights into business strategies can lead to sustainable growth and enhanced competitiveness in the ride-sharing industry. Ultimately, leveraging data-driven decisions will create a more efficient and profitable ride-sharing ecosystem.

**GitHub Link:**

**Dataset Link:**

**LinkedIn Link:**