Report on Analytic Solution of NYC Taxi Fare

Written by Sahil Thummar
Professor Dr. Junaid Qazi

1. <u>Introduction</u>

1.1 Overview of the Open-Source Data

The dataset used in this project is publicly available from NYC Open Data, which provides detailed taxi trip records. The dataset includes trip details such as pickup and drop-off times, fare amounts, surcharges, and payment methods. It is a rich source for understanding transportation dynamics in New York City.

1.2 Purpose and Goals

1.2.1 Purpose

The purpose of this project is to analyze NYC taxi trip data to optimize fare strategies and identify congestion periods. By examining trip patterns and fare components, we aim to enhance pricing models, improve resource allocation, and maximize revenue during peak times.

1.2.2 Goals

- Analysis traffic patterns and congestion periods
- Track KPI'S
- Predict trip times and fare amounts
- Identify key areas for service improvement and fare optimization

1.3 Source and Dataset Selection

This dataset was selected from NYC Open Data because it offers comprehensive information on taxi rides, including fare breakdowns and travel patterns. It was chosen due to its relevance in studying urban mobility, fare optimization, and traffic congestion in one of the busiest cities in the world.

1.4 Expectations

From this dataset, we expect to find clear patterns in passenger demand, peak travel times, and the impact of surcharges on total revenue. Additionally, we aim to uncover insights on fare optimization by analysing how different factors, such as trip distance and congestion charges, contribute to overall fares.

2. <u>Research Questions</u>

2.1 Key Question 1

What are the busiest hours of the day?

This shows the times when taxi demand is highest, usually during morning and evening rush hours and late at night. Knowing these trends helps make sure there are enough taxis available when they're most needed.

2.2 Key Question 2

Which are top 3 location for pickup?

This shows the most popular pickup spots, like airports, business areas, or tourist attractions. Knowing these spots helps taxi companies place their cars in high-demand areas to serve more customers and run more efficiently.

2.3 Key Question 3

What are the busiest day of the week?

This analysis identifies which days have the most taxi trips. Weekdays may have more trips for commuting, while weekends might show a different trend with more leisure or event-related travel. This helps in planning and predicting the demand for taxi services.

2.4 Key Question 4

What are the top 2 vendor ?

This compares how well taxi service providers are doing to find the top two based on the number of trips or total earnings. Knowing the leading companies helps understand the competition and see trends in what customers prefer or how good the service is.

3. Metrics and KPIs

Metrics are data points used to measure performance, while KPIs (Key Performance Indicators) are important metrics that show how well a business is doing in reaching its goals.

3.1 Total Revenue

This metric represents the total income generated from all taxi rides, encompassing base fares, tips, surcharges, and any additional fees. It is crucial for evaluating the financial performance of the taxi business and identifying areas for potential revenue growth, helping operators make informed decisions to enhance profitability.

3.2 On-Time Performance

On-time performance measures the percentage of trips that start and finish as scheduled. High on-time performance indicates reliability, which is vital for ensuring customer satisfaction and efficient operations, ultimately leading to improved service quality and operational effectiveness.

3.3 Total Rides

This KPI measures the total number of completed trips over a specific time frame. It's essential for evaluating demand and operational scale, helping to pinpoint growth trends and opportunities for improving services.

3.4 Average Trip Duration

This metric measures the average time taken for taxi trips from pickup to drop-off. It is useful for analysing traffic patterns and service efficiency, helping to

understand how quickly drivers can complete trips under various conditions and improve overall operations.

4. Data Model

4.1 Overview of the Data

The dataset includes important details like pickup and drop-off times, location IDs, passenger counts, and total fares. These elements are vital for understanding taxi service patterns and customer habits. Analysing this data helps identify busy times, popular locations, and pricing strategies, providing valuable insights for improving service and increasing customer satisfaction.

4.2 Entity-Relationship Diagram (ERD)
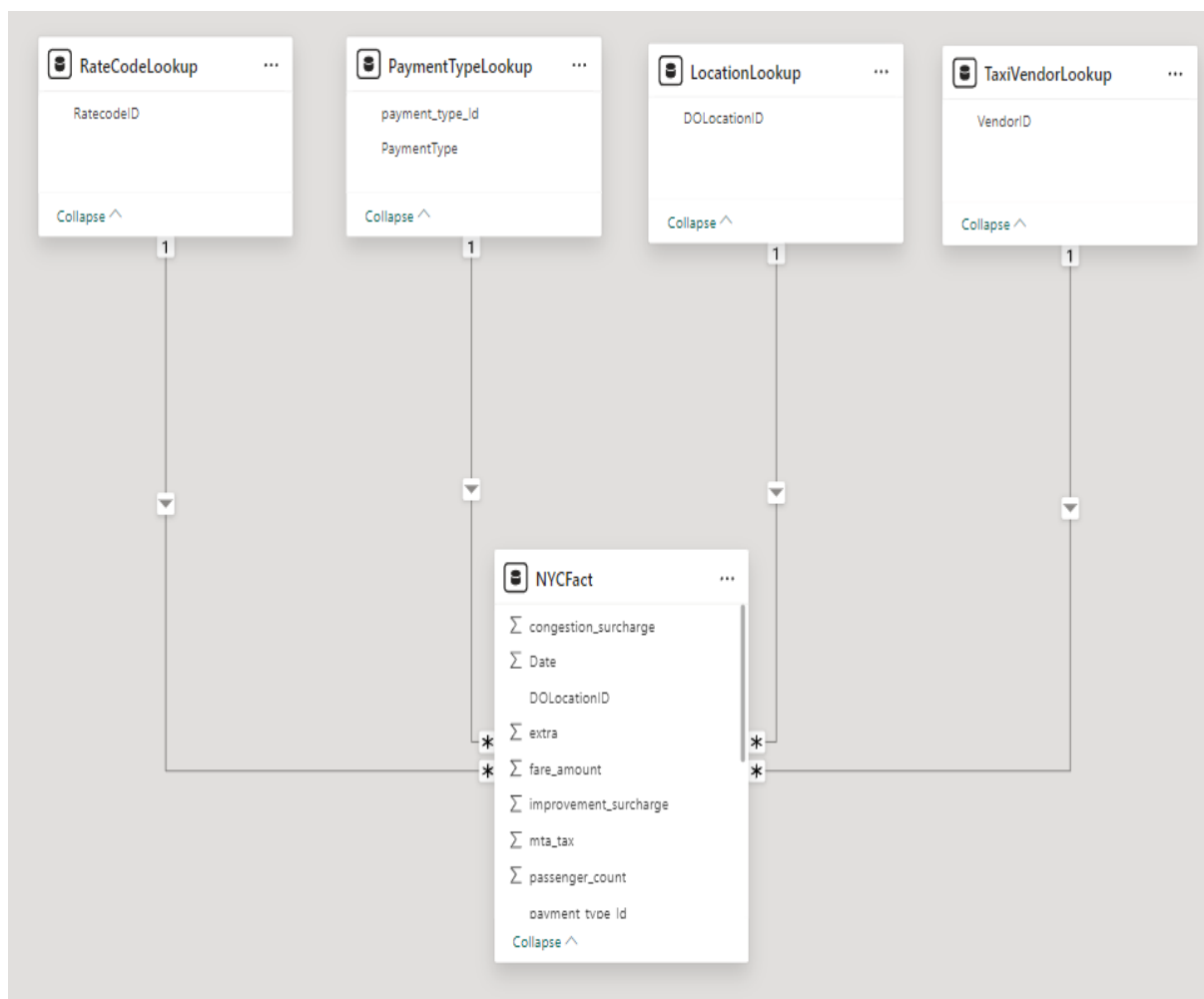
4.3 Other Tables (Measure Table, Group By Table)
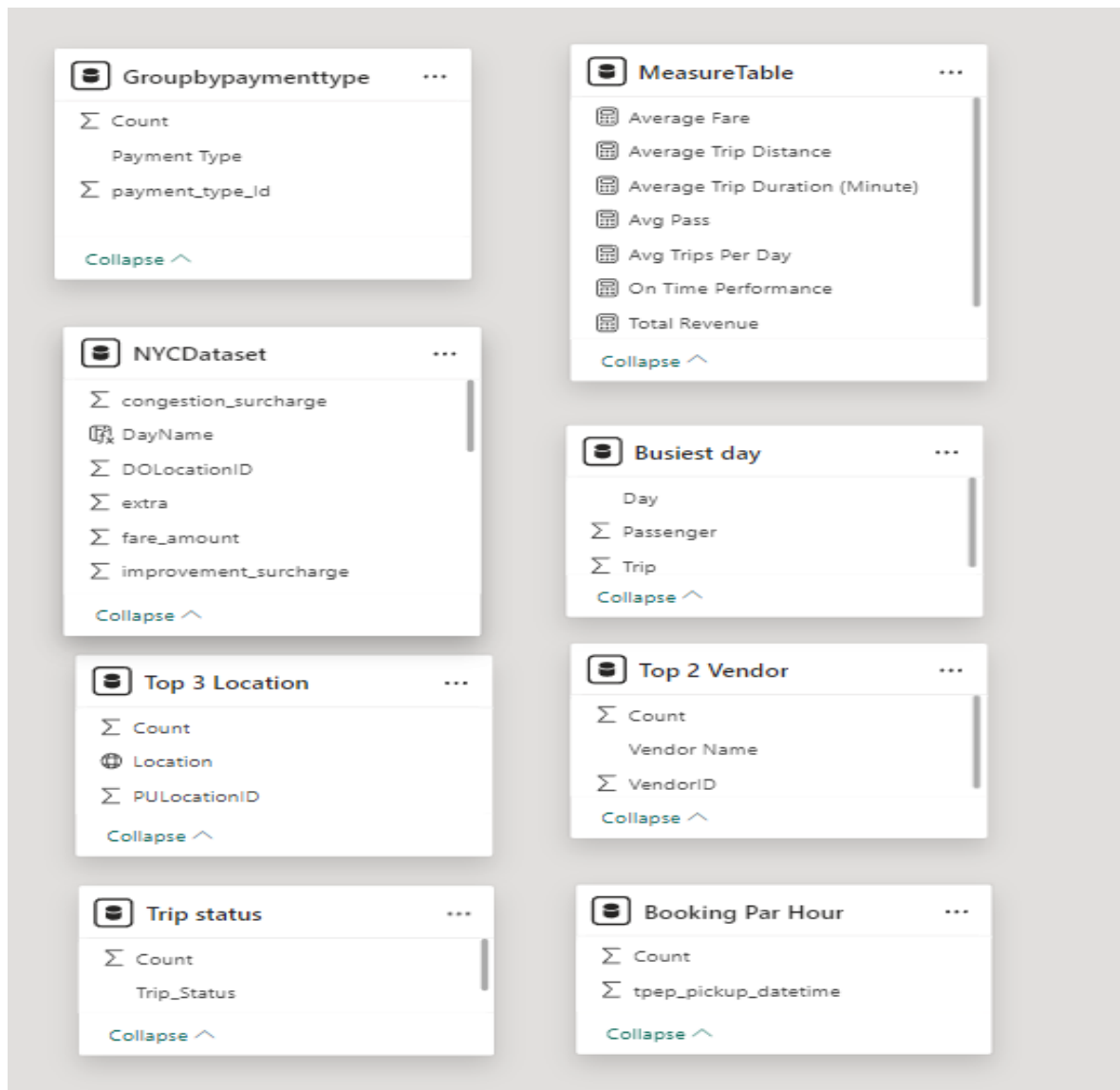
5. Results

Here's a brief explanation of the Each and Every Card And Graph:

1. **Total Revenue Card**
*   A card visualization for Total Revenue has been created, which dynamically updates based on the selected date range.
*   This enables quick access to the total income generated from taxi rides within the chosen period, offering a clear and concise view of revenue performance that aids in financial analysis and decision-making.

2. Average Trip Duration Card

- A card visualization for Average Trip Duration has been created, which dynamically updates based on the selected date.
- This allows users to see how the average duration of taxi trips varies with different date selections, providing insights into changes in travel times over specific periods. This feature enhances interactivity and aids in effectively analyzing trends in trip durations.


3. Total Rides Card

- A card visualization displays the Total Rides metric, dynamically updating to reflect the number of rides based on the selected date or date range from a slicer.
- This interactive feature enables users to easily observe how ride volumes change over time, offering valuable insights into trends and fluctuations in demand.

4. Slicer for select Date

- A slicer was created to allow users to select a specific date or date range, enabling dynamic updates to metric like the Total Rides, Total revenue and Average Trip Duration in the card visualization.
- This feature enhances interactivity and helps analyze ride volumes over different time periods.
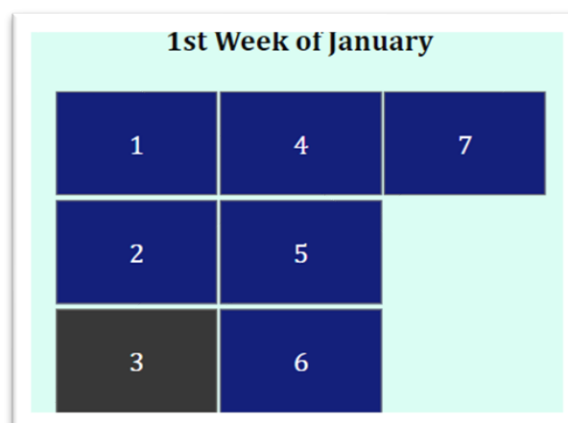
5. On Time Performance Card

• A card visualization for On-Time Performance has been developed, which updates dynamically based on the selected date or date range.

• This interactive feature allows users to monitor and analyse how on-time rates fluctuate over different periods, providing valuable insights into service reliability and performance trends in the taxi service.

6. Bar Graph Of Top 2 Vendor

• A bar graph was created to display the top two vendors, NYC Yellow Taxi and NYC Green Taxi, with the vendors represented on the x-axis and the ride count on the y-axis.

• This visualization helps compare the performance of these vendors based on the number of rides they provided.
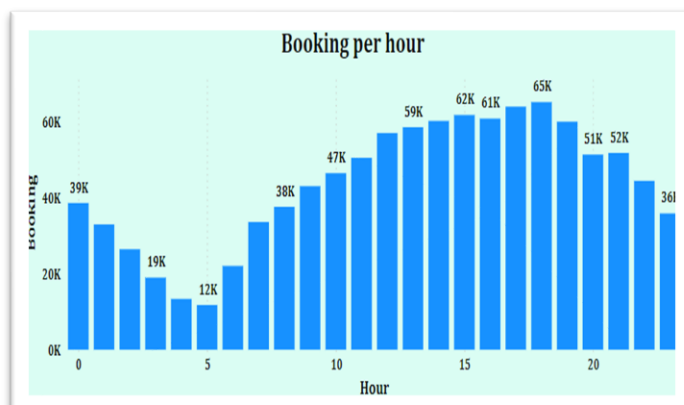

7. Bar graph for Booking par hour

• A bar graph was created to display bookings per hour, with hours represented on the X-axis and the number of bookings on the Y-axis.

• This visualization effectively illustrates the distribution of bookings throughout the day, making it easy to identify peak hours of activity.

8. Pie chart for Trip Status

• A pie chart was created to illustrate trip status, categorizing trips as completed, cancelled, or rejected.

• This visual representation provides a clear overview of the distribution of trip outcomes, helping to quickly assess service efficiency and customer satisfaction.
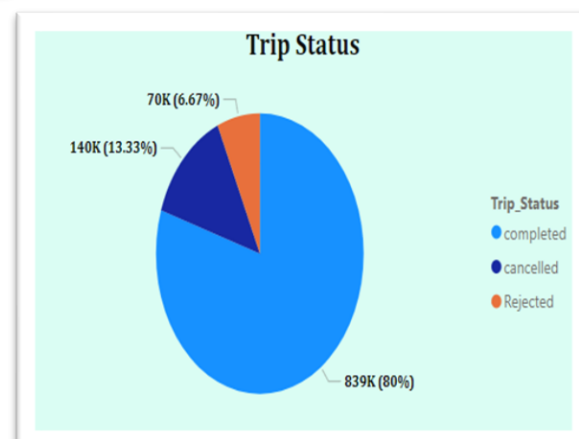
9. Map visualization for Top 3 pickup/ Drop-off

- A map visualization was created to display the top three pickup and drop-off locations.
- This visual representation highlights the most popular areas for taxi services, making it easy to identify key hotspots for both pickups and drop-offs in the dataset.

10. Clustered Bar Graph For Busiest Day

- A clustered bar graph was created to visualize the busiest days of the week, featuring days on the y-axis.

- The x-axis represents both the sum of trips and the sum of passengers, allowing for a comprehensive comparison of ride volumes and passenger counts.

- This visualization helps identify trends in service demand and provides insights into customer behaviour across different days..
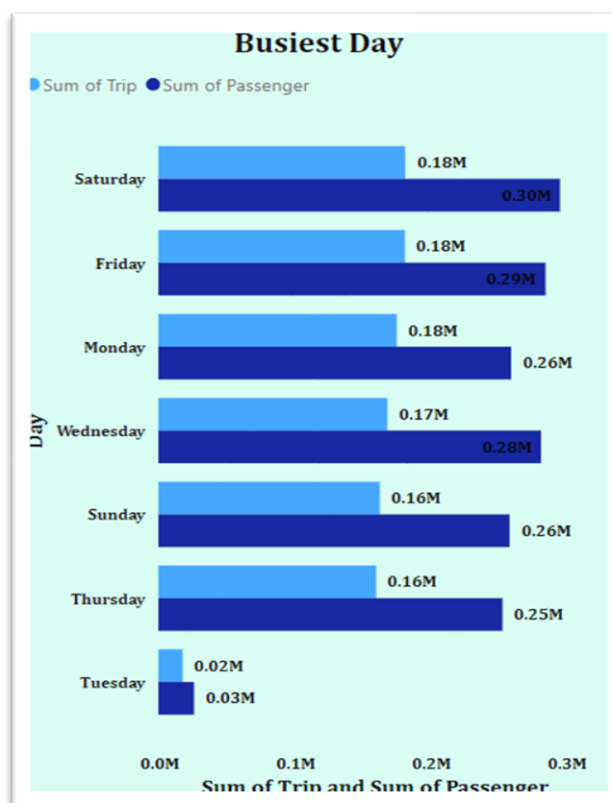
6. Conclusion

The dashboard offers a concise overview of key performance indicators (KPIs) for NYC taxi trips, visualizing essential metrics like total revenue, on-time performance, total rides, and average trip duration. These metrics help assess performance and operational efficiency. Visualizations, including a bar graph for the top two vendors, a graph for the busiest day, and a pie chart for trip status, facilitate easy comparison and understanding of service effectiveness.

In conclusion, the dashboard is vital for monitoring taxi service performance and provides stakeholders with insights to enhance efficiency and customer satisfaction. By focusing on important metrics, it supports informed decision-making and strategic planning, ensuring effective resource allocation to meet demand and improve overall service quality.

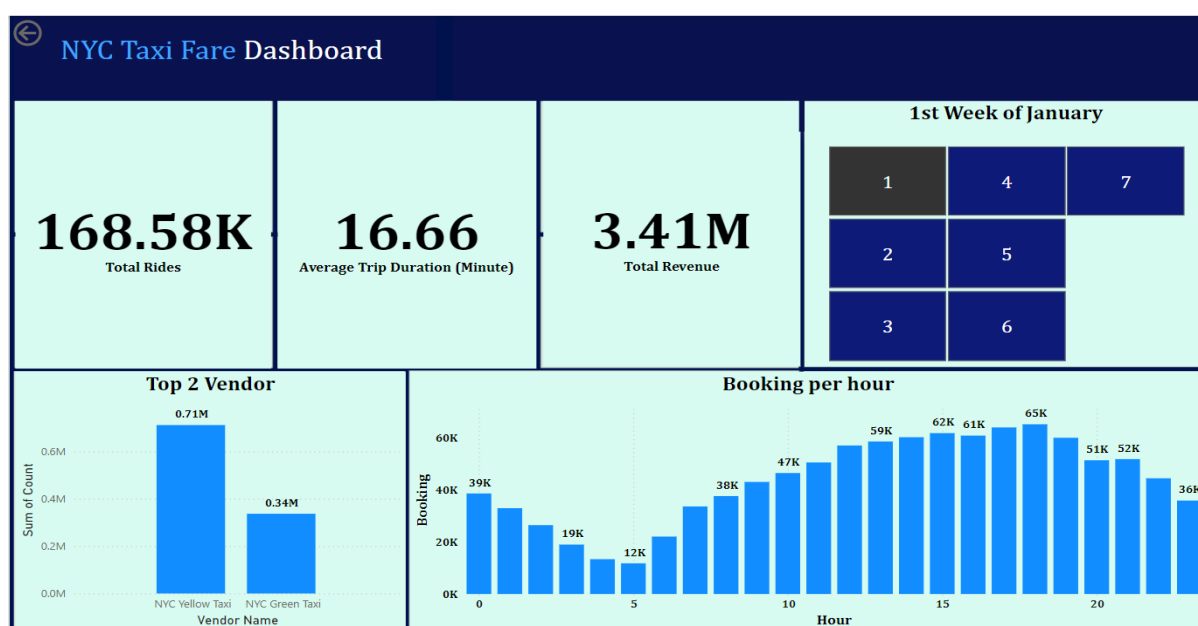## To see Power BI DashBoard Click here Power BI

7.Appendix

7.1 Power BI Dashboard

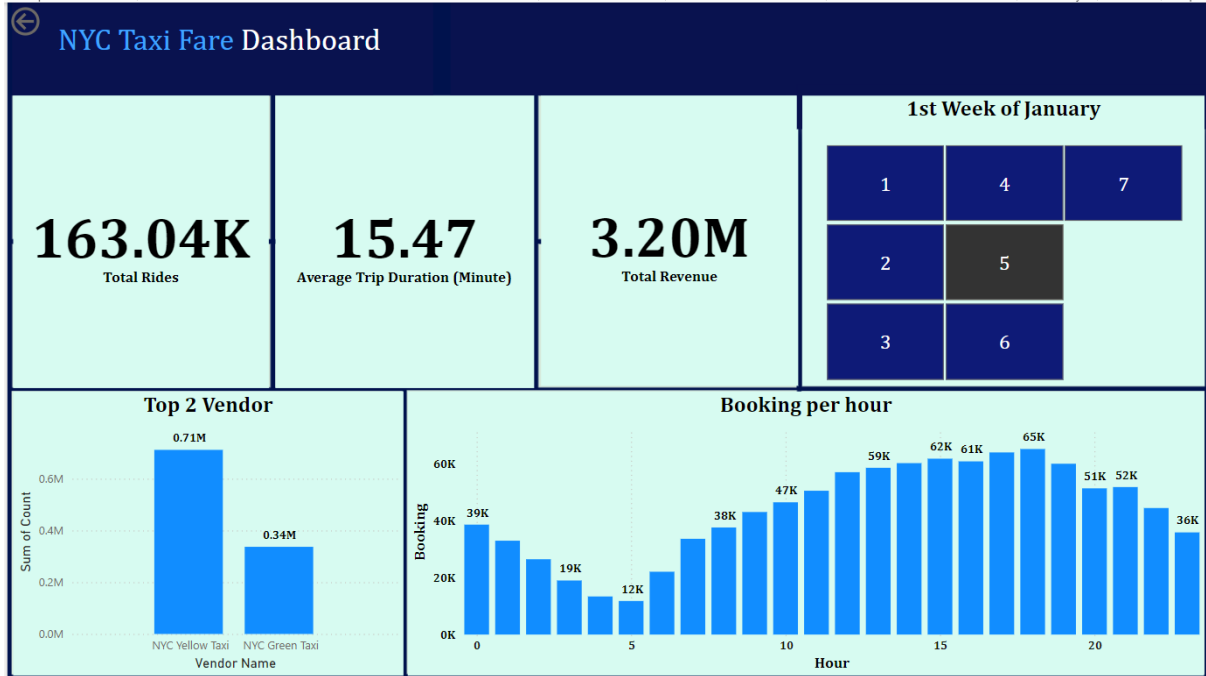Here's an example of how the dashboard behaves for two different dates:

Example for Date January 1
- Total Revenue: $3.41M
- Total Rides: 168.58K
- Average Trip Duration: 16.66 minutes

Example for Date January 5
- Total Revenue: $3.20M
- Total Rides: 163.04K
- Average Trip Duration: 15.47 minutes

By selecting different dates, the dashboard updates the values for these metrics, allowing for a quick comparison of performance across days.

Second Dashboard