NYC TAXI FARE – REPORT ON DATA DRIVEN SOLUTION

ABSTRACT

This article examines NYC taxi fare data using a Power BI dashboard, emphasizing crucial variables such as total revenue, average trip distance and many more. The visualizations demonstrate substantial patterns in passenger behaviour and attempts to improve taxi services' decision-making and operational efficiency.

Arlin George

GitHub

https://www.linkedin.com/in/arlin-george-382b1519b/
https://github.com/arlingeo99/DATA_SCIENCE_PROJECTS

Table of Contents

https://www.linkedin.com/in/arlin-george-382b1519b/
https://github.com/arlingeo99/DATA_SCIENCE_PROJECTS

NYC Taxi Fare - Report on Data Driven Solution

Introduction

The NYC Taxi Fare dataset provides a comprehensive picture of the cab transportation system in New York City and is derived from Kaggle. Numerous details are included in it, such as records of taxi trips, times of pick-up and drop-off, travel lengths, fare amounts, modes of payment and many more. Researchers and companies may investigate important aspects affecting transportation efficiency and service quality in one of the busiest cities in the world by using this dataset to do in-depth assessments of urban mobility. We can make data-driven decisions that enhance operations and maximise client experiences by examining this dataset to find trends and patterns in taxi usage.

Purpose And Goals

Purpose

This report analyses the data to construct a dashboard useful for weekly NYC taxi trip planning. It will focus on historical data to provide insight into the total number of trips, average trip duration, total revenue, on-time performance, busiest days and times of the week, popular pick-up location, and top two vendors. To provide data-driven solutions for the taxi sector, the **main purpose of this analysis is to improve taxi fare methods and assess times of heavy congestion**. This research specifically aims to comprehend how well NYC's taxi services operate and how different elements affect ride efficiency, consumer demand, and total income.

Goals

1. Analysis of traffic patterns and congestion periods: The traffic flow patterns were identified to understand which times of the day or week are more prone to congestion and how this would affect the rides and fare amounts.
2. Track KPIs: KPIs can be tracked to measure the performance of the NYC service through KPIs such as total revenue, trip duration, cancellation rate and more to identify the areas that need improvement.
3. Predict trip times and fare amounts: NYC Taxi service can anticipate the operating expense and income based on trip distance and durations.
4. Identify key areas for service improvement and fare optimization: Identifying areas in need of improvement in terms of service quality can be highlighted by optimizing the service during peak time or rush time or also by adjusting trip fares when the demand is high.

Key Questions for Analysis

Some of the key questions that we asked ourselves while exploring the dataset are:

1. *Which times of day are most popular for taxi rides, and how do they relate to trends in traffic?*
   We hope to determine times of high demand and assess how congestion affects travel times and fare amounts by looking at ride volume at various times of the day.

https://www.linkedin.com/in/arlin-george-382b1519b/
https://github.com/arlingeo99/DATA_SCIENCE_PROJECTS

2. ***What reasons lead to delays, and how does the average journey length vary by borough and hour?***
We can better understand how weather, traffic, and time of day impact service efficiency by looking at journey duration.

3. ***How much money was made overall on each day, and how does the demand for rides affect revenue trends?***
By examining revenue patterns, we can determine the most lucrative days and hours for taxi services, which provides information for resource allocation and strategic planning.

KPIs and Metrics

During the exploration and analysis of the dataset, we identified some important KPIs to track both operational and strategic objectives.

i. ***Total Revenue***
This metric depicts the total sum of fare amounts over the period (1st week of January 2020) being analysed. This is a key indicator that measures business performance, showing how much income they generate from all trips. This reflects the overall business health and demand for NYC Taxi service.

ii. ***Total Rides***
The total number of taxi rides is tracked by this KPI. It helps in tracking demand over time, and customer usage patterns helping us to identify the trends that can be very useful in managing taxi availability and also plan service schedules.

iii. ***Average Trip Duration (in Minutes)***

This KPI reflects trip efficiency and can highlight traffic or route optimization opportunities. Longer travel times are frequently a sign of more traffic, which may be reduced with more effective route planning or dynamic pricing techniques.

iv. ***On-Time Performance***
This metric measures how efficient the taxi trips are, specifically in terms of the arrival and drop-off. This KPI tracks how often the taxis reach their destination on scheduled or estimated time. This would help us to identify the dependability of the taxi service.
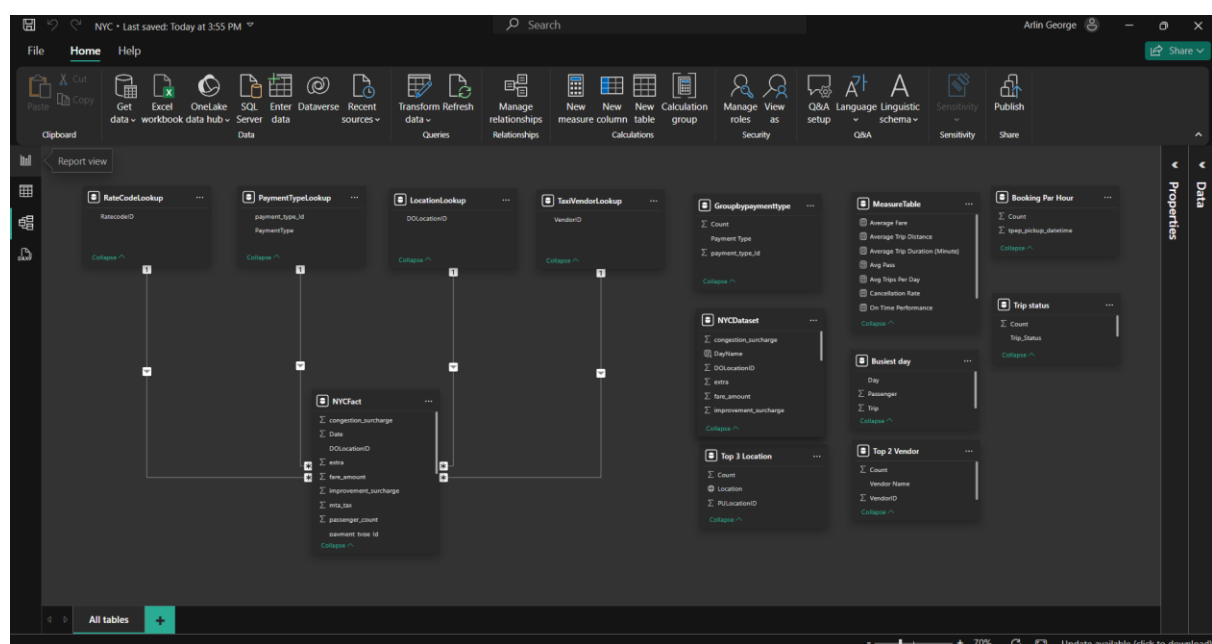
v. ***Cancellation Rate***
This KPI tracks the percentage of trips that were cancelled either by the driver or the passenger. A high cancellation rate may be a sign of problems like bad customer service, erroneous wait time estimates, or traffic jams. Taxi drivers may increase client satisfaction and service dependability by tracking the cancellation rate and implementing remedial measures.

https://www.linkedin.com/in/arlin-george-382b1519b/
https://github.com/arlingeo99/DATA_SCIENCE_PROJECTS

Objectives

1. Determine Ride Patterns: To maximize the availability and resource allocation, examine peak times and places.

2. Evaluate Service Efficiency: To make operational effectiveness, monitor KPIs such as average trip time, total rides, and cancellation rates.

3. Monitor Revenue: To comprehend profitability across various trip kinds, keep track of total revenue as well as revenue per mile.

4. Analyse Congestion Impact: With an emphasis on enhancing on-time performance, investigate how traffic impacts travel times and service dependability.

5. Create models to predict journey times and charges in order to improve customer service and planning.

6. Minimize Cancellations: Determine the reasons behind cancellations and suggest ways to increase service dependability.

7. Improve the Customer Experience: Make use of data to cut down on wait times, increase the number of available rides, and modify prices in response to demand.

8. Reduce Expenses: Find operational and route inefficiencies to save expenses and enhance overall service quality.

Data Model

This data model includes lookup tables for RateCode, PaymentType, Location, TaxiVendor and all linked to the NYCFact through RateCode ID, PaymentType ID, Location ID, Vendor ID.

https://www.linkedin.com/in/arlin-george-382b1519b/
https://github.com/arlingeo99/DATA_SCIENCE_PROJECTS

Measure Table: I created measures such as Average Fare, Average Trip Distance, Average Trip Duration (in Minutes), Average Trips per Day, Cancellation Rate, On Time Performance, Total Revenue, Total Rides and Average Pass for visualizations

GroupBy Tables: I built four GroupBy Tables:

1. Booking Per Hour
2. Trip Status
3. Top 2 Vendor
4. Top 3 Location

These were used to create the dashboard insights.

Results

A date slicer is used to represent and analyse the data for specific time period (1st week of January for the year 2020).

The above result is for **7th January 2020.**

- A total of 17.77K rides were completed, indicating a high level of demand following the Christmas season.
- The average travel duration was 13.26 minutes, indicating that traffic flow was efficient and that most trips were short.
- Taxi operations brought in $345.59K, demonstrating strong profitability even in the early January.

https://www.linkedin.com/in/arlin-george-382b1519b/
https://github.com/arlingeo99/DATA_SCIENCE_PROJECTS

Inference: Taxi services are consistently used, even outside of peak seasons, as evidenced by the great demand and significant earnings. Because of the efficient traffic circumstances implied by the low average journey length, businesses run more smoothly and customer turnover is higher. These results have the potential to enhance fleet management and boost overall service effectiveness.
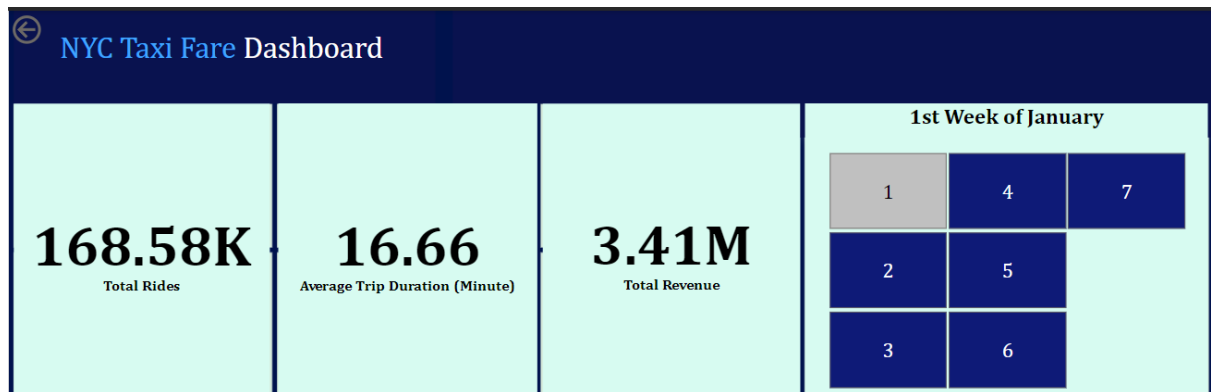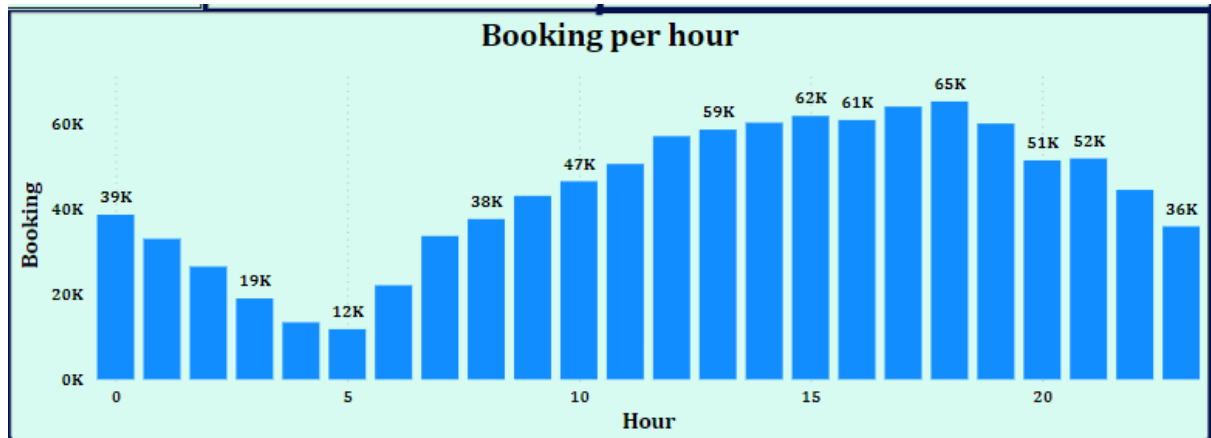

The above result is for **1st January 2020**

- A total of **168.58K rides** were completed, indicating a strong demand that was probably caused by the necessity for public transportation and the celebrations of the New Year.
- The average trip took **16.66 minutes**, which is longer than normal and may be the result of more traffic over the holidays.
- Taxi services brought in a total of **$3.41M**, a notable increase that was probably brought on by the high demand over the holidays.

Inference: The significance of taxis as the main form of transportation on New Year's Day is demonstrated by the large number of rides and income. Higher total ride numbers and robust income production offset the longer trip durations, which indicate greater traffic. This data emphasizes how important it is to schedule for holidays when demand is at its highest in order to optimize service.
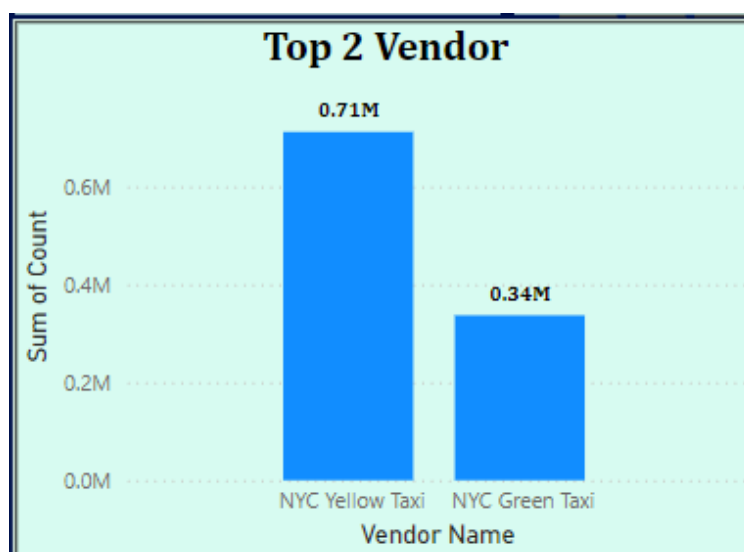

Vendor performance analysis reveals that **NYC Yellow Taxis completed 0.71M** bookings, substantially exceeding **NYC Green Taxis's 0.34M** bookings. This demonstrates how Yellow Taxis dominates the transportation market in NYC.

https://www.linkedin.com/in/arlin-george-382b1519b/
https://github.com/arlingeo99/DATA_SCIENCE_PROJECTS

According to the hourly booking trend, activity peaks between **15:00 and 18:00, with 65K trips booked at 18:00.** With just **12K trips, the lowest booking** count occurred at 5:00 AM, most likely as a result of low early morning demand.

This makes it obvious that market share advantage over Green Taxis, Yellow Taxis are crucial to NYC's transit network. The busiest hours are in the afternoon and early evening, indicating that fleet allocation and operational planning should concentrate on these periods.
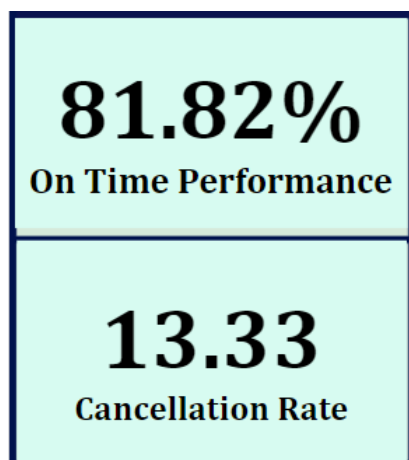

According to the data, **81.82%** of the trips were **finished on schedule**, suggesting that the service is typically dependable. Although this degree of punctuality indicates effective operations, there is still opportunity for improvement, particularly during peak hours.

13.33% is a rather high **cancellation rate**. This implies that a sizable percentage of journeys were not finished as planned, maybe as a result of traffic, client conduct, or other outside variables.

Although the on-time performance is good, the cancellation rate suggests that scheduling, customer service, or driver availability may need some work. Lowering the cancellation ratemay improve operational effectiveness and customer happiness.
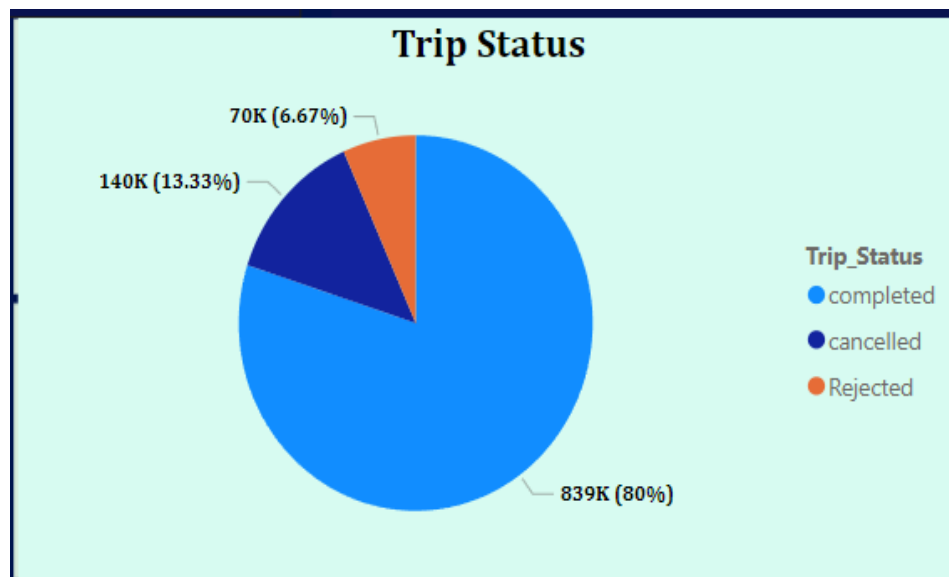
- Completed: 839,000 trips (80%)
- Cancelled: 140,000 trips (13.33%)
- Rejected: 70,000 trips (6.67%)
- The fact that 80% of journeys are finished shows that most requests are fulfilled.
- The 13.33% cancellation rate raises the possibility that there are problems driving drivers/passengers to cancel their journeys, which may need more research.
- A 6.67% rejection rate means that very few requests are turned down, which is a good indicator of service dependability.
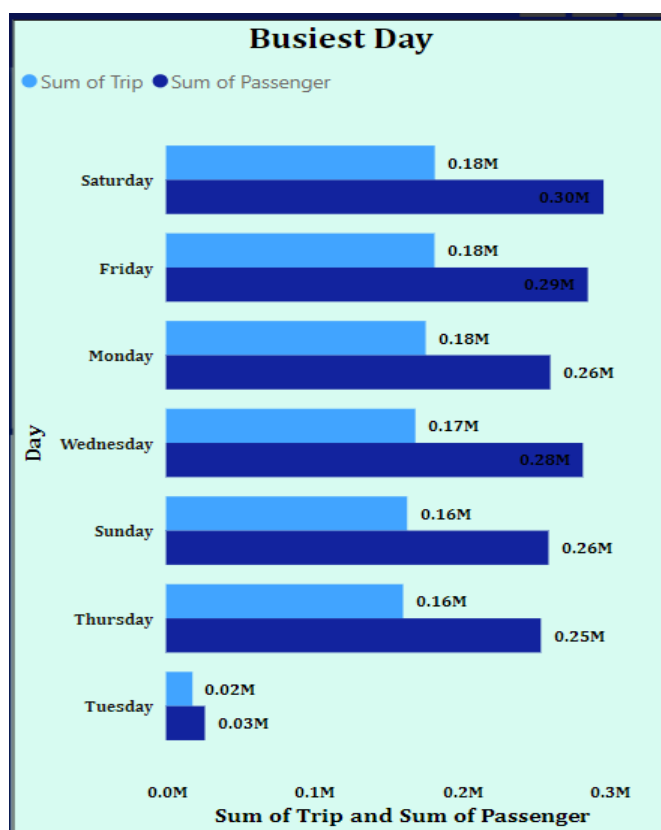
The Busiest day is visualised in a clustered bar chart. The chart depicts that the busiest day is Saturday followed by Friday. On Saturday, a total of 0.18M trips and 0.30M total passengers travelled whereas on Friday, 0.18M trips and 0.29M passengers travelled, indicating a higher demand for rides during the weekend. The least busy day is Tuesday with a total of 0.02M trips and 0.03M passengers.

This information helps to analyse the user behaviour and make decisions for operational improvement.

https://www.linkedin.com/in/arlin-george-382b1519b/
https://github.com/arlingeo99/DATA_SCIENCE_PROJECTS

The heat map shows top three busy locations for pickup/drop-off. The highest area in demand in terms of taxi rides is Manhattan Midtown with a total of 47923 rides. The second most busy zone is with a total of 38638 rides followed by 38656 rides to Lenox Hill. According to this statistics, Manhattan Midtown is the main location for taxi activity, highlighting the necessity of improved services and resources to meet the increasing demand.
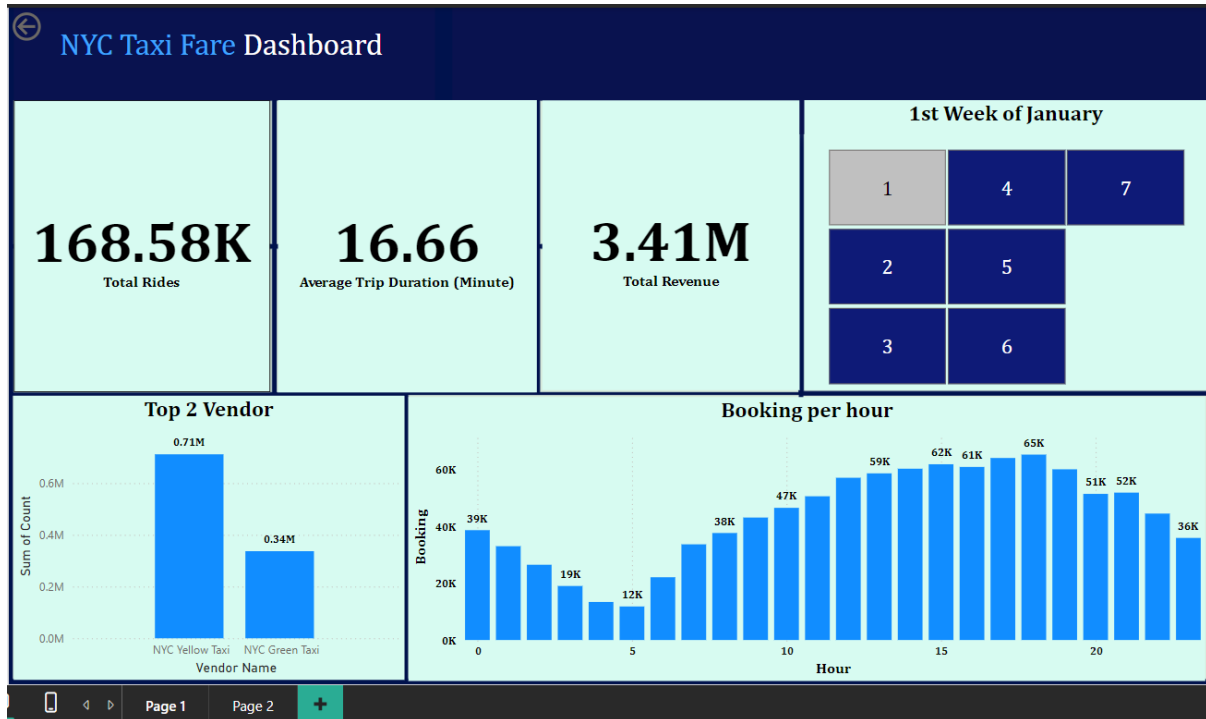
Conclusion

To conclude, the NYC Taxi Fare Analytics dashboard offers a concise and useful perspective on taxi performance across a number of important variables, such as total income, total trips, average journey length, and peak hours. Managers of taxi services may use these insights to inform data-driven choices that will increase client pleasure, optimize fare strategies, and improve operational efficiency. Taxi operators can employ tactics that strike a balance between demand and profitability by concentrating on peak hours, trip duration trends, and revenue production. This research demonstrates the possibility for increasing service delivery by modifying resources based on real-time traffic patterns and congestion data, allowing taxi services to function more successfully in a highly competitive urban setting.

References
Dataset from Kaggle: NYC TAXI FARE Dataset

https://www.linkedin.com/in/arlin-george-382b1519b/
https://github.com/arlingeo99/DATA_SCIENCE_PROJECTS

Appendix

The dashboard shows the report for 1st January 2020

This the second dashboard.

https://www.linkedin.com/in/arlin-george-382b1519b/
https://github.com/arlingeo99/DATA_SCIENCE_PROJECTS