INFOSYS MILESTONE-2

Milestone 2 – Infosys Internship

TOPIC: Uber Data Analytics using Power BI

Introduction:

To uncover insights into Uber's trip patterns, popular pickup/drop-off locations, peak hours, and trip frequency over time. These insights can help Uber in operational planning, resource allocation, and improving user satisfaction. By using Power BI's powerful data visualization and analytical tools, we can generate insights that are not only valuable for operations but also beneficial in enhancing service quality.

DAX Measures and Calculations:

1. DateTable:

```
DateTable =
ADDCOLUMNS (
    CALENDAR(DATE(2024,1,1), DATE(2024,12,31)),
    //CALENDARAUTO(),
    "Year", YEAR([Date]),
    "Quarter", "Q" & FORMAT(CEILING(MONTH([Date])/3, 1), "#"),
    "Quarter No", CEILING(MONTH([Date])/3, 1),
    "Month No", MONTH([Date]),
    "Month Name", FORMAT([Date], "MMMM"),
    "Month Short Name", FORMAT([Date], "MMM"),
    "Month Short Name Plus Year", FORMAT([Date], "MMM,yy"),
    "MonthSort", FORMAT([Date], "yyyyMM"),
    "DateSort", FORMAT([Date], "yyyyMMdd"),
    "Day Name", FORMAT([Date], "dddd"),
    "Details", FORMAT([Date], "dd-MMM-yyyy"),
    "WeekSort", FORMAT([Date], "w"),
    "Day Number", DAY ( [Date] )
)
```

This DAX code creates a DateTable in Power BI by defining a calendar range from January 1, 2024, to December 31, 2024, using the CALENDAR function. It then adds a series of date-related attributes with ADDCOLUMNS, enhancing the table for various time-based analyses. Here's a breakdown of each column:

1. **Year** - Extracts the year from the Date field.
2. **Quarter** - Generates a label for each quarter in "Q1," "Q2," etc., by using CEILING to calculate the quarter based on the month.
3. **Quarter No** - Provides the numeric representation of the quarter (1, 2, 3, or 4).
4. **Month No** - Extracts the numeric month value from each date (1 for January, 2 for February, etc.).
5. **Month Name** - Displays the full month name (e.g., "January").
6. **Month Short Name** - Shows the abbreviated month name (e.g., "Jan").
7. **Month Short Name Plus Year** - Combines the month abbreviation and two-digit year (e.g., "Jan,24").
8. **MonthSort** - Formats the date as "yyyyMM" to ensure accurate chronological sorting by month.
9. **DateSort** - Provides a sortable full date format in "yyyyMMdd."
10. **Day Name** - Displays the full name of the day (e.g., "Monday").
11. **Details** - Shows the complete date in "dd-MMM-yyyy" format for readability.
12. **WeekSort** - Generates the week number in the year for organizing dates by week.
13. **Day Number** - Extracts the day of the month (1 to 31).

2. Toggel:

```
Toggel = {

   ("Total Trips", NAMEOF('Trip Details'[Total Trips]), 0),
   ("Total Fare", NAMEOF('Trip Details'[Total Fare]), 1),
   ("Total Duration", NAMEOF('Trip Details'[Total Duration]), 2),
   ("Total Distance", NAMEOF('Trip Details'[Total Distance]), 3) }
```

Toggle table can be used in a slicer or as part of a DAX SWITCH function to allow users to switch dynamically between metrics such as **Total Trips**, **Total Fare**, **Total Duration**, and **Total Distance**.

3. Night Shift %:

```
Night Shift % =  VAR Nightcount = CALCULATE([Total Trips],'Trip Details'[Shift]= "Night")
RETURN
DIVIDE(Nightcount,[Total Trips],0)
```

The Night Shift % measure calculates the percentage of trips that occur during the night shift. It filters the total trips to include only those with a "Night" shift label, then divides this count by the overall total trips.

4. Total Distance:

```
Total Distance = SUM('Trip Details'[Distance])
```

It is a measure calculates the sum of the Distance column in the Trip Details table.

5. Total Duration:

Total Duration = SUM('Trip Details'[Duration])

It is a measure calculates the sum of the Duration column in the Trip Details Table,

6. Total Fare:

Total Fare = SUM('Trip Details'[Fare])

It is a measure calculates the sum of the Fare column in the Trip Details Table,

7. Total Trips:

Total Trips = COUNT('Trip Details'[Trip ID])

It is a measure calculates the count of the Trip ID column in the Trip Details Table,

8. Shift:

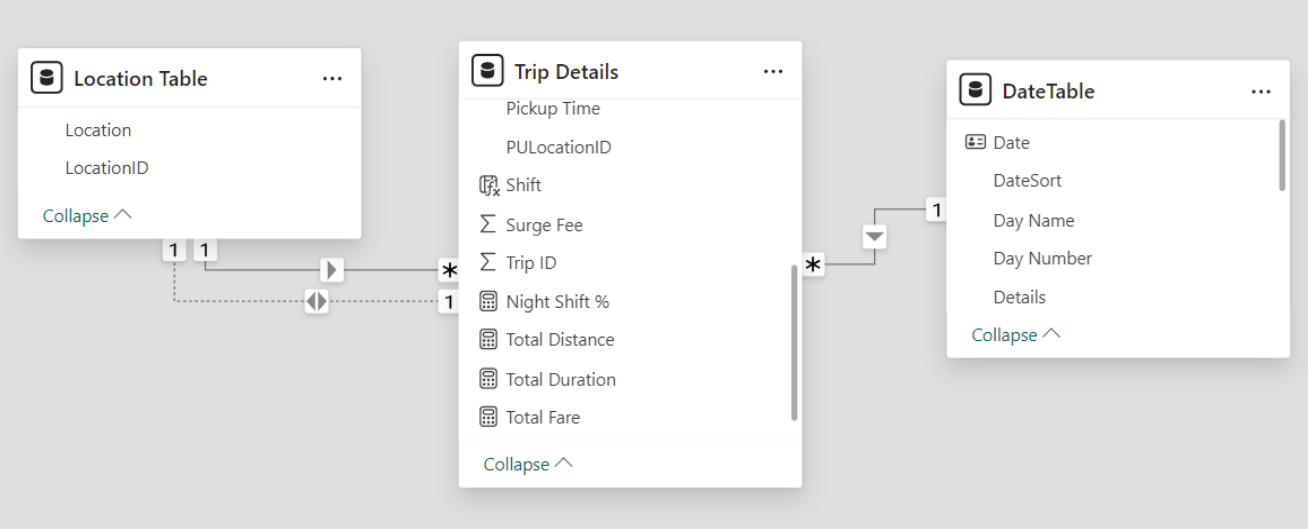
Shift =

```
VAR _hr = HOUR('Trip Details'[Pickup Time])
return
SWITCH(
TRUE(),
   _hr>=6 && _hr<=21, "Day",
   _hr>21 && _hr<=23, "MidNight",
   _hr>=0 && _hr<6, "Night",
BLANK())
```

The Shift measure determines whether a trip falls during the day or night based on the pickup time. It first extracts the hour of the Pickup Time using the HOUR function and stores it in the variable _hr. The SWITCH function then evaluates this hour to categorize the shift:

- If the hour is between 6 AM and 9 PM (_hr >= 6 && _hr <= 21), the shift is labeled as "Day".
- If the hour is between 9 PM and midnight (_hr > 21 && _hr <= 23), or between midnight and 6 AM (_hr >= 0 && _hr < 6), the shift is labeled as "Night".

Understanding Relationships in Power BI:



Many-to-One RelationShip:

The **Location Table** has a many-to-one relationship with the **Trip Details** table, meaning each unique location in the **Location Table** can be associated with multiple trips in the **Trip Details** table. This allows one location to appear across various trip records. Similarly, the **Date Table** has a many-to-one relationship with the **Trip Details** table, where each unique date in the **Date Table** can correspond to multiple trips.

One-to-One RelationShip:

The Trip ID column in the Trip Details Table and Location ID column in the Location Table are connected using One-One Relationship.

Distance Hierarchies:

By using the **Distance Hierarchy**, you can analyze the data by starting with a broad look at Distance and then drilling down to see details by DOLocationID. This setup is useful for analyzing distances per drop-off location and understanding location-based travel patterns.

Row-Level Security and Data Modeling:

In Power BI, **Row-Level Security (RLS)** allows you to control access to data at the row level, so that users can only view the data relevant to them. RLS is implemented by creating **roles** with specific DAX (Data Analysis Expressions) filter rules. For more complex scenarios, a security table mapping users to access rights can provide dynamic filtering based on user credentials. RLS ensures data security, a personalized user experience, and compliance with data privacy standards.

1. Distance:

For the **Distance** role, a filter is applied to the **Trip Details** table, where only records with a Distance equal to 1.2. This means users assigned to the **Distance** role will only see trip details where the distance is exactly 1.2, helping control access to specific data based on distance criteria.

2. Dynamic Shift:

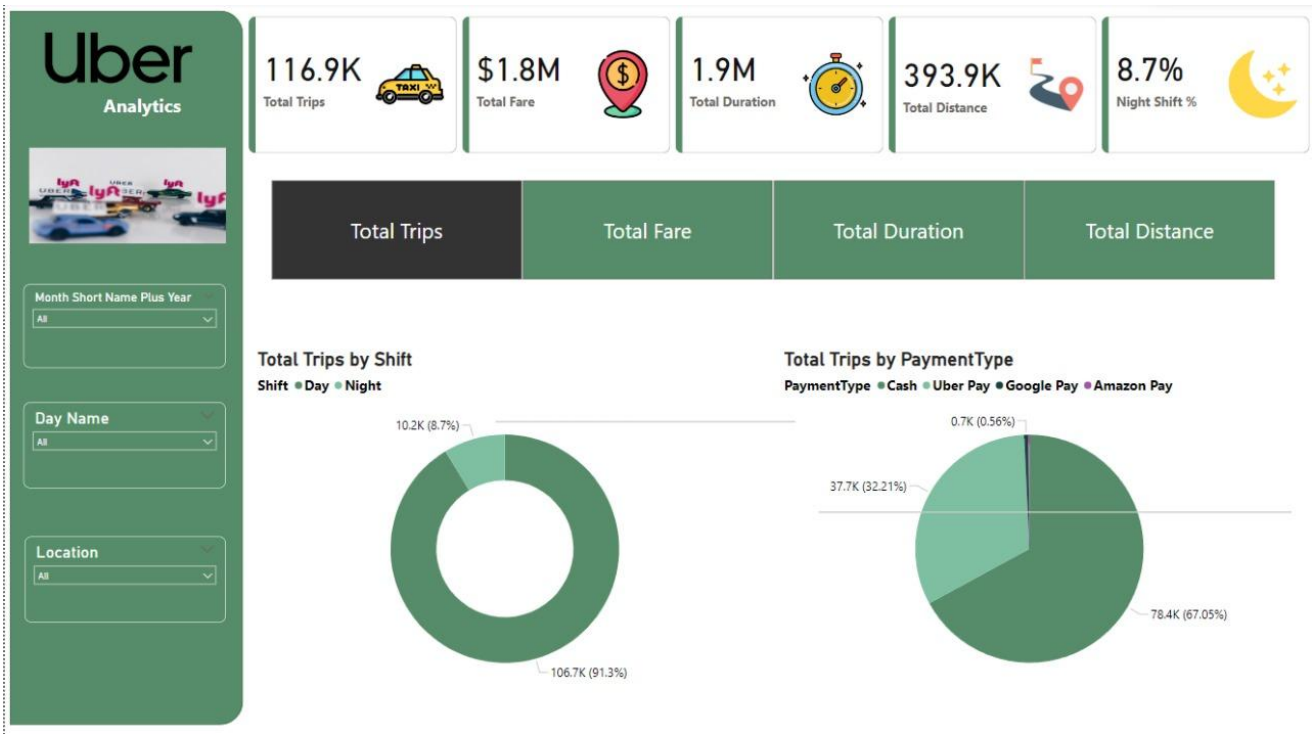
For the **Dynamic Shift** role, a filter is applied to the **Trip Details** table, where only records with a "Shift" equal to "Day". This means users assigned to the **Dynamic Shift** role will only see trip details where the Shift is exactly Day, helping control access to specific data based on Shift criteria.

View As Role:

We have to select the role, the report will refresh to display only the data that is accessible based on the defined condtions. Additionally, test other roles as necessary to confirm that they also function correctly and that no unauthorized data is accessible.

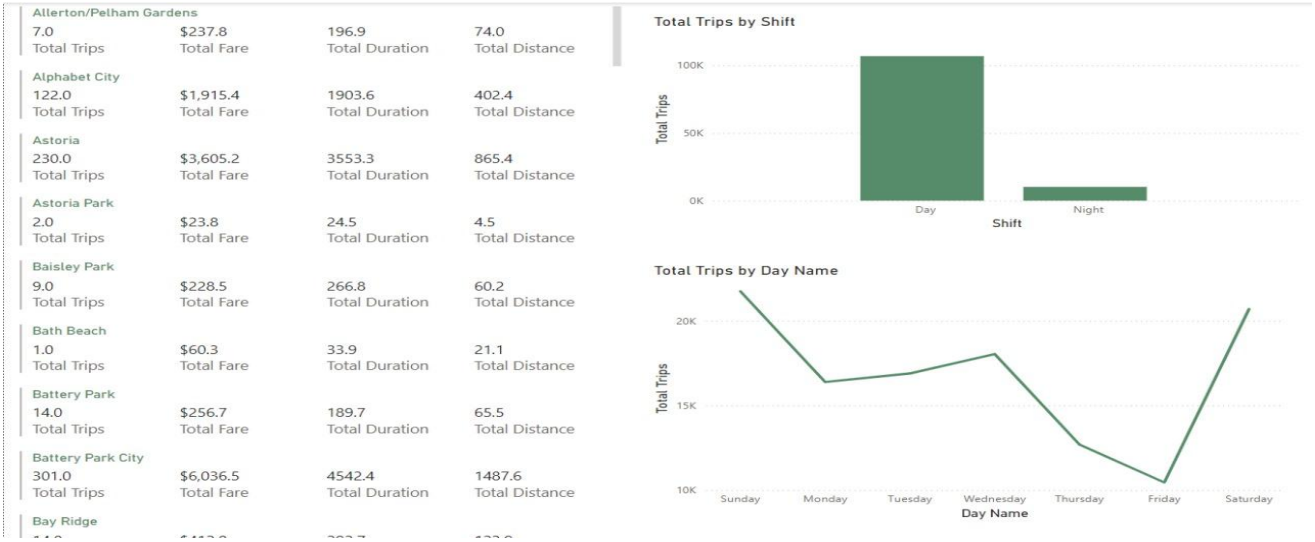
Report:

Report Image 1:



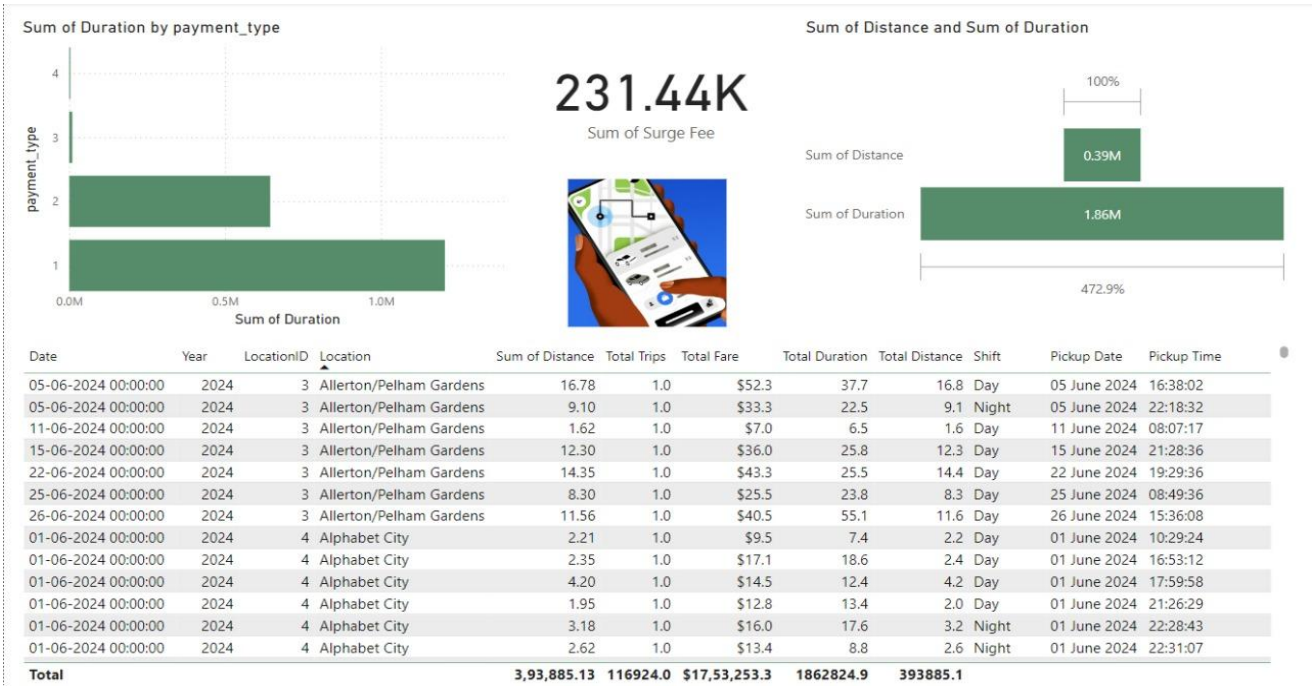
The image shows an Uber Analytics dashboard that provides key metrics about trips, fares, duration, distance, and shifts. The top section displays high-level statistics such as the total number of trips, total fare, total duration, total distance, and the percentage of night trips. Below, there are two pie charts: one showing the distribution of trips by shift (91.3% during the day and 8.7% at night) and the other showing payment types used, where cash is the most popular, followed by Uber Pay and smaller shares for Google Pay and Amazon Pay. The sidebar on the left includes filters for month, day, and location to refine the data displayed.

Report Image 2:



This dashboard provides an overview of trip data by neighborhood, showing metrics like total trips, fare, duration, and distance for each area. A bar chart reveals that most trips occur during the day, while a line chart highlights daily patterns, with fewer trips on Fridays and a peak on Saturdays. This data helps identify high-demand neighborhoods and optimal times for trips.

Report Image 3:

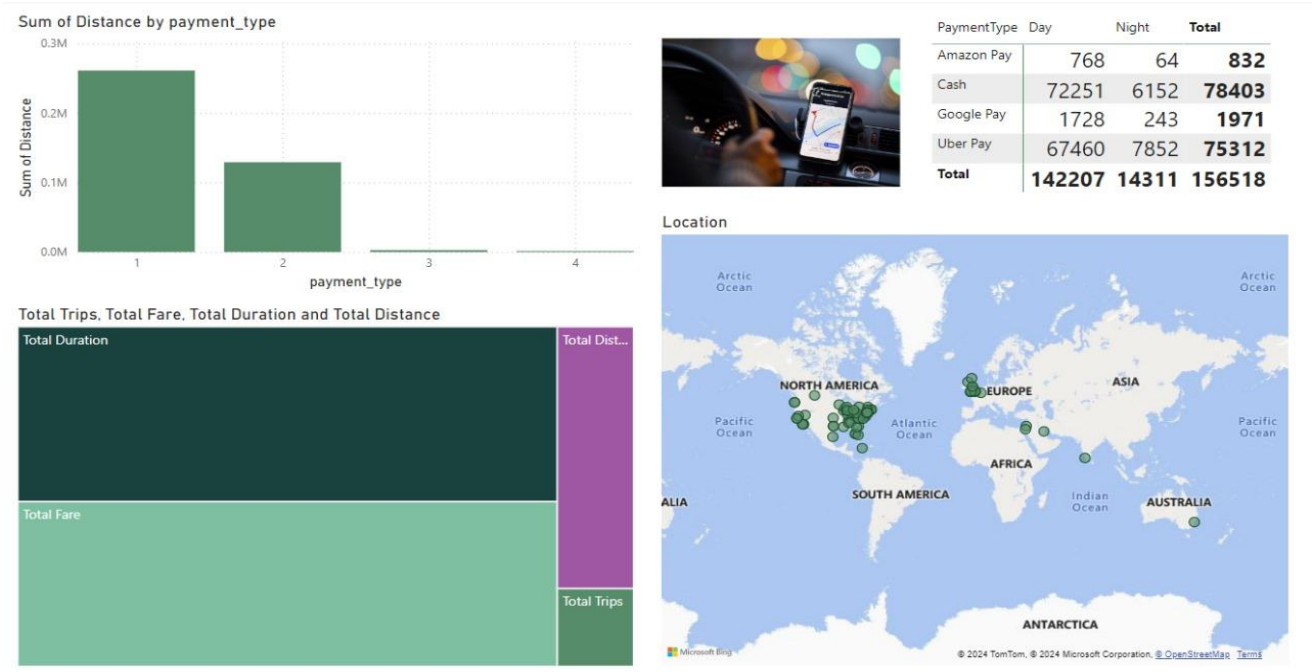


Date	Year	LocationID	Location	Sum of Distance	Total Trips	Total Fare	Total Duration	Total Distance	Shift	Pickup Date	Pickup Time
05-06-2024 00:00:00	2024	3	Allerton/Pelham Gardens	16.78	1.0	$52.3	37.7	16.8	Day	05 June 2024	16:38:02
05-06-2024 00:00:00	2024	3	Allerton/Pelham Gardens	9.10	1.0	$33.3	22.5	9.1	Night	05 June 2024	22:18:32
11-06-2024 00:00:00	2024	3	Allerton/Pelham Gardens	1.62	1.0	$7.0	6.5	1.6	Day	11 June 2024	08:07:17
15-06-2024 00:00:00	2024	3	Allerton/Pelham Gardens	12.30	1.0	$36.0	25.8	12.3	Day	15 June 2024	21:28:36
22-06-2024 00:00:00	2024	3	Allerton/Pelham Gardens	14.35	1.0	$43.3	25.5	14.4	Day	22 June 2024	19:29:36
25-06-2024 00:00:00	2024	3	Allerton/Pelham Gardens	8.30	1.0	$25.5	23.8	8.3	Day	25 June 2024	08:49:36
26-06-2024 00:00:00	2024	3	Allerton/Pelham Gardens	11.56	1.0	$40.5	55.1	11.6	Day	26 June 2024	15:36:08
01-06-2024 00:00:00	2024	4	Alphabet City	2.21	1.0	$9.5	7.4	2.2	Day	01 June 2024	10:29:24
01-06-2024 00:00:00	2024	4	Alphabet City	2.35	1.0	$17.1	18.6	2.4	Day	01 June 2024	16:53:12
01-06-2024 00:00:00	2024	4	Alphabet City	4.20	1.0	$14.5	12.4	4.2	Day	01 June 2024	17:59:58
01-06-2024 00:00:00	2024	4	Alphabet City	1.95	1.0	$12.8	13.4	2.0	Day	01 June 2024	21:26:29
01-06-2024 00:00:00	2024	4	Alphabet City	3.18	1.0	$16.0	17.6	3.2	Night	01 June 2024	22:28:43
01-06-2024 00:00:00	2024	4	Alphabet City	2.62	1.0	$13.4	8.8	2.6	Night	01 June 2024	22:31:07
Total				**3,93,885.13**	**116924.0**	**$17,53,253.3**	**1862824.9**	**393885.1**			

This dashboard displays trip duration by payment type, with one payment method showing a much higher total duration. The total surge fees amount to $231.44K, and a chart shows that total trip duration (1.86M) far exceeds the distance (0.39M), highlighting the time spent on trips compared to miles traveled. The data table

below provides detailed trip records by date, location, fare, duration, distance, and shift, offering a comprehensive view of individual trips and their attributes.

Report Image 4:



This dashboard analyzes trip data by payment type, with a bar chart showing the total distance for each method. Cash and Uber Pay are the most popular options, dominating both day and night shifts in the table. A treemap illustrates total trips, fare, duration, and distance, highlighting total duration as the largest metric. Additionally, a world map displays the geographic distribution of trips, mainly concentrated in North America and Europe.

Conclusion:

The Uber trip analysis dashboard, provides critical insights into trip distribution, revenue generation, and customer preferences. The dashboard allows stakeholders to identify peak trip times, top-performing locations, and the most popular payment methods, thereby enabling data-driven decision-making. By utilising historical data, Uber can forecast future demand trends.