**Deliverable 2**

**1. Dataset:**

This dataset includes trip records from all trips completed in yellow taxis in NYC in January. Records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts.

Source: <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

**2. Tool Used for Data Analysis:** Amazon QuickSight.

**3. Pre-Processing of Data:**

1. The data is cleaned to remove null values.
2. Dataset is prepared with the Yellow Taxi data using CSV file, A calculated field to get WeekDay for a date is added.

A picture containing text

Description automatically generated

**4. Data Exploration:**

**Analysis 1:**

Chart, bar chart

Description automatically generated

This vertical bar chart shows the correlation between total fare on each day of the week. It is observed that maximum collection happens on Friday ($2.8 Million), and minimum happens on Sunday with $1.95 Million.

**Analysis 2:**

Graphical user interface

Description automatically generated

This Line Chart shows the total fare amount on each week day grouped by the passenger count. It can be seen that passenger 1 contribute more to the total fare amount. Maximum fare is accumulated on Friday by passenger travelling alone.

**Analysis 3:**

Chart, bar chart

Description automatically generated

This Clustered Bar Combo Chart gives information about tip amounts given by passengers traveling alone, traveling in a group of 2, 3, 4, 5, 6, 7, 8. The chart also groups the tip amount by the vendor and shows which group of passengers travels the longest distance.

From the analysis, it is observed that:

1. The maximum tip amount is given by passengers who are traveling in groups of 7.
2. Passengers in groups of 7 travel the longest distance.
3. Passengers in group of 7

**Analysis 4:**

Chart, waterfall chart

Description automatically generated