

# Deliverable 2

## Group-7

Akshdeep Rajawat  
Prasenjeet Paul

Rishabh Mehta  
Rajat Maheshwari

Dhvani Patel

### 1. Dataset:

The 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.

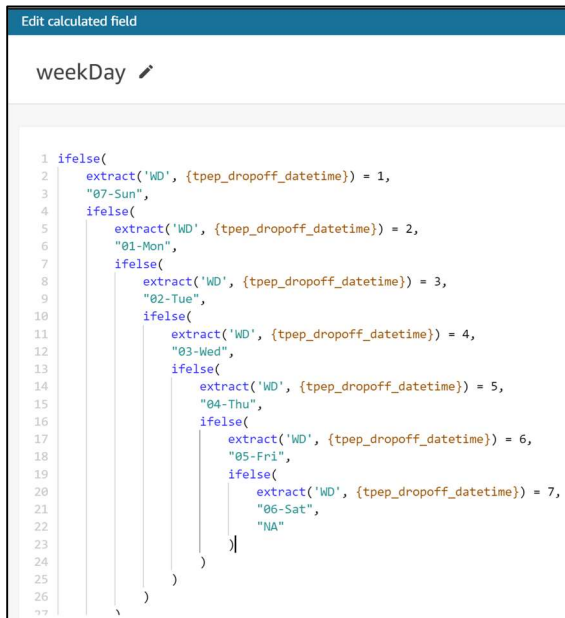
Source: <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

### 2. Tool Used for Data Analysis:

Amazon QuickSight.

### 3. Pre-Processing of Data:

1. The data is cleaned to remove null values.
2. In the pre-processing step, a calculated field is added to the dataset named `weekDay` which calculates the week of the trip date.



```
1  ifelse(
2    extract('WD', {tpep_dropoff_datetime}) = 1,
3    "07-Sun",
4    ifelse(
5      extract('WD', {tpep_dropoff_datetime}) = 2,
6      "01-Mon",
7      ifelse(
8        extract('WD', {tpep_dropoff_datetime}) = 3,
9        "02-Tue",
10       ifelse(
11         extract('WD', {tpep_dropoff_datetime}) = 4,
12         "03-Wed",
13         ifelse(
14           extract('WD', {tpep_dropoff_datetime}) = 5,
15           "04-Thu",
16           ifelse(
17             extract('WD', {tpep_dropoff_datetime}) = 6,
18             "05-Fri",
19             ifelse(
20               extract('WD', {tpep_dropoff_datetime}) = 7,
21               "06-Sat",
22               "NA"
23             )
24           )
25         )
26       )
27     )
28   )
```

#### 4. Data Exploration:

##### Analysis 1:

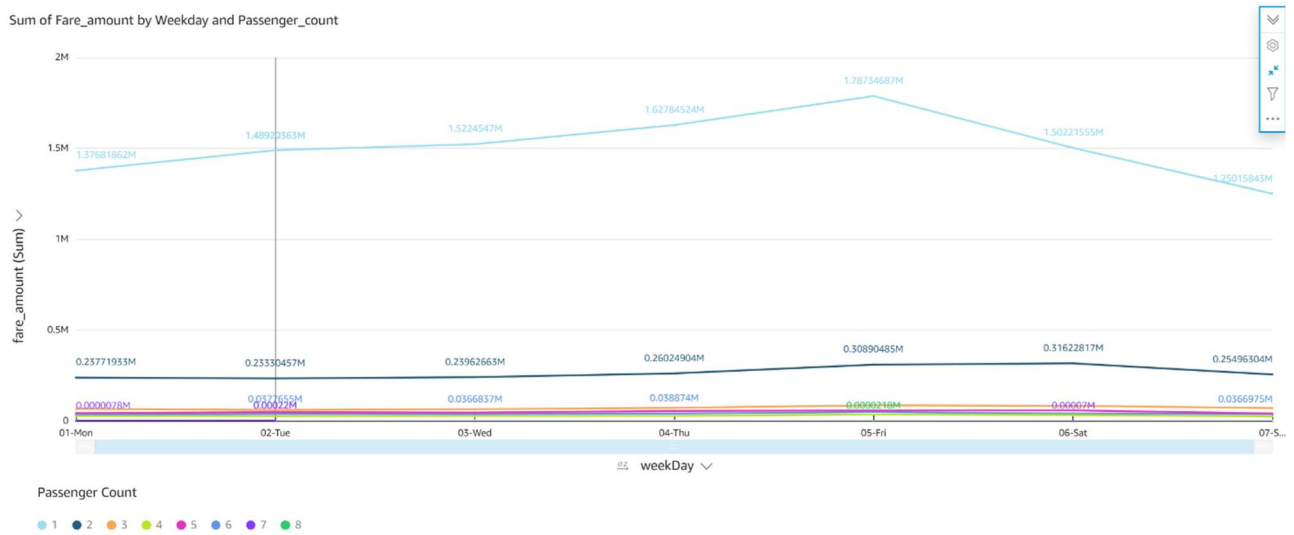


This vertical bar chart shows the correlation between total fare on each day of the week. It is observed that:

1. Maximum collection happened on Friday (\$2.8 Million).
2. Minimum happened on Sunday with \$1.95 Million.

## Analysis 2:

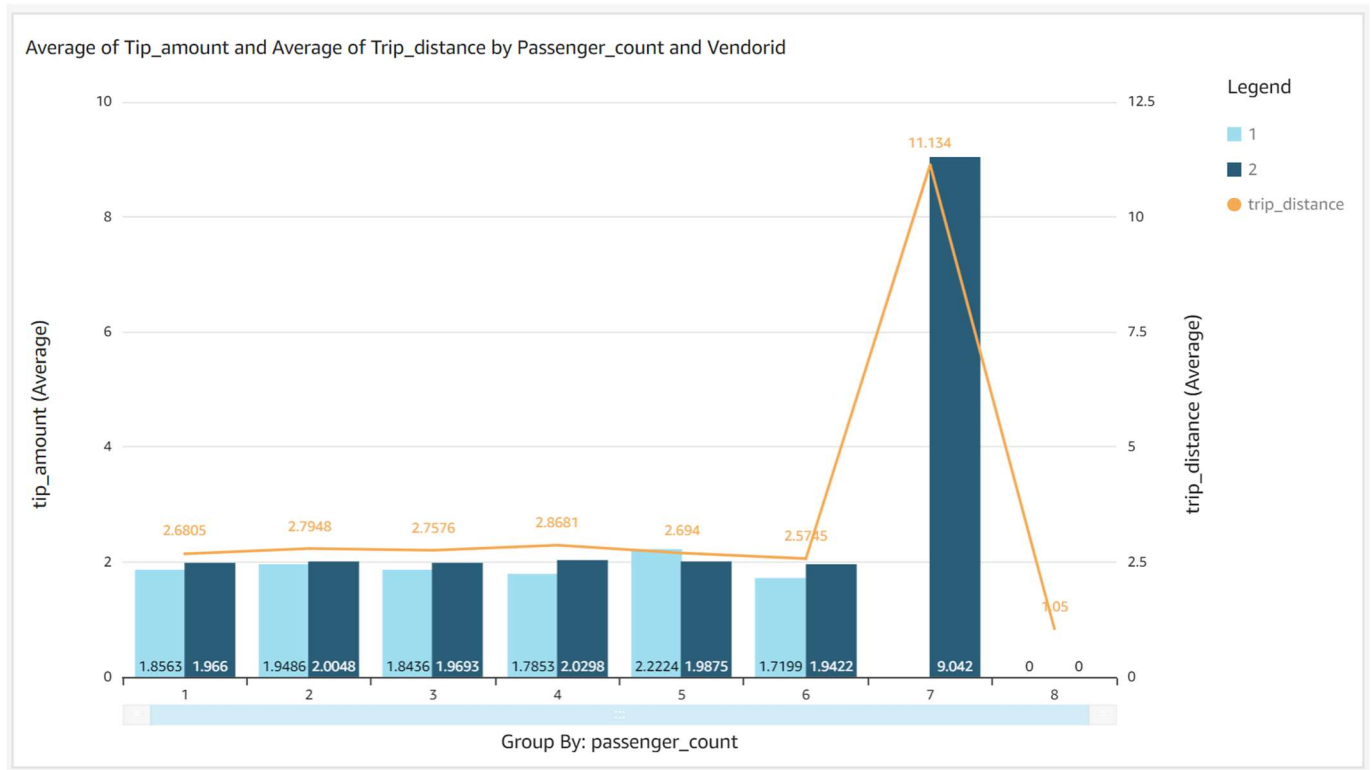
Sum of Fare\_amount by Weekday and Passenger\_count



This Line Chart shows the total fare amount on each weekday grouped by the passenger count.

1. Passengers traveling alone contribute more to the total fare amount, and Friday is seen to have the highest collection.
2. Passengers traveling in group of 2 contribute the second highest to the total fare amount.

### Analysis 3:

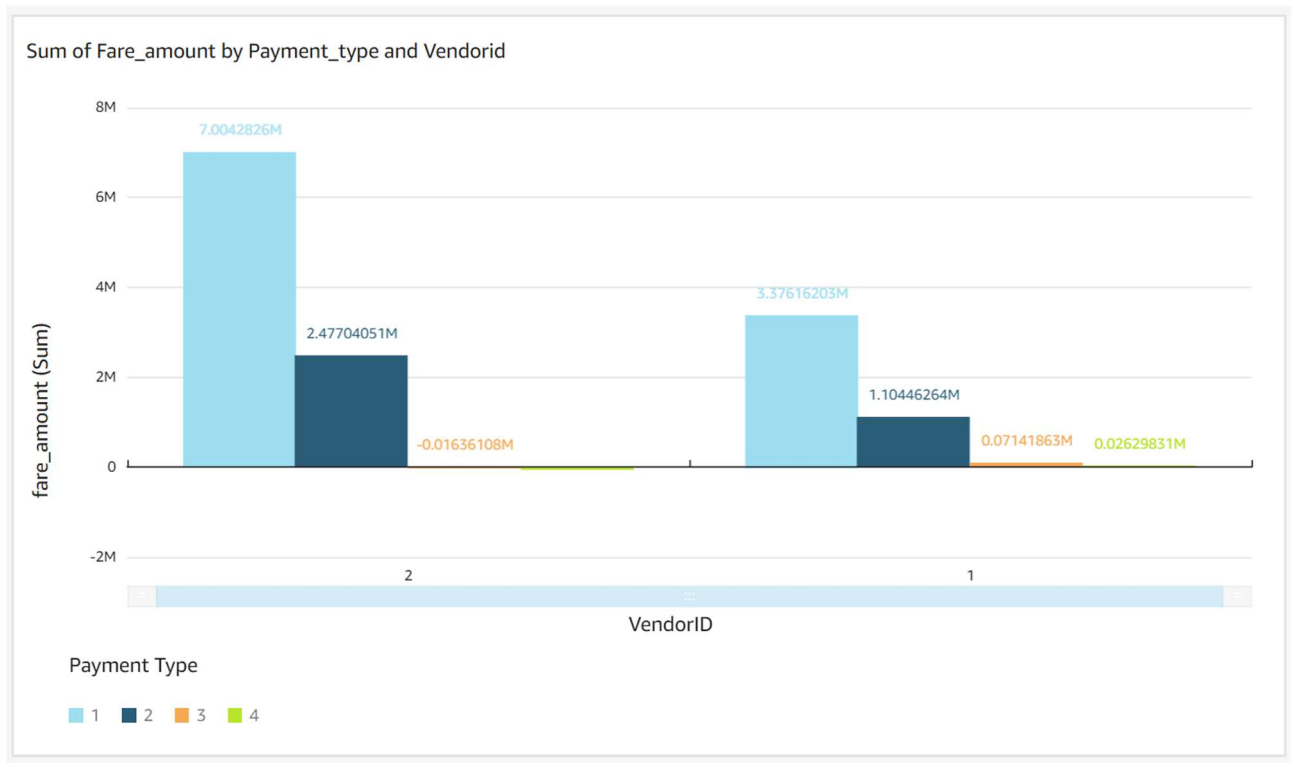


This Clustered Bar Combo Chart gives information about average tip amounts given by passengers traveling alone/traveling in a group of 2, 3, 4, 5, 6, 7, 8. The chart also groups the average tip amount received by different vendors. The line chart represents the average trip distance buy these vendors.

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. Only vendor 2 has passengers who travel in group of 7.

#### Analysis 4:



This vertical bar chart shows the total fare amount by vendor and payment type.

1. In both the vendors, payment type 1 has highest share among the 4 payment types.
2. Vendor 2 collects more fare amount than vendor 1.