NYC TAXI RIDES

ANALYTIC SOLUTION REPORT

RICHIK
MUKHERJEE

Table of Contents

NYC TAXI RIDES DATASET

The data source chosen for the project is NYC Taxi Rides, the dataset comes from the NYC Taxi and Limousine Commission (TLC). The reason for the selection of this dataset is that it provides a rich source of information into the urban mobility, taxi usage and revenue patterns. This dataset contains the detailed information on the taxi rides including the pick-up and drop off locations, fare amounts, tips per ride and several other ride details. The dataset aligns the project goal of optimizing the revenue and driver allocation by collectively analyzing the trip patterns. This open-source dataset entails a fair understanding of the geographical coverage of in demand trip areas, provides the temporal coverage of the data where there are millions of records per month, it is also considered suitable for both short- and long-term trend analysis.

PURPOSE/GOAL

The purpose of selecting the NYC Taxi Rides dataset is to optimize the revenue and driver allocation by thorough analysis of the trip patterns.

KEY QUESTIONS

The key questions are listed below:

- What are the major pick-up locations?
- How does the taxi trip patterns, and demand of taxi rides fluctuate across different times in each day, week and seasons of the city?
- What is the maximum number of passengers served?

METRICS and KPIs

- The frequency of the pick-ups per location
- The maximum number of passengers being served at a particular phase of time
- Number of pick-ups per hour/minute
- The approximate total revenue earned

DATA MODEL

Data Contents

Field	Particulars
Vendor ID	Identifies the type of taxi service vendor (1 for Creative Mobile Technologies, 2 for VeriFone Inc.)
Pick-up DateTime	The exact date and time the passenger was picked up
Dropoff DateTime	The exact data and time the passenger was dropped off
Passenger Count	The total number of passengers taking a ride in the taxi
Trip Distance	Total distance and duration of the trip in miles
Pickup Location ID	A unique code signifying the pickup location/zone
Dropoff Location ID	A unique code signifying the drop-off location/zone
Rate Code ID	Determines the rate type (standard, JFK, group ride or Newark)
Fare Amount	The basic fare being charged per trip, excluding surcharges and extras
Tip Amount	The amount of tip being paid
Tolls Amount	Amount paid for tolls during the trip (if applicable)
Surcharge	Surcharge added to the trips for funding the infrastructural improvements
Total Amount	A total amount being added to the trips that includes fares, tips, tolls, taxes and surcharges
Payment Type	The type of payment used by passengers (credit card, cash, etc.)
Trip Duration	The total difference of the distance between pickups and drop-offs.

Facts Table

vendorid	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	ratecodeid	store_and_fwd_flag	pulocationid	dolocationid	payment_type	fare_amount	extra	tip_amount
2	2023-01-01 12:59:24 AM	2023-01-01 1:14:26 AM	1	0	1	N	141	193	2	13.5	1	0
2	2023-01-01 12:58:00 AM	2023-01-01 1:12:24 AM	1	0.64	1	N	186	164	2	12.8	1	0
2	2023-01-01 12:57:50 AM	2023-01-01 1:02:03 AM	1	0.55	1	N	246	68	2	5.8	1	0
2	2023-01-01 12:56:54 AM	2023-01-01 1:01:52 AM	1	0.09	1	N	249	249	2	5.8	1	0
2	2023-01-01 12:55:17 AM	2023-01-01 1:03:27 AM	1	2.14	1	N	264	264	2	11.4	1	0
2	2023-01-01 12:55:03 AM	2023-01-01 1:00:09 AM	1	0.69	1	N	142	239	2	7.2	1	0
2	2023-01-01 12:54:19 AM	2023-01-01 1:10:21 AM	1	3.29	1	N	13	246	2	17	1	0
2	2023-01-01 12:53:11 AM	2023-01-01 12:57:25 AM	1	0.35	1	N	239	239	2	5.8	1	0
2	2023-01-01 12:52:53 AM	2023-01-01 1:07:25 AM	1	2.61	1	N	141	107	2	15.6	1	0
2	2023-01-01 12:52:42 AM	2023-01-01 12:59:15 AM	1	0.72	1	N	68	68	2	7.2	1	0
2	2023-01-01 12:51:17 AM	2023-01-01 12:55:08 AM	1	1.49	1	N	263	74	2	7.9	1	0
2	2023-01-01 12:50:34 AM	2023-01-01 12:58:45 AM	1	1.52	1	N	107	162	2	10	1	0
2	2023-01-01 12:49:56 AM	2023-01-01 1:30:46 AM	1	1.94	1	N	237	141	2	18.4	1	0
2	2023-01-01 12:48:53 AM	2023-01-01 12:59:50 AM	1	2.58	1	N	229	75	2	13.5	1	0
2	2023-01-01 12:48:30 AM	2023-01-01 1:01:26 AM	1	2.1	1	N	43	48	2	14.2	1	0
2	2023-01-01 12:48:19 AM	2023-01-01 1:10:17 AM	1	0.79	1	N	161	164	2	17.7	1	0
2	2023-01-01 12:47:57 AM	2023-01-01 1:12:21 AM	1	7.41	1	N	163	87	2	32.4	1	0
2	2023-01-01 12:47:54 AM	2023-01-01 12:53:04 AM	1	0.63	1	N	13	261	2	7.2	1	0
2	2023-01-01 12:47:34 AM	2023-01-01 12:58:22 AM	1	3.39	1	N	233	74	2	15.6	1	0
2	2023-01-01 12:47:29 AM	2023-01-01 12:53:05 AM	1	1.2	1	N	229	237	2	7.9	1	0
2	2023-01-01 12:47:17 AM	2023-01-01 12:55:30 AM	1	1.24	1	N	90	158	2	9.3	1	0
2	2023-01-01 12:47:09 AM	2023-01-01 1:00:08 AM	1	1.42	1	N	140	140	2	9.3	1	0
2	2023-01-01 12:46:57 AM	2023-01-01 1:03:48 AM	1	2.51	1	N	234	148	2	15.6	1	0
2	2023-01-01 12:46:39 AM	2023-01-01 1:06:46 AM	1	7.19	1	N	43	87	2	31.7	1	0
2	2023-01-01 12:44:52 AM	2023-01-01 12:56:52 AM	1	3.16	1	N	79	237	2	15.6	1	0
2	2023-01-01 12:44:16 AM	2023-01-01 12:49:28 AM	1	0.67	1	N	239	238	2	7.2	1	0
2	2023-01-01 12:43:48 AM	2023-01-01 12:58:12 AM	1	4.68	1	N	148	112	2	21.9	1	0

Data Modelling
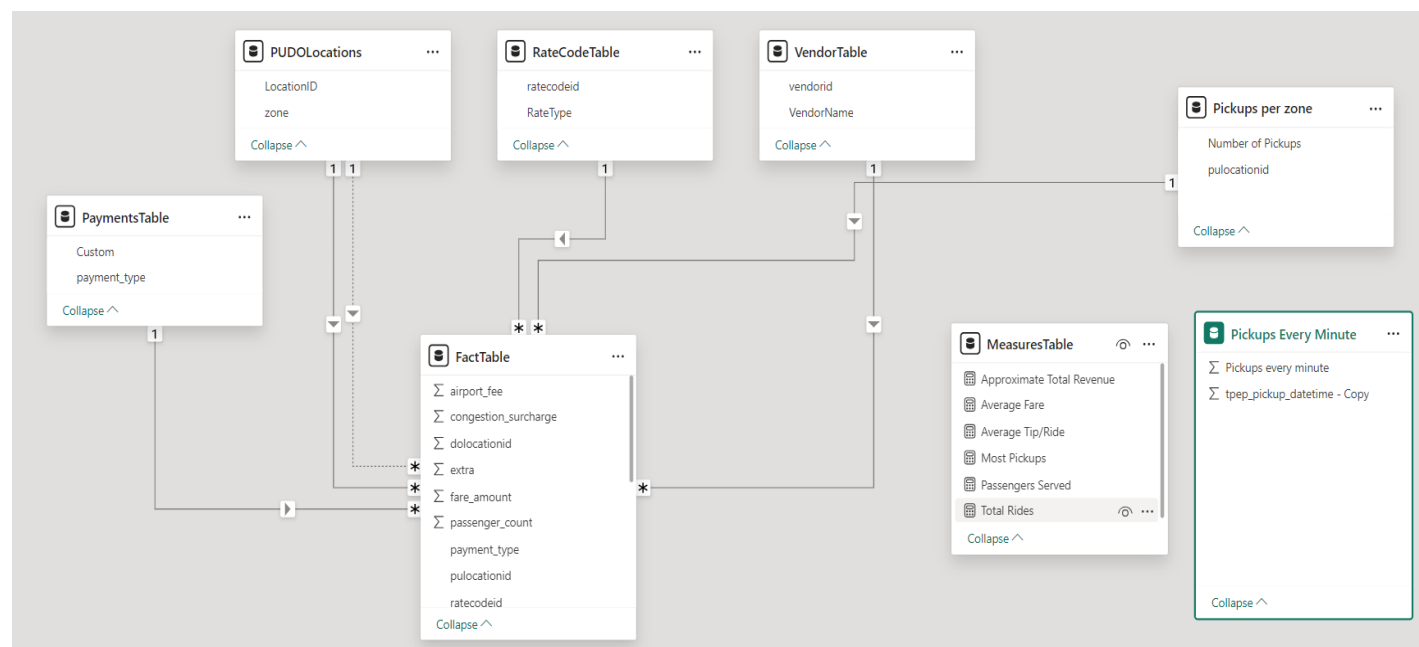
RESULTS

Below is the screenshot of the Power BI dashboard showing the analysis of the key factors to help evaluate the goal of the report. The major pickups are mentioned by a card showing 62 pickup frequencies in East Village, moreover, the total number of passengers being served is also highlighted which stands at a strong number of 1517. The dataset is based on a historical form and this data pertains to the year 2023. Moving forward, the other purpose was also to

analyse and revenue and give a boost to it, so the average total revenue in every hour, this will provide with the approximate valuation of revenue for the day.

Moving to the graphs, pickups every minute starting from 12am to 1.30am is presented below as a line graph, through this the trend patterns of the frequent pickup can be analysed, this shall help allocate more taxi drivers in areas that have higher demand, the peak time can be seen in 12.30am with around 28-30 passengers asking for quick rides. Moving forward, the next graph shows the total amount of fare charged based on total passengers in an hour.

The sum of total fare is at its highest with approximate $16k for every passenger in an hour. The pie chart organizes the data for payment type used by passengers to conclude their rides, the highest mode of payment is by credit cards with an approximate valuation of 77.46%. The latter shows cash and dispute settlement with minorities under 25%. The final representation includes the map of USA showing the distribution of the pick-ups and drop offs marking the demand and transport patterns, this mapping helps to divert taxi drivers to needed locations which have heavy demand to increase revenue.

CONCLUSION

The NYC Taxi Rides dataset is quite a lucrative and important resource for creating a profound understanding of urban mobility patterns, transportation systems being used, socioeconomic trend patterns and other aspects in one of the busiest cities in the world. The detailed level of trip data ranging from hundreds of taxi rides, long term operations provide a strong opportunity to policy makers, researchers, analysts, and several businesses to delve into the insights of traffic flow, behavior of passengers and the overall infrastructure of the transportation industry. The solution towards maximising the driver allocation to needed areas of the city and increasing revenue can therefore be achieved by analyzing the traffic patterns, channel alternative routes through fastened GPS tracking and increase the base salaries of the tax drivers to ensure loyalty and motivation for longer shifts in the future.