INTRODUCTION

In the project, we have the dataset:

NYC Taxi Trips:
Provides travel records for yellow taxis, green taxis, and for-hire automobiles in New York City in January 2015.
Files:
fhv_tripdata_2015-01,
green_tripdata_2015-01,
yellow_tripdata_2015-01

NYC Weather Conditions:
Hourly weather data for January 2015 are available, as well as daily data for the full year.
Files:
New York City Weather Information Hourly 2015-01-01 to 2015-01-30.xls,
New York City Weather Information Daily 2015-01-01 to 2016-01-01.xls

US Flights Information:
It tracks the on-time performance of domestic flights, including delays, cancellations, and diversions.
Files:
airlines.csv,
airports.csv,
flights.csv

We developed a PowerBI report that includes 12 dashboards:

1. Green Taxi Travel Information (January 2015)
2. Yellow Taxi Travel Information (January 2015)
3. For-Hire Vehicles Travel Information (January 2015)
4. New York City Taxi Travel Information (January 2015)
5. New York City Weather Information (January 2015)
6. New York City Flights Information (2015)
7. Dataset Anomaly Centre
8. Summary Statistics Centre
9. Comparison Centre
10. Analytics Dashboard
11. Hypothesis Centre
12. My Story

TRANSFORMATION:

After loading the dataset in PowerBI, we added the new column and measures to the data table to make our project report more convenient to visualize and analyze.

green_tripdata_2015_01:

Calculated Column	DAX
DO Hour from Airport	DO Hour From Airport = IF([DOLocationID] = 1, HOUR([DO Time]),BLANK())
PU DO Location ID	PU DO Location ID = [PULocationID] & " - " & [DOLocationID]
PU Hour	PU Hour = HOUR([PU Time])
PU Hour To Airport	PU Hour To Airport = IF([PULocationID] = 1, HOUR([PU Time]),BLANK())
To Airport	To Airport = IF([DOLocationID] = 1,1,0)
Trip Duration	Trip Duration = DATEDIFF([pickup_datetime],'green_tripdata_2015-01'[dropoff_datetime],MINUTE)
Type	Type = "Green"
Weekday	Weekday = IF(WEEKDAY([pickup_datetime]) = 1, 0, IF(WEEKDAY([pickup_datetime]) = 7,0,1))

Yellow_tripdata_2015_01: Renamed as "Yellow Taxi Tripdata Jan 2015"

Calculated Column	DAX
PU Hour	PU Hour = HOUR([PU Time])
Trip Duration	Trip Duration = DATEDIFF([pickup_datetime],'Yellow Taxi Tripdata Jan 2015'[dropoff_datetime],MINUTE)
Type	Type = "Yellow"

Measures	DAX
Busiest Zone on 1st Sat	Busiest Zone on 1st Sat = CALCULATE(Count('Yellow Taxi Tripdata Jan 2015'[PULocationID]), 'Date Table'[IS 1st Saturday] = "Yes")
Yellow Taxi Avg Fare	Yellow Taxi Avg Fare = CALCULATE(AVERAGE('Yellow Taxi Tripdata Jan 2015'[fare_amount]))
Yellow Taxi Count by PU LocationID	Yellow Taxi Count by PU LocationID = COUNT('Yellow Taxi Tripdata Jan 2015'[PULocationID])

fhv_tripdata_2015-01: Renamed as "FHV Tripdata 2015 Jan"

The dataset is incomplete with the comparison of the other two trip datasets. DropOff Location Id , Fare Amount , Trip Duration are not available on the dataset.

Calculated Column	DAX
MissingFlag	MissingFlag = IF(ISBLANK('FHV Tripdata 2015 Jan'[DOlocationID]), "Missing", "Not Missing")
PU Date	PU Date = 'FHV Tripdata 2015 Jan'[pickup_datetime].[Date]
PU Hour	PU Hour = HOUR([pickup_datetime])
PU Location Count	PU Location Count = COUNT([PULocationID])
PU Time	PU Time = 'FHV Tripdata 2015 Jan'[pickup_datetime].[Date]
Type	Type = "FVH"
Weather Conditions	Weather Conditions = LOOKUPVALUE('NYC Weather 2015 Daily'[conditions],'NYC Weather 2015 Daily'[Date],'FHV Tripdata 2015 Jan'[PU Date])

Measures	DAX
MissingCount	MissingCount = COUNTAX(FILTER('FHV Tripdata 2015 Jan', ISBLANK('FHV Tripdata 2015 Jan'[DOlocationID])), 'FHV Tripdata 2015 Jan'[dropOff_datetime])

Airlines:

Measures	DAX
Avg Arr Delays	Avg Arr Delays = AVERAGEX('flights',[ARRIVAL_DELAY])
Avg Dep Delays	Avg Dep Delays = AVERAGEX('flights',[DEPARTURE_DELAY])

Airport:
We divided the airport table into two tables: Airport (D)—Departures and Airport (O)—Origin.

Airport(O):

New Column	DAX
Location	Location = [CITY] & " " & [STATE] & " " & [COUNTRY]

Date Table: We created a new table to deal with the time-period for the analysis and visualization of the dataset.

Column	DAX
Day of Week	Day of Week = Switch(WEEKDAY([Date]), 1,"Sun", 2,"Mon", 3,"Tue", 4,"Wed", 5,"Thu", 6,"Fri", 7,"Sat")
Day of Week No	Day of Week No = WEEKDAY([Date])
IS 1st Saturday	IS 1st Saturday = IF('Date Table'[Week No] = 1, IF([Day of Week No] = 7, "Yes", "No"),"No")
Month	Month = FORMAT([Date],"mmmm")

Month No	Month No = MONTH([Date])
Quarter	Quarter = "Q" & QUARTER([Date])
Weather Conditions	Weather Conditions = LOOKUPVALUE('NYC Weather 2015 Daily'[conditions],'NYC Weather 2015 Daily'[datetime],'Date Table'[Date])
Week	Week = "Week " & [Week No]
Week No	Week No = WEEKNUM([Date])

Weekday or Weekend	Weekday or Weekend = Switch(
	WEEKDAY([Date]),
	    1,"Weekend",
	    2,"Weekday",
	    3,"Weekday",
	    4,"Weekday",
	    5,"Weekday",
	    6,"Weekday",
	    7,"Weekend"
	)

Year	Year = YEAR([Date])

In the table, Date Hierarchy is created as:
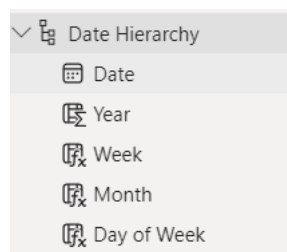
Fig: Representation of Hierarchy.

Flights:

| Calculated Column	DAX
Day of Week Name	Day of Week Name = FORMAT([Flight Date],"ddd")
Delay Above Avg	Delay Above Avg = IF([DEPARTURE_DELAY] > 'airlines'[Avg Dep Delays],1,0)
Flight Date	Flight Date = DATE([YEAR],[MONTH],[DAY])
Flight Delay	Flights Delayed = IF('Flights'[DEPARTURE_DELAY] > 0,1,0)
Weather Conditions	Weather Conditions = LOOKUPVALUE('NYC Weather 2015 Daily'[conditions],'NYC Weather 2015 Daily'[Date],'flights'[Flight Date])

Measures	DAX
Arr Delay	Arr Delay = DIVIDE('Flights'[Arr Delays],CALCULATE(COUNT('Flights'[ARRIVAL_DELAY]),ALL('flights'),Flights[ARRIVAL_DELAY] > 0))

Arr Delays	Arr Delays =
	AVERAGEX('Flights',
	CALCULATE(
	    SUM('Flights'[ARRIVAL_DELAY]),
	    'Flights'[ARRIVAL_DELAY] > 0
	))

Dep Delay	Dep Delay = DIVIDE('flights'[Dep Delays],CALCULATE(COUNT('flights'[DEPARTURE_DELAY]),ALL('flights'),flights[DEPARTURE_DELAY] > 0))

Dep Delays	Dep Delays =
	AVERAGEX('Flights',
	CALCULATE(
	    SUM('flights'[DEPARTURE_DELAY]),
	    'Flights'[DEPARTURE_DELAY] > 0
	))

New York City Weather Information Hourly 2015-01-01 to 2015-01-30: Renamed as "NYC Weather 2015 Jan Hourly"

Calculated Column	DAX
Temp Variance	Temp Variance = [temp] – [feelslike]

Data Modelling:
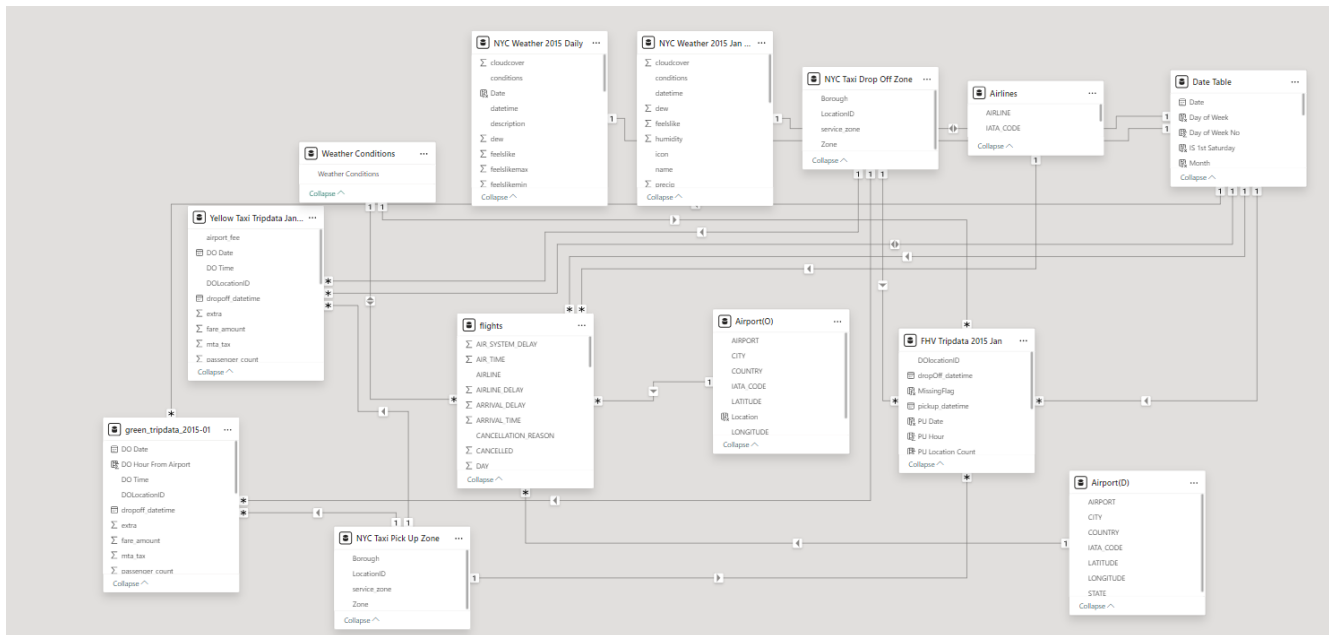
Fig: Representation of Relationships in between the dataset tables

From: table (column)	Relationship ↓	To: table (column)	Status
NYC Weather 2015 Daily (date...	1 ◀▶ 1	📅 Date Table (Date)	Active
NYC Weather 2015 Jan Hourly ...	1 ◀▶ 1	📅 Date Table (Date)	Active
flights (Weather Conditions)	* ◀▶ 1	Weather Conditions (Weather ...	Active
Yellow Taxi Tripdata Jan 2015 (...	* ◀▶ 1	📅 Date Table (Date)	Active
green_tripdata_2015-01 (DOL...	* ◀ 1	NYC Taxi Drop Off Zone (Locat...	Active
Yellow Taxi Tripdata Jan 2015 (...	* ◀ 1	NYC Taxi Drop Off Zone (Locat...	Active
FHV Tripdata 2015 Jan (DOloc...	* ◀ 1	NYC Taxi Drop Off Zone (Locat...	Active
FHV Tripdata 2015 Jan (PUloca...	* ◀ 1	NYC Taxi Pick Up Zone (Locati...	Active
green_tripdata_2015-01 (PULo...	* ◀ 1	NYC Taxi Pick Up Zone (Locati...	Active
Yellow Taxi Tripdata Jan 2015 (...	* ◀ 1	NYC Taxi Pick Up Zone (Locati...	Active
FHV Tripdata 2015 Jan (PU Date)	* ◀ 1	📅 Date Table (Date)	Active

	FHV Tripdata 2015 Jan (Weath...	✱—◀—1		Weather Conditions (Weather ...	Active
	flights (AIRLINE)	✱—◀—1		Airlines (IATA_CODE)	Active
	flights (DESTINATION_AIRPORT)	✱—◀—1		Airport(D) (IATA_CODE)	Active
	flights (Flight Date)	✱—◀—1	📅	Date Table (Date)	Active
	flights (ORIGIN_AIRPORT)	✱—◀—1		Airport(O) (IATA_CODE)	Active
	green_tripdata_2015-01 (PU D...	✱—◀—1	📅	Date Table (Date)	Active

Fig: Representation of Relationships in between the Tables

green_tripdata_2015-01:

- green_tripdata_2015-01(PULocationID) is linked to NYC Taxi Pick Up Zone(LocationID) with Many to one Cardinality and Single Cross-filter direction.
- green_tripdata_2015-01(DOLocationID) is linked to NYC Taxi Drop Off Zone(LocationID) with Many to one Cardinality and Single Cross-filter direction.
- green_tripdata_2015-01(PU Date) is linked to Date Table(Date) with Many to one Cardinality and Single Cross-filter direction.

Yellow Taxi Trip Jan 2015:

- Yellow Taxi Trip Jan 2015 (PULocationID) is linked to NYC Taxi Pick Up Zone(LocationID) with Many to one Cardinality and Single Cross-filter direction.
- Yellow Taxi Trip Jan 2015 (DOLocationID) is linked to NYC Taxi Drop Off Zone(LocationID) with Many to one Cardinality and Single Cross-filter direction.
- Yellow Taxi Trip Jan 2015 (PU Date) is linked to Date Table(Date) with Many to one Cardinality and Single Cross-filter direction.

FHV Tripdata 2015 Jan:

- FHV Tripdata 2015 Jan (PULocationID) is linked to NYC Taxi Pick Up Zone(LocationID) with Many to one Cardinality and Single Cross-filter direction.
- FHV Tripdata 2015 Jan (Weather Conditions) is linked to Weather Conditions (Weather Conditions) with Many to one Cardinality and Single Cross-filter direction.
- FHV Tripdata 2015 Jan (PU Date) is linked to Date Table(Date) with Many to one Cardinality and Single Cross-filter direction.
- FHV Tripdata 2015 Jan (DOLocationID) is linked to NYC Taxi Drop Off Zone(LocationID) with Many to one Cardinality and Single Cross-filter direction.

Flights:

- flights (ORIGIN_AIRPORT) is linked to Airport(O)(IATA_CODE) with Many to one Cardinality and Single Cross-filter direction.
- flights (Weather Conditions) is linked to Weather Conditions (Weather Conditions) with Many

to one Cardinality and Both Cross-filter direction.
- flights (Flight Date) is linked to Date Table (Date) with Many to one Cardinality and Single Cross-filter direction.
- flights(DESTINATION_AIRPORT) is linked to Airport(D) (IATA_CODE) with Many to one Cardinality and Single Cross-filter direction.
- flights(AIRLINE) is linked to Airlines (IATA_CODE) with Many to one Cardinality and Single Cross-filter direction.

NYC Weather 2015 Jan Hourly:
- NYC Weather 2015 Jan Hourly(datetime) is linked to Date Table (Date) with One to one Cardinality and Both Cross-filter direction.

NYC Weather 2015 Jan Daily:
- NYC Weather 2015 Jan Daily(datetime) is linked to Date Table (Date) with One to one Cardinality and Both Cross-filter direction.

REPORTS:

Green Taxi Travel Information (January 2015)

In the Green Taxi Information Page, we build six visuals which includes: card, slicer, clustered bar chart, clustered column chart.
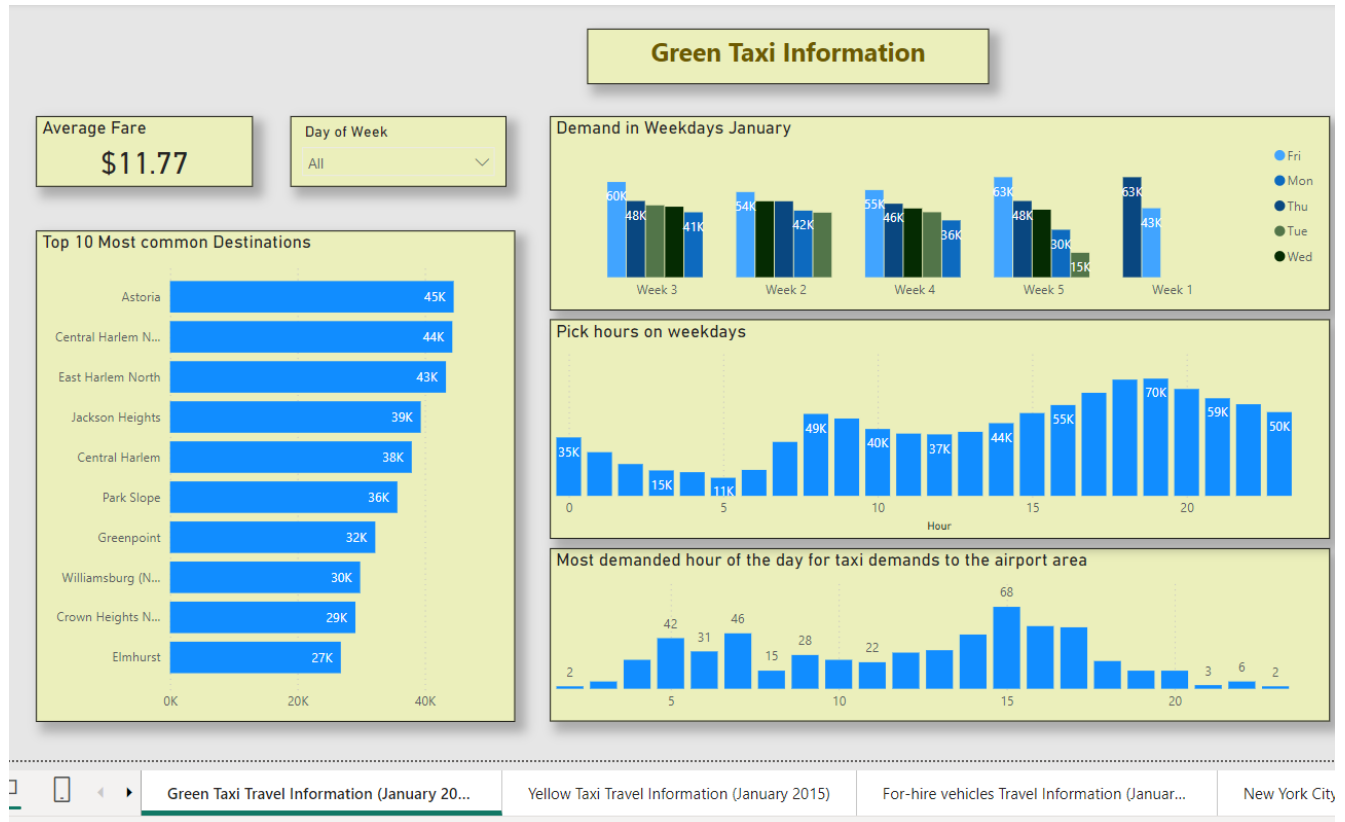

Fig: Green Taxi Travel Information (January 2015)

Reason for selecting those visuals: Summary statistics are provided by a card, a slicer is used for data filtering, and clustered bar and column charts are available for side-by-side comparison of categorical data in Power BI, which widens data analysis and visualization.
- Average Fare is calculated on a card.
- Selection of day of week done using slicer.
- Top 10 Most common Destinations are presented in the clustered bar chart.
- In clustered column chart, Demand in Weekdays January, Pick hours on weekdays, and Most demanded hour of the day for taxi demands to the airport area are presented.

Findings:
- At 44,530, Astoria had the highest number of destination and was 66.10% higher than Elmhurst, which had the lowest number of destinations 26,809.
- Astoria accounted for 12.26% of total number of destinations.
- Across all 10 Zone, number of destinations ranged from 26,809 to 44,530.
- Week 5 in Day of Fri of January made up 6.32% of demand.
- Average Fare is $11.77 for the Green Taxi.

Yellow Taxi Travel Information (January 2015)

In the page, we build six visual which includes: Slicer, Area chart, Heat Map, clustered column chart and clustered bar chart.

Reason for selecting those visuals: A slicer enables interactive filtering, an area chart shows trends over time, a heat map highlights intensity of data, and clustered column and bar charts allow for side-by-side comparison of categorical data, enhancing overall data analysis and visualization.

- Selection of Borough is done using Slicer.
- Highest use of yellow taxis by borough is observed on Area chart.
- Yellow taxis usage in different Zones of NYC is present in Heat Map.
- In Column chart, Top 3 destinations having higher fare and Yellow CAB Avg Fare by PU Hour are presented.
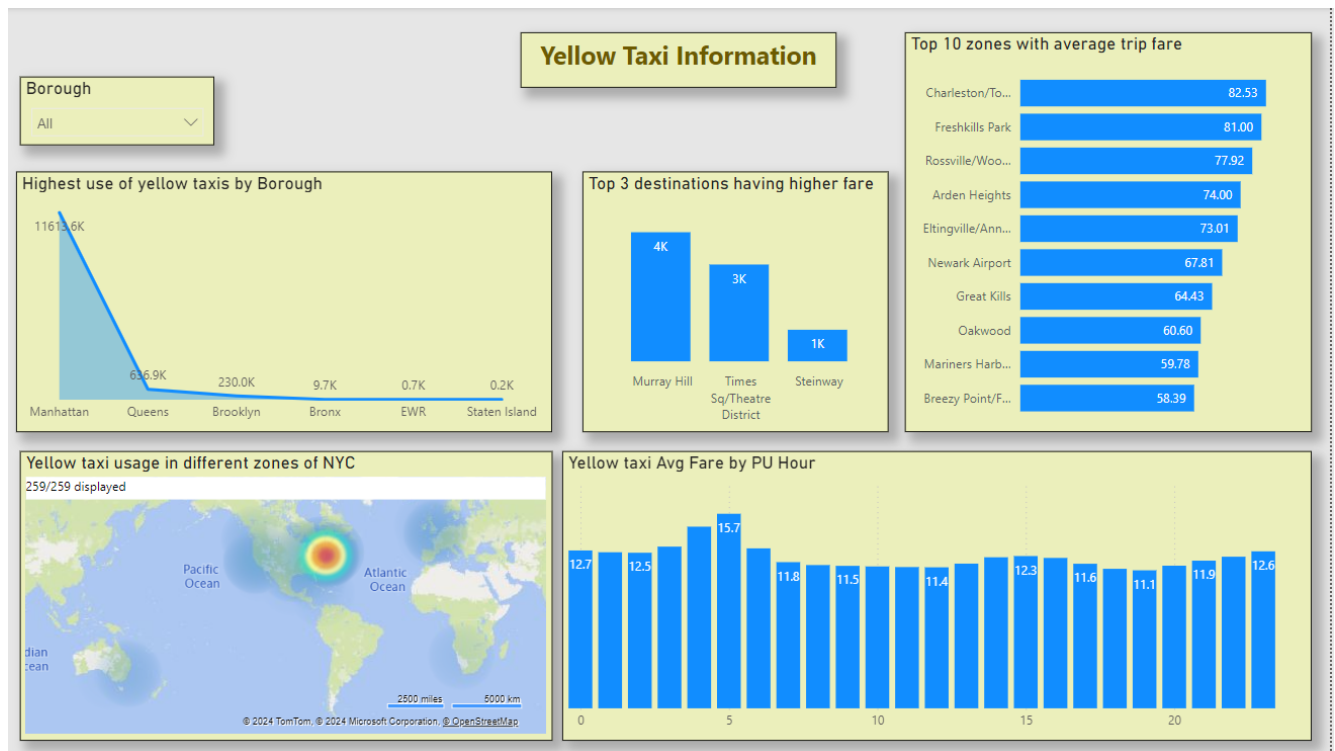- The top 10 zones with average trip fare are calculated on clustered bar chart.

Fig: Yellow Taxi Travel Information (January 2015)

Findings:

- At 11,613,627, Manhattan had the highest number of uses of yellow taxi and was 6,048,664.06% higher than Staten Island, which had the lowest use at 192.
- Manhattan accounted for 92.98% of total use of yellow taxi.
- Murray Hill has the maximum fare at $4,008.
- The average trip fare is 82.53, which is the highest average fare in the Charleston/Tootenville.
- At 5 am morning, the yellow taxi avg fare is $15.7 which is highest average.

For-hire vehicles Travel Information (January 2015)

In the page, we used four visual which includes: card, clustered bar chart and stacked column chart.
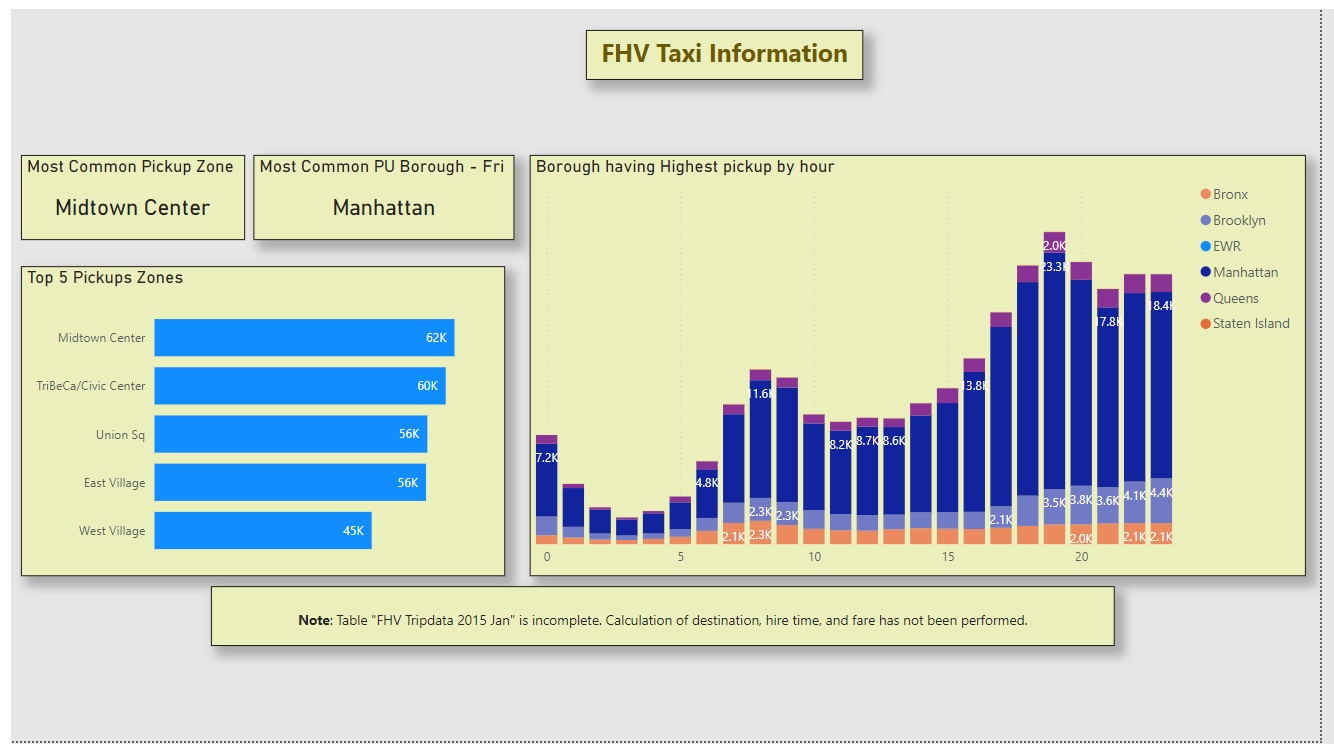
Fig: For-hire vehicles Travel Information

Reason for selecting those visuals: A card highlights key metrics, a clustered bar chart allows side-by-side comparison of categorical data, and a stacked column chart shows the contribution of each category.
- Most Common Pickup Zone and Most Common Pickup Borough – Fri are presented in card.
- The top 5 pickups zone is presented in clustered bar chart
- Borough having Highest pickup by hour is shown using the stacked column chart.

Findings:

- At 62,069, Midtown Center had the highest number of pickups and was 38.13% higher than West Village, which had the lowest number of pickups at 44,935.
- Midtown Center accounted for 22.18% of total number of pickups.
- Across all 5 Zone, number of pickups ranged from 44,935 to 62,069.
- Manhattan made up most common pickup borough on Friday.
- Table "FHV Tripdata 2015 Jan" is incomplete. Calculation of destination, hire duration and fare information has not been performed.

New York City Taxi Travel Information (January 2015)

In the page, we build five visuals which includes: slicer, stacked bar chart, stacked column chart.
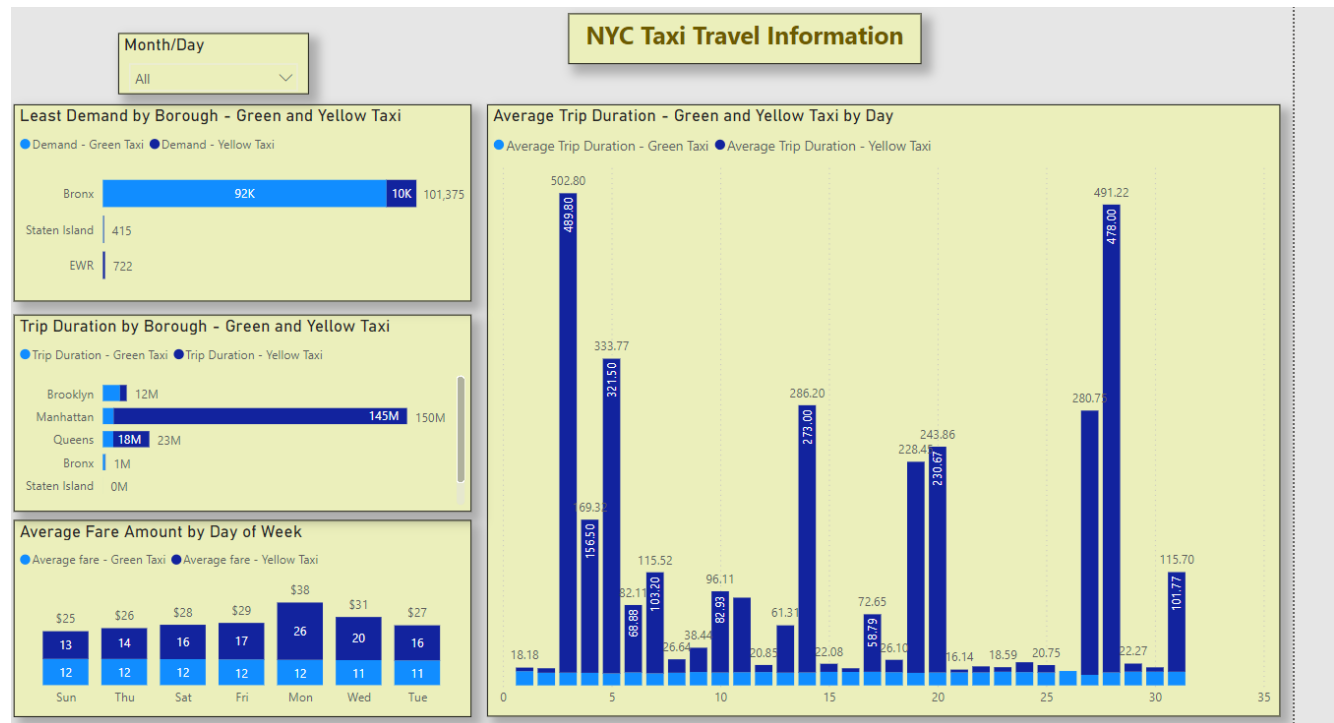

Fig: NYC Taxi Travel Information

Reason for selecting those visuals: A slicer filters data interactively, while stacked bar and column charts display the composition and comparison of categories which helps to analyze both individual contributions and overall trends.

- Month/day filter is done by using slicer.
- Least Demand by borough – Green Taxi and Yellow Taxi, Trip Duration by Borough – Green and Yellow Taxi and Average Fare Amount by Day of Week are display on Stack bar chart.
- In the stacked column chart, Average Trip Duration – Green and Yellow Taxi by Day is displayed.

Findings:
- The lowest demand for taxis is in Staten Island at 415.
- 150 million total trip duration is in Manhattan which is the highest among the boroughs.
- At $38, Monday has the highest average fare and at $25, Sunday has lowest average fare.
- Average Trip duration is highest in Day 3 with 502.80 and lowest is 14.74 in day 26.

New York City Weather Information (January 2015)

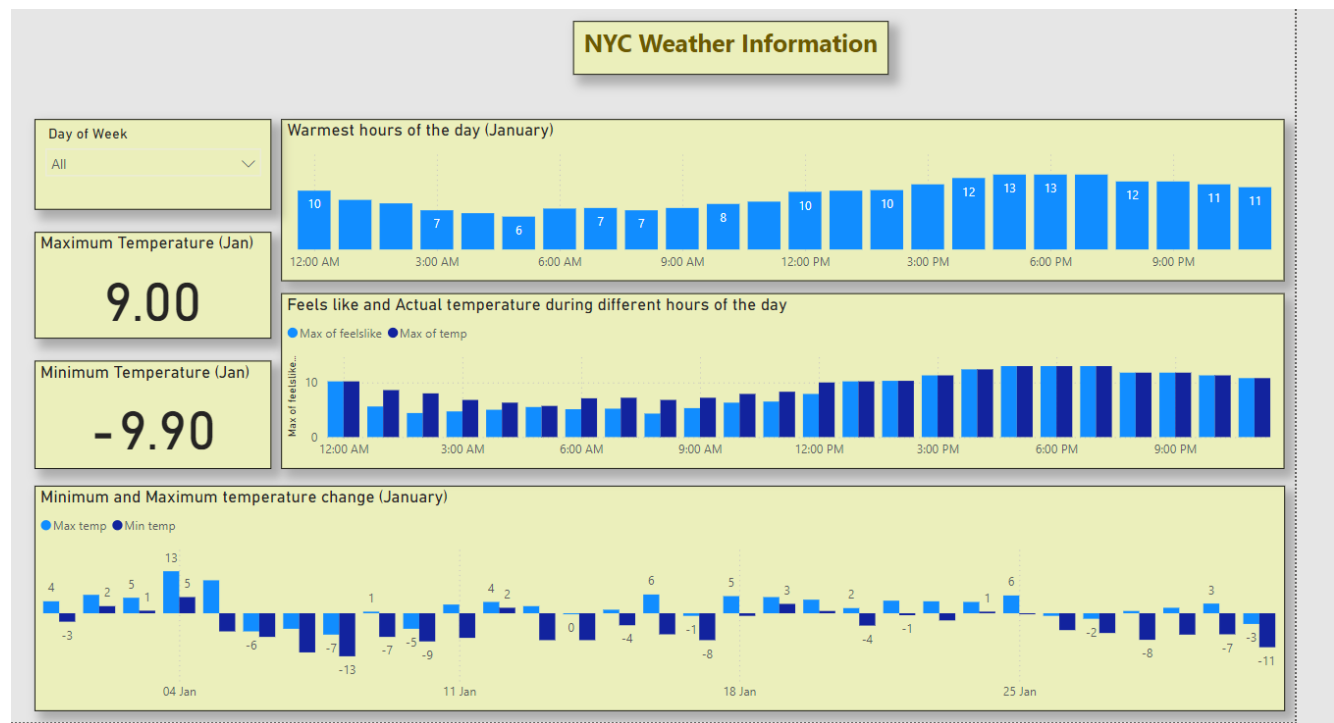In the page, we build six visuals which includes: slicer, card, clustered column chart.

Fig: New York City Weather Information (January 2015)

Reasons for selecting those visuals: A slicer filters data interactively, a card highlights key metrics, and a clustered column chart compares categories side-by-side which enhances data analysis and visualization.

- In slicer, Day of week filter is used.
- Maximum Temperature and Minimum Temperature of January is presented in the card
- Warmest hours of the day (January), Feels like and Actual temperature during different hours of the day and Minimum and Maximum temperature change (January) are displayed using clustered column chart.

Findings:

- Minimum Temperature (303.85% decrease) and Maximum Temperature (189.19% decrease) both trended down between Thursday, January 1, 2015 and Saturday, January 31, 2015.
- Minimum Temperature started trending down on Sunday, January 25, 2015, falling by 5,150.00% (10.30) in 6 days.
- Maximum temperature dropped from 3.70 to 2.70 during its steepest decline between Thursday, January 1, 2015, and Sunday, January 11, 2015.
- All metrics fell in the last Day, each decreasing by 0.50.
- At 2 AM, the temp variance between feels like and actual temperature is 4.

New York City Flights Information (2015)

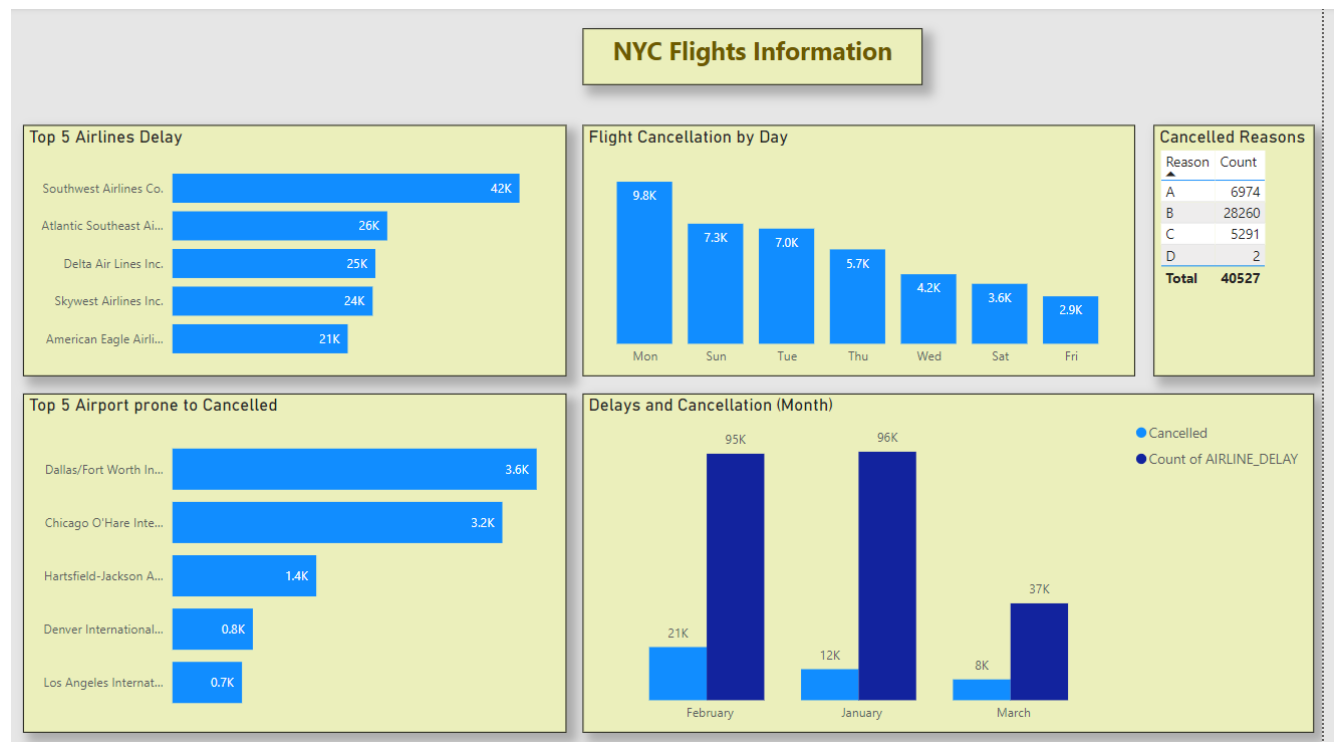In the page, we build five visuals which includes: clustered bar chart, clustered column chart and Table.

Fig: New York City Flights Information (2015)

Reason for selecting those visuals: A table provides detailed data, a clustered column chart compares categories by height, and a clustered bar chart compares categories by length which offers different perspectives for data analysis.
- Cancelled Reasons are displayed in the table.
- The Top 5 Airlines Delay and The Top 5 Airport prone to cancelled is presented in clustered bar chart
- Flight Cancellation by day and Delays and Cancellation (Month) is displayed in clustered column chart.

Findings:

- At 42,195, Southwest Airlines Co. had the highest number of delays and was 98.00% higher than American Eagle Airlines Inc., which had the lowest number of delays at 21,311 among the top 5 airlines delays.
- Dallas/Fort Worth International Airport has the highest number of cancelled flights with 3578 total cancellations.
- On Monday, the highest number of cancellations of flight with number 9823 and Friday has lowest cancellation with number of 2876.
- Number of delays and Cancelled diverged the most when the Month was January, number of delays were 83969 higher than cancelled.

Data Anomaly Centre

In the Data Anomaly Centre, we built the visuals using cards to represent the data outliers.
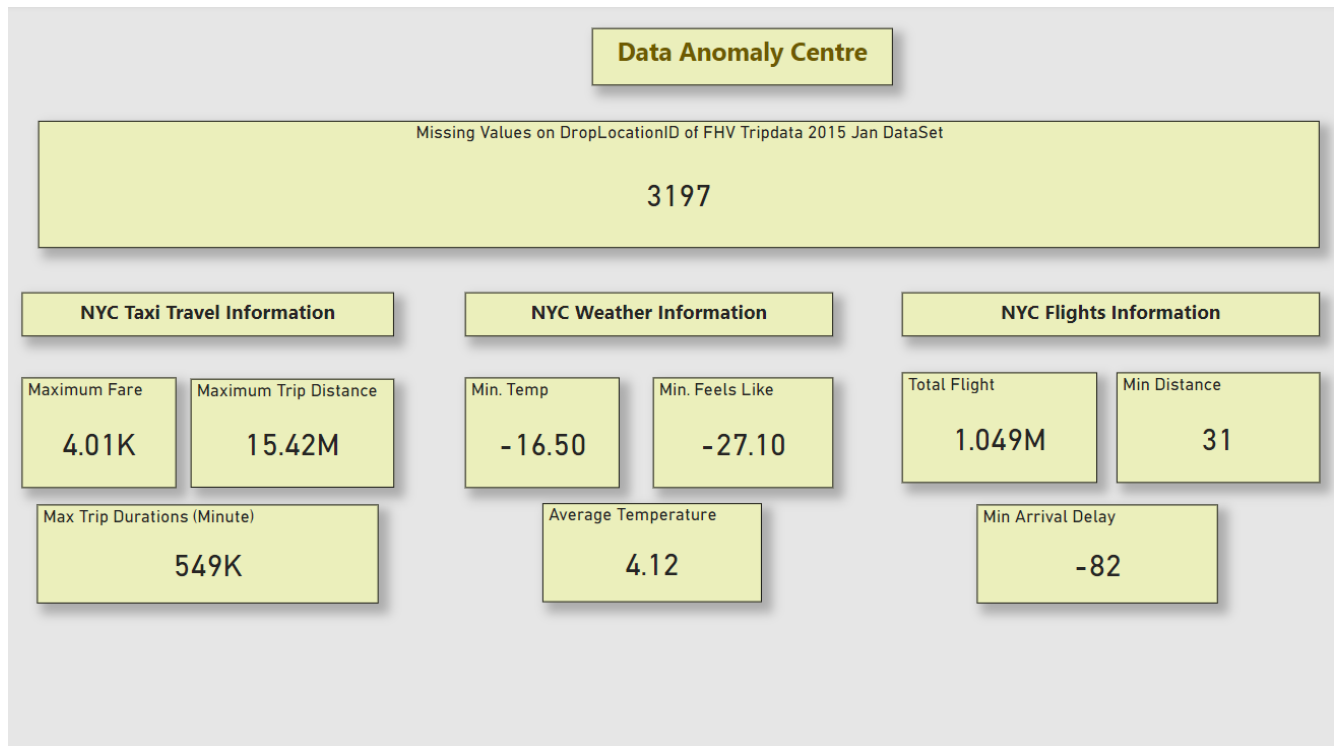
Fig: Data Anomaly Centre

For NYC taxi Travel Information:
- Maximum Fare = 4.01K
- Maximum Trip Distance = 15.42M
- Maximum Trip Duration = 549K

For NYC Weather Information:
- Min. Temperature = -16.50
- Min. Feels Like Temperature = -27.10
- Average Temperature = 4.12

NYC Flights Information:
- Total Flight = 1.049M
- Min. Distance = 31
- Min. Arrival Delay = -82

FHV Tripdata 2015 Jan dataset has missing values on DropLocationID with 3197.

Summary Statistics Centre:

For the summary, we took yellow taxi, temperature daily and flights datasets where we found:
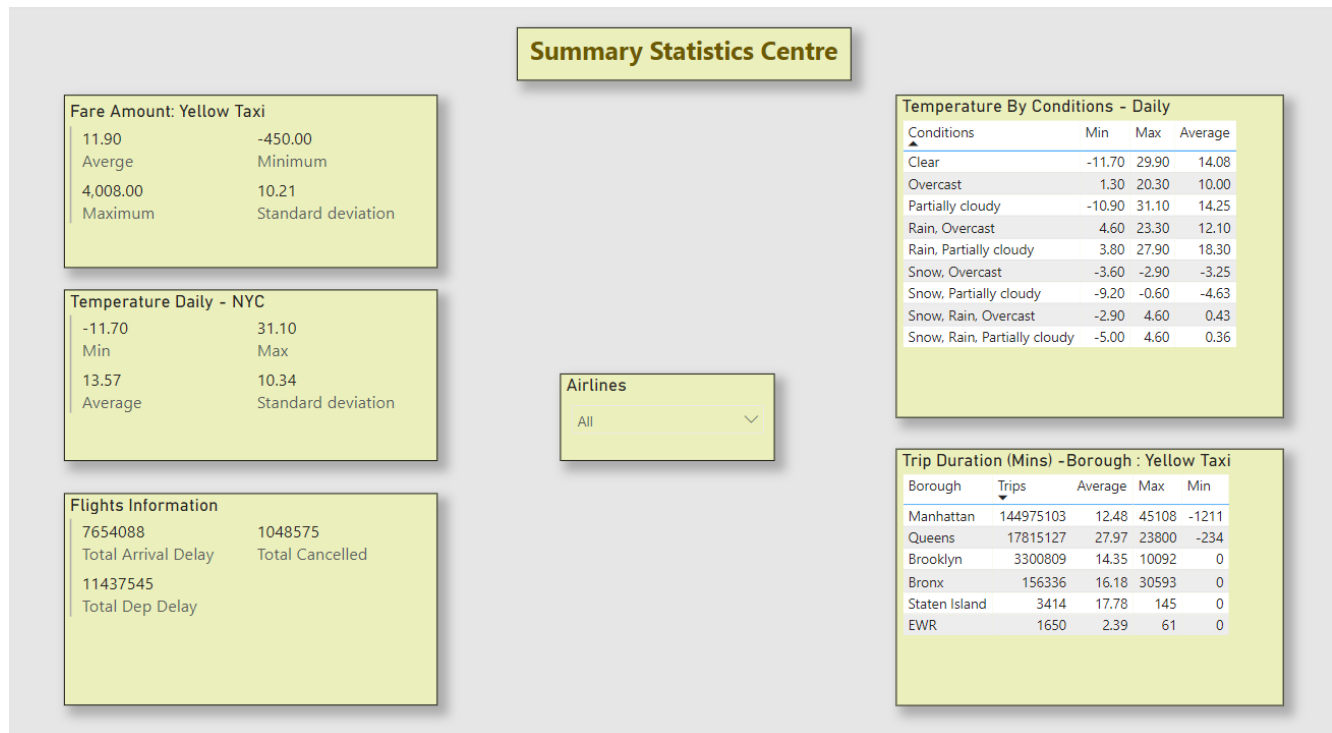

Summary Statistics Centre

Fare Amount: Yellow Taxi

11.90	-450.00
Averge	Minimum
4,008.00	10.21
Maximum	Standard deviation

Temperature Daily - NYC

-11.70	31.10
Min	Max
13.57	10.34
Average	Standard deviation

Flights Information

7654088	1048575
Total Arrival Delay	Total Cancelled
11437545	
Total Dep Delay	

Airlines

All

Temperature By Conditions - Daily

Conditions	Min	Max	Average
Clear	-11.70	29.90	14.08
Overcast	1.30	20.30	10.00
Partially cloudy	-10.90	31.10	14.25
Rain, Overcast	4.60	23.30	12.10
Rain, Partially cloudy	3.80	27.90	18.30
Snow, Overcast	-3.60	-2.90	-3.25
Snow, Partially cloudy	-9.20	-0.60	-4.63
Snow, Rain, Overcast	-2.90	4.60	0.43
Snow, Rain, Partially cloudy	-5.00	4.60	0.36

Trip Duration (Mins) -Borough : Yellow Taxi

Borough	Trips	Average	Max	Min
Manhattan	144975103	12.48	45108	-1211
Queens	17815127	27.97	23800	-234
Brooklyn	3300809	14.35	10092	0
Bronx	156336	16.18	30593	0
Staten Island	3414	17.78	145	0
EWR	1650	2.39	61	0

Fig: Summary Statistics Centre

Fare Amount: Yellow Taxi
- Average: 11.90
- Minimum: -450
- Maximum: 4,008
- Standard Deviation: 10.21

Temperature Daily – NYC
- Average: 13.57
- Minimum: -11.70
- Maximum: 31.10
- Standard deviation: 10.34

Flights Information
- Total Arrival Delay: 7654088
- Total Cancellation: 1048575
- Total Dep Delay: 11437545

Comparison Centre

- Total demand of yellow Taxi in January Weekends was higher for Sat (56) than Sun (38).
- Demand for Sat and Sun diverged the most in the second week weekends, when Sat was 7 higher than Sun.
- FHV and total Green Taxi are positively correlated with each other.

Fig: Comparsion Centre
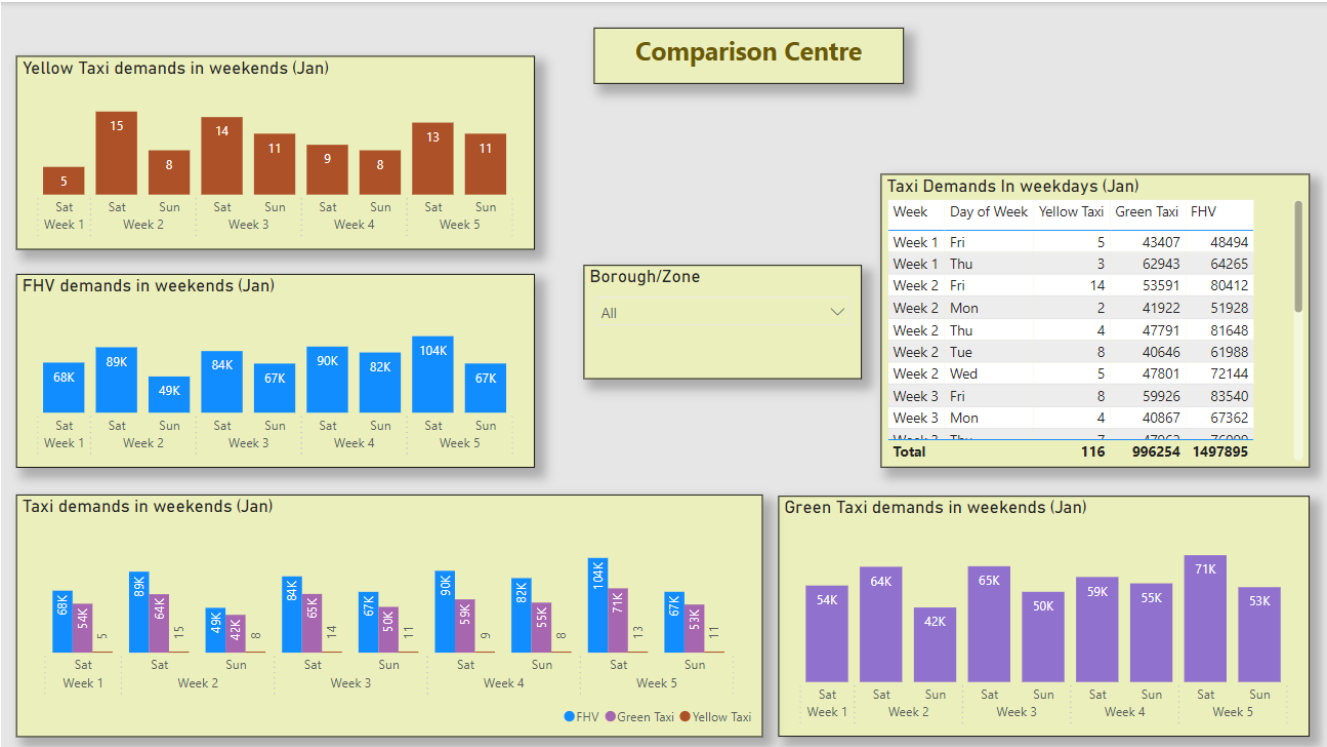
Analytics Dashboard:

- Arrival Delay and Dep. Dealy are positively correlated with each other.
- Across all 31 DAY, Arrival Delay ranged from 1 to 5, Dep. Dealy ranged from 1 to 5, and Cancelled ranged from 0 to 3.
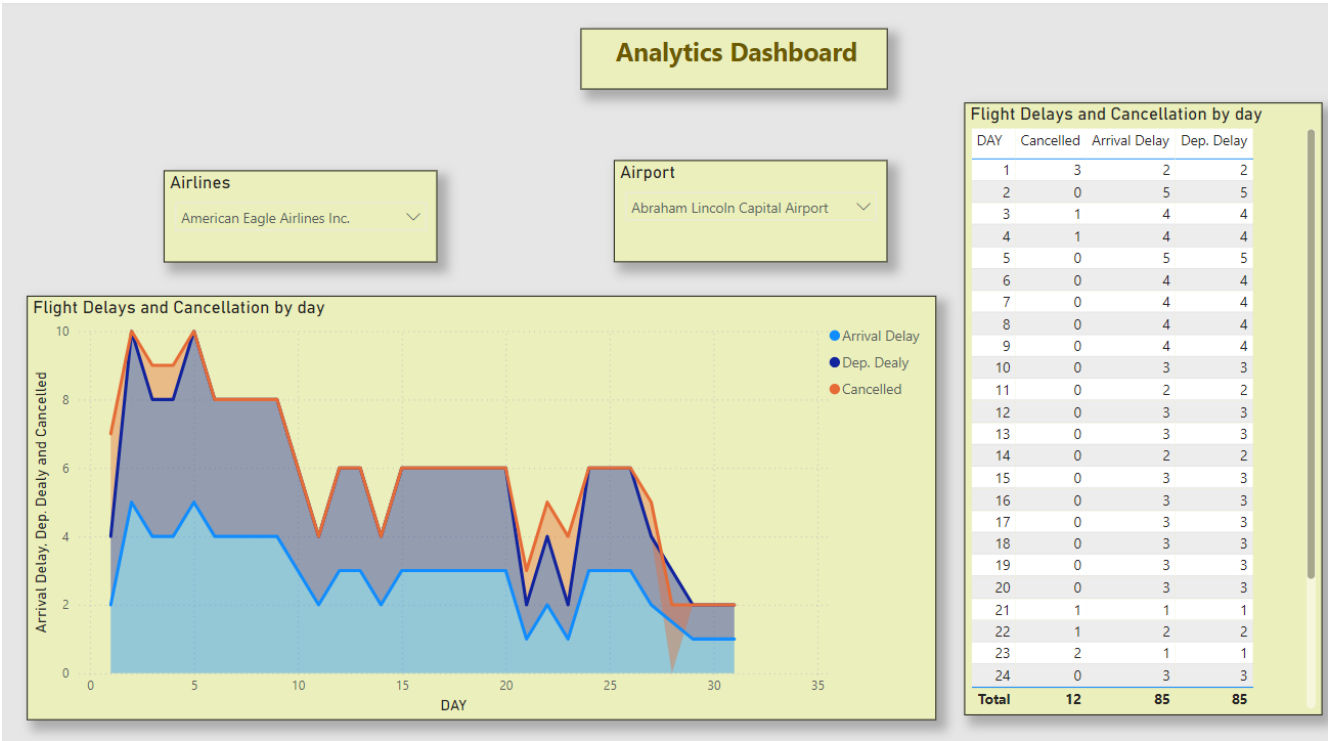

Fig: Analytics Dashboard

Hypothesis Centre:

We can look at the data to get the information of delay and cancellation as per weather condition. In the page there is the filter of airlines, weather condition and month/day.
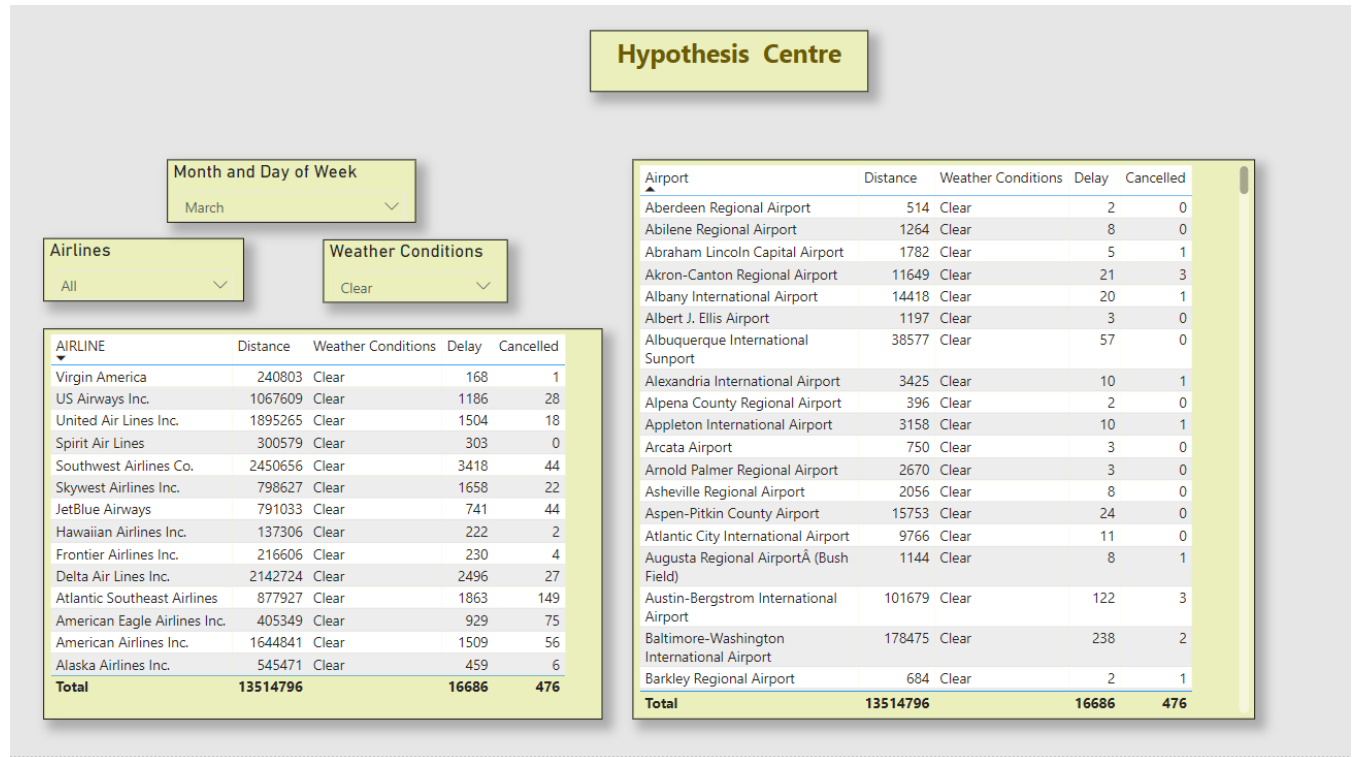

Fig: Hypothesis Centre

My Story:

In the page, we have four visual which represents total flights by day and weather conditions, taxi and Flights by day, taxi demand by Borough and Taxi and Flights by Day (Jan).
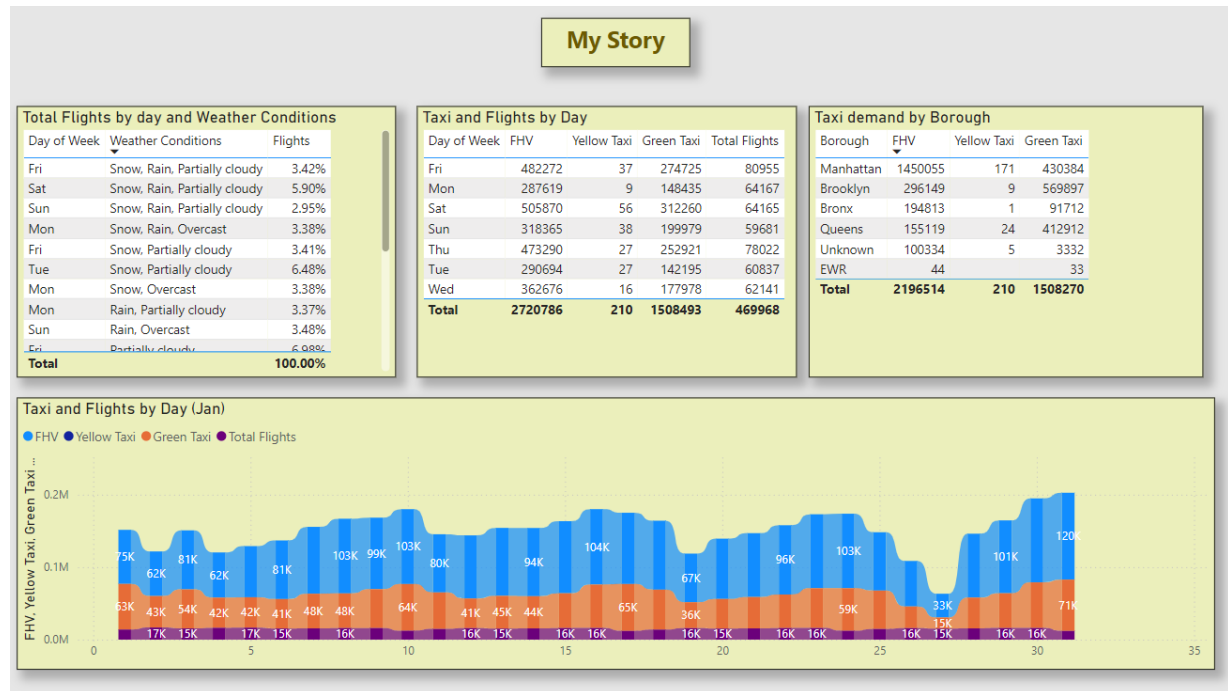

Fig: My Story

Findings:

- In the Friday, there is high demand of Taxi and also there is a greater number of flights with number of 80955.
- 10.80% of flights are takeoff in Thursday and weather is partially cloudy.
- In January, the demand for yellow taxi is low and for hire vehicles are highly demanded.
- Most demanded of taxi on Manhattan area and least one is EWR.