Airline Flight Delay & Price Analysis

Pre-Placement Project

By SACHIN B C

Content

Project Objective

Data Cleaning and Preparation

Connecting to SQL & Updating Data

EDA Using Python

MySQL Queries

Tableau Dashboard

Project Objective

The primary objective is to enhance operational efficiency, flight price, stay price, revenue generated and elevate customer satisfaction by implementing targeted strategies derived from the insights obtained through the analysis of the airline dataset.

Data Cleaning and Preparation

Getting datatypes

```
info = data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48884 entries, 0 to 48883
Data columns (total 26 columns):
 #   Column                     Non-Null Count  Dtype  
---  ------                     --------------  -----  
 0   flight                     48884 non-null  int64  
 1   flight_code                48884 non-null  object 
 2   cust_id                    48884 non-null  int64  
 3   cust_name                  48862 non-null  object 
 4   city                       48884 non-null  object 
 5   neighbourhood              48884 non-null  object 
 6   latitude                   48884 non-null  float64
 7   longitude                  48884 non-null  float64
 8   stay_type                  48884 non-null  object 
 9   stay_price                 48884 non-null  int64  
 10  flight_price               48884 non-null  int64  
 11  night_flight               48884 non-null  int64  
 12  total_number_travel        48884 non-null  int64  
 13  total_cust_listings_count  48884 non-null  int64  
 14  availability_365           48884 non-null  int64  
 15  year                       48884 non-null  int64  
 16  month                      48884 non-null  int64  
 17  day                        48884 non-null  int64  
 18  sched_dep_time             48884 non-null  object 
 19  dep_time                   48171 non-null  object 
 20  dep_delay_min              48171 non-null  float64
 21  sched_arr_time             48884 non-null  object 
 22  arr_time                   48149 non-null  object 
 23  arr_delay_min              48062 non-null  float64
 24  air_time_min               48062 non-null  float64
 25  distance                   48884 non-null  int64  
dtypes: float64(5), int64(12), object(9)
memory usage: 9.7+ MB
```

Checking for Missing Values, Updating & Removing null values

- Total 7 columns are having missing values
- 1. cust_name 22
- 2. dep_time 713
- 3. dep_delay_min 713
- 4. arr_time 735
- 5. arr_delay_min 822
- 6. air_time_min 822

Connecting to SQL & Updating Data

Connecting Flight Details Table

```
flights_details = pd.DataFrame(data, columns=['flight', 'flight_code', 'sched_dep_time', 'dep_time', 'dep_delay_min'
                                              'sched_arr_time', 'arr_time', 'arr_delay_min', 'air_time_min', 'distance'])
```

```
flights_details.head()
```

	flight	flight_code	sched_dep_time	dep_time	dep_delay_min	sched_arr_time	arr_time	arr_delay_min	air_time_min	distance
0	1545	N14228	05:15:00	05:17:00	2.0	08:19:00	08:30:00	11.0	227.0	1400
1	1714	N24211	05:29:00	05:33:00	4.0	08:30:00	08:50:00	20.0	227.0	1416
2	1141	N619AA	05:40:00	05:42:00	2.0	08:50:00	09:23:00	33.0	160.0	1089
3	725	N804JB	05:45:00	05:44:00	-1.0	10:22:00	10:04:00	-18.0	183.0	1576
4	461	N668DN	06:00:00	05:54:00	-6.0	08:37:00	08:12:00	-25.0	116.0	762

```
try:
    data.to_sql('flights_details', engine, if_exists='replace', index=False)
    print("Table updated successfully.")
except Exception as e:
    print("Error occurred:", e)
```

```
Table updated successfully.
```

Connecting Customer Details Table

```
customers_details = pd.DataFrame(data, columns=['cust_id', 'cust_name', 'city', 'neighbourhood', 'latitude', 'longit
```

```
customers_details.head()
```

	cust_id	cust_name	city	neighbourhood	latitude	longitude
0	2787	John	Brooklyn	Kensington	40.64749	-73.97237
1	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377
2	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190
3	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976
4	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399

```
try:
    data.to_sql('customers_details', engine, if_exists='replace', index=False)
    print("Table updated successfully.")
except Exception as e:
    print("Error occurred:", e)
```

```
Table updated successfully.
```

Connecting Booking Details Table

```
bookings_details = pd.DataFrame(data, columns=['flight', 'flight_code', 'cust_id', 'stay_type', 'stay_price', 'fligh
                                      'night_flight', 'total_number_travel', 'total_cust_listings_count',
                                      'availability_365', 'year', 'month', 'day'])
```

```
bookings_details.head()
```

	flight	flight_code	cust_id	stay_type	stay_price	flight_price	night_flight	total_number_travel	total_cust_listings_count	availability_365	year	month	day
0	1545	N14228	2787	Private room	149	714	1	9	6	365	2013	1	1
1	1714	N24211	2845	Entire home/apt	225	1079	1	45	2	355	2013	1	1
2	1141	N619AA	4632	Private room	150	719	3	2	1	365	2013	1	1
3	725	N804JB	4869	Entire home/apt	89	427	1	270	1	194	2013	1	1
4	461	N668DN	7192	Entire home/apt	80	384	10	9	1	0	2013	1	1

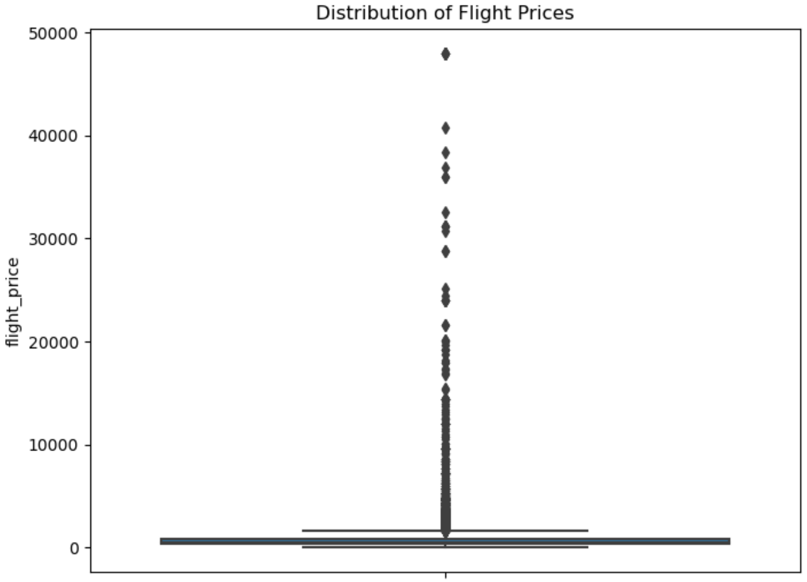
```
try:
    data.to_sql('bookings_details', engine, if_exists='replace', index=False)
    print("Table updated successfully.")
except Exception as e:
    print("Error occurred:", e)
```

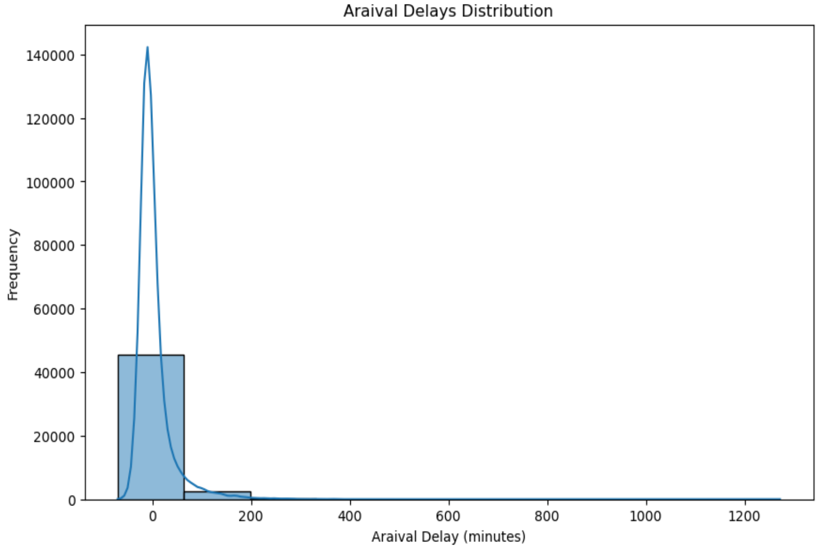
```
Table updated successfully.
```

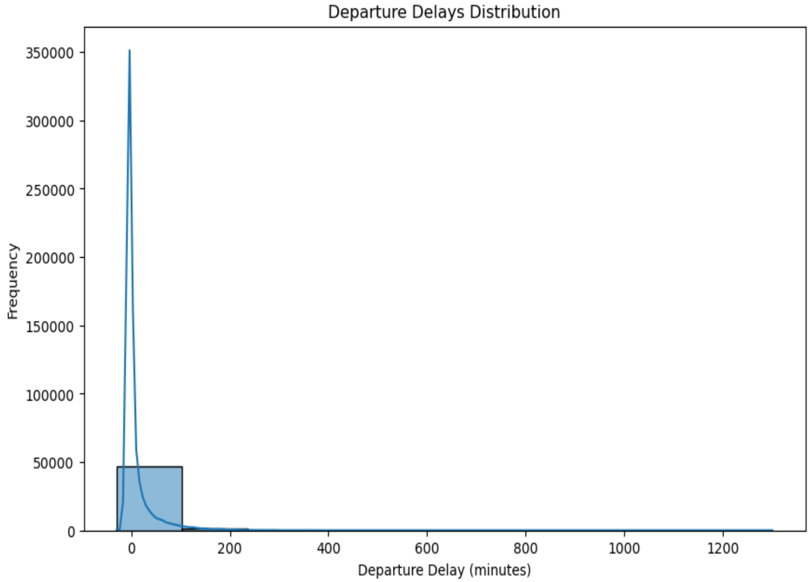
EDA Using Python

Univariate Analysis:

Distribution of Flight Prices, Departure & Arrival Delays Distribution

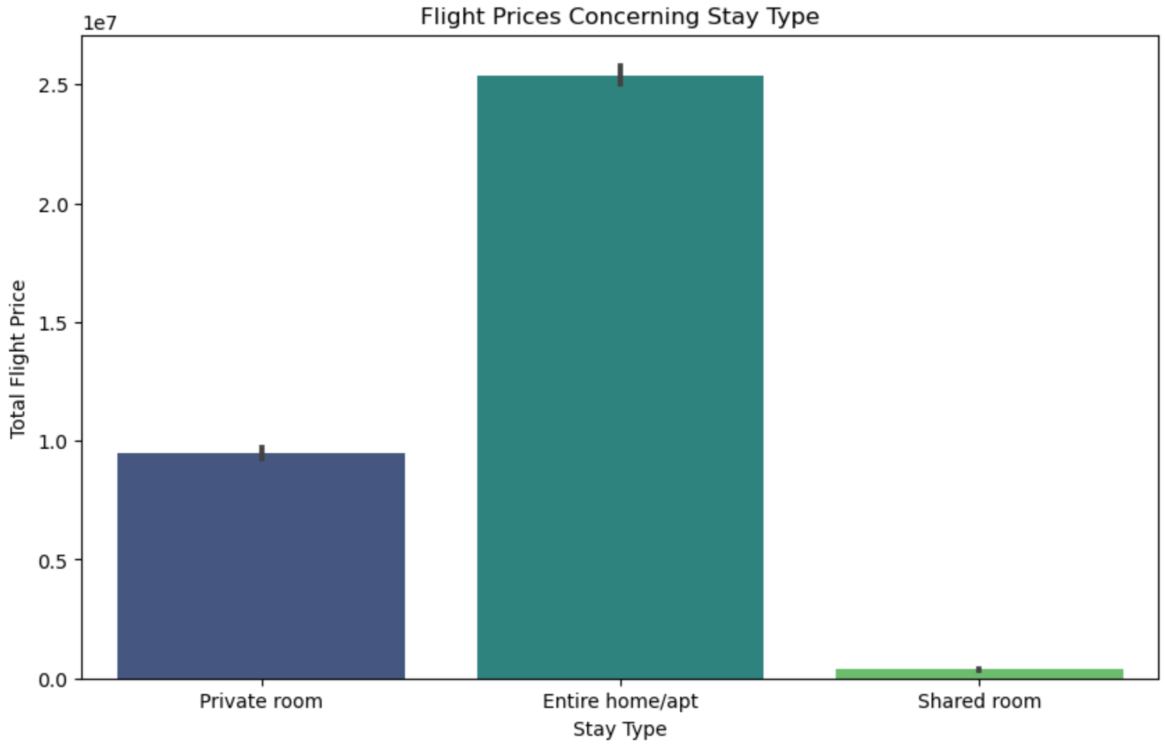


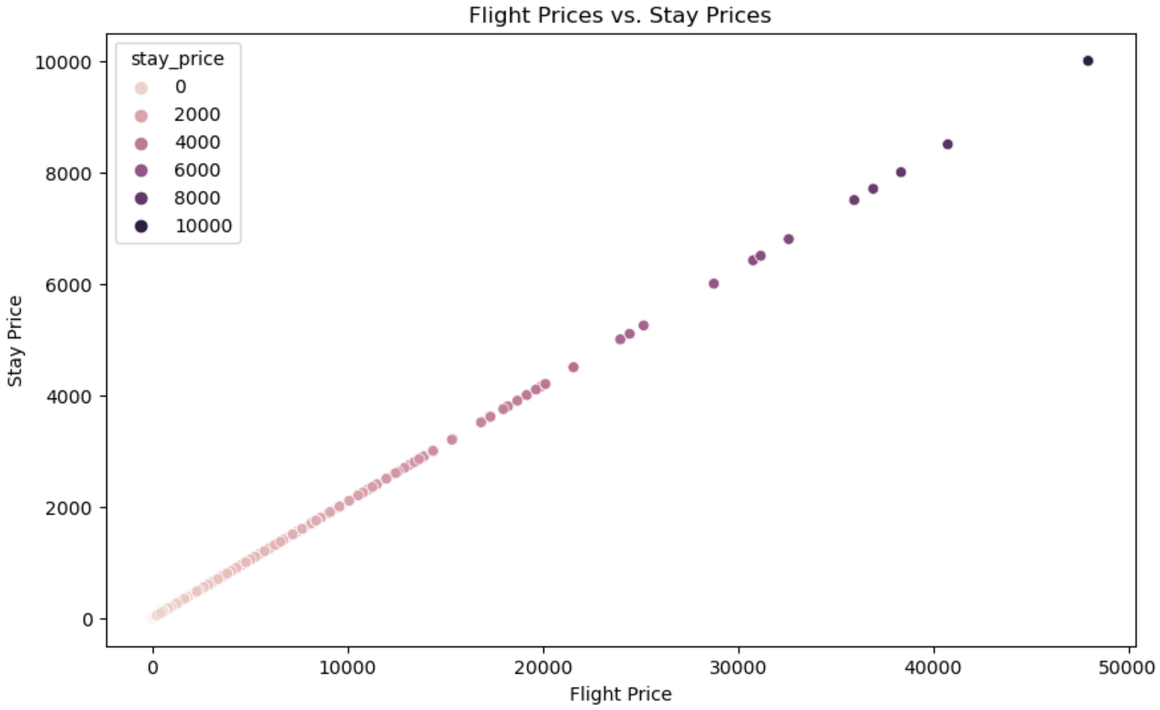




Bivariate Analysis:

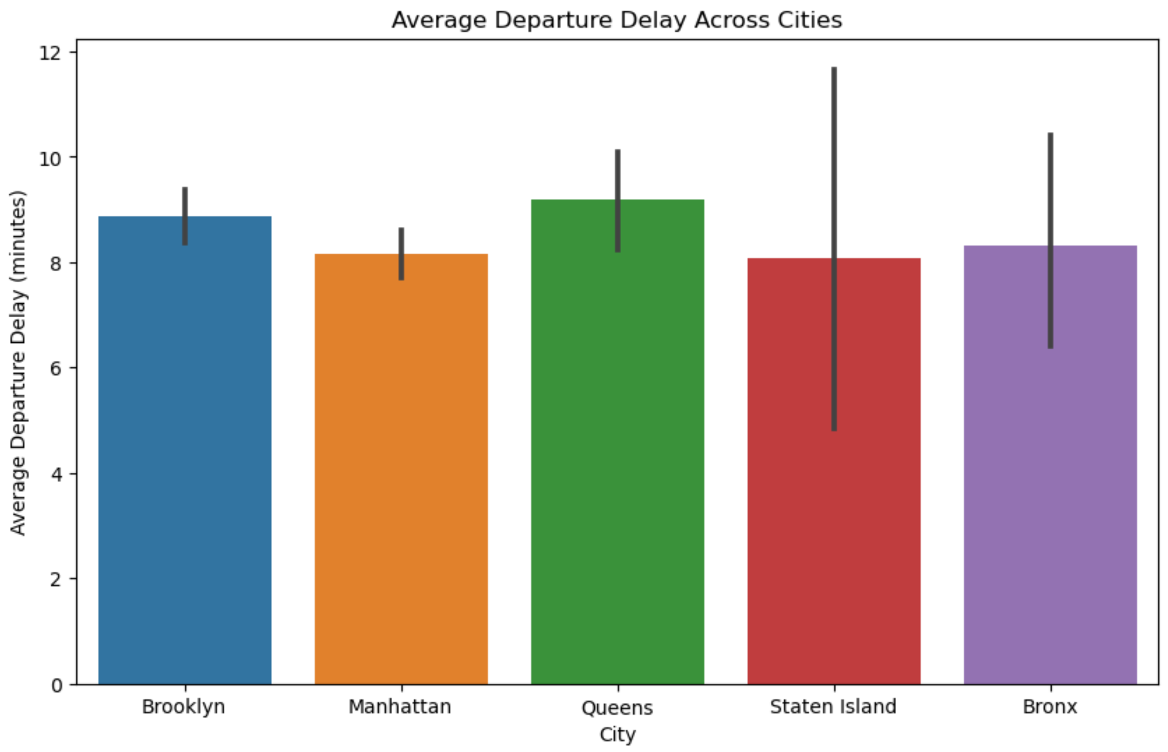
Flight Prices vs. Stay Prices & Flight Prices Concerning Stay Type

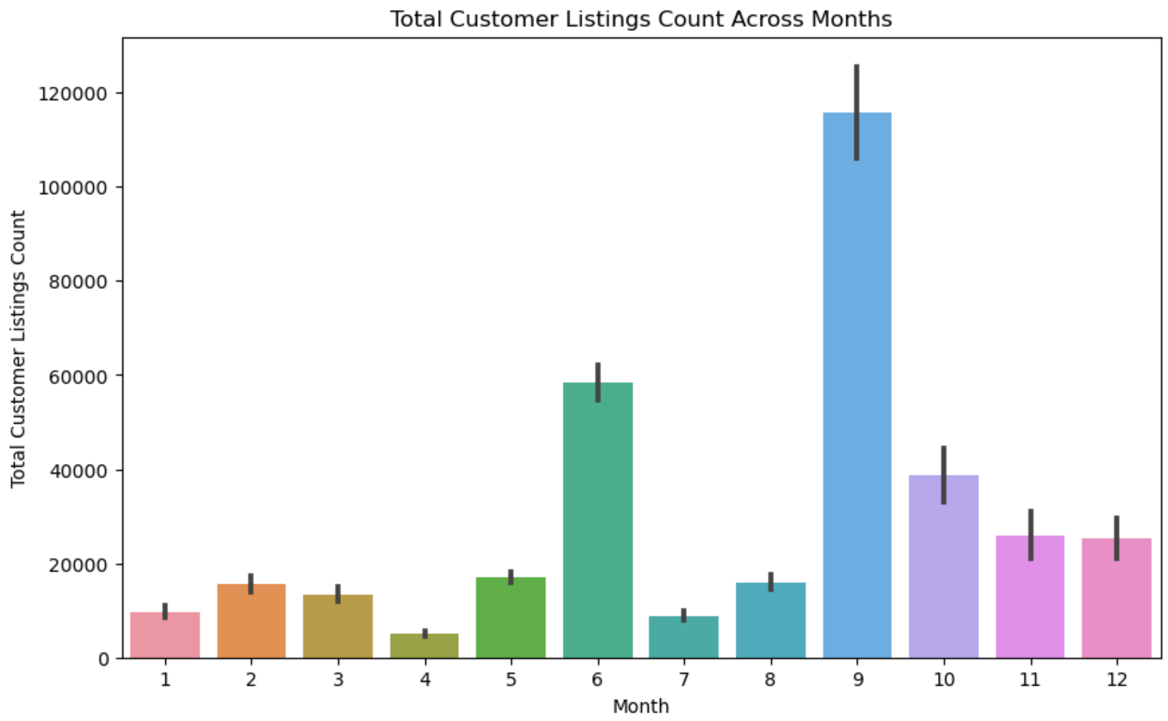




Bivariate Analysis:

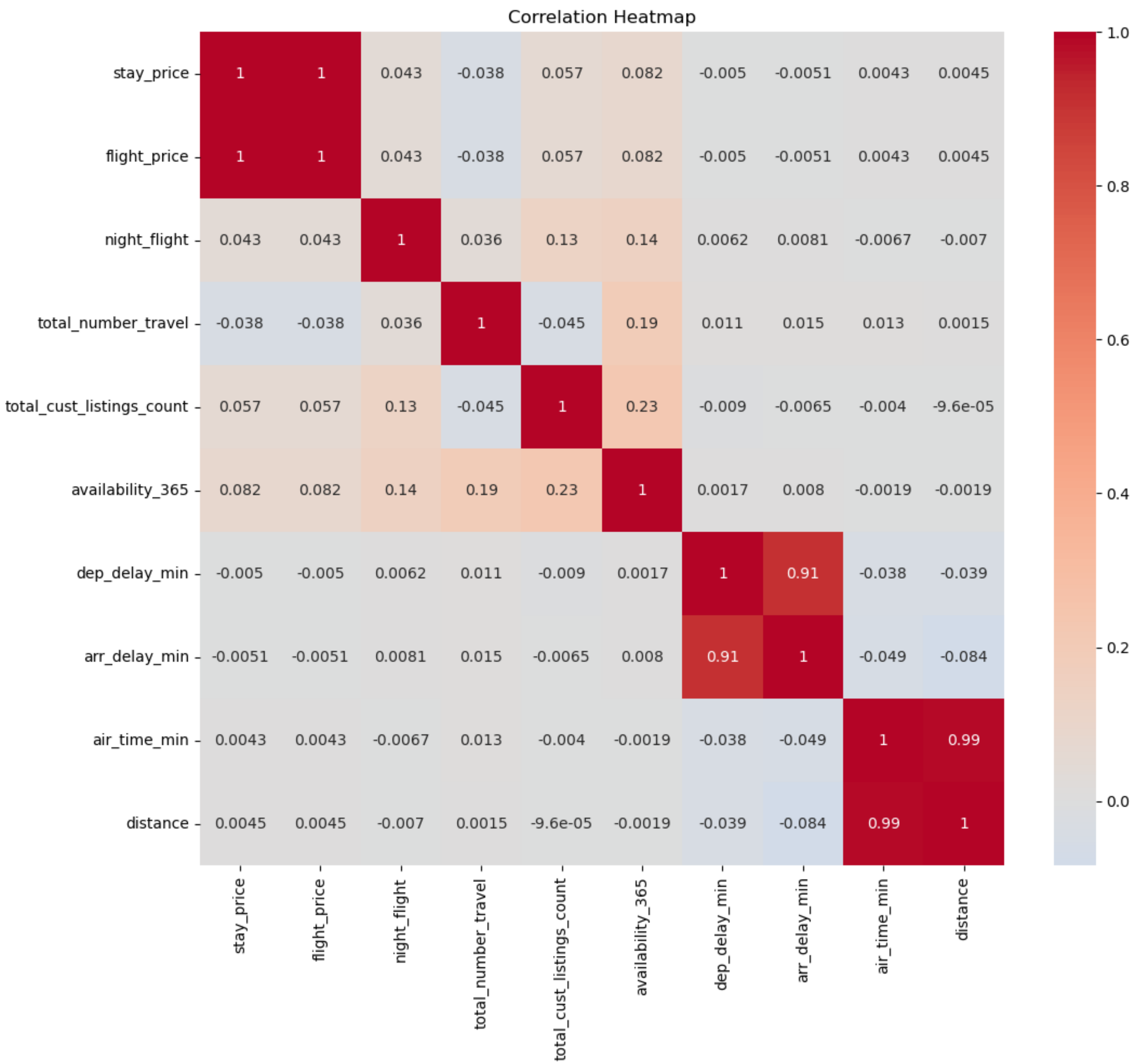
Average Departure Delay Across Cities & Total Customer Listings Count Across Months





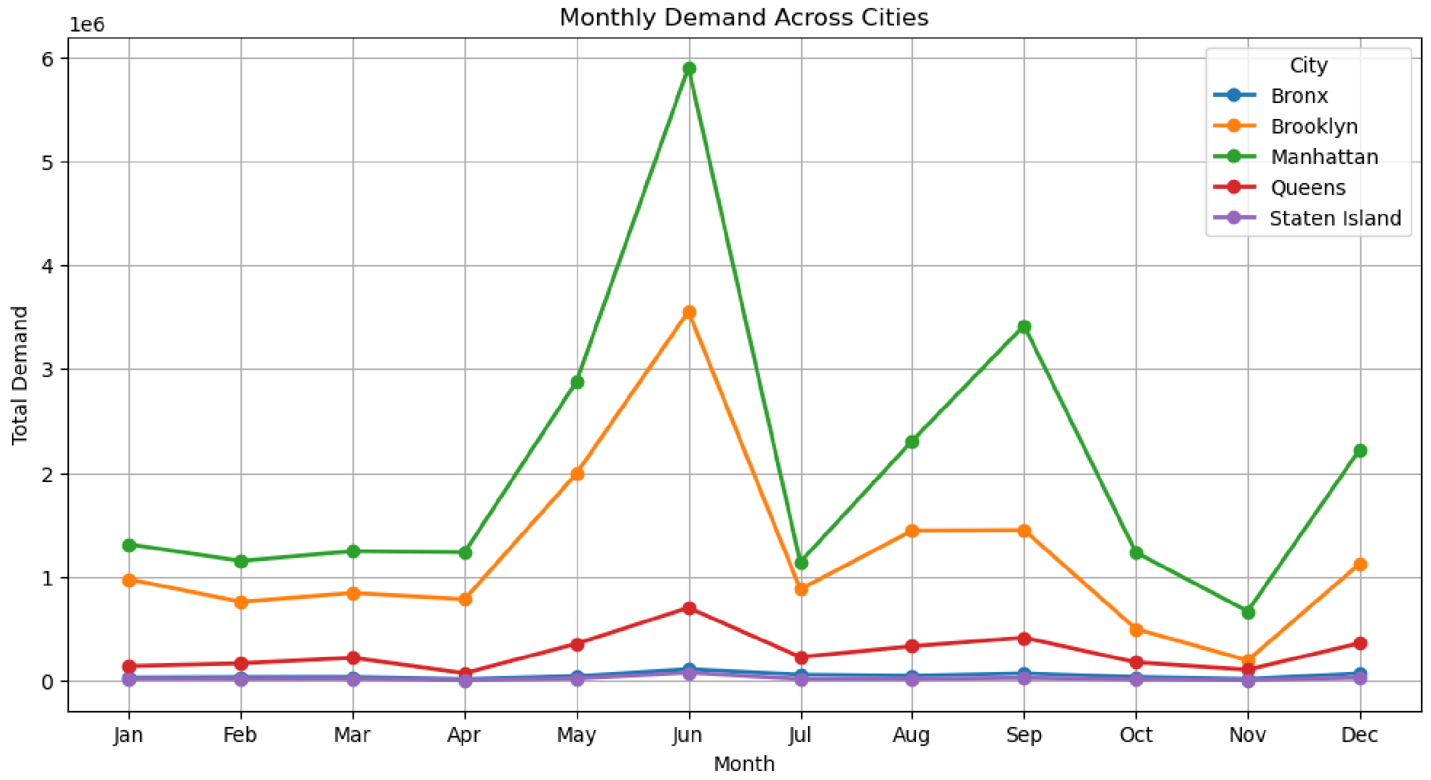
Multivariate Analysis:

Correlation Heatmap

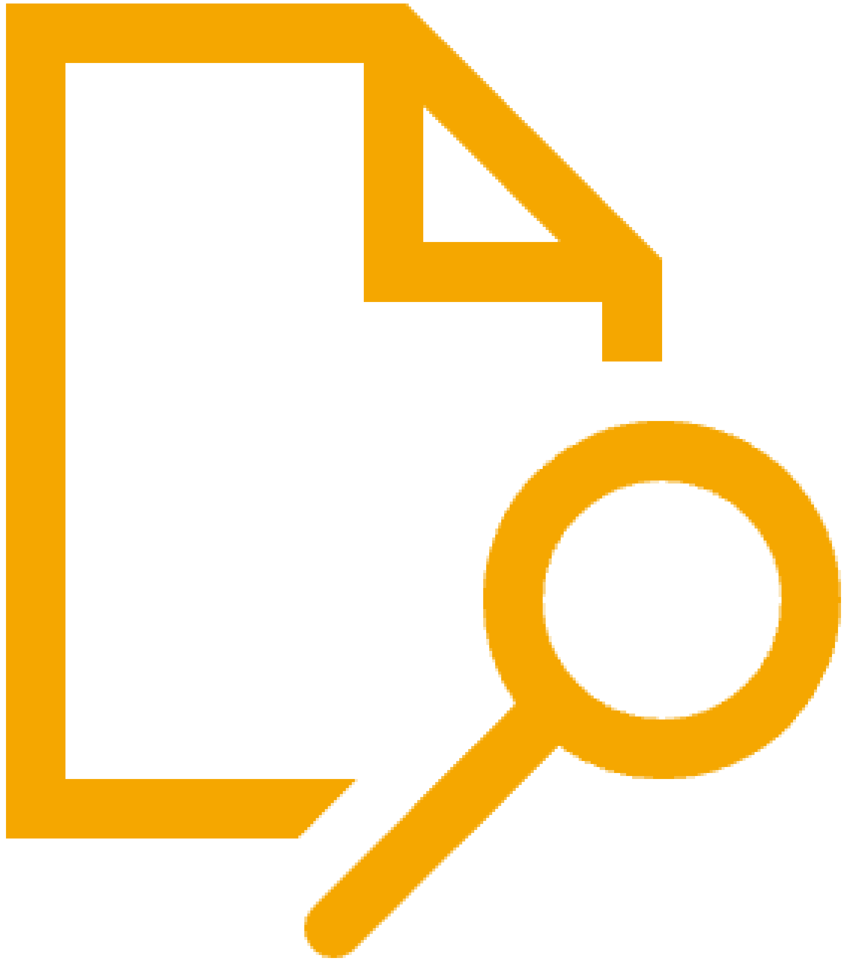


Multivariate Analysis:

Monthly Demand Across Cities



MySQL Queries



1. Retrieve the total number of bookings made by customers = '48062'
2. average flight delay in minutes = '8.5630'
3. number of customers per city
4. top 5 flights with the most bookings.
5. top 10 customer details along with their corresponding flight and flight code

city	customer_cou...
Brooklyn	19754
Manhattan	21317
Queens	5553
Staten Island	366
Bronx	1072

3

flight_code	booking_cou...
N725MQ	97
N526MQ	76
N283JB	76
N0EGMQ	75
N947UW	73

4

cust_name	flight	flight_code
John	1545	N14228
Jennifer	1714	N24211
Elisabeth	1141	N619AA
LisaRoxanne	725	N804JB
Laura	461	N668DN
Chris	1696	N39463
Garon	507	N516JB
Shunichi	5708	N829AS
MaryEllen	79	N593JB
Ben	3393	N912XJ

5

6. total revenue generated from bookings by city

7. number of bookings for each stay type

8. flight details for flights that have availability for 365 days

9. light details for flights that have bookings

city	total_revenue
Manhattan	24341951
Brooklyn	14228409
Queens	3209642
Bronx	546311
Staten Island	244141

6

stay_type	booking_cou...
Entire home/apt	25000
Private room	21928
Shared room	1134

7

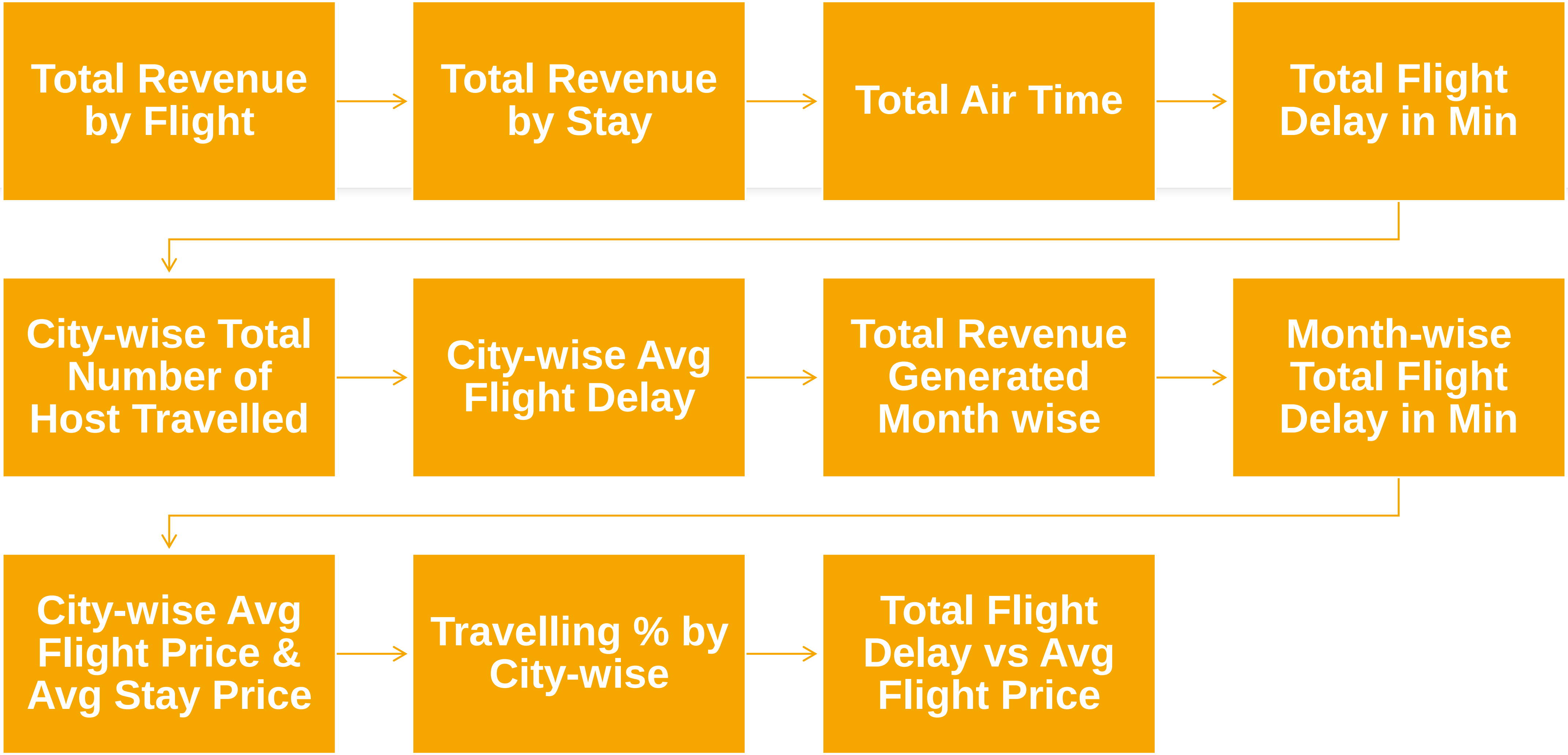
flight	flight_code	availability_365
1545	N14228	365
1141	N619AA	365
413	N3BAAA	365
303	N3CYAA	365
2263	N325US	365
443	N554UA	365
485	N371NB	365
1519	N24715	365
766	N957WN	365
251	N641VA	365
4	N503JB	365

8

flight	flight_code	cust_id
1545	N14228	24496111
1545	N14228	860636
1545	N14228	7503643
1545	N14228	2787
1714	N24211	2845
1141	N619AA	14537404
1141	N619AA	1607111
1141	N619AA	20261309
1141	N619AA	42623155

9

Tableau Dashboard


Total Revenue by Flight
Total Revenue by Stay
Total Air Time
Total Flight Delay in Min
City-wise Total Number of Host Travelled
City-wise Avg Flight Delay
Total Revenue Generated Month wise
Month-wise Total Flight Delay in Min
City-wise Avg Flight Price & Avg Stay Price
Travelling % by City-wise
Total Flight Delay vs Avg Flight Price

Airline Flight Delay & Price Dashboard

Total Revenue by Flight

$ 35,802K — Avg Price: $ 732

Total Revenue by Stay

$ 7,466K — Avg Price: $ 153

Total Air Time

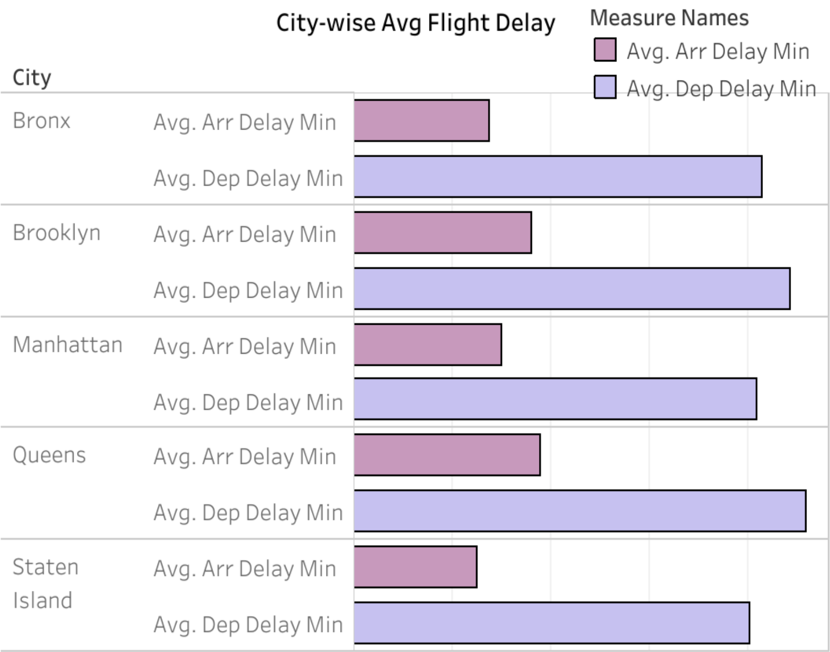
7,281K min — Avg Air Time: 151.5

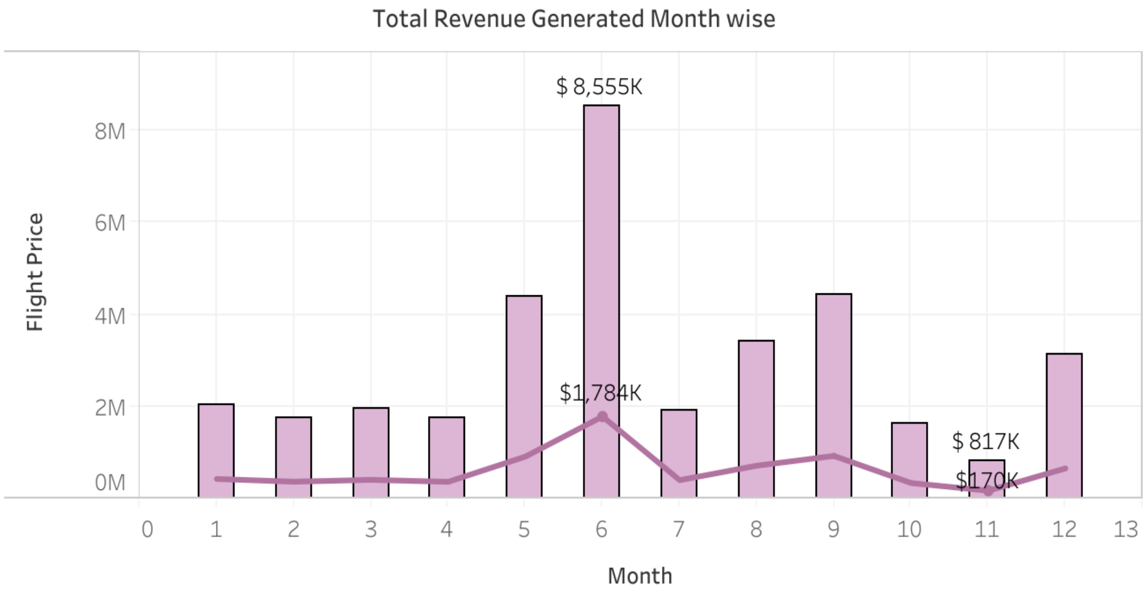
Total Flight Delay in Min

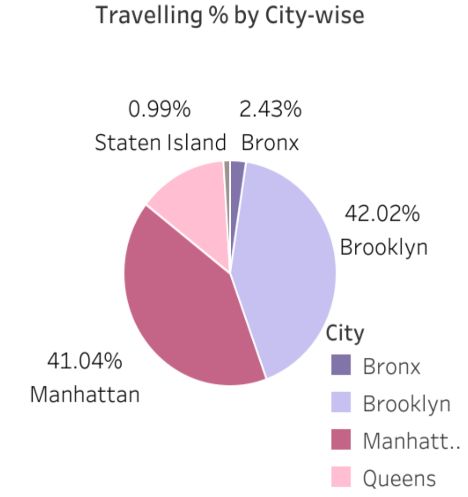
7,695K min — Dep Delay: 414K min, Arr Delay: 161K min

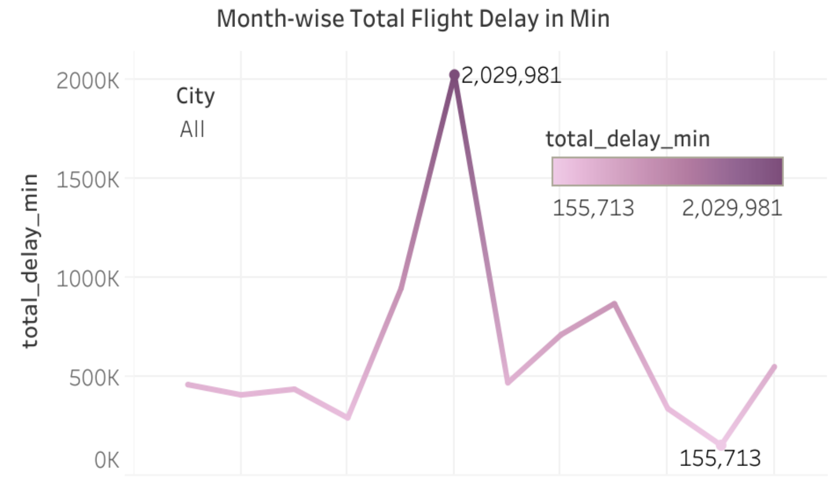
City-wise Total Number of Host Travelled

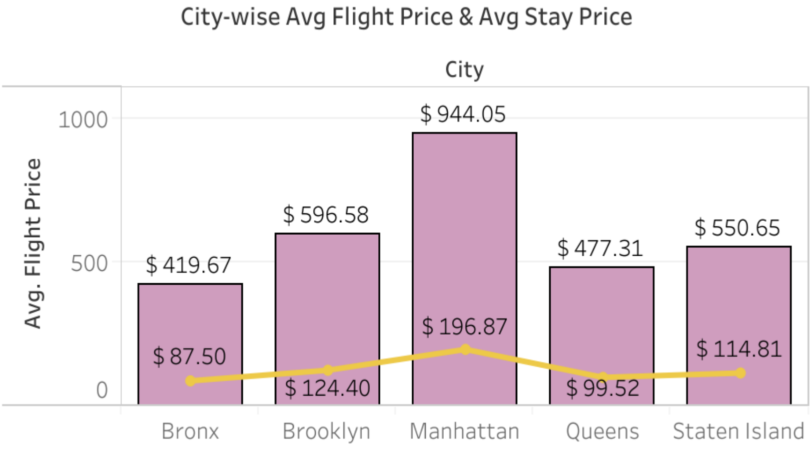
Bronx	2,437
Brooklyn	45,914
Manhattan	277,066
Queens	23,005
Staten Island	865

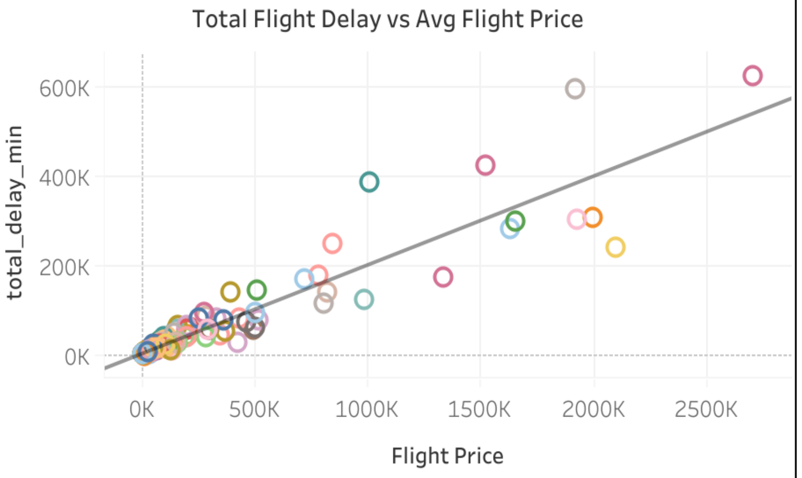












Conclusion

The analysis of the airline dataset has provided valuable insights into various aspects of flight operations, customer behavior, and accommodation preferences. Key findings from the exploration of the dataset include:

- Flight Delays Impact
- Total Delay Time and Average Delay
- City wise Analysis of Flight
- Revenue Analysis
- Availability of Flights
- Stay Types & Stay Price Insights

In conclusion, the dataset analysis has provided a comprehensive understanding of customer behavior, operational efficiency, and revenue dynamics for the airline. The insights gained can serve as a foundation for data-driven decision-making, allowing the airline to enhance its services, optimize resource allocation, and tailor marketing strategies to meet customer preferences.

THANK YOU