Analysis of Base Flight Fare

By Neha Awasthi

Objective

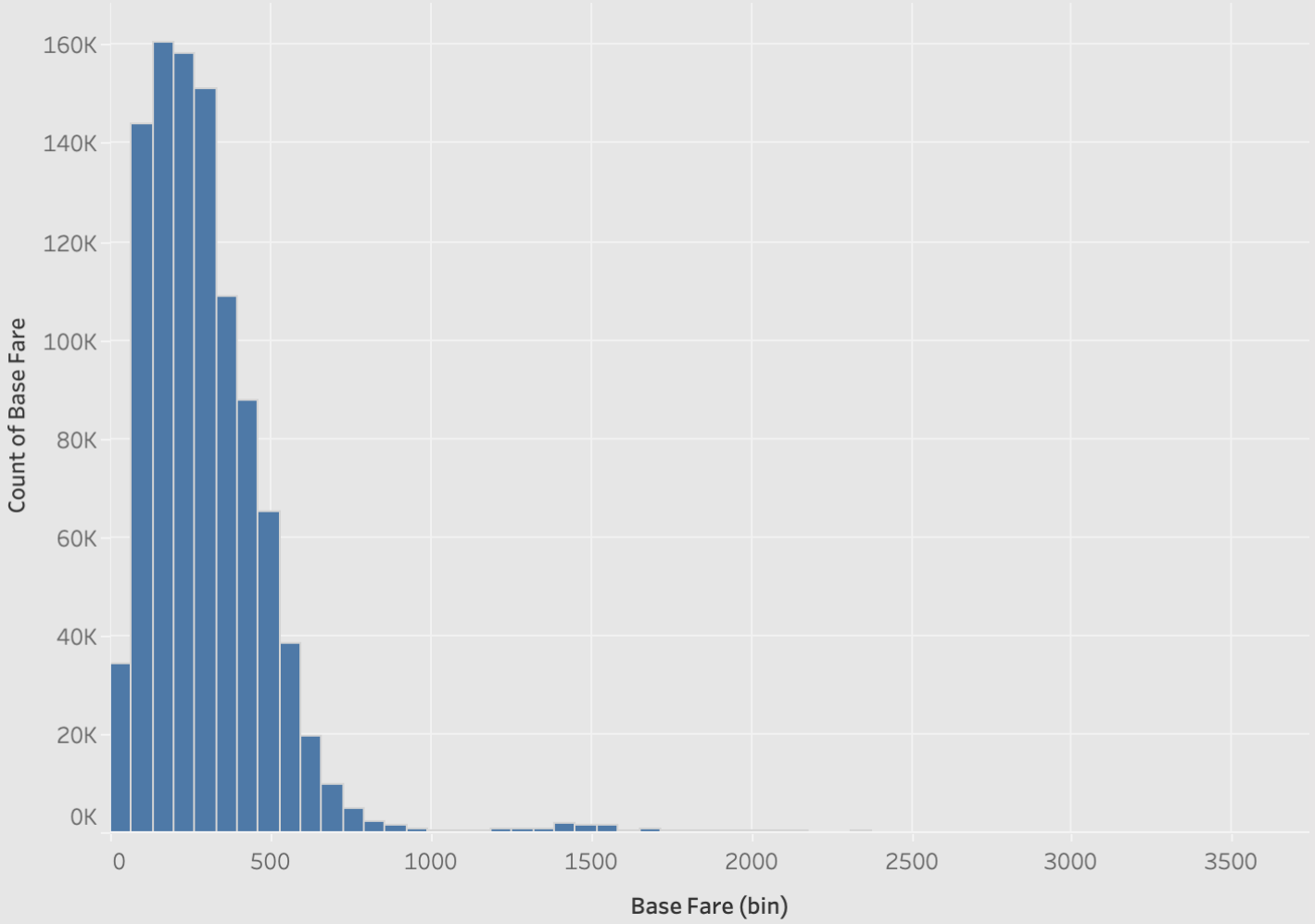
- In this Analysis, We try to understand various factors impacting the Base fares of flights in the United States of America for a given period of time.

- We also analyze the relationships between each factor and the Flight Base Fare.

- This is primarily an exploratory Data analysis and we try to answer the questions regarding the various factors contributing to change in Base fares of the flights.

How is Base Fare Distributed?

Distribution of Base Fare



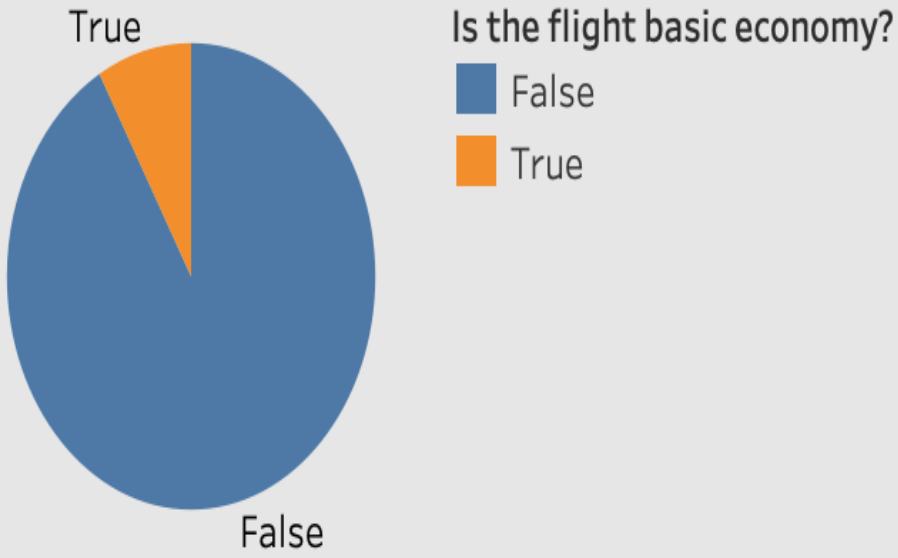
Histogram of Base Flight Fare to understand the distribution of Base Fare.

Base Fare ranges up to 1000 $ with a few outliers around 1500$.

Base Fare averages around 300$

How are Flights distributed according to economic classes?

Distribution of Flights according to Class

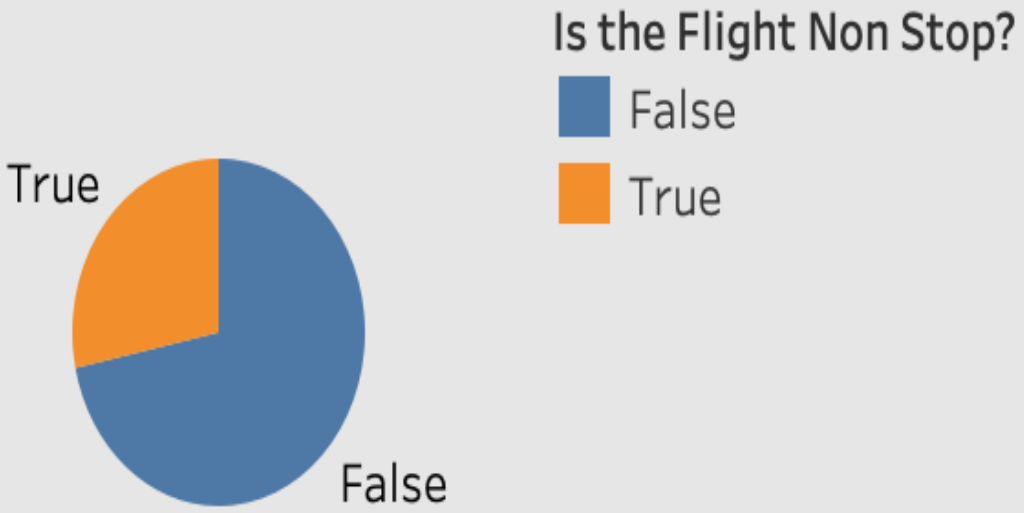


Pie Chart of Flight Classes to understand the distribution.

Majority of the data is from Flights that are not basic economy

How are Flights distributed according to stops in the Flights?

Distribution of Flights Considering whether they are Direct or not



Pie Chart of Flights to understand the distribution of Direct and Non-Direct Flights.

Most of the Flights (Almost 3/4th) are Flights with one or more Stops.

How are Flight distributed according to the refundability of flights?

Distribution of Flights according to the refundability of the Flights

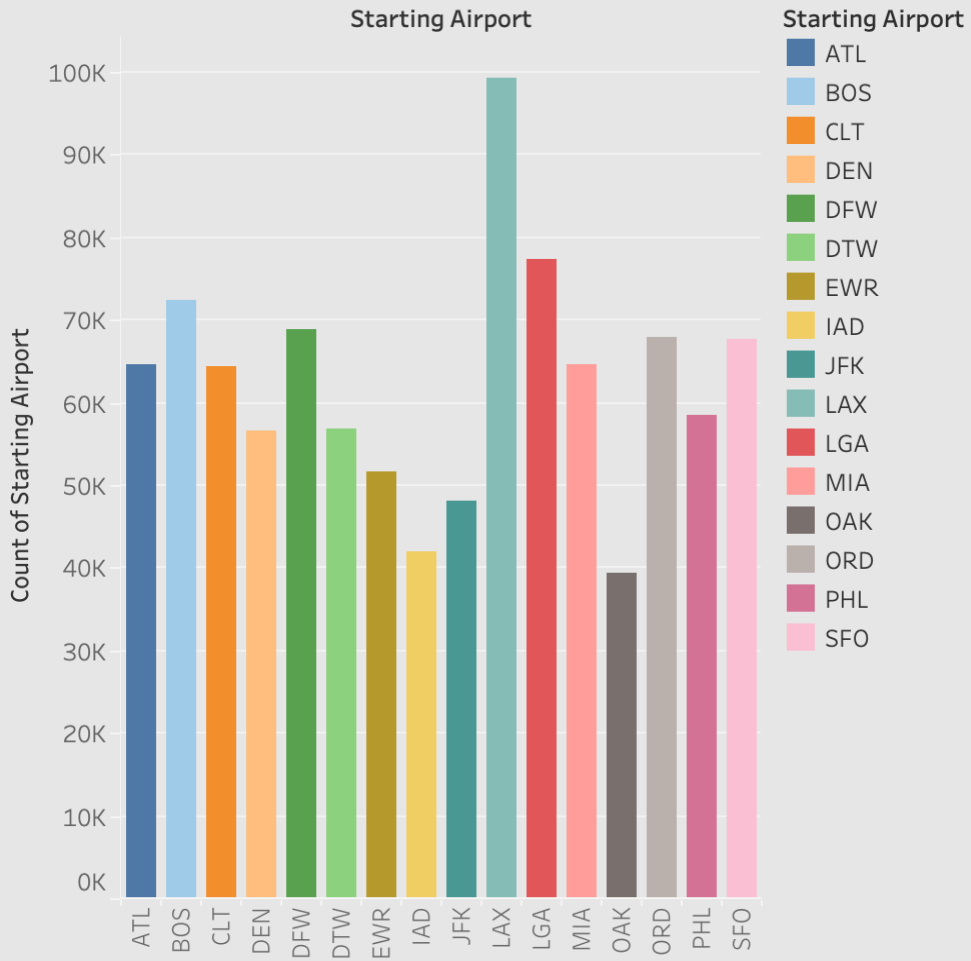
Is Refundable	
False	999,994
True	6

Pivot Table to understand the distribution of Flights among Refundable and Non-Refundable.

Overwhelming no. of flights are non-refundable with a negligible count for Refundable flights in the data.

How are Flights distributed according to originating Airports?

Flights originating from Airports



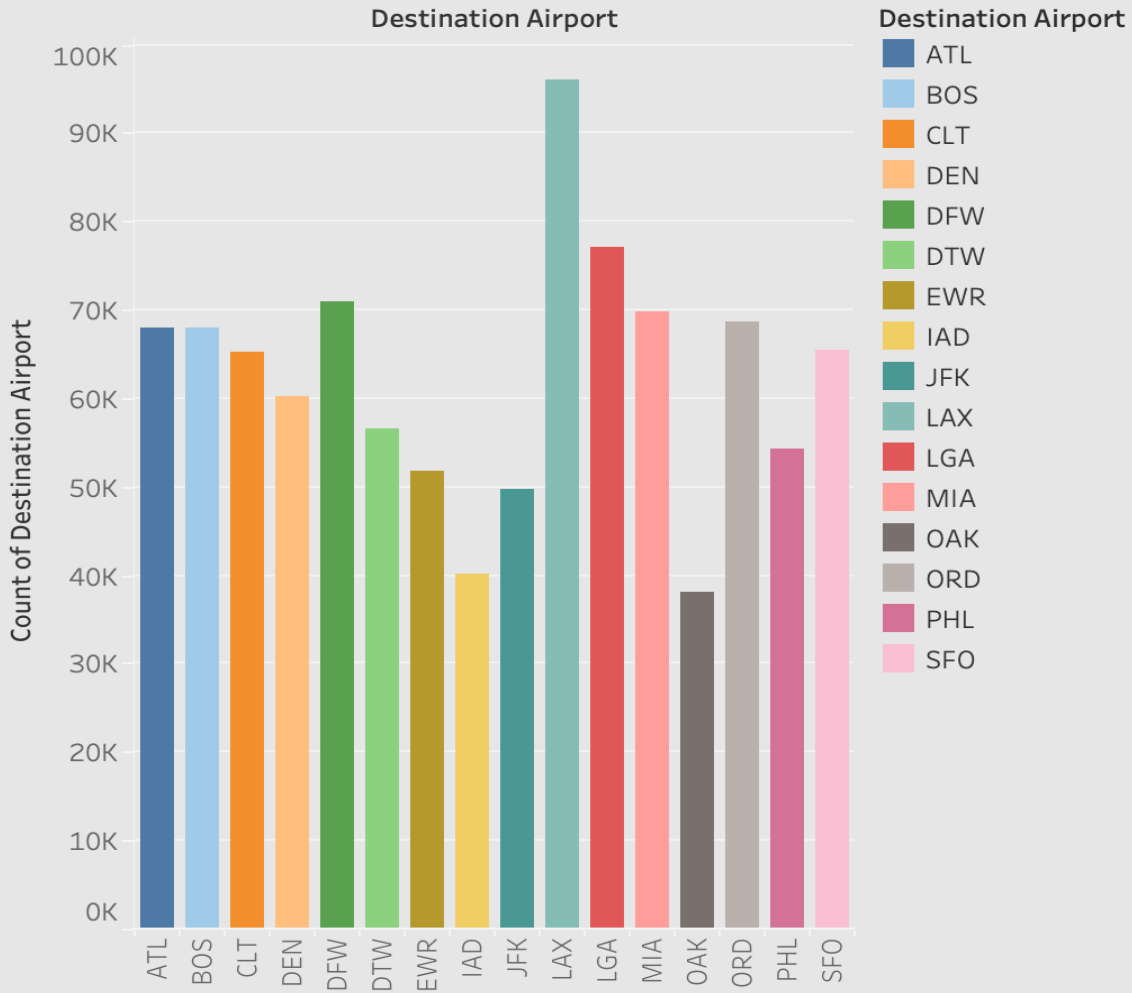
Bar Graph of Flights originating from Various Airports to understand the distribution.

Los Angeles International Airport sees the most traffic at around 100K flights, followed by LaGuardia Airport and Boston International Airport respectively on the given days.

Oakland Airport sees the least Traffic.

How are Flights distributed according to Destination Airports?

Flights Landing at Airports



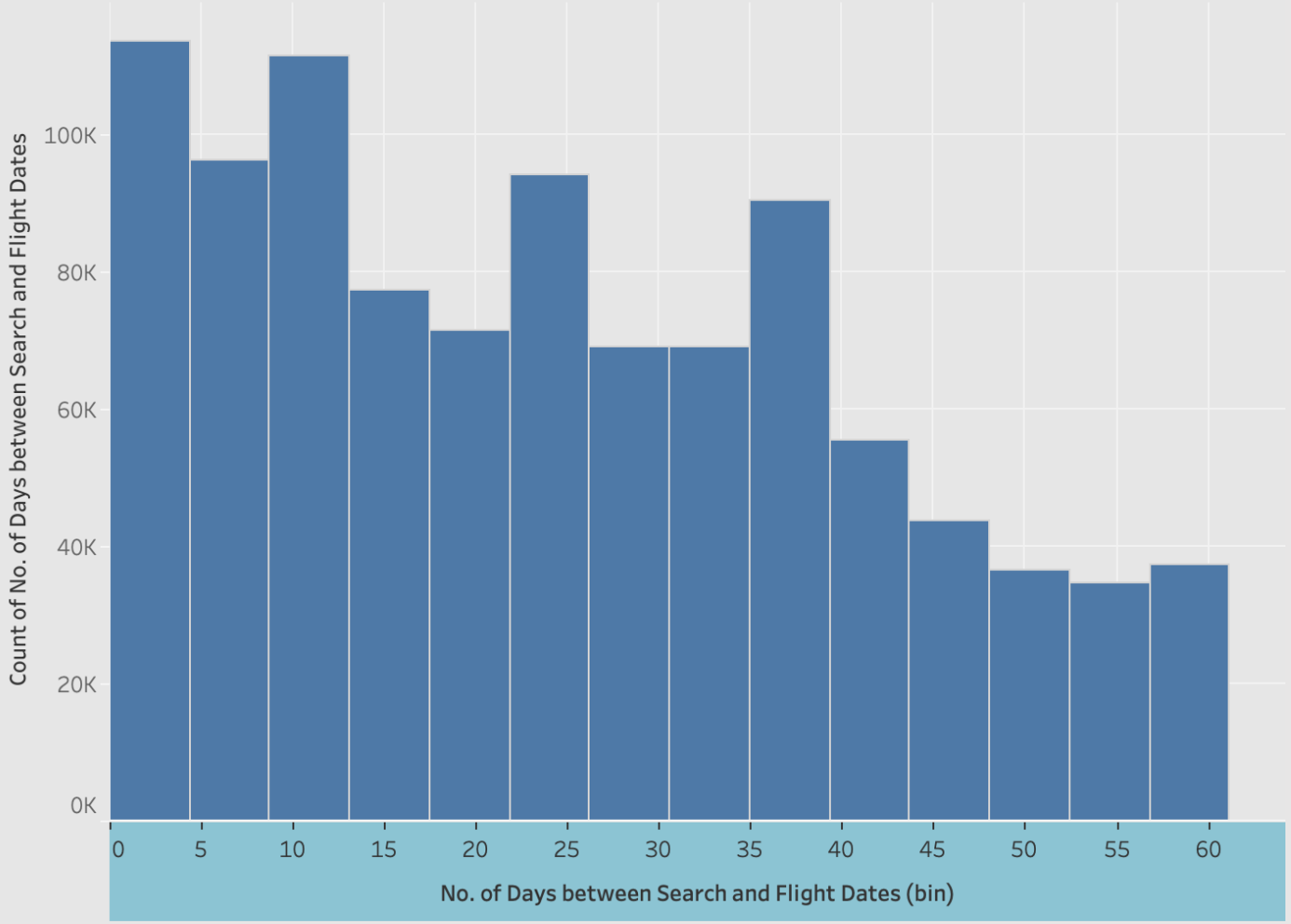
Bar Graph of Flights Landing at Various Airports to understand the distribution.

Los Angeles International Airport again sees the most traffic at around 95K flights, followed by LaGuardia Airport and Dallas Fort Worth Airport respectively on the given days.

Oakland Airport again sees the least Traffic.

How are Flights distributed according to Days left between Search and Flight Dates?

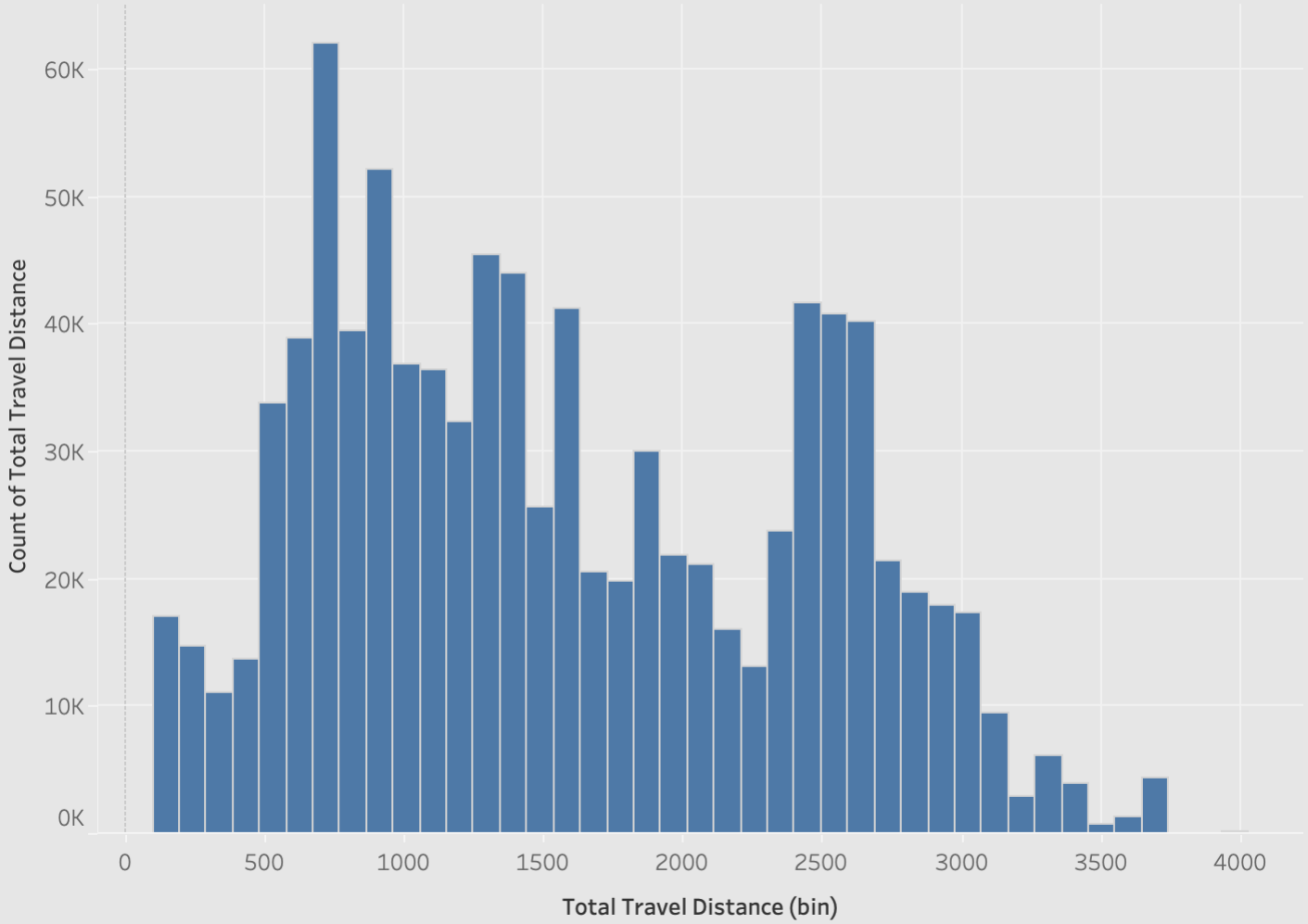
Distribution of the difference of days between Search and Flight Dates



Histogram of the difference in days of Search and Flight date to understand the distribution.

How are Flights distributed according to Distance Travelled?

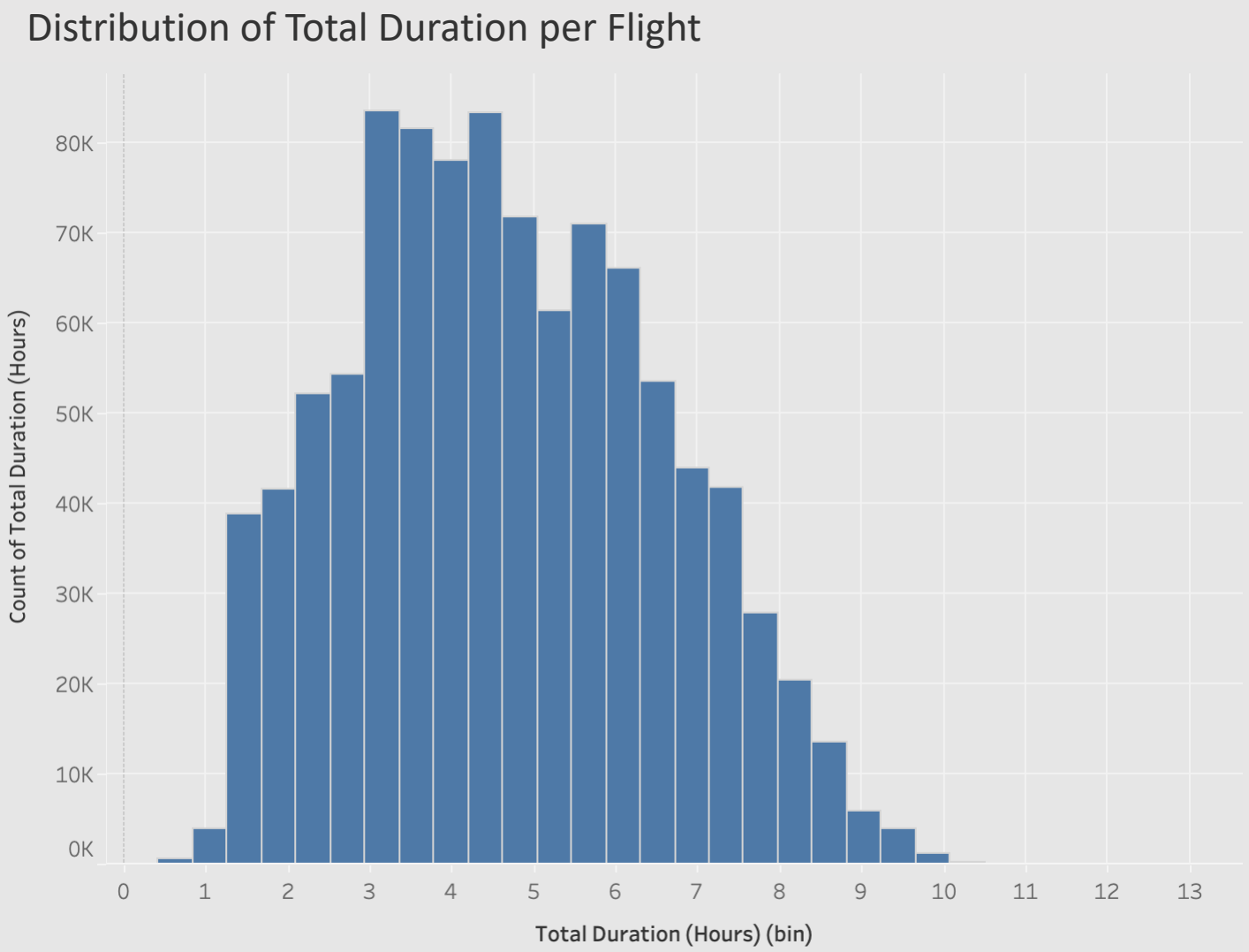
Distribution of Total Distance travelled per flight



Histogram of Total Distance Travelled per flight to understand the distribution.

Average Distance travelled seems to be around 1500 miles ranging till around 3700 miles

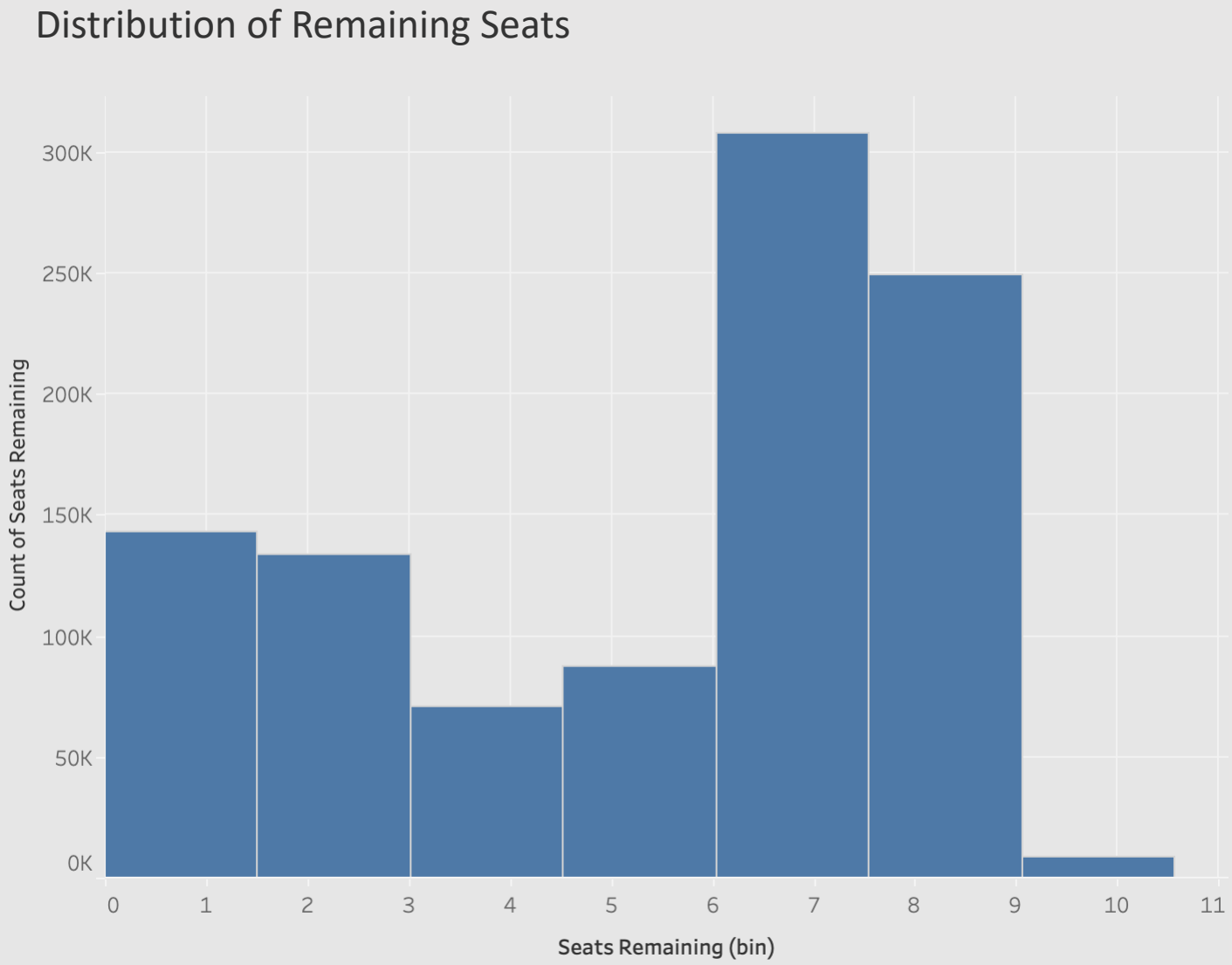
How are Flights distributed according to Total Duration of Flights?



Histogram of Total Duration of Flights in hours to understand the distribution.

Average time seems to be around 5 hours with range going up to 10 hours.

How are remaining seats per flights Distributed?

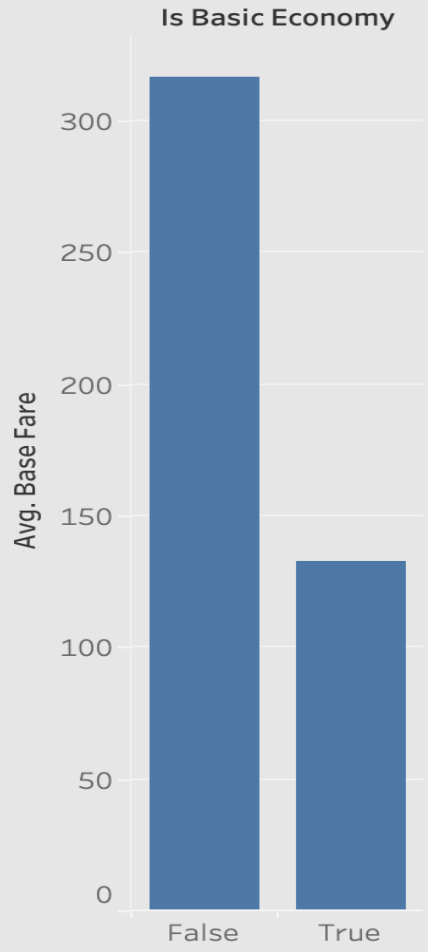


Histogram of Remaining seats to understand it's distribution.

Average empty seats seem to be around 5-6 with ranging up to 10.

What is the impact of flight class on Base Fare?

Average Base Fare According to the Class

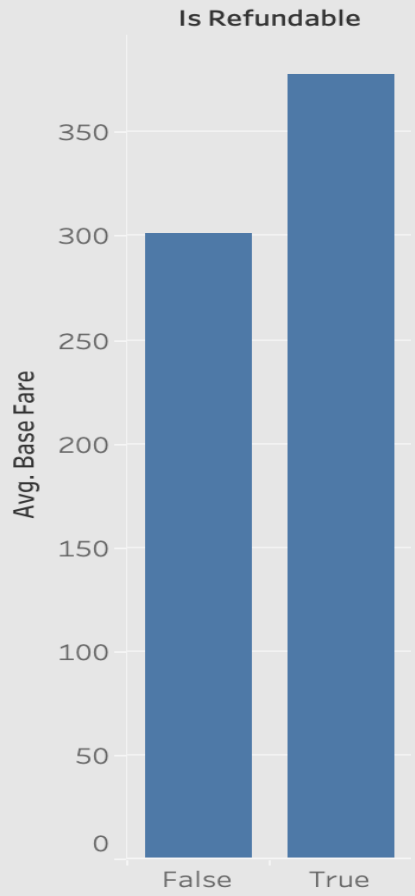


Comparison of Average Base Fare with the flight class to understand the relationship between them.

Not surprisingly, Average Base Fare is higher for Flights that are not basic Economy.

What is the impact of flight Refundability on Base Fare?

Average Base Fair according the Refundability of Flight

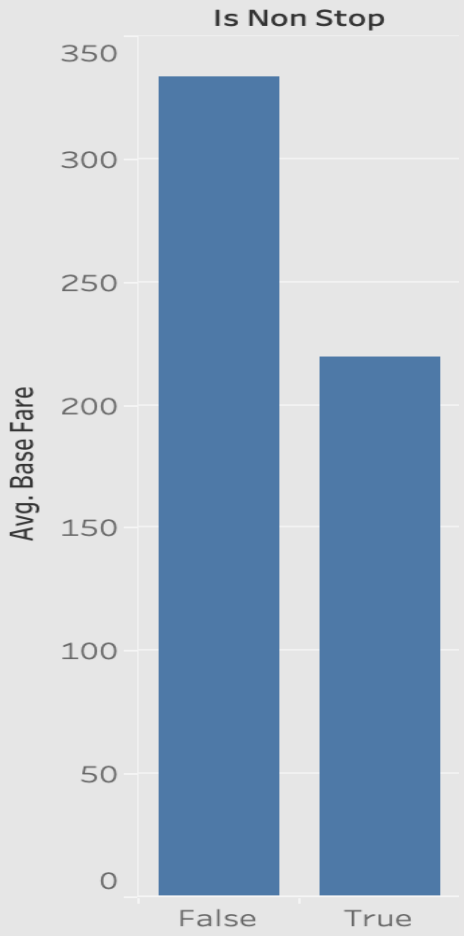


Comparison of Average Base Fare with the Refundability of flights to understand the relationship between them.

As expected, Average Base Fare is higher for Flights that are refundable.

What is the impact of stops in Flights on Base Fare?

Effect of Direct Route on Average Base Fare

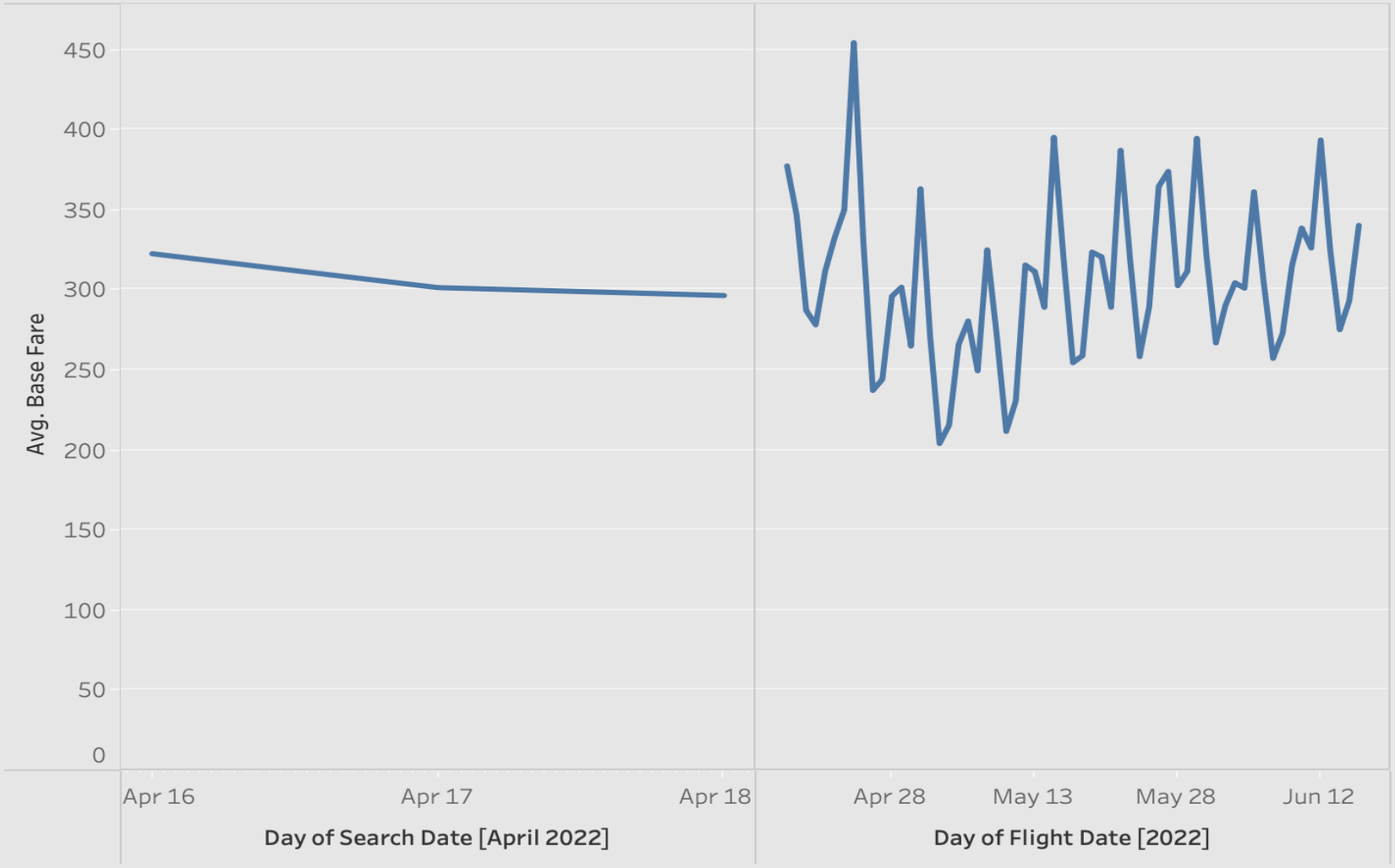


Bar Graph of Average Base Fare against the possibility of flight being non-stop to understand the effect a flight being non–stop has on Average Base Fare.

Average Base Fare is lower for Flights that are non-stop. This might be due to the fact the non-stop flights cost the airlines less if the distance is not high. Since, we considered Average Fare , this might skew it in favor of non-stop flights.

What is the impact of Search and Flight Dates on Base Fare?

Comparison of Average Base Fare with Search and Flight Dates



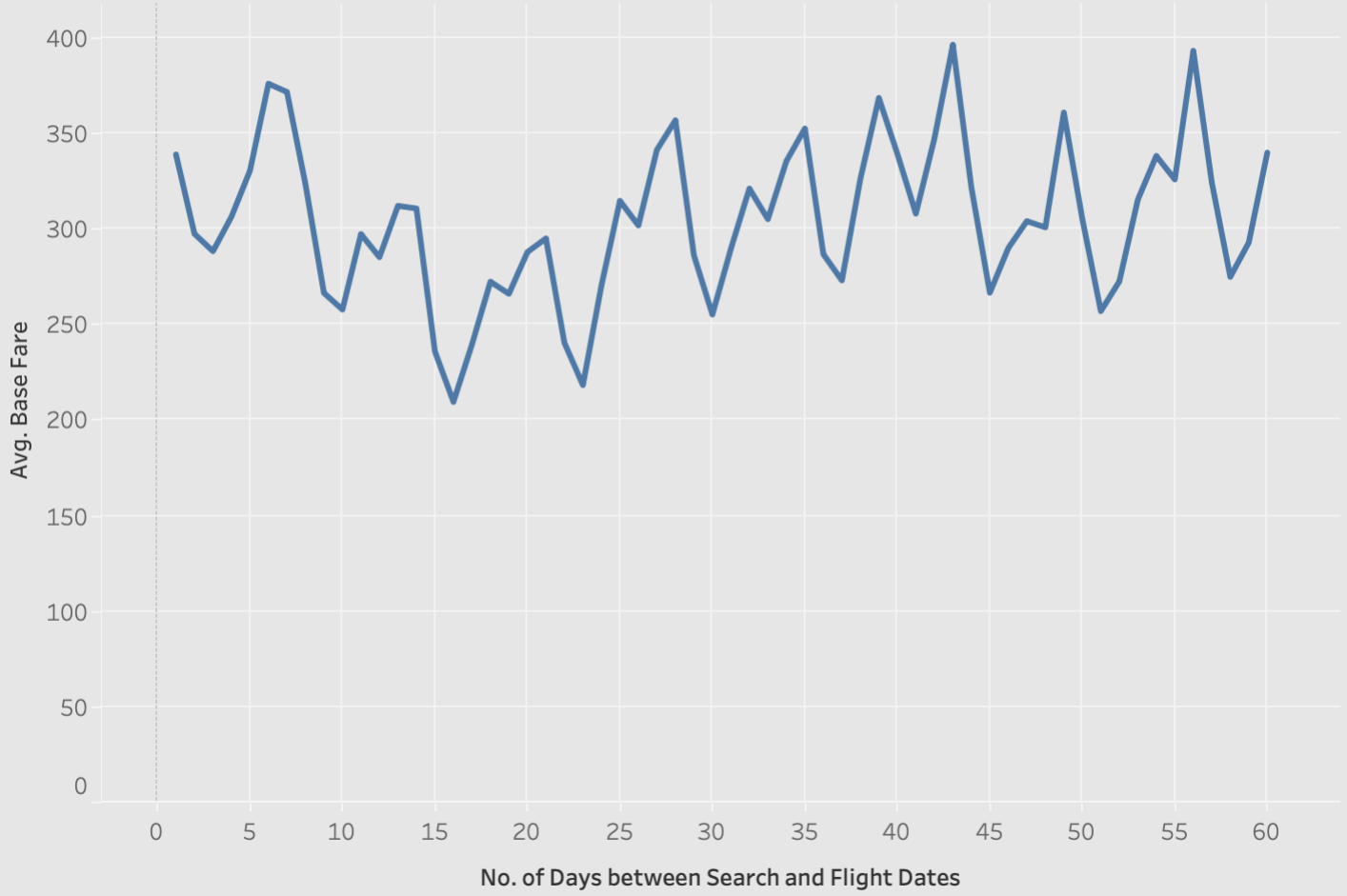
Line Graphs of Average Base Fare to compare the relationships with Search and Flight dates respectively.

There is almost no linear relationship between Search date and the Average Base Fare. This might be impacted by the fact that we only have this data for 3 days.

Similarly, There is almost no linear relationship between Flight date and the Average Base Fare. There is however a very clear surge in average Base Fare around the Easter holiday.

How does the difference in Search and Flight dates contribute to Base Fare?

Average Base Fare according to the difference of days between Search and Flight Dates

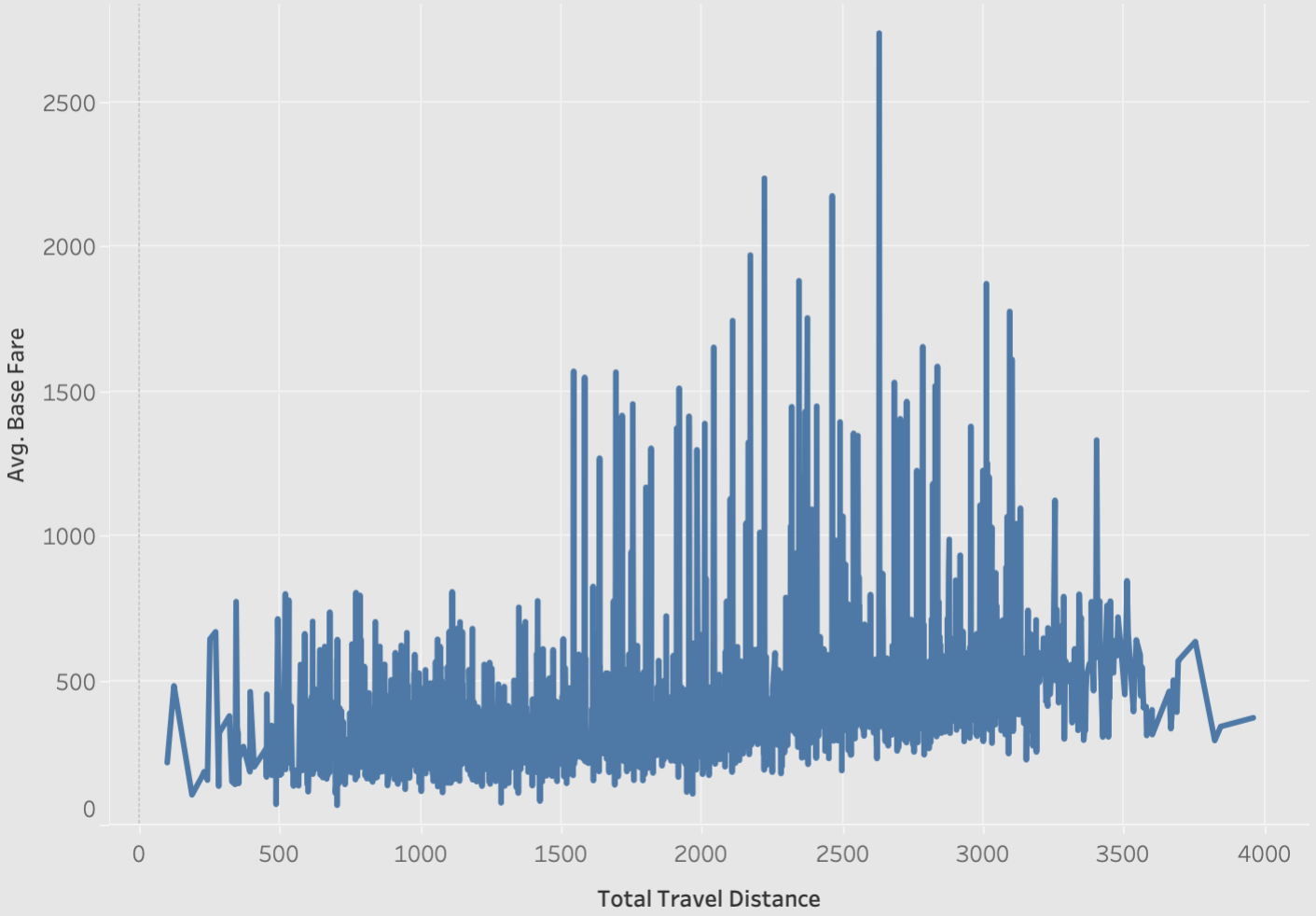


Line Graph of Average Base Fare to compare the relationship with no. of days between search and flight dates.

Surprisingly, there isn't a clear linear relationship between the two. Again, this might be explained by the limited availability of data

What is the impact of Total distance travelled in a flight on Base Fare?

Average Base Fare according to the Total distance covered per Flight

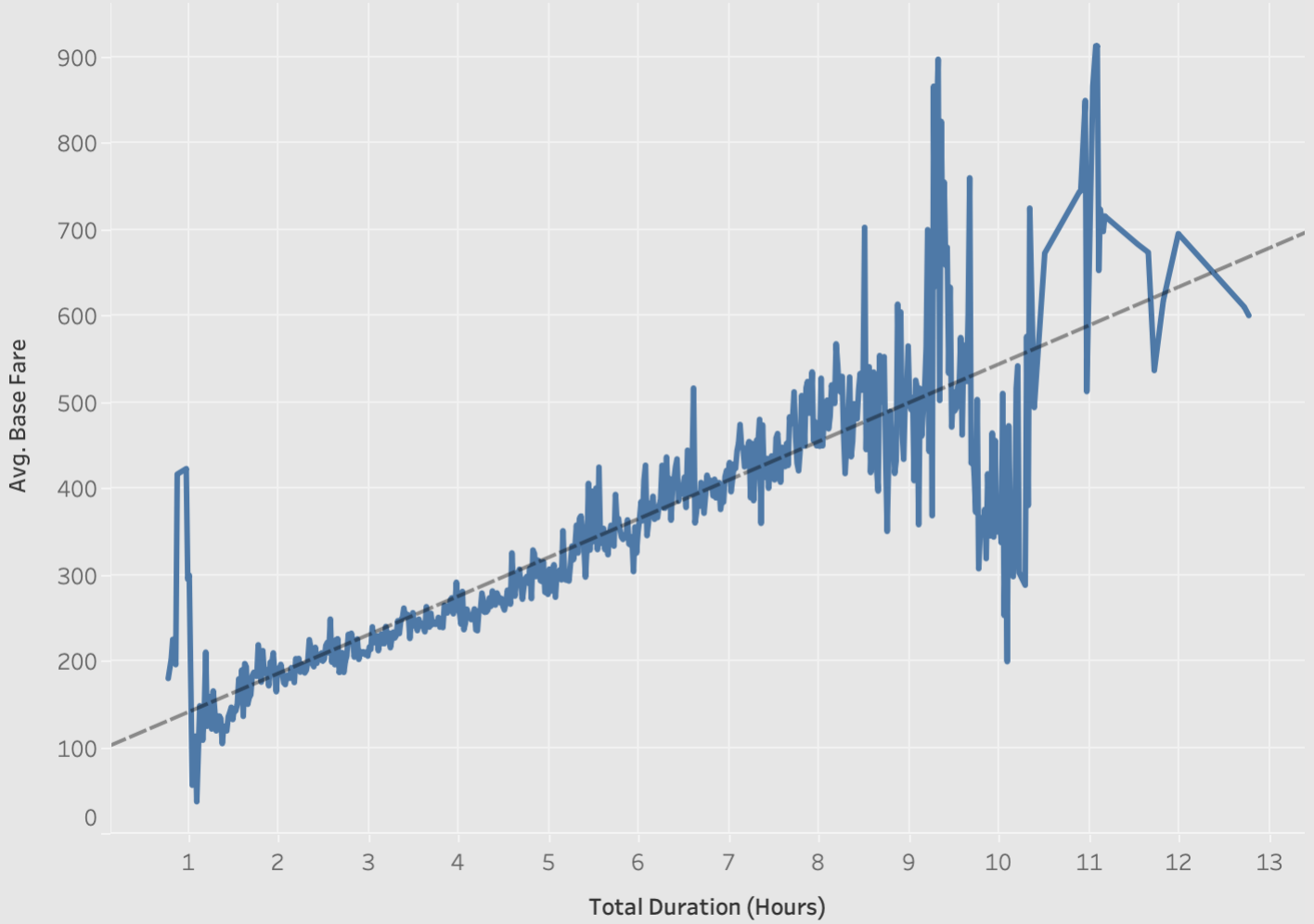


Line Graph of Average Base Fare with Total travelled distance to understand the relationship between the two.

There is a slightly positive linear relationship between the two.

What is the impact of Total flight duration on Base Fare?

Average Base Fare according to the Total Duration of Flights



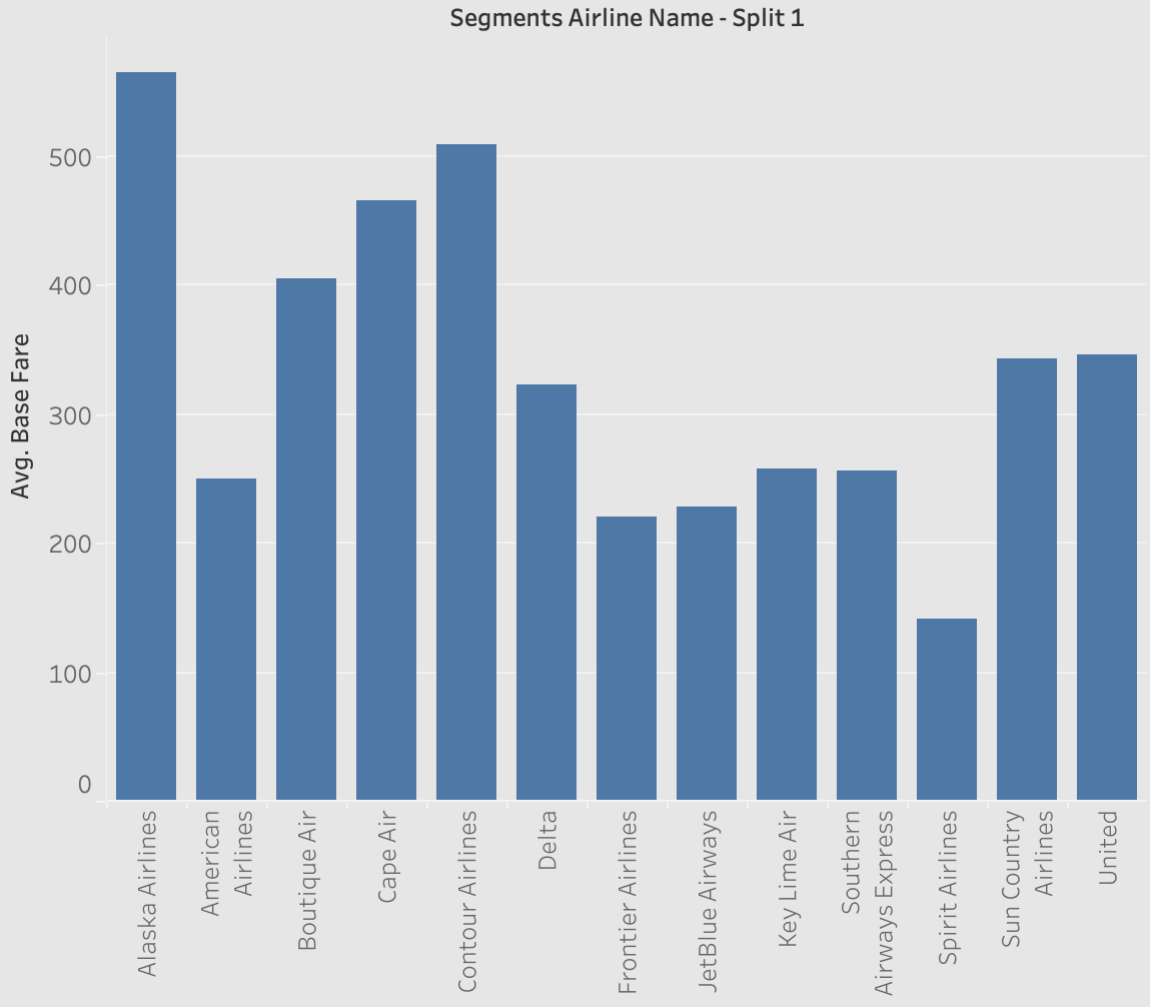
Line Graph of Average Base Fare with Total duration of the flight.

There is a clear linear relationship between the two.

As the time increases, the Average base fare increases too.

What is the relationship between Airlines and Base flight fares?

Average Base Fare per Airline

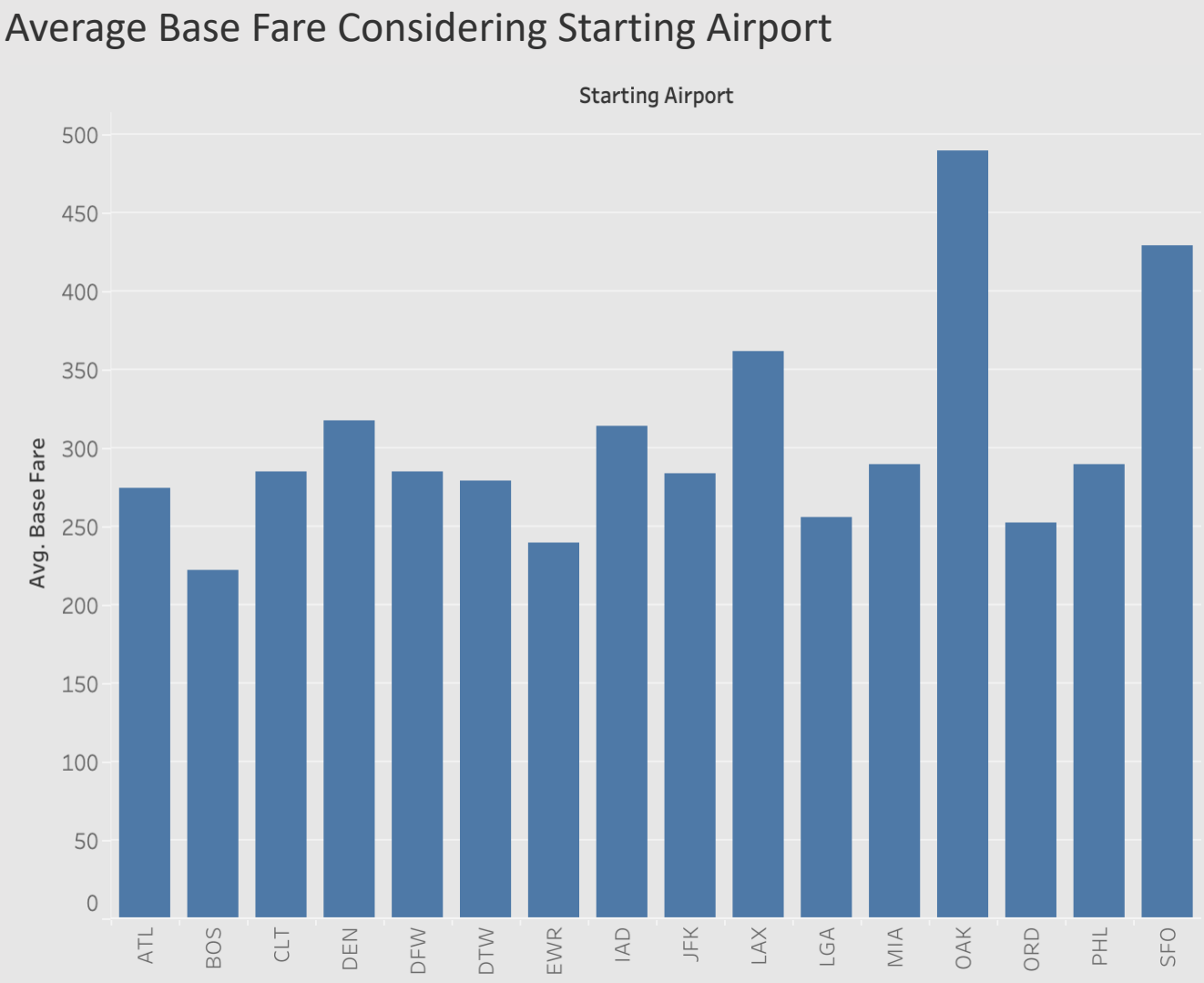


Bar Graph of Average Base Fare against the originating airlines to see the relationship between the two.

Alaska Airlines has the highest Average Base Fare which could be explained by both the distance and the rarity of the flights.

Spirit Airlines have the lowest Average base fare which could be explained by the low count of flights on the given dates.

What is the relationship between Starting Airport and Base flight fare?



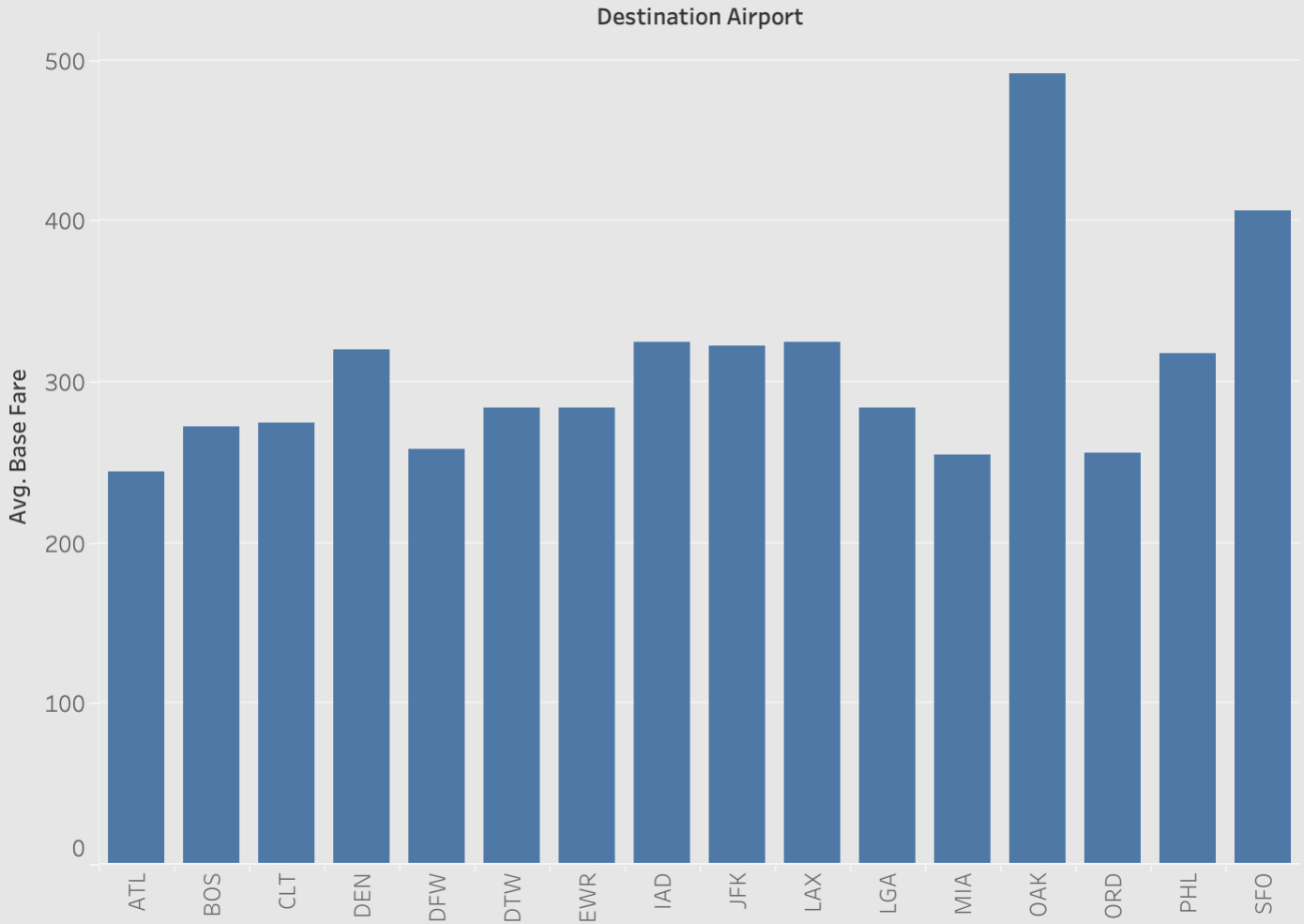
Bar Graph of Average Base Fare against the Starting Airports.

Oakland has the highest average base fare which seems to be indirectly proportional to the counts of both originating and landing flights at the airport.

Boston Airport has the lowest average base fare.

What is the relationship between Destination Airport and Base flight fare?

Average Base Fare Considering Destination Airports



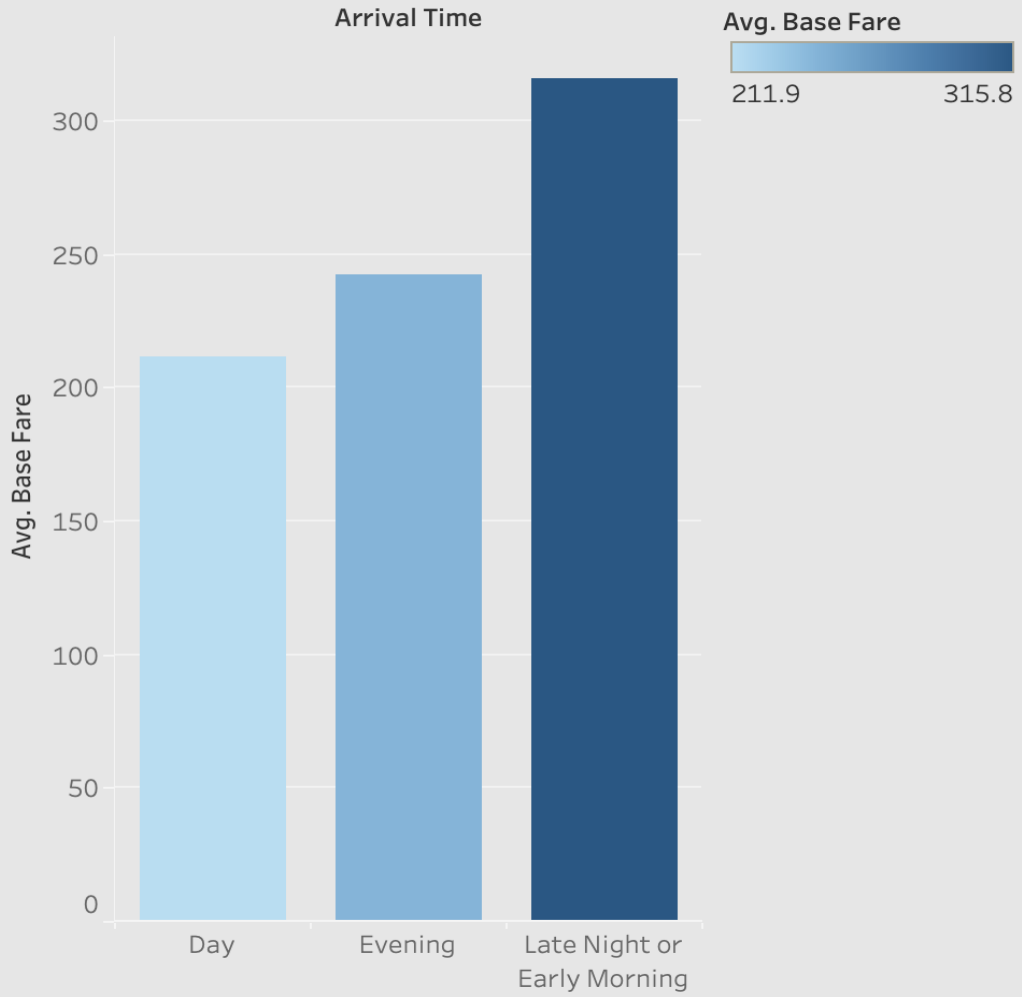
Bar Graph of Average Base Fare against the Destination Airports.

Oakland again has the highest average base fare which seems to be indirectly proportional to the counts of both the starting and landing flights.

Chicago O'Hare Airport has the lowest average base fare.

What is the relationship between Departure time of Flights and Base flight fare?

Average Base Fare according to Departure Time



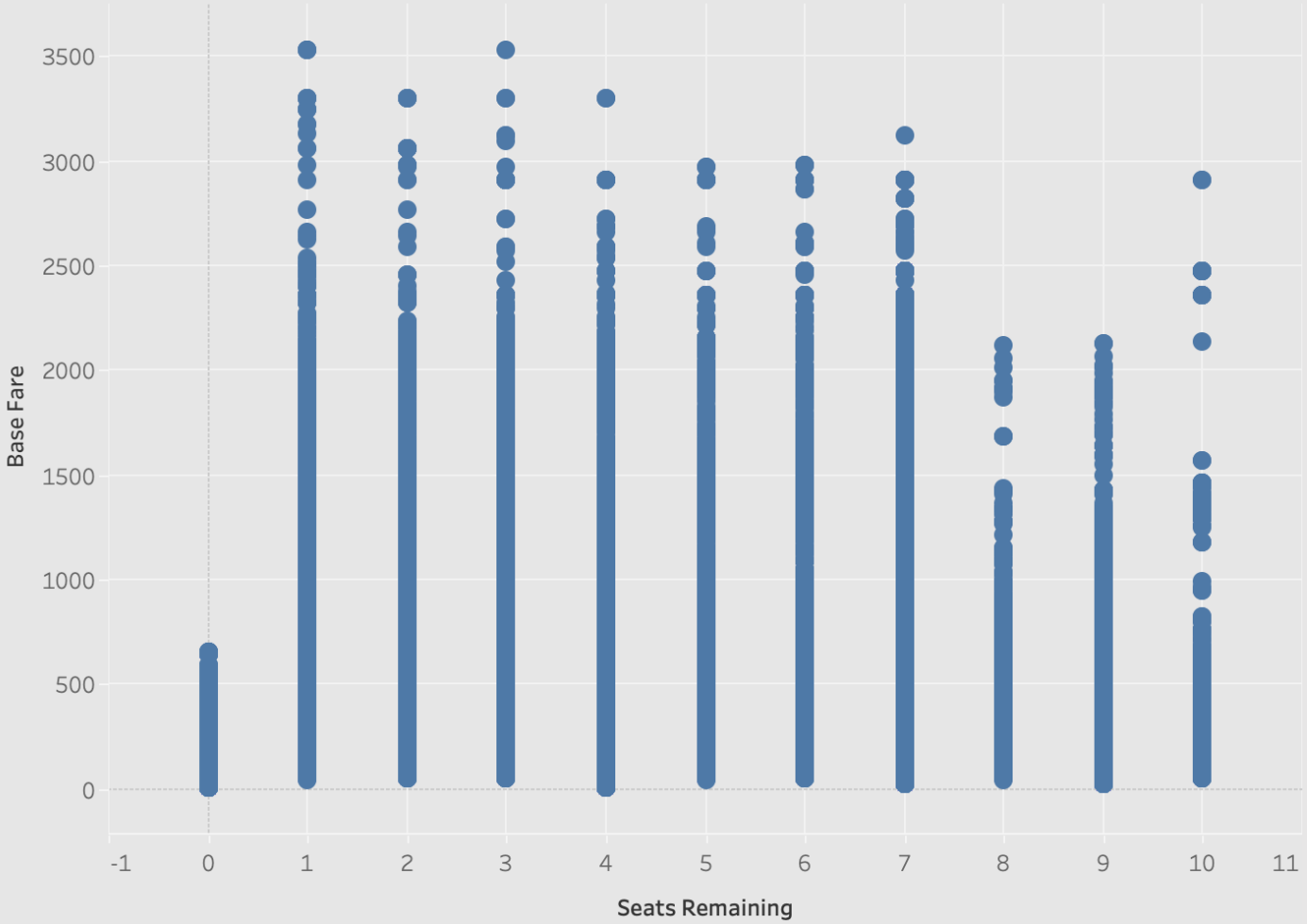
Bar Graph of Average Base Fare against the departure time of the day to understand the relationship between the two.

As expected, Late night or early morning flights (between midnight and 9 am) have the highest average base fare which makes sense since most people want to start and before the start of the days.

Hence, Flights during the day (9 am to 4 pm) have the lowest average base Fare.

What is the relationship between Remaining seats and Base flight fare?

Base Fare According to the No. of seats remaining

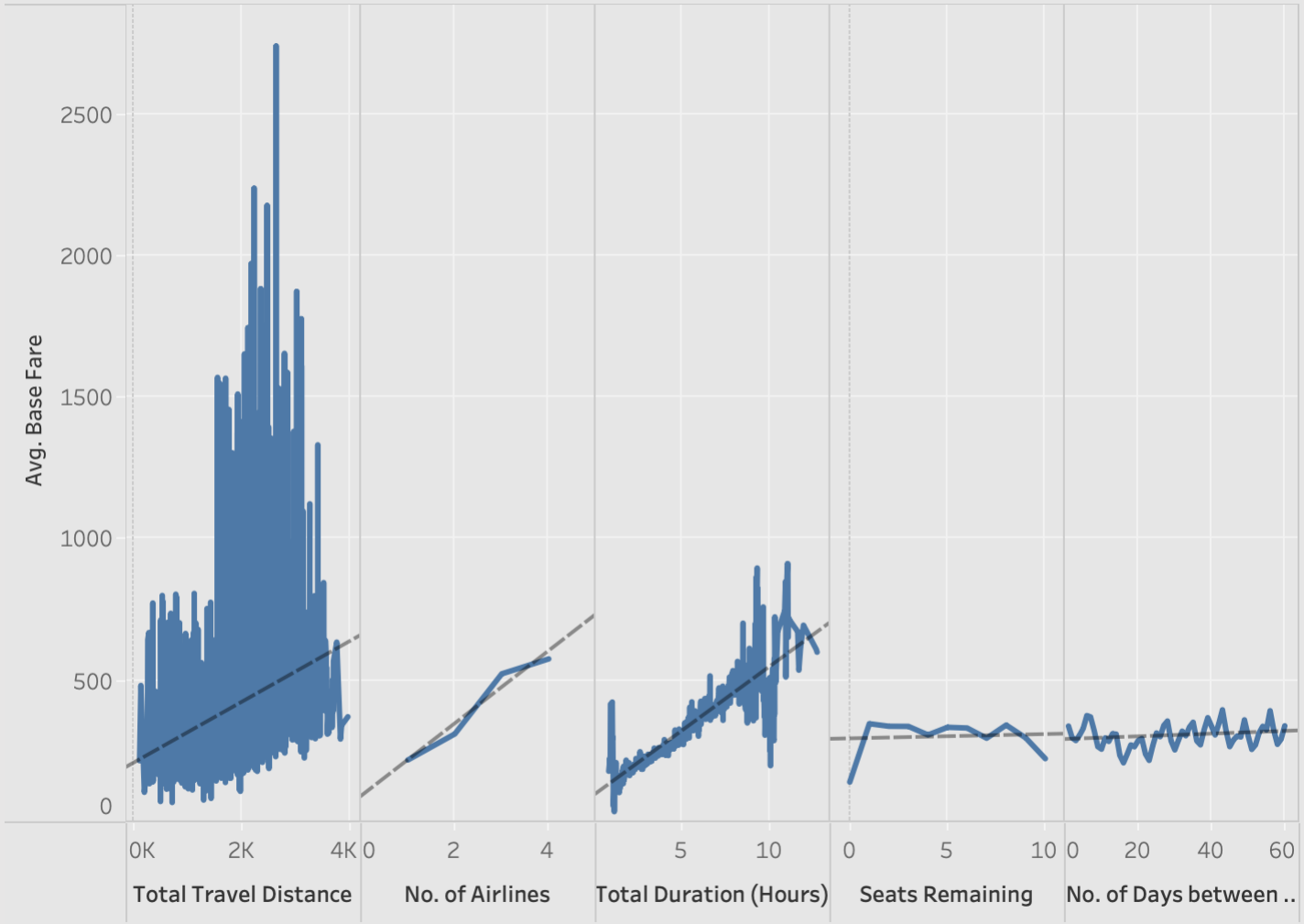


Graph to understand the relationship between remaining seats with Base Fare

There doesn't seem to be a direct relationship between the two. However, if there are no pending seats, the Base Fare seems to be significantly lower.

What trends can be seen in Base Fare?

Base Fare Trends



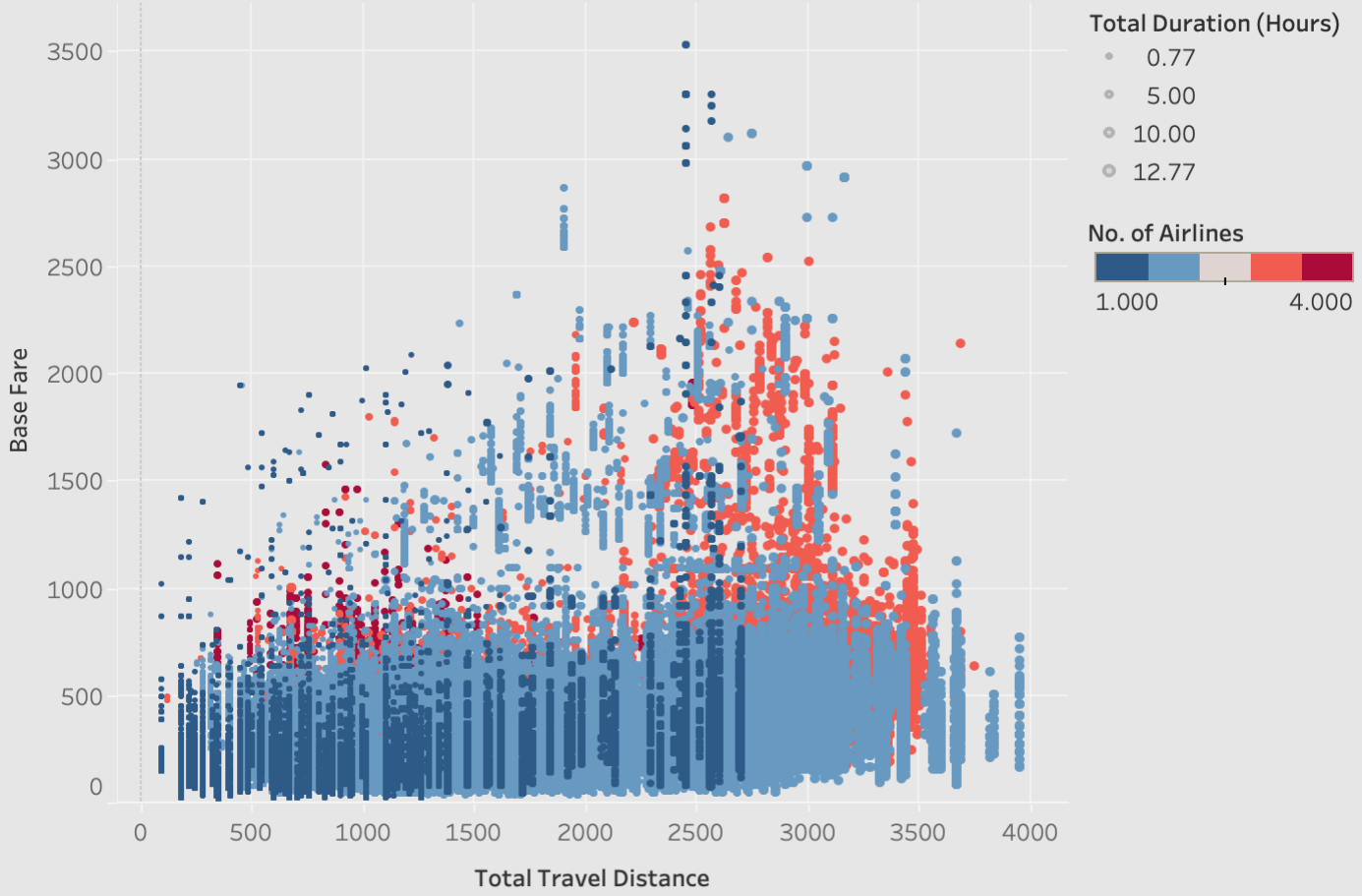
Line Graphs to understand Base fare trends in relation to total distance travelled by the flights, No. of airlines used in the flight, total duration of the flight, seats remaining in the flight and No. of days between search and flight dates respectively.

There are clear positive linear relationships between average base fare and total distance travelled, no. of airlines and total duration of the flights.

It is also evident that there are no relationships of Average Base fare with Seats remaining and the difference of days from search and flight dates.

How do total travel distance, total time taken and no. of airlines used in a flight collectively impact base fare?

Base Fare considering Total Distance Travelled, Total Time Taken and No. of airlines used per Flight



Plot to understand the impact of Total distance travelled, total time taken and total no. of airlines used in a flight on the Base Fare.

Non-stop flights with lower flight duration have the highest base fares for the same distance travelled.

Notable Observations

- Base fare has direct relationships with distance travelled, duration of the flights, and no. of airlines used.
- Non-stop flights with lower flight duration have the highest base fares for the same distance travelled.
- Flights starting and ending in late nights or early mornings have higher average base price.
- Surprisingly, there isn’t a clear linear relationship between the the difference of days in search and flight dates and the average base price.
- Again Surprisingly, Average Base Fare is lower for Flights that are non-stop.


THANK
YOU