NYC FLIGHTS 2013

TABLEAU ANALYSIS

NOÉMIE GAUTIER - ROHIT BHALERAO - LAURA EL AOUFIR

BUSINESS REPORTING TOOLS

Introduction

This data set illustrates all the flights leaving from the three New York Airports in 2013. With this data, we have been able to create multiple worksheets and dashboards illustrating the key information from the dataset.

I. Comparing Departure and Arrival Delays (dashboard 1)

A. Overview

One of the key focus of this data set is the delays of the flights. We therefore started by creating two pie charts that illustrate how many flights departed on time, compared to delayed departures. Similarly, we compared how many flights arrive on time or delayed. To do so we created new variables "Arr: Delayed or on time" based on the If-Then statement bellow and applied the same principle for departure.

```
If [Sched Arr Time]=[Arr Time]
  Then "On time"
  Else "Delayed"
```

B. Findings

These charts highlight that there are many more flights that depart or arrive delayed compared to the ones that are on time. Additionally, there is a smaller difference between on time arrivals and delayed arrivals then for departures. This illustrates that even amongst the flights that depart on time, only a few manage to arrive at the destination on time.

II. Delay by Airport (dashboard 2)

A. Delay by Origin

Another key output of this data is to compare the delays by airports. In order to do so we first created a bar chart illustrating the departure and arrival delays for each of the three New York airports from which the flights leave from. This chart illustrates that flights from the Newark Liberty International are the ones with the most delays for both departure and arrival. The departure delay for the other two airports are close. Moreover, this graph illustrates the strong correlation between departure delays and arrival delays. Indeed, the more departure delays the more arrival delays for that airport's flights.

B. Delay by Destination

We also created a similar bar graph for the destination airport. Here we can see that the destination airport with the most delay is the Columbia Metropolitan. We can also see that some airport delays are negative, meaning all flights arriving at this airport have arrived early. Some airports have departure delay but no arrival delay which also shows that the departure delay can be caught up during the flight.

To illustrate the delays by destination we also created a map, which represents the same data as the bar graph. This map allows us to see if flights that have the most delay are further or closer to the origin airport. On the map we can see that the airports that are the furthest from their departure airport are not the ones with the most delays. Furthermore, the closest airports do not have strong delays either. The destination airports with the most delays are spread throughout the USA. From this one we could say that distance does not necessarily correlate with delay. On the other hand, it could be argued that the flights that have the longest distance also have the most time to catch up any delay, and the flights that are the closest won't have this opportunity.

It should also be noted that there are not the same number of flights for each destination. This means that some of these bars are the reflection of 1 or less than 10 flights and other reflect an average over thousands of flights for this destination.

III. Delays Over Time (dashboard 3)

A. By month

We created a line graph to illustrate the delays through the year 2013. The graph highlights that in June and July the delays are at their peak. This could be due to the summer holidays, as there are more people traveling and therefore more flights, this could create some traffic and delays for flights. Following this peak there is the least delays in August and September when US citizens are traveling less as school has started again for children. We created a calculated field called Total Delay which is simply the addition of Arrival and Departure delays for each flight. We use this calculated field in many of the following sheets.

[Arr Delay] + [Dep Delay]

B. By hour

We created a similar line graph for the hours of the day. This chart highlights that the delays keep increasing during the day. Indeed, the more delay a flight has the more it will impact the next flight from the same airplane. Therefore, the delay cumulates until the end of the day where it is peaking. Morning flights generally leave on time and arrive early, but around 12h they start arriving late, meaning they are not able to catch up their departure delays during the flight anymore. As the day goes by delays increase. It should be noted that there is a decrease in delays at 22h but then delays rise again at 23h.

IV. Weather (dashboard 4)

A. Visibility

We created a scatter plot for the average total delay for each value of visibility. A clear inverse correlation between the two variables is observed. As visibility increases, delays in flights decrease. This is because when the visibility is low, flights remain grounded.

B. Precipitation

Again, a scatter plot is used to show average delay of flights for each airport. We filter the plot by airports to know how precipitation affects each airport individually. Newark and LaGuardia airports see more precipitation than JFK, hence flights from these airports are delayed when high precipitation is recorded.

C. Humidity

There is no strong correlation between humidity and flight delays. Flight delays are irregular for the humidity scale. Humidity does not really affect flight schedule a lot.

D. Wind Speed

We used a line plot to show the change in average delay and wind speed. The plot follows a trend where when the windspeed increases, delays in flights increase.

E. Pressure

We used a scatter plot to show the average delay for various values for pressure. Although the correlation is not strong, we can see that low pressure causes flights to delay. Hence, when the pressure is high, flights tend to depart early, sometimes before schedule.

F. Temperature

The line plot of temperature with average delay is parabolic. Temperature when plotted against the average Low (< 20 °C) and extreme high (>80 °C) temperatures cause flights to delay, while the delay is less for moderate temperatures between 20 °C and 80 °C.

V. Comparison from Airlines (dashboard 5)

A. Overview

We plotted a line graph and a bar graph in the same sheet on similar axes. The plots are not coherent, showing that some airlines have a higher arrival delay while some have a higher departure delay. This could be because different airlines depart from different airports in New York. This could cause the flights to delay due to weather conditions (precipitation) as seen in previous dashboards. On the other hand, arrival delay could be a result of the planes used by different airlines among other reasons. A bar chart represents the departure per airport for each Airline. We also plot the number and percentage of cancelled flights for each airline.

B. Findings

There is no relation between the arrival and departure delay of different airlines. Similarly, there is no correlation between delay and the number of flights for the airlines. The plots help analyse each airline individually. We can see that Frontier Airlines has the highest arrival and departure delay. US Airways has the lowest departure delay while American Airlines has the lowest arrival delay. Among larger airlines with many flights, Delta Airways sees comparatively lesser delays.

VI. Delay by Airplanes (dashboard 6)

A. Overview

We plotted the engine type, manufacturer and the different models with the average delays.

B. Findings

Most flights run on Turbo-fan engine while 4 cycle engines see the highest average delay. Planes manufactured by Augusta SPA have highest delay. Among the leading airplane manufacturers (Airbus, Boeing and Embraer), Embraer has highest average delay while Airbus Industries performs slightly better than Boeing.

VII. Routes (dashboard 7)

A. Overview

We plot the routes from each of the three New York airports to the various destinations in the US. We also plot the delays and the number of flights to each of the various destinations. Then, we plot the average delays for departure and arrival for each of the flight routes.

We created a calculated field called ‘Route’ which uses the function MAKELINE() with variables origin and destination. For the origin, we have used only one common origin as all flights are departing from New York City. For the ‘Route’, we use variable ‘Destination Point’ which uses function MAKEPOINT() with arguments Latitude and Longitude of destination airports. Another variable the ‘Route’ calculated field needs is the Origin Point, which are the coordinates for New York, the origin city.

Another calculated field, ‘Route Details’ is just the string concatenation of the Origin and the Destination strings to get the unique route name for each route between origin and destination airports.

B. Findings

We can see the plot of routes showing a lot of cross-country as well as shorter flights from the different airports. We see larger circles for San Francisco, Los Angeles, Atlanta and Orlando Airports among a few others. These airports have many flights from NYC. The red circles show a lower average delay (ex. Los Angeles) while the dark blue circles depict a higher average delay. Destinations closer to New York have highest average delays. Destinations closer to west and east coast have moderate and lower delays.

As the distance of destination from New York increases, we can say that the departure delay decreases. The same is not true for Arrival delays.

The best routes are:

- LGA – Dallas
- LGA – Louiseville
- LGA – Dayton
- EWR - South Bay Indiana
- EWR - John Wayna Aiport California

The worst routes are:

- EWR – Columbia
- EWR - Alcoa Tennesse
- EWR – Oklahama
- EWR – Tulsa
- JFK - Colorado

Conclusion

Overall, this data set allowed us to create multiple graphs, dashboards and a story analysing the flights leaving New York in 2013. Our Tableau analysis lets us understand various reasons for flight delays considering different criteria such as distance, weather, time of the day, airport, airline and so on. To conclude, there are different reasons why flights face delays, but with our data we have an idea when these delays can happen. This could allow us to predict potential delays and deal with them in the best possible way. Indeed, we can see that different periods of the year or of the day have higher delays than others. Therefore, for these periods maybe more planes could be made available to reduce the cumulative delay. Moreover, weather is a key external factor of delay. It is harder to predict the exact weather, but we can have some predictions which could allow the adaptation of flight schedules. Our analysis could therefore bring some predictions and help NY airports limit their flight delays.