NYC Flights Data

Prepared by:

Hina HUSSAIN

Elina KELLY  
Mui Han MA

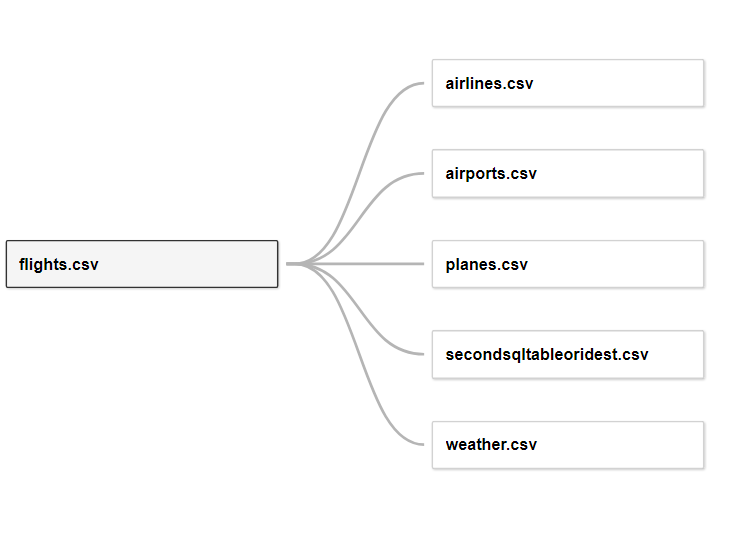
This document provides a technical overview of the publicly available Tableau Workbook “ProjectTableauFinal” found at <https://public.tableau.com/views/ProjectTableauFinal_16344909833850/Group_Project_Story?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link>

The workbook uses the following tables:

|  |  |
| --- | --- |
| **CSV File Name** | **Source** |
| Airlines | Nycflights13 |
| Airports | Nycflights13 |
| Flights | Nycflights13 |
| Planes | Nycflights13 |
| Weather | Nycflights13 |
| SQLOriginDestinationTable | Created on SQL using data from Flights & Airports |

The tables were connected in Tableau using the following connections:

* carrier in flights = carrier in airlines
* origin in flights = faa in airports
* talinum in flights = talinum in planes
* origin in flights = origin in weather
* timehour in flights = timehour in weather
* flights in flights = flights in SQLOriginDestinationTable (secondsqltableoridest)



The sections below provide the description of each slide of the story

[Maps for Average Delay per Route 2](#_Toc85396369)

[The Best Airlines 3](#_Toc85396370)

[Distance matters…? 4](#_Toc85396371)

[Ohh…Weather 4](#_Toc85396372)

[And…Seasonal Factors! 5](#_Toc85396373)

[Book a ticket for Saturdays in September! 6](#_Toc85396374)

# Maps for Average Delay per Route

Background

A new table was created in SQL using the following query:

CREATE Table OriginDestination as

SELECT o.flight, o.origin, o.dest, o.dep\_delay, o.originlat, o.originlon, case when o.dest = o.dest then ar.lat end as destlat, case when o.dest = o.dest then ar.lon end as destlon

FROM (SELECT f.flight, f.dep\_delay, f.origin, f.dest, case when f.origin = f.origin then a.lat end as originlat, case when f.origin = f.origin then a.lon end as originlon

FROM airports a, flights f

WHERE a.faa = f.origin) o LEFT JOIN airports ar ON o.dest = ar.faa;

This table provides separate columns for origin latitude, origin longitude, destination latitude, destination longitude, and departure delay for each flight. Using the flight column, this table was linked to the Flights table part of the main dataset in Tableau.

Origin and destination airports were defined in calculated fields by using the following queries:

MAKEPOINT([Destlat], [Destlon])

MAKEPOINT([Originlat], [Originlon])

These points were used to make the lines for each route using the following query:

MAKELINE([Origin Airport Points], [Destination Airport Points])

The map was then created by dragging the routes to details, origin and destination airport names to labels, and average departure delay to colors. The origin airport and average departure delay are also available as a filter.

Analysis

The map displays a darker line for the routes with a high average departure delay and a lighter line for the ones with a shorter or no departure delay. It can be seen that the departure delay has no direct correlation with the distance of the flight.

The filter quickly allows the user to see the worst and best routes.

# The Best Airlines

Background

This diagram plots the position of the airlines on the average departure delays and arrival delays.

Average departure time (in minutes) is placed on the x-axis while the average arrival time (in minutes) is placed on the y-axis.

The followings are the extra calculation fields:

Average time (in minute) of arrival delays = SUM([Arr Delay])/COUNT([Flight])

Average time (in minute) of departure delays = SUM([Dep Delay])/COUNT([Flight])

Analysis

The diagram shows that, Hawaiian Airlines Inc, Alaska Airlines Inc. and US Airways Inc. are the top 3 airlines with the less time being delayed. Customers can also have an expectation from the diagram to know how long they should expect for the delays.

# Distance matters…?

Background

To see the link between the delay and the distance, the destination name from the flights table was placed as a dimension in the columns of the graph.

To display the number of delays per airport, the count of delays and the distance flown between each origin and destination airport from the flights table was included in the rows of the graph.

The count of delays was an additional calculated field:

count(if [Dep Delay] > 0 then [Dep Delay] END)

Analysis

Although people normally have assumptions on the distance of flight to the arrival delay, this graph provides evidence that the delays have no link with the distance between the origin and destination airports.

# Ohh…Weather

Background

To find out the correlation between the weather and delays, 4 graphs were built.

For the Precipitation, Visibility, Wind Speed, and Temperature graphs, all the rows consisted of the average departure delay from the flights table. As for the columns, the precipitation, visibility, wind speed and temperature measures from the weather table were selected respectively for the columns for each graph.

As a last step, a linear trend line was added to facilitate the visualisation.

Analysis

The Precipitation and Wind Speed graphs demonstrate that the more it rains or the windier it is, the delay increases. This could be interpreted by a logical reason which would be that this could cause a challenging take off or landing.

The Visibility graph indicates that whenever there is a considerable drop in visibility, in other words, the less the pilot can see ahead of him/her in miles, the delay increases.

As for the Temperature graph, an article from Forbes (<https://www.google.fr/amp/s/www.forbes.com/sites/lealane/2017/06/22/extremely-hot-temperatures-mean-bumpier-plane-rides-more-flight-delays-for-travelers/amp/>) states that “hot temperatures mean bumpier plane rides… as temperatures increase, air density decreases”.

This explains why the increase in temperature in the graph causes an increase in delay as pilots and airplane companies would prioritise the customers’ safety.

# And…Seasonal Factors!

Background

The two line graphs show the trend of the number departure delays and the average departure delays (in minutes) by month.

The followings are the extra calculation fields:

For Number of Departure Delays:

Date = Makedate([Year],[Month],[Day])

Number of departure delays = COUNT(IF [Dep Delay]>0 THEN [Dep Delay] END)

For Average Departure Delays (in minutes):

Date = Makedate([Year],[Month],[Day])

Average time (in minute) of departure delays = SUM([Dep Delay])/COUNT([Flight])

Analysis

The graphs show that there are seasonal factors for the number of delays and the average time of delay in different months. For instance, July and August are the summer vacation period and December are the Christmas holiday, the graphs show that there are more number of delays and the average time of delay in these period as more people will spend their holiday travelling.

# Book a ticket for Saturdays in September!

Background

Both bar charts show the chance of getting departure delays by days and by month.

The followings are the extra calculation fields:

For Chance of Departure Delays (Day):

Weekday = DATENAME("weekday",[Date])

Chance of departure delays = [Number of departure delays]/COUNT([Flight])

For Chance of Departure Delays (Month):

Date = Makedate([Year],[Month],[Day])

Chance of departure delays = [Number of departure delays]/COUNT([Flight])

Analysis

The bar charts show that we have a lower chance of having delays in September and on Saturday. One possible reason is that people love to travel on long holidays and take one or two-day leaves before the weekend starts to avoid the crowd. And for people having a business trip, they usually go back to their home on Thursday or Friday so that they can enjoy the weekend. Therefore, it has a higher traffic on Thursdays and Fridays while less on Saturdays, and thus higher chance of getting departure delays on Thursday and Fridays and less on Saturdays.

For the months, as the school term starts in September, students are back to school and the families already have their vacations during summer holidays, and thus less chance to have a leisure trip when the school term starts. Therefore, it is normal to have a slack season for the airline industry from September to November, thus less traffic in the airport and lower the chance for having a departure delay.

Due to above observations, it is suggested that customers buy air-tickets for Saturdays in September.