

Flight Data Analysis

by Mihai Chiva

Investigation Overview

The main goal of the analysis will be to identify trends and correlations in the data, factors that increase the risk of cancellations or delays, so as to be able to provide recommendations to potential flyers on how to plan their travelling to minimize the risk of issues.

The focus will fall on how each airline performs in terms of cancellations or delays, what airports are most impacted by these issues and in which period of the year. In addition, there will also be a top of flight routes which consistently report problems.

Dataset Overview

The dataset which will be explored here contains flight information from 2008 from the US. The data is split per flight and besides generic info on the carrier and origin/destination, it is focused on cancellations and delays. In this regards it provides details on the reason for cancellations as well as delay times split by category.

In addition to the main information, two additional sets of data were used, containing extra details on the carriers (e.g. full name) and the airports (e.g. name, geographical location)

Airlines With Highest Risk of Cancellations

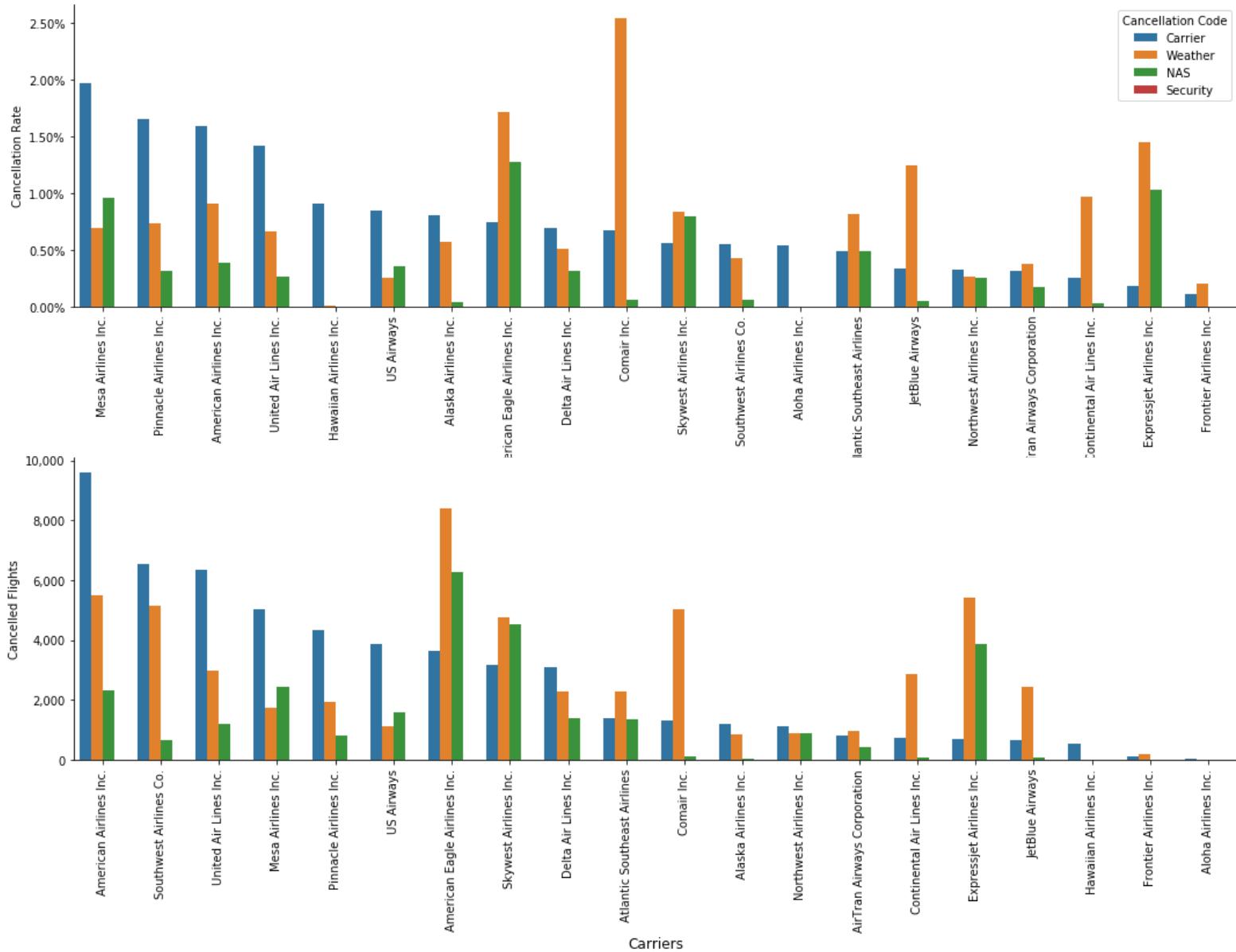
Objectives

A grouped barplot will be used to display for each airline the ratio/number of cancellations. The airlines will be ordered based on the values specific to carrier cancellations and not the total values. The full airline names will be used in the x-axis to ensure they are easily identifiable. Ratio and absolute amounts will be displayed in order to put into light any biases from the number of flights associated to a carrier

Findings

For most of the airlines, the risk of a cancelled flight due to the carrier's own fault, stands around 1%. There are four exceptions, out of which Mesa Airlines stands out due to a combination of both high number and percentage of cancellations, in the context of a low number of flights. Two other carriers which can be noted are American Airlines and United which are also in the top 4 on both types of measurements. But in their case, they also record a large volume of flights which partly explains the number of cancellations. One positive example is the low cost carrier Southwest Airlines, which has a cancellation rate of less than half of American Airlines while having double the number of flights.

Flight Cancellations Per Category



Airports With Highest Risk of Cancellations

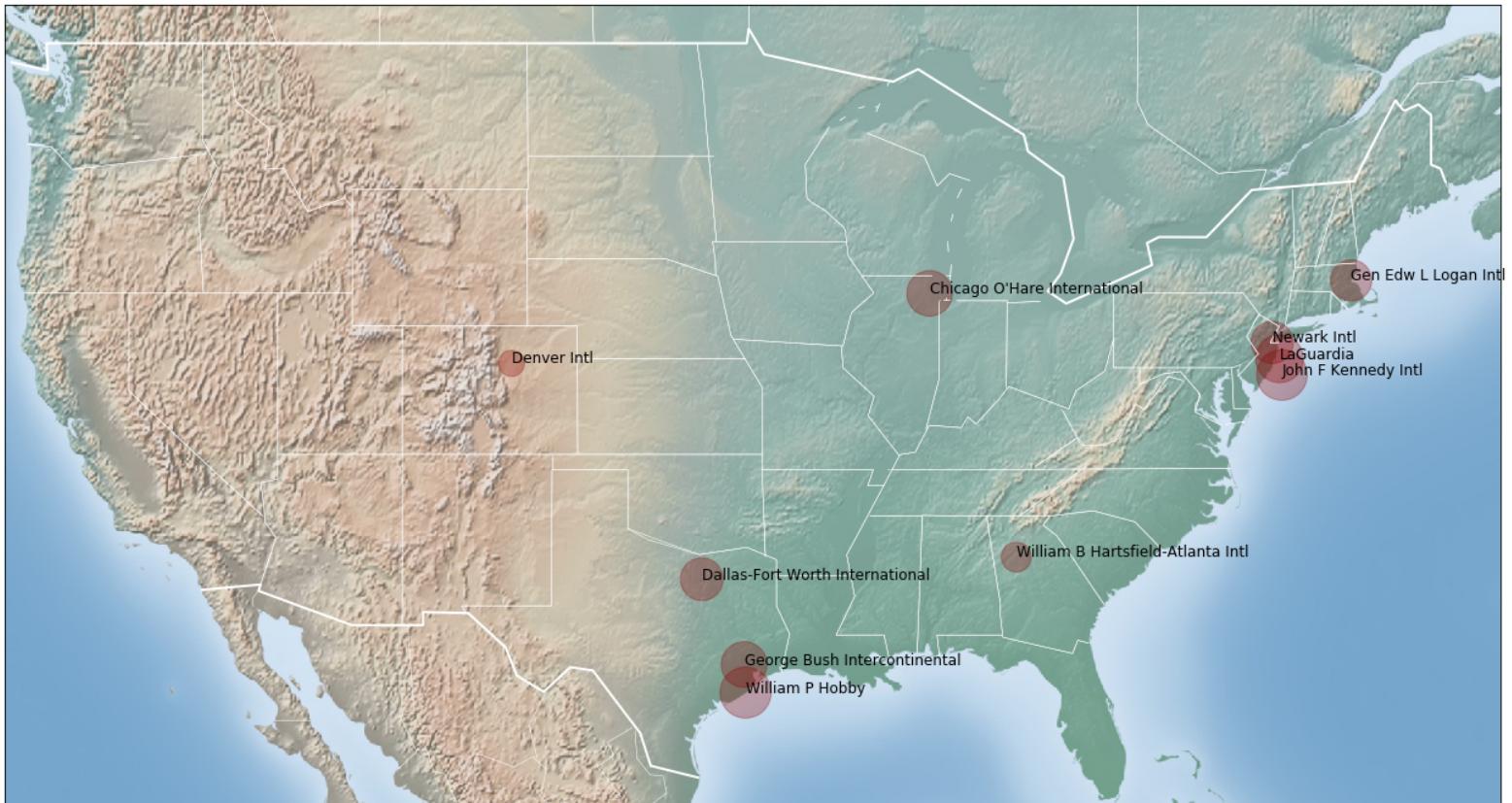
Objectives

In order to put into context the airports with the highest rate of cancellations, we will use a map and add markers with varying size based on the results for each airport.

Findings

In terms of the airports with a higher rate of cancellations, there are two regions which stand out: New York and Houston with 3 respectively 2 entries each. In particular, the New York airports record worst stats as there is combination of lower flight volumes with high number of cancellations. Atlanta and Chicago also make it to the top 10 but in their case the total volumes are also high, which contribute to the high number of cancellations.

Departure Airports with Highest Rate Of Cancellations



Cancellation Causes Monthly Evolution

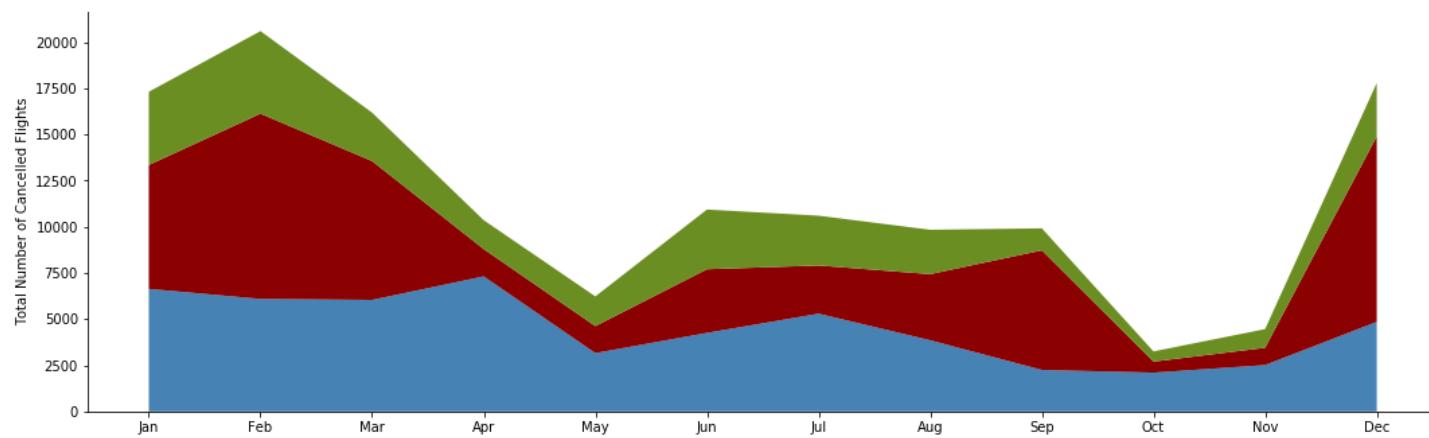
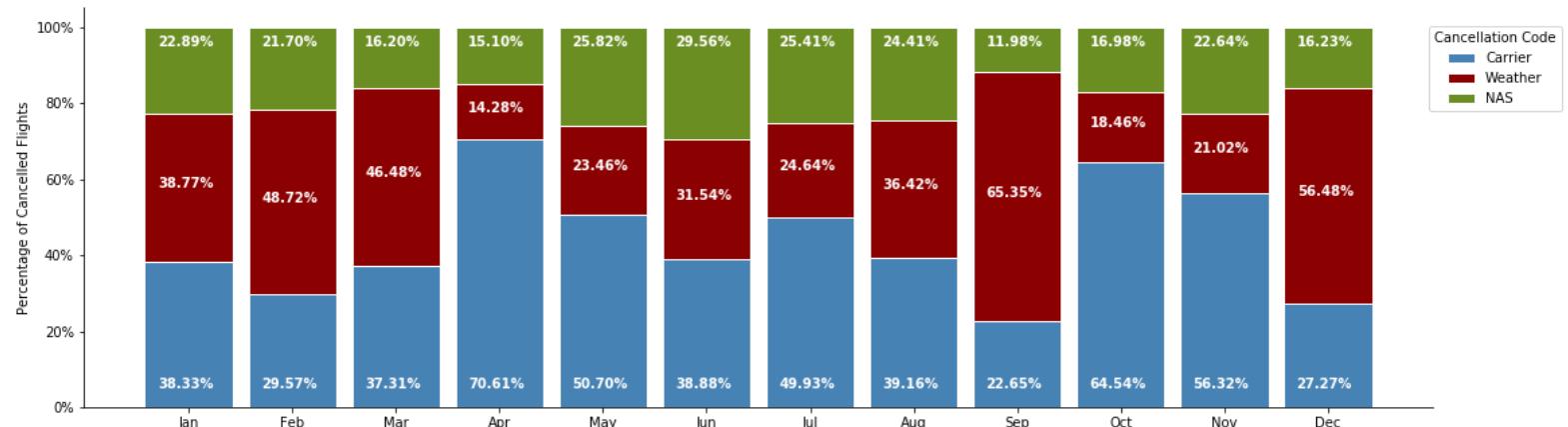
Objectives

A stacked barchart will be used to show the distribution of cancellation causes over each month. In order to correlate this with the number of flights per month, a second area chart will be created.

Findings

In terms of the total number of cancellations, the best months to travel are late spring (April, May) and late autumn (October, November). These are also the months where there are the least frequent weather caused cancellations. At the opposite end we find the winter months where the high rate of weather cancellations are correlated with an overall high percentage. During Summer, weather cancellations are less frequent, but the carrier cause delays record a spike in July which might indicate some issues in coping with the volumes of the holiday season.

Flight Cancellations Distribution Per Month



Delay Rates And Average Waiting Times

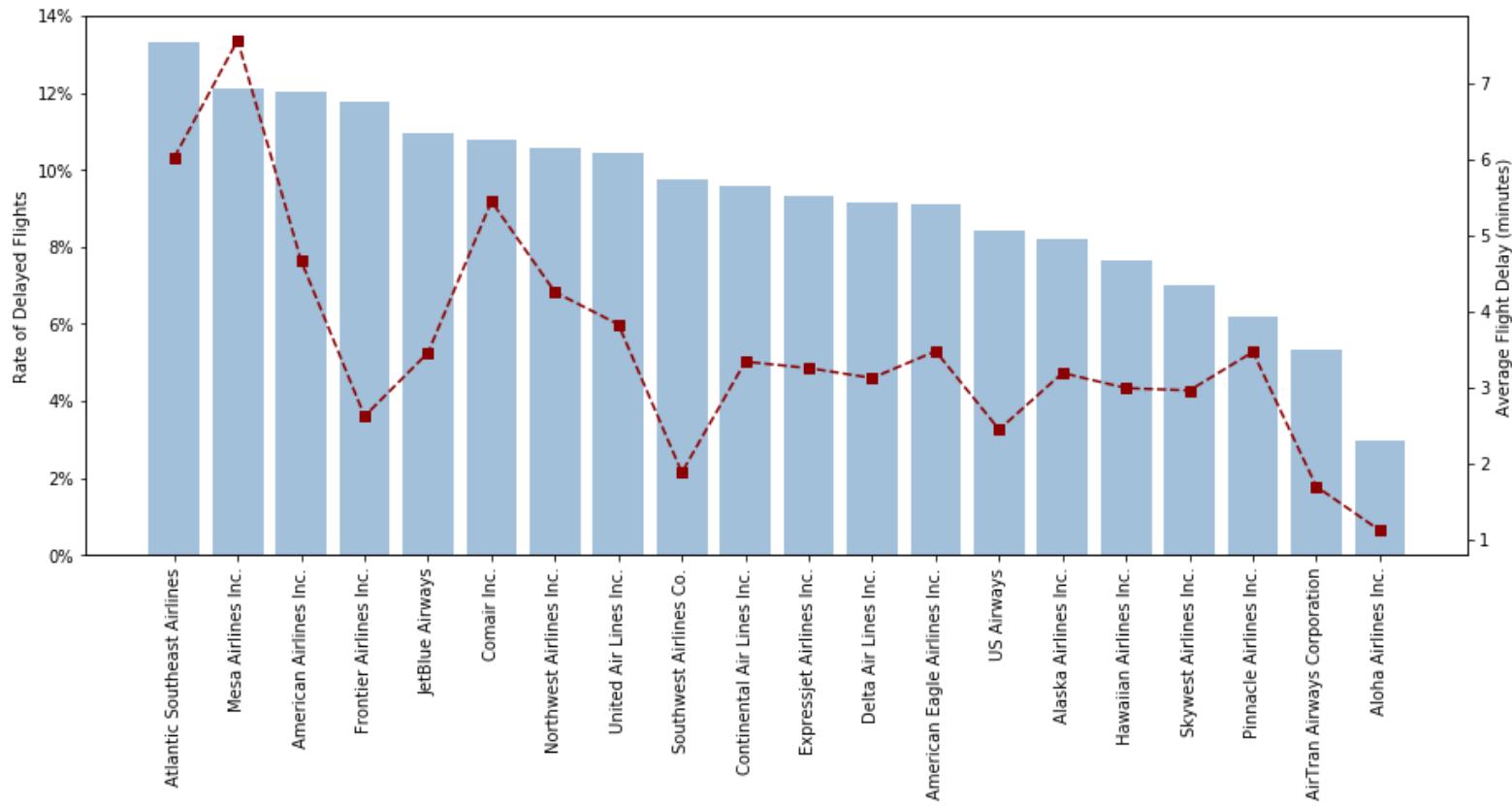
Objectives

In order to classify the carriers based on their delays, we will use a dual axis chart. On the main x axis a barplot will display the rate of flights delayed. While for the second x axis, we will draw a line a line chart to show the change in average delay from one company to another. This will help in identifying companies with high rate of cancellations and also high average waiting times.

Findings

In terms of the average chance of having a flight delay, most companies stand within the 8%-14% interval. Although in some cases (Frontier Airlines, Southwest Airlines) the average wait time is relatively low, close to 2 minutes. On the negative end, there are some similitudes with the cancellation charts. Mesa Airlines again scores second in terms of delay chances and first in terms of average wait time. American Airlines also gets bad marks in terms of the delays probabilities, but scores better on the average wait time. On average, for all carriers, the wait time falls somewhere between 2 and 5 minutes.

Delay Rate and Average Time Per Carrier



Flight Delays Distribution

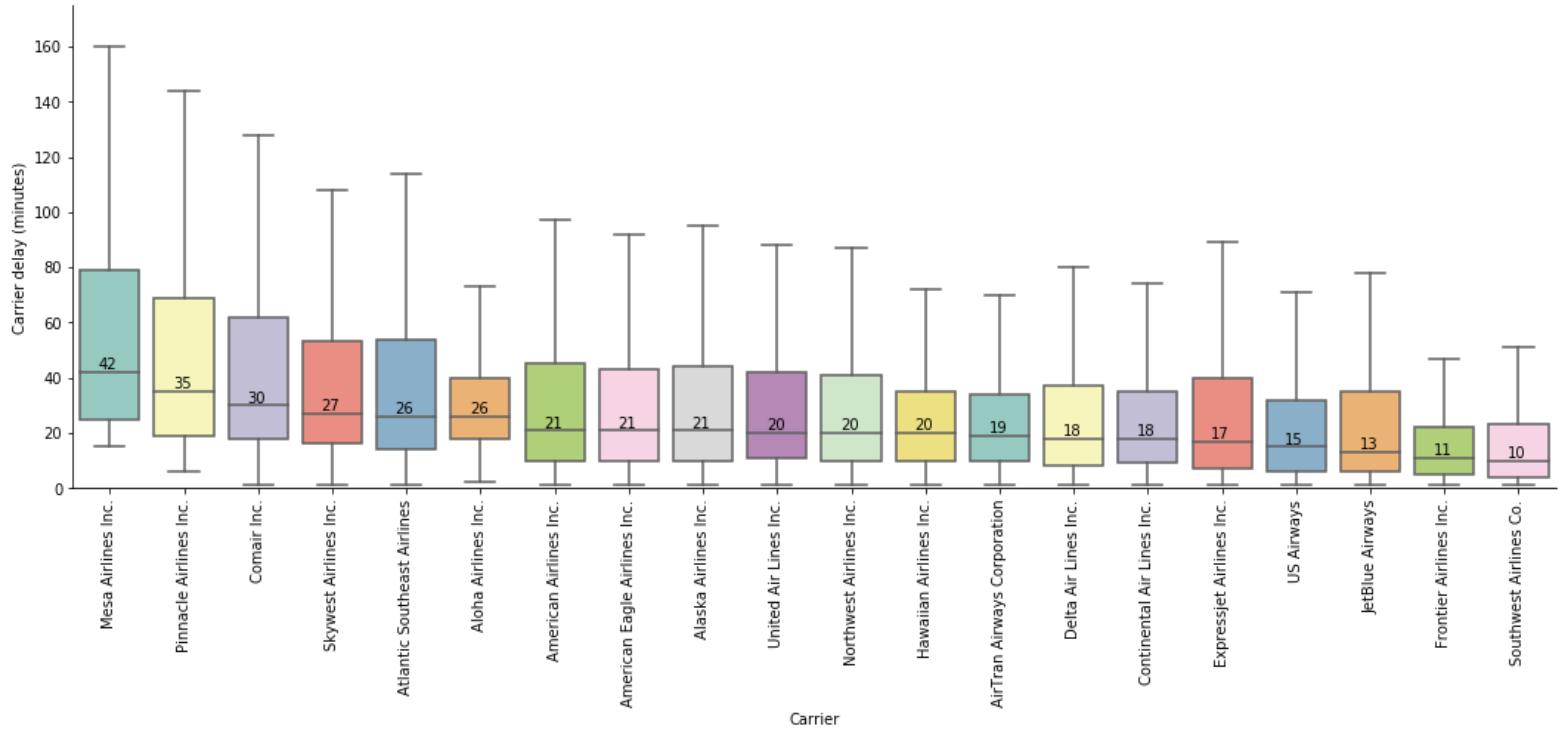
Objectives

To complete the information on the delays at airline level, we will use a series of boxplots, to show the range in which the delay times are floating. The values of the medians will be shown on the chart for easier comparisons. Outliers were removed as they were polluting the visualisation and also since they are present for all companies, they were not bringing any added value.

Findings

The difference in waiting times are more pronounced when starting only from the delayed flights and looking at the median. Most values range between 10 and 20 minutes. The result is quite remarkable for Southwest Airlines which records the lowest value while having the most number of flights. Frontier airlines also reports a good result with a median of only 11 minutes and tight ranges for the delays. For the major airlines such as United and American airlines, the values are relatively close, with differences of only +/- 1min. On the negative end, again Mesa airlines tops the chart, with a median of 42 and likely values up to 80 minutes.

Flight Delays Distribution Per Carrier



Monthly Average Delay Per Airline

Objectives

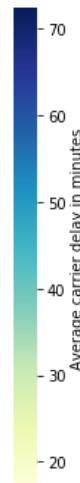
In order to reflect the seasonal correlations, we will use two heatmaps (for carriers and airports). The full name of the airline/airport will be used for easier comprehension.

Findings

The breakdown of delays per month complements the boxplot chart and shows that the airlines performing badly in terms of delays (Mesa or Pinnacle) are consistent throughout the year. Likewise the good performers are consistent in their results without any seasonal spikes. Overall there is no clear trend for all companies. We can see that with some exceptions (Atlantic, Skywest) most keep withining a tight range of values, all over the year. Atlantic is quite an interesting example as they have managed to almost half their average delay for the final 3 months in the year.

Average Flight Delay Per Airline and Month

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Pinnacle Airlines Inc.	62	60	55	54	54	56	59	49	54	54	48	55
American Airlines Inc.	36	38	38	36	35	39	41	39	43	41	43	41
Aloha Airlines Inc.	38	37										
Alaska Airlines Inc.	37	41	40	45	41	39	39	37	39	37	36	37
JetBlue Airways	24	26	25	26	27	29	38	44	37	32	29	34
Continental Air Lines Inc.	32	34	34	29	32	34	38	37	43	37	35	38
Delta Air Lines Inc.	36	36	33	37	35	30	42	37	37	32	29	29
Atlantic Southeast Airlines	48	50	51	49	46	48	53	60	50	31	28	27
Frontier Airlines Inc.	23	24	19	21	20	20	23	26	23	27	27	22
AirTran Airways Corporation	27	29	33	32	31	32	44	37	28	28	26	29
Hawaiian Airlines Inc.	40	41	56	38	47	37	33	40	46	36	37	38
American Eagle Airlines Inc.	40	39	38	38	37	37	38	36	43	41	35	40
Northwest Airlines Inc.	35	38	40	44	42	41	40	45	49	47	43	34
Comair Inc.	45	48	53	47	43	52	53	52	49	48	46	60
Skywest Airlines Inc.	58	51	46	34	37	40	44	41	42	33	31	34
United Air Lines Inc.	37	38	35	35	32	36	38	38	38	39	45	36
US Airways	26	27	28	31	30	31	34	32	31	28	27	26
Southwest Airlines Co.	17	18	17	19	18	19	23	20	21	20	22	21
Expressjet Airlines Inc.	32	34	38	30	31	35	34	41	37	33	32	38
Mesa Airlines Inc.	70	72	61	62	58	66	59	57	60	54	54	69

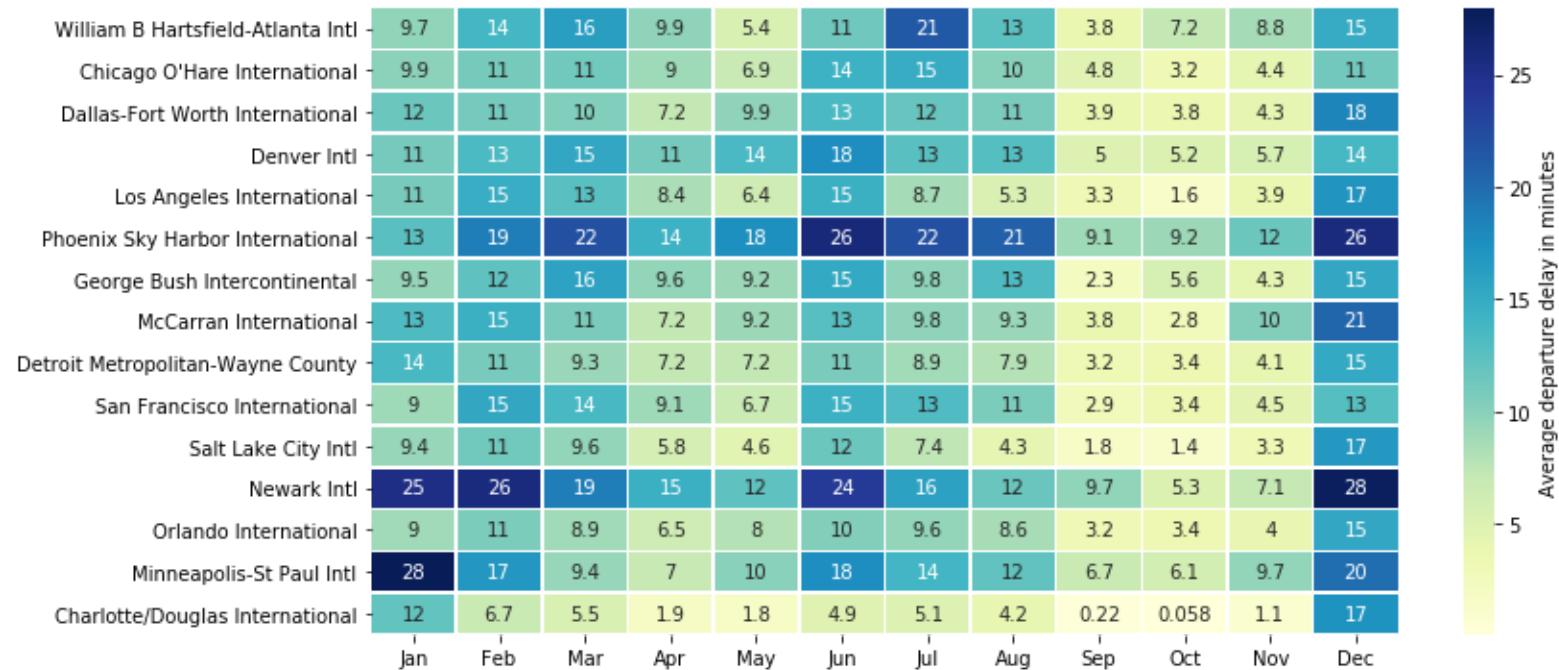


Monthly Average Delay Per Airport

Findings

As opposed to the airlines, the airport show common trends across months. Inline with previous results, the autumn months as well as late spring record low average delays. While on the opposing end there are spikes during mid summer and December. Amongst the airports with low delay times we can find Charlotte International, Orlando, or LA. On the opposite side, Newark and Phoenix consistently score amongst the highest values.

Average Flight Delay Per Origin Airport and Month



Weather Caused Departure Delays

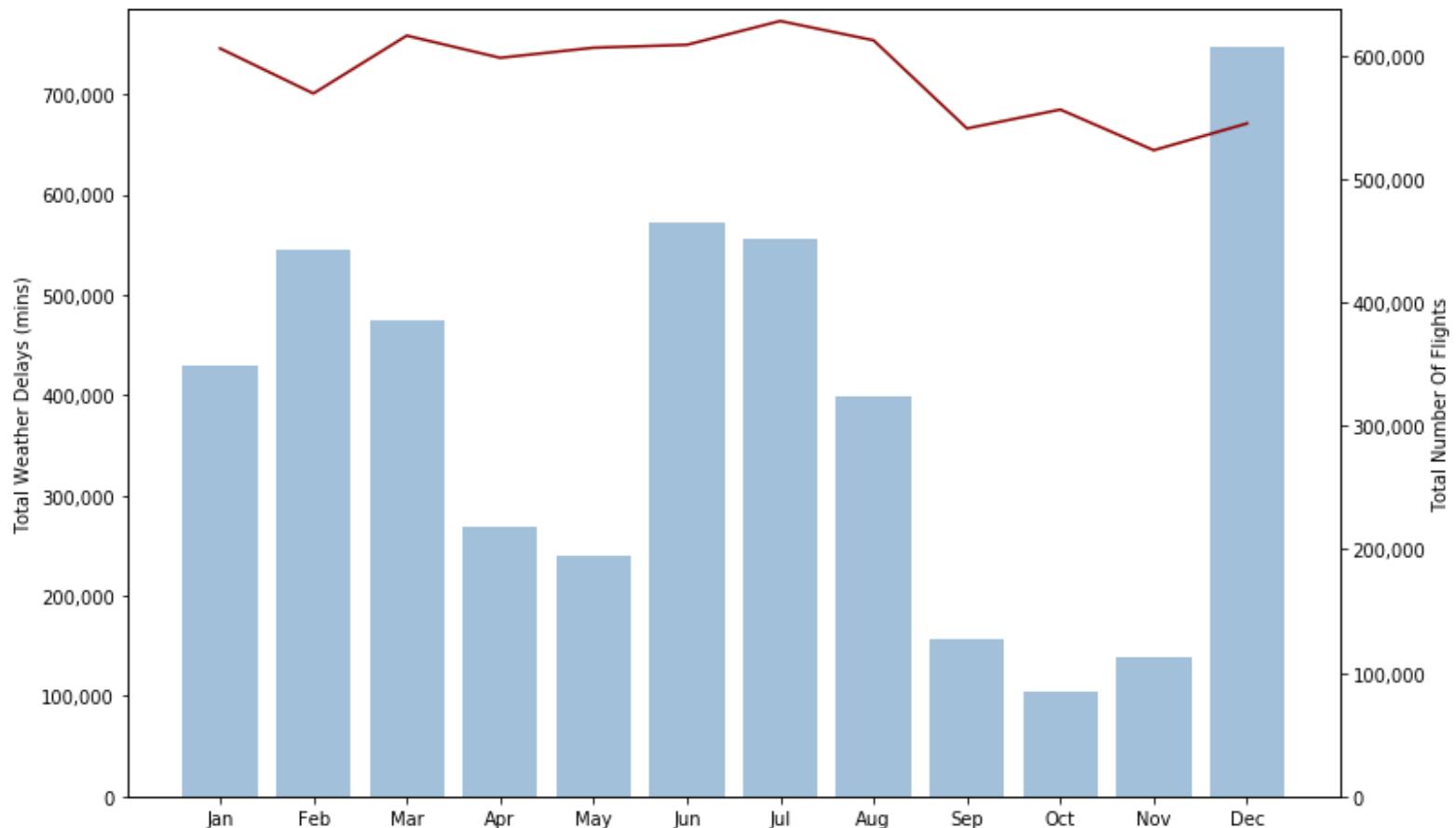
Objectives

In order to reflect the risk of weather delays for each month, a dual axis chart will be used. Main x axis will be shown as a barchart with the total time of delays per month, while on the second axis we will show the number of flights per month (as a line chart) to correlate the two amounts and see if the delays are potentially caused by an increased number of flights.

Findings

To complement the previous results, weather specific delays, confirm that autumn and spring are safer to travel. Interestingly, December records a major spike compared to November (almost six times higher) despite a relatively close total number of flights. During the summer season, June and July perform significantly worst than August, recording almost 25% more delay time.

Departure Delays Due To Weather Conditions



Routes With The Highest Risk of Delays or Cancellations

Objectives

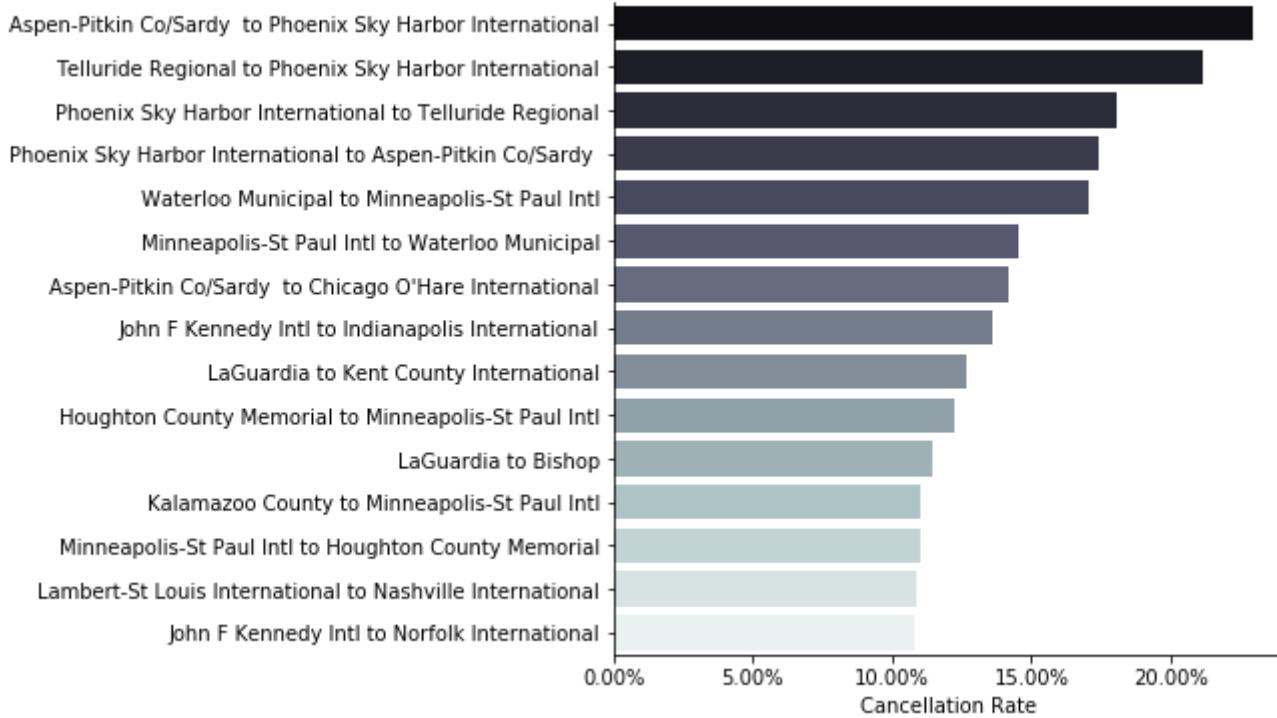
For the final section of the analysis we aim to identify the routes which record on average the highest number of issues. For each criteria (cancellation or delay) we will create two charts: a barchart showing the values and names, and a map drawing the actual routes.

Findings

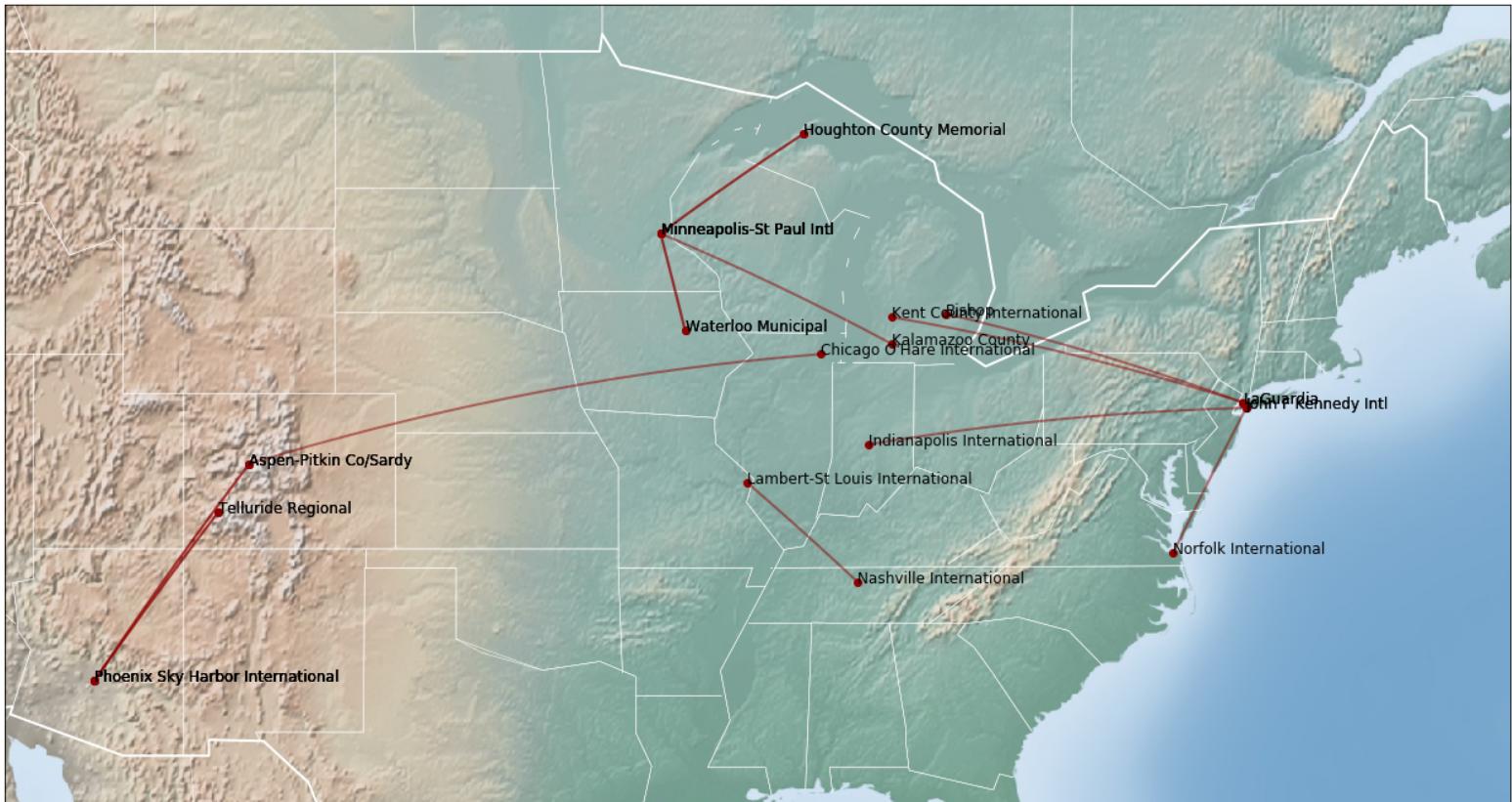
In terms of cancellations, the least desirable routes are between Phoenix on one side and Aspen and Telluride, with rates close to 20%. New York airports also seem to have issues with some of the routes towards Indianapolis or Norfolk. Overall the rates range between 10% and 20% and as it can be seen on the map, the routes are relatively short distance and concentrated in the eastern coast.

On a different note, the routes which record the highest delays are specific to a higher degree to longer flights (Chicago to LA or San Francisco). There are also several common locations which appear in the list, such as Newark, Washington or Chicago. On average the delays range between 40 and 60 minutes.

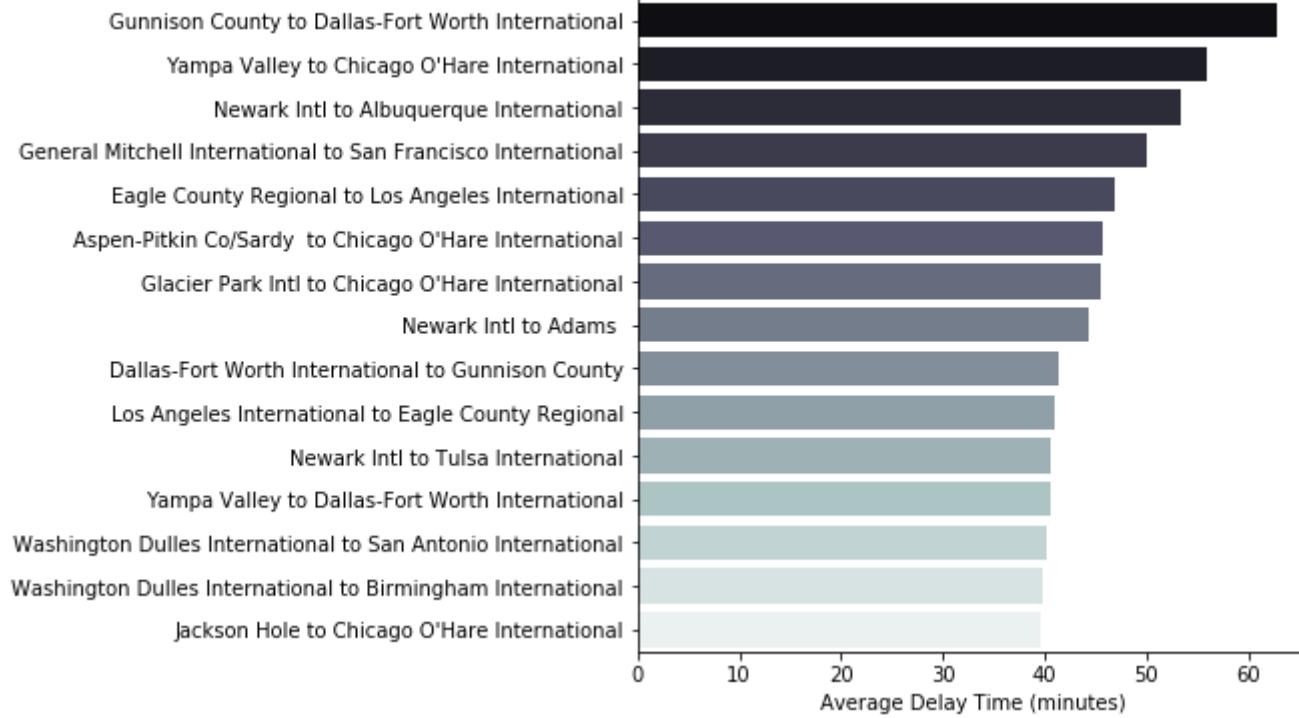
Flight Routes with Highest Rate Of Cancellations



Flight Routes with Highest Rate Of Cancellations



Flight Routes with Highest Average Delay



Flight Routes with Highest Delays

