Business blog assignment

Flight delays and cancellations can be expensive for businesses and ruinous for vacation travellers. With millions of flights departing annually in the United States, odds are pretty good that you've experienced the frustration and helplessness associated with interrupted travel.

The problem

Stakeholders are hungry for better departure and arrival information for US flights. For example, flight crews are only paid when aircraft doors are closed and flight underway. Departure delays directly impact overall time at work, upsetting work-life balance. Prospective employees want to work for airlines who have proven track-records.

Airport authorities and airline service providers are very interested in predicting performance. For non-hub airports, airlines lease transient gate spots and hire third-party ground crews for fuelling and baggage handling. When flights arrive late at the airport, they face stiff penalties as alternative gates need to be assigned, causing tarmac delays. Ground crews have to alter schedules, increasing time and stress associated with providing services. If groups could anticipate when a flight would likely be delayed, based on predictive modelling, they could have contingency plans in place to ease the burden on crews and other airlines.

The goal of our research, analysis and modelling is to provide travellers and other stakeholders with a tool that can be used to better predict on-time performance, allowing the public and those associated with airline travel to make informed decisions about which airline to best suit their needs.

The data

The dataset being used to develop our predictive models includes nearly six million lines of information from 2015 flights, and is sourced from the Department of Transportation (https://www.kaggle.com/usdot/flight-delays). It takes into account departures, arrivals, as well as the frequency of flight cancellations and categorizes the reasons why.

The data was very imbalanced, which created difficulties for early iterations of the models, and had to be compiled from multiple files and sources in order to be ready for analysis. It also contained many "NA" entries that had to be removed and reconciled, again adding to the challenge for our team.

Methods used

Our approach was to analyze data from millions of flights to create predictive models that use past performance to indicate the likelihood of future on-time airline performance.

We were able to run various models that showed the likelihood of cancellation, using multiple variables. But noticed the Random Forest model (74% accuracy) was providing more accurate results than the SVM (66% accuracy) with this dataset.

Our various regression models estimate the chances of a flight being cancelled as a percentage, and also determine the likelihood of delays, and if so, the chance of it being delayed longer than 15 minutes.

We intend to create a front-end application that will allow users to enter intended travel dates and selected airlines, providing them with simple to read predictive airline performance.

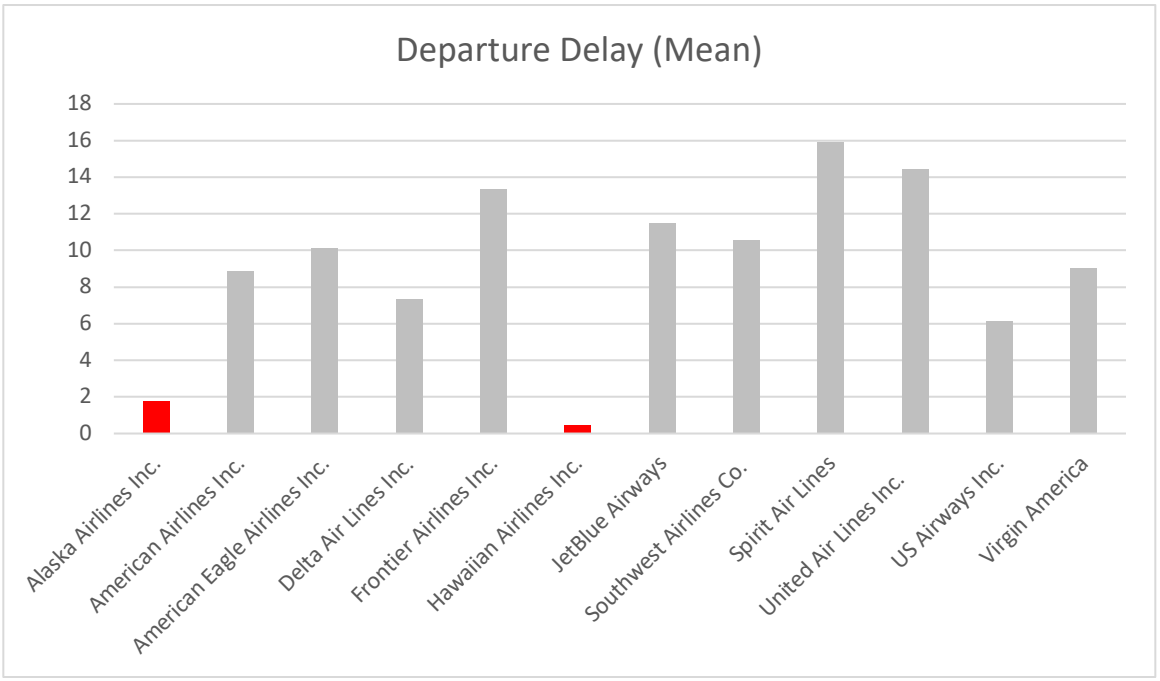
The results

With air traffic congestion and weather events set, are we able to adequately evaluate individual airline performance?

Yes. Air traffic congestion, severe weather and security events are all something experienced by national carriers, and are considered outside the airlines' control. However, by comparing on-time arrivals, departures and the amount of cancelled flights, we're able to evaluate the differences between carriers. These differences will be valuable in predicting which airlines have the best on-time flight performance and likelihood of cancellations, and will allow us to make accurate suggestions to users of our application. This information can then be compared against current seat prices, to find the best value for cost-conscious travellers.

Is there a significant difference between national carriers?

Yes. Our analysis showed a wide gap between performances, with Alaska Airlines having the most on-time arrivals, second best on-time departures and second least amount of cancelled flights in the group. On the other end of the scale was Frontier Airlines and Spirit Airlines. They were bottom of the pack in both arrivals and departures, and Spirit Airline's cancelled flights were in the bottom quarter of all airlines.



What percentage of delays/cancellations are due to airlines versus weather or air traffic congestion? Is it significant?

According to our analysis, the biggest contributing factors for delays are "late_aircraft delay" (40% of the total delay time) followed by "airline_delay" (32%), "air system_delay" (23%) and "weather_delay (5%). There is no significant delay related to security. For cancellations, weather is the biggest driver, followed by Airline issues.

Do flight speeds impact on-time arrivals?

Most airlines fly at similar speeds (400-450 miles/h), and it seems that there is not any strong correlation between speed and on-time performance. Hawaiian Airlines has the slowest speed in the group, but has the third best arrivals times. However, this may mean that Hawaiian Airlines is just better at estimating their arrival time. To better answer this question, we should compare the time it took for different airlines to travel the same route. Maybe Hawaiian Airlines just have longer flight time compare to other airlines.

Can we predict which airline will have the greatest chance of an on-time departure?

Based on past performance, we feel confident that future performance can be accurately predicted. The airlines with the greatest chance of an on-time departure are Hawaiian Airlines (mean score of 0.48) and Alaska Airlines (1.78).

Can we predict which airlines will have the greatest chance of an on-time arrival? what about cancellations?

Based on past performance, we feel confident that future performance can be accurately predicted. The airlines with the greatest chance of an on-time arrival are Alaska Airlines (-0.97), Delta Airlines (0.18), and Hawaiian Airlines (2.02). As for the cancellations, more than 5% of "American Eagle Airlines" flights got cancelled! In contrast, only 0.2% of Hawaiian Airlines and 0.38% of Alaska Airlines flights were cancelled in 2015.

Conclusion

Travellers have different objectives when selecting who to fly with. Business fliers aren't as cost-conscious, and may put a greater emphasis on on-time arrivals due to important meetings. Companies will pay a higher rate to ensure their employees arrive when required.

Vacation travellers on cruises or all-inclusive holidays, with set start and stop dates, are very time sensitive. They may choose to pay slightly more for a greater likelihood of on-time arrivals, increasing the odds they'll get to enjoy the vacation they've already invested so much in.

Price-sensitive users will look for the cheapest way to get from point A to point B, but when prices are similar between airlines, will find the predictive score valuable to choose the best airline for the money.

Predictive knowledge puts important information at the fingertips of travellers, providing value and insight for making informed decisions. Easing the stress associated with flight.