**AIRLINE DELAY ANALYSIS**

**DISTINGUISHING BETWEEN SYSTEMATIC AND RANDOM DELAY FACTORS**

Angela Fogel for Remedy Partners

**OVERVIEW**

Accurate flight scheduling of individual flights is critical in managing airport flight traffic. Even a slight flight delay can create a cascading effect of airport delays as crew members are late for departing flights, late passengers delay connecting flights and late aircraft equipment interrupts following flights. Understanding the relationships between flight components such as taxiing, carrier delays and weather factors can help distinguish systematic from random events and identify predictable factors that affect flight delays.

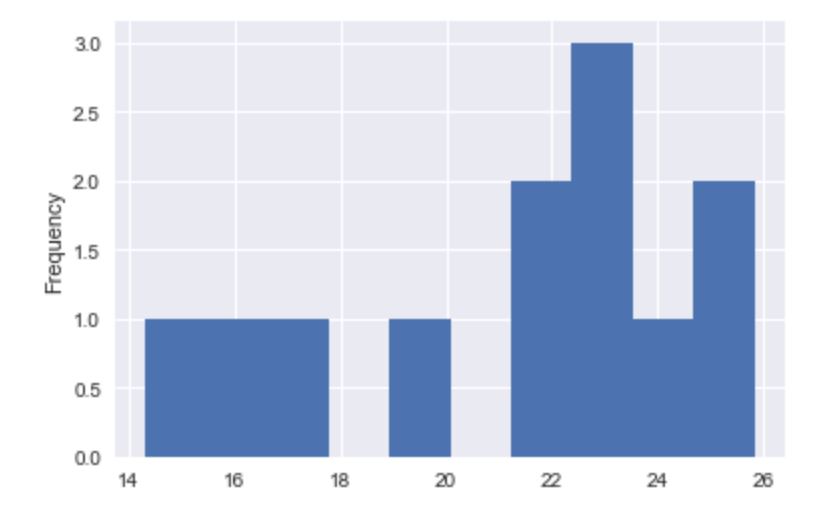
Systematic delays such as carrier delays are a significant proportion of airline on-time performance. Carrier delays include maintenance, aircraft cleaning, crew problems, baggage loading, fueling, etc. Ultimately, scheduling can be optimized by incorporating information about routes having a high probability of systematic delay.

This analysis generates models that identify key factors in arrival delay. Data was gathered from Department of Transportations On-Time Performance database from the period January 2017 to May 2018. Pearson correlation and linear regression methods were used to extract flight features that are highly correlated with arrival delay.

**METHODOLOGY**

Data Filtering

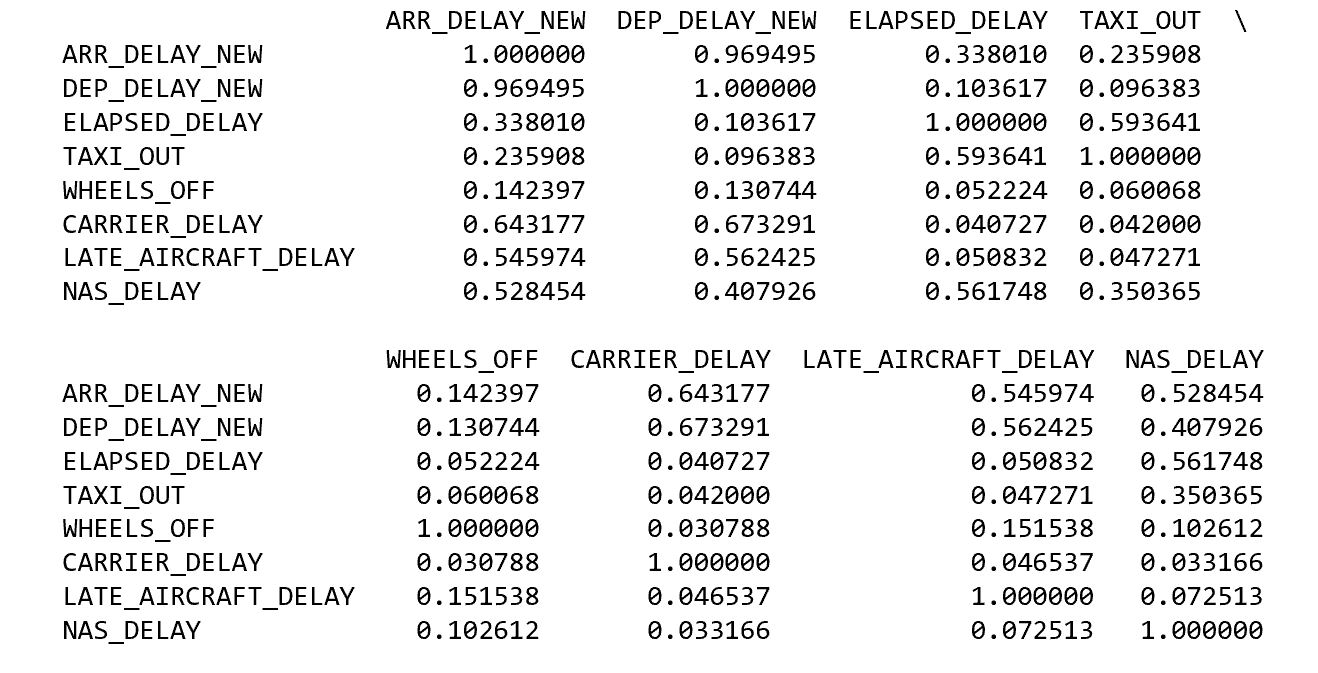
Records missing arrival and departure times were removed. In the remaining records, NaN values where converted to zeros. Outliers were filtered based on average monthly delay as illustrated in the following histogram. A flight is considered delayed if its actual time varies from scheduled time by more than 15 minutes. The majority of the delays lie between 15 and 24 minutes. Accordingly, analysis data was limited to delays between 15 and 24 minutes. Inspection of actual records to understand data definitions and gathering methods revealed that delays are not always categorized by reason. Also delay reasons are gathered only for arrival delays. For an accurate correlation to delay reasons, the analysis focused on arrival delay performance.

**Average Monthly Arrival Delay Histogram**

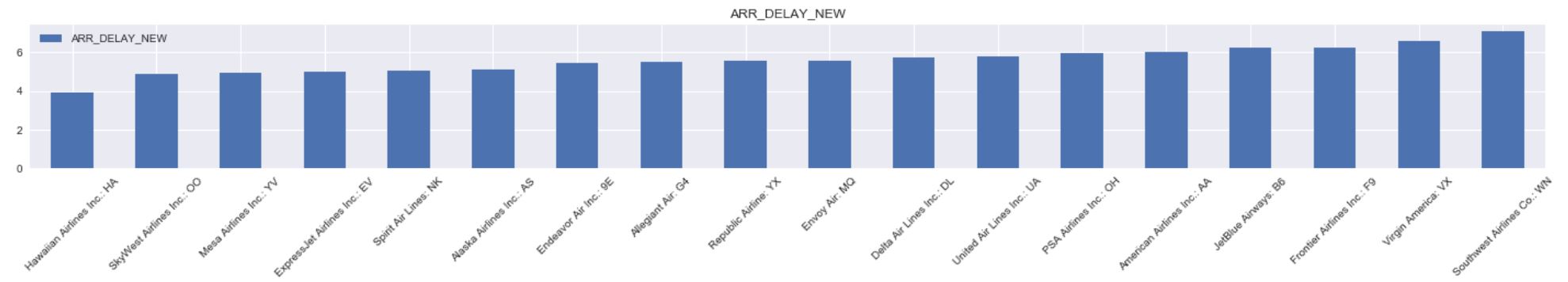
Monthly average delay was used in calculating summary statistics. The resulting data set is efficiently processed while still at an adequate granularity to capture any seasonality effects.

A pairwise correlation matrix of all independent variables resulted in a narrow list of correlated variables (shown below).

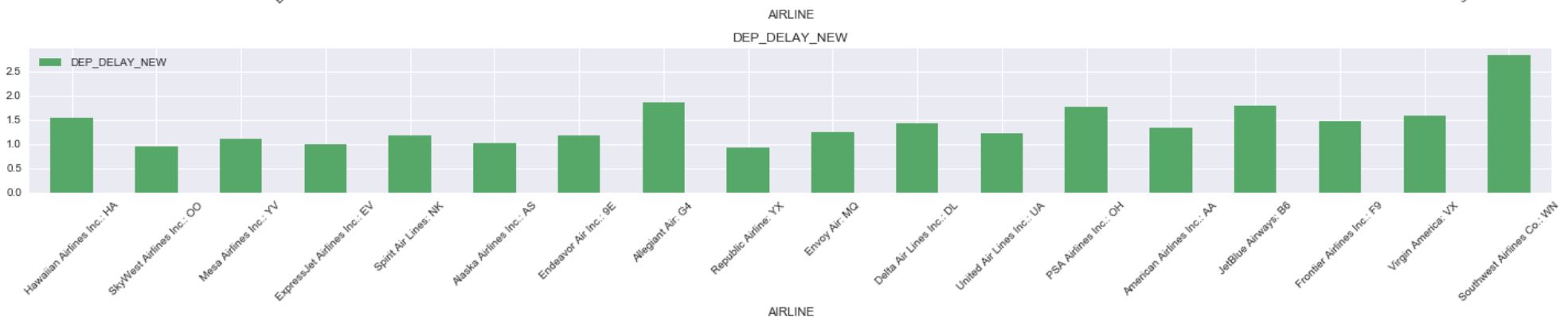
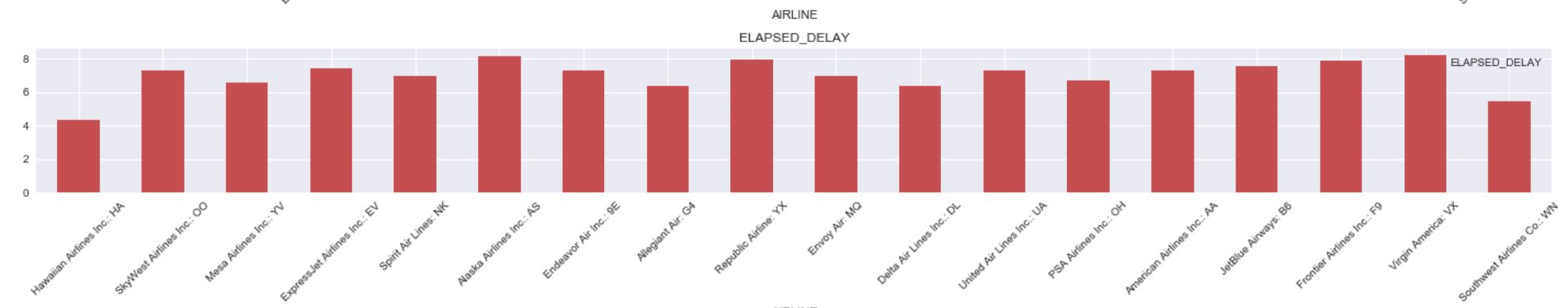
**Correlated Independent Variables**



**DISCUSSION**

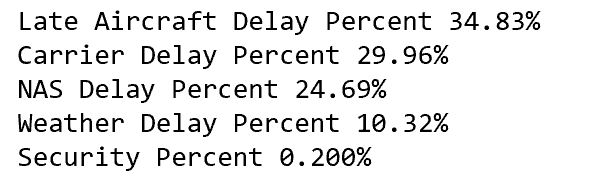
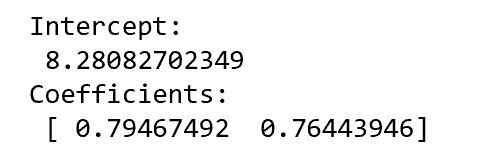
Arrival delays are highly correlated to departure delay. Since delay reasons are not available for departures it is assumed that the reasons for departure delay will follow similar proportions as arrival delays. One factor specific to departure delay correlations is the impact of taxi out time. Since taxi out time is only captured within elapsed time delays, a complete picture of delay characteristics would include all variables related to elapsed time delay.

The following charts illustrate a divergent pattern between departure and arrival delays when compared to elapsed time delays. The difference suggests factors not captured in arrival and departure delay metrics. Specifically, Southwest Airlines has a low average elapsed time delay but the highest average arrival and departure delays of all airlines. Further analysis of elapsed time delay outside the scope of this paper could uncover that factors such as flight time, distance and taxi time have an inverse relationship to overall delay for some routes.



This analysis focused on the reasons for arrival delay presented in the On-time Performance database. Of the five causes of delay (Carrier Delay, Weather Delay, NAS Delay, Security Delay, Late Flight Delay), Carrier Delay and Late Aircraft Delay are the most often reported reasons for arrival delays. Carrier Delay and Late Aircraft Delay comprise 30% and 35% of all delays, respectively. These delays are systematic, predictable and controllable by airlines. They also exhibit a significant correlation to arrival delay.

**Delay Reason as Percent of Total Arrival, Carrier, Late Aircraft Regression Coefficients**



**CONCLUSIONS**

A significant proportion of factors related to flight arrival delays are under airline control. Continuous improvement of flight processes such as maintenance, aircraft cleaning, fueling, baggage loading, etc. can positively impact on-time flight performance. At the same time, incorporating delay probability metrics and prediction capabilities for historically problematic routes can improve scheduling and more effectively manage passenger expectations. A detailed analysis of carrier delay factors will be an effective next step toward this goal. Also, addressing overall flight performance through an analysis of elapsed time delay components is likely to uncover other areas of performance improvement.