INTRODUCTION

Last year, there were a total of 7 million domestic flights, with over 20% of them being delayed. An additional 3% were outright cancelled. Which airlines are the best at keeping delays at a lower rate? Are there any differences between the airlines in the COVID era? There is anecdotal evidence that the general population likes to rely on but in this study, we attempt to identify clusters of airlines when modeling to delays and cancellations.

METHODS

Flight delay and cancellation data from March 2020-November 2022 were obtained from the Bureau of Transportation Website.

- More recent data were not available at time of collection.
- 8 Variables were selected to be analyzed in a Clustering Model
  - Arrival Delay, Departure Delay, Cancelled Flight, Carrier Delay, Security Delay, NAS Delay, and Late Aircraft Delay

Airlines were selected as the reference factor. Observations for each variable were summed to create an overall snapshot per Airline.

- Summations were then standardized to the normal distribution.

Hierarchical clustering was chosen over k-means clustering due to the small number of observations.

3 types of clustering techniques were used, Complete, Single and Ward's to determine the best clustering model. Ward's method was chosen as the best clustering model.

The clustering analysis results in the identification of five distinctive groups of airlines.

Figure 1. Cluster Dendrogram of Airlines

Figure 2. 2D Representation of Airline Clusters

**RESULTS**

Table 1: Cluster Analysis

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Alaska Airlines</th>
<th>American Airlines</th>
<th>Delta Air Lines</th>
<th>Frontier Airlines</th>
<th>ExpressJet</th>
<th>Allegiant Air</th>
<th>Spirit Airlines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>Alaska Airlines</td>
<td>American Airlines</td>
<td>Delta Air Lines</td>
<td>Frontier Airlines</td>
<td>ExpressJet</td>
<td>Allegiant Air</td>
<td>Spirit Airlines</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>Alaska Airlines</td>
<td>American Airlines</td>
<td>Delta Air Lines</td>
<td>Frontier Airlines</td>
<td>ExpressJet</td>
<td>Allegiant Air</td>
<td>Spirit Airlines</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>Alaska Airlines</td>
<td>American Airlines</td>
<td>Delta Air Lines</td>
<td>Frontier Airlines</td>
<td>ExpressJet</td>
<td>Allegiant Air</td>
<td>Spirit Airlines</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>Alaska Airlines</td>
<td>American Airlines</td>
<td>Delta Air Lines</td>
<td>Frontier Airlines</td>
<td>ExpressJet</td>
<td>Allegiant Air</td>
<td>Spirit Airlines</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>Alaska Airlines</td>
<td>American Airlines</td>
<td>Delta Air Lines</td>
<td>Frontier Airlines</td>
<td>ExpressJet</td>
<td>Allegiant Air</td>
<td>Spirit Airlines</td>
</tr>
</tbody>
</table>

The Ward's method dendrogram provides a visual representation of the hierarchical relationships between the different airlines.

Figure 1: Cluster Dendrogram

Figure 2: 2D Representation of Airline Clusters

Using the clusplot function, a two-dimensional graph was derived from Ward's method cluster analysis. This plot easily illustrates which airlines belong to which cluster through the different shapes and colors.

Figure 3: Variable Means

In the ‘Variable Means by Cluster’ graph, the sign of the standardized values indicates whether a delay variable is above or below the mean for all the clusters combined. For example, a positive value for the ‘Departure Delay’ variable indicates that flights in cluster 5 tend to have longer departure delays than the mean departure delay time.

DISCUSSION

We were surprised by the diverse clustering of airlines, which did not follow our initial hypothesis that big brand names and regional lines would be clustered together. ExpressJet’s high rate of cancellations could be explained by their recent bankruptcy, which was discovered through further research. Southwest Airlines also had higher than expected delays, but future analysis with more recent data may show them belonging to a different cluster especially with the higher number of cancellations they experienced in December 2022. To improve predictions of delays, future studies should analyze additional flight factors. Customers should book with Alaska, Delta, Endeavor, Horizon, or Republic, and avoid Allegiant, Frontier, JetBlue, Southwest, and Spirit Airlines if possible.