**Final Report**

Team TEN: Ilter Çakmak, Michelle Guan, Cyndi Ng, Tony Yu
International Business School, Brandeis University
BUF 240F: Information Visualization
Professor Kamis
27 October 2021

## Introduction

Civil aviation is a large industry in the U.S. In 2016, civil aviation, directly and indirectly, contributed 5.2% of the US GDP (Federal Aviation Administration). Commercial airlines employed roughly 750,000 people in the U.S. in December 2019 (Bureau of Transportation Statistics). Being a prominent contributor to the U.S. economy and the source of livelihood for millions of Americans, the current standing and the future of the civil aviation industry is important to understand.

Many flights were canceled due to the risk of COVID-19. The industry shrank in terms of the number of employees due to the travel restrictions during the pandemic. Commercial airline companies as well as employees serving in the industry have been hit hard financially. As of December 2020, the number of airline employees is down at 695,000. In domestic flights alone, airline companies have lost $120 billion in revenue in 2020 compared to 2019 (International Civil Aviation Organization). While 2021 will be a better year compared to last year, the industry is not expected to catch up with its pre-pandemic financial strength.

## Our Consultancy's Focus

Our consultancy aims to curate data relevant to the air travel industry in the wake of the recent pandemic. Before the impact of COVID-19, civil air transportation contributed $850 billion in economic activity in direct sectors only, and $1.8 trillion across both direct and catalytic sectors (Federal Aviation Administration). As business slowly regenerates, looking at existing data can provide historical insights and help guide current decisions. In our reports, we seek to advise clients who are airline companies themselves by providing perspective about industry trends and exploring relationships in delays, airfare, and flight volume. By focusing on airfare and flight delay data from the recent past, we can provide a sense of how the flow of the industry can be altered to the benefit of airline companies.

## Data Quality Assurance

We used two datasets from the United States Department of Transportation Bureau of Transportation Statistics. The first dataset is 'Airline On-Time Statistics and Delay Causes'. It divides all the scheduled flights of a carrier company in an airport into seven categories: On Time, Air Carrier Delay, Aircraft Arriving Late, Security Delay, National Aviation System Delay, Extreme Weather, and Cancelled & Diverted. The second dataset we used is Table 5 from Consumer Airfare Report which is named 'Detailed Fare Information for Highest and Lowest Fare Markets Under 750 miles.' We chose the distance range of 'under 750 miles' due to the reason that passengers in short-haul markets are more sensitive to fare disparities and are more likely to act differently. Any airline that carries 10 percent or more passengers in a city-pair market is considered a competitor and recorded in this report.

As these data were summarized and reported by the U.S. Department of Transportation, accuracy, consistency, validity, and uniqueness were assumed. We conducted some data cleaning to match our needs for further analysis and visualization. The data cleaning we conducted included the arrangement of datasets to have consistent date information, the addition of carrier name to 'Table 5', the separation of the city, state, and airport information into separate columns, and removal of missing data.

The table below shows a summary of the two datasets.

| Table name | Delay Table | Table 5 |
|---|---|---|
| No. of records | 302291 | 13617 |
| Distinct Carrier Companies | 19 | 9 |
| Date Data Received | Jun 2003 – Jun 2021 | 1996 Q1 – 2021 Q1 |

For consistency, we used data points from 2016 Q1 up to 2021 Q1 for both datasets. By using these data, we can investigate the relationship between fares, passengers, and delays, in a manner of how market fares and flight delays affect passengers' behavior of choosing carrier companies in a city-pair market.

## Airline Categories

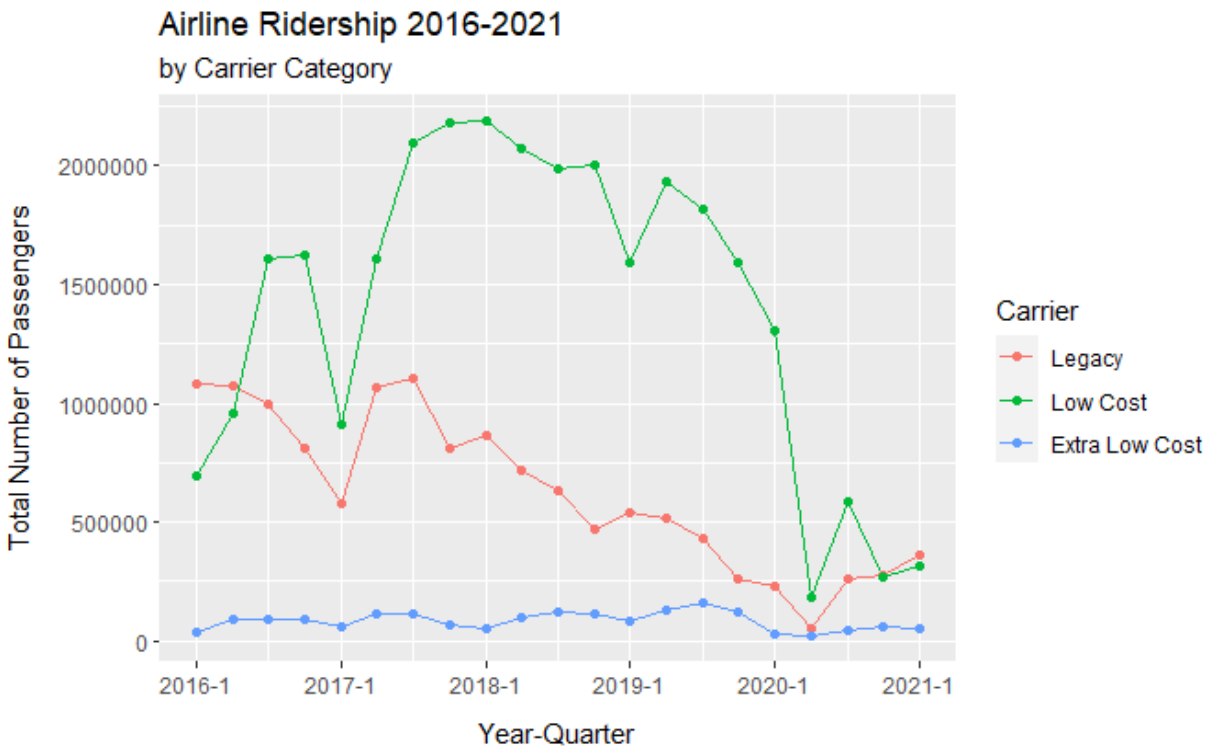
The three categories we will refer to in this report are Legacy carriers, Low-Cost carriers, and Extra Low-Cost carriers. These categories are defined by the airlines themselves, besides Alaska Airlines. Below is a chart summarizing the categorization of carriers discussed in Report 1.

| | Legacy | Low-Cost | Extra Low-Cost |
|---|---|---|---|
| Airlines | American, Delta, United | Alaska, Jetblue, Southwest | Allegiant, Frontier, Spirit |
| Average Fare Range | $200+ | $120-150 | $70-$100 |

While Alaska Airlines is not officially classified as a Low-Cost airline, we have placed them in the Low-Cost airline category as their flight volume, airfare, and price range are closer to the range of Low-Cost Carriers than other categories.

## Visualizations

### Ridership in the recent past




In investigating how the pandemic has affected the airline industry, we can look to the change in ridership. It can be seen from this line chart that the number of passengers, particularly for Legacy and Low-Cost carriers, in these markets hit a peak in mid-2017 and the beginning of 2018. (The trend for Extra Low-Cost carriers has not changed significantly in years prior to 2020 aside from seasonality.) After this peak, total passengers on these routes began declining at an increasing rate. This was before the full effect of the pandemic in the United States, implying that the airline industry showed signs of distress independent of COVID-19, underscoring the need for an upturn.

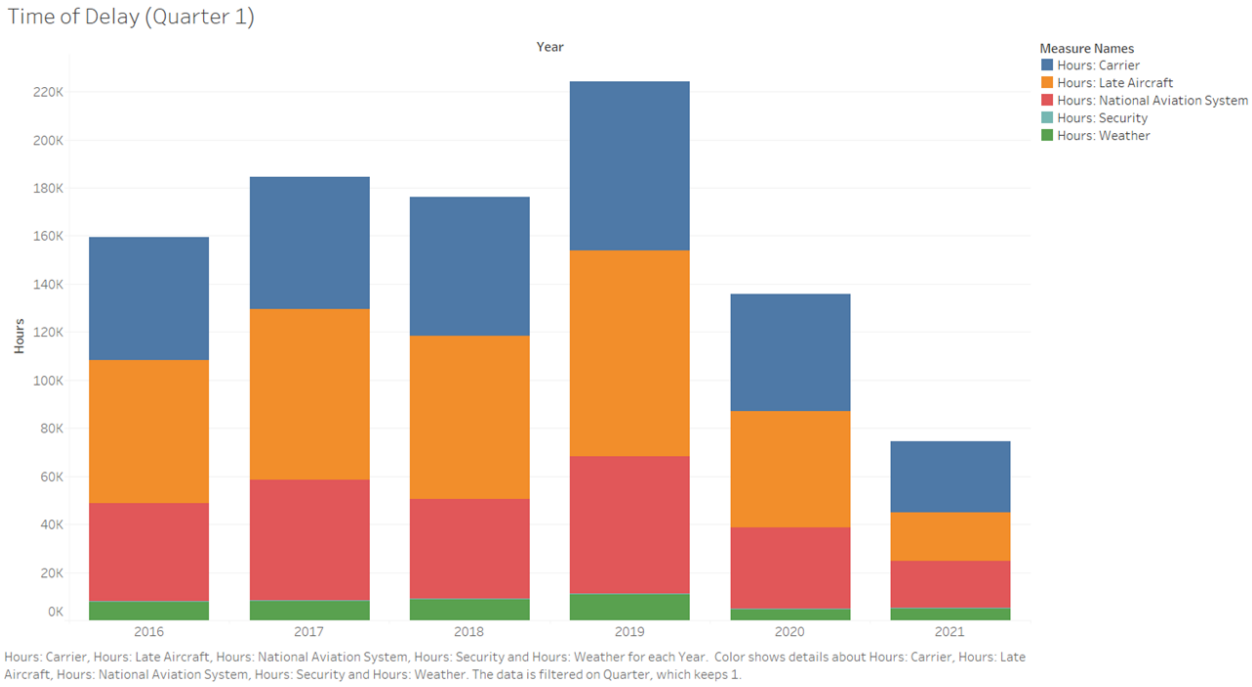
While the industry as a whole felt the impact of the pandemic, it is clear that Low-Cost carriers saw the largest drop in ridership in 2020, followed by Legacy carriers. In comparison, Extra Low-Cost carriers experienced a much smaller dip, possibly due to less market share in general. Because of the dramatic drop of the two larger categories, all carriers now have a chance to rebuild and grow their ridership numbers by taking strategic actions to gain a better position within the market.

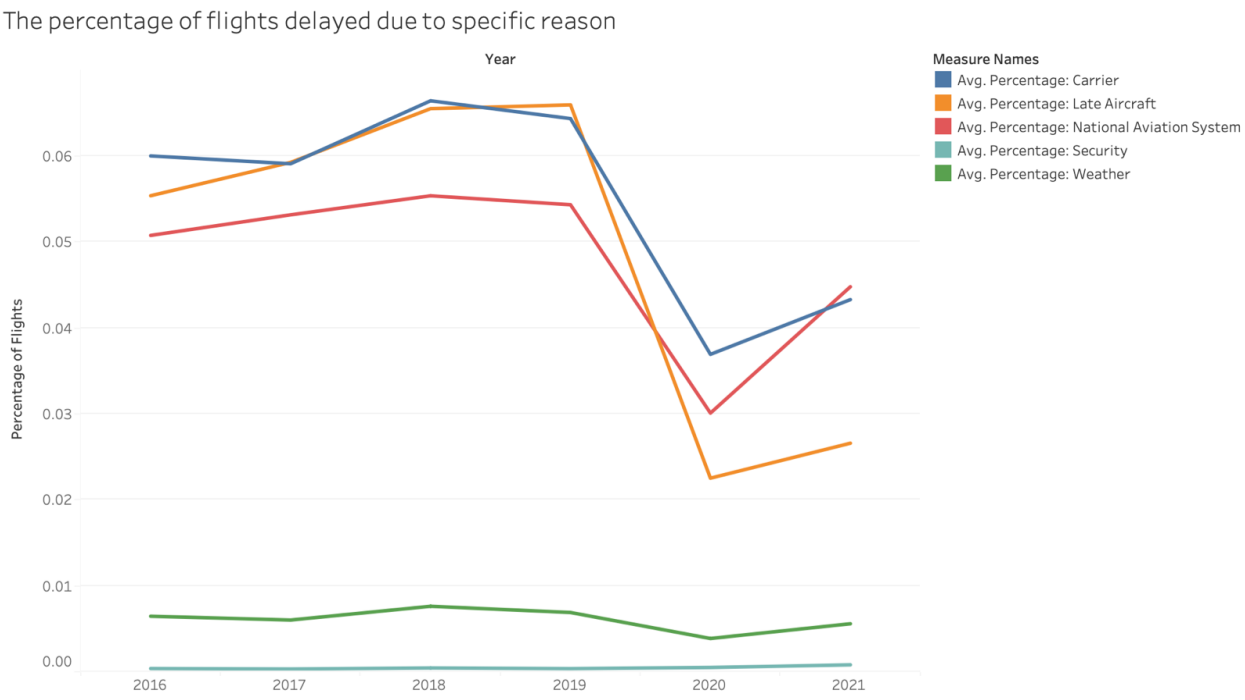
**Delay Reasons and the Opportunity Cost of Time**

According to the Federal Aviation Administration, time delays cost the airline industry an estimated total of $113.5 billion in the years 2016-2019 (“Cost of Delay Estimates”), with $71 billion of that in passenger costs and lost demand. Airlines can assess these costs by studying delay times across the industry. To evaluate performance and explore the relationship between

prices, delays, and passenger behavior, it is important to understand the severity and reasons behind the delays.

In this stacked bar chart, we are showing the total time of delays by delay reason in the first quarter of the years 2016-2021. The reason we are presenting the first quarter of every year is to be able to compare the delay time in each year to the most recent piece of data available, Quarter 1 of 2021. In the line chart, we can see a breakdown of the percentage of flights delayed by specific reason, in relation to the total number of scheduled flights in that year. This gives a wider look at how certain delays compare to all flights. In interpretation, it is important to remember that even though the percentage of delays of a specific reason is comparatively small to the total number of flights, the total delay times are a significant factor in flight (and customer) experience and have a larger impact on flights experiencing delays. To put it another way, delay length matters more in the context of a delayed flight rather than an on-time flight.

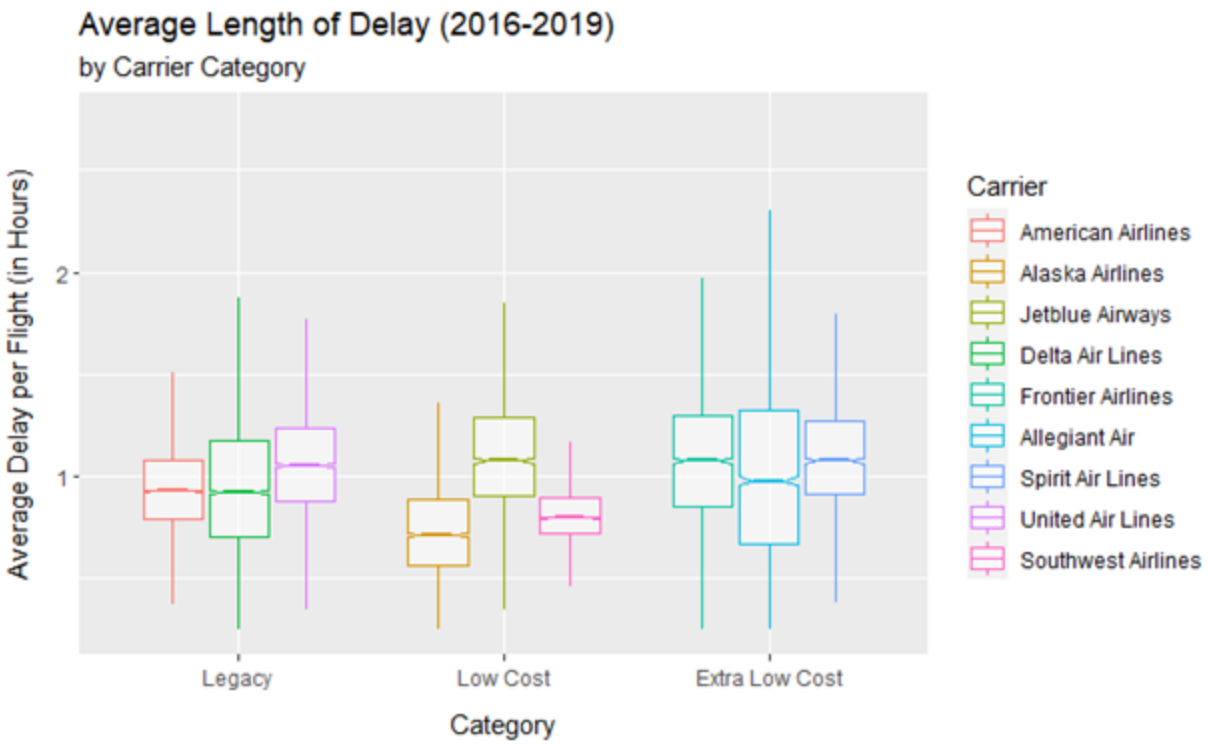


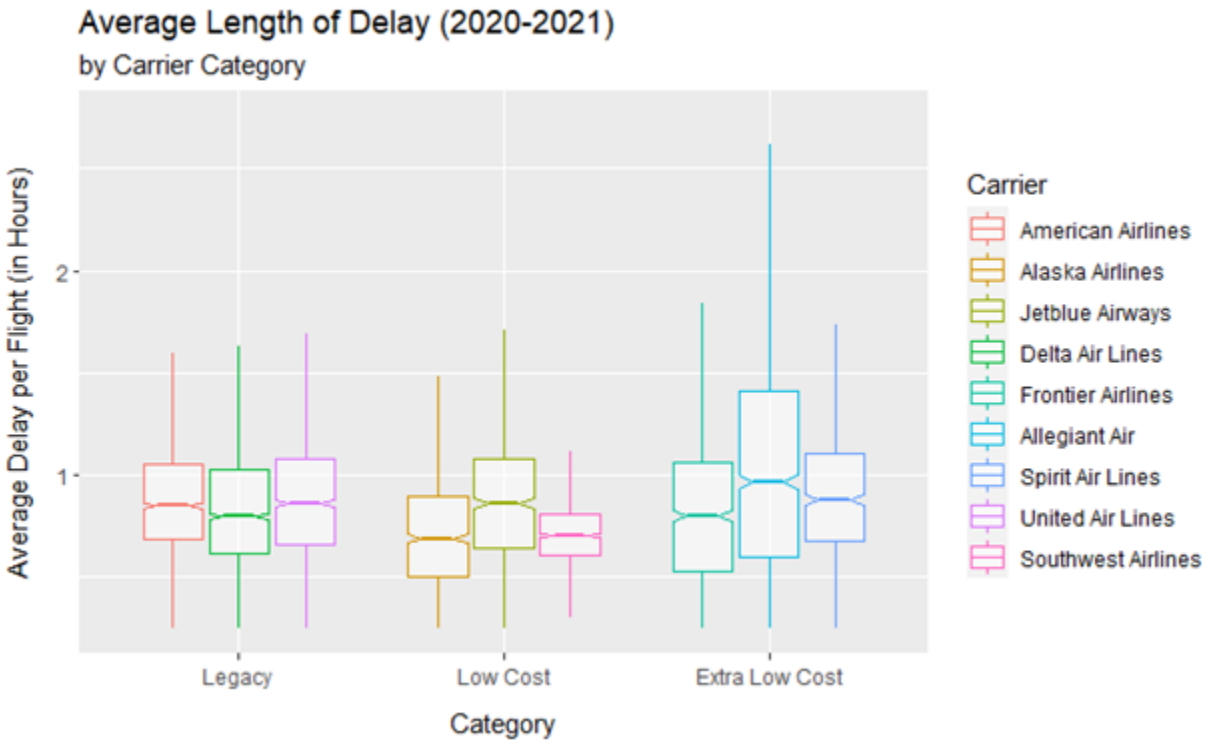



Please note that delays have been shorter in 2020 because the number of flights has been lower due to the COVID-19 pandemic. Similarly, for 2021, we observe significantly shorter delays because the aviation industry has not gotten back to pre-pandemic levels of flight volume and as a result, the duration of delays is low.

The bar graph indicates that total delay time has increased from 2016 to 2019. In the first quarter of 2019, domestic flights were delayed by roughly 220 thousand hours. This number corresponds to 92 thousand days or 25 years. The relative magnitude of the delay reasons in Quarter 1 of each year has not changed from 2016 to 2019. Late aircraft, problems rooting from the carrier, and the national aviation system have been the most prominent reasons behind delays respectively from 2016 through 2019. In comparison, delays due to security reasons have remained relatively stable, except for 2019. In 2021, delays due to carriers are the largest reason for delay by volume, and the second largest reason for delay by percentage of all scheduled flights (beat out by delays due to the National Aviation System by a difference of less than 1%).

**Implications of Expectations set by the Pandemic**




Most delays that passengers experience will fall within a certain range rather than in outlying data points. In these boxplots, outliers showing extreme delays have been removed, allowing a clearer picture of more common delay times. The measure of delay per flight starts at 0.25 hours because a flight is considered on time if it leaves within 15 minutes of its scheduled

departure. Between 2016-2019 (pre-pandemic) and 2020-2021 (post-pandemic), the median average length of delay decreases across all categories of carriers. Notably, the notches on these boxplots are not very large pre-pandemic across all carriers, suggesting that the 95% confidence interval for the median delay covers a smaller range, which is slightly less variable and more predictable. This holds for post-pandemic Legacy Carriers and Southwest Airlines. The other Low-Cost and Extra Low-Cost carriers have slightly bigger notches post-pandemic, indicating that their median delay time is more likely to fluctuate than the previously mentioned airlines. This is further indicated in the longer whiskers of Extra Low-Cost carriers, representing delay times outside the interquartile range (above 75$^{th}$ percentage) for these carriers. Whiskers for Legacy and Low-Cost Carriers do not change noticeably, meaning that carriers of these categories are remaining within the minimum and maximum range that they had prior to the pandemic.

For airlines that had decreased time delays post-pandemic, the average change is a reduction of about 15 minutes. Because passengers returning to air travel see that shorter delay times are possible, even with the resource challenges of the pandemic, carriers will need to meet and maintain customer expectations as recovery progresses. If airlines are able to maintain reduced delays, the time savings could encourage passengers that have the option of other travel modes to fly instead, especially in the short-haul markets we explored in our reports.

## Relationship between Fares, Carriers, and Delays

Average Ticket Price by Carrier

| | Year / Quarter | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2016 | | | | 2017 | | | | 2018 | | | | 2019 | | | | 2020 | | | | 2021 |
| Carrier Na.. | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 |
| Allegiant Air | $87.74 | $81.21 | $81.78 | $89.85 | $86.88 | $85.02 | $81.25 | $87.59 | $87.75 | $85.17 | $87.61 | $91.18 | $84.13 | $81.44 | $82.33 | $87.50 | $88.35 | $81.16 | $82.30 | $86.61 | $85.50 |
| Frontier Air.. | | $86.51 | $83.02 | | | $91.06 | $85.08 | | | | $86.56 | $97.14 | $92.07 | | $87.03 | $89.33 | $80.54 | $88.19 | | $85.36 | $96.97 |
| Spirit Air Li.. | $87.90 | $86.00 | $85.41 | $94.15 | $88.56 | $85.90 | $81.58 | $89.07 | $80.70 | $85.36 | $87.29 | $90.94 | $80.77 | $92.02 | $87.28 | $107.56 | $90.59 | | $80.81 | $80.03 | $79.52 |
| JetBlue Air.. | $125.30 | | $130.80 | $124.49 | $117.05 | $105.87 | $110.88 | $117.50 | $129.91 | $121.59 | $119.07 | $125.80 | $119.71 | $123.03 | $125.92 | $115.60 | $130.51 | $163.10 | $169.68 | $113.82 | $103.86 |
| Southwest .. | $156.81 | $150.48 | $149.64 | $132.59 | $126.12 | $144.33 | $127.08 | $126.27 | $129.24 | $137.71 | $140.65 | $134.33 | $139.41 | $146.80 | $147.03 | $139.37 | $147.98 | $119.65 | $114.15 | $120.77 | $123.16 |
| Alaska Airli.. | $129.86 | $131.63 | $133.02 | $124.50 | $128.43 | $122.98 | $124.08 | $122.75 | $121.83 | $127.19 | $126.90 | $124.98 | $128.07 | $129.77 | $124.59 | $124.56 | $124.94 | $170.24 | $139.55 | $149.63 | $113.81 |
| American Al.. | $283.61 | $268.37 | $257.47 | $253.30 | $231.23 | $284.54 | $247.55 | $256.31 | $275.08 | $280.61 | $276.31 | $285.73 | $281.58 | $296.60 | $278.79 | $288.73 | $276.64 | $212.71 | $169.27 | $188.16 | $189.43 |
| Delta Air Li.. | $286.17 | $280.22 | $265.03 | $233.87 | $239.62 | $260.99 | $196.51 | $226.29 | $258.18 | $253.48 | $230.61 | $221.43 | $253.58 | $220.96 | $183.75 | $228.93 | $220.77 | $237.07 | $202.93 | $206.77 | $207.52 |
| United Air L.. | $314.55 | $284.71 | $254.44 | $276.06 | $262.18 | $271.29 | $220.82 | $226.44 | $243.40 | $240.94 | $227.57 | $279.57 | $326.15 | $260.06 | $277.44 | $319.56 | $301.17 | $205.88 | $153.05 | $156.19 | $186.87 |

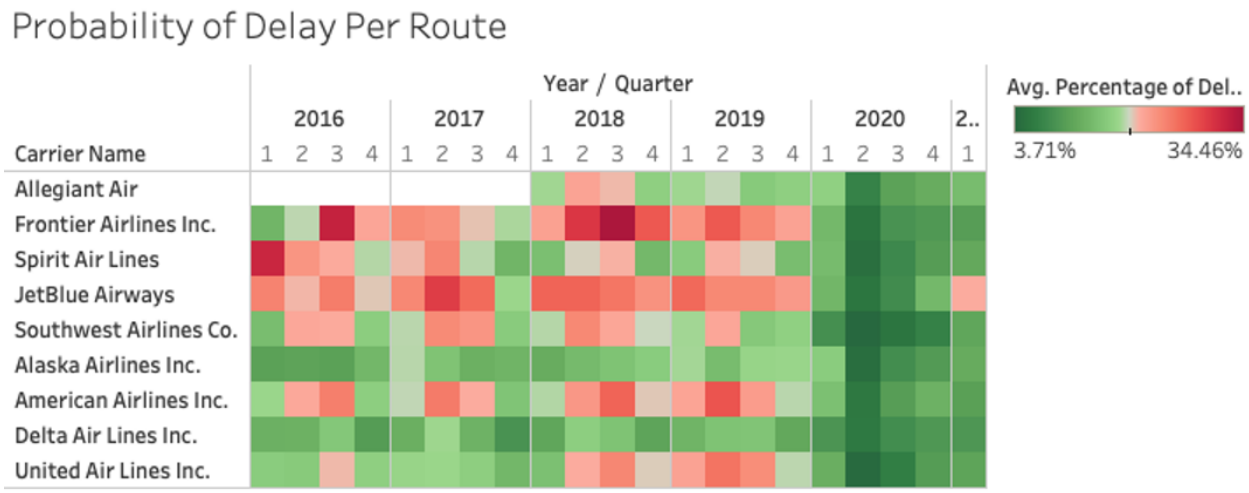
Avg. Average Fare: $79.52 – $326.15

Average of Average Fare broken down by Year and Quarter vs. Carrier Name. Color shows average of Average Fare. The marks are labeled by average of Average Fare.

*Average Ticket Price by Carrier* above shows the average ticket price for each airline. When the price exceeds $200, which is close to the total average fare of $192.02, the color will cross over from green to red.

From the table, we found price information is consistent, with some changes in the year 2020. The table above gave more explicit price information that could be helpful for potential customers to determine whether the prices in their search result is fair. It is also worth noting that Legacy Carriers' average fare is more than twice the average fare for Extra Low-Cost Carriers.

Customers might ask which airline is best in each when they are most concerned about on-time performance.



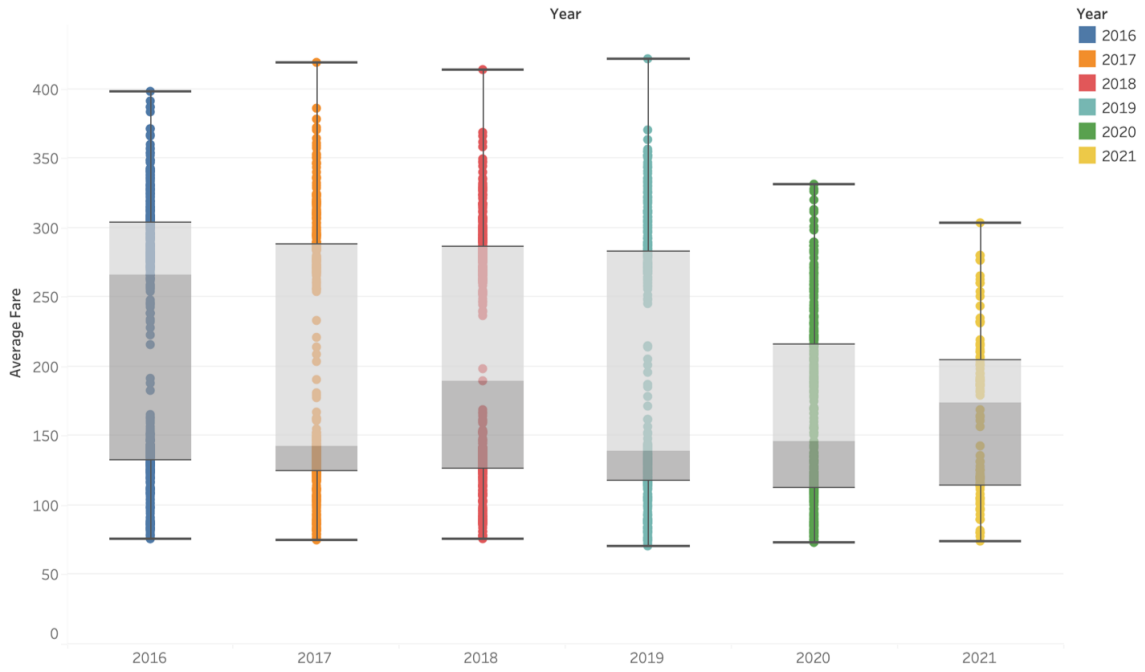

Average of Percentage of Delayed Flights (color) broken down by Year and Quarter vs. Carrier Name.

To break it down by carrier categories using the heatmap, *Probability of Delay Per Route*, we observed the following:

- For Extra Low-Cost carriers, Allegiant is observed to perform well however lacking two years of data. Spirit AirLines placed second.
- For Low-Cost carriers, Southwest appears to have a better on-time performance compared to Jetblue. Alaska Airlines has an excellent on-time performance; however, many travelers may not have this option as it is a regional airline.
- For Legacy carriers, Delta has the best on-time performance, exceeding other carriers by a significant amount as it does not have any “red” squares.

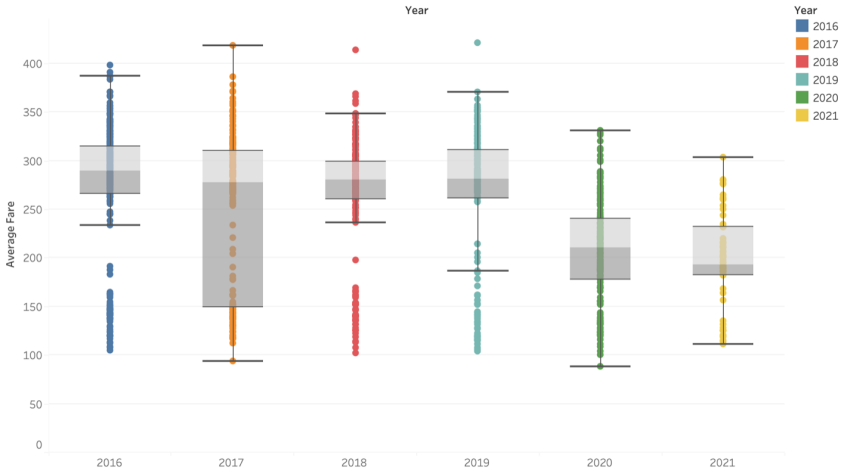
Overall, you could see that the Extra Low-Cost carriers tend to have more severe delays; Delta and Alaska have never exceeded the average probability of delay in the past five years, making them the best airlines for on-time performance.

**Exploring Average Airfares**

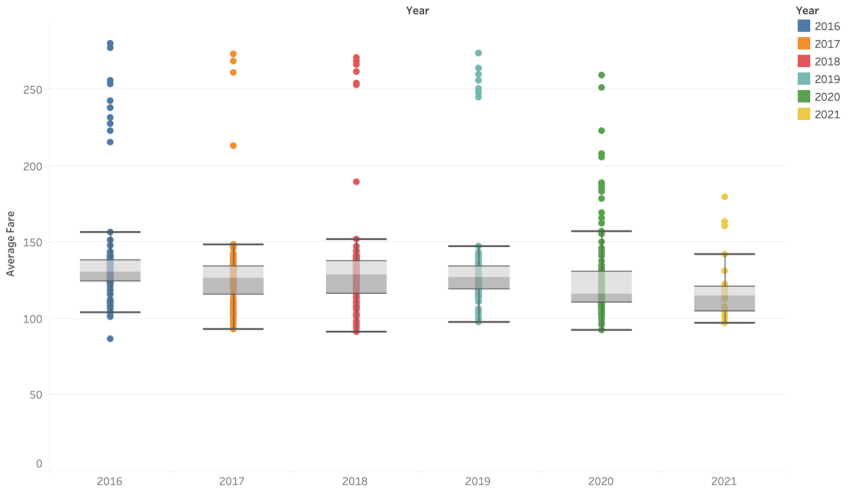



When we look at the distribution of average airfare over the years, we can clearly see that the range of average airfare is becoming more convergent in 2020 and 2021 than the previous years. By seeing the third quartile and the maximum value of average airfare decreasing, we could infer that some airlines have been reducing their price since the outbreak of the pandemic.

Legacy Airlines



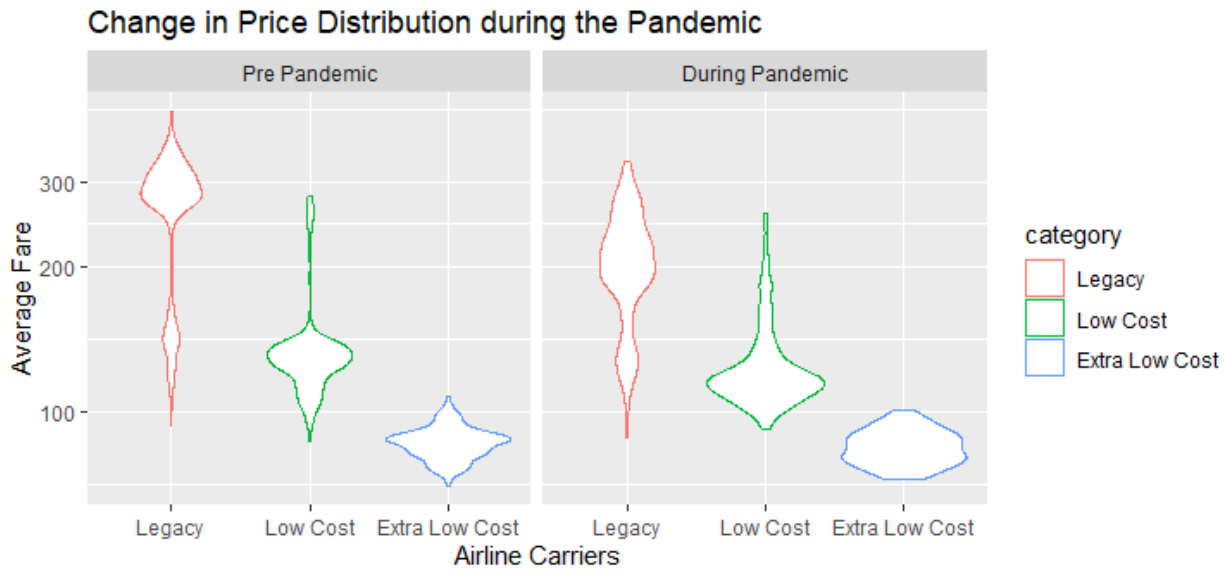

Low-Cost Airlines




Extra-Low-Cost Airlines

After checking the distribution of average airfare for each category, we can attribute the majority of this change of average airfare to the reduction of airfare of Legacy airlines. This is likely due to the fact that there are less business travellers since the pandemic started. Legacy airlines faced some difficulties selling their tickets in this situation and needed to lower their airfare.

**Effect on Airfare Distribution**




During the pandemic, due to travel restrictions, there has been a sharp decline in demand for air travel. The faceted violin plot above shows the reflection of this decrease in demand on average airfare. Both the Legacy and Low-Cost carriers have experienced a downward shift in their prices. While their minimum prices before the pandemic have barely changed, the distribution of their average fare shifted towards the lower end of their price range.

Given there will be more challenges such as pandemics in the future, it would be wise for carriers to be prepared for the demand shocks which will also be reflected as changes in average prices. Going forward, airline companies should expect a decrease in their revenue at a relatively unpredictable point in time and should adjust their operations, especially those requiring investment, accordingly. In other words, they should be prepared to postpone such activities.

## Business Value, Suggestions, and Conclusion

The observations and recommendations in this report can help airline companies further grow, remain competitive against their competitors and also prepare for a future with shocks,

such as the COVID-19 pandemic. We would like to reiterate that airlines in different categories have different market conditions and it is advisable to remain mindful of the differences based on the categories.

On the demand side of the airline business, we have seen that airline ridership has slowed down even before the pandemic. This suggests that airlines needed to make strategic decisions to increase the demand without the presence of a shock.

We observed a high variance in average fare for Extra Low-Cost carriers with relatively lower ridership (measured in number of passengers). This means that Extra Low-Cost carriers are operating in a market with limited demand. Since these airlines appeal to passengers who would like to fly on a budget, they must keep their operating costs low and cannot realistically offer better additional services than airlines in other categories. Each airline within this category, however, has the opportunity to court one another's passengers. We have also observed that these airlines cannot further lower their prices to respond to demand shocks. Therefore remaining competitive through increased service quality while staying as Extra Low-Cost airlines and also implementing creative marketing strategies such as bundle selling are more feasible strategies.

It can be seen that Low-Cost carriers are more distinct among similarly-priced competitors than carriers within other categories. Southwest is uniquely large and enjoys earning revenue in its ridership and average price outliers. Southwest should innovate in case other Low-Cost airlines in the same category decide to compete for Southwest market share.

In terms of handling delays, it is worth highlighting Legacy airlines. We have seen that when it comes to delays per flight, Delta performs better than American Airlines. However, on average, American Airlines charges higher prices. American Airlines could either improve its operations to handle delays better or decrease their prices. Otherwise, there is an opportunity for both Delta and United to increase their prices and still offer better service than American Airlines.

We would like to conclude by emphasizing that the pandemic has changed the target market conditions beyond ridership and price range. We have observed the distribution of prices have shifted and customer expectations have also changed. These changes signal that target markets may have shifted. Airlines should consider redefining their target market and their positions within the market as they are proceeding with strategic decisions. Our visualizations helped inform suggestions for airlines to take action on. Overall, our recommendations in this report can help airline companies further grow and remain competitive within their industry. Our insights may also help airlines in preparing for future prediction in air travel where shocks similar to the pandemic may occur. It is our hope that our reports provided useful information for a variety of reasons and that the implementation of some of our suggestions will gain positive results for the airline companies.

## Reflection on Our Process

Our team was initially frustrated that we couldn't find a dataset that was easy to work with. The various methods of dataset searching could be contradicting. On the one hand, we wondered if it would be more efficient for us to aimlessly look for a dataset then consider what stories we could tell based on the data. On the other hand, perhaps it could be more helpful in meeting report requirements if we were to come up with specific questions first then search for data based on our guidelines.

After spending hours looking for the desired dataset, it is not surprising that no dataset would completely satisfy our desires in what we want in our data. One of the things is that we wish our datasets allowed us to incorporate geospatial route data into our analysis regarding flight delays. Our team would have enjoyed showing information on maps. Incorporating geographical information could have further provided insights to the general audience and allowed the team to learn more in this area. The problem with not being able to do that was not because there wasn't any geographical information in any dataset, but rather because the details weren't consistent across datasets. In the future, rather than coming up with the narrative first, our team might consider exploring datasets with visualization tools first, then brainstorm based on those visualization samples. This would allow us to be creative and make sure that we can fit our data into the stories we want to tell.

It impresses us each time when we reflect on how frustrated we were and what a long way we've come. Our initial three-hour meeting in the early days of the semester was shorter compared to some later meetings, but it still felt like the longest because we were aimlessly and stubbornly yearning and looking for that "perfect" dataset. It was a meeting that started a fire in our hearts which set us up to work hard and work well as a team. Team TEN will never forget the midnight chilled air, the IBS picnic, and the early morning PowerPoint workshops we've experienced together. Since these hard-working two months, our team gained data identifying skills, trained on developing insights, and, most importantly, formed lasting friendships.

## Works Cited


Bureau of Transportation Statistics. *Mid-December Airline Industry Employment Up 16,000 Employees from Mid-November*. 9 February 2021. <https://www.bts.gov/newsroom/mid-december-airline-industry-employment-16000-employees-mid-november>.

Federal Aviation Administration. "The Economic Impact of Civil Aviation on the US Economy." January 2020. <https://www.faa.gov/about/plans_reports/media/2020_jan_economic_impact_report.pdf>.

Bureau of Transportation Statistics, "Flight Delays at a Glance." <www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp?20=E>.

Department of Transportation Office of the Assistant Secretary for Aviation and International Affairs. "Consumer Airfare Report: Table 5 - Detailed Fare Information For Highest and Lowest Fare Markets Under 750 Miles: Department of Transportation - Data Portal." Data.transportation.gov, 23 July 2021,

<data.transportation.gov/Aviation/Consumer-Airfare-Report-Table-5-Detailed-Fare-Info/bkh6-tj42>.

International Civil Aviation Organization. *Economic Impacts of COVID-19 on Civil Aviation*. n.d. 22 September 2021. <https://www.icao.int/sustainability/Pages/Economic-Impacts-of-COVID-19.aspx>.

Federal Aviation Administration. "Cost of Delay Estimates." 8 July 2020. <https://www.faa.gov/data_research/aviation_data_statistics/media/cost_delay_estimates.pdf>.