

Cleveland Travel Data

Flights in and out of Cleveland Hopkins Unrestrictedational
Airport

Where is the data coming from?

- Most of the data was downloaded as CSVs from the Bureau of Transportation Statistics
 - These were downloaded month by month and then appended together
 - These were then cleaned to reflect only data regarding flights to and from CLE
 - The resulting data frame was 349795 rows x 24 columns
 - Not every column was used to answer every question, so additional cleaning was done for each set of data
- Google geocoding API was used to convert destination cities into Latitude and Longitude values to then use with gmaps to mark the destinations on a map
- Weather data was obtained from WorldWeatherOnline historical weather data API

Data Cleaning Examples

```
In [4]: jan_2019_data=pd.read_csv(jan_2019)
feb_2019_data=pd.read_csv(feb_2019)
mar_2019_data=pd.read_csv(mar_2019)
apr_2019_data=pd.read_csv(apr_2019)
may_2019_data=pd.read_csv(may_2019)
jun_2019_data=pd.read_csv(jun_2019)
jul_2019_data=pd.read_csv(jul_2019)
aug_2019_data=pd.read_csv(aug_2019)
sep_2019_data=pd.read_csv(sep_2019)
oct_2019_data=pd.read_csv(oct_2019)
nov_2019_data=pd.read_csv(nov_2019)
dec_2019_data=pd.read_csv(dec_2019)
```

```
In [5]: jan_feb_2019=jan_2019_data.append(feb_2019_data)
feb_mar_2019=jan_feb_2019.append(mar_2019_data)
mar_apr_2019=feb_mar_2019.append(apr_2019_data)
apr_may_2019=mar_apr_2019.append(may_2019_data)
may_jun_2019=apr_may_2019.append(jun_2019_data)
jun_jul_2019=may_jun_2019.append(jul_2019_data)
jul_aug_2019=jun_jul_2019.append(aug_2019_data)
aug_sep_2019=jul_aug_2019.append(sep_2019_data)
sep_oct_2019=aug_sep_2019.append(oct_2019_data)
oct_nov_2019=sep_oct_2019.append(nov_2019_data)
data_2019=oct_nov_2019.append(dec_2019_data)
```

```
In [7]: data_2019_cle=data_2019.loc[(data_2019["ORIGIN"]=="CLE")|(data_2019["DEST"]=="CLE")]
```

```
In [ ]: output="../../cleaned_csvs/2019_data.csv"
data_2019_cle.to_csv(output)
```

```
In [2]: data_total=pd.read_csv("../cleaned_csvs/total_data.csv", parse_dates=["FL_DATE"])
data_total_reduced=data_total[["FL_DATE", "ORIGIN", "DEST", "DEST_CITY_NAME"]]

/Applications/anaconda3/lib/python3.7/site-packages/IPython/core/interactiveshell.py:3063:
DtypeWarning: Columns (17) have mixed types.Specify dtype option on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

```
In [3]: #only flights leaving cleveland
data_total_departures=data_total_reduced.loc[data_total_reduced["ORIGIN"]=="CLE"]
data_total_departures
```

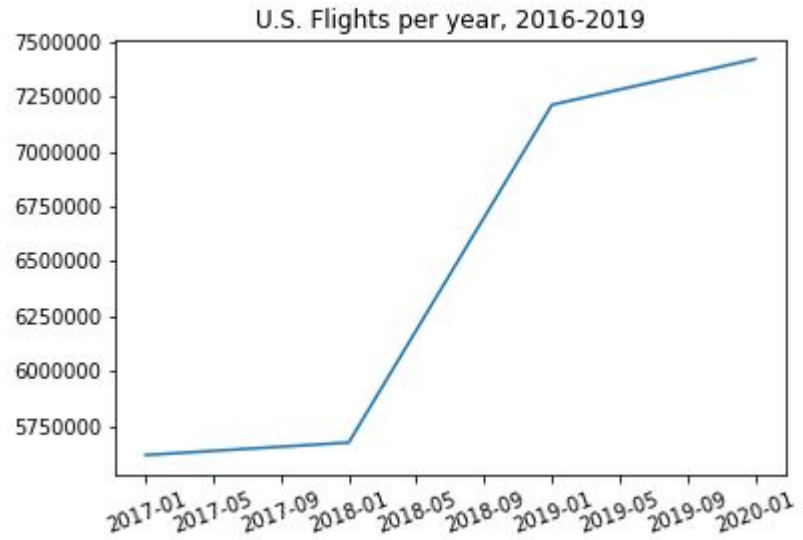
Out[3]:

	FL_DATE	ORIGIN	DEST	DEST_CITY_NAME
30	2016-04-01	CLE	DFW	Dallas/Fort Worth, TX
31	2016-04-02	CLE	DFW	Dallas/Fort Worth, TX
32	2016-04-03	CLE	DFW	Dallas/Fort Worth, TX
33	2016-04-04	CLE	DFW	Dallas/Fort Worth, TX
34	2016-04-05	CLE	DFW	Dallas/Fort Worth, TX
...
349788	2020-03-11	CLE	ORD	Chicago, IL
349790	2020-03-11	CLE	DEN	Denver, CO
349792	2020-03-11	CLE	ORD	Chicago, IL
349793	2020-03-11	CLE	SFO	San Francisco, CA
349794	2020-03-11	CLE	ORD	Chicago, IL

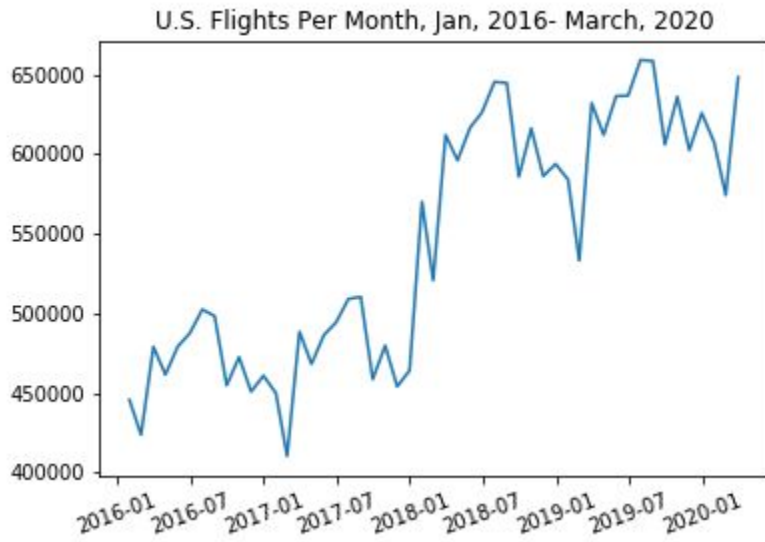
Total Flights Over Time

Our data starts January 2016 and goes to March 2020.

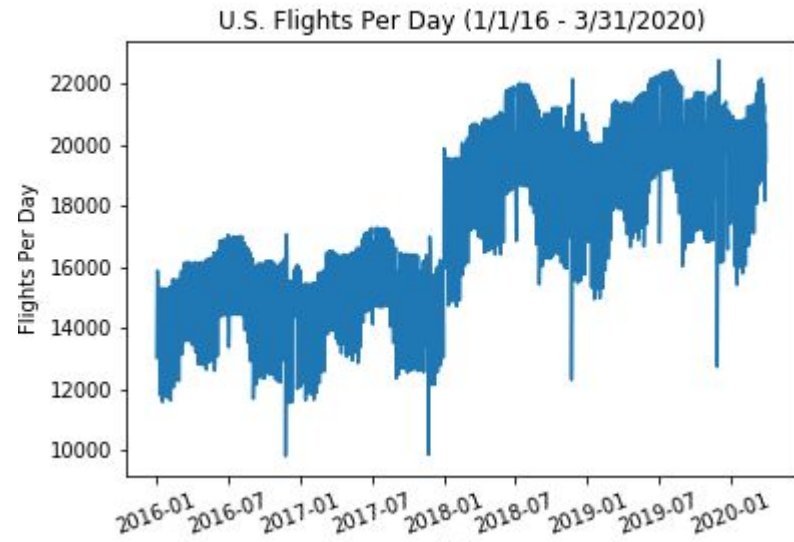
Over this period time we, see an increase of domestic flights, from 5.75 million domestic flights in 2016 to almost 7.5 million by the end of 2019, a 30% increase.



We can see this reflected in flights per month.



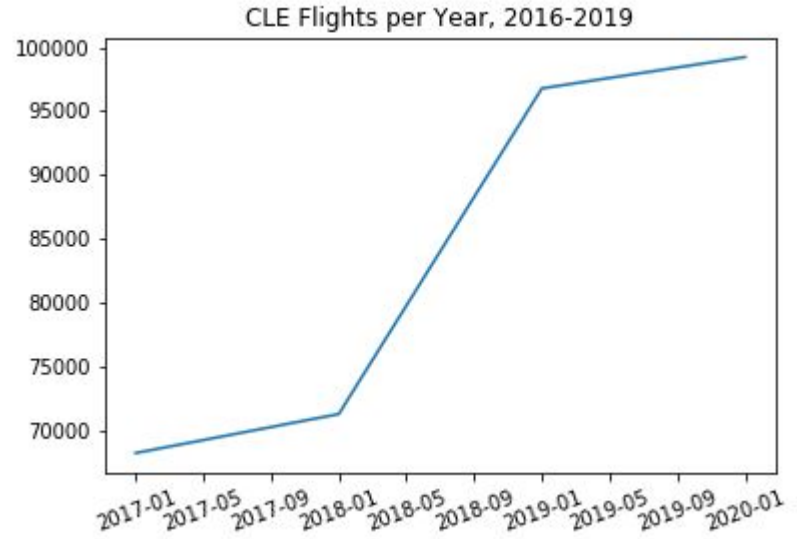
And flights per day.



Flights to and from CLE

CLE airport data runs parallel from this trend

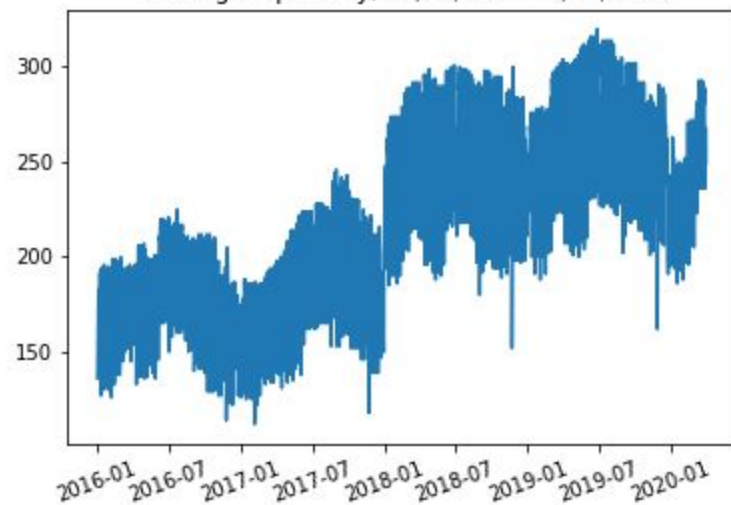
	U.S. Flights	CLE Fights	% of overall flights
Total Flights (2016-2020)	27,757,609	358,033	1.29%
2020	1829847	22539	1.23%
2019	7422037	99228	1.34%
2018	7213446	96774	1.34%
2017	5674621	71272	1.26%
2016	5617658	68220	1.21%
Year w/ Most Flights	2019	2019	
/least (excluding 2020)	2016	2016	



CLE flights per month, 01/01/2016 -03/31/2020



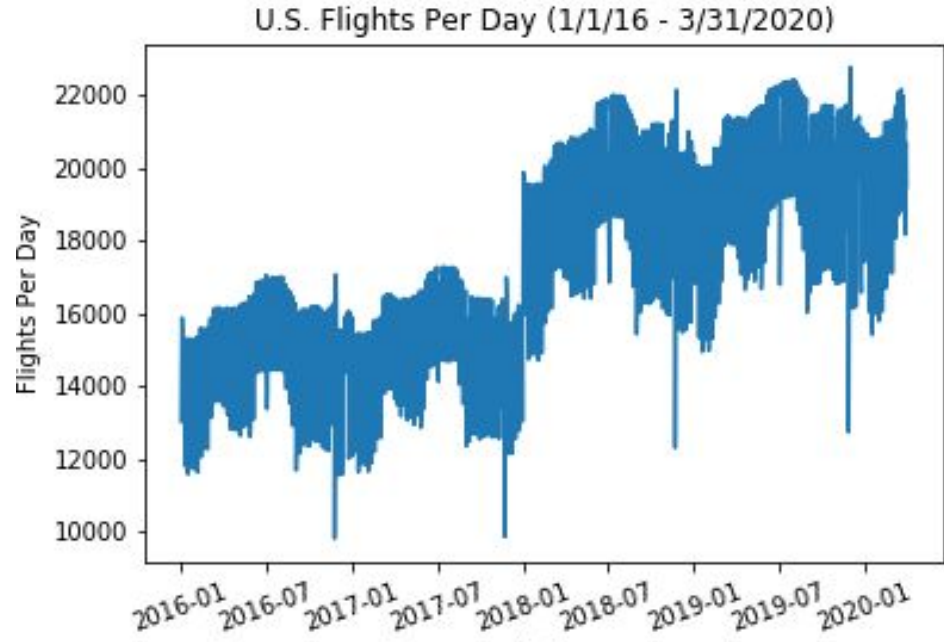
CLE Flights per Day, 01/01/2016-03/31/2020



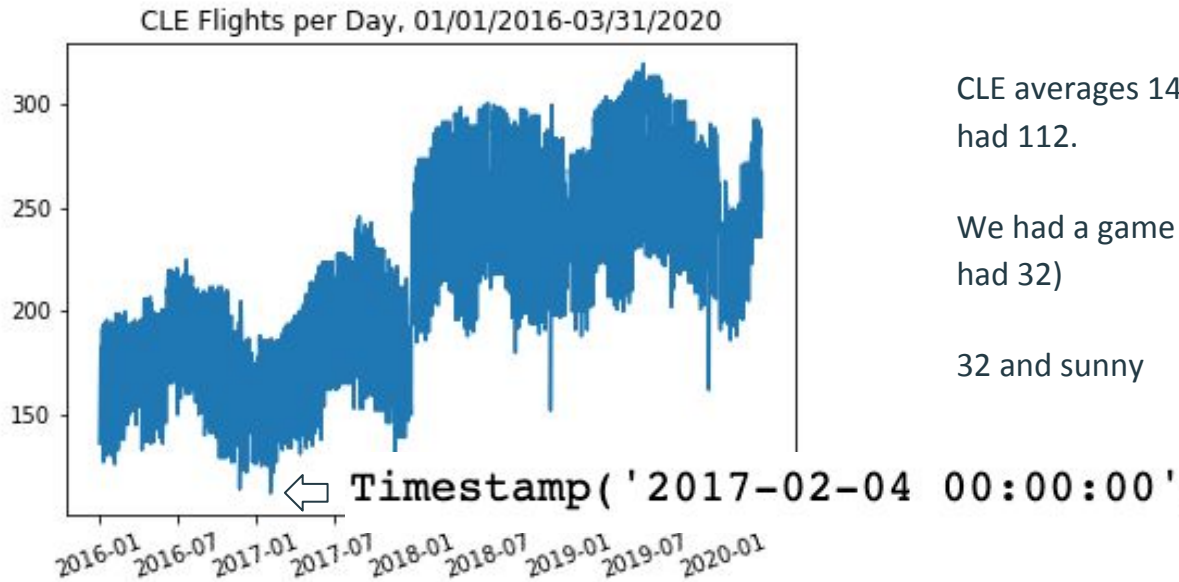
Interesting things to note

You can actually see the drop in travel on Thanksgiving Day each year

DATE	
2016-12-31	2016-11-24
2017-12-31	2017-11-23
2018-12-31	2018-11-22
2019-12-31	2019-11-28
2020-12-31	2020-01-18



CLE has one lower day...

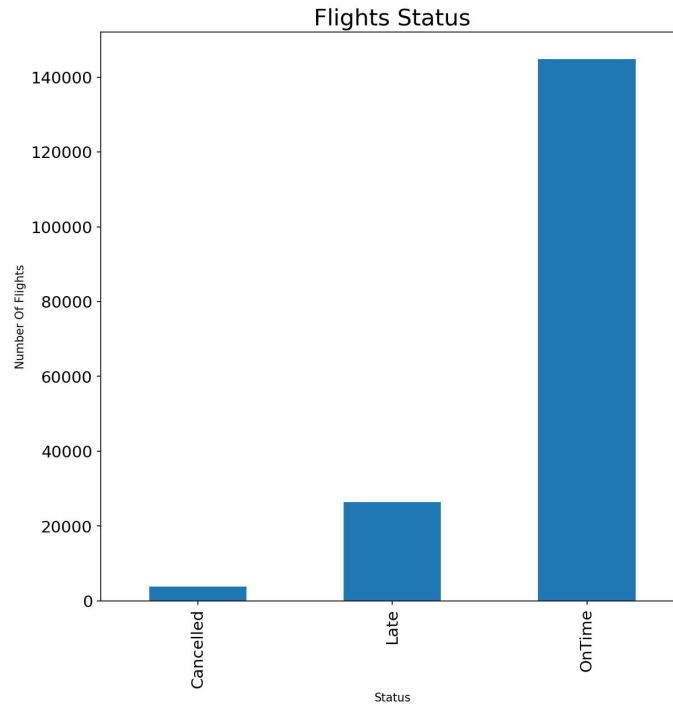
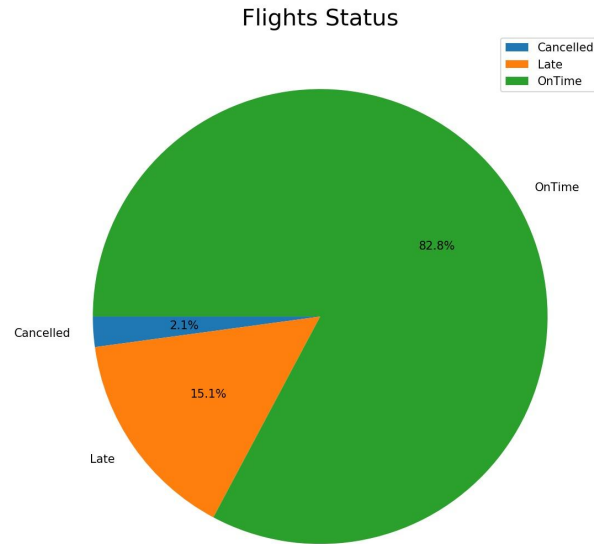


CLE averages 148 flights per day in 2017, that day had 112.

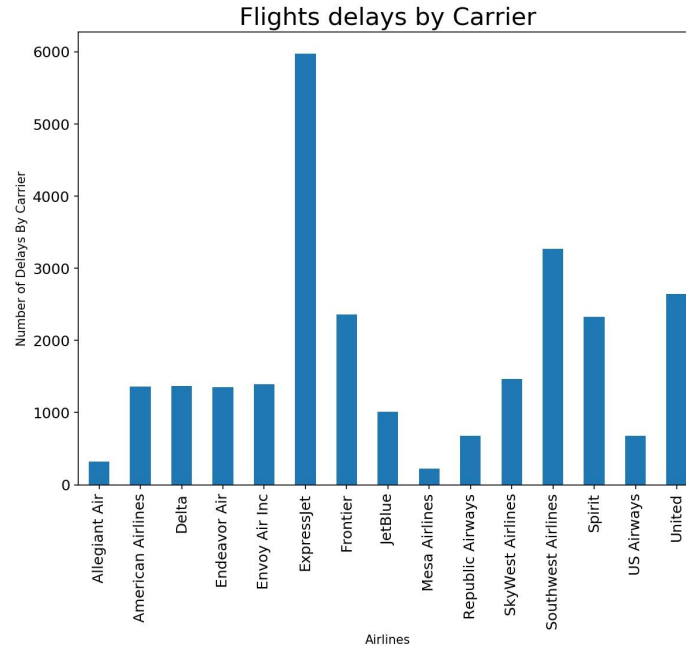
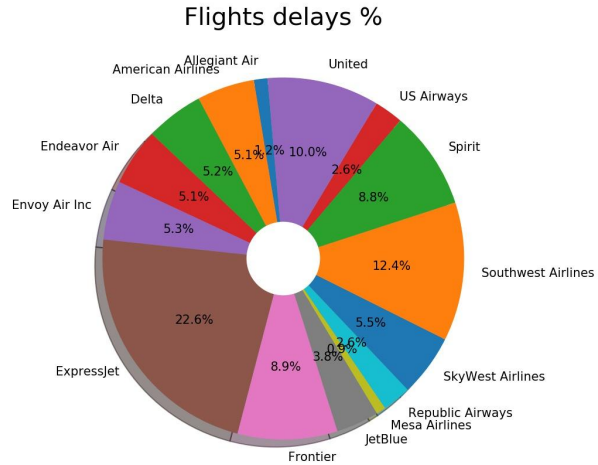
We had a game against the Knicks (we won, LeBron had 32)

32 and sunny

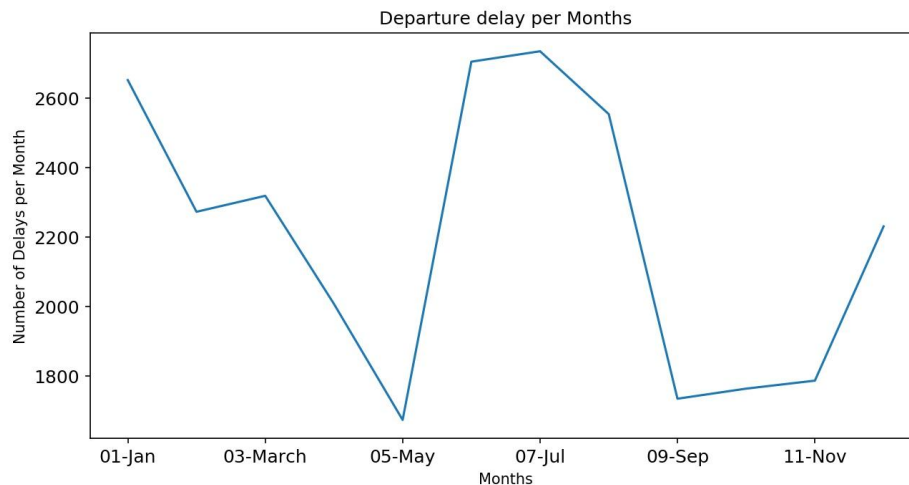
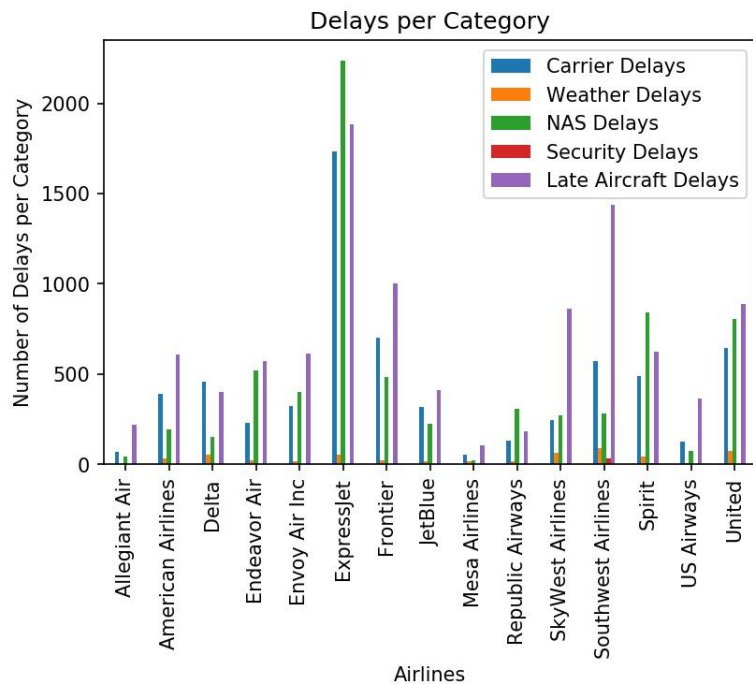
What about Delays



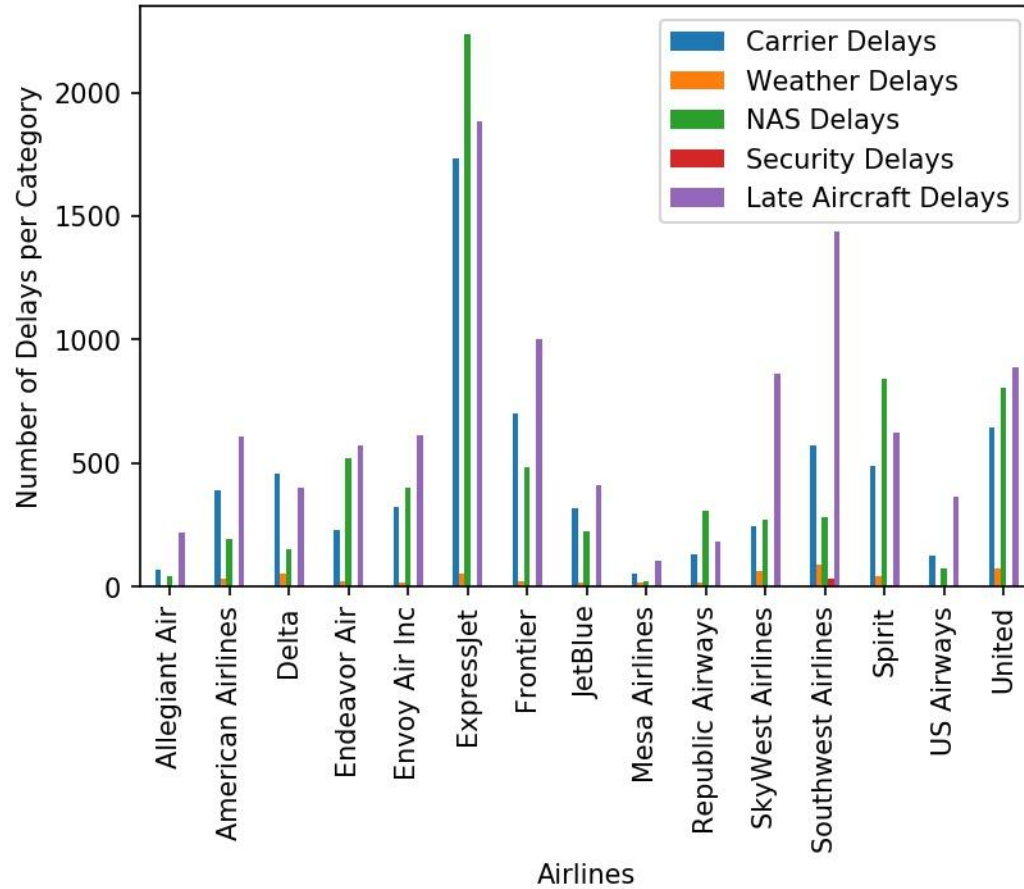
What about Delays?



Delay by category and month



Delays per Category



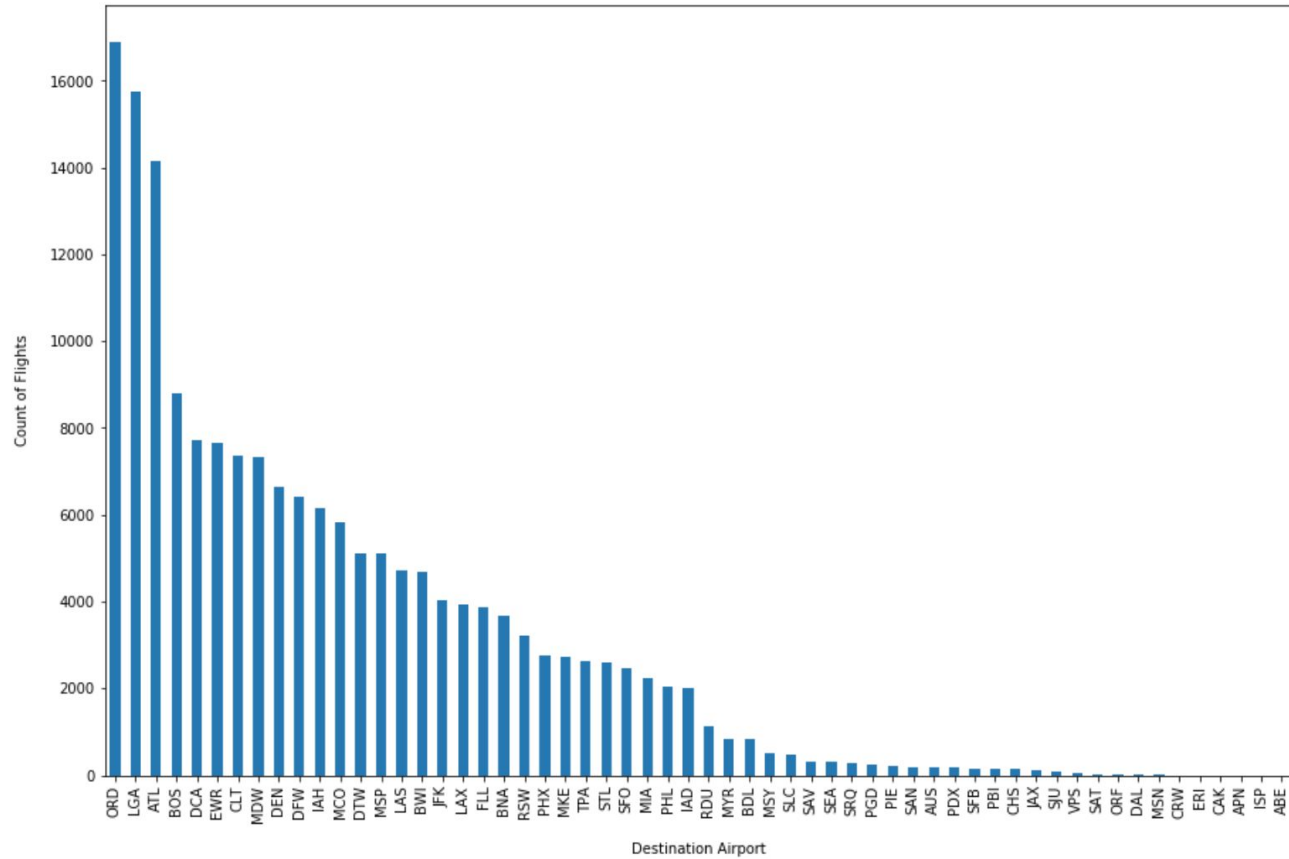


But Where are We Going?

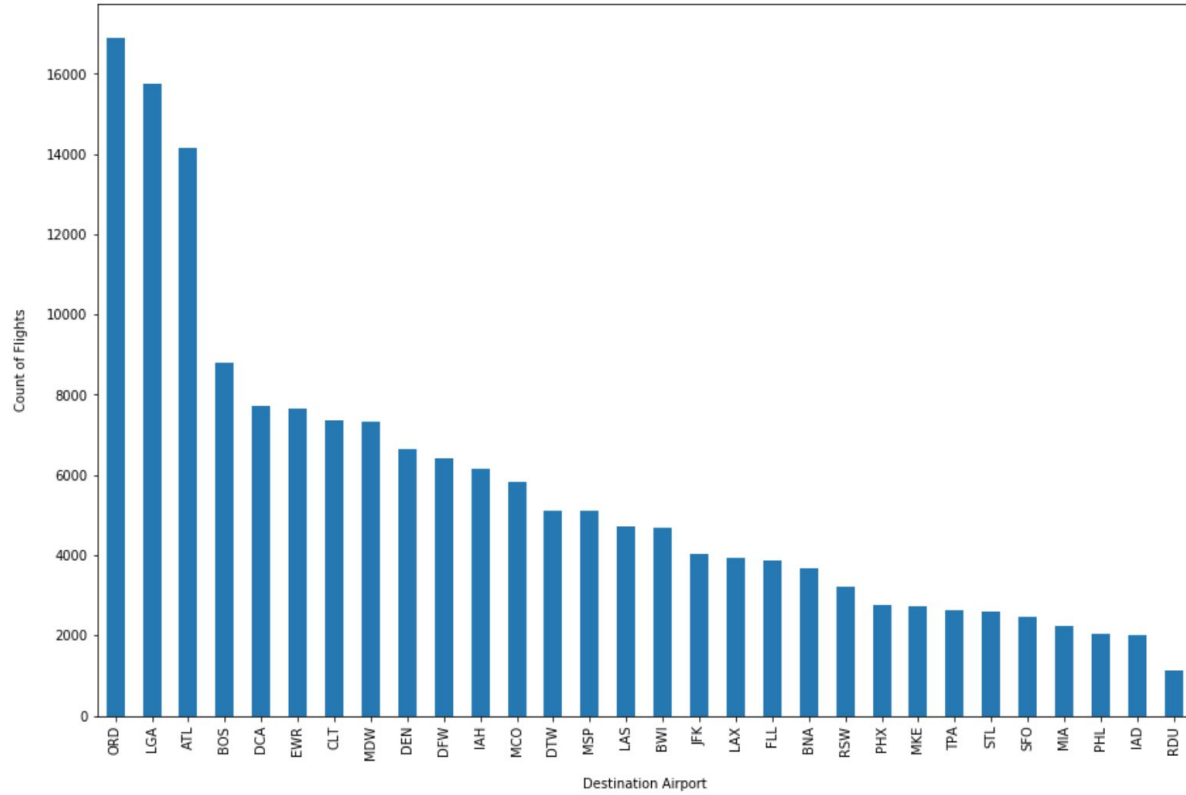
Over to Ruthy, with our destinations...



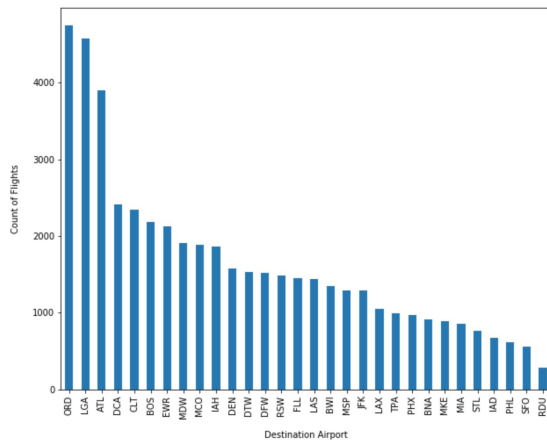
Count of Flights from CLE by Destination 2016-2020



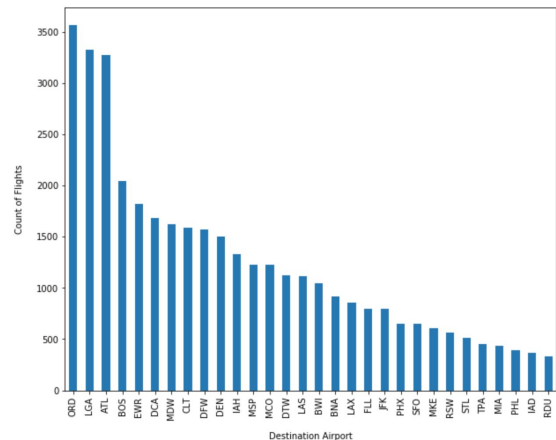
Count of Flights from CLE by Destination 2016-2020 (30 Most Popular Destinations)



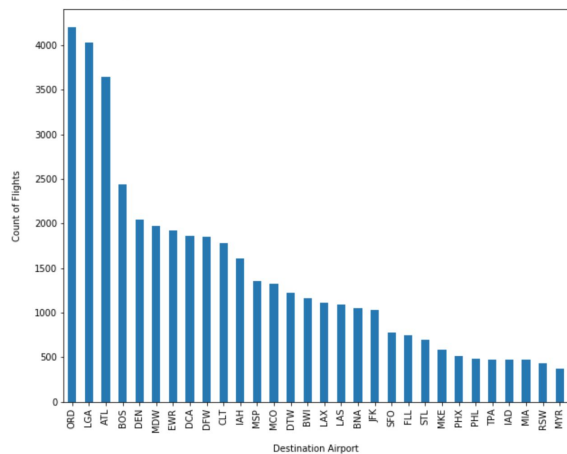
Count of Flights from CLE by Destination Winters 2016-2020 (30 Most Popular Destinations)



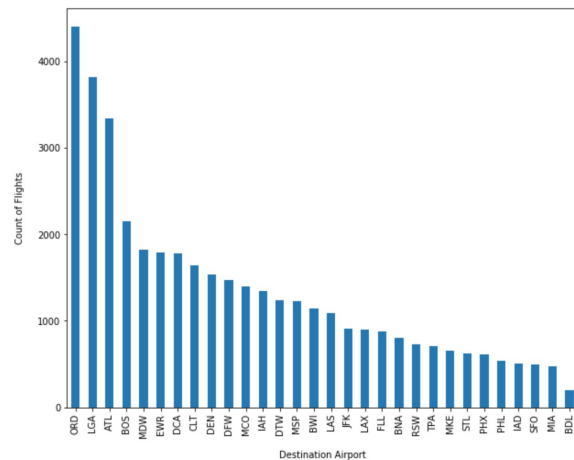
Count of Flights from CLE by Destination Spring 2016-2020 (30 Most Popular Destinations)



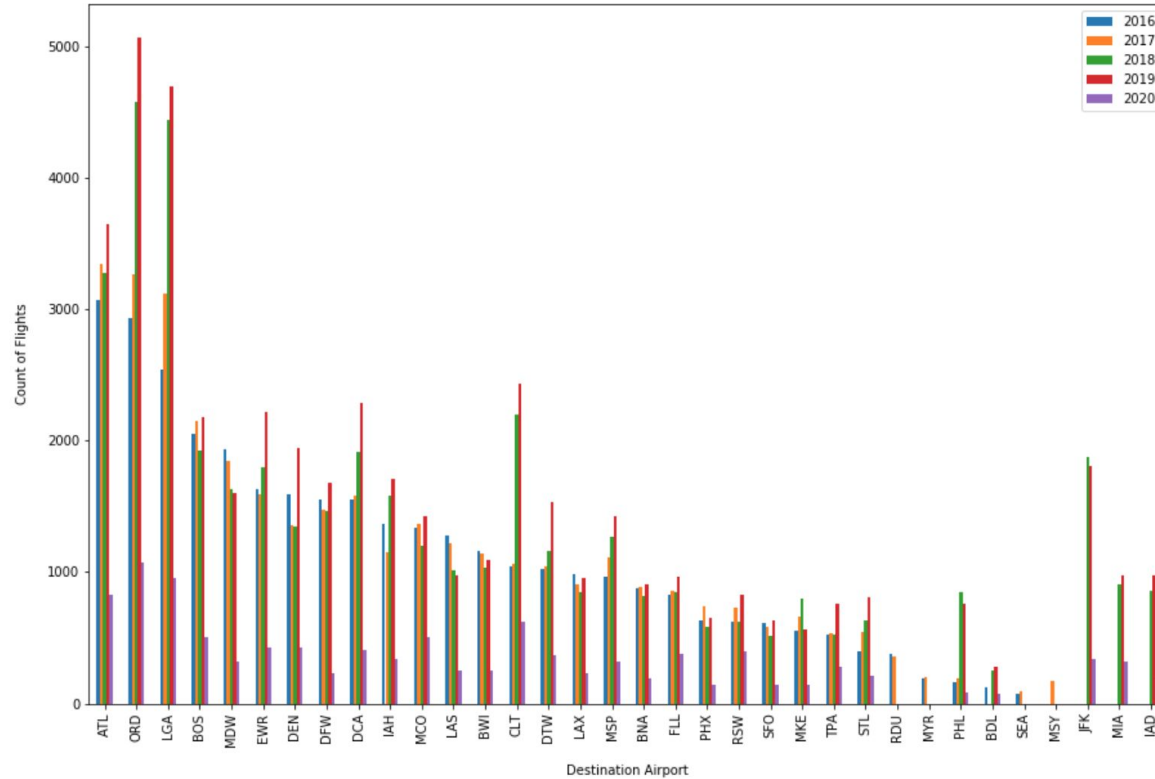
Count of Flights from CLE by Destination Summers 2016-2020 (30 Most Popular Destinations)



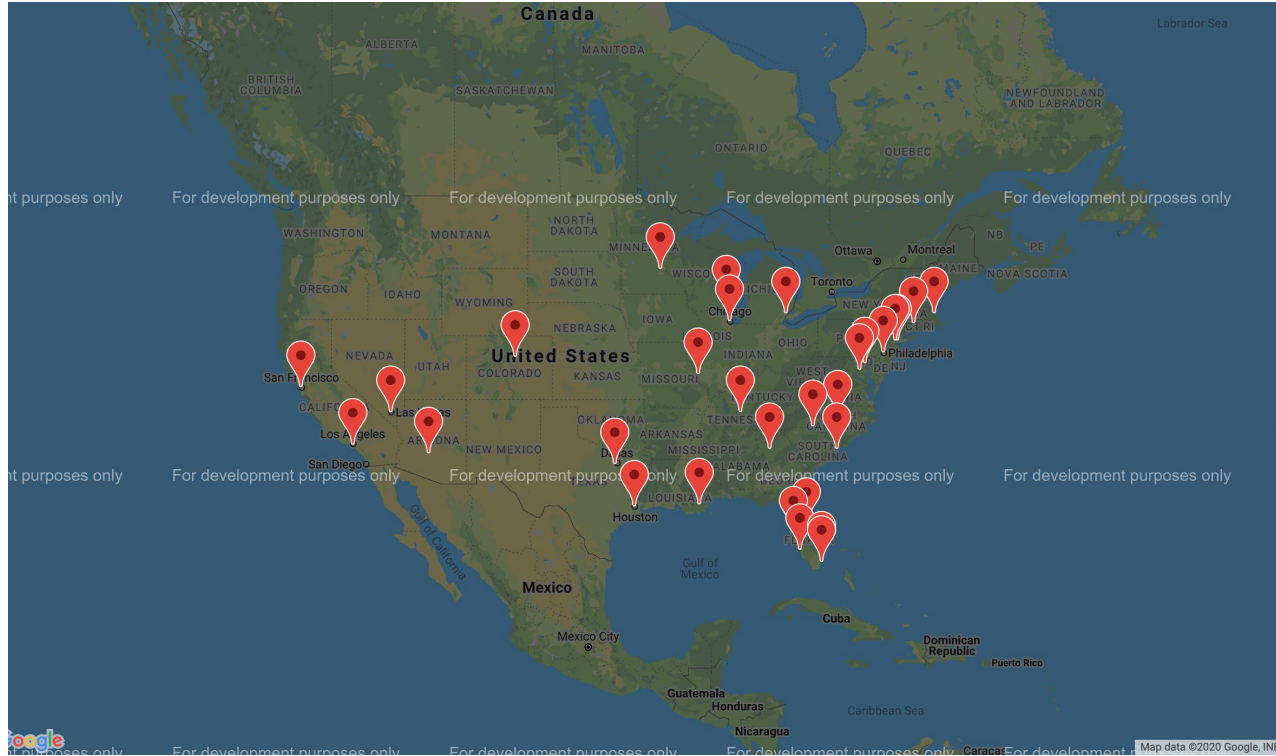
Count of Flights from CLE by Destination Falls 2016-2020 (30 Most Popular Destinations)



Count of Flights from CLE by Destination 2016-2020



Top 30 Flight Destinations from CLE

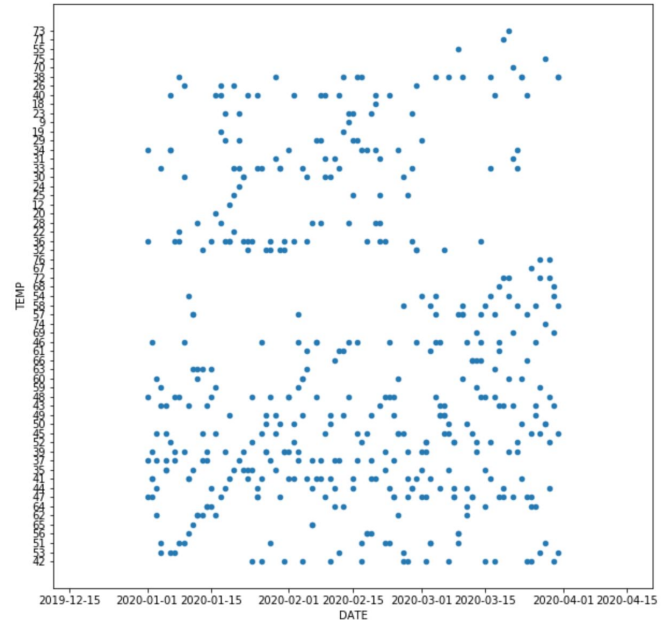


Takeaways!

- There is minimal seasonal and yearly change in the top three destinations: ATL, ORD, and LGA
 - These airports are major hubs for a number of airlines which suggests that a good number of the flights from CLE into these airlines are carrying passengers that then connect to other destinations
- Within the other top destinations there is more variation within flight destination
 - For example: In the summers MYR (Myrtle Beach Airport) is one of the top 30 most popular destinations but not during any other season
 - For example: In 2018 and 2019 there were more flights going to CLT (Charlotte Douglass International Airport) than in 2016,17, and 20.
- However, even with small shifts, the top 30 most popular flying destinations for travelers from CLE are relatively stable across season and year.

Bonus: Temperature Analysis of Top Destinations

- We looked into the temperatures in the top 5 destination locations for 2020 to see if there were any trends.
- The lower quartile of temperatures is: 37.0
- The upper quartile of temperatures is: 52.0
- The interquartile range of temperatures is: 15.0
- The the median of temperatures is: 44.0
- Values below 14.5 could be outliers.
- Values above 74.5 could be outliers.
- There did not seem to be any trends in destination temperature, the graph of date of flight vs. destination temperature showed high variability.



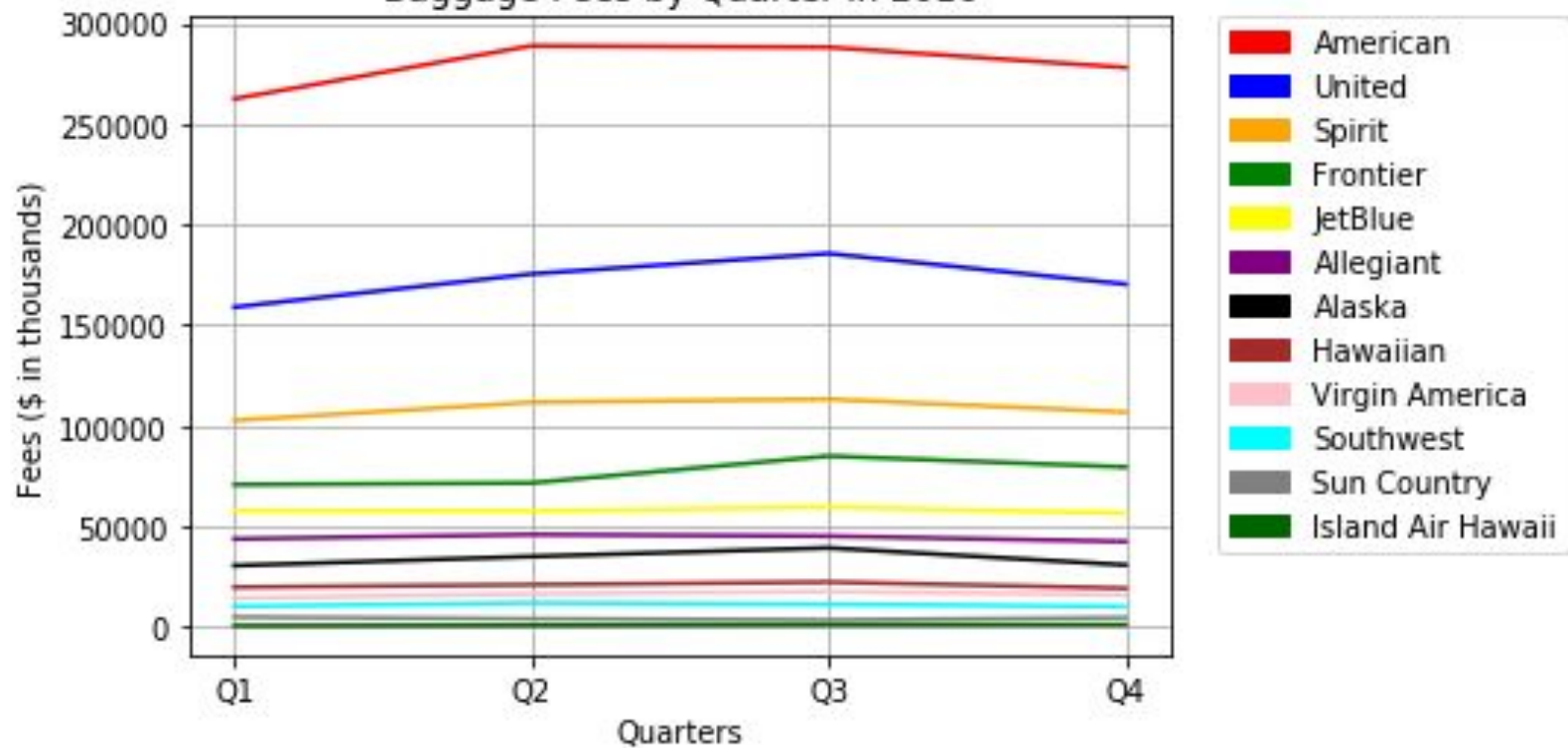


What are we paying in baggage fees? Who is charging the most?

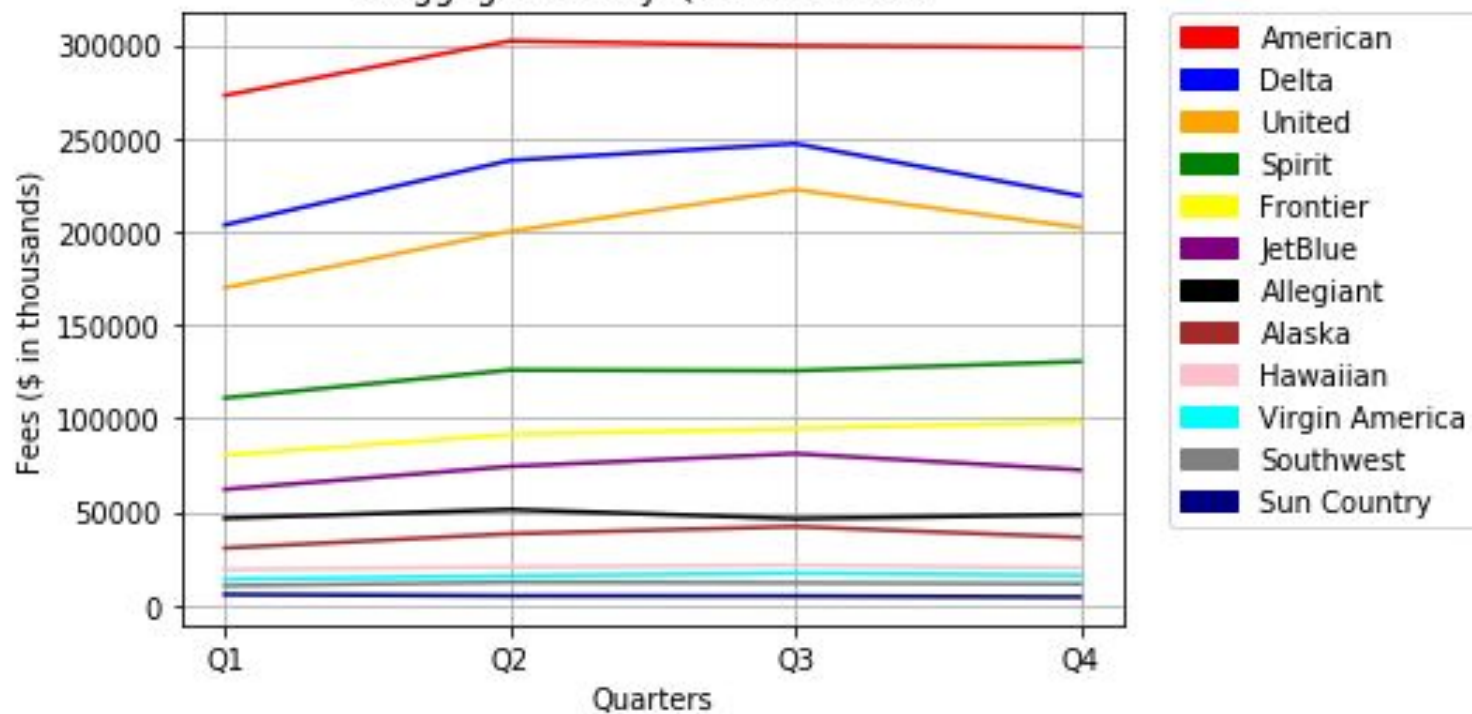
Fees over 2016-2019



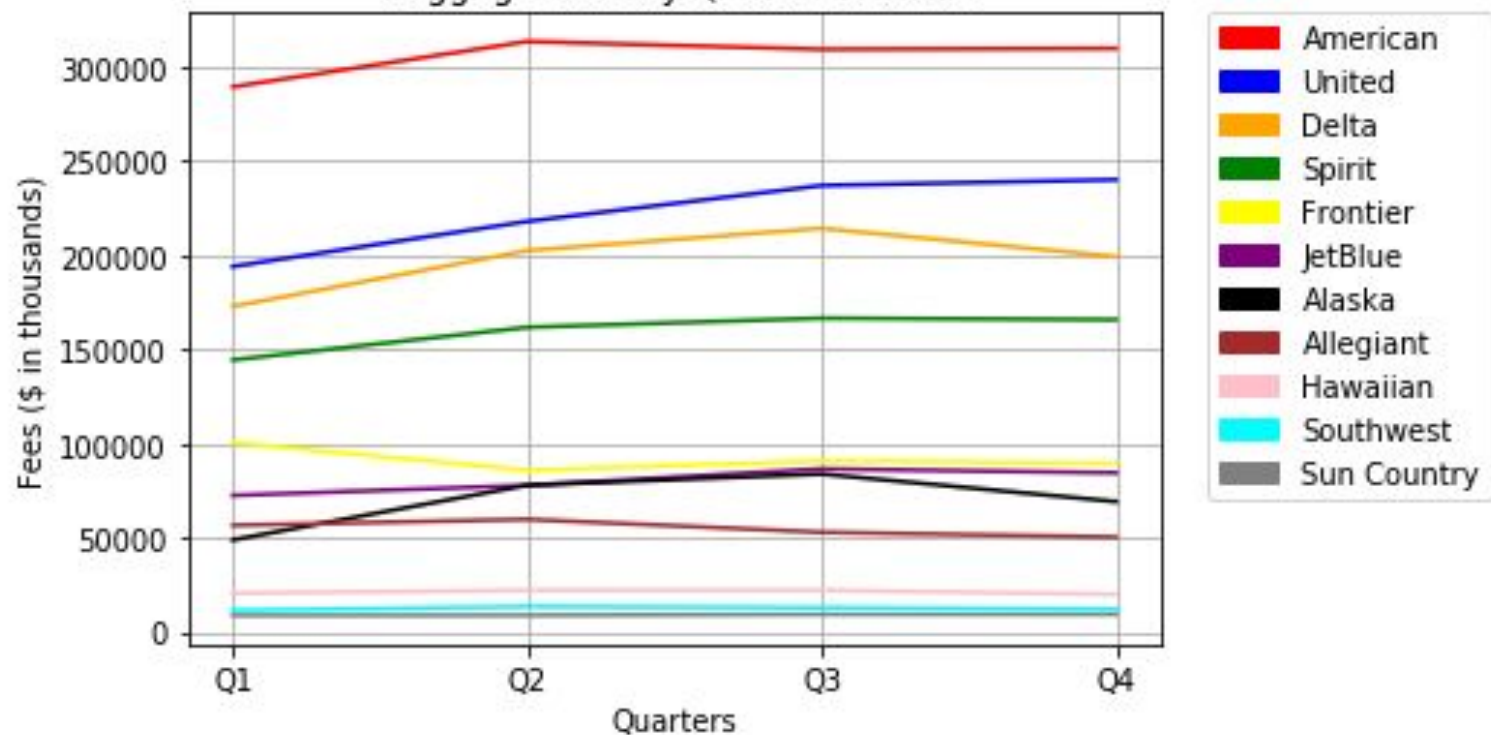
Baggage Fees by Quarter in 2016



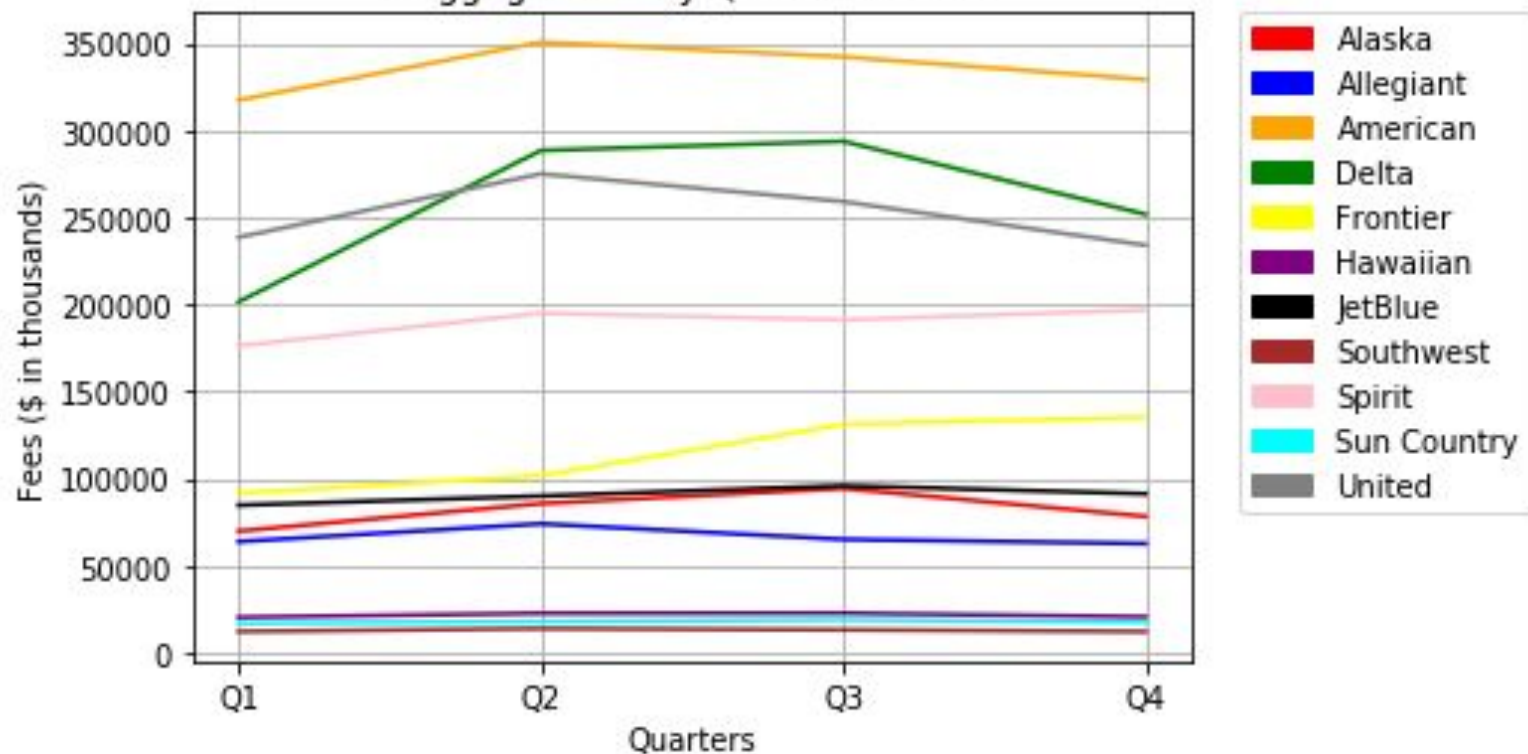
Baggage Fees by Quarter in 2017



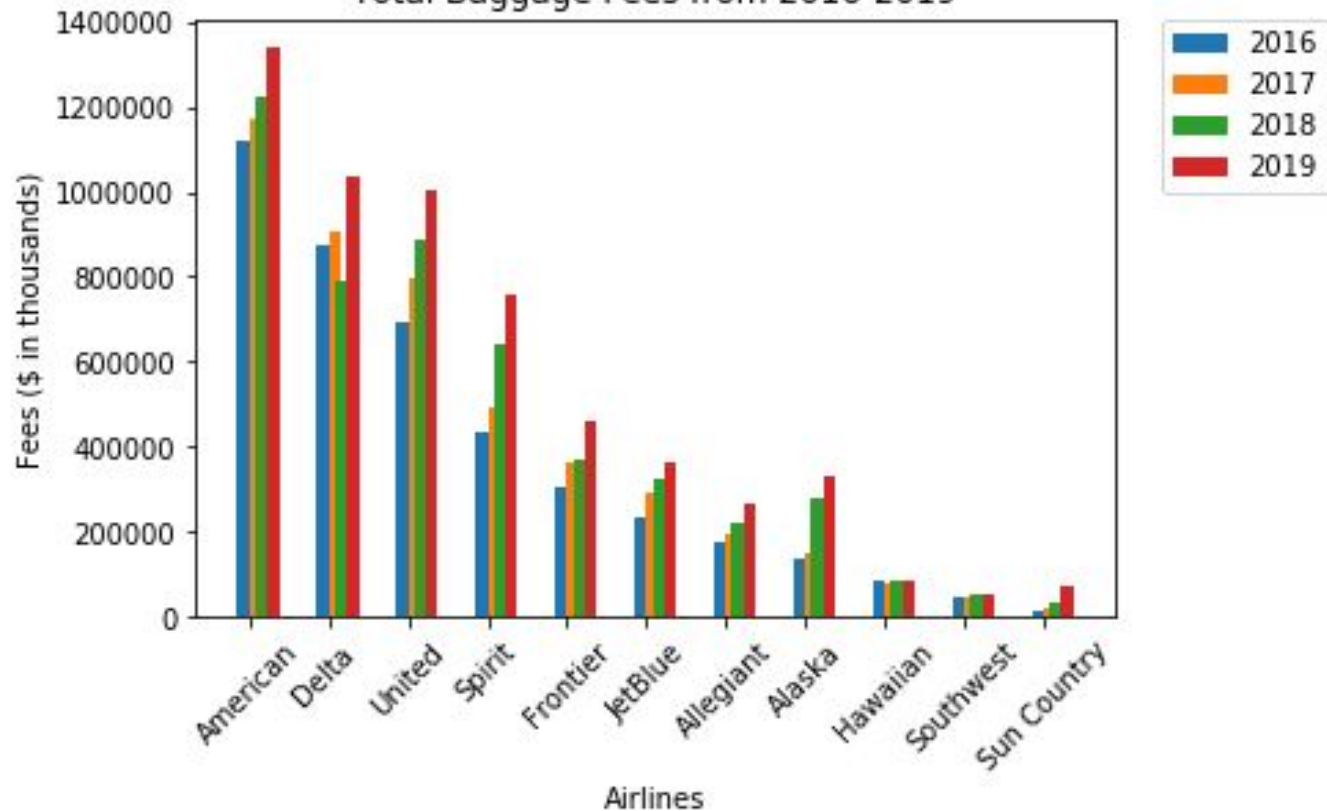
Baggage Fees by Quarter in 2018





Baggage Fees by Quarter in 2019





Total Baggage Fees from 2016-2019





American increased their fees by 198%
while Southwest only increased their
fees by 16%



According to an article in the Washington Post on Feb 26, 2020, airlines are encouraging passengers to save money by checking bags in online, travel by Southwest (which is the only airline that doesn't charge for the first 2 checked bags) or to travel with a carry-on bag.

The article attributes the increase in fees to the decline of cargo revenue among other reasons. Another news article from Travel and Leisure, published in Sept 2018, linked the increase to the increase in fuel prices, while helping to pay for other amenities such as free tv and movies during the flight, and will help keep the price of airline tickets low.