**Michael English**

**Data Wrangling Report and Notes**

***Three Most Pertinent Data Points***

1. **Flight Routes(Bureau of Transportation Statistics[BTS]):** The flight routes to the most popular cities and least popular cities will give a good indicator of peak times based on the density of flights to and from these cities on a daily basis
2. **Maximum number of seats (from the Releasable Aircraft Database)**: While some routes might be more popular, it doesn’t completely mean that those flights will have certain aircraft to be used. Need to get the *type, tail numbers,* and *number of seats* for each aircraft being used to give overestimate of population based on the flights analyzed in a given time frame.
3. **BTS Database of Flights:**  This will provide most pertinent data as it is giving the precise dates for flights departures, which will give a more accurate description of the population in smaller intervals. The dataset has been refined to focus only on two cities that have the highest traffic of flights out of Atlanta: Fort Lauderdale, FL and New York, NY.

* From the **T-100D Database**, have parsed data from the collective data on carriers and enplaning on particular routes from all of America to Atlanta, GA. This database is also restricted to March 2019
* From the Atlanta subset, have removed all rows that contain data on passenger capacity less than 100, in such a large timeframe of a month, routes with less than 100 passengers per month are mostly like indicative of freight vehicles/planes and the employees that work them, and are negligible in broader calculations.
* Added a column, *Average Passengers per Day,* in order to provide a static context for the data that is more specific to the day as opposed to the month.
* For the **OnTime Reporting Database** from the Bureau of Transportation Statistics, there is similar data, have removed redundant columns and leaving data that is pertinent to the particular timeframe of March 2019, where both datasets are consistent.
* This dataset lists dates and tail numbers for all flights enplaning in Atlanta, GA in March 2019.
* So far, in comparison, the T-100D Dataset shows total passengers for the various *routes* departing from Atlanta, whereas the OnTime Reporting Database gives access to tail numbers that can be used to determine the load capacity of each flight enplaning in Atlanta for the entirety of March 2019.