**Airline Delays - EDA Summary**

**Time-of-year data**

* Year - ranges 2015-2019
* Quarter -- data well represented across quarters & years
* Month -- data well represented across months & years
* Day\_of\_Month - ranges 1-31, fewer records on 31st
* Day\_of\_week - ranges 1-7
* FL\_Date - flight data (as a string)

**Airline info**

* Op\_Unique\_Carrier - airline Id, total of 19 distinct airlines

**Origin Airport Info**

* Origin\_Airport\_Id - 371 distinct numeric values
* Origin\_Airport\_Seq\_Id - 688 distinct numeric values
* Origin\_City\_Market\_Id - 344 distinct values
* Origin - 371 distinct alphanumeric values (e.g. "SEA")
* Origin\_city\_name - 362 distinct values (e.g. "Seattle, WA") ----- seems odd 362 < 371...
* Origin\_state\_abr - 53 distinct values (e.g. "WA")
* Origin\_state\_FIPS - 53 distinct values (ranges 1-78)
* Origin\_state\_nm - 53 distinct values (e.g. "Washington")
* Origin\_WAC - 53 distinct values (ranges 1-93)

**Destination Airport Info**

* Dest\_Airport\_id - 369 distinct values (numeric)
* Dest\_Airort\_Seq\_id - 686 distinct value (numeric)
* Dest\_City\_Market\_Id - 343 distinct values
* Dest - 369 distinct values (e.g. "SEA")
* Dest\_City\_Name - 361 distinct values (e.g. "Seattle, WA") ------ seems odd 361 < 369...
* Dest\_State\_Abr - 53 distinct values (e.g. "WA")
* Dest\_State\_FIPS - 53 distinct values (ranges 1-78)
* Dest\_State\_Nm - 53 distinct values (e.g. "Washington")
* Dest\_WAC - 53 distinct values (ranges 1-93)

**Departure-Related Info**

* CRS\_Dep\_Time - ranges 0001 - 2359, 1433 values
* Dep\_Time - 472,320 nulls, ranges 0001 - 2400, 1441 values
* Dep\_Delay - 477,296 nulls, ranges -234 to 2755, 1749 values
* Dep\_Delay\_New - 477,296 nulls, ranges 0 to 2755, 1655 values (zero-ed out all negatives)
* Dep\_Del15 - 477,296 nulls, 25576004 with "0" value, 5693541 with "1" value
* Dep\_Delay\_Group - 477,296 nulls, ranges -2 to 12, 16 values
* Dep\_Time\_Blk - hour time blocks, no nulls, 19 values (e.g. 2300-2359); aggregated variable

**Taxi & Wheels Info**

* Taxi\_out - 486,417 nulls, ranges from 0 to 227, 195 distinct values
* Wheels\_Off - 486,412 nulls, ranges from 1 to 2400, 1441 distinct values
* Wheels\_On - 501,924 nulls, ranges from 1 to 2400, 1441 distinct values
* Taxi\_in - 501,924 nulls, ranges from 0 to 414, 291 distinct values

**Arrival-Related Info**

* CRS\_Arr\_Time - ranges 0001 - 2400, 1440 values
* Arr\_Time - 501,922 nulls, ranges 0001-2400, 1441 values
* Arr\_Delay - 570,640 nulls, ranges -1238 to 2695, 1761 values
* Arr\_Delay\_New - 570,640 nulls, ranges 0 to 2695, 1641 values
* Arr\_Del15 - 570,640 nulls, 25377086 with "0" value, 5799115 with "1" value
* Arr\_Delay\_Group - 570,640 nulls, ranges -2 to 12, 16 values, seems to follow a normal distribution
* Arr\_Time\_Blk - hour time blocks, no nulls, 19 values (e.g. 2300-2359); aggregated variable

**Cancelled/Diverted Indicators**

* Cancelled - 31,256,894 not cancelled; 489,947 cancelled
* Diverted - 31,668,733 not diverted; 78,108 diverted

**Elapsed Time Info**

* CRS\_Elapsed\_Time - 164 null values, range -99 to 948, 633 distinct values
* Acutal\_Elapsed\_Time - 568,042 null values, range 14 to 1604, 739 values
* Air\_Time - 568,042 null values, ranges 4 to 1557, 705 values

**Flights**

* Flights - number of flights, which is always 1

**Distance Info**

* Distance - ranges from 21 to 5095, 1689 distinct values
* Distance\_Group - aggregated variable, ranges 1-11, 11 total values

**Delay Info**

* Carrier\_Delay - 25,947,727 nulls, ranges 0 to 2695, 1597 distinct values
* Weather\_Delay - 25,947,727 nulls, ranges 0 to 2692, 1241 distinct values
* NAS\_Delay - 25,947,727 nulls, ranges 0 to 1848, 1249 distinct values
* Security\_Delay - 25,947,727 nulls, ranges 0 to 1078, 315 distinct values
* Late\_Aircraft\_Delay - 25,947,727 nulls, ranges 0 to 2454, 1262 distinct values