**REPORT**

Data Cleaning:

1. Converting strings to int (‘347’ to 347, ’1,240’ to 1240)
2. Date-Time conversion

Data Pre-Processing:

1. Reducing dates to day, month
2. Removing missing values (NaN) – ‘FIRST\_DEP\_TIME’ feature contained only few values and was a measure of the time that has elapsed since the flight was set off from the first departure gate. This feature seemed irrelevant with missing values and was hence dropped.
3. Converting string objects to numerals (Label Encoding and Hash-map)
4. Removing in-important features
5. ‘AIRLINE\_ID’ and ‘UNIQUE\_CARRIER’ represent the same thing so I dropped the ‘UNIQUE\_CARRIER’ feature.