SUMMARY REPORT
Assignment #01

Emaan Ali
22i2325
CS-Z

Dataset details:
Domain : Transportation
Target variable : ARIVAL_DELAY (numeric val)
No of Records: 1048576 (1.4 million)
Number of features : 30
Problem type: Regression
Categorical Feature for One-Hot Encoding: AIRLINE
Test Feature: ARRIVAL_DELAY (or DEPARTURE_DELAY for alternative tasks).

```
Missing Values After Handling:
YEAR                          0
MONTH                         0
DAY                           0
DAY_OF_WEEK                   0
AIRLINE                       0
FLIGHT_NUMBER                 0
TAIL_NUMBER               14721
ORIGIN_AIRPORT                0
DESTINATION_AIRPORT           0
SCHEDULED_DEPARTURE           0
DEPARTURE_TIME                0
DEPARTURE_DELAY               0
TAXI_OUT                      0
WHEELS_OFF                    0
SCHEDULED_TIME                0
ELAPSED_TIME                  0
AIR_TIME                      0
DISTANCE                      0
WHEELS_ON                     0
TAXI_IN                       0
SCHEDULED_ARRIVAL             0
ARRIVAL_TIME                  0
ARRIVAL_DELAY                 0
DIVERTED                      0
```

```
]  Outliers in DEPARTURE_DELAY: 736242
   Outliers in ARRIVAL_DELAY: 539002
Σ  Outliers in SCHEDULED_ARRIVAL: 0
   Outliers in DISTANCE: 349511
   Outliers in SCHEDULED_DEPARTURE: 0
   Outliers in WHEELS_ON: 0
   Correlation with ARRIVAL_DELAY:
    ARRIVAL_DELAY          1.000000
   DEPARTURE_DELAY         0.937710
   AIRLINE_DELAY           0.611132
   LATE_AIRCRAFT_DELAY     0.600018
   AIR_SYSTEM_DELAY        0.394284
   WEATHER_DELAY           0.270213
   TAXI_OUT                0.227141
   DEPARTURE_TIME          0.159409
   WHEELS_OFF              0.155548
   TAXI_IN                 0.116692
   SCHEDULED_DEPARTURE     0.098872
   SCHEDULED_ARRIVAL       0.087636
   WHEELS_ON               0.058405
   ARRIVAL_TIME            0.049451
   ELAPSED_TIME            0.030194
   SECURITY_DELAY          0.028125
   FLIGHT_NUMBER           0.017042
   DAY                    -0.002983
   AIR_TIME               -0.005933
   DIVERTED               -0.012136
```