

Flight Delay Predictor

LHL Midterm Project B
Nitin Kohli
Gerard Agada



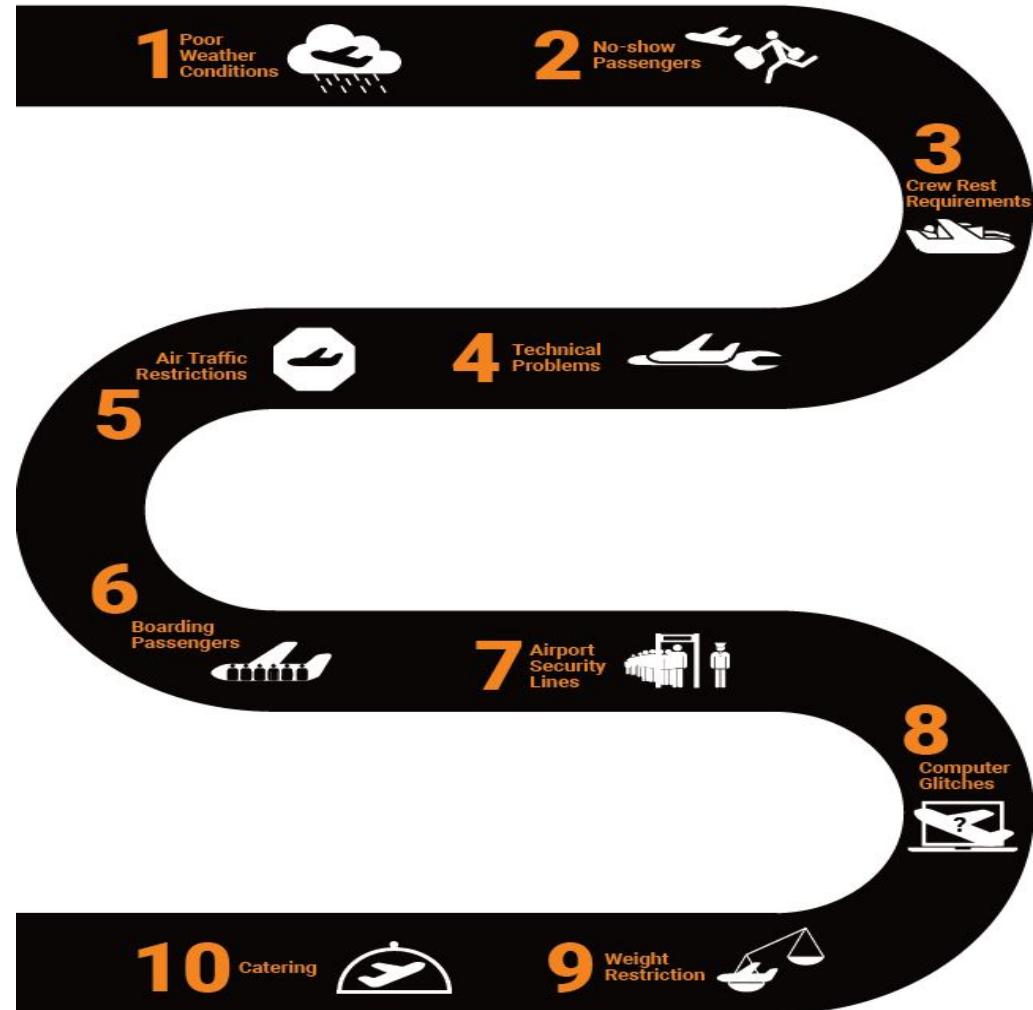
Why are Flight Delays Bad?



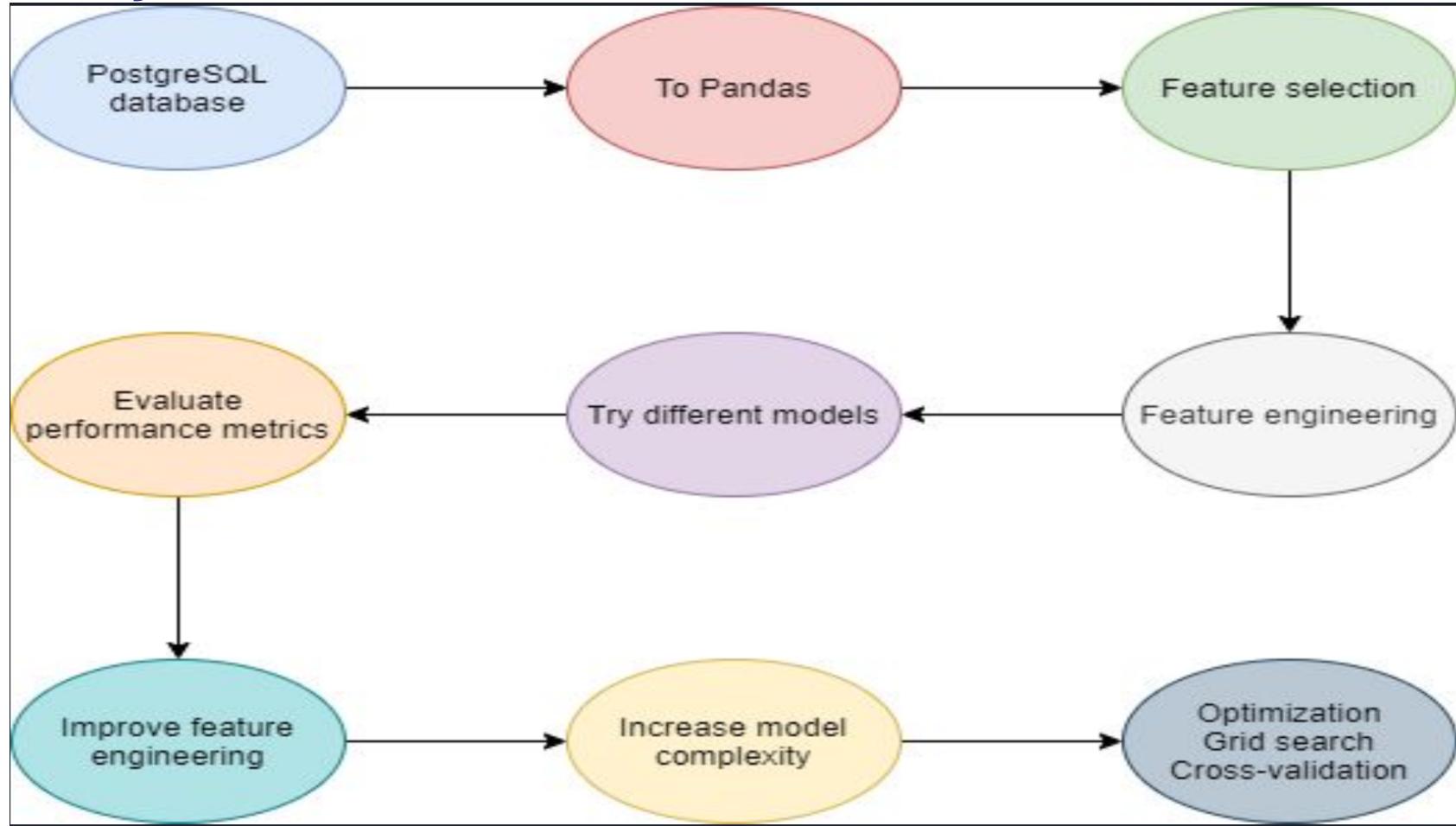
- Costs A LOT of Money.
 - FAA estimated the annual cost of delays to be \$28 BILLION in the USA alone.
 - Annoyed passengers.
 - Ruined plans.
 - Reduced customer base.
 - Make travellers to miss connecting flights.
-

What causes a Flight to be Delayed?

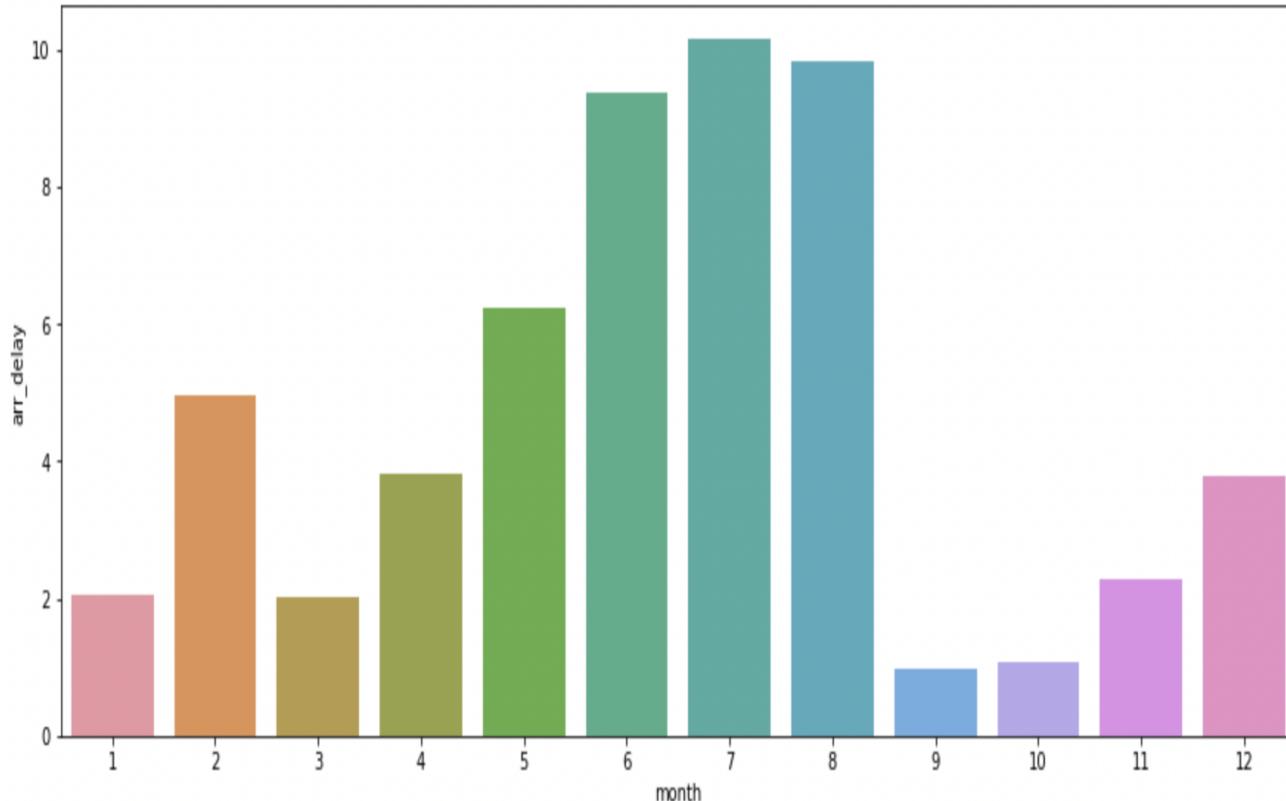
- Adverse Weather Conditions
- Waiting for Passengers
- Air Traffic Restrictions
- Technical Problems
- Airport Security
- Catering
- Crew Rest Requirements



Project Workflow :

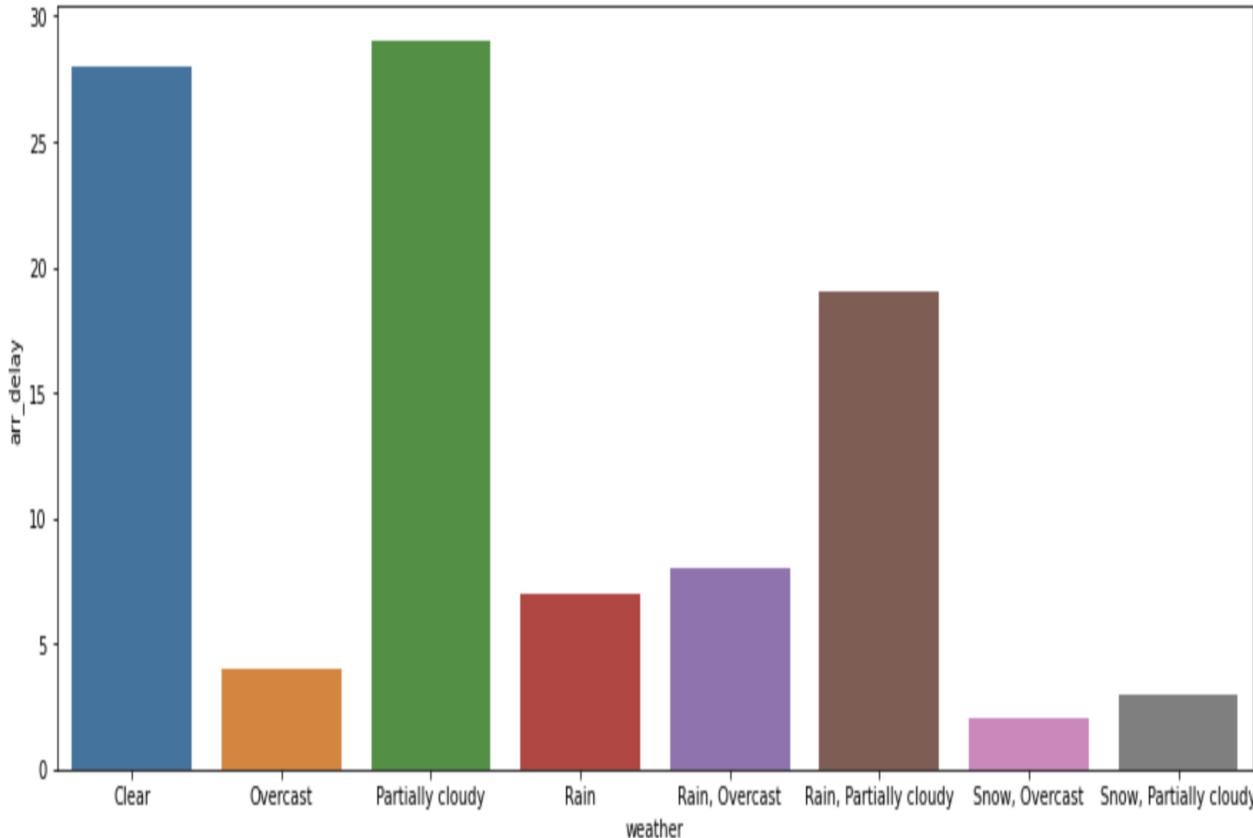


Findings from Dataset Analysis:



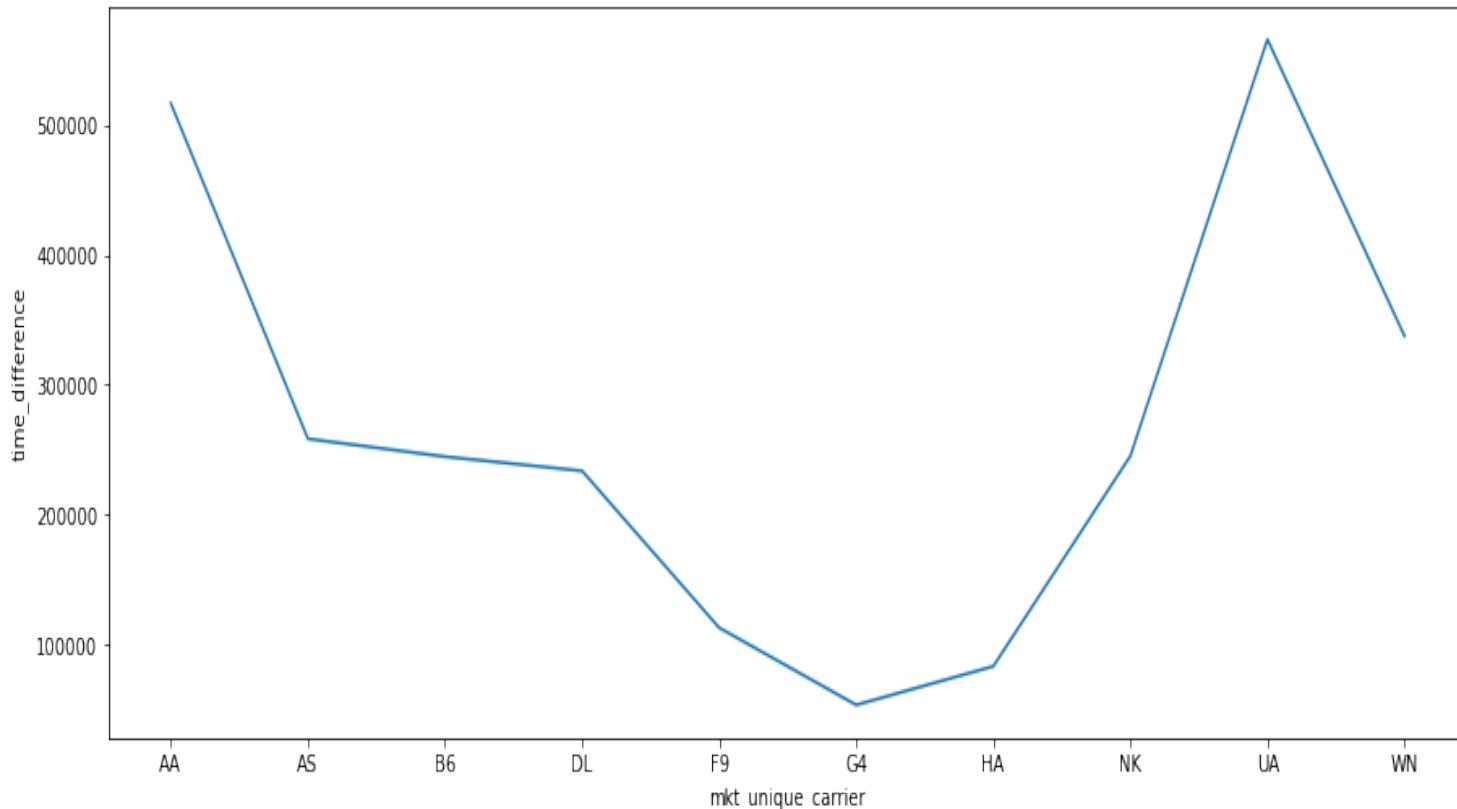
Arrival Delays increase during months of June, July and August as the number of travellers increase in Summer.

Findings from Dataset Analysis:



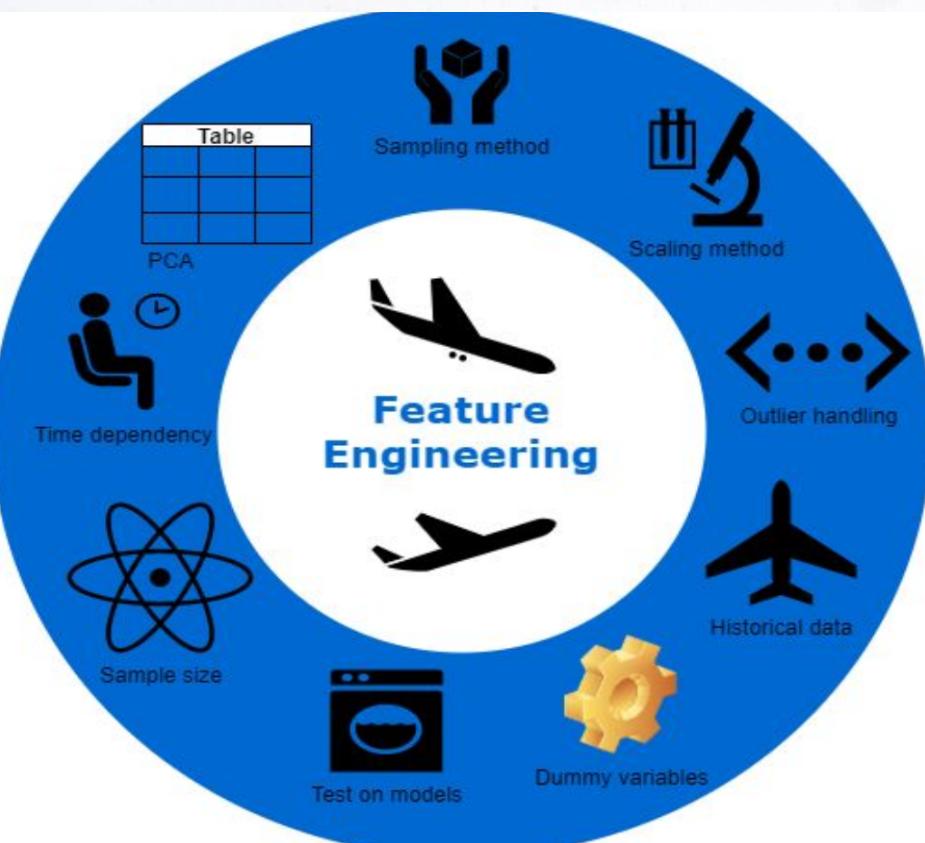
Based on the sample size, weather is not a significant predictor of Arrival Delays.

Findings from Dataset Analysis:

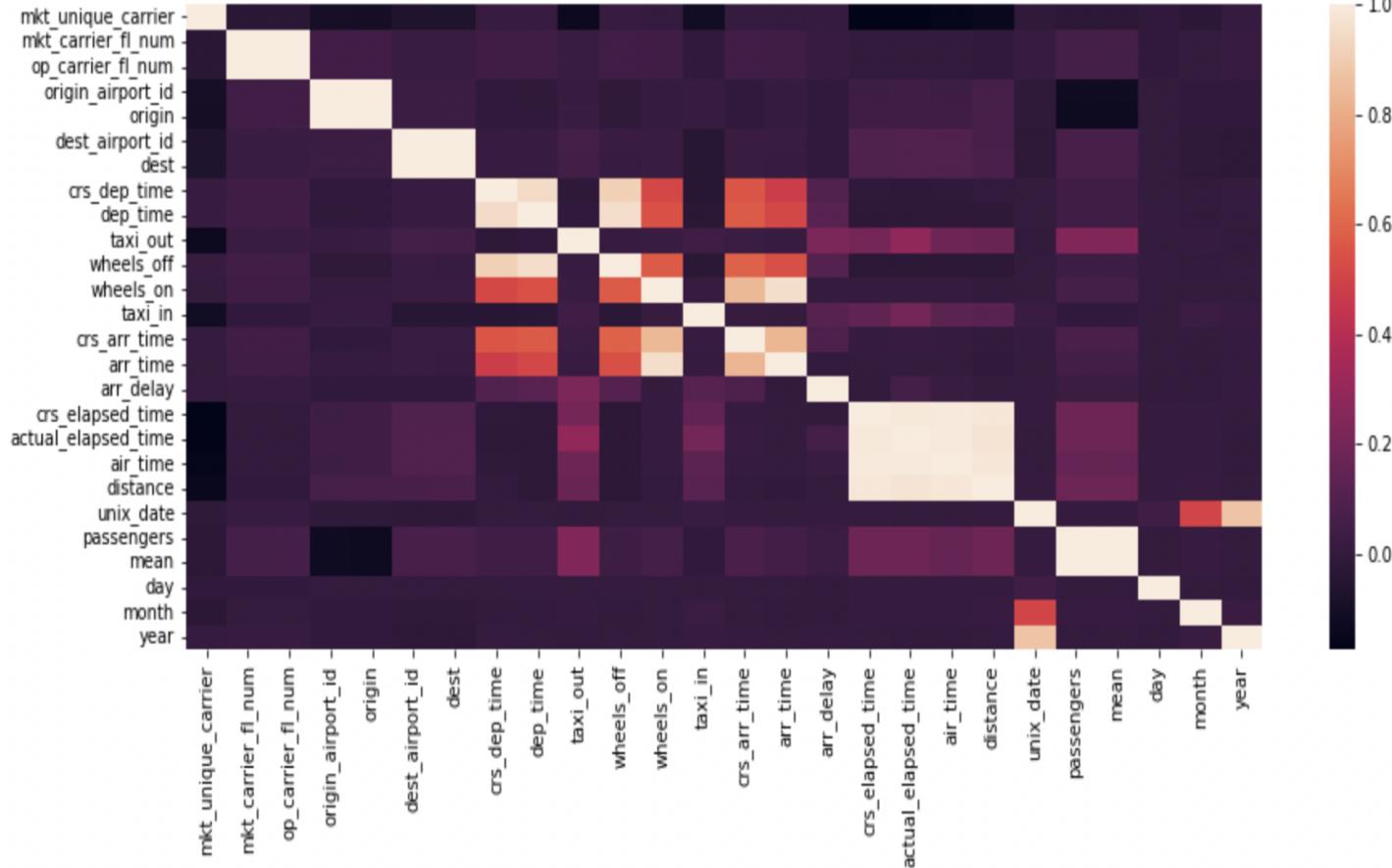


Upon comparing, we found that American Airlines and United Airlines have the highest amount of delayed flights.

Getting Predictions One Week in Advance

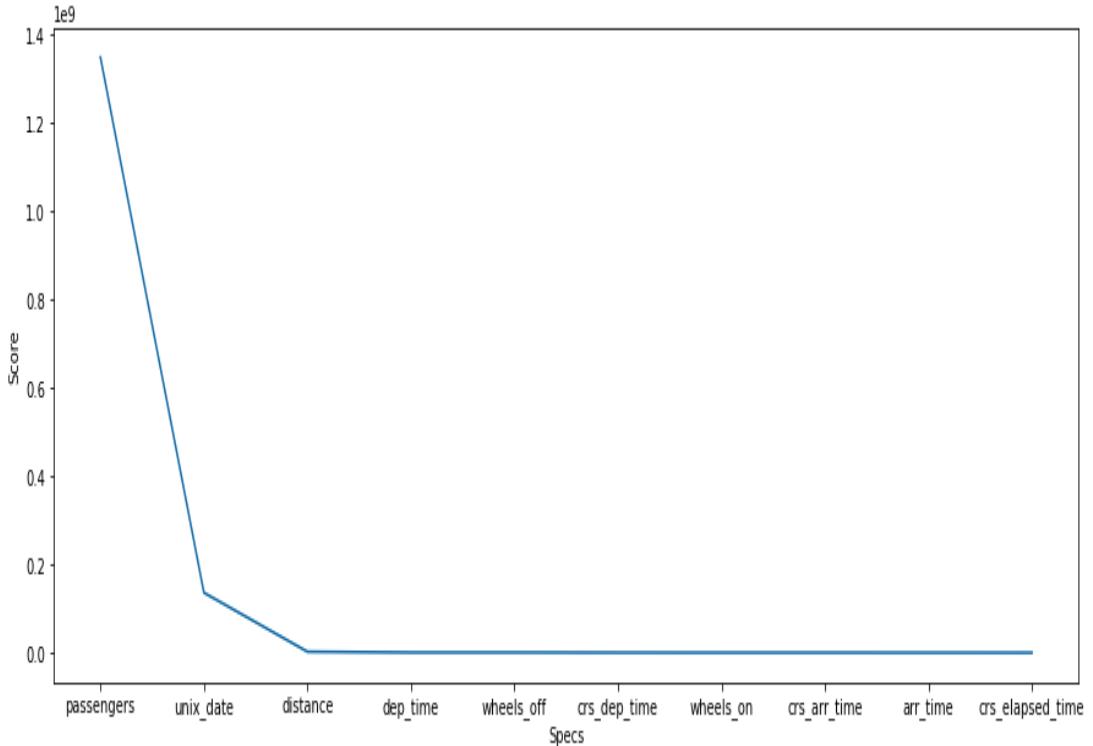


Findings from Dataset Analysis:



*Analyzing
correlation
between
features.*

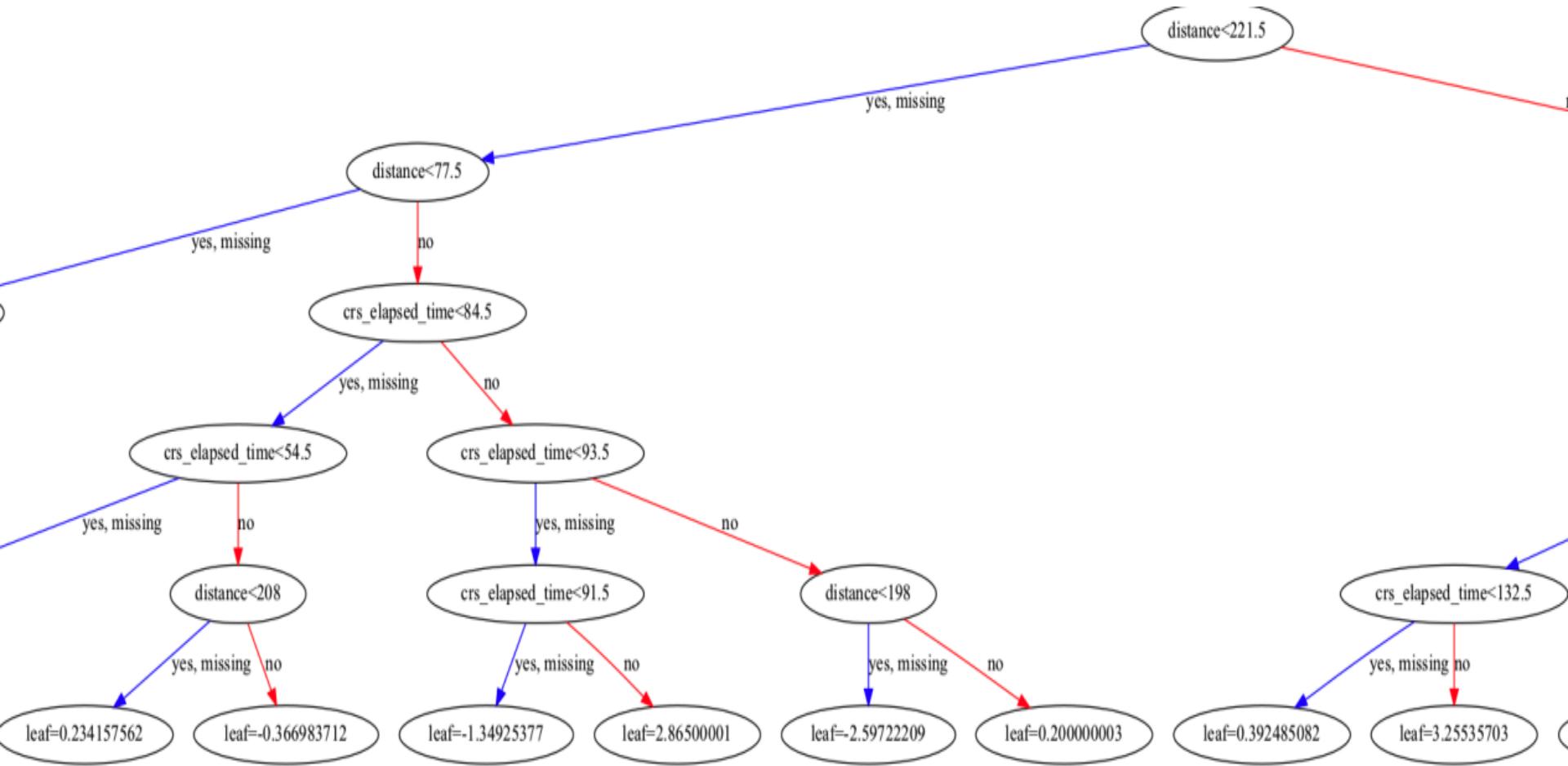
Findings from Dataset Analysis:



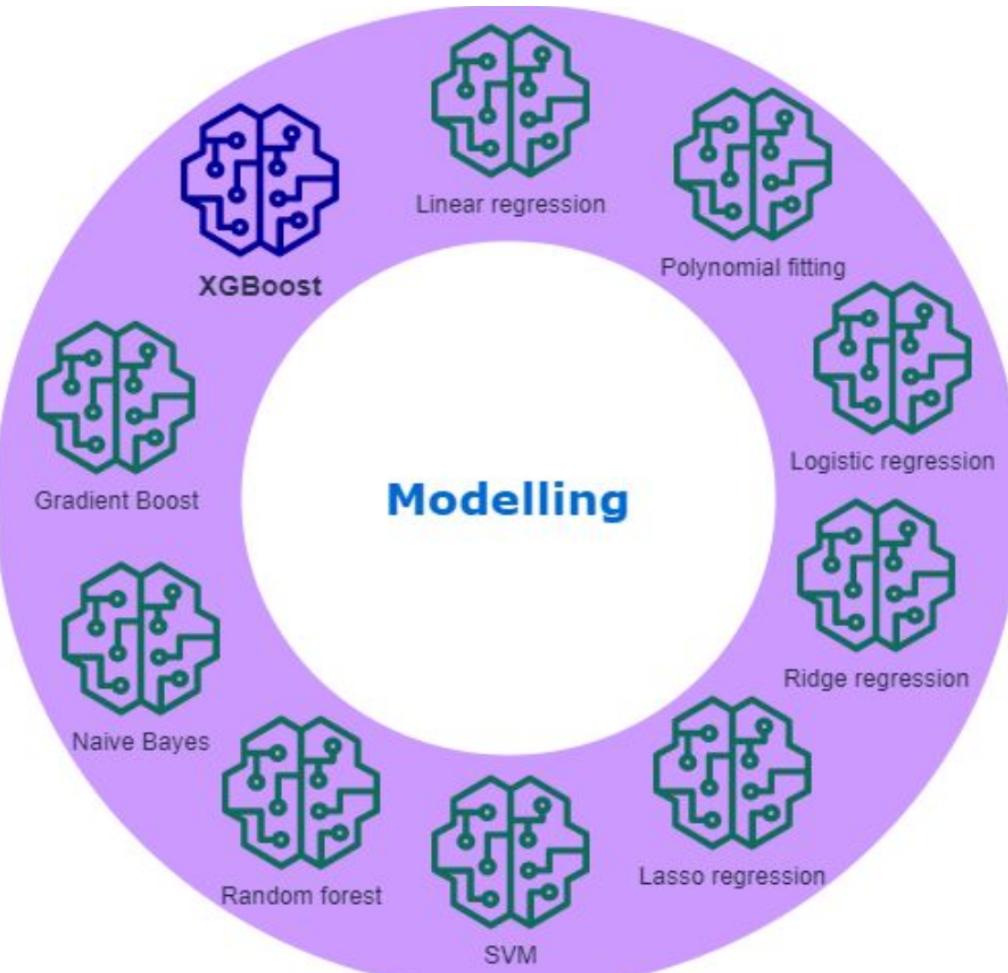
| Specs | Score |
|------------------|--------------|
| passenger | 1.347744e+09 |
| unix_date | 1.361190e+08 |
| distance | 2.546102e+06 |
| dep_time | 8.634091e+05 |
| wheels_off | 8.534938e+05 |
| crs_dep_time | 5.360602e+05 |
| wheels_on | 4.335217e+05 |
| crs_arr_time | 4.302145e+05 |
| arr_time | 4.020535e+05 |
| crs_elapsed_time | 3.034180e+05 |

*Determining importance
of features to predict
arrival delays.*

Modelling with Random Forest :



Different Models Used For Predictions :



Arrival Delays Predicted One Week In Advance :

| fl_date | mkt_carrier | mkt_carrier_fl_num | origin | dest | predicted_delay |
|------------|-------------|--------------------|--------|------|-----------------|
| 2020-01-01 | WN | 5888 | ONT | SFO | 5.0236435 |
| 2020-01-01 | WN | 6276 | ONT | SFO | 3.3203819 |
| 2020-01-01 | WN | 4598 | ONT | SJC | 4.6955895 |
| 2020-01-01 | WN | 4761 | ONT | SJC | 3.1810355 |
| 2020-01-01 | WN | 5162 | ONT | SJC | 1.0680859 |
| 2020-01-01 | WN | 5684 | ONT | SJC | 0.58055025 |
| 2020-01-01 | WN | 6152 | ONT | SJC | 4.306197 |
| 2020-01-01 | WN | 1679 | ONT | SMF | 4.306197 |
| 2020-01-01 | WN | 3479 | ONT | SMF | 3.5444014 |
| 2020-01-01 | WN | 4069 | ONT | SMF | 0.5836405 |
| 2020-01-01 | WN | 4905 | ONT | SMF | 1.9249954 |
| 2020-01-01 | WN | 4918 | ONT | SMF | 4.6955895 |
| 2020-01-01 | WN | 5144 | ONT | SMF | 5.5759635 |
| 2020-01-01 | WN | 5722 | ONT | SMF | 5.0236435 |
| 2020-01-01 | WN | 3719 | ORF | BWI | 1.9712534 |
| 2020-01-01 | WN | 4863 | ORF | BWI | 0.7644057 |
| 2020-01-01 | WN | 4924 | ORF | BWI | 5.532424 |
| 2020-01-01 | WN | 5321 | ORF | BWI | 6.197722 |
| 2020-01-01 | WN | 4618 | ORF | MCO | 2.4593706 |
| 2020-01-01 | WN | 5992 | ORF | MDW | 5.43926 |

Hurdles Faced & Future Goals:



- Data cleaning
- Feature transformation
- Feature selection
- Model optimisation
- Multiclass Predictions
- Increase model complexity
- Binary Classification

Thank You for your time and attention.