2022

LITERATURE SURVEY REPORT

LITERATURE SURVEY REPORT

FLIGHT DELAY ANALYSIS

Team Members:

PES2UG20CS387 - VISHNU D

PES2UG20CS568 - VARUN

PES2UG20CS571 – VEERAVALLI RUTHVIK

Project Data Set – AIRLINES DELAY (Kaggle)

LITERATURE SURVEY REPORT

ABOUT DATASET:

The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS) tracks the on-time performance of domestic flights operated by large air carriers. Summary information on the number of on-time, delayed, canceled and diverted flights appears in DOT's monthly Air Travel Consumer Report, published about 30 days after the month's end, as well as in summary tables posted on this website. BTS began collecting details on the causes of flight delays in June 2003. Summary statistics and raw data are made available to the public at the time the Air Travel Consumer Report is released.

DATASET GLOSSARY:

Year :2008

Month :1-12

DayofMonth : 1-31

DayOfWeek :1 (Monday) - 7 (Sunday)

DepTime : actual departure time (local, hhmm)

CRSDepTime : scheduled departure time (local, hhmm)

ArrTime : actual arrival time (local, hhmm)

CRSArrTime : scheduled arrival time (local, hhmm)

UniqueCarrier : unique carrier code

FlightNum : flight number

TailNum : plane tail number: aircraft registration, unique aircraft identifier

ActualElapsedTime : in minutes

LITERATURE SURVEY REPORT

CRSElapsedTime : in minutes

AirTime : in minutes

ArrDelay : arrival delay, in minutes: A flight is counted as "on time" if it operated less than 15 minutes later the scheduled time shown in the carriers' Computerized Reservations Systems (CRS).

DepDelay : departure delay, in minutes

Origin : origin IATA airport code

Dest : destination IATA airport code

Distance : in miles

TaxiIn : taxi in time, in minutes

TaxiOut : taxi out time in minutes

Cancelled :*was the flight cancelled

CancellationCode : reason for cancellation (A = carrier, B = weather, C = NAS, D = security)

Diverted : 1 = yes, 0 = no

CarrierDelay : in minutes: Carrier delay is within the control of the air carrier. Examples of occurrences that may determine carrier delay are: aircraft cleaning, aircraft damage, awaiting the arrival of connecting passengers or crew, baggage, bird strike, cargo loading, catering, computer, outage-carrier equipment, crew legality (pilot or attendant rest), damage by hazardous goods, engineering inspection, fueling, handling disabled passengers, late crew, lavatory servicing, maintenance, oversales, potable water servicing, removal of unruly passenger, slow boarding or seating, stowing carry-on baggage, weight and balance delays.

WeatherDelay : in minutes: Weather delay is caused by extreme or hazardous weather conditions that are forecasted or manifest themselves on point of departure, enroute, or on point of arrival.

LITERATURE SURVEY REPORT

NASDelay : in minutes: Delay that is within the control of the National Airspace System (NAS) may include: non-extreme weather conditions, airport operations, heavy traffic volume, air traffic control, etc.

SecurityDelay : in minutes: Security delay is caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.

LateAircraftDelay : in minutes: Arrival delay at an airport due to the late arrival of the same aircraft at a previous airport. The ripple effect of an earlier delay at downstream airports is referred to as delay propagation.

STATISTICAL SUMMARY:

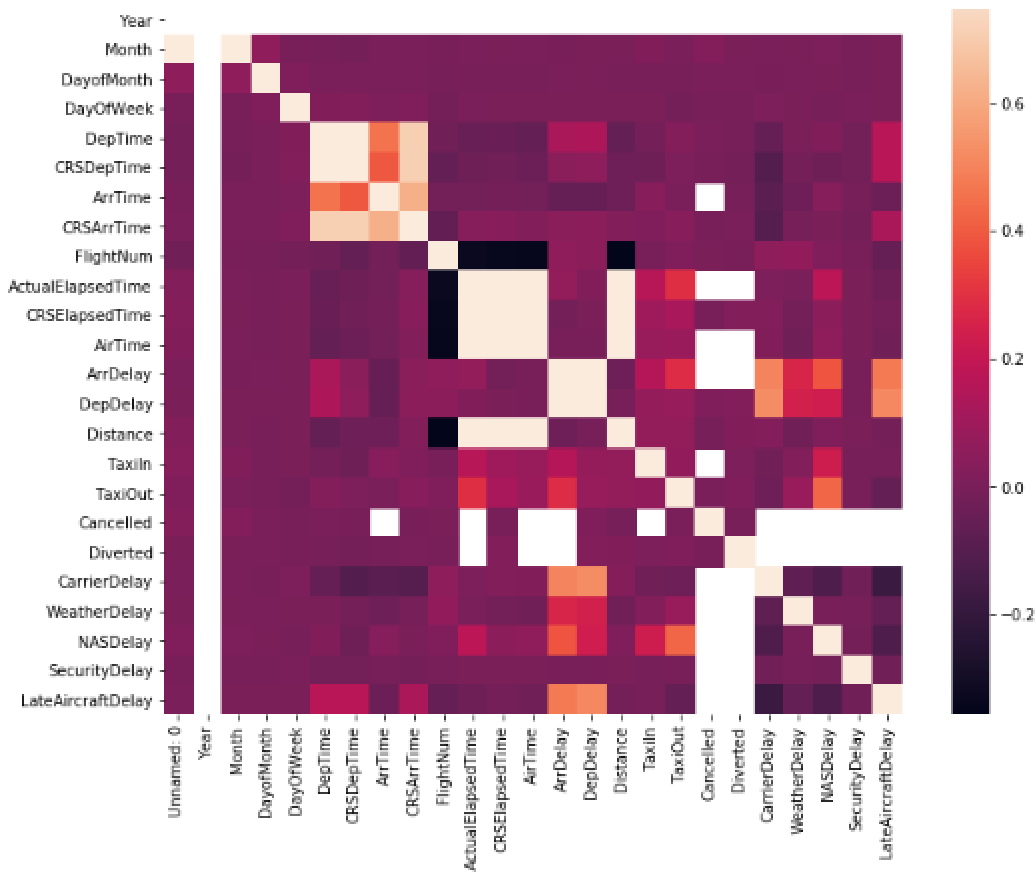
```
       X1              Year               Month            DayofMonth
 Min.   :      0   2008:1936758   12      :203385   Min.   : 1.00
 1st Qu.:1517452                  6       :200914   1st Qu.: 8.00
 Median :3242558                  3       :200842   Median :16.00
 Mean   :3341651                  2       :189534   Mean   :15.75
 3rd Qu.:4972467                  1       :183527   3rd Qu.:23.00
 Max.   :7009727                  7       :182945   Max.   :31.00
                                  (Other):775611
   DayOfWeek         DepTime        CRSDepTime        ArrTime        CRSArrTime
 Min.   :1.000   Min.   :   1   Min.   :   0   Min.   :   1   Min.   :   0
 1st Qu.:2.000   1st Qu.:1203   1st Qu.:1135   1st Qu.:1316   1st Qu.:1325
 Median :4.000   Median :1545   Median :1510   Median :1715   Median :1705
 Mean   :3.985   Mean   :1519   Mean   :1467   Mean   :1610   Mean   :1634
 3rd Qu.:6.000   3rd Qu.:1900   3rd Qu.:1815   3rd Qu.:2030   3rd Qu.:2014
 Max.   :7.000   Max.   :2400   Max.   :2359   Max.   :2400   Max.   :2400
                                               NA's   :7110
 UniqueCarrier      FlightNum        TailNum          ActualElapsedTime
 WN     :377602   Min.   :   1   Length:1936758     Min.   :  14.0
 AA     :191865   1st Qu.: 610   Class :character   1st Qu.:  80.0
 MQ     :141920   Median :1543   Mode  :character   Median : 116.0
 UA     :141426   Mean   :2184                      Mean   : 133.3
```

LITERATURE SURVEY REPORT

```
OO      :132433   3rd Qu.:3422                          3rd Qu.: 165.0
DL      :114238   Max.   :9742                          Max.   :1114.0
(Other):837274                                          NA's   :8387
CRSElapsedTime      AirTime          ArrDelay          DepDelay
Min.   :-25.0   Min.   :   0.0   Min.   :-109.0   Min.   :   6.00
1st Qu.: 82.0   1st Qu.:  58.0   1st Qu.:   9.0   1st Qu.:  12.00
Median :116.0   Median :  90.0   Median :  24.0   Median :  24.00
Mean   :134.3   Mean   : 108.3   Mean   :  42.2   Mean   :  43.19
3rd Qu.:165.0   3rd Qu.: 137.0   3rd Qu.:  56.0   3rd Qu.:  53.00
Max.   :660.0   Max.   :1091.0   Max.   :2461.0   Max.   :2467.00
NA's   :198     NA's   :8387     NA's   :8387
    Origin             Dest            Distance           TaxiIn
ATL     : 131613   ORD     : 108984   Min.   :  11.0   Min.   :  0.000
ORD     : 125979   ATL     : 106898   1st Qu.: 338.0   1st Qu.:  4.000
DFW     :  95414   DFW     :  70657   Median : 606.0   Median :  6.000
DEN     :  74323   DEN     :  63003   Mean   : 765.7   Mean   :  6.813
LAX     :  58772   LAX     :  59969   3rd Qu.: 998.0   3rd Qu.:  8.000
IAH     :  56847   EWR     :  55861   Max.   :4962.0   Max.   :240.000
(Other):1393810   (Other):1471386                      NA's   :7110
   TaxiOut          Cancelled          CancellationCode       Diverted
Min.   :  0.00   Min.   :0.0000000   Length:1936758      Min.   :0.000000
1st Qu.: 10.00   1st Qu.:0.0000000   Class :character    1st Qu.:0.000000
Median : 14.00   Median :0.0000000   Mode  :character    Median :0.000000
Mean   : 18.23   Mean   :0.0003268                       Mean   :0.004004
3rd Qu.: 21.00   3rd Qu.:0.0000000                       3rd Qu.:0.000000
Max.   :422.00   Max.   :1.0000000                       Max.   :1.000000
NA's   :455
 CarrierDelay      WeatherDelay        NASDelay       SecurityDelay
Min.   :   0.0   Min.   :   0.0   Min.   :   0     Min.   :  0.0
1st Qu.:   0.0   1st Qu.:   0.0   1st Qu.:   0     1st Qu.:  0.0
Median :   2.0   Median :   0.0   Median :   2     Median :  0.0
Mean   :  19.2   Mean   :   3.7   Mean   :  15     Mean   :  0.1
3rd Qu.:  21.0   3rd Qu.:   0.0   3rd Qu.:  15     3rd Qu.:  0.0
Max.   :2436.0   Max.   :1352.0   Max.   :1357     Max.   :392.0
NA's   :689270   NA's   :689270   NA's   :689270   NA's   :689270
LateAircraftDelay
Min.   :   0.0
1st Qu.:   0.0
Median :   8.0
Mean   :  25.3
3rd Qu.:  33.0
Max.   :1316.0
NA's   :689270
```

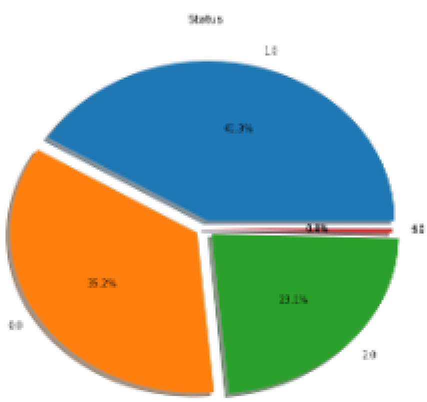
LITERATURE SURVEY REPORT

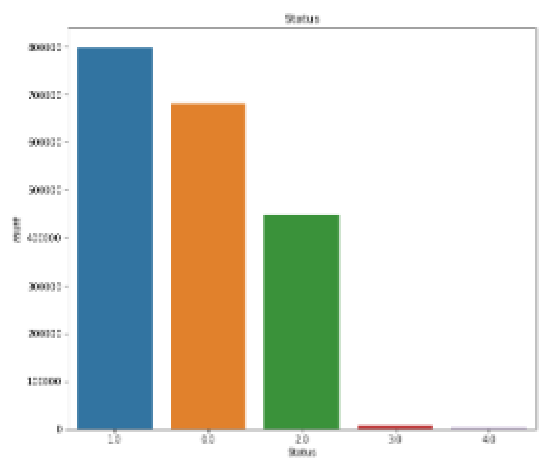
CORRELOGRAM:



LITERATURE SURVEY REPORT

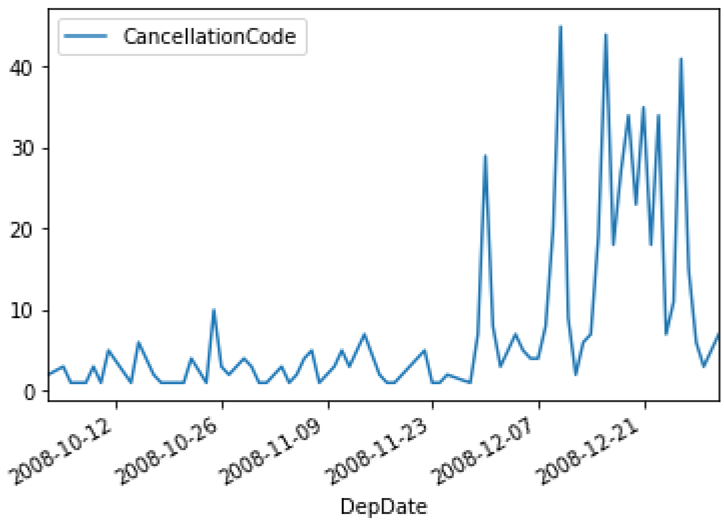
Status represents whether the flight was on time (0), slightly delayed (1), highly delayed (2), diverted (3), or cancelled (4)



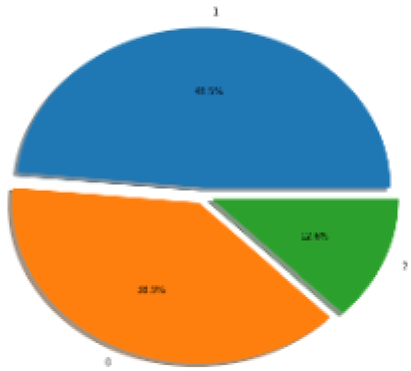


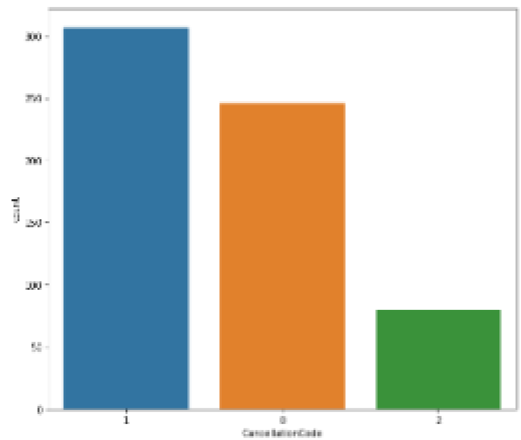
LITERATURE SURVEY REPORT

Cancelled Flights:



LITERATURE SURVEY REPORT





0 = carrier, 1 = weather, 2 = NAS

MODELS USED:

We Have Used Linear Regression and Logistic Regression
After applying both the models for predicting whether
a flight should be delayed, as well as how much one
would expect a flight should be delayed, we found the
following factors to be important: week, month, airline
carrier reference, planned elapsed time (in airtime),
distance between two departure and destinations,
flight planned departure time, departure airport code,
and taxi-in and taxi-out4 time. By applying our model,

LITERATURE SURVEY REPORT

on the data collected, one could be able to predict whether a flight might be delayed, and more importantly, how long delayed time she/he would expect.

However, there is some limitation in our model, first, our model only included one-year data due to our computation capability, as more years of data included, the prediction could be easier. In addition, some other related information such as airplane type, e.g., detailed weather data specific to airport were not included. Therefore, researchers could try to collect more related data and deploy better computational powers to build a better model. This paper presented a methodology for predicting aggregate flight departure delays in airports by exploring supervised learning methods. This way, we may be able to predict the delays of a new flight, without needing several months of data to build a prediction model. Another step forward would be to generalize the model to flights of the entire world, or at least to exploit more data sources, to build more complete predictions. Finally, the most interesting step would be to integrate such a model into a flight booking tool, to provide the delay prediction to future passengers, even this would require a strong confidence in the information provided, considering the possible impact in terms of reservations.