South Central SAS Users Group Educational Forum 2016

An Analysis of US Domestic Flight delays using SAS Enterprise Miner®

Piyush Gupta, Anuj Dwivedi & Aditya Agrawal

ABSTRACT

A flight delay is one of the major concern for the Aviation industry in the United States. Out of 8 million domestic flights, about 25% get delayed, and 1.5% flights get canceled every year. Delays lead to excess cost resulting in billions of dollars in loss to the flight carriers and the entire aviation industry. At the same time, it restricts the airline companies to deliver better services to the passengers. In today's data-driven world, with the availability of tonnes of on time flights data, the delays can be optimized. The first part of this paper analyzes the on-time performance data between the period of November 2015 to May 2016 collected from the open data center of US Bureau of Transport Statistics and the daily weather from the Weather Underground© to estimate the significant factors responsible for the flight delays such as Weather delay, Carrier delay, National Airspace System(NAS) delay and Late Aircraft delay. The second part of the paper is devoted to improving the delay predictions based on the various weather conditions.

1. INTRODUCTION

We selected the top five busiest airports in the USA by the air-traffic: Atlanta(ATL), Los Angeles(LAX), Chicago(ORD), Dallas(DFW) and Denver(DEN). And combined the open weather data with the on-time performance data to increase the granularity of the climatic reasons such as wind speed and direction, Air Density, visibility, dew, humidity, and precipitation.

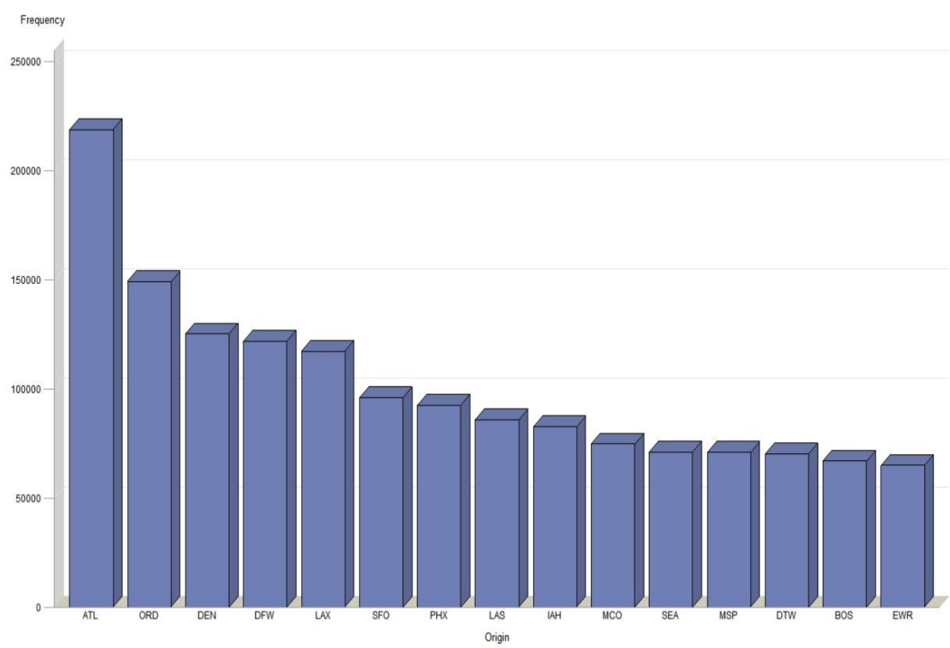


Fig.1 Busiest Airports in the United States

2. DATA PREPARATION USING SAS ENTERPRISE GUIDE AND MINER

Firstly, we performed descriptive analysis on the flight data and removed the irrelevant variables from the data set using SAS Enterprise Miner. We sampled the 3 million rows into a smaller stratified sample by the five busiest airports Atlanta, Dallas, Denver, Chicago, Los Angeles. We filtered the data using the SAS Filter node into five separate data sets for the airports and merged the individual weather data for the five airports using the Merge Node in SAS Enterprise Miner.

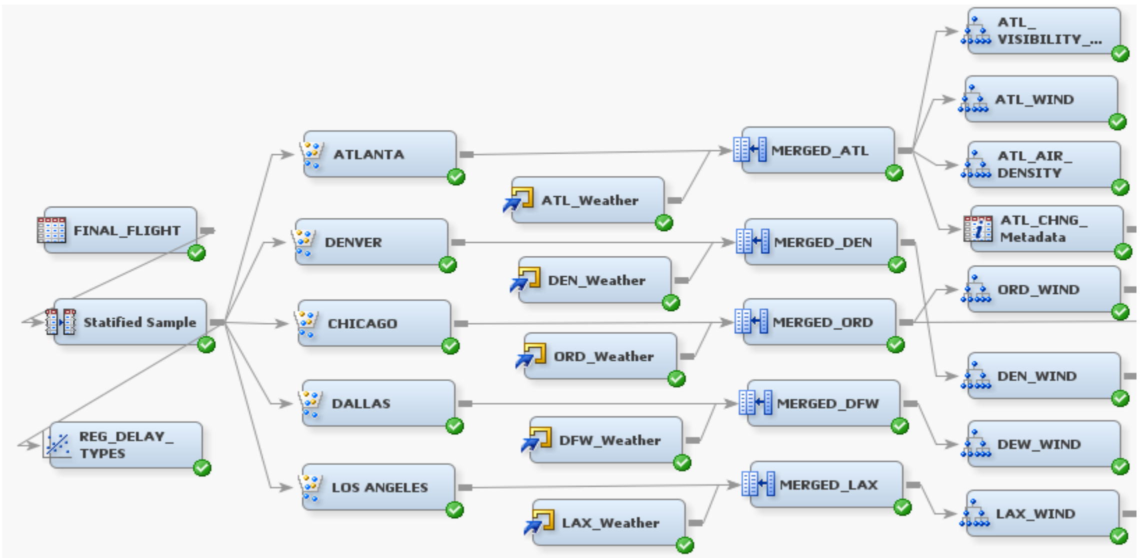


Fig.2 Data Preparation

3. DEPARTURE DELAY CATEGORIES

It is prominent first to analyze the various delay types before going deeper to causes behind the delays. Taking several delay types such as weather delay, security delay, carrier delay, national airspace system(NAS) delay, and late aircraft delays into consideration; we ran a regression on the departure delay minutes. All types of delays appeared significant in the analysis, and they all have a positive impact on the departure delay minutes.

Model Fit Statistics

R-Square	0.7598	Adj R-Sq	0.7598
AIC	1492843.2710	BIC	1492845.2714
SBC	1492904.7879	C(p)	6.0000

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10.0754	0.1015	99.30	<.0001
CarrierDelay	1	0.9763	0.00156	626.86	<.0001
LateAircraftDelay	1	0.9685	0.00185	522.23	<.0001
NASDelay	1	0.5713	0.00290	196.86	<.0001
SecurityDelay	1	0.8660	0.0378	22.90	<.0001
WeatherDelay	1	0.9334	0.00383	243.39	<.0001

According to US Federal Aviation Administration report, the largest aircraft delay factor is the weather. The pie chart shows that the weather causes 69% of the aircraft delay. It has also been stated in the report that the delay due to weather conditions added 10 million minutes of delay in 2013. Thus it is imperative to analyze the various weather elements associated with a delay in aircraft departure.

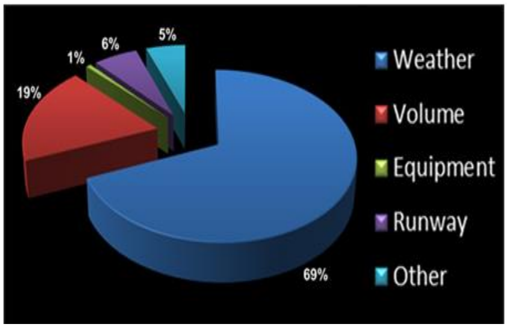
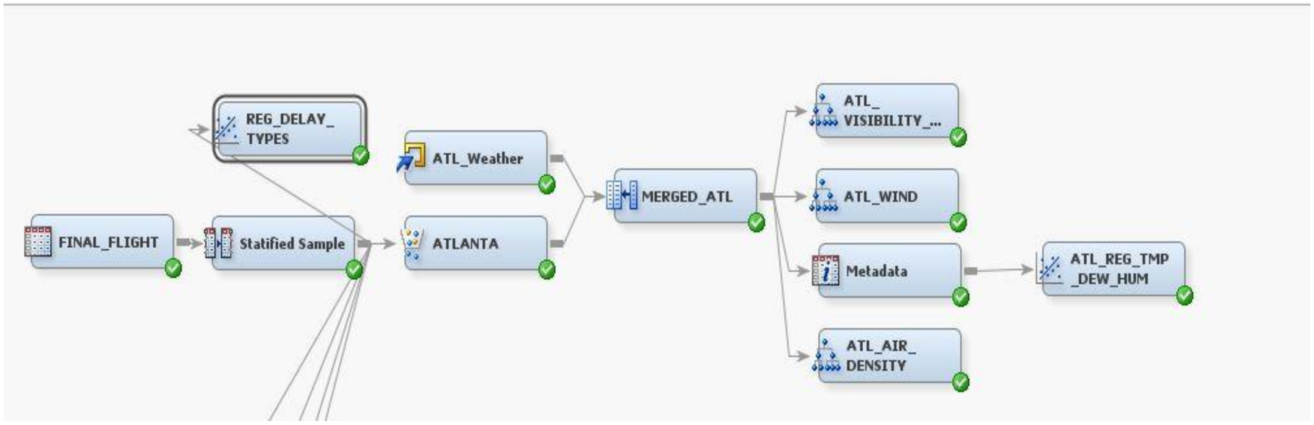


Fig. 3 Causes of air traffic delay

4. WEATHER FACTORS AFFECTING THE FLIGHT DELAYS

This paper provides a detailed characterization of delay distribution due to weather conditions. To expand our analysis, we have considered weather components such as the wind, temperature, humidity, dew point, air pressure, visibility and cloud cover. We have used decision tree model in an attempt to correlate multiple conditions with departure.



4.1 THE WIND EFFECT

One of the primary weather element behind the flight delays is the nature of the wind (Speed and Direction) at the runway. Flight delays occur when the winds are not parallel to the runways (Crosswinds) at the time of take-off. When the winds travel in the direction of the aircraft(Tailwinds), it increases the speed of the flight but decreases the lift of the flight. Headwinds travel against the direction of the aircraft and decrease the speed of the aircraft. However, out of these three effects, Crosswinds has a major impact on the flights to delay and hence it needs further**.** We deeply analyzed the runway map of each airport to find out the arch or angle to measure exact crosswind direction. We used the Decision Tree model in the SAS Enterprise Miner to explore the effect of the crosswinds on the delays for the five airports. We separately analyzed the effect of wind speed and direction on the delay minutes for all the five airports.

HARTSFIELD JACKSON AIRPORT, ATLANTA, GA

Being the busiest airport in the world, the Hartsfield-Jackson Airport has five runways to manage its takeoffs and landing. All the five runways are in the east to west direction and hence the winds coming from the 292.5º – 67.5º (northwest, north, north-east) and 112.5º - 247.5º (southeast, south, southwest) will be the crosswinds for the runways. Hence, there is a significant increase in average departure delay observed for the crosswinds with wind speed higher than 25 miles per hour as compared to other directions.

DENVER INTERNATIONAL AIRPORT, DENVER, CO

It is the largest airport in the United States in land area that expands to 33,531 acres. It has five runways, four in the east-west direction and one in the north-south. It is observed that there is a significant effect of the crosswinds and tailwinds on the delays. Denver experiences a higher average delay of 7.55 minutes due to crosswind and 7.40 minutes due to tail winds.

O'HARE INTERNATIONAL AIRPORT, CHICAGO, IL

The O'Hare International Airport at the Windy City is the busiest airport regarding the passenger traffic. It is the hub for American and Delta Airlines. It has eight runways to manage its air-traffic. Out of the eight runways, five are constructed in the east-west direction and two at inclined angles specially designed for crosswinds at the time of severe weather conditions. 70% of the departures are routed to the two east-west runways (28R and 27L). Due to severe weather conditions in Chicago, the Gust speed (sudden increase in the wind speed) is also taken into consideration.

It is observed that at higher Gust Speed (greater than 30 mph) the average delay for the inclined runway (4R-22L) and east-west runways (28R and 27L) is increased significantly due to crosswinds as well as the tailwinds. For the east-west runways, the winds coming from the north or south will be crosswinds. The northern crosswinds (less than 45 degrees) for the east-west runways is producing an average delay of 10.13 minutes. Adding to that, the tailwinds (45-112.5 degrees) coming from the north east and south west is causing an average delay of 7.55 minutes.

For the inclined runways (4R-22L) the winds coming from north becomes the tail winds and cause an average delay of around 10 minutes which is a significant increase in delay and can affect the take-offs. The crosswinds for the same runways cause an average delay of 7 minutes.

DALLAS-FORTWORTH INTERNATIONAL AIRPORT, DALLAS, TX

Dallas/Fort Worth International Airport is the primary international airport serving the Dallas–Fort Worth area in the U.S. state of Texas. DFW Airport has eight Major runways, six in north, south direction and two in Northwest and South East directions. Hence, the wind coming from east and west will be cross winds for the six runways and winds coming from North and South are crosswinds for the other two runways. It can be observed from the results that the average delay for East and West crosswinds increases the average delay to 10.5 minutes and the North and South crosswinds are increasing the delay to 6.7 minutes.

LOS ANGELES INTERNATIONAL AIRPORT, LOS ANGELES, CA

Los Angeles International Airport(LAX) is the largest and busiest airport in the Greater Los Angeles Area and the state of California, as well as being one of the largest international airports in the United States. LAX airport has four major runways, all in east-west directions. Hence, the winds coming from North and South will be cross wind the four runways. It is observed from the results that North winds increase the average delay time to 11.59 minutes and South winds increase the delay to 9.06 minutes. It is also observed that head winds and tail winds are also increasing the delay to 8 minutes.

AN ALTERNATIVE SOLUTION TO REDUCE THE IMPACT OF CROSSWINDS

After carefully observing the impact of the crosswinds on the departure delays, we can say that the crosswinds have a significant impact on the departure delays for all the five airports, and we would strongly recommend the aviation industry that there should be more runways created for crosswinds as they did at Chicago where they constructed the inclined runways which intersect with the other runways to minimize the delays during cross windy times.

4.2 EFFECT OF VISIBILITY ON AVERAGE DELAY

Aviation, probably more than any other mode of transportation, is greatly affected by weather. Visibility is one of the major climatic element, which affects the aviation industry. Visibility is one of the major cause of accidents and delays of flights. Visibility can be defined as the greatest horizontal distance at which a suitable object can be seen and recognized with the unaided eye. Visibility depends on the number of factors like the change in temperature, Dew point, and humidity. As the temperature increases, visibility also decreases.

Fog is one of the major factors that affects the visibility, more the fog lesser the visibility. When the dew point temperature and air temperature are equal, the air is said to be saturated. At that point, the air condenses into water droplets, which we see as fog. The lower the dew point, the drier the air, and vice versa. Dew point temperature is never greater than the air temperature. Therefore, if the air cools, moisture must be removed from the air, and this is accomplished through condensation. This process results in the formation of tiny water droplets that can lead to the development of fog, frost, clouds, or even precipitation. The air around us contains invisible moisture known as water vapor. The difference between a location's temperature and its dew point indicates how well the air can currently hold moisture. Because cool air has a harder time holding moisture than warm air, pilots use this information to determine how likely it is that fog, clouds, or precipitation will form.

The analysis is done on the five airports on the basis of higher Mean temperature (greater than 56.5 F), higher Mean Dew Points (greater than 55 F) which generates fog in the atmosphere causing the Mean Visibility less than 6.1 miles. The analysis confirms the above theory and it is observed that the closer the Mean Temperature and Mean Dew Points the more is the fog in the atmosphere which decreases the visibility and thus increases the is the Average mean delay.

4.3 EFFECT OF THE AIR DENSITY ON THE DELAYS

On a hot and humid day, the aircraft will accelerate more slowly down the runway, will need to move faster to attain the same lift, and will climb more slowly. The less dense the air, the less lift, the more lackluster the climb, and the longer the distance needed for takeoff and landing. These factors can lead to greater average delays. Density is directly proportional to pressure and inversely proportional to temperature and Humidity.

After analyzing the effect of air density on the average delays for Chicago, Atlanta and Los Angeles A temperature more than 56 F, humidity greater than 78% and mean sea level pressure less than 29.9 psi causes very low density in the air which leads to an average delay of 13 minutes at Atlanta. Similarly, in Chicago, the average delay is 10.39 minutes at a low density caused by higher temperature (greater than 65 F), more humidity (greater than 86.5%) and lower sea pressure (less than 29.435 psi). In Los Angeles, the average delay goes up to 13.11 minutes at a low density caused by high temperature (greater than 56.5 F), high humidity (78%) and low sea pressure (29.975 psi).

CONTACT ADVICE FOR THE AUTHORS

Please contact the authors of this paper

Piyush Gupta, MS Management Information System, is a Graduate student at Oklahoma State University with experience in Data Analytics, Visualization and Predictive Modeling using SAS and Tableau across Finance and Digital Marketing domain. He is pursuing a career in Advanced Analytics.

Contact- piyush.gupta@okstate.edu Phone- (206)-334-7251

Anuj Dwivedi, MS Management Information System, is a Graduate Student at Oklahoma State University with 9+ Certifications in Data Science and Machine Learning. He is a data enthusiast and an aspiring data scientist planning to complete his graduate studies in May 2017.

Contact- anuj.dwivedi@okstate.edu Phone- (405)-762-3865

Aditya Agrawal, MS Industrial Engineering & Management, is a Graduate Student at Oklahoma State University is a budding Industrial Engineer with experience in Additive Manufacturing, Data Analytics, Supply Chain Management and Operations Research. He has a penchant in Supply Chain Management and Data Analytics

Contact- aditya.agrawal@okstate.edu Phone- (405)-762-3778

REFERENCES

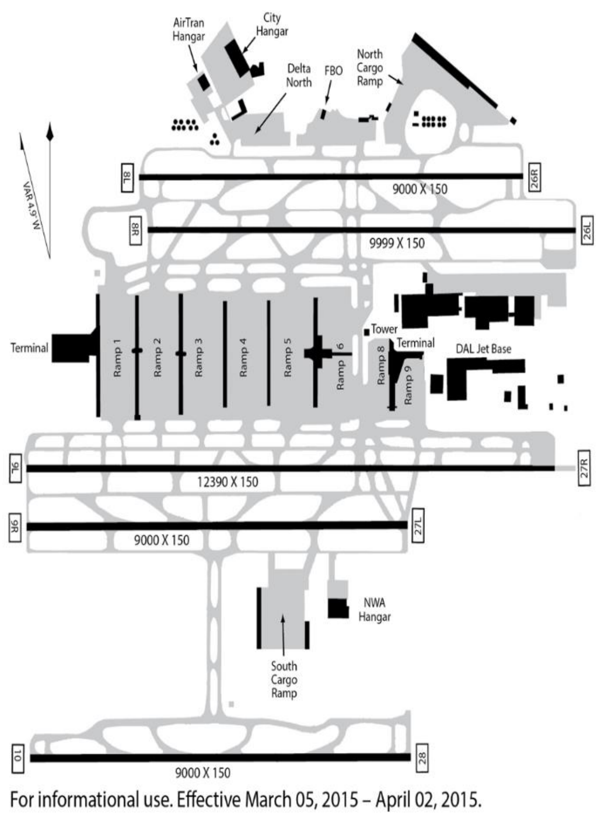
SULE BULUS (2014). "*The Effects of Visibility On Flight Operations In Kaduna International Airport, Kaduna"*

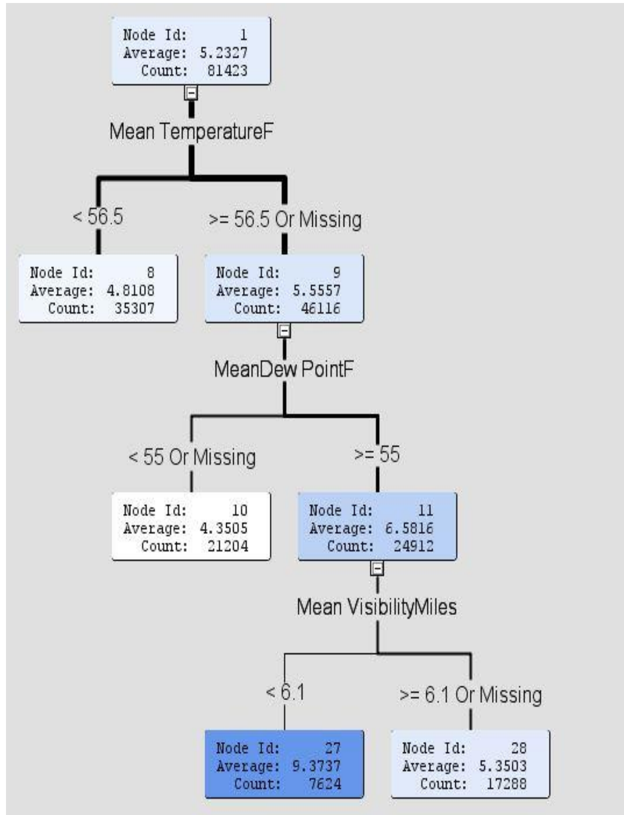
https://www.faa.gov/airports/planning_capacity/profiles/media/ORD-Airport-Capacity-Profile-2014.pdf

Gano B. Chatterji and Yun Zheng. "*Wheels-off Time Estimation at Non-ASDE-X Equipped Airports."*

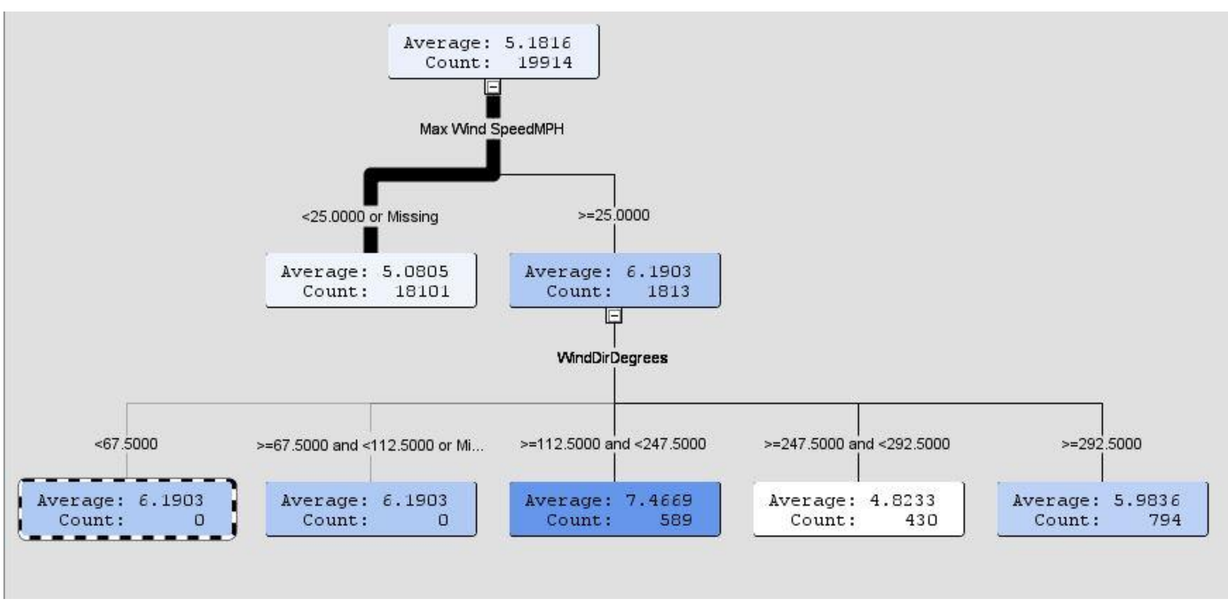
APPENDIX

ATLANTA

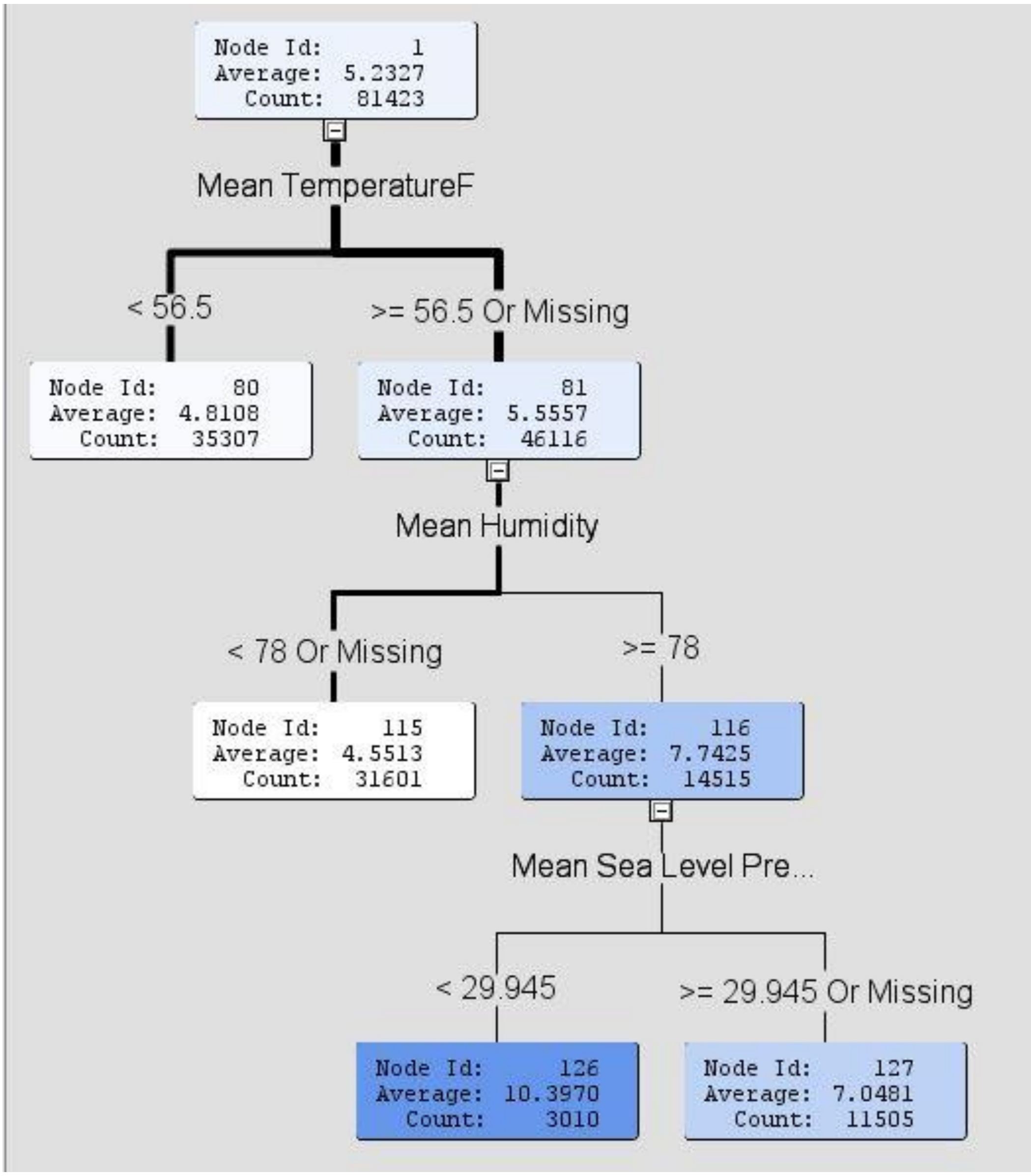




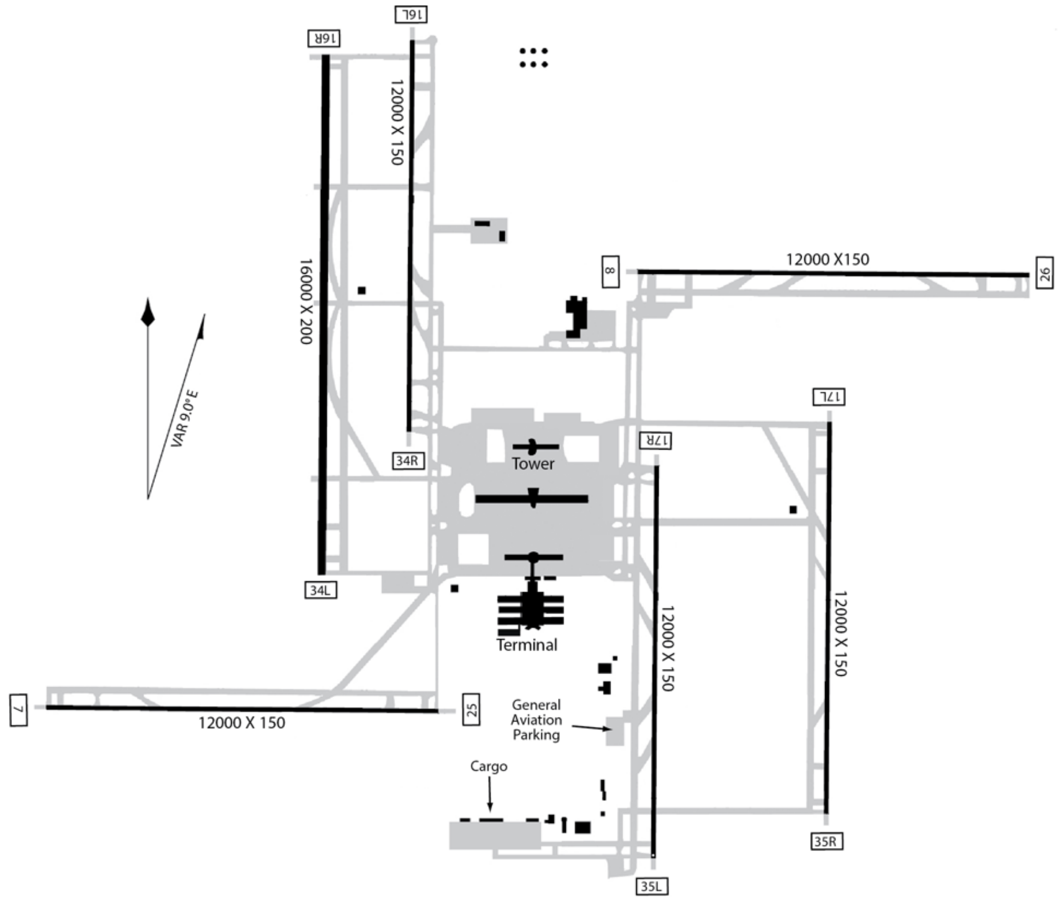
http://www.faa.gov/NextGen/snapshots/airport/?locationId=54



DENSITY ANALYSIS ON AVERAGE DELAY

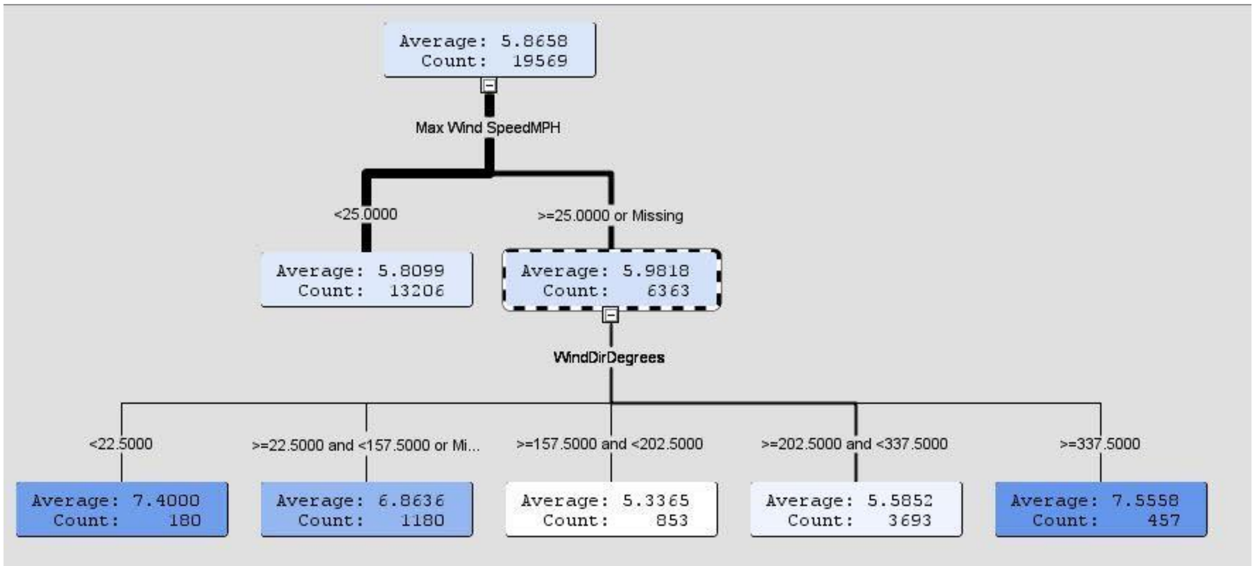


DENVER

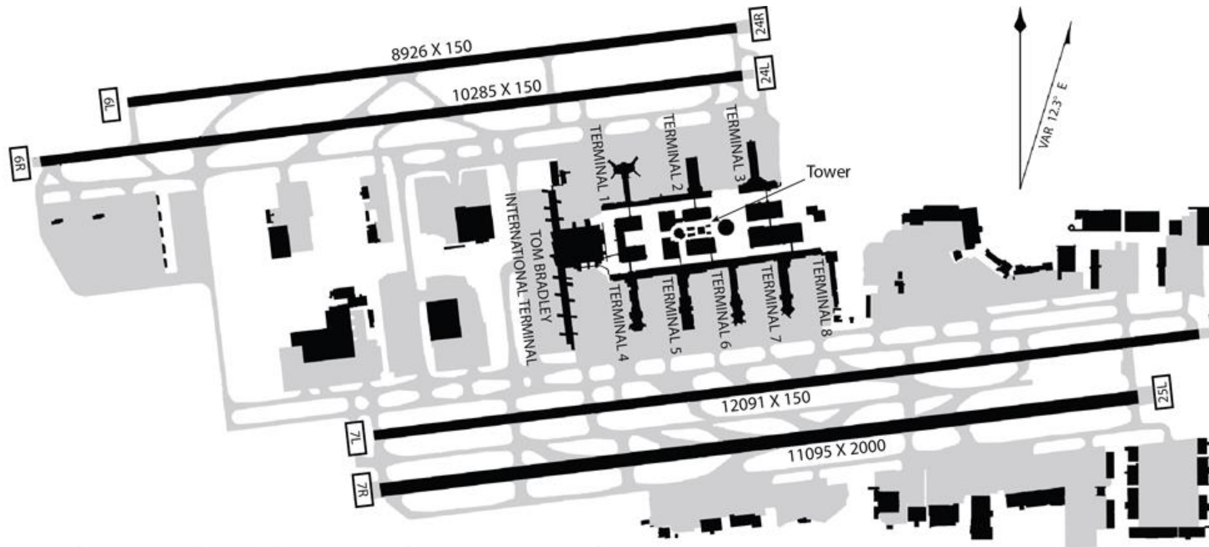


For informational use. Effective March 05, 2014 – April 02, 2015.

http://www.faa.gov/nextgen/snapshots/assets/img/airports/maps/DEN.png

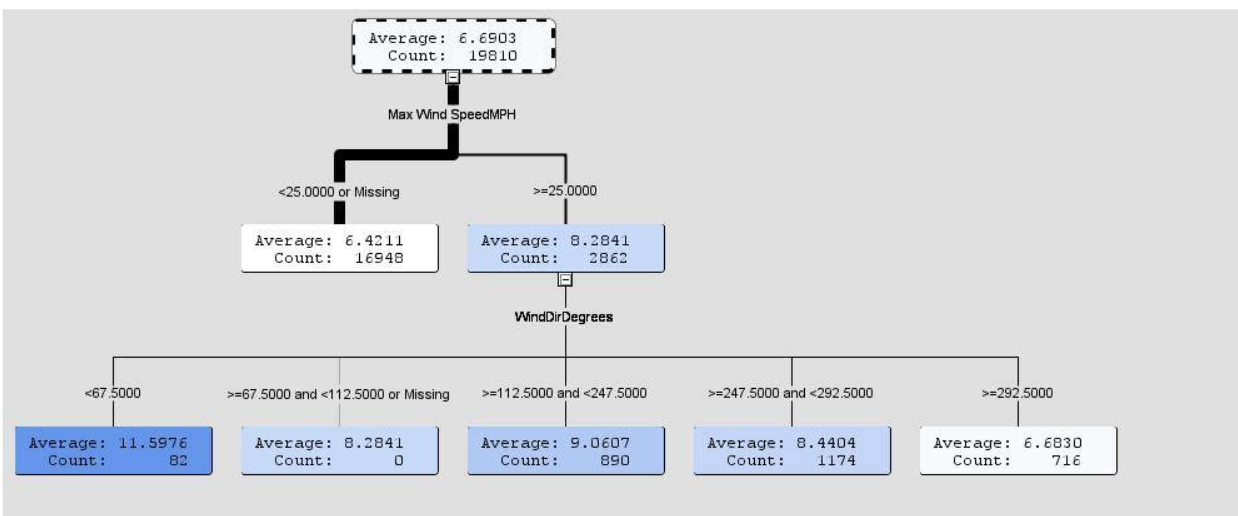


LOS ANGELES

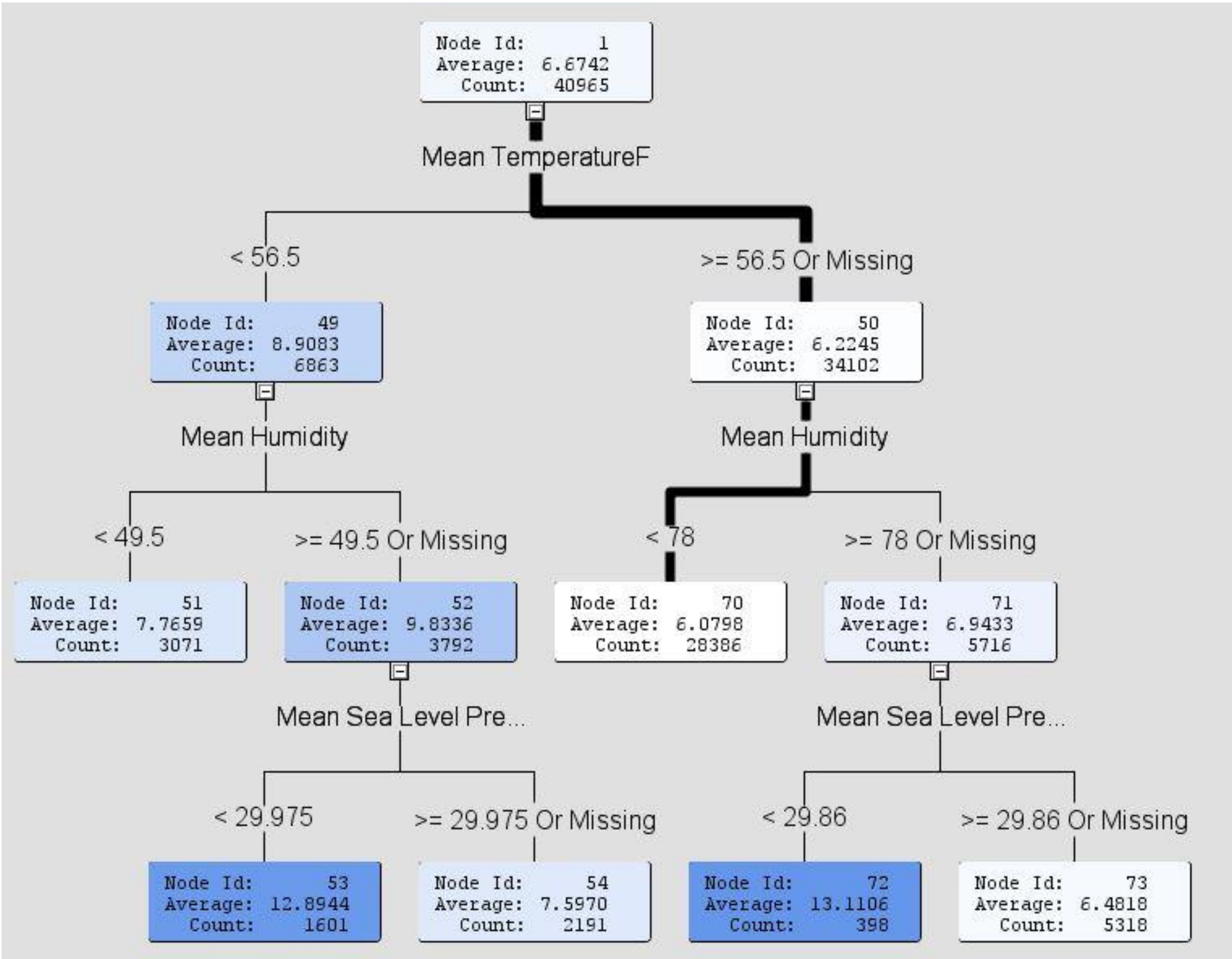


For informational use. Effective March 05, 2015 – April 02, 2015.

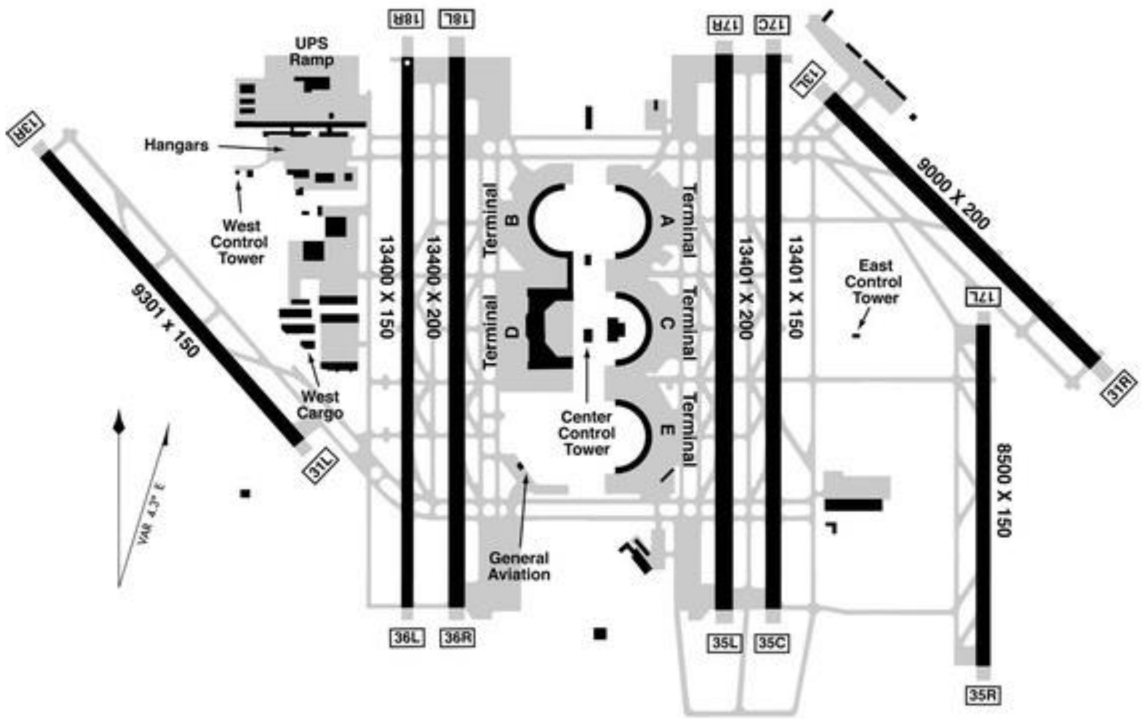
http://www.faa.gov/nextgen/snapshots/assets/img/airports/maps/LAX.png



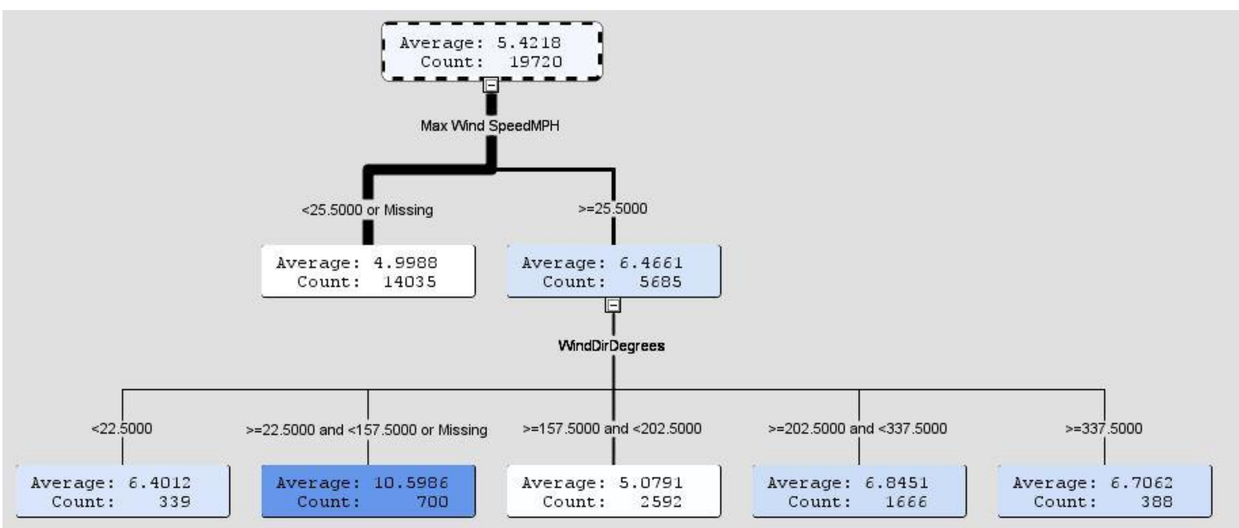
DENSITY ANALYSIS ON AVERAGE DELAY



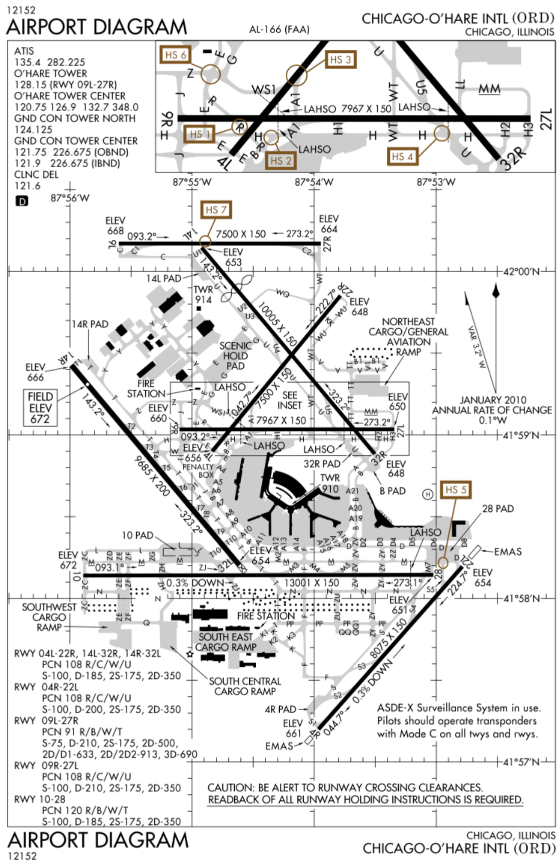
DALLAS

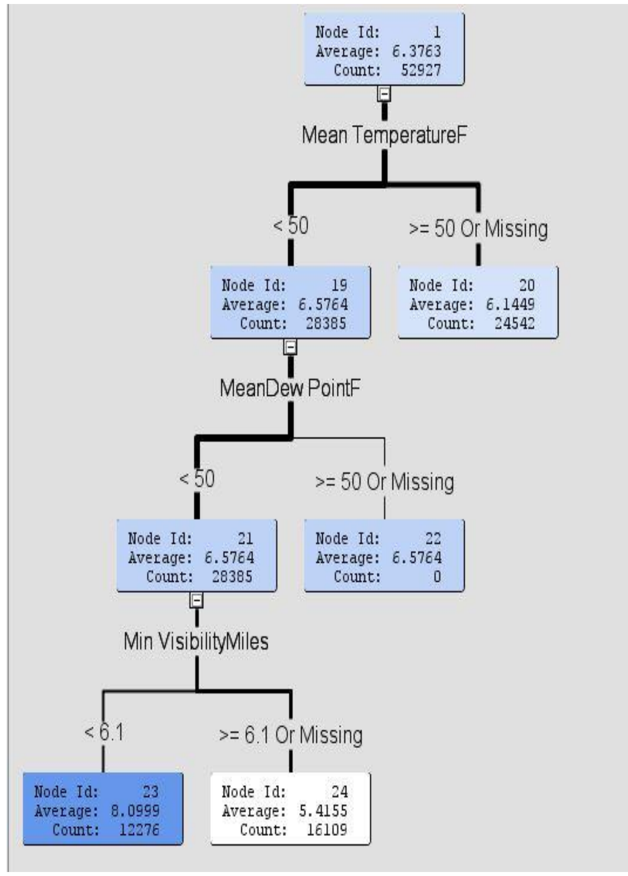


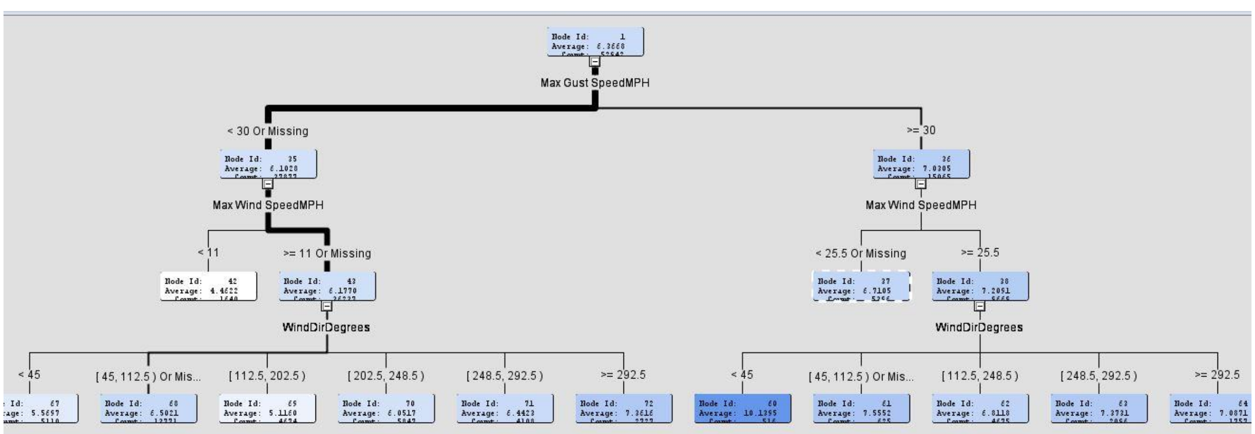
https://qph.ec.quoracdn.net/main-qimg-59a8c1f24abc1326d4e11abbdfc8c78b-c?convert_to_webp=true

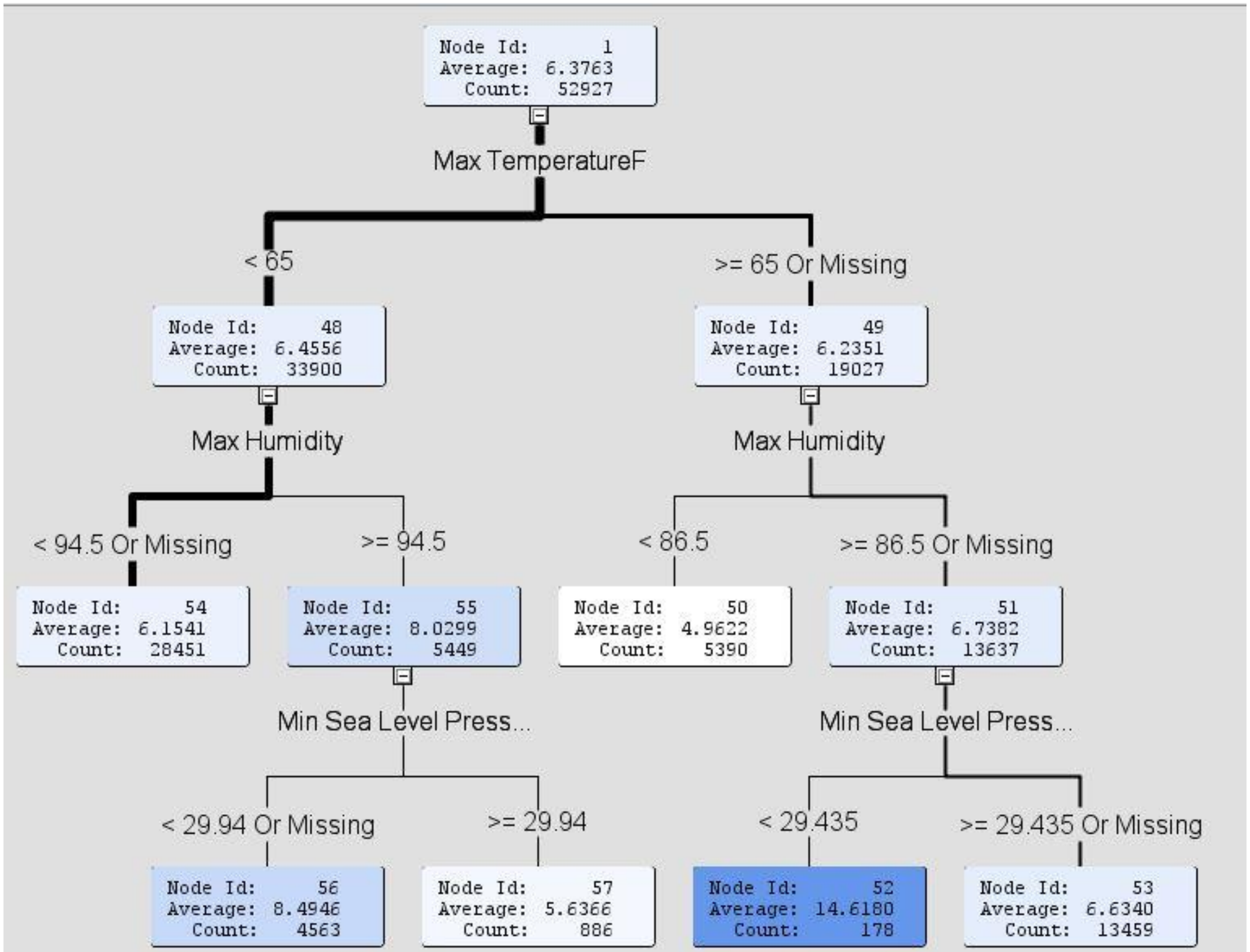


CHICAGO


12152
AIRPORT DIAGRAM
AL-166 (FAA)
CHICAGO-O'HARE INTL (ORD)
CHICAGO, ILLINOIS
ATIS
135.4 282.225
O'HARE TOWER
128.15 (RWY 09L-27R)
O'HARE TOWER CENTER
120.75 126.9 132.7 348.0
GND CON TOWER NORTH
124.125
GND CON TOWER CENTER
121.75 226.675 (OBND)
121.9 226.675 (IBND)
CLNC DEL
121.6
JANUARY 2010
ANNUAL RATE OF CHANGE
0.1°W
NORTHEAST CARGO/GENERAL AVIATION RAMP
SOUTHWEST CARGO RAMP
SOUTH EAST CARGO RAMP
SOUTH CENTRAL CARGO RAMP
FIRE STATION
SCENIC HOLD PAD
FIELD ELEV 672
RWY 04L-22R, 14L-32R, 14R-32L
PCN 108 R/C/W/U
S-100, D-185, 2S-175, 2D-350
RWY 04R-22L
PCN 108 R/C/W/U
S-100, D-200, 2S-175, 2D-350
RWY 09L-27R
PCN 91 R/B/W/T
S-75, D-210, 2S-175, 2D-500, 2D/D1-633, 2D/2D2-913, 3D-690
RWY 09R-27L
PCN 108 R/C/W/U
S-100, D-210, 2S-175, 2D-350
RWY 10-28
PCN 120 R/B/W/T
S-100, D-185, 2S-175, 2D-350
ASDE-X Surveillance System in use. Pilots should operate transponders with Mode C on all twys and rwys.
CAUTION: BE ALERT TO RUNWAY CROSSING CLEARANCES. READBACK OF ALL RUNWAY HOLDING INSTRUCTIONS IS REQUIRED.
EC-3, 28 JUN 2012 to 26 JUL 2012
AIRPORT DIAGRAM
12152
CHICAGO, ILLINOIS
CHICAGO-O'HARE INTL (ORD)


Node Id: 1
Average: 6.3763
Count: 52927
Mean TemperatureF
< 50
>= 50 Or Missing
Node Id: 19
Average: 6.5764
Count: 28385
Node Id: 20
Average: 6.1449
Count: 24542
MeanDew PointF
< 50
>= 50 Or Missing
Node Id: 21
Average: 6.5764
Count: 28385
Node Id: 22
Average: 6.5764
Count: 0
Min VisibilityMiles
< 6.1
>= 6.1 Or Missing
Node Id: 23
Average: 8.0999
Count: 12276
Node Id: 24
Average: 5.4155
Count: 16109


Max Gust SpeedMPH
< 30 Or Missing
>= 30
Max Wind SpeedMPH
< 11
>= 11 Or Missing
Max Wind SpeedMPH
< 25.5 Or Missing
>= 25.5
WindDirDegrees
< 45
[45, 112.5) Or Mis...
[112.5, 202.5)
[202.5, 248.5)
[248.5, 292.5)
>= 292.5
WindDirDegrees
< 45
[45, 112.5) Or Mis...
[112.5, 248.5)
[248.5, 292.5)
>= 292.5



RELATIONSHIP BETWEEN CLOUD COVER AND HUMID