



AIR CRASH ANALYSIS

BOEING VS AIRBUS

A. Introduction

A.1. Description & Discussion of the Background

In our day to day life we encounter several accidents. Accidents may happen on road, water and air. However, Air accidents cause severe damages and brings more risks for whom travel by air. Statistically, air travel is safer than other source but when something goes wrong during air travel, it can be catastrophic with hundreds of lives at stake.

Since there are plenty of aircrafts, **Boeing and Airbus are considered for this study**. Boeing and Airbus are the most powerful commercial jets in the world. Even these aircraft manufacturers have different types of aircrafts with lots of safety measurements, Accidents are been reported every year. When there is a 0% reported air accidents over the certain period, it can be considered as 100% growth in aircraft technology.

From **1990 to 2020**, there were many major accidents happened due to many reasons. In this analysis, The Cause of the accident, damages, Airlines and other facts are been analyzed. From the analyzed data, Safest airlines and aircrafts can be found using data science.

A.2. Data Description

Collected data are from below sources.

- Accident data source - aviation-safety.net
- Aircraft Registrations - Wikipedia.com

Accident rates between Boeing and airbus aircrafts, accident rates between each types of aircraft, locations where accident happened, reasons for the accidents are analyzed in this study.

Below tools are been used for the analysis,

- geocoder to get location address, latitude and longitude.

- Folium to point the accident places in the map.
- Foursquare to find near affectable places.

A.3. Data Acquisition and Cleaning

Most of the dataset are downloaded from official websites. After downloading the data sets, major problem was transforming the data into readable csv or excel file as those data were not in proper format. Several unwanted columns and all null value rows were dropped to avoid complication. 50% of records were invalid data, so preprocessing each data row by row consumed more time.

While generating latitude and longitude of locations, Attribute errors and Geocoder Timed Out errors were popped up continuously. Exception error handling method was used to avoid these errors. All data were transferred to csv files and imported as csv file through pandas library to overcome the above discussed errors. Some modifications were done on csv files to obtain the valid numbers. After processed the data, Boeing database reduced from 800 to 300 records.

B. Data Analysis and Visualization

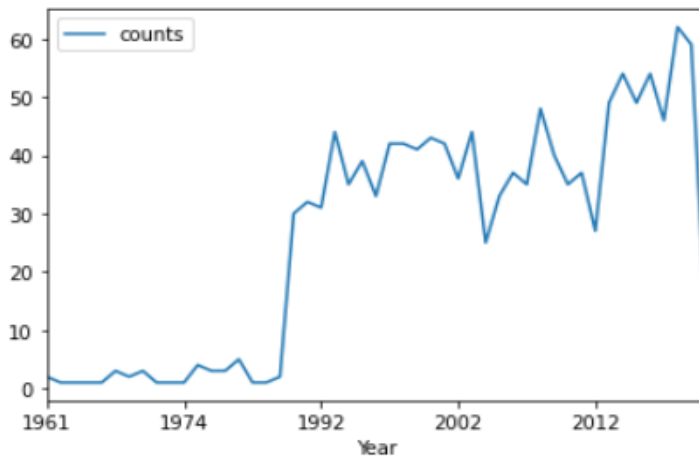


Figure 1

Figure 1 shows the variations in air accidents per year in Airbus and Boeing. Reported total death is higher than 15000, starting from year 1990. Air accidents are considered to be riskier than land accidents. Death due to air accident can't be predicted as counts of death per accidents are really high. Aircrafts manufactured by Airbus and Boeing encounter 30 air accidents averagely per year. Necessary precautions have to be taken by Authority and Manufacturing companies.

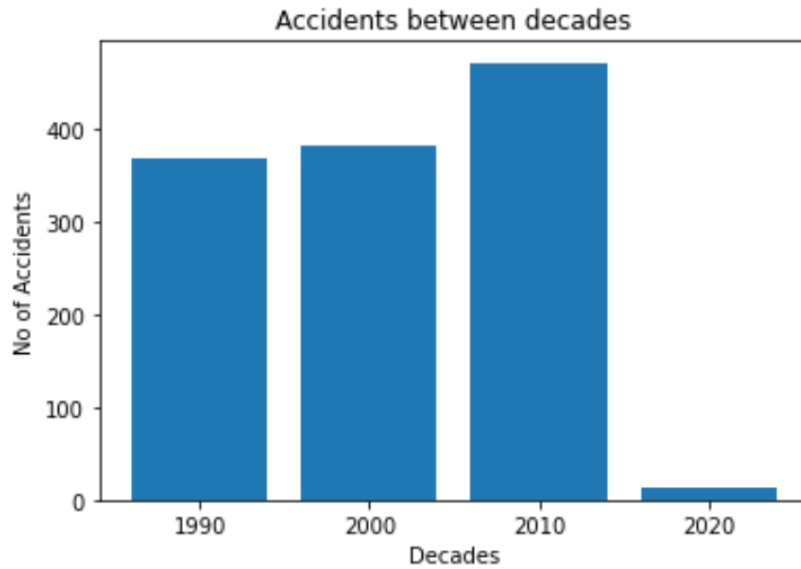


Figure 2

Figure 2 shows the variation of accidents happened in each decade.

According to the above graph no of Airbus and Boeing Accidents are given below,

1990's: 369

2000's: 383

2010's: 472

2020 January to April : 13

Still we have 9 ½ years to go to pass this decade. Can you imagine how the rate going to increase? Every decade, rate of change in accidents is low. But number of air accidents are increasing.



Figure 3

Below figure represents the places of accident in world map. Overall, most of the accidents happened in American Continent as per figure 3. If cities are considered, London and New York would be the cities reporting higher number of accidents.

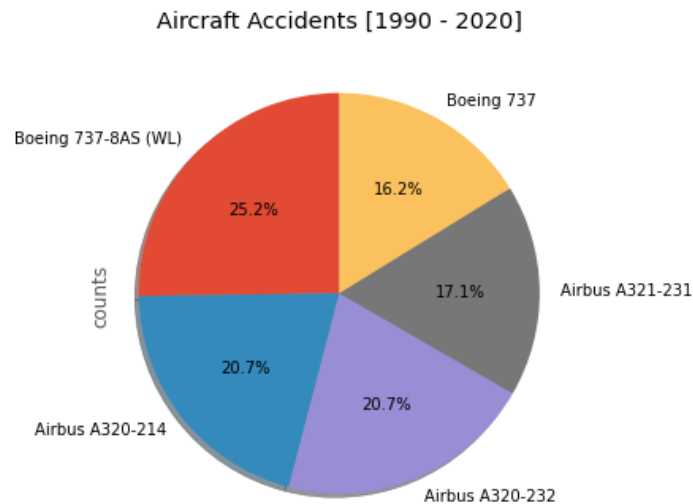


Figure 4

If we analyse accidents on aircraft type wise, Boeing model 737 series were reported for more accidents, approximately 41.7% of overall accidents. Also, in 2019 these types of aircraft had error in autopilot stall warning. Other than Boeing, Airbus A320 series reported for 41.4% of overall accidents.

Aircraft Accidents Categories [1990 - 2020]

Accident Categories

A = Accident

I = Incident

H = Hijacking

C = Criminal occurrence (sabotage, shoot down)

O = other occurrence (ground fire, sabotage)

1 = hull-loss

2 = repairable damage

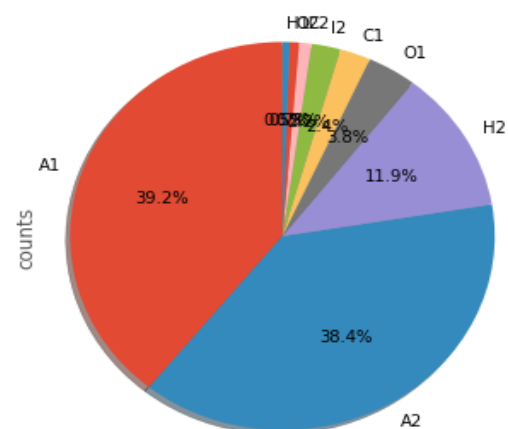


Figure 5

According to the above results, Most of the Accidents are hull-loss. Also 151 cases of Hijacking.

C. Model Development and Methodology

All the data used in model development based on Boeing Aircrafts. Two decision tree models were developed to predict the damage and fatal loss, if incident or Accident happened.

Independent Variables(X) are

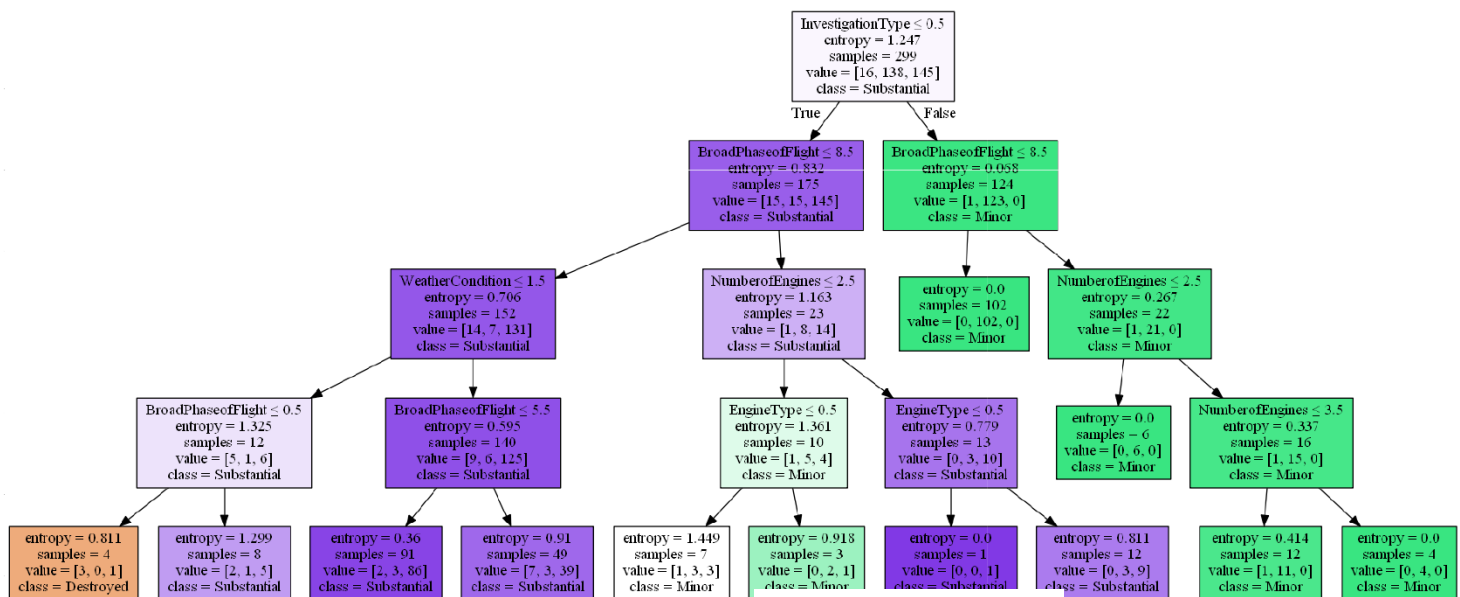
- Investigation Type - Accident or Incident
- Engine Type - Turbofan or Reciprocating or Turbojet or Turboprop
- Weather Condition - VMC (visual meteorological conditions) or IMC (Instrument meteorological conditions) or UNK(Unknown)
- Broad Phase of Flight -Landing or Takeoff or Taxi or Climb or Cruise or Approach or Standing or Decent or Go-Around or Maneuvering
- Number of Engines - 1 or 2 or 3 or 4

C.1. Decision Tree – Aircraft Damage

Damage can be Minor or Substantial or Destroyed. These cases taken as target variable. The Model have to predict the nature of damage by using independent variables.

Accuracy of this model is 90.66%. It means, model is good to make the predictions.

In train, test split 80% of data taken for training and 20% for testing.



Prediction testing,

```
Prediction = ['Substantial' 'Minor' 'Minor' 'Minor' 'Minor' 'Minor'
              'Substantial' 'Substantial' 'Substantial' 'Substantial']
Original = [ 'Substantial' 'Minor' 'Minor' 'Minor' 'Minor' 'Minor'
            'Substantial' 'Substantial' 'Substantial' 'Substantial']]
```

In prediction, all 10 target test variables matches with Predicted variable.

C.2. Decision Tree – Injury Severity

Target Variable is injury severity. It can be Fatal or Non-Fatal. The model predict if accident happened, injury will be fatal or non-fatal or incident.

Accuracy of this model is 94.66%. It means, model is good to make the predictions.

In this model also train, test split 80% of data taken for training and 20% for testing.

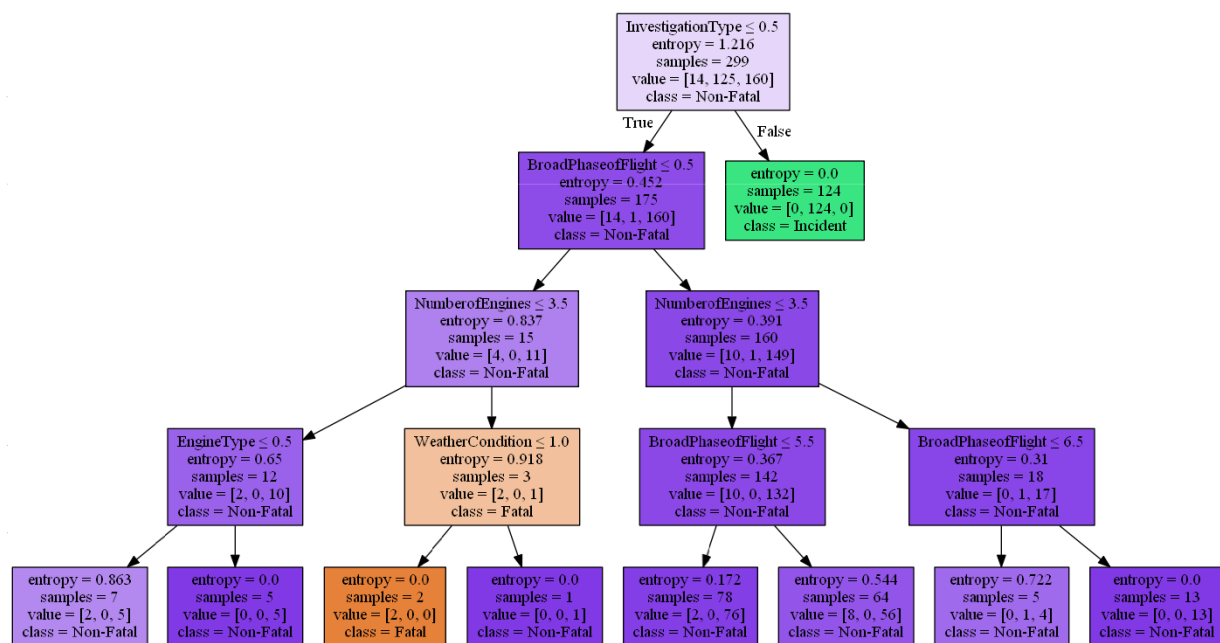


Figure 7

Prediction testing,

```
Prediction = ['Non-Fatal' 'Incident' 'Incident' 'Incident' 'Incident']
Original    = 'Fatal'    'Incident' 'Incident' 'Incident' 'Incident'
```

Fatal is predicted as non-fatal, others are predicted correctly.

Foursquare

Foursquare used to mark the red zone where most of the accidents are happened. According to the data most of the accidents happened in Miami. About 19 accidents happened in a year while landing. So 300m radius places around Miami Airport spotted as danger zone.

This area includes 5 public places including children's park. Those areas pointed in red mark in below map.



Figure 8

D. Results and Discussion

In this analysis two databases were used to develop the model and get overall idea about Air crash accidents. Analysis and Visualization part focused on basic idea on what happened in past. According to the figure 2, Air accidents are increasing every decade. Basic reason for increasing in Accident rate is increase in air transport users.

In our analysis most of the accidents happened in Europe and America, where busiest airports are situated. In these accidents more than 20% of Accidents were happened while landing.

from the data we can see some factors give huge impact on these accidents, those are Investigation Type, Engine Type, Weather Condition, Broad Phase of Flight and Number of Engines. from these data we can predict if accident or incident happens, will it minor or major accident.

To implement the model, Decision Tree method was used. Also, Foursquare method was used to get surrounded area of accident.

E. Conclusion

Purpose of this project is analysis the Air accidents and its factors. Decision Tree method was used and accuracy of the model is more than 90%. Here two models were made to predict Aircraft Damage and Injury Severity, respectively both have 90.66% and 94.66% accuracy. So model is good for predictions.

Most of the Accidents gives fatal damages to the Aircraft than Incident. number of Engines and Engine type also plays major roles in accidents. If number of engines are 4, then most of the accidents are unrecoverable.

Weather condition is also another key factor. if weather is under IMC, then damage will be huge and unpredictable weathers also will cause disaster. other than that, most of the accidents were happened while landings.