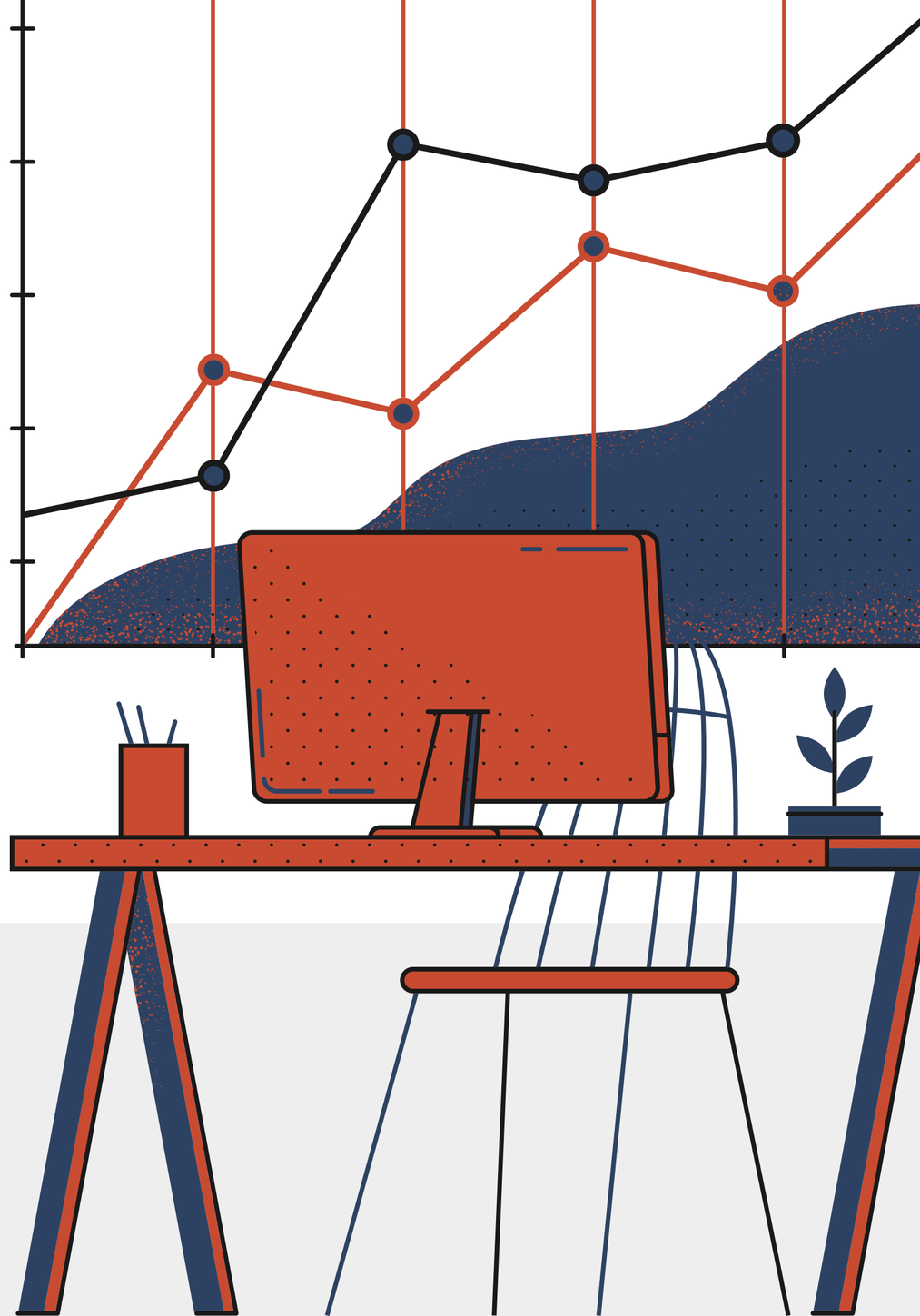
SC1015 MINI-PROJECT

DATASET: AVIATION INCIDENT



GROUP 3

MEMBERS: 1. Koh kai jie,Farrel
2. Wong Xiang Rui
3. Quek Wei Quan

Table of Contents

I Research Background & Motivation

II Cleaning of Data set

III Explorative Data Analysis

IV Machine learning

V Conclusion & insights driven

BACKGROUND

Latest plane crash news

Top Stories Latest News Discover Singapore Asia Commentary Sustainability CNA Insider Lifestyle Watch Listen + All Sections

World

DHL plane breaks in two during emergency landing in Costa Rica airport

A DHL cargo plane broke in two after an emergency landing at the Juan Santamaria internati
Rica, on Apr 7, 2022 (Photo: AFP/Ezequiel Becerra)

SAN JOSE: A DHL cargo plane carrying mail and packages skidded off th
and broke in two during an emergency landing in Costa Rica on Thursda
causing the temporary closure of the international airport in San Jose.

South China Morning Post

China / Politics

China Eastern Airlines flight MU5735: second black box sent for decoding

- The plane's data recorder could reveal details of its speed and altitude before it crashed in southern China
- The black box was found buried under 1.5 metres of soil, state broadcaster CCTV reports

Luna Sun and Jack Lau
Published: 11:09am, 27 Mar, 2022

Why you can trust SCMP

China Eastern Airlines flight MU5735 crash

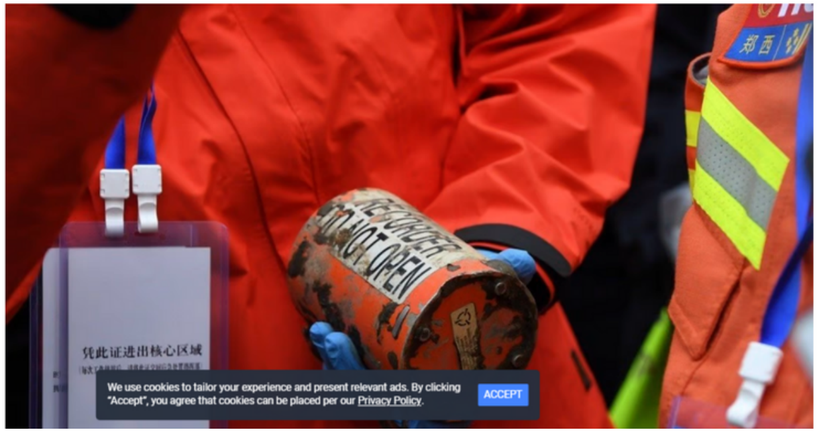
+ FOLLOW

TOP PICKS

US-China Relations
China's support for Russia galvanises West as US examines action against Beijing
12 Apr 2022

News
China delivered FK-3 missile system to Serbian military, state media says
11 Apr 2022

China Macro Economy
China speeds up inward economic push with plan for



We use cookies to tailor your experience and present relevant ads. By clicking "Accept", you agree that cookies can be placed per our Privacy Policy.
ACCEPT

Top Stories Latest News Discover Singapore Asia Commentary Sustainability CNA Insider Lifestyle Watch Listen + All Sections

Second black box found in China Eastern plane crash

27 Mar 2022 11:41AM
(Updated: 27 Mar 2022 07:36PM)

reckage at the site of the China Eastern plane
/Zhou Hua, Xinhua)

red from the crash of a China
week, Chinese state media said

MU5735 was recovered on March

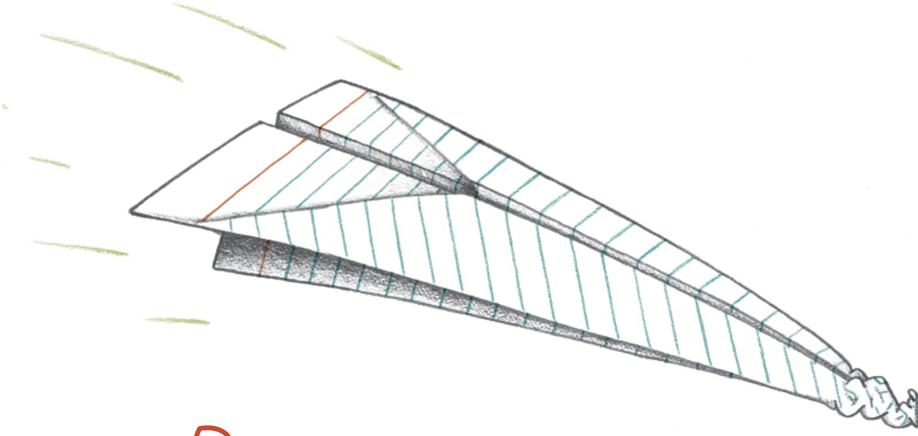
: One in the rear passenger cabin
recorder.

Related Topics

China Eastern Airlines crash | China | aviation

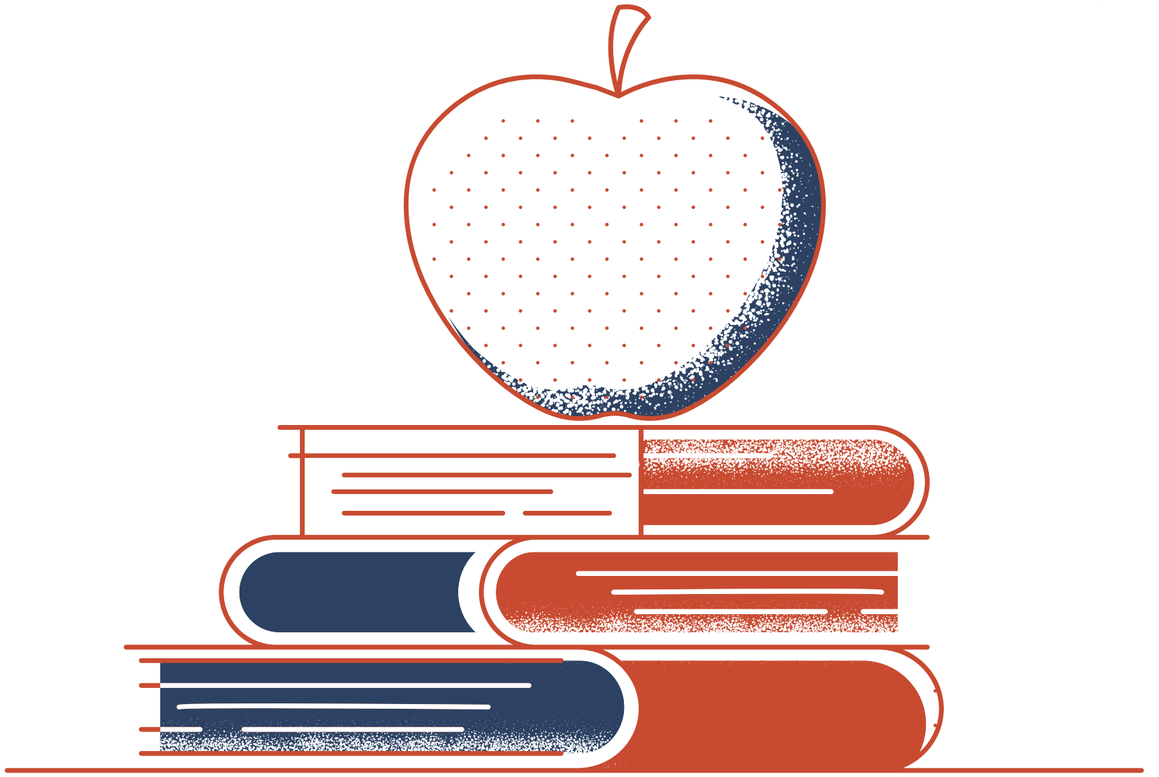
Also worth reading

Research Background & Motivation



PROBLEM STATEMENT

To identify the factor that has the most impact on aviation incident.



CLEANING DATA SET

Selecting appropriate data:

- Selecting data columns with the least amount of missing values
- Data that have a great influence on answering our problem statement
- Data inputs that could be used for machine learning

Selected Data sets

- Event Date
- Location
- Country
- Company
- Number of Engines
- Injury Severity
- Total Fatal Injuries
- Weather Condition
- Broad phase of flight
- Investigation Type

CLEANING DATA SET

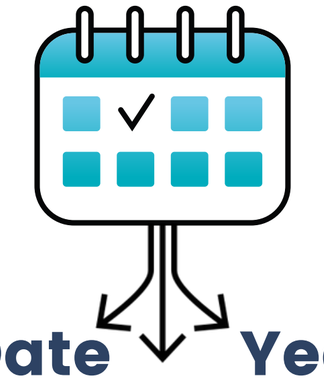
Data

Accident

Incident

Number of Engine & Severity of injury

EMPTY



Date
Year
Month

Convert Empty Rows into

UNKNOWN

Conversion similar terms to same terms

Y/N → YES/NO

Objects to

12345

67890

Int 64 and Category

CLEANING DATA SET

Over removal of data

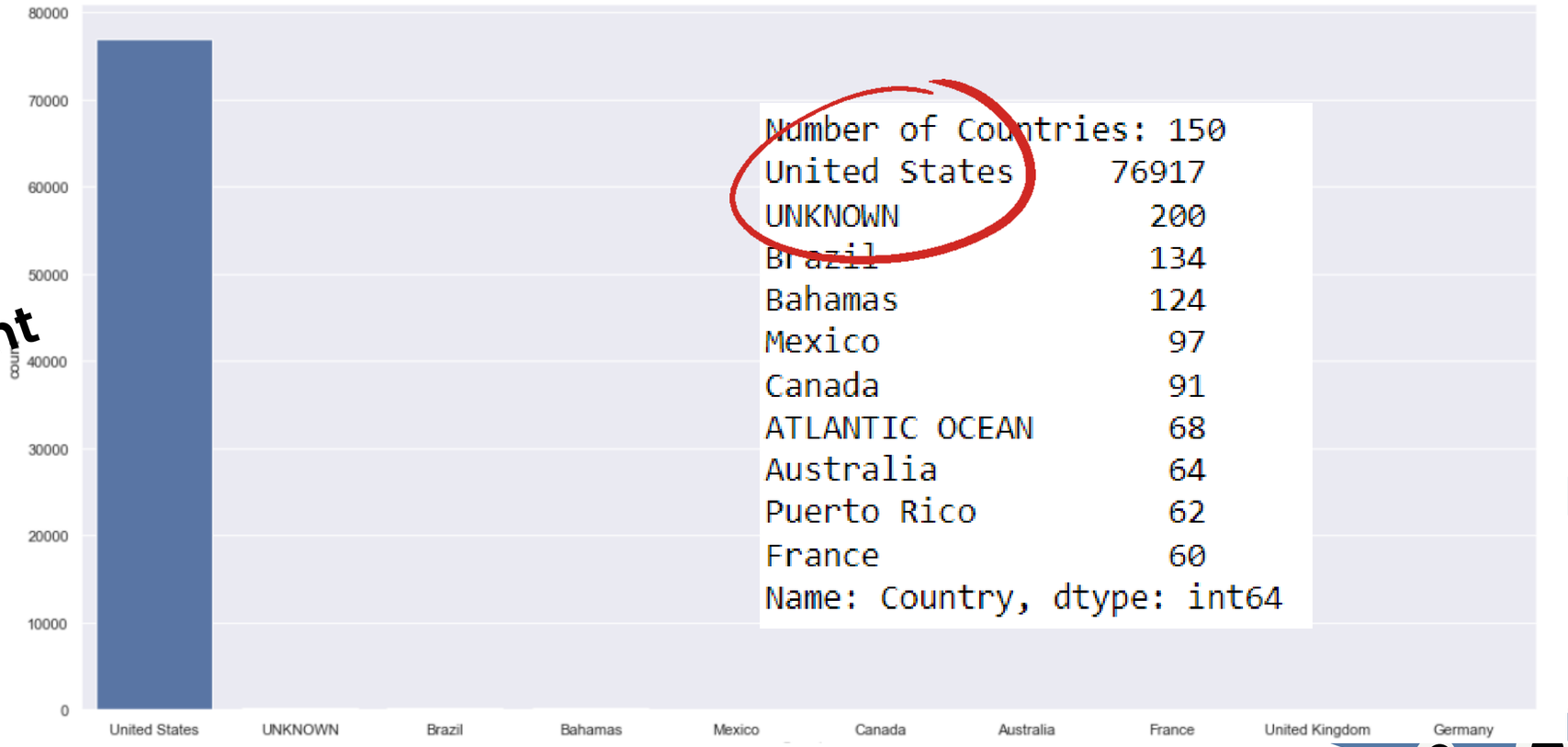
- This could result in false data appearing
- Data might appear too "fake" after confusion matrix is done

Removal of selected data set

- Other variables such as the phase of flight were considered.
  - However, they were removed due to the high number of unknowns.

Explorative Data Analysis

Top 10 countries with the highest number of accidents



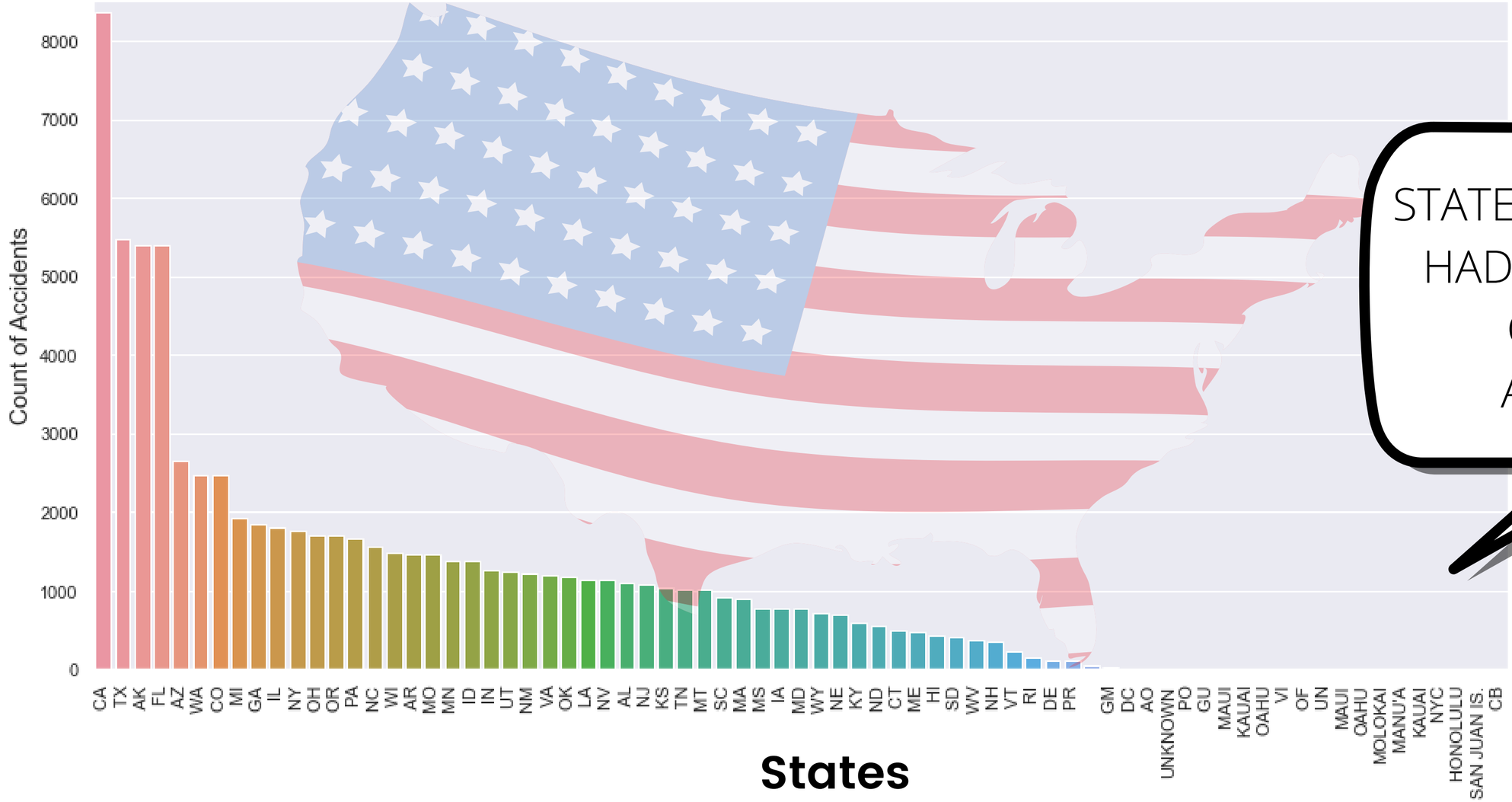

```
Number of Countries: 150
United States     76917
UNKNOWN             200
Brazil              134
Bahamas             124
Mexico               97
Canada               91
ATLANTIC OCEAN       68
Australia            64
Puerto Rico          62
France               60
Name: Country, dtype: int64
```

No. of accident counts

Countries

Explorative Data Analysis

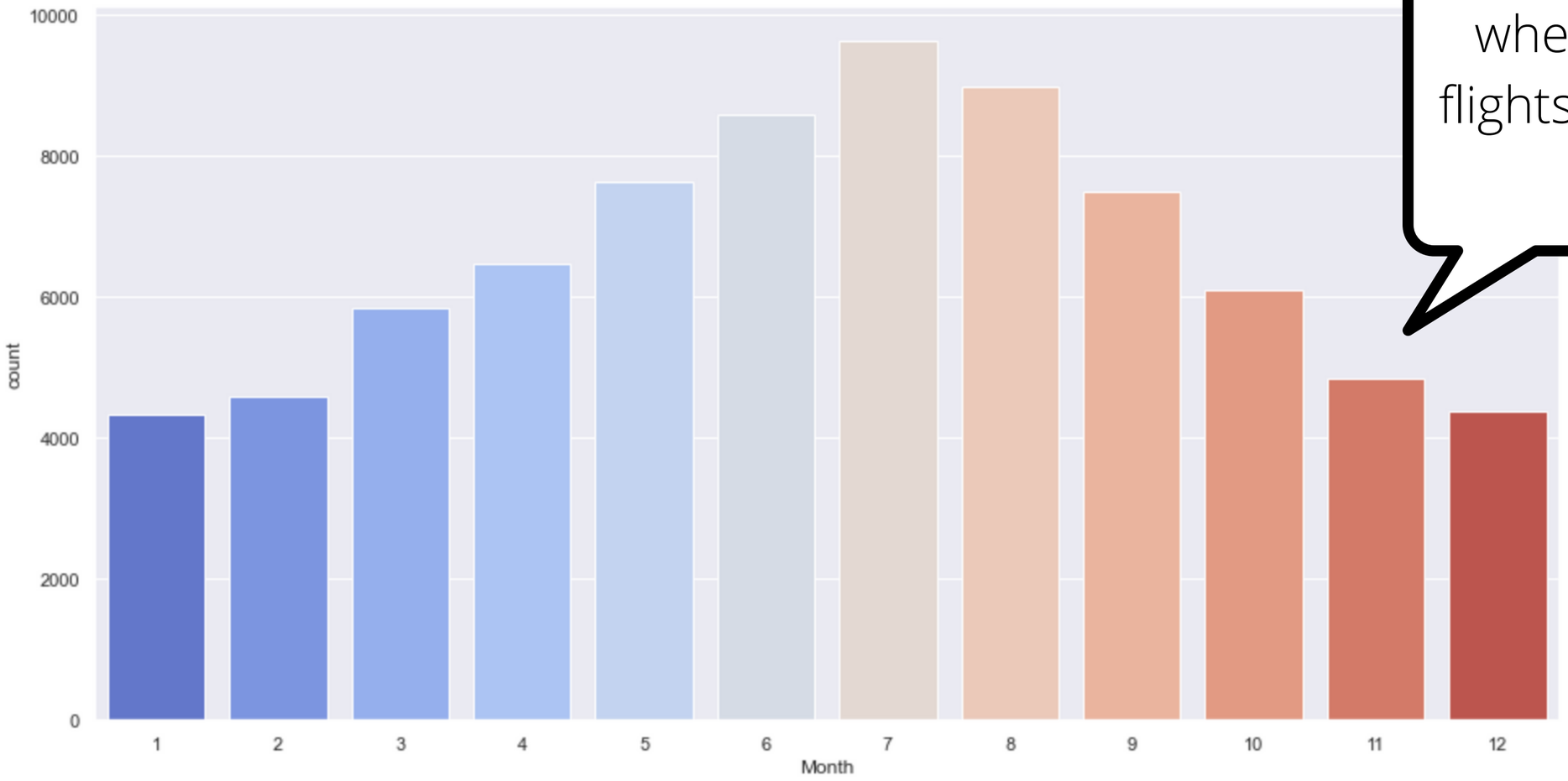
No. of Accident occurred for each state in the US



STATE OF CALIFORNIA HAD THE **HIGHEST** COUNT OF ACCIDENTS

III Explorative Data Analysis

Time Series of Injuries



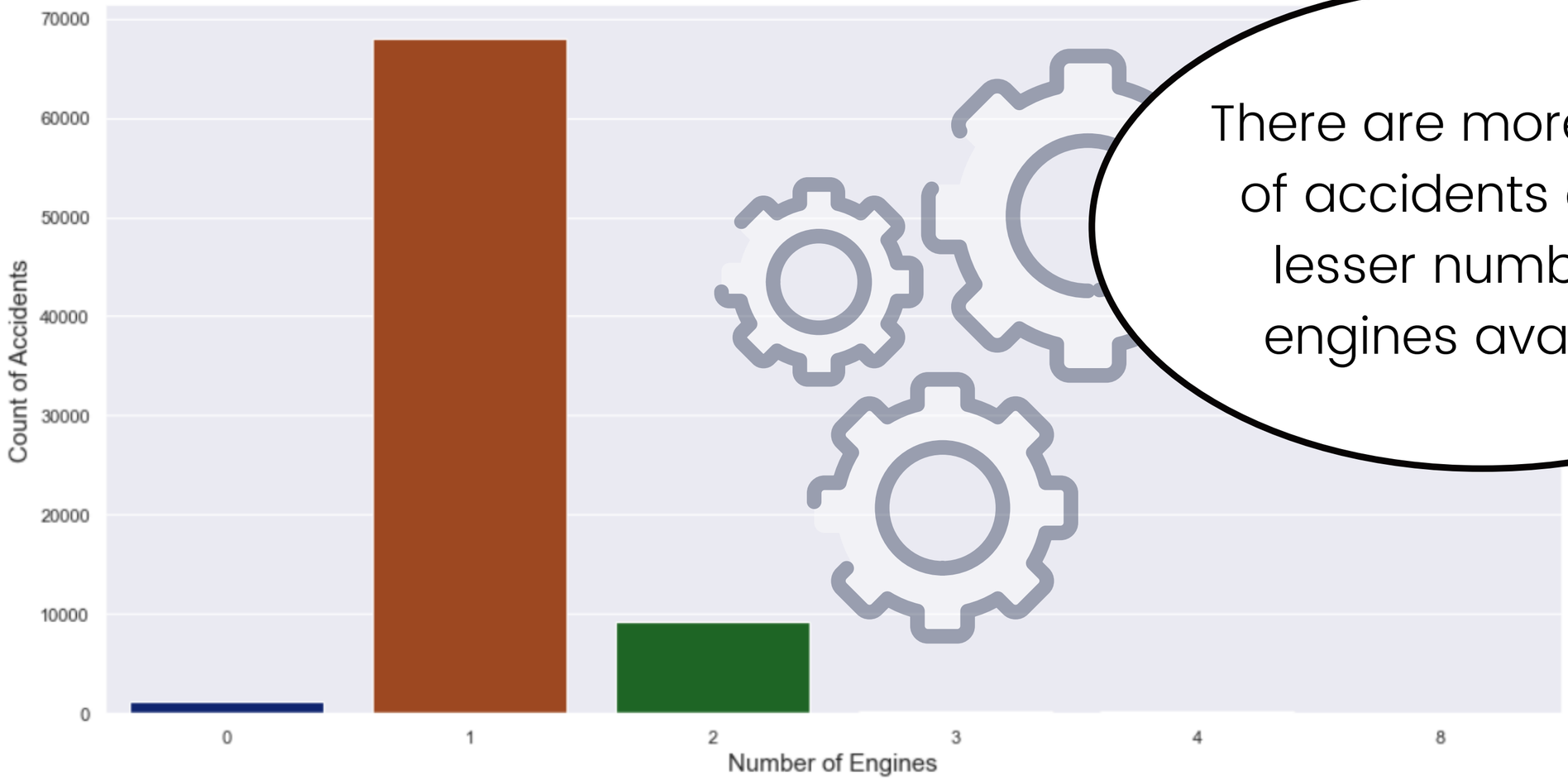
There is a correlation to the summer season when there are more flights during this period

III Explorative Data Analysis

Top 10 Companies with the highest number of accidents

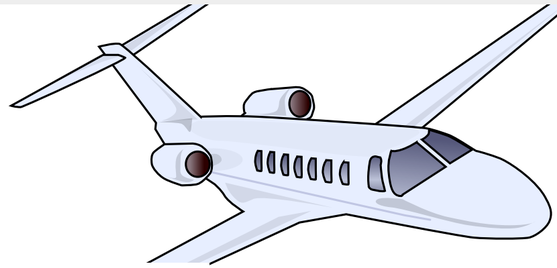
Explorative Data Analysis

Number of Engines in relation to the count of accident

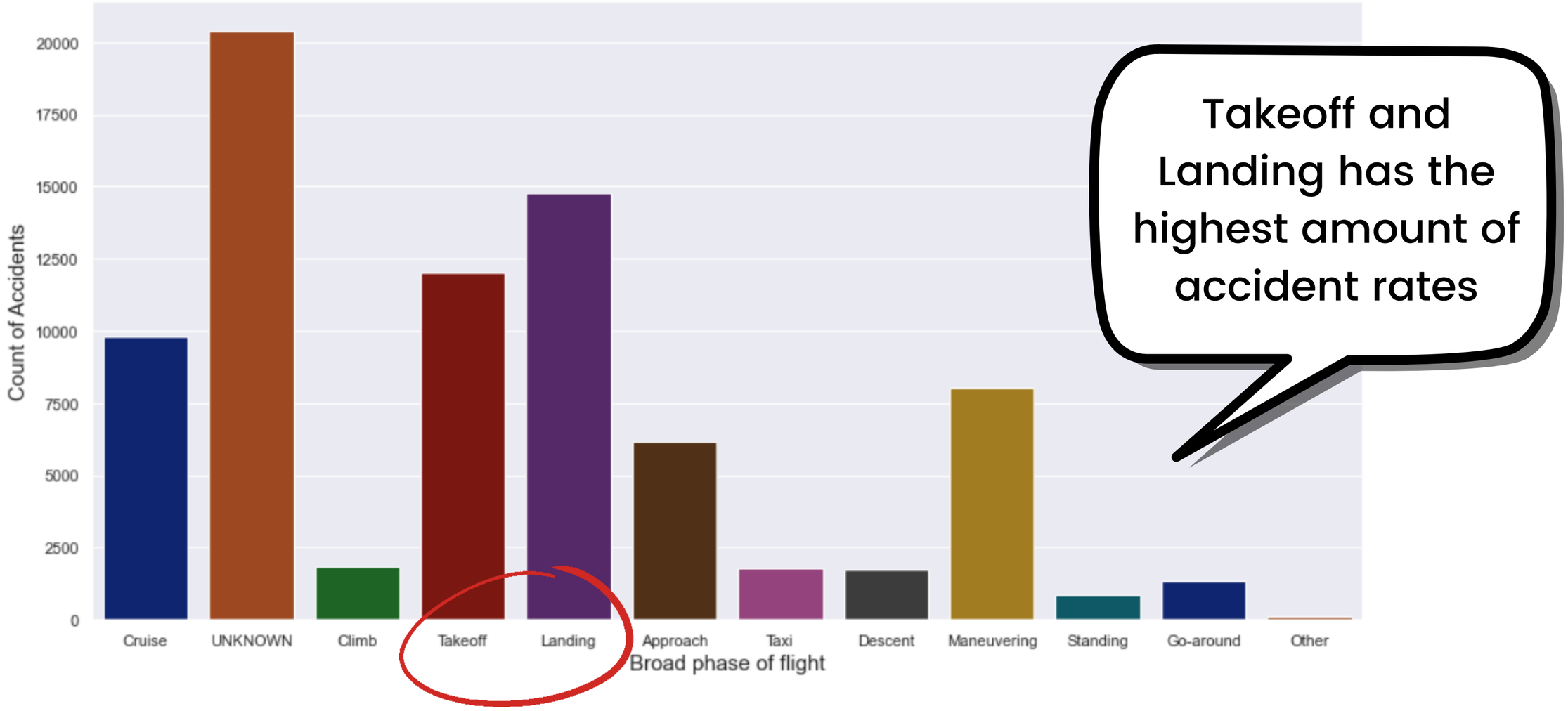


There are more cases of accidents due to lesser number of engines available

III Explorative Data Analysis



Phase of flight accident scenarios

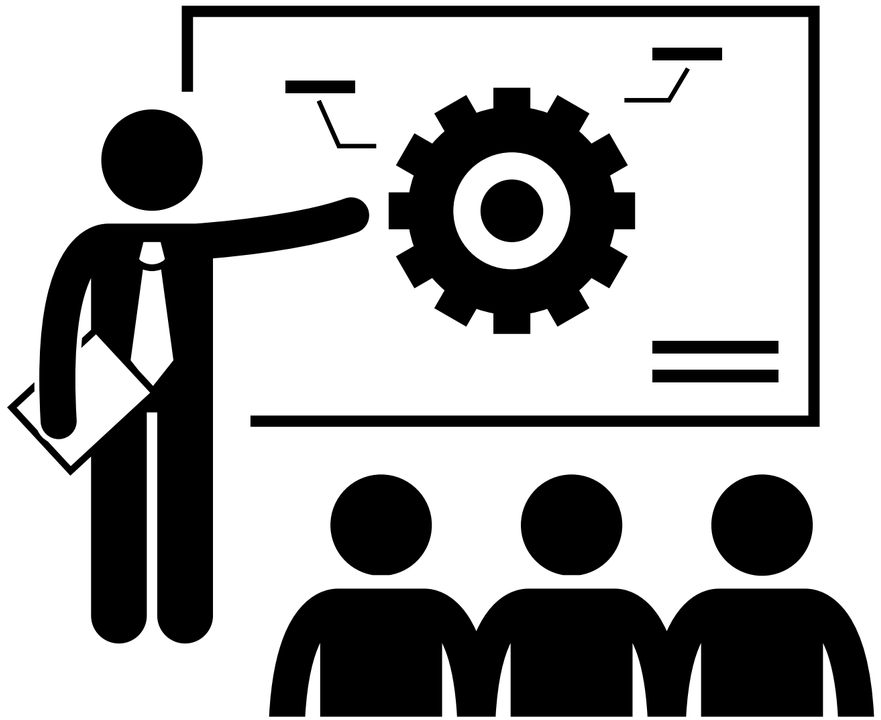


Takeoff and Landing has the highest amount of accident rates

CLASSIFICATION TREE

Comparing factors such as the Number of engine, Weather condition, Company & Month with injury severity.

CLASSIFICATION TREE



Predictor

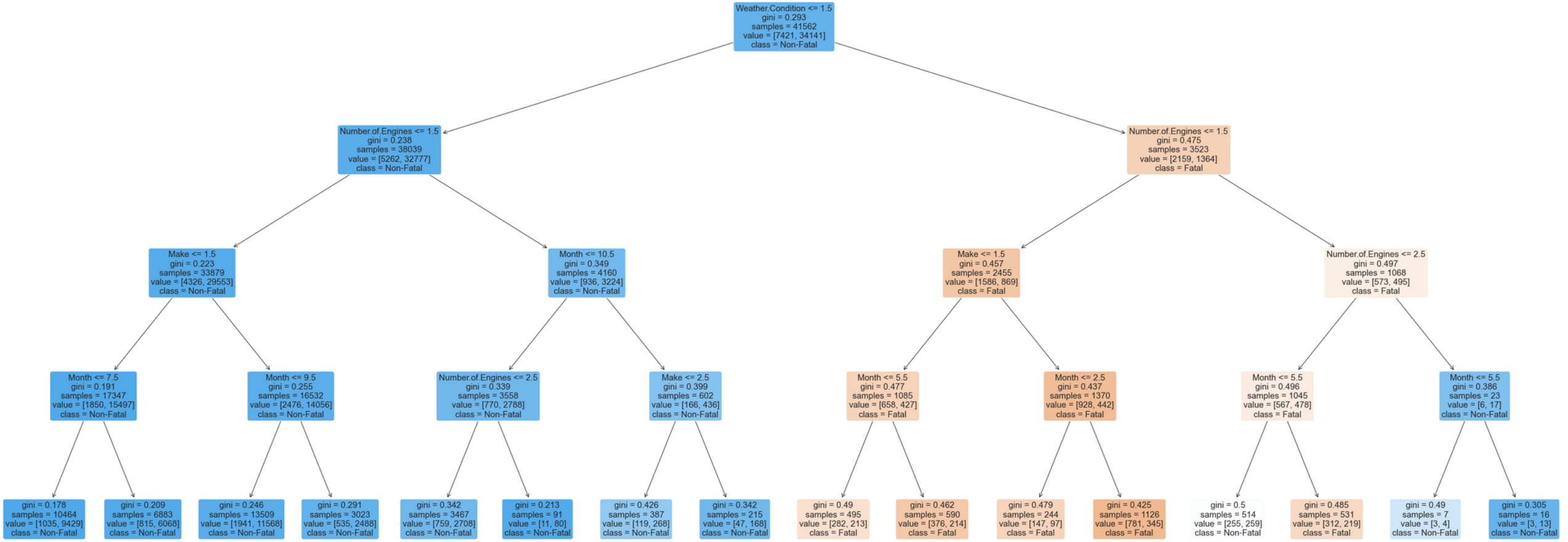
y = Injury Severity

#Response

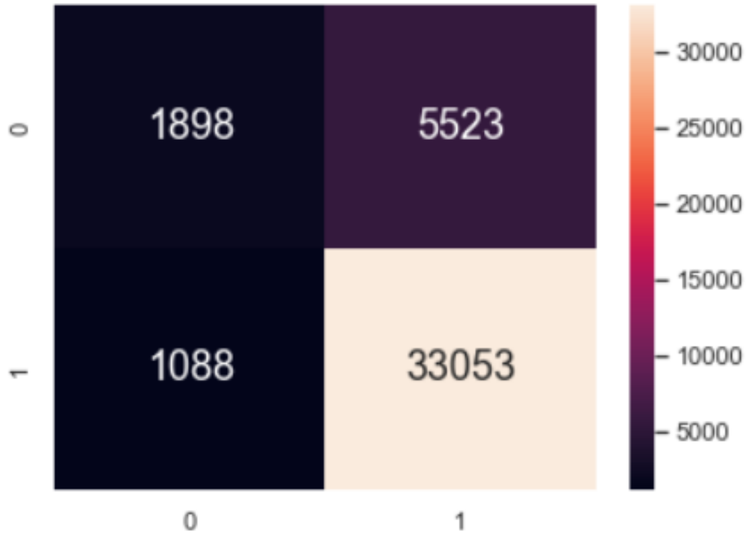
X = Number.of.Engines,Weather.Condition,Make,Month

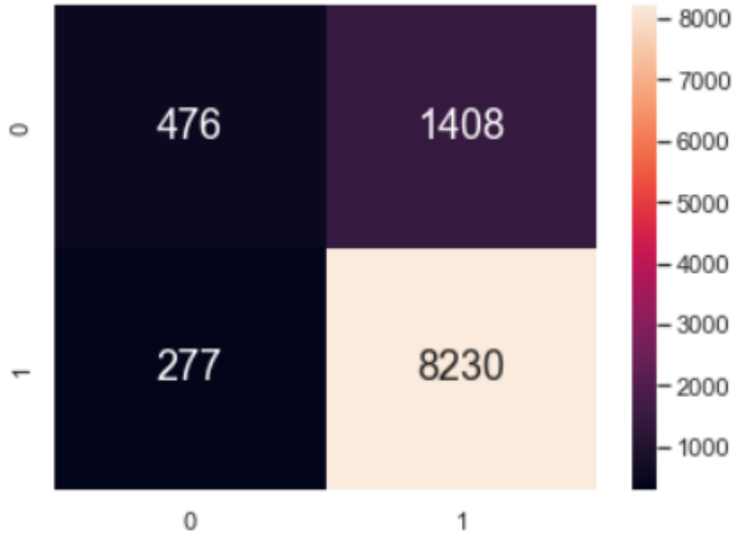
MACHINE LEARNING(CLASSIFICATION TREE)

- Number of engine, weather condition, make & month with injury severity



IV MACHINE LEARNING(CLASSIFICATION TREE)





TRAIN SAMPLE

- TRUE POSITIVE=33053
- TRUE NEGATIVE=1898
- FALSE POSITIVE=5523
- FALSE NEGATIVE=1088

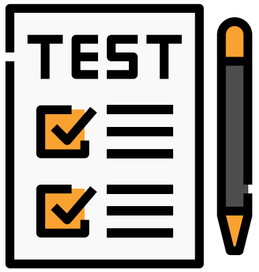
TEST SAMPLE

- TRUE POSITIVE=8230
- TRUE NEGATIVE=476
- FALSE POSITIVE=1408
- FALSE NEGATIVE=277

MACHINE LEARNING(CLASSIFICATION TREE)

TRAIN DATASET

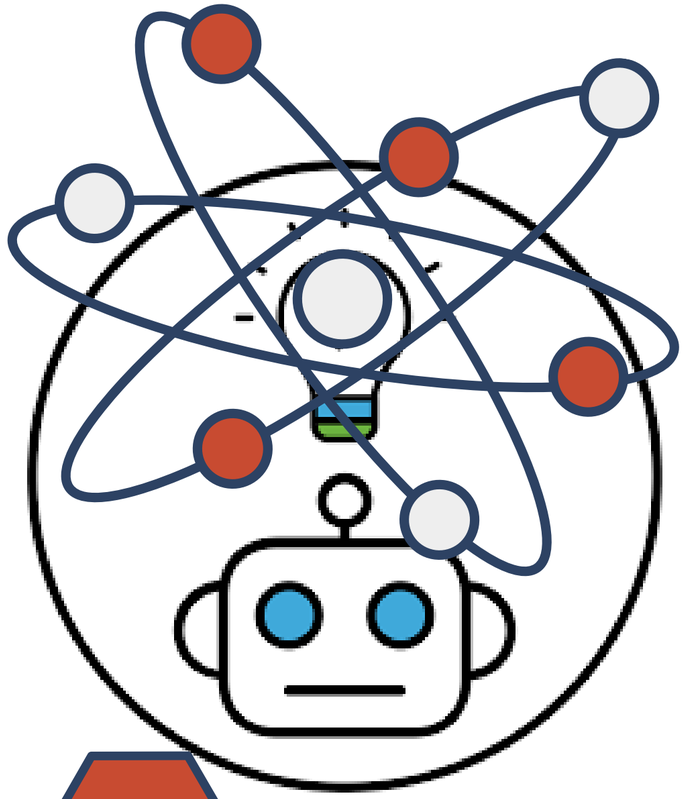
- CLASSIFICATION ACCUARACY=0.825
- TRUE POSITIVE RATE =0.968
- TRUE NEGATIVE RATE=0.255



TEST DATASET

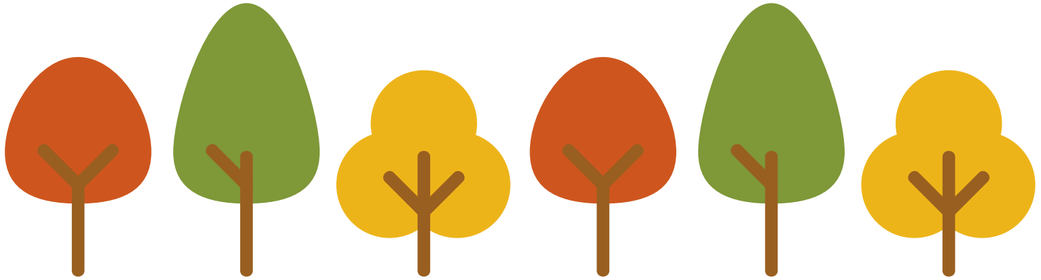
- CLASSIFICATION ACCUARACY=0.824
- TRUE POSITIVE RATE = 0.967
- TRUE NEGATIVE RATE= 0.252

IV MACHINE LEARNING



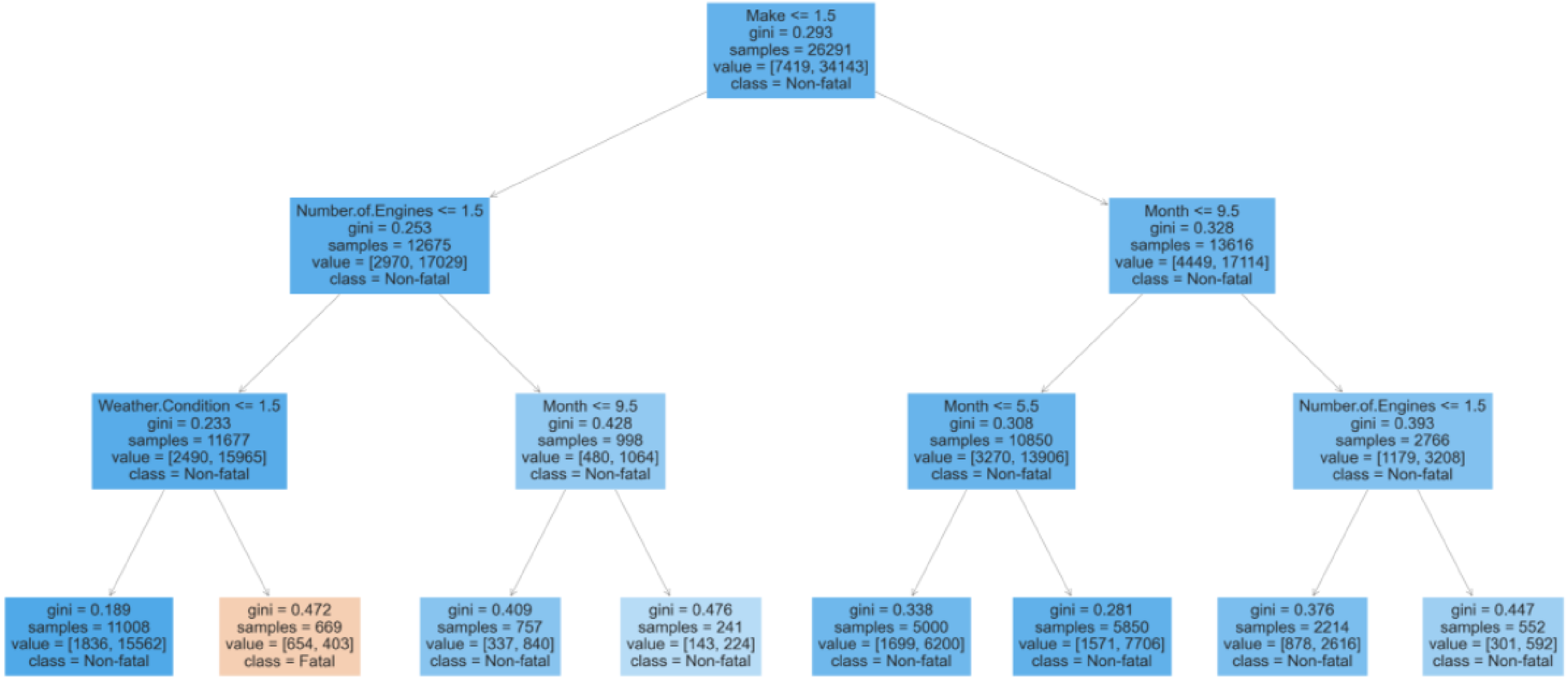
TO FURTHER OPTIMISE OUR CLASSIFICATON TREE

WE DECIDED TO USE RANDOM FOREST CLASSIFIER



MACHINE LEARNING(RANDOM FOREST CLASSIFIER)

- Number of engine, weather condition, make & month with injury severity

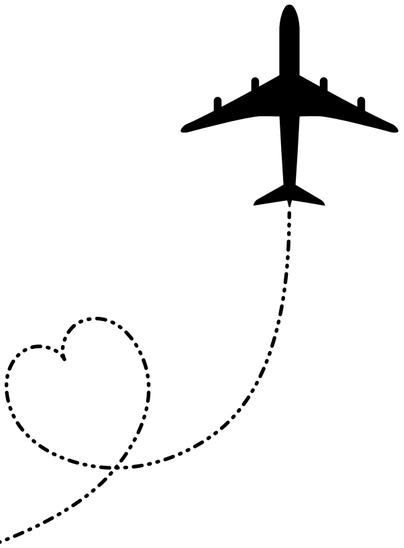


MACHINE LEARNING(CLASSIFICATION TREE)

	Varname	Imp
1	Weather.Condition	0.849349
0	Number.of.Engines	0.091880
2	Make	0.034458
3	Month	0.024313

- Random Forest Classifier resulted in **better confidence** in the decision tree as seen from the gini coefficient
- It helped us identify the variable with the **most impact** on aviation incident

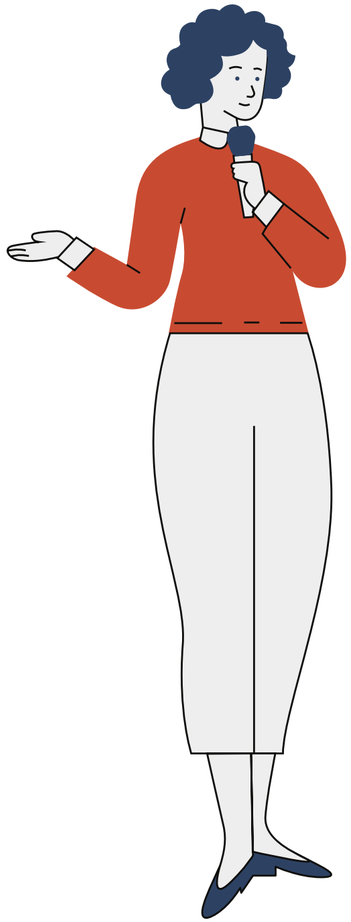
V CONCLUSION AND INSIGHTS DRIVEN



WEATHER CONDITION

- WEATHER CONDITION HAS THE GREATEST IMPACT ON AVIATION INCIDENT
- IT IS HARD TO PREDICT AND REQUIRE CONSISTENT MONITORING
- USE MACHINE LEARNING MODELS TO BETTER PREDICT WEATHER CONDITIONS
- WITH BETTER WEATHER PREDICTION AIRLINES CAN MAKE MORE INFORMED DECISIONS AND PILOTS CAN BE MORE WARY OF MORE SUITABLE FLIGHT PATHS

V CONCLUSION AND INSIGHTS DRIVEN



COUNT OF ENGINE IS A MAJOR FACTOR CONTRIBUTING TO AVIATION INCIDENTS

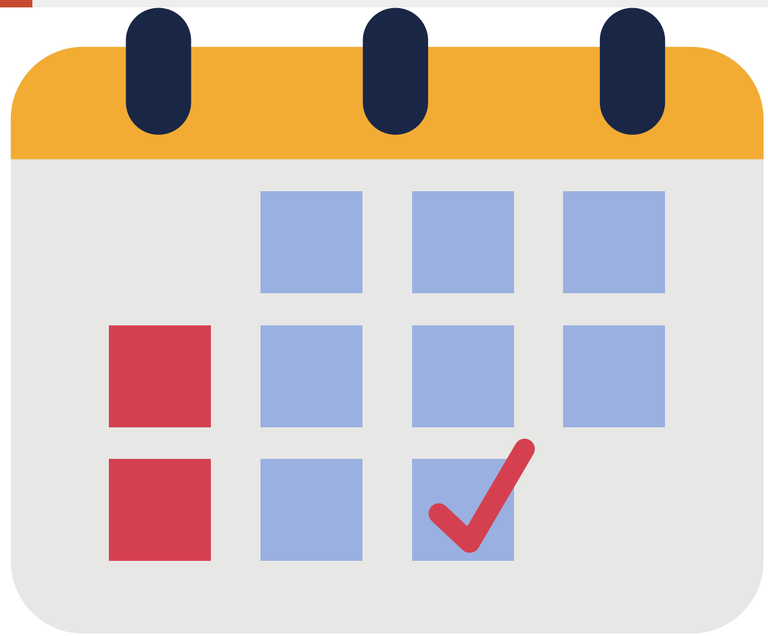
- Manufacturers of planes should consider improving the design of their planes i.e to include more engines
- Although it may be less economical, it results in better ethical practice and lower chance of aviation accidents

V CONCLUSION AND INSIGHTS DRIVEN

Company

- WHILE THE MANUFACTURER DOES NOT HAVE A SIGNIFICANT IMPACT ON AVIATION INCIDENTS BASED ON OUR DATA ANALYSIS , THEY STILL HAVE AN IMPACT ON AVIATION INCIDENTS
- GOVERNMENT BODIES SHOULD MAKE AN EFFORT TO MAKE SURE MANUFACTURER UPHOLD A CERTAIN STANDARD FOR THE PARTS OF THE PLANE THEY PRODUCE

CONCLUSION AND INSIGHTS DRIVEN



MONTH

- WE REALISE THE VARIABLE MONTH HAD THE LEAST IMPACT ON AVIATION INCIDENTS AS COMPARED TO THE OTHER VARIABLES
- MOST LIKELY BECAUSE THE CERTAIN MONTH HAS MORE FLIGHTS BUT THERE IS LESS AVIATION INCIDENTS

Thank you for listening!