PROJECT REPORT
CSE5007 EXPLORATORY DATA ANALYSIS

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

EMBEDDED PROJECT EXPLORATORY DATA ANALYSIS ON AIRLINE FATAILITIES USING PYTHON

Under Guidance of
Dr. Rushi Kumar

Team Members: -
Sneha Yadav [21MDT0028]
Divye Gupta [21MDT0063]
Abhay Singh B Patil [21MDT0075]

Abstract:

A lot of air incidents are happening in these days which become inspiration of our project.
In this project we find various relations and patterns behind the airline incidents.
From our work the airlines can benefit by taking more precautions against the factors due to which more number of incidents are occurred.

Introduction:

Airline Incident is an occurrence associated with the operation of an aircraft that affects or could affect the safety of operation.
Such incidents are increasing day by day due to n number of reasons, so we try to analyse few of them so as to help in bringing reduction of such incidents.
Python is a very robust language and contains wide variety of modules and packages which makes analysis much more easy process. So, we use python Jupyter Lab to carry out our analysis.
Specifically we use numpy, pandas, seaborn and matplotlib packages of python.

Literature Review:

Paper I:

Authors: M. Shahriari and M.E. Aydin

Article title: Aviation Accident Analysis: A Case Study

Journal: Advances in intelligent system and computing

Description

This study is focused on an airplane crash case to analyze and identify the accident contributing factors. The accident occurred on 27th of December1991 in a few minutes after a Scandinavian Airlines System plane departed from Stockholm on a route to Copenhagen, Denmark. It was found that the cause of this accident is a combination of several factors. Errors can result from ambiguously written

procedures, inadequate training, unexpected operational situation or individual judgments. Situational awareness, environmental and crew coordination factors, as well as shortcomings in pilot technical knowledge, skills and experience, also can cause incidents. Other mistakes might be the result of improper airspace design or crew coordination .

Paper II:

Authors: Ved Prakash Gupta, M Sajid Mansoori, Jitendra Shreemali, Payal Paliwal

Article title: Predicting Causes of Airplane Crashes using Machine Learning Algorithms

Journal: International Journal of Recent Technology and Engineering (IJRTE)

Description

Considering the immense cost of air crashes, the study examines the causes of crashes of aircrafts based on reported findings for the crash. The dataset used for this study included data for all reported air crashes across the globe for the period from 1981 to 2019. The causes were classified into seven categories. Multiple machine learning algorithms were used to identify the best for predicting the likely cause of accident based on features available. The Machine Learning Models used are Auto Classifier, Tree-AS and XGBoost. Also the key predictors are identified for use by planners.

Paper III:

Authors: Nikita Pande, Devyani Gupta, Jitendra Shreemali , Prasun Chakrabarti

Article title: Predicting Fatalities in Air Accidents using CHAID XGBoost Generalized Linear Model Neural Network and Ensemble Models of Machine Learning

Journal: International Journal of Engineering and Advanced Technology (IJEAT)

Description

The study examines the historical data of about 4700 air crashes all over the world since the first recorded air crash

of 1908. Given the immense impact on human beings as well as companies, the study aimed at utilizing Machine Learning principles for predicting fatalities. The train-test partition used was 75-25. Employing the IBM SPSS Modeler, the machine learning models used included CHAID model, Neural Network, Generalized Linear Model, XGBoost, Random Trees and the Ensemble model to predict fatalities in air crashes. The best results (90.6% accuracy) were achieved through Neural Network with one hidden layer. The results presented also include comparison of the predicted versus observed results for the test data.

Paper IV:

Authors: Jamsher Bhanbhro, Faiez Yousuf, Sanam Narejo, Mursal Furqan

Article title: PIA Accidents Analysis Using Naïve Bayes Classifier

Journal: International Conference on Computational Sciences and Technologies

Description

There have been many airline accidents especially in Pakistan, the nation has lost many precious lives and had a bad impact on the economy. This research study is about finding reasons and statistical analysis of airline accidents. There is not recent study available in this area from Pakistan Region, although the research studies are done on accident analysis of the international airline. The data science algorithms and the statistical analysis techniques are used to find out the reasons and data abnormalities. Mostly the model used in this study is based upon a supervised learning algorithm, i.e., Naïve based classifier (because it is very efficient and produces better predictions when there is fewer data in the dataset and in this case, dataset contains only 22 data entries) since it is used to classify analysis. Histogram analysis is used for graphical representation. The dataset used in this research is data of all airline accidents in Pakistan since 1947. The study gives the best results and helps to find out causes and suggests how accidents can be reduced.

Problem Formulation:

We Try to Find :-

which aircraft type faced the highest number of incidents and fatalities, like Domestic Passenger, Cargo, Military etc.

The Aircraft Phase where most number of Fatalities occurred.

Number of Fatalities depending upon the damage of the Aircraft.

The aircraft operators linked with the most number of incidents and Fatalities.

The Number of Fatalities corresponding to each month.

The Number of Fatalities corresponding to each Date.

The Number of Fatalities corresponding to each Day of the Week.

Methodology:

In this project, we have taken the secondary data as a part of our research. So, we assume that this data is verified and authenticated. The Data has been taken from Kaggle.

We used Python Jupyter Notebook to carry out the analysis as it one of the best available tools to do Data Analysis.

Specifically, we used numpy, pandas, seaborn and matplotlib packages of python.

We imported the Data and carried out the pre-processing like cleaning the data, removing/treating outliers and made it ready for the analysis.

We further grouped the data on various features like aircraft types, aircraft operators and so on, and summed up all the Fatalities.

Then, we used multiple number of bar plots and line plots to find patterns and relations amongst these grouped attributes of the features.

Software Requirements:

Python: Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently whereas other languages use punctuation, and it has fewer syntactical constructions than other languages.

Jupyter: The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text.

Libraries used in projects are Pandas, NumPy, Seaborn and Matplotlib.

Pandas:

Pandas is a Python library that is used for **faster data analysis, data cleaning and data pre-processing**.

Pandas is built on top of NumPy.

So, NumPy gets some superpower with pandas.

You might have heard about data-frames, which is a common term in machine learning.

This word comes from pandas.

NumPy:-

NumPy is a Python library used for working with arrays.

It also has functions for working in domain of linear algebra, Fourier transform, and matrices.

NumPy was created in 2005 by Travis Oliphant. It is an open-source project, and you can use it freely.

NumPy stands for Numerical Python.

Shape manipulation and Fourier transformation

Logical as well as mathematical operations.

Linear algebra operations using inbuilt functions.

Seaborn:-

Seaborn is a library in Python predominantly used for making **statistical graphics**.

Seaborn is a data visualization library built on top of matplotlib and closely integrated with pandas data structures in Python.

Visualization is the central part of Seaborn which helps in exploration and understanding of data.

Seaborn provides many colour palettes and defaults beautiful styles to make the creation of many statistical plots in Python more attractive.

The Seaborn library is built on top of Matplotlib and offers many advanced data visualization capabilities.

Matplotlib:-

Matplotlib is an amazing visualization library in Python for 2D plots of arrays.

Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack.

It was introduced by John Hunter in the year 2002.

It allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.

Code:

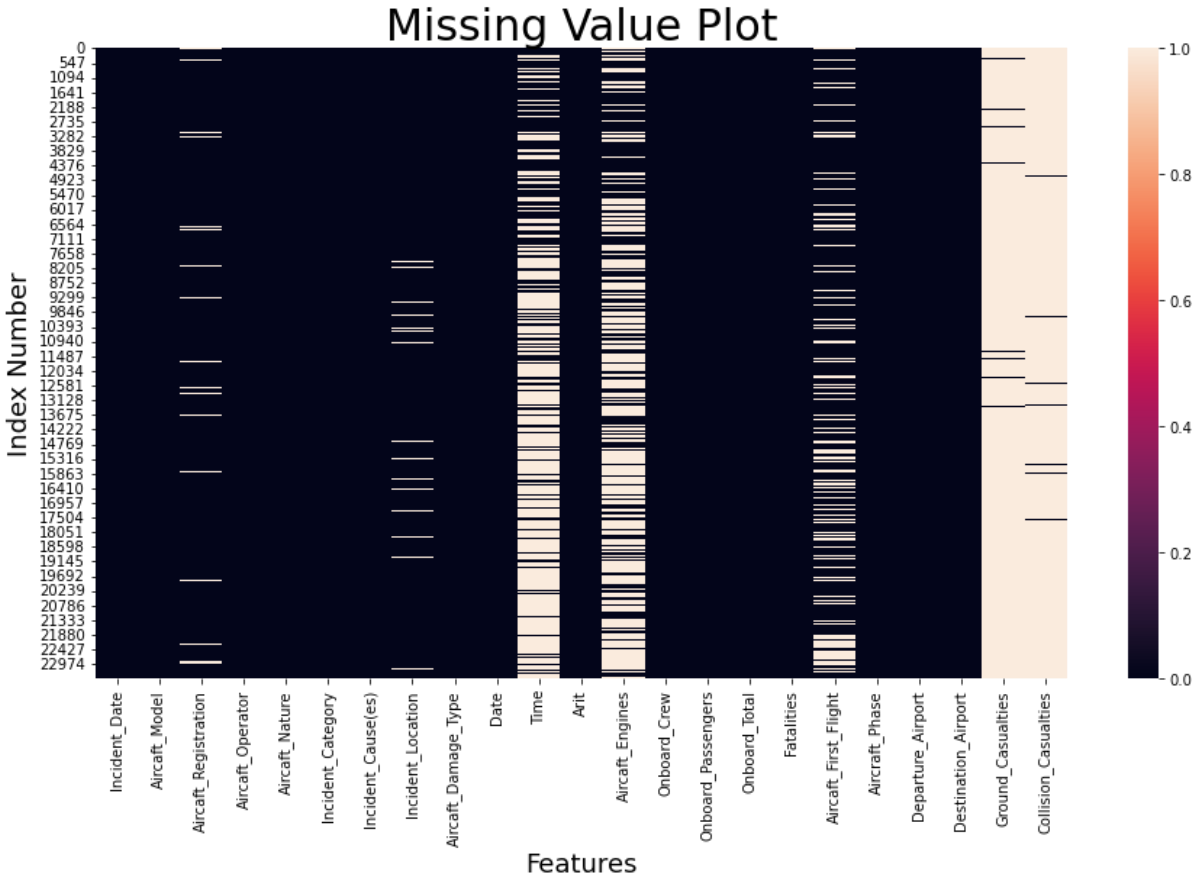
```
In [1]: import os
os.chdir("D:/CSV files")
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

data1 = pd.read_csv('Aircraft_Incident_Dataset.csv')
data1.head()
```

Out[1]:

	Incident_Date	Aircaft_Model	Aircaft_Registration	Aircaft_Operator	Aircaft_Nature	Incident_Category	Incident_Cause(es)	Incident_Location	Aircaft_Damage_
0	3-Jan-22	British Aerospace 4121 Jetstream 41	ZS-NRJ	SA Airlink	Domestic Non Scheduled Passenger	Accident | repairable-damage	Airplane - Engines, Airplane - Engines - Prop/...	near Venetia Mine...	Substa
1	4-Jan-22	British Aerospace 3101 Jetstream 31	HR-AYY	LANHSA	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Roatán-Juan ...	Substa
2	5-Jan-22	Boeing 737-4H6	EP-CAP	Caspian Airlines	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Isfahan-Shah...	Substa
3	8-Jan-22	Tupolev Tu-204-100C	RA-64032	Cainiao, opb Aviastar-TU	Cargo	Accident | hull-loss	Cargo - Fire/smoke, Result - Damaged on the gr...	Hangzhou-Xia...	Destr
4	12-Jan-22	Beechcraft 200 Super King Air	NaN	private	Illegal Flight	Criminal occurrence (sabotage, shoot down) | h...	Result - Damaged on the ground	Machakilha, ...	Damaged be n

5 rows × 23 columns

```
plt.rcParams['figure.figsize']=[15,8]
sns.heatmap(data1.isnull())
plt.title('Missing Value Plot',fontsize=30)
plt.xlabel("Features",fontsize=18)
plt.ylabel('Index Number',fontsize=18)
plt.show()
```




```
# trying to simplify the Incident_Date Column
Incident_date,Incident_month,Incident_year=[],[],[]
for i in data2['Incident_Date']:
    j=i.split('-')
    Incident_date.append(j[0])
    Incident_month.append(j[1].lower())
    Incident_year.append(j[2])
data2['Incident_date'] = Incident_date
data2['Incident_month'] = Incident_month
data2['Incident_year'] = Incident_year
data2.head()
```

	Incident_Date	Aircaft_Model	Aircaft_Registration	Aircaft_Operator	Aircaft_Nature	Incident_Category	Incident_Cause(es)	Incident_Location	Aircaft_Damage_T
0	3-Jan-22	British Aerospace 4121 Jetstream 41	ZS-NRJ	SA Airlink	Domestic Non Scheduled Passenger	Accident | repairable-damage	Airplane - Engines, Airplane - Engines - Prop/...	near Venetia Mine...	Substa
1	4-Jan-22	British Aerospace 3101 Jetstream 31	HR-AYY	LANHSA	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Roatán-Juan ...	Substa
2	5-Jan-22	Boeing 737-4H6	EP-CAP	Caspian Airlines	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Isfahan-Shah...	Substa
3	8-Jan-22	Tupolev Tu-204-100C	RA-64032	Cainiao, opb Aviastar-TU	Cargo	Accident | hull-loss	Cargo - Fire/smoke, Result - Damaged on the gr...	Hangzhou-Xia...	Destr
4	12-Jan-22	Beechcraft 200 Super King Air	NaN	private	Illegal Flight	Criminal occurrence (sabotage, shoot down) | h...	Result - Damaged on the ground	Machakilha, ...	Damaged be r

5 rows × 26 columns

```
# trying to simplify the Date column

Date_day=[]
for i in data2['Date']:
    j = i.split(' ')
    Date_day.append(j[0])
print(np.unique(Date_day))
data2['Day'] = Date_day
data2.head()
```

```
['Friday' 'Monday' 'Saturday' 'Sunday' 'Thursday' 'Tuesday' 'Wednesday'
 'XX' 'xx' 'xx.xx.1945']
```

	Incident_Date	Aircaft_Model	Aircaft_Registration	Aircaft_Operator	Aircaft_Nature	Incident_Category	Incident_Cause(es)	Incident_Location	Aircaft_Damage_
0	3-Jan-22	British Aerospace 4121 Jetstream 41	ZS-NRJ	SA Airlink	Domestic Non Scheduled Passenger	Accident | repairable-damage	Airplane - Engines, Airplane - Engines - Prop/...	near Venetia Mine...	Substa
1	4-Jan-22	British Aerospace 3101 Jetstream 31	HR-AYY	LANHSA	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Roatán-Juan ...	Substa
2	5-Jan-22	Boeing 737-4H6	EP-CAP	Caspian Airlines	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Isfahan-Shah...	Substa
3	8-Jan-22	Tupolev Tu-204-100C	RA-64032	Cainiao, opb Aviastar-TU	Cargo	Accident | hull-loss	Cargo - Fire/smoke, Result - Damaged on the gr...	Hangzhou-Xia...	Destr
4	12-Jan-22	Beechcraft 200 Super King Air	NaN	private	Illegal Flight	Criminal occurrence (sabotage, shoot down) | h...	Result - Damaged on the ground	Machakilha, ...	Damaged be r

5 rows × 27 columns

```
print(np.unique(data2['Incident_month']))
data2['Incident_month'].value_counts()
data2[data2['Incident_month']=='14']
data2['Incident_month'].replace('14','feb',inplace=True)
data2[data2['Incident_month']=='20']
data2['Incident_month'].replace('20','aug',inplace=True)
data2['Incident_month'].value_counts()

data2=data2.drop(data2[data2['Incident_month']=='???'].index,axis=0)
```

```
['14' '20' '???' 'apr' 'aug' 'dec' 'feb' 'jan' 'jul' 'jun' 'mar' 'may'
 'nov' 'oct' 'sep']
```

```
data2['Incident_date'].value_counts()
data2=data2.drop(data2[data2['Incident_date']=='??'].index,axis=0)

year=np.unique(data2['Incident_year'])
year

data2['Day'].value_counts()
data2.drop(axis=1,columns=['Incident_Date','Date'],inplace=True)
data2.head()
```

	Aircaft_Model	Aircaft_Registration	Aircaft_Operator	Aircaft_Nature	Incident_Category	Incident_Cause(es)	Incident_Location	Aircaft_Damage_Type	Time	A
0	British Aerospace 4121 Jetstream 41	ZS-NRJ	SA Airlink	Domestic Non Scheduled Passenger	Accident | repairable-damage	Airplane - Engines, Airplane - Engines - Prop/...	near Venetia Mine...	Substantial	8:10	Ja
1	British Aerospace 3101 Jetstream 31	HR-AYY	LANHSA	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Roatán-Juan ...	Substantial	ca 12:00	Ja
2	Boeing 737-4H6	EP-CAP	Caspian Airlines	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Isfahan-Shah...	Substantial	17:07	Ja
3	Tupolev Tu-204-100C	RA-64032	Cainiao, opb Aviastar-TU	Cargo	Accident | hull-loss	Cargo - Fire/smoke, Result - Damaged on the gr...	Hangzhou-Xia...	Destroyed	4:40	Ja
4	Beechcraft 200 Super King Air	NaN	private	Illegal Flight	Criminal occurrence (sabotage, shoot down) | h...	Result - Damaged on the ground	Machakilha, ...	Damaged beyond repair	c. 02:00	1 Ja

5 rows × 25 columns

```
data2['Time'].value_counts()
data2[data2['Time'].isnull() == True].shape
# As we cannot determine what ca/ca./PST etc is in reality and the amount of null is very high
# so we treat it by removing this particular column
```

```
(13809, 25)
```

7]:
```
Arit_date,Arit_month,Arit_year=[],[],[]
for i in data2['Arit']:
    j=i.split('-')
    Arit_date.append(j[0])
    Arit_month.append(j[1].lower())
    Arit_year.append(j[2])
np.unique(Arit_month)
if data2['Incident_date'].tolist() == Arit_date:
    print("Same")
# As the Arit dates is also same as Incident dates, therefore dropping Arit column too
data2.drop(columns=['Time','Arit'],axis=1,inplace=True)
data2.head()
```

```
Same
```

7]:

	Aircaft_Model	Aircaft_Registration	Aircaft_Operator	Aircaft_Nature	Incident_Category	Incident_Cause(es)	Incident_Location	Aircaft_Damage_Type	Aircaft_E
0	British Aerospace 4121 Jetstream 41	ZS-NRJ	SA Airlink	Domestic Non Scheduled Passenger	Accident | repairable-damage	Airplane - Engines, Airplane - Engines - Prop/...	near Venetia Mine...	Substantial	2 TPE331-
1	British Aerospace 3101 Jetstream 31	HR-AYY	LANHSA	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Roatán-Juan ...	Substantial	
2	Boeing 737-4H6	EP-CAP	Caspian Airlines	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Isfahan-Shah...	Substantial	2 CFMI C
3	Tupolev Tu-204-100C	RA-64032	Cainiao, opb Aviastar-TU	Cargo	Accident | hull-loss	Cargo - Fire/smoke, Result - Damaged on the gr...	Hangzhou-Xia...	Destroyed	2 Solov
4	Beechcraft 200 Super King Air	NaN	private	Illegal Flight	Criminal occurrence (sabotage, shoot down) | h...	Result - Damaged on the ground	Machakilha, ...	Damaged beyond repair	

]:
```
def fat_occup(data):
    l1,l2=[],[]
    for i in data:
        j=i.split('/')
        k=j[0].split(':')
        l1.append(k[1])
        f = j[1].split(':')
        l2.append(f[1])
    return l1,l2
data2['Crew_fatal'],data2['Crew_Occup'] = fat_occup(data2['Onboard_Crew'])
data2['Pass_fatal'],data2['Pass_Occup'] = fat_occup(data2['Onboard_Passengers'])

data2['Crew_fatal'].replace('',0,inplace=True)
data2['Crew_Occup'].replace('',0,inplace=True)
data2['Pass_fatal'].replace('','0',inplace=True)
data2['Pass_Occup'].replace('','0',inplace=True)

data2.head()
```

]:

	Aircaft_Model	Aircaft_Registration	Aircaft_Operator	Aircaft_Nature	Incident_Category	Incident_Cause(es)	Incident_Location	Aircaft_Damage_Type	Aircaft_E
0	British Aerospace 4121 Jetstream 41	ZS-NRJ	SA Airlink	Domestic Non Scheduled Passenger	Accident | repairable-damage	Airplane - Engines, Airplane - Engines - Prop/...	near Venetia Mine...	Substantial	2 TPE331-
1	British Aerospace 3101 Jetstream 31	HR-AYY	LANHSA	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Roatán-Juan ...	Substantial	
2	Boeing 737-4H6	EP-CAP	Caspian Airlines	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Isfahan-Shah...	Substantial	2 CFMI C
3	Tupolev Tu-204-100C	RA-64032	Cainiao, opb Aviastar-TU	Cargo	Accident | hull-loss	Cargo - Fire/smoke, Result - Damaged on the gr...	Hangzhou-Xia...	Destroyed	2 Solov
	Beechcraft				Criminal				

```
data2['Ground_Casualties'].fillna('0',inplace=True)
data2['Collision_Casualties'].fillna('0',inplace=True)
data2['Ground_Casualties'].head()
data2['Ground_Casualties'].value_counts()
data2.info()
```

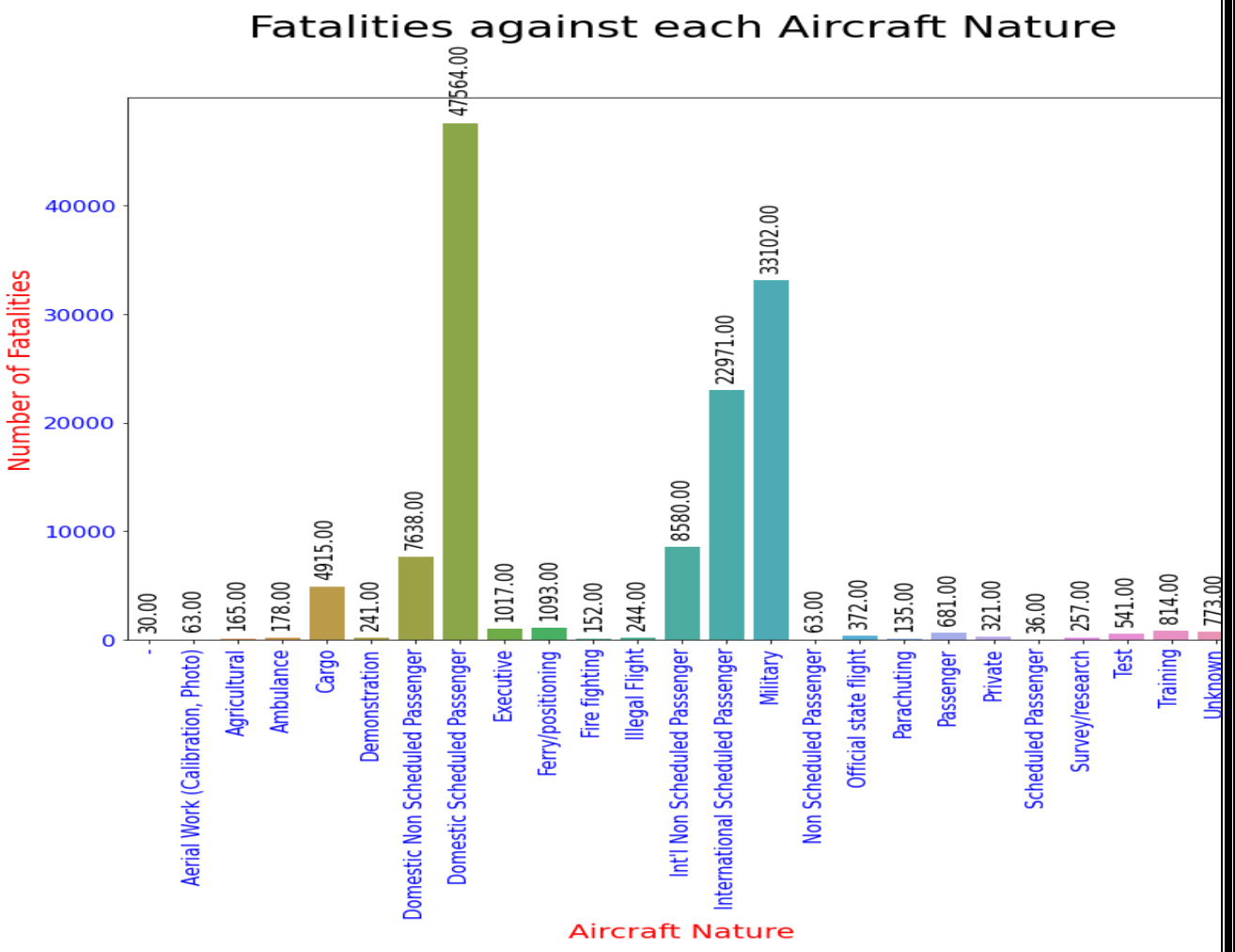
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 23012 entries, 0 to 23518
Data columns (total 27 columns):
 #   Column                Non-Null Count  Dtype 
---  ------                --------------  ----- 
 0   Aircaft_Model         23012 non-null  object
 1   Aircaft_Registration  21499 non-null  object
 2   Aircaft_Operator      23012 non-null  object
 3   Aircaft_Nature        23012 non-null  object
 4   Incident_Category     23012 non-null  object
 5   Incident_Cause(es)    23012 non-null  object
 6   Incident_Location     22253 non-null  object
 7   Aircaft_Damage_Type   23012 non-null  object
 8   Aircaft_Engines       10528 non-null  object
 9   Onboard_Crew          23012 non-null  object
 10  Onboard_Passengers    23012 non-null  object
 11  Onboard_Total         23012 non-null  object
 12  Fatalities            23012 non-null  int64 
 13  Aircaft_First_Flight  17705 non-null  object
 14  Aircraft_Phase        23012 non-null  object
 15  Departure_Airport     23012 non-null  object
 16  Destination_Airport   23012 non-null  object
 17  Ground_Casualties     23012 non-null  object
 18  Collision_Casualties  23012 non-null  object
 19  Incident_date         23012 non-null  object
 20  Incident_month        23012 non-null  object
 21  Incident_year         23012 non-null  object
 22  Day                   23012 non-null  object
 23  Crew_fatal            23012 non-null  object
```

```
#data2['Ground_Casualties']
l1,l2=[],[]
allowed_chars = set('1234567890')
for i in data2['Ground_Casualties']:
    if set(i).issubset(allowed_chars)==False:
        j=i.split(' ')
        l1.append(j[1])
    else:
        l1.append(i)
data2['Ground_Casualties']=l1
for i in data2['Collision_Casualties']:
    if set(i).issubset(allowed_chars)==False:
        j=i.split(' ')
        l2.append(j[1])
    else:
        l2.append(i)
data2['Collision_Casualties']=l2
data2.head()
```

	Aircaft_Model	Aircaft_Registration	Aircaft_Operator	Aircaft_Nature	Incident_Category	Incident_Cause(es)	Incident_Location	Aircaft_Damage_Type	Aircaft_E
0	British Aerospace 4121 Jetstream 41	ZS-NRJ	SA Airlink	Domestic Non Scheduled Passenger	Accident | repairable-damage	Airplane - Engines, Airplane - Engines - Prop/...	near Venetia Mine...	Substantial	2 TPE331
1	British Aerospace 3101 Jetstream 31	HR-AYY	LANHSA	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Roatán-Juan ...	Substantial	
2	Boeing 737-4H6	EP-CAP	Caspian Airlines	Domestic Scheduled Passenger	Accident | repairable-damage	Airplane - Undercarriage, Airplane - Undercarr...	Isfahan-Shah...	Substantial	2 CFMI C
3	Tupolev Tu-204-100C	RA-64032	Cainiao, opb Aviastar-TU	Cargo	Accident | hull-loss	Cargo - Fire/smoke, Result - Damaged on the gr...	Hangzhou-Xia...	Destroyed	2 Solov

Criminal

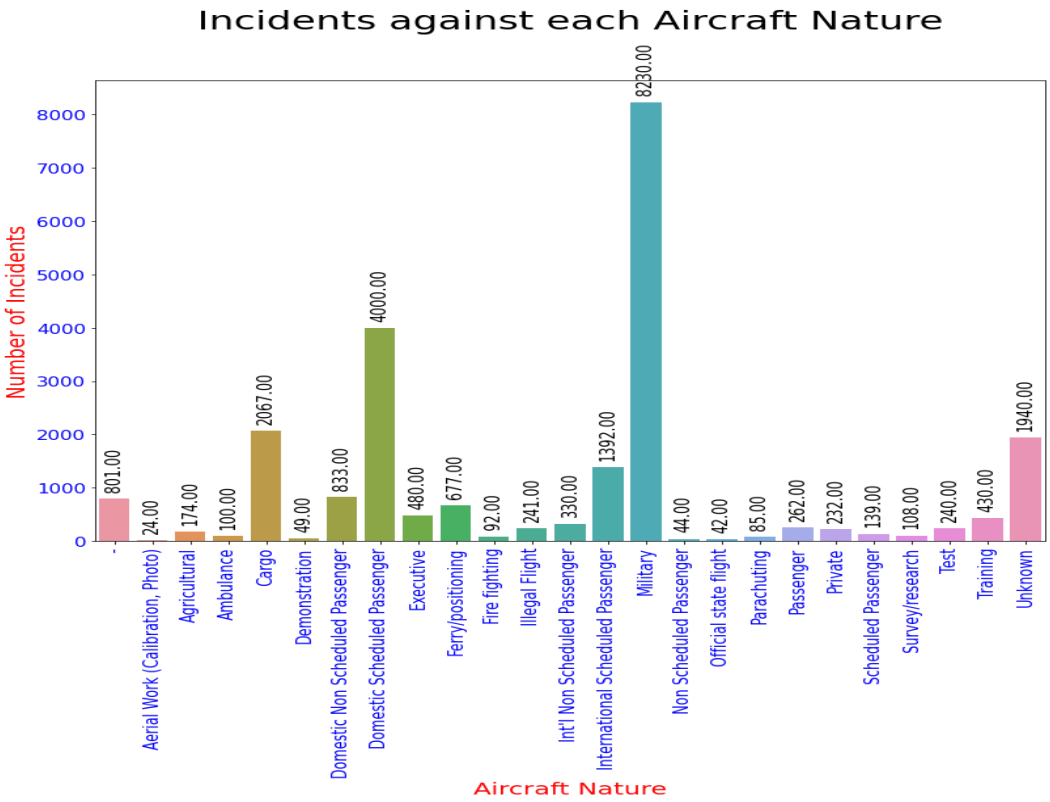
In [67]:
```
plt.rcParams['figure.figsize'] = [13,10]
plots=sns.barplot(x=l1,y=l2)
for bar in plots.patches:
    plots.annotate(format(bar.get_height(), '.2f'),
                   (bar.get_x() + bar.get_width() / 2,
                    bar.get_height()), ha='center', va='bottom',
                   size=15, xytext=(0, 8),
                   textcoords='offset points',rotation=90)
plt.xticks(rotation=90,fontsize=15,color='blue')
plt.yticks(fontsize=15,color='blue')
plt.xlabel('Aircraft Nature',fontsize=20,color='red')
plt.ylabel('Number of Fatalities',fontsize=20,color='red')
plt.title('Fatalities against each Aircraft Nature',fontsize=30,y=1.1)
plt.show()
```



We are trying here to display the Number of Fatalities based on the nature of the aircraft whether it be a commercial, cargo or military aircraft.

We can ascertain from the graph above that most passengers were killed in domestic scheduled passenger aircraft followed by military aircraft.

```
In [68]: plots=sns.barplot(x=l1,y=l3)
         for bar in plots.patches:
             plots.annotate(format(bar.get_height(), '.2f'),
                            (bar.get_x() + bar.get_width() / 2,
                             bar.get_height()), ha='center', va='bottom',
                            size=15, xytext=(0, 8),
                            textcoords='offset points',rotation=90)
         plt.xticks(rotation=90,fontsize=15,color='blue')
         plt.yticks(fontsize=15,color='blue')
         plt.xlabel('Aircraft Nature',fontsize=20,color='red')
         plt.ylabel('Number of Incidents',fontsize=20,color='red')
         plt.title('Incidents against each Aircraft Nature',fontsize=30,y=1.1)
         plt.show()
```

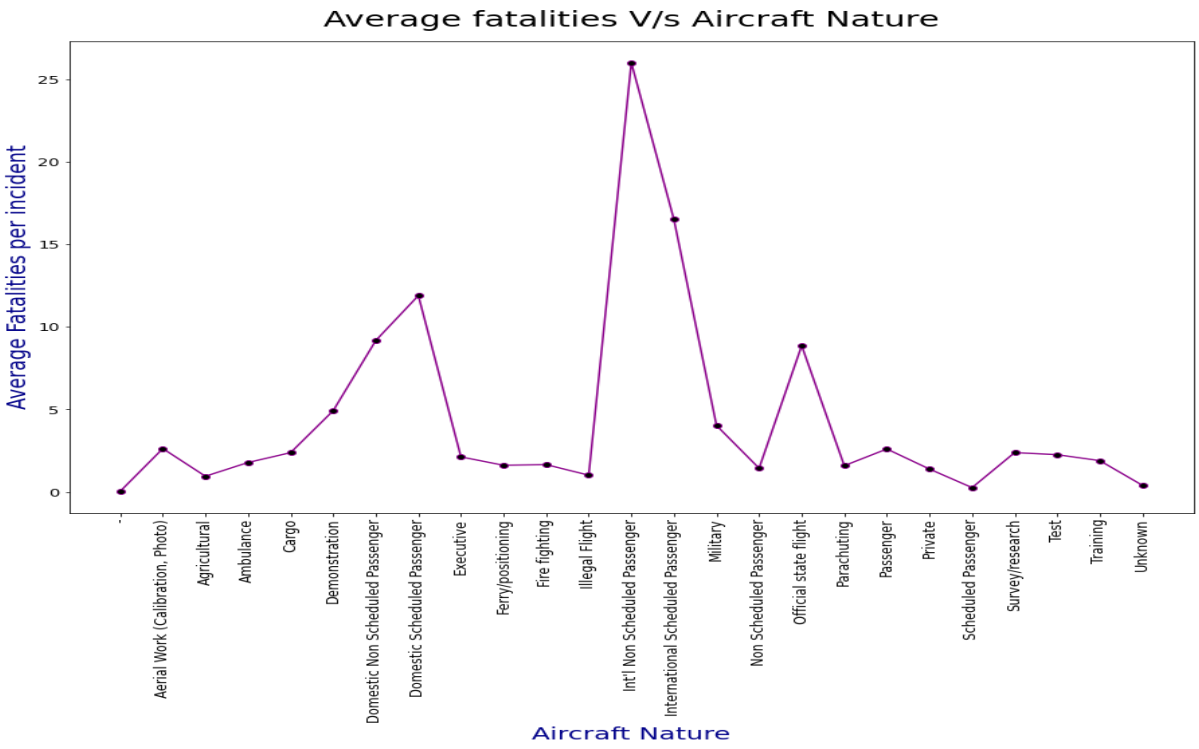


In this graph, we have attempted to display the Number of incidents based on the type of aircraft, whether it be a commercial, cargo, or military aircraft.

As we can see from the graph, most incidents occurred with military aircraft.

In light of the graph above, we can conclude that military type aircraft have fewer fatalities than domestic scheduled passenger aircraft but have more incidents.

```
l4=[]
for i in range (0,len(l1)):
    l4.append(l2[i]/l3[i])
plt.rcParams['figure.figsize'] = [15,10]
plt.plot(l1,l4,color='darkmagenta',marker='o',markerfacecolor='black')
plt.title('Average fatalities V/s Aircraft Nature',fontsize=25,color='black',y=1.02)
plt.xticks(rotation=90,fontsize=13,color='black')
plt.yticks(fontsize=13,color='black')
plt.xlabel('Aircraft Nature',fontsize=20,color='darkblue')
plt.ylabel('Average Fatalities per incident',fontsize=20,color='darkblue')
plt.show()
```



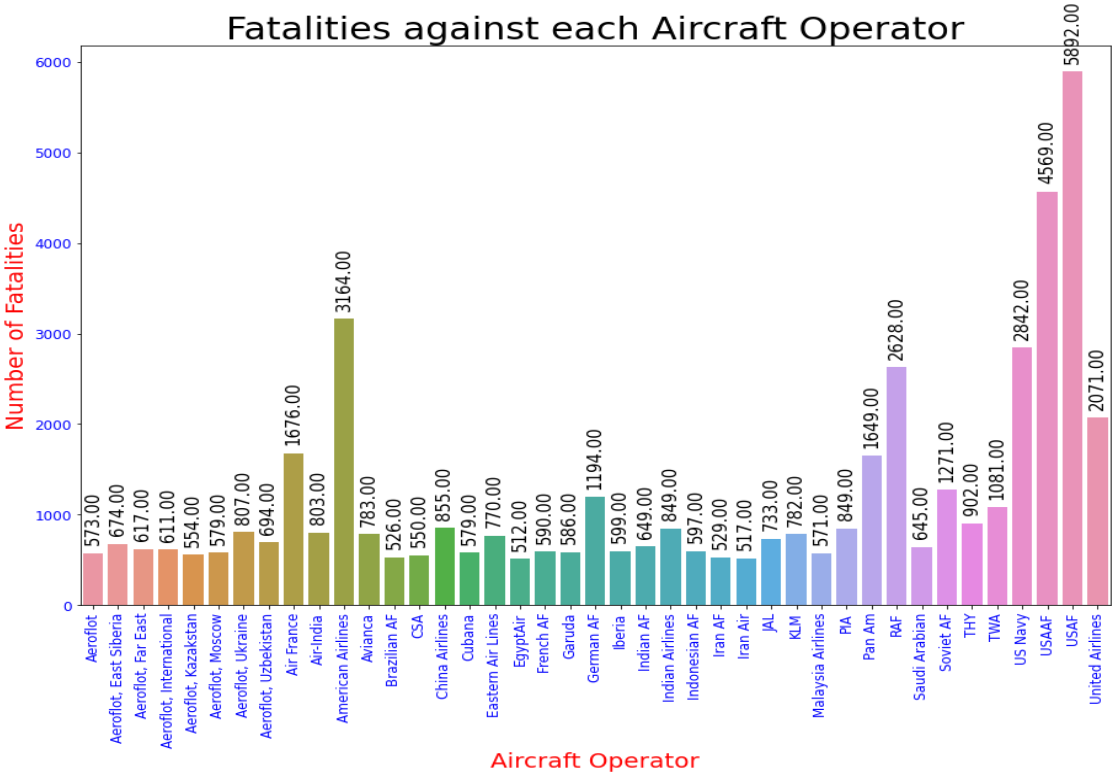
Based on the type of Aircraft Nature, we have attempted to display the average number of fatalities per incident.

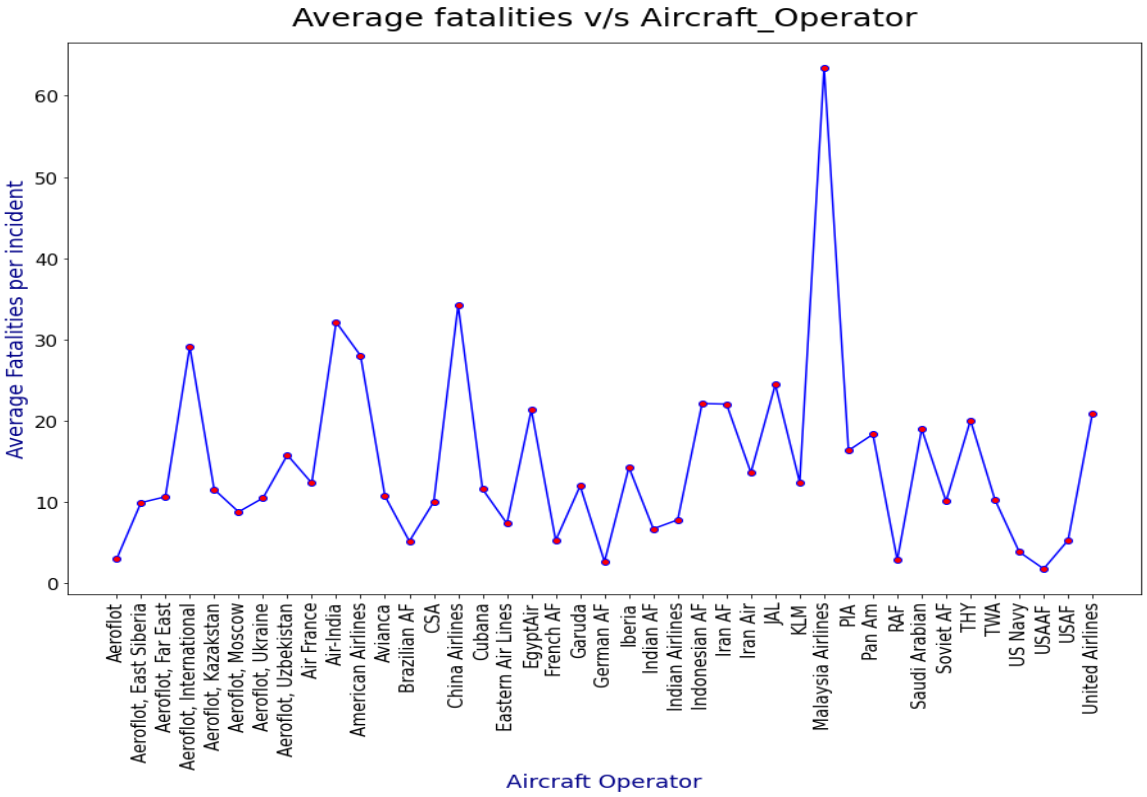
By examining this scatter plot, we can conclude that the average fatality occurs in International Non-Scheduled Passenger Aircrafts.

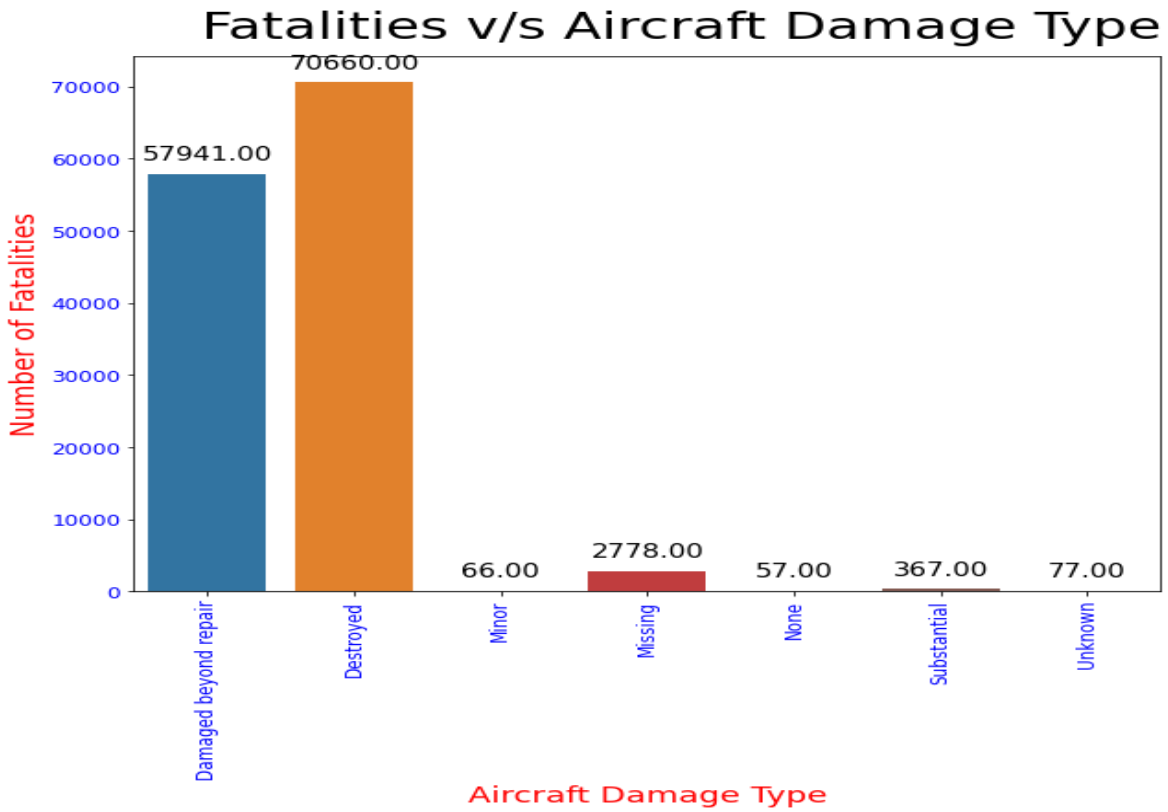
Non-scheduled passenger aircraft are probably the biggest victims of terrorist hijackings.

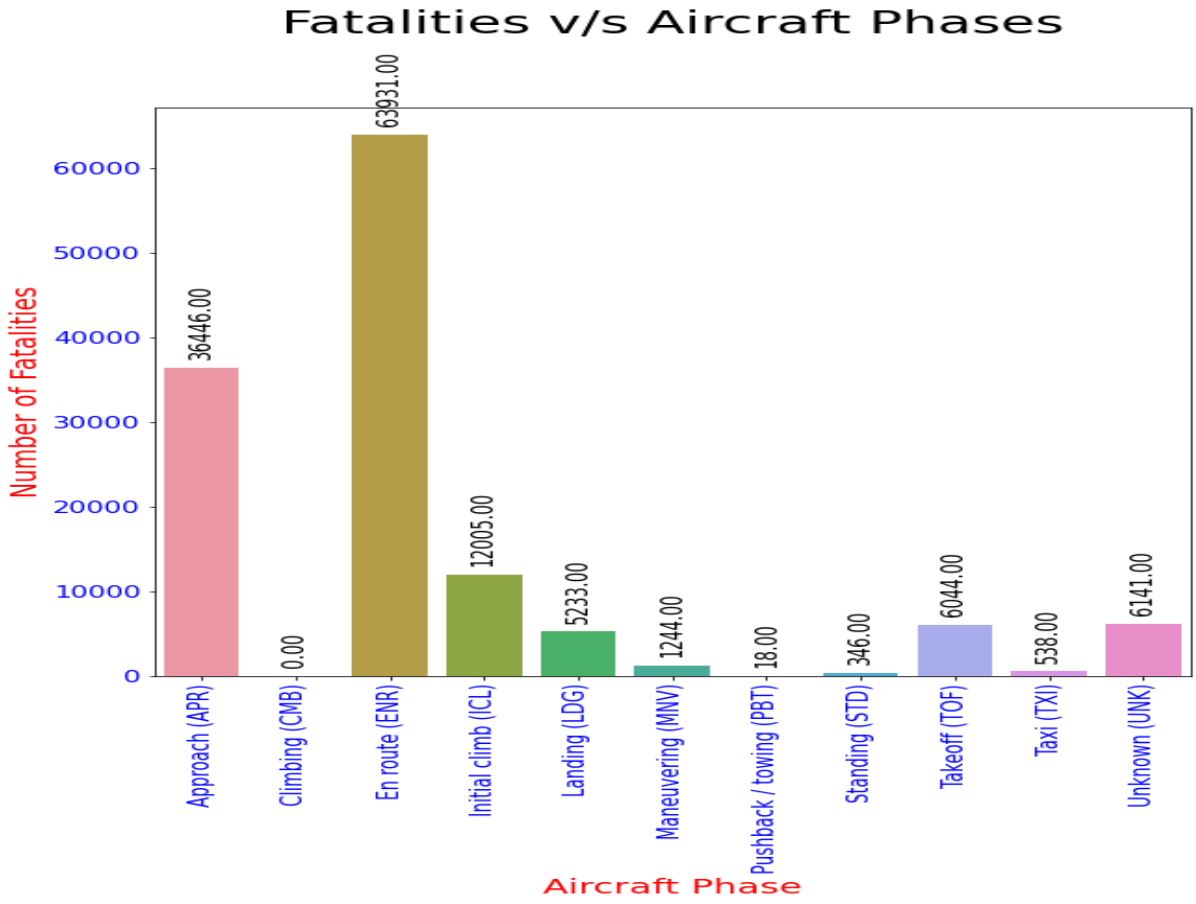
As stated, this is more focused on non-scheduled passenger airliners, which are probably a bigger cause of hijackings by terrorists.

8]:
```
data2['Aircaft_Operator'].value_counts()
a=np.unique(data2['Aircaft_Operator'])
b=0
l1=[]
l2=[]
l3=[]
for i in a:
    data3=data2[data2['Aircaft_Operator']==i]
    c=sum(data3['Fatalities'])
#     print(i,c)
    if c>500:
        l1.append(i)
        l2.append(c)
        #print(i,c)
        d=data2[data2['Aircaft_Operator']==i].shape[0]
        l3.append(d)
        b=b+1
#print(b)
```


Fatalities against each Aircraft Operator
Number of Fatalities
Aircraft Operator
Aeroflot
Aeroflot, East Siberia
Aeroflot, Far East
Aeroflot, International
Aeroflot, Kazakstan
Aeroflot, Moscow
Aeroflot, Ukraine
Aeroflot, Uzbekistan
Air France
Air-India
American Airlines
Avianca
Brazilian AF
CSA
China Airlines
Cubana
Eastern Air Lines
EgyptAir
French AF
Garuda
German AF
Iberia
Indian AF
Indian Airlines
Indonesian AF
Iran AF
Iran Air
JAL
KLM
Malaysia Airlines
PIA
Pan Am
RAF
Saudi Arabian
Soviet AF
THY
TWA
US Navy
USAAF
USAF
United Airlines
573.00
674.00
617.00
611.00
554.00
579.00
807.00
694.00
1676.00
803.00
3164.00
783.00
526.00
550.00
855.00
579.00
770.00
512.00
590.00
586.00
1194.00
599.00
649.00
849.00
597.00
529.00
517.00
733.00
782.00
571.00
849.00
1649.00
2628.00
645.00
1271.00
902.00
1081.00
2842.00
4569.00
5892.00
2071.00

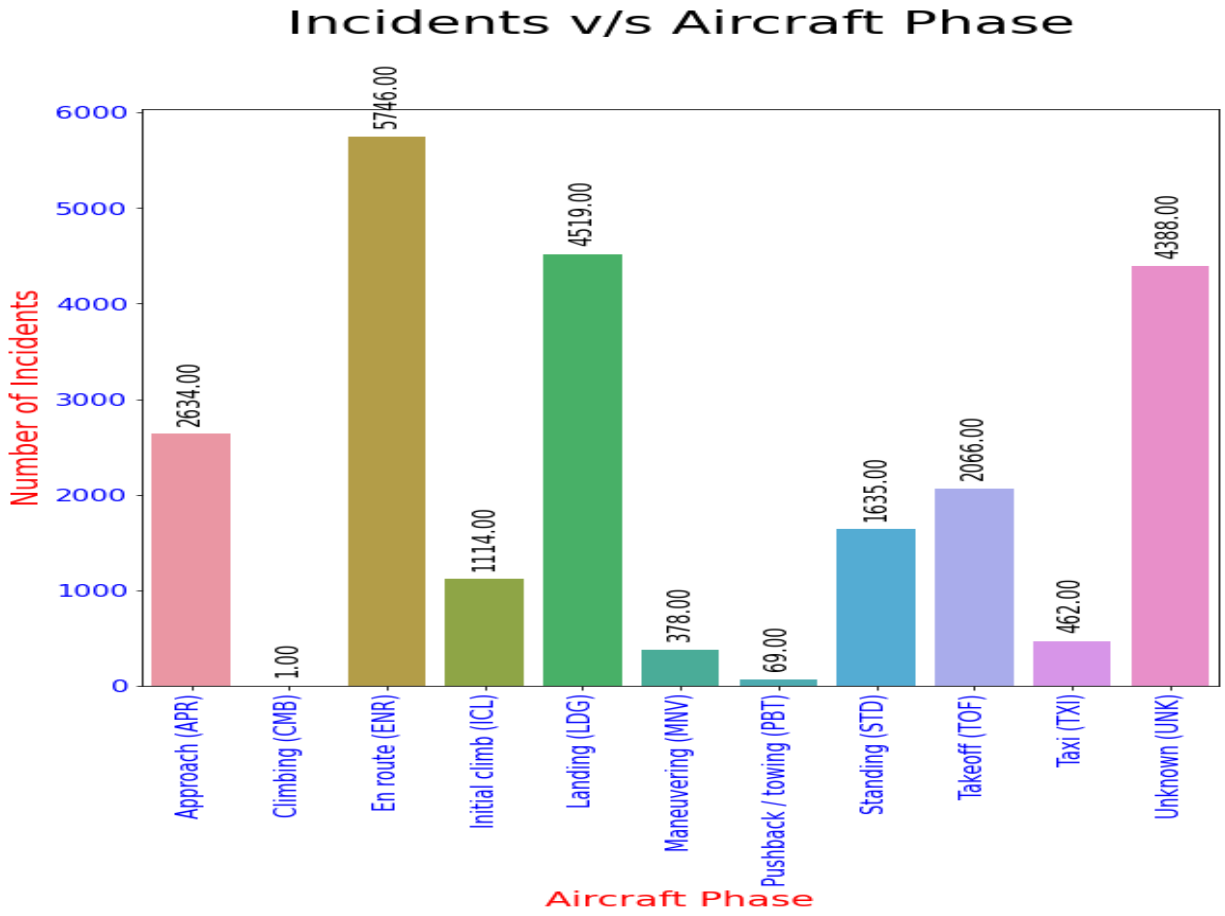

Average fatalities v/s Aircraft_Operator
Average Fatalities per incident
Aircraft Operator
Aeroflot
Aeroflot, East Siberia
Aeroflot, Far East
Aeroflot, International
Aeroflot, Kazakstan
Aeroflot, Moscow
Aeroflot, Ukraine
Aeroflot, Uzbekistan
Air France
Air-India
American Airlines
Avianca
Brazilian AF
CSA
China Airlines
Cubana
Eastern Air Lines
EgyptAir
French AF
Garuda
German AF
Iberia
Indian AF
Indian Airlines
Indonesian AF
Iran AF
Iran Air
JAL
KLM
Malaysia Airlines
PIA
Pan Am
RAF
Saudi Arabian
Soviet AF
THY
TWA
US Navy
USAAF
USAF
United Airlines





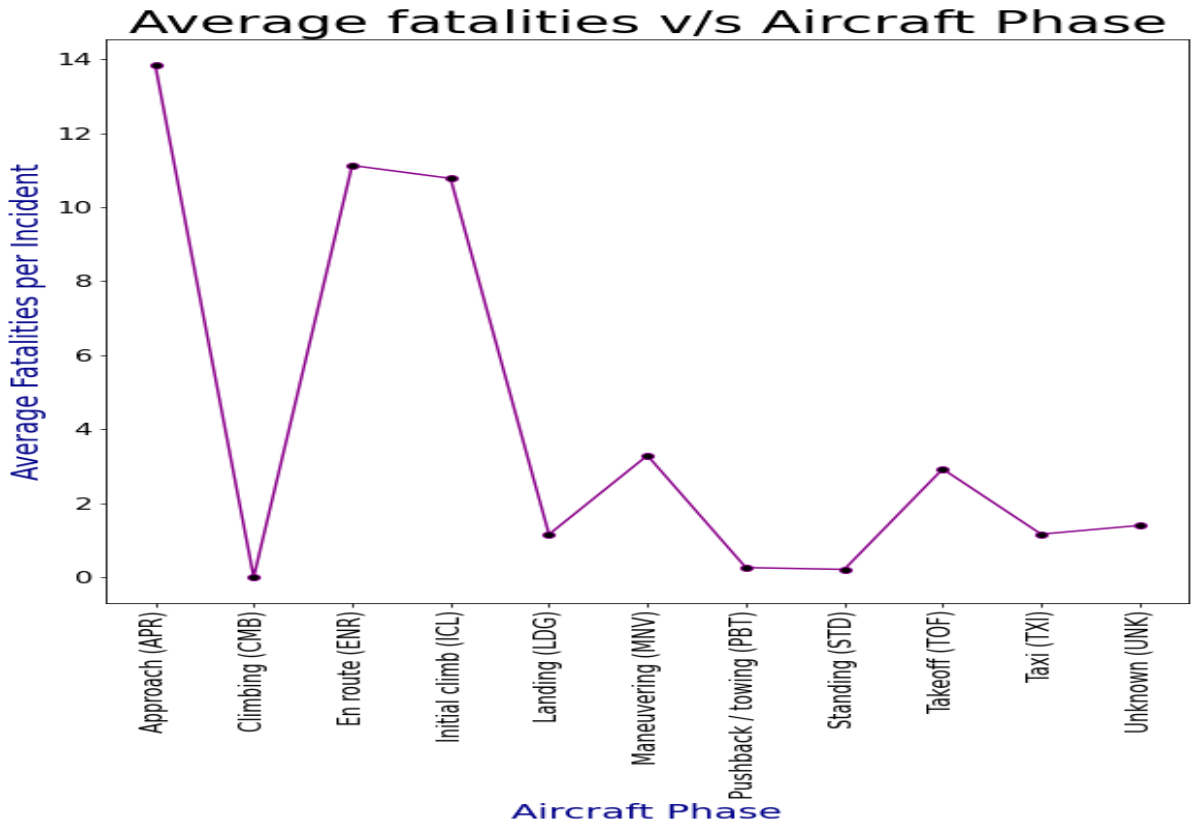
According to the type of aircraft phase, we have attempted to display the Amount of fatality.

In analysing this Bar plot, we can determine that the Number of fatality occurs during the approach phase, during the initial climb phase, and during the En route phase.



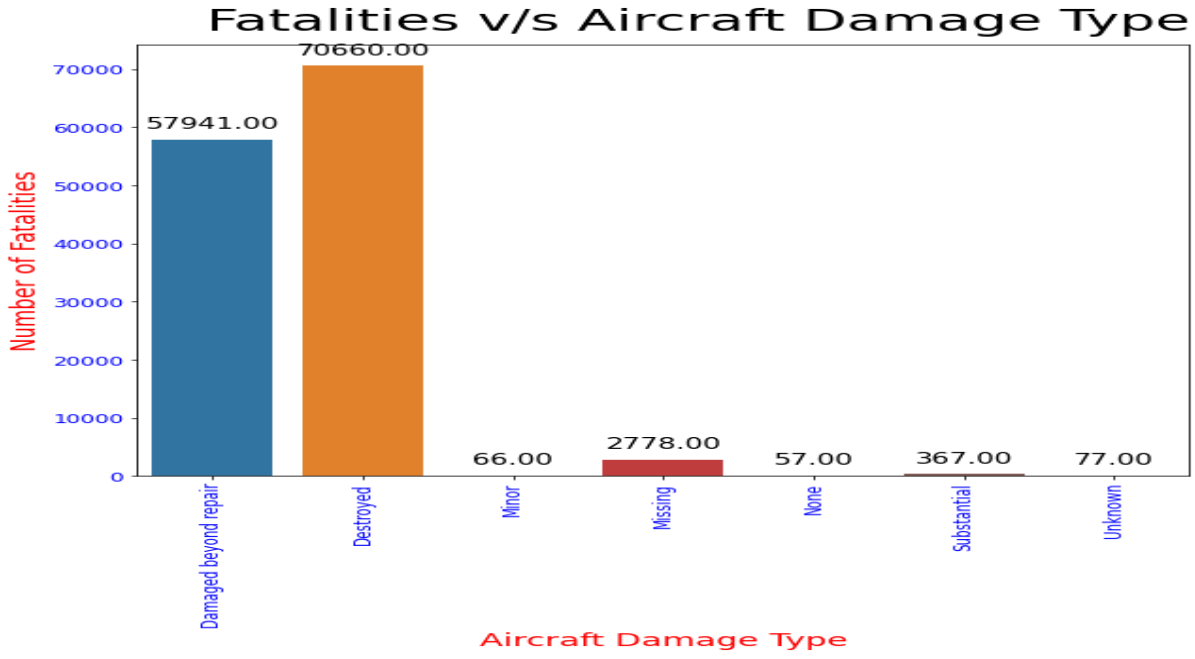
This following chart illustrates which Aircraft Phase has the highest incidence of aircraft incidents.

Here we can visualize that Number of incidents while the Aircraft is in En Route phase is pretty high followed by Landing Phase and Unknown phase.



According to the type of aircraft phase, we have attempted to display the average fatality per incident.

In analyzing this scatter plot, we can determine that the average fatality occurs during the approach phase, during the initial climb phase, and during the En route phase.

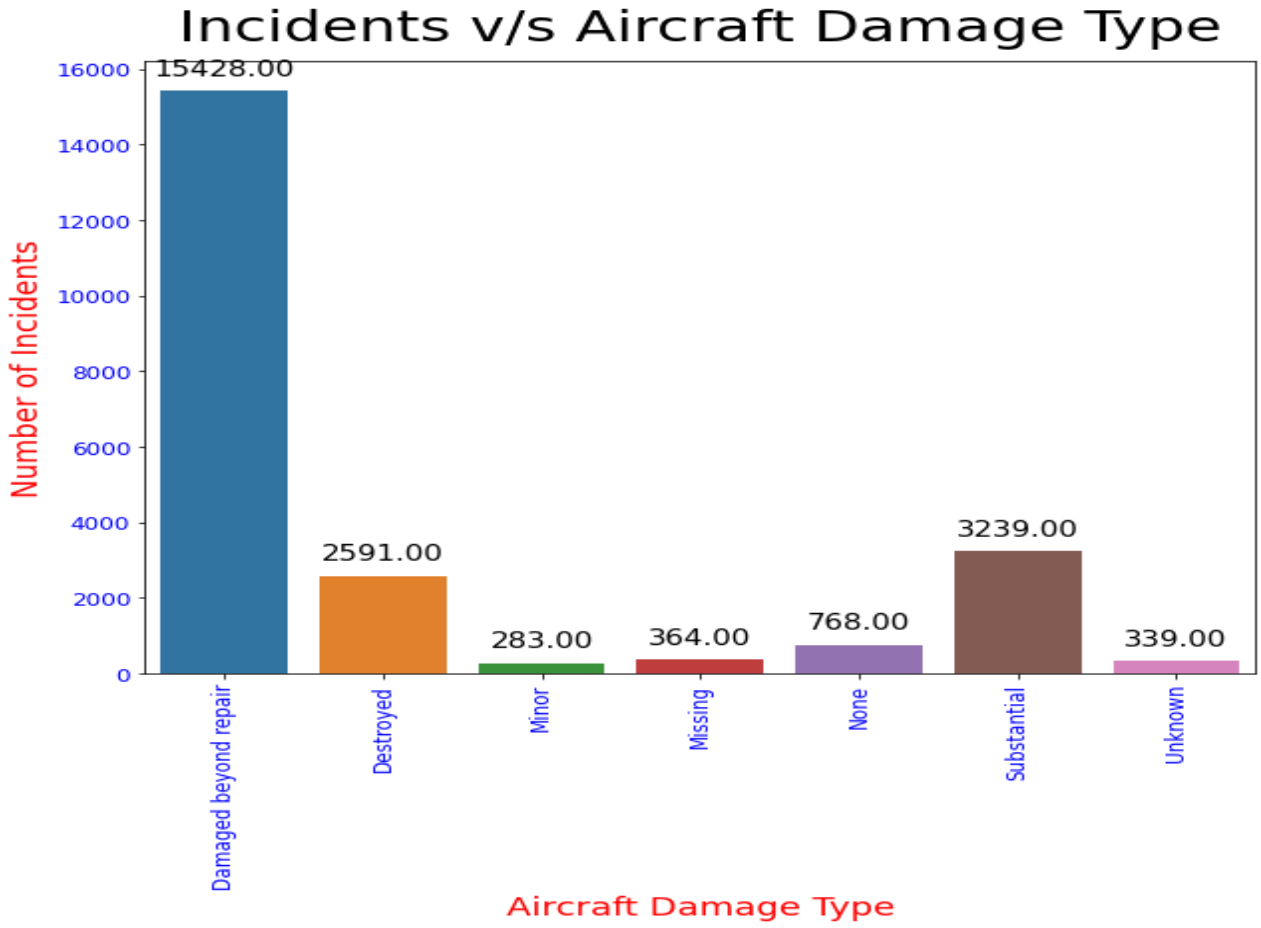


For each type of Aircraft Damage, we have attempted to display the fatality count.

Our conclusion is that when DESTROYED is the damage type of aircraft, the fatalities are pretty high compared to any other type of damage.

Our conclusion is that when an aircraft's damage type is "DESTROYED," the number of fatalities is pretty high.

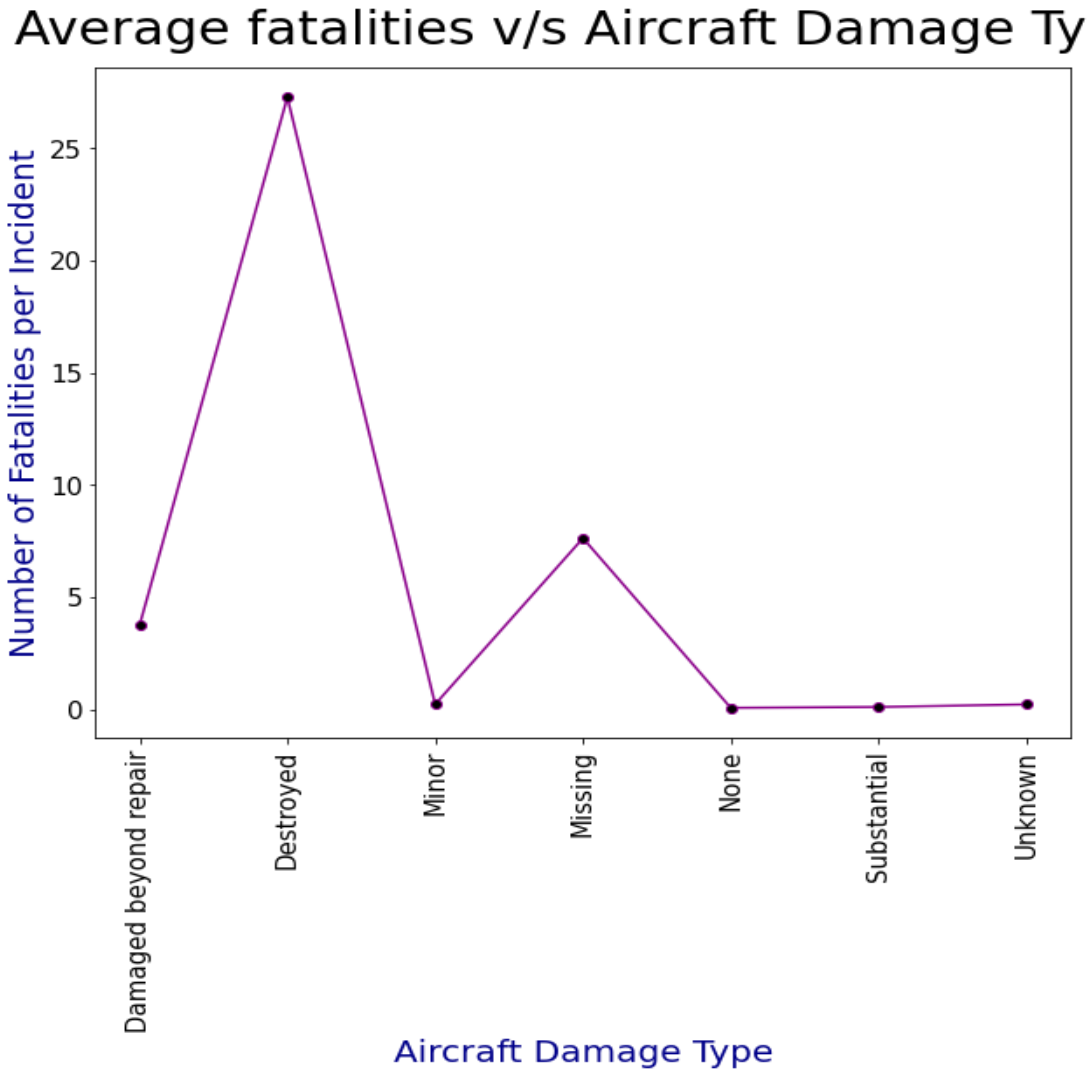
As when the Aircraft gets destroyed it's a very rare chance of survival of the passengers.



For each type of Aircraft Damage, we have attempted to display the incident count.

Our conclusion is that when" DESTROYED BEYOND REPAIR " is the damage type of aircraft damage, the incidents are pretty high compared to any other type of damage.

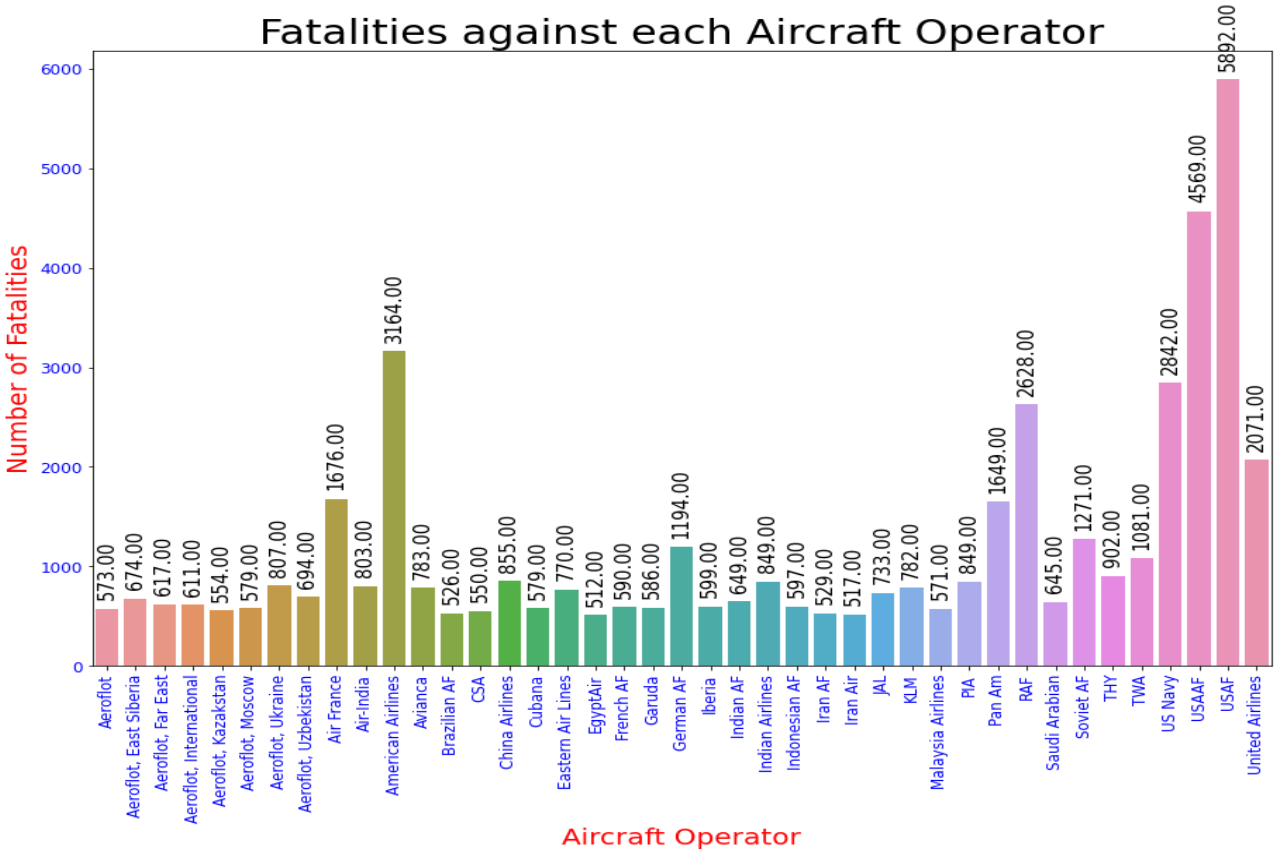
Our conclusion is that when an aircraft's damage type is "DESTROYED BEYOND REPAIR" the number of incidents is pretty high.



For each type of Aircraft Damage, we have attempted to display the fatality count by incident.

Our conclusion is that when DESTROYED is the damage type of aircraft, the fatalities are pretty high compared to any other type of damage.

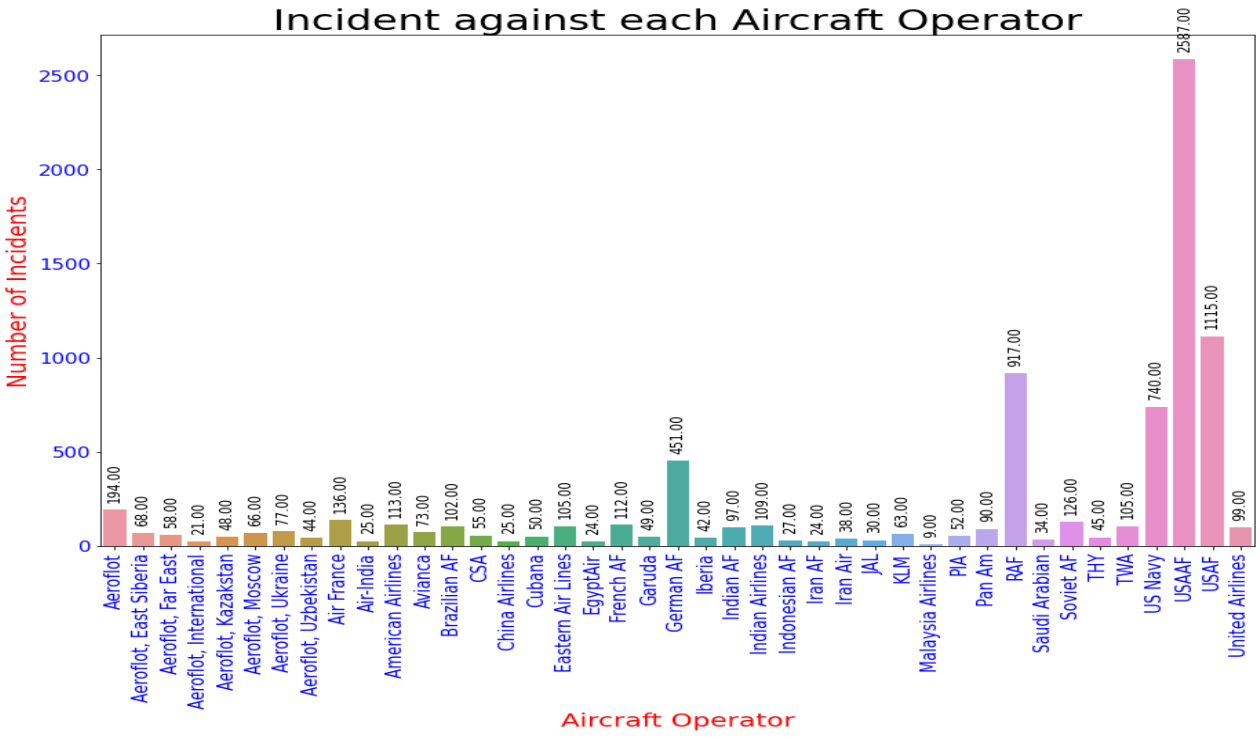
Our conclusion is that when an aircraft's damage type is "DESTROYED," the number of fatalities is pretty high.



According to the type of aircraft operator, we have attempted to display the number Fatalities.

In Analysing this Bar plot, we can determine that the number of incident are pretty high in USAF (United states Air Force).

In light of the graph above, we can conclude that USAF type aircraft have fewer incidents than USAAF aircraft but have more fatalities.



According to the type of aircraft operator, we have attempted to display the number incident.

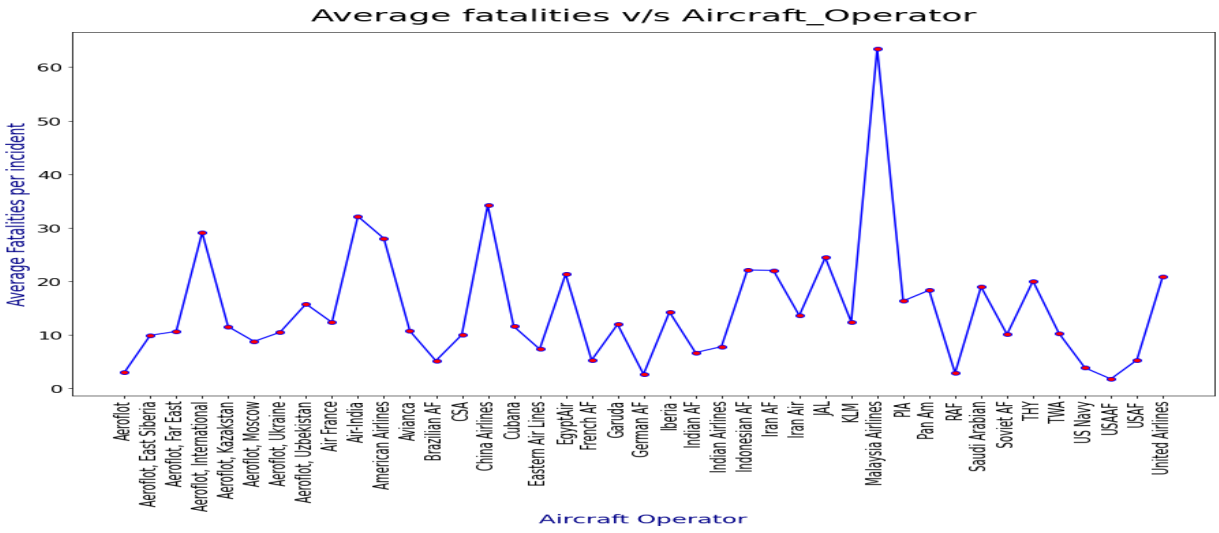
In Analysing this Bar plot, we can determine that the number of incident are pretty high in USAAF (United states Army Air Force).

Following may the reasons for the Cause of USAAF Airline Incidents: -

Due to Training

Due to Human Error

Due to Missile training



According to the type of aircraft operator, we have attempted to display the average fatality per incident.

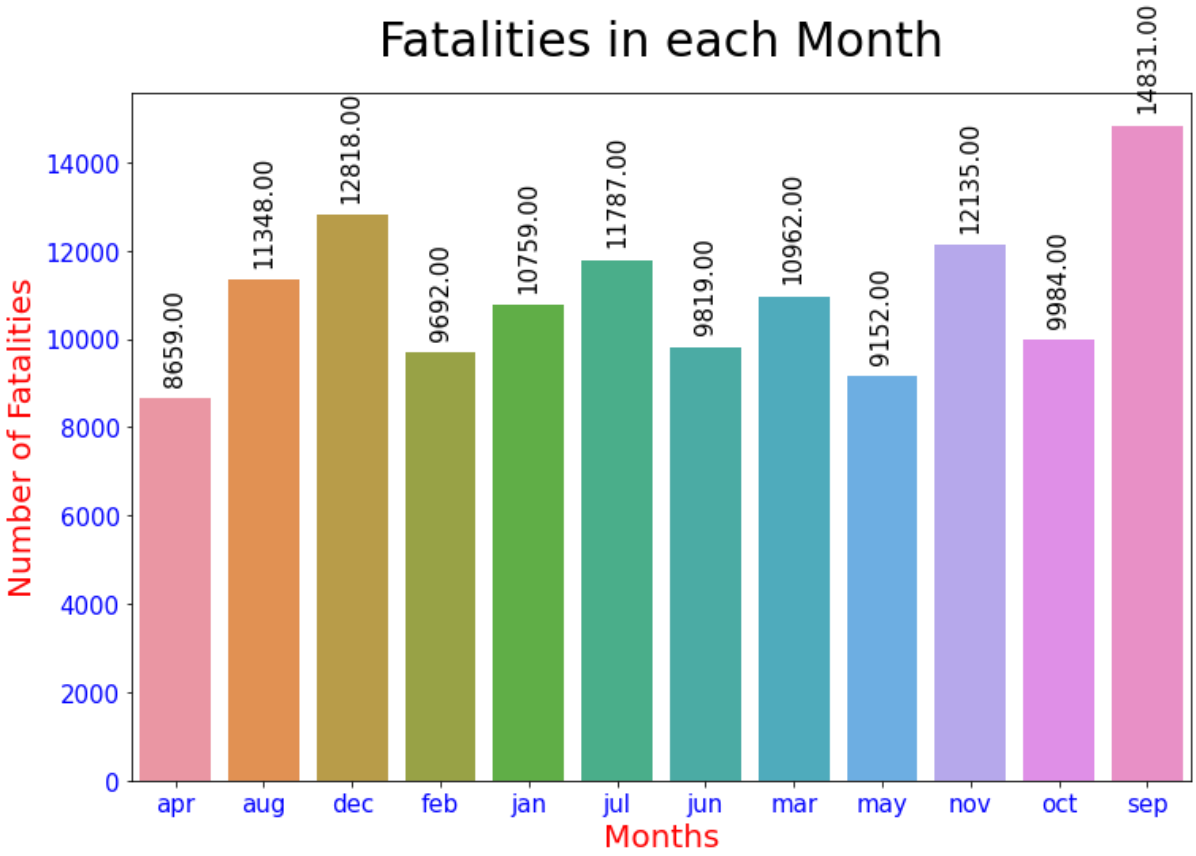
In Analysing this scatter plot, we can determine that the average fatality occurs by the Malaysia Airlines.

Following may the reasons for the Cause of Malaysian Airline Incidents: -

Bad weather

Pilot error

Mechanical Failure



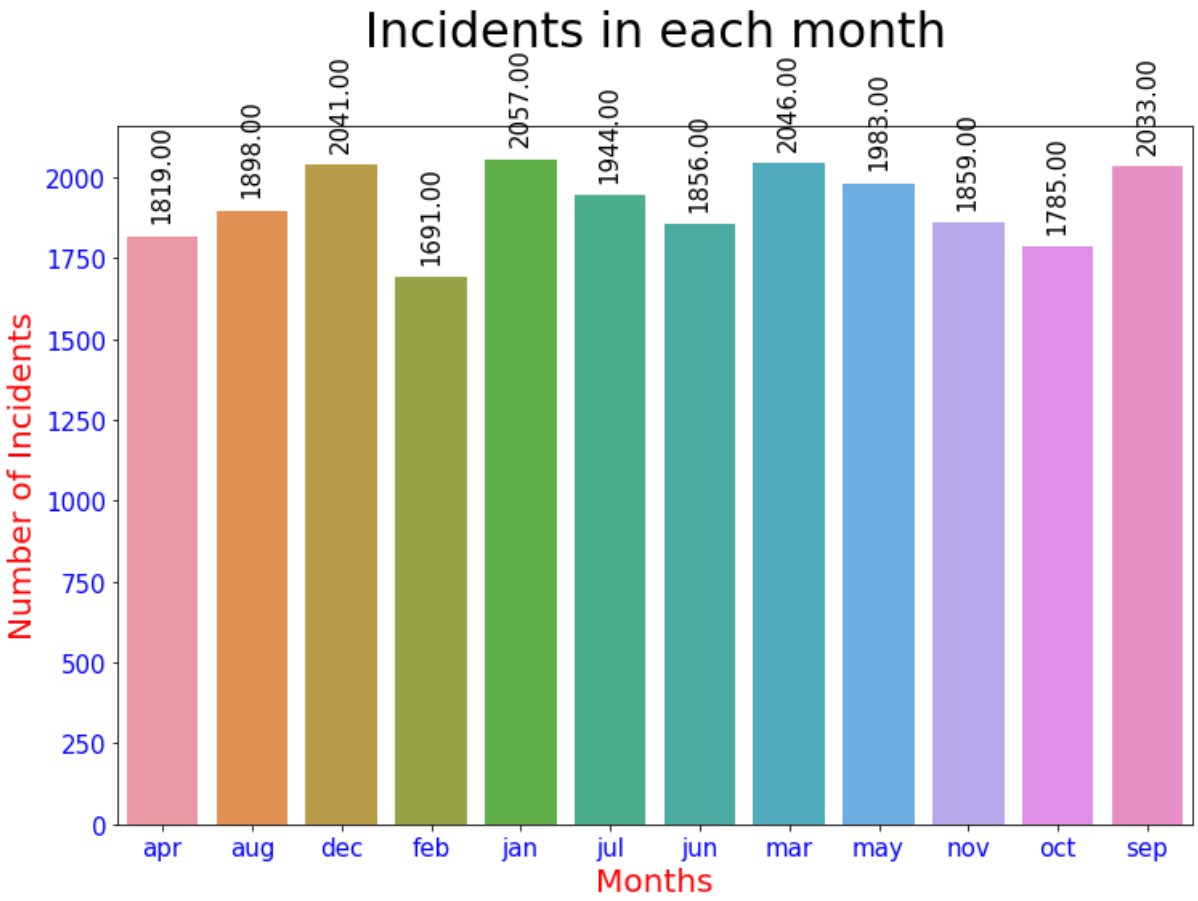
Based on the Which Month, we have attempted to display the number of fatalities per incident.

As evident from this Bar plot, the majority of fatalities occur in the months of November and September and December due to bad weather as these two months have pretty bad weather conditions in this part of the year.

Reasons: -

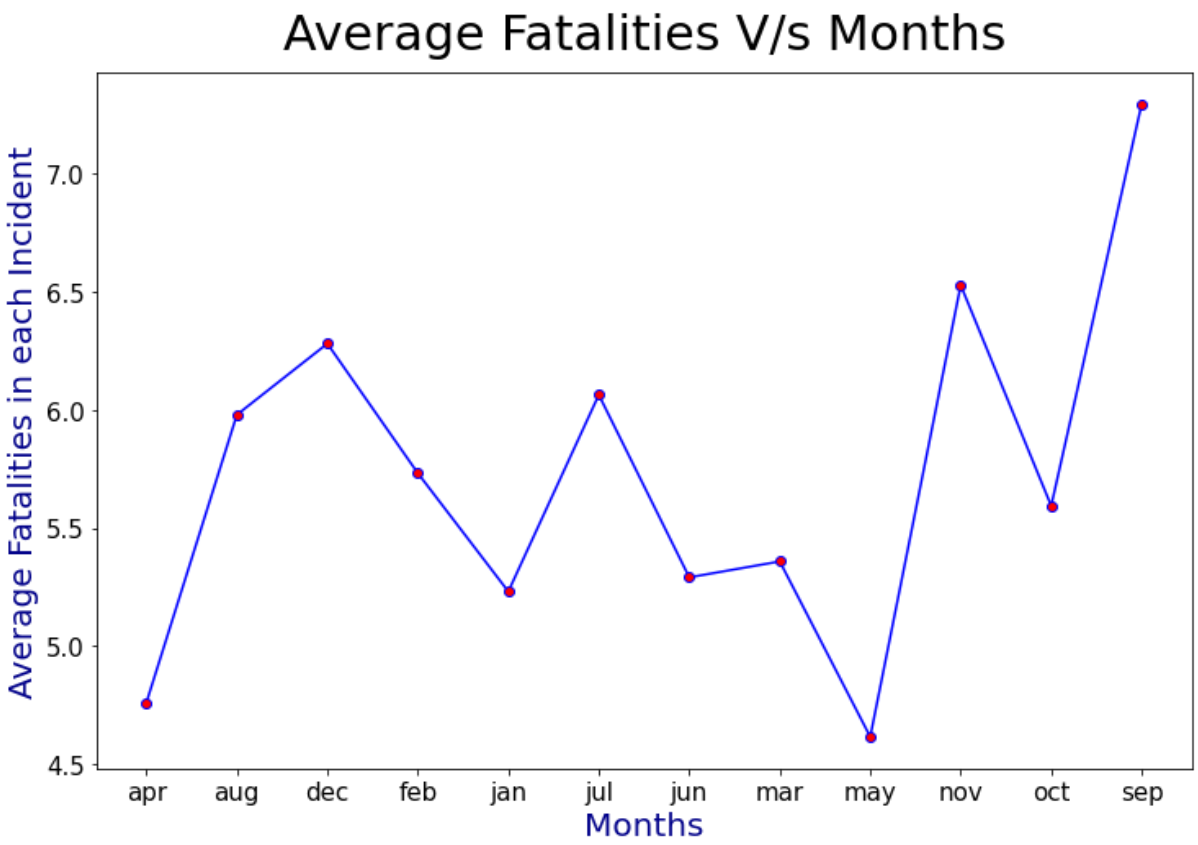
Iceing due to the cold weather

The Fogg in the winter days



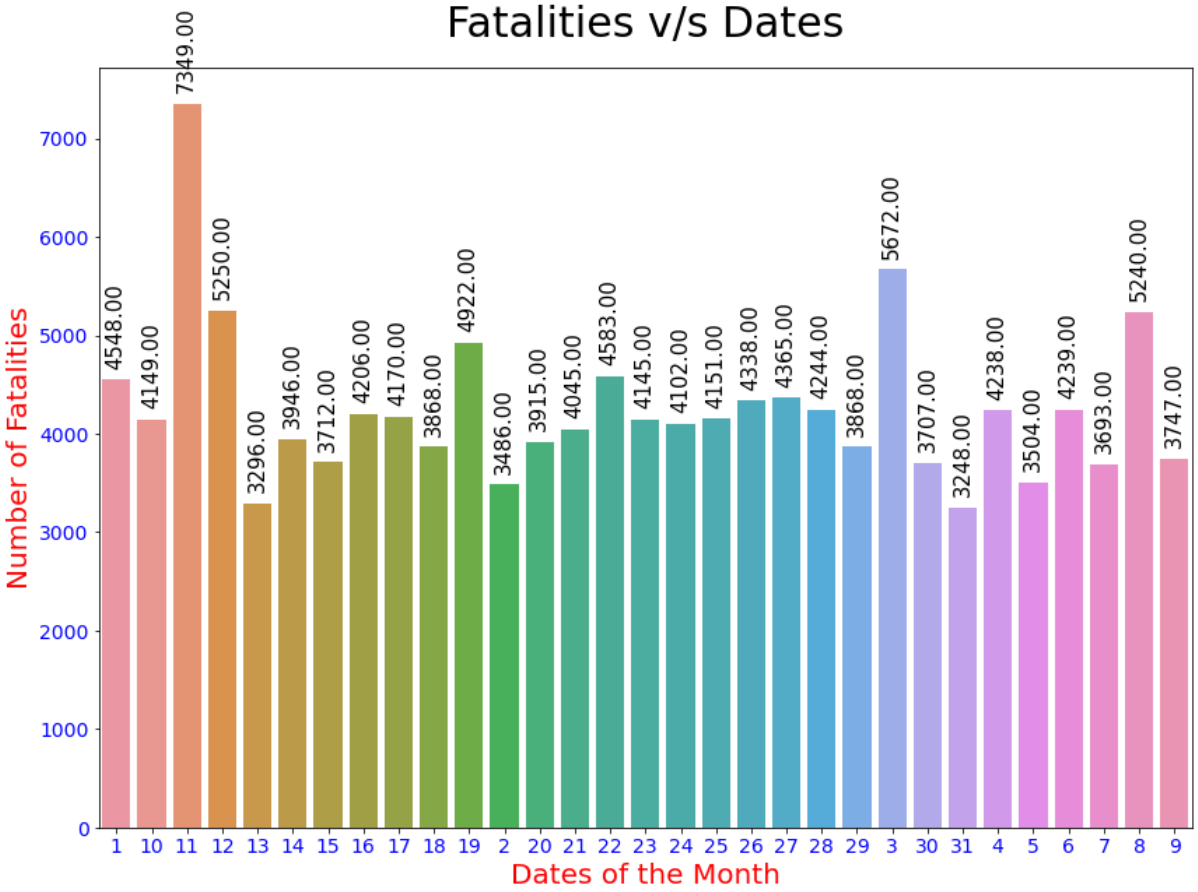
This following chart illustrates which month of the year has the highest incidence of aircraft accidents.

In conclusion, we can see that there is no particular difference between the number of incidents per month but we can gradually see that there are more incidents in the months of December/September/January.



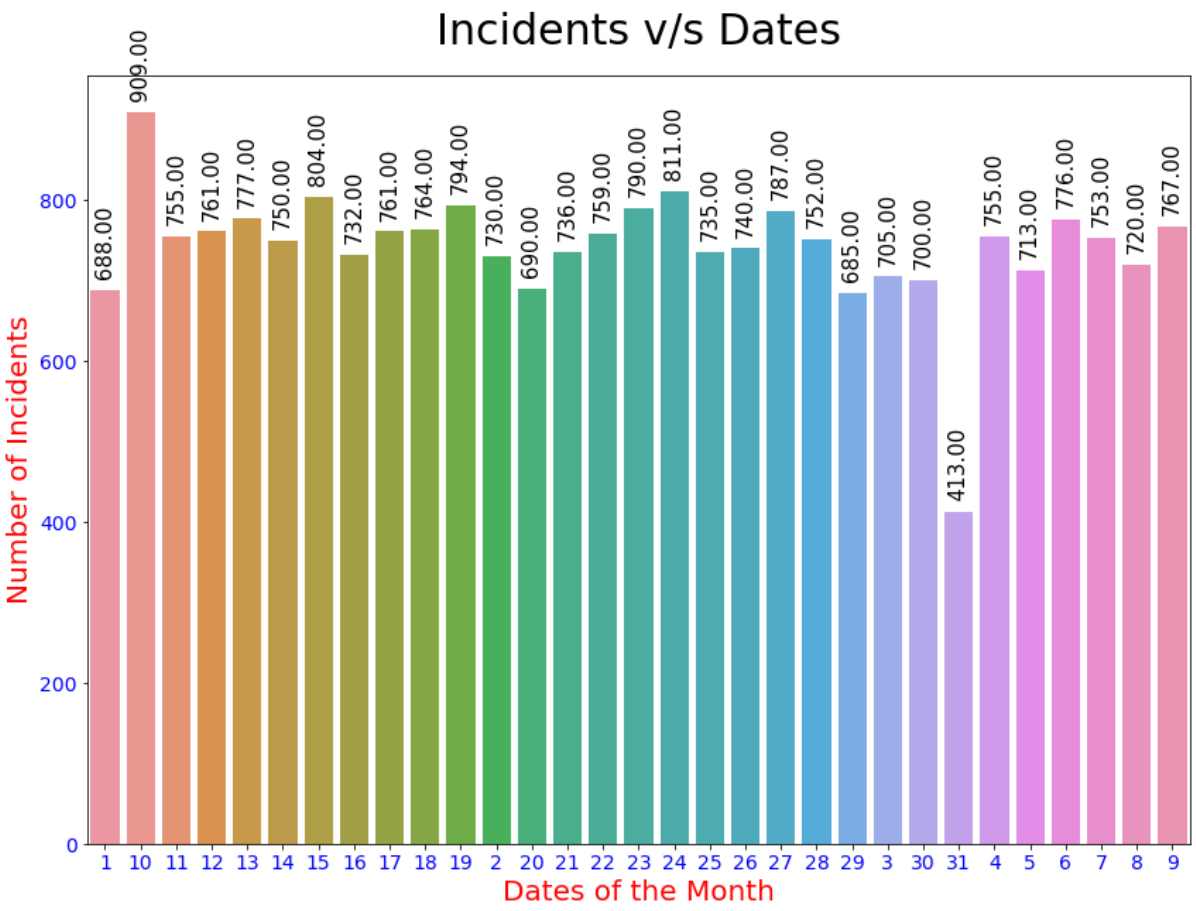
Based on the type of Month, we have attempted to display the average number of fatalities per incident.

As evident from this scatter plot, the majority of fatalities occur in the months of November and September due to bad weather as these two months have pretty bad weather conditions in this part of the year.



Based on the type of Date of the Month, we have attempted to display the Number of fatalities.

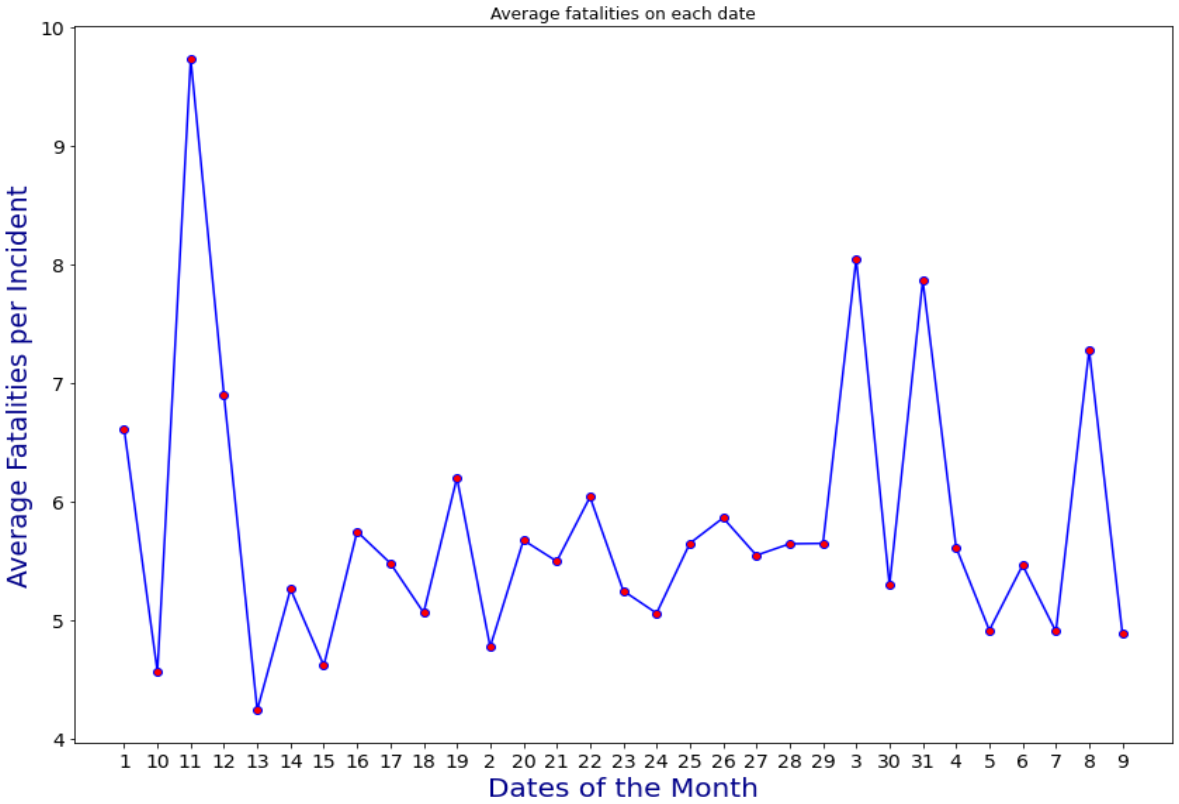
By examining this Bar plot, we can conclude that the average fatality occurs in the second week of a month, or exactly from 10th to 13th.



In this following Bar Chart, we enlighten that which Dates of the month of is verry peak for Number of Incidents

And as we can conclude that there is no particular difference between number of incidents per Date here but we can gradually see that the incidents on the Date of 31st of every Month are less than the other Dates.

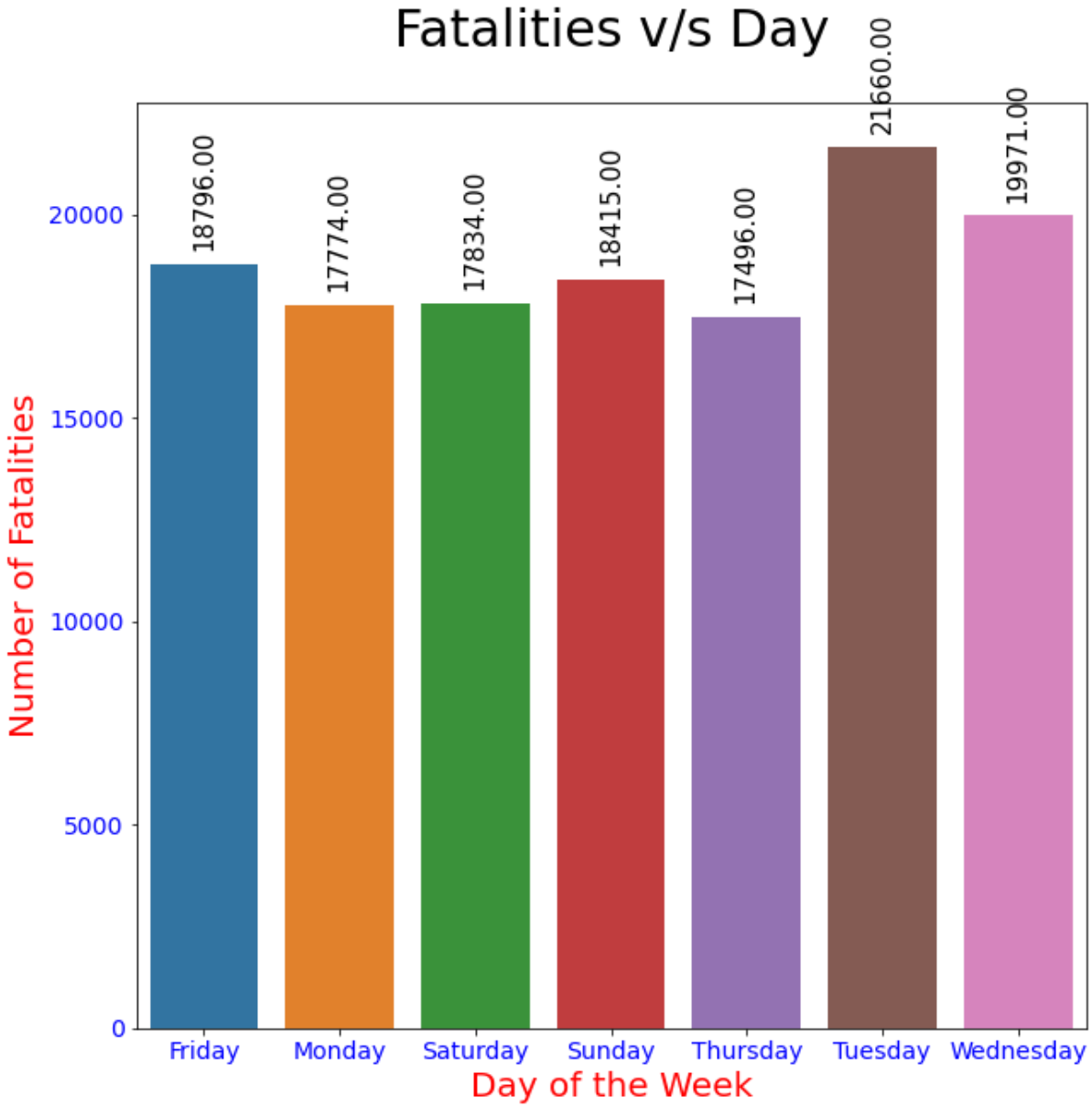
And Most incidents occur on the 10th of every month which is probably a start of the month.



Based on the type of Date of the Month, we have attempted to display the average number of fatalities per incident.

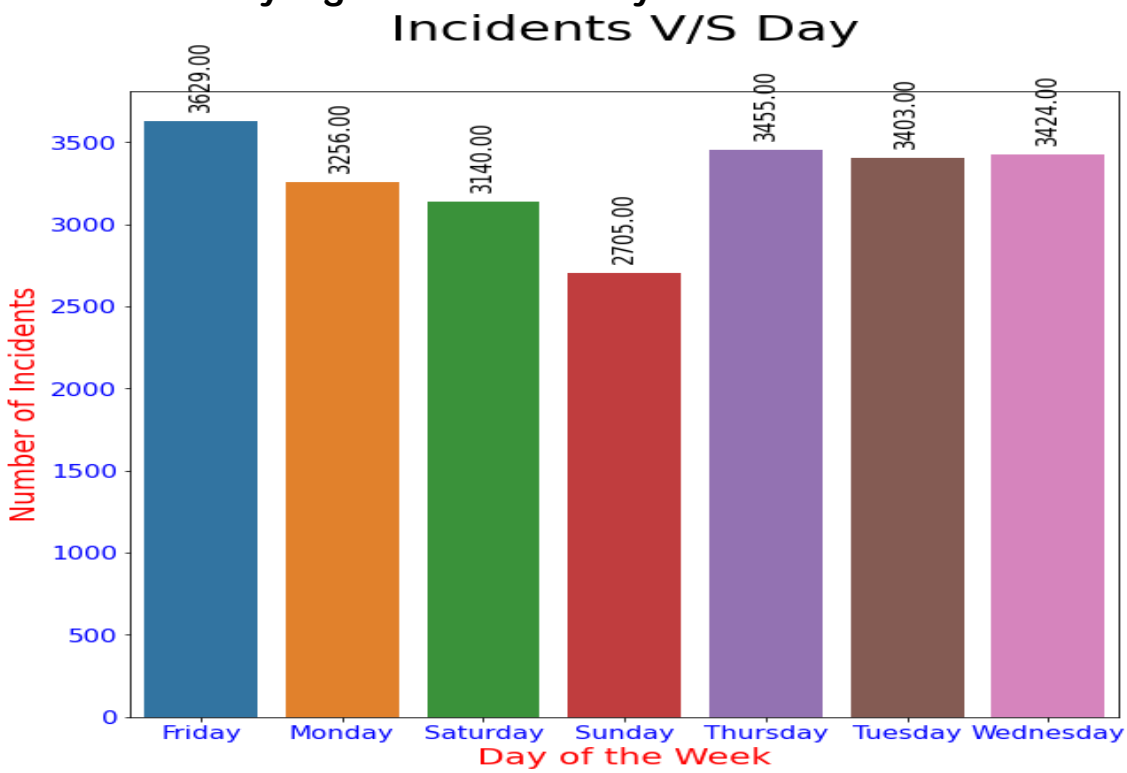
By examining this scatter plot, we can conclude that the average fatality occurs in the second week of a month, or exactly from

10th to 13th.



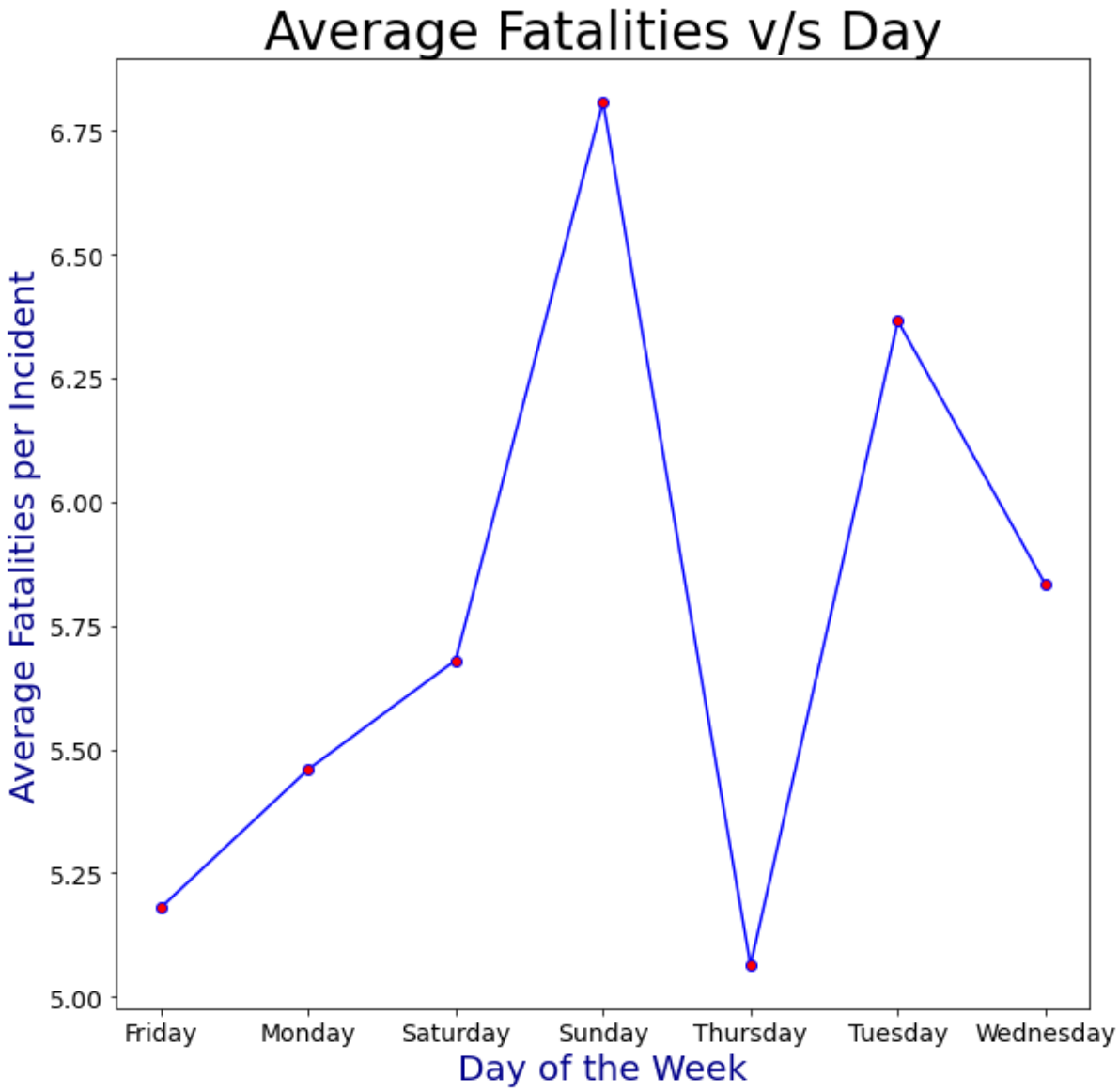
Based on the type of Date of the Week, we have attempted to display the Number of fatalities.

By examining this Bar plot, we can conclude that the Number of fatalities is very high on the Tuesday.



Based on the type of Date of the Week, we have attempted to display the Number of fatalities.

Based on the Bar plot, it appears that the number of incidents does not follow any upward or downward trend. In addition, Sunday is the only day when number of incidences are lower than the other days of the week.



Based on the type of Date of the Week, we have attempted to display the average number of fatalities per incident.

By examining this scatter plot, we can conclude that the average fatality occurs on the Sunday.

Results and Discussions :-

- **The aircraft has a close proximity to the ground and is in a more vulnerable configuration than it would be in other phases of flight: the crew has to deal with a high workload and small maneuver margins during these phases.**
- **It makes sense: Ice on wings, slippery runways, etc. are common during winter months. Are there other factors that contribute to the higher accident rate during the winter**
- **There were many causes for Malaysia Airlines incidents, including bad weather, pilot errors, and mechanical failures.**
- **There may be three main reasons why USAAF airliners crash:-Due to Training, Due to Human Error, Due to Ambush training.**

LIMITATIONS :-

- In this project, we considered only a few of the Features. One could further work on the locations of the incidents and could find the hotspots.
- Further one could judge whether these hotspots were static or changing over time.
- We didn't considered time of the incidents in our project as we couldn't get proper and complete information about the time, which could be an important factor.
- Also , we didn't considered those Aircraft Operators in which Total Fatalities were less than 500. So, one could even try to find patterns under these.

Conclusions :-

- **When an aircraft's damage type is "DESTROYED," the number of fatalities is pretty high.**
- **The fact that winters have a very high rate of accidents can be prevented by taking safety precautions like checking the wings and engine for ice, and properly cleaning the runways.**
- **Often times, the reason for aircraft incidents is not only weather-related but human error as well, so aircraft companies need trained personnel to perform quality checks.**
- **A military aircraft incident may occur during training or an ambush, but has fewer fatalities due to good emergency exits.**

References :-

- Data Sets:- https://www.kaggle.com
- Pandas :-https://pandas.pydata.org/docs
- Scipy :-https://docs.scipy.org/doc/scipy
- Literature review 1:- https://www.researchgate.net/publication/318154819_Aviation_Accident_Analysis_A_Case_Study
- Literature review 2 :- https://www.ijrte.org/wp-content/uploads/papers/v8i6s/F10270386S20.pdf
- Literature review 3 :- https://www.ijeat.org/wp-content/uploads/papers/v9i3s/C10090393S20.pdf
- Literature review 4 :- https://www.researchgate.net/publication/348429441_PIA_Accidents_Analysis_Using_Naive_Bayes_Classifier