

# Final Project Submission

Please fill out:

- Student name: Claire Njeri Nduhiu
- Student pace: part time
- Instructor name: Brian Chacha and Lians Wanjiku

## Business Understanding

This project aims to advise the head of the new aviation division, on aircraft that have lowest risk potential to accidents. Through Data Analysis the project will provide actionable insights to the venture, on which Aircrafts enhance air safety and optimise resource allocation.

## Importing the necessary Libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

## Data Staging

```
# Set the option to display all columns
pd.set_option('display.max_columns', None)
```

```
df=pd.read_csv('data\Aviation_Data.csv')
df.head()
```

```
c:\Users\Nduhiu\anaconda3\just_anaconda\envs\learn-env\lib\site-
packages\IPython\core\interactiveshell.py:3145: DtypeWarning: Columns
(6,7,28) have mixed types.Specify dtype option on import or set
low_memory=False.
```

```
has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

|   | Event.Id       | Investigation.Type | Accident.Number | Event.Date | \ |
|---|----------------|--------------------|-----------------|------------|---|
| 0 | 20001218X45444 | Accident           | SEA87LA080      | 1948-10-24 |   |
| 1 | 20001218X45447 | Accident           | LAX94LA336      | 1962-07-19 |   |
| 2 | 20061025X01555 | Accident           | NYC07LA005      | 1974-08-30 |   |
| 3 | 20001218X45448 | Accident           | LAX96LA321      | 1977-06-19 |   |
| 4 | 20041105X01764 | Accident           | CHI79FA064      | 1979-08-02 |   |

|   | Location        | Country       | Latitude | Longitude | Airport.Code | \ |
|---|-----------------|---------------|----------|-----------|--------------|---|
| 0 | MOOSE CREEK, ID | United States | NaN      | NaN       | NaN          |   |
| 1 | BRIDGEPORT, CA  | United States | NaN      | NaN       | NaN          |   |
| 2 | Saltville, VA   | United States | 36.9222  | -81.8781  | NaN          |   |
| 3 | EUREKA, CA      | United States | NaN      | NaN       | NaN          |   |
| 4 | Canton, OH      | United States | NaN      | NaN       | NaN          |   |

| Airport.Name Injury.Severity Aircraft.damage Aircraft.Category \ |               |          |           |          |
|------------------------------------------------------------------|---------------|----------|-----------|----------|
| 0                                                                | NaN           | Fatal(2) | Destroyed | NaN      |
| 1                                                                | NaN           | Fatal(4) | Destroyed | NaN      |
| 2                                                                | NaN           | Fatal(3) | Destroyed | NaN      |
| 3                                                                | NaN           | Fatal(2) | Destroyed | NaN      |
| 4                                                                | NaN           | Fatal(1) | Destroyed | NaN      |
| Registration.Number Make Model Amateur.Built                     |               |          |           |          |
| Number.of.Engines \                                              |               |          |           |          |
| 0                                                                | NC6404        | Stinson  | 108-3     | No       |
| 1.0                                                              |               |          |           |          |
| 1                                                                | N5069P        | Piper    | PA24-180  | No       |
| 1.0                                                              |               |          |           |          |
| 2                                                                | N5142R        | Cessna   | 172M      | No       |
| 1.0                                                              |               |          |           |          |
| 3                                                                | N1168J        | Rockwell | 112       | No       |
| 1.0                                                              |               |          |           |          |
| 4                                                                | N15NY         | Cessna   | 501       | No       |
| NaN                                                              |               |          |           |          |
| Engine.Type FAR.Description Schedule Purpose.of.flight           |               |          |           |          |
| Air.carrier \                                                    |               |          |           |          |
| 0                                                                | Reciprocating | NaN      | NaN       | Personal |
| NaN                                                              |               |          |           |          |
| 1                                                                | Reciprocating | NaN      | NaN       | Personal |
| NaN                                                              |               |          |           |          |
| 2                                                                | Reciprocating | NaN      | NaN       | Personal |
| NaN                                                              |               |          |           |          |
| 3                                                                | Reciprocating | NaN      | NaN       | Personal |
| NaN                                                              |               |          |           |          |
| 4                                                                | NaN           | NaN      | NaN       | Personal |
| NaN                                                              |               |          |           |          |
| Total.Fatal.Injuries Total.Serious.Injuries Total.Minor.Injuries |               |          |           |          |
| \                                                                |               |          |           |          |
| 0                                                                | 2.0           |          | 0.0       | 0.0      |
| 1                                                                | 4.0           |          | 0.0       | 0.0      |
| 2                                                                | 3.0           |          | NaN       | NaN      |
| 3                                                                | 2.0           |          | 0.0       | 0.0      |
| 4                                                                | 1.0           |          | 2.0       | NaN      |
| Total.Uninjured Weather.Condition Broad.phase.of.flight          |               |          |           |          |
| Report.Status \                                                  |               |          |           |          |
| 0                                                                | 0.0           | UNK      | Cruise    | Probable |
| Cause                                                            |               |          |           |          |

|                  |            |     |          |          |
|------------------|------------|-----|----------|----------|
| 1                | 0.0        | UNK | Unknown  | Probable |
| Cause            |            |     |          |          |
| 2                | NaN        | IMC | Cruise   | Probable |
| Cause            |            |     |          |          |
| 3                | 0.0        | IMC | Cruise   | Probable |
| Cause            |            |     |          |          |
| 4                | 0.0        | VMC | Approach | Probable |
| Cause            |            |     |          |          |
| Publication.Date |            |     |          |          |
| 0                | NaN        |     |          |          |
| 1                | 19-09-1996 |     |          |          |
| 2                | 26-02-2007 |     |          |          |
| 3                | 12-09-2000 |     |          |          |
| 4                | 16-04-1980 |     |          |          |

## Data Understanding

Here i will be exploring the Data to get an understanding of the columns, shape of the data ,data types,missing values and duplicates.

```
df.shape
```

```
(90348, 31)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 90348 entries, 0 to 90347
```

```
Data columns (total 31 columns):
```

| #  | Column              | Non-Null | Count    | Dtype  |
|----|---------------------|----------|----------|--------|
| 0  | Event.Id            | 88889    | non-null | object |
| 1  | Investigation.Type  | 90348    | non-null | object |
| 2  | Accident.Number     | 88889    | non-null | object |
| 3  | Event.Date          | 88889    | non-null | object |
| 4  | Location            | 88837    | non-null | object |
| 5  | Country             | 88663    | non-null | object |
| 6  | Latitude            | 34382    | non-null | object |
| 7  | Longitude           | 34373    | non-null | object |
| 8  | Airport.Code        | 50249    | non-null | object |
| 9  | Airport.Name        | 52790    | non-null | object |
| 10 | Injury.Severity     | 87889    | non-null | object |
| 11 | Aircraft.damage     | 85695    | non-null | object |
| 12 | Aircraft.Category   | 32287    | non-null | object |
| 13 | Registration.Number | 87572    | non-null | object |
| 14 | Make                | 88826    | non-null | object |
| 15 | Model               | 88797    | non-null | object |
| 16 | Amateur.Built       | 88787    | non-null | object |

|    |                        |       |          |         |
|----|------------------------|-------|----------|---------|
| 17 | Number.of.Engines      | 82805 | non-null | float64 |
| 18 | Engine.Type            | 81812 | non-null | object  |
| 19 | FAR.Description        | 32023 | non-null | object  |
| 20 | Schedule               | 12582 | non-null | object  |
| 21 | Purpose.of.flight      | 82697 | non-null | object  |
| 22 | Air.carrier            | 16648 | non-null | object  |
| 23 | Total.Fatal.Injuries   | 77488 | non-null | float64 |
| 24 | Total.Serious.Injuries | 76379 | non-null | float64 |
| 25 | Total.Minor.Injuries   | 76956 | non-null | float64 |
| 26 | Total.Uninjured        | 82977 | non-null | float64 |
| 27 | Weather.Condition      | 84397 | non-null | object  |
| 28 | Broad.phase.of.flight  | 61724 | non-null | object  |
| 29 | Report.Status          | 82508 | non-null | object  |
| 30 | Publication.Date       | 73659 | non-null | object  |

dtypes: float64(5), object(26)

memory usage: 21.4+ MB

df.isnull().sum()

|                        |       |
|------------------------|-------|
| Event.Id               | 1459  |
| Investigation.Type     | 0     |
| Accident.Number        | 1459  |
| Event.Date             | 1459  |
| Location               | 1511  |
| Country                | 1685  |
| Latitude               | 55966 |
| Longitude              | 55975 |
| Airport.Code           | 40099 |
| Airport.Name           | 37558 |
| Injury.Severity        | 2459  |
| Aircraft.damage        | 4653  |
| Aircraft.Category      | 58061 |
| Registration.Number    | 2776  |
| Make                   | 1522  |
| Model                  | 1551  |
| Amateur.Built          | 1561  |
| Number.of.Engines      | 7543  |
| Engine.Type            | 8536  |
| FAR.Description        | 58325 |
| Schedule               | 77766 |
| Purpose.of.flight      | 7651  |
| Air.carrier            | 73700 |
| Total.Fatal.Injuries   | 12860 |
| Total.Serious.Injuries | 13969 |
| Total.Minor.Injuries   | 13392 |
| Total.Uninjured        | 7371  |
| Weather.Condition      | 5951  |
| Broad.phase.of.flight  | 28624 |
| Report.Status          | 7840  |

```
Publication.Date      16689
dtype: int64
```

```
df.isnull().mean()*100
```

```
Event.Id      1.614867
Investigation.Type  0.000000
Accident.Number  1.614867
Event.Date    1.614867
Location      1.672422
Country       1.865011
Latitude      61.944924
Longitude     61.954886
Airport.Code   44.382831
Airport.Name   41.570372
Injury.Severity  2.721698
Aircraft.damage  5.150086
Aircraft.Category  64.263736
Registration.Number  3.072564
Make          1.684597
Model         1.716695
Amateur.Built  1.727764
Number.of.Engines  8.348829
Engine.Type    9.447913
FAR.Description  64.555939
Schedule       86.073848
Purpose.of.flight  8.468367
Air.carrier    81.573471
Total.Fatal.Injuries  14.233851
Total.Serious.Injuries  15.461327
Total.Minor.Injuries  14.822686
Total.Uninjured   8.158454
Weather.Condition  6.586753
Broad.phase.of.flight  31.681941
Report.Status    8.677558
Publication.Date  18.471909
dtype: float64
```

```
df.duplicated().sum()
```

```
1390
```

```
df.columns
```

```
Index(['Event.Id', 'Investigation.Type', 'Accident.Number',
      'Event.Date',
      'Location', 'Country', 'Latitude', 'Longitude', 'Airport.Code',
      'Airport.Name', 'Injury.Severity', 'Aircraft.damage',
      'Aircraft.Category', 'Registration.Number', 'Make', 'Model',
      'Amateur.Built', 'Number.of.Engines', 'Engine.Type',
      'FAR.Description',
```

```

'Schedule', 'Purpose.of.flight', 'Air.carrier',
'Total.Fatal.Injuries',
'Total.Serious.Injuries', 'Total.Minor.Injuries',
'Total.Uninjured',
'Weather.Condition', 'Broad.phase.of.flight', 'Report.Status',
'Publication.Date'],
dtype='object')

```

## Data Preparation

From the data understanding done earlier, we found missing values, duplicates. To deal with the missing values Columns with High number of missing observations will be dropped, we also drop the duplicated rows.

For continuous data with missing values we imputed with the mean, while, for categorical columns with missing values, we dropped the rows.

```

df=df.drop_duplicates()
df.duplicated().sum()
0
df.isnull().mean().sort_values(ascending=False)*100

```

|                        |           |
|------------------------|-----------|
| Schedule               | 85.856247 |
| Air.carrier            | 81.285550 |
| FAR.Description        | 64.002113 |
| Aircraft.Category      | 63.705344 |
| Longitude              | 61.360417 |
| Latitude               | 61.350300 |
| Airport.Code           | 43.513793 |
| Airport.Name           | 40.657389 |
| Broad.phase.of.flight  | 30.614447 |
| Publication.Date       | 17.198004 |
| Total.Serious.Injuries | 14.140381 |
| Total.Minor.Injuries   | 13.491760 |
| Total.Fatal.Injuries   | 12.893725 |
| Engine.Type            | 8.033004  |
| Report.Status          | 7.250613  |
| Purpose.of.flight      | 7.038153  |
| Number.ofEngines       | 6.916747  |
| Total.Uninjured        | 6.723398  |
| Weather.Condition      | 5.127139  |
| Aircraft.damage        | 3.668023  |
| Registration.Number    | 1.558039  |
| Injury.Severity        | 1.201691  |
| Country                | 0.331617  |
| Amateur.Built          | 0.192226  |
| Model                  | 0.180984  |

```
Make                0.148385
Location            0.136019
Event.Date          0.077565
Accident.Number     0.077565
Event.Id            0.077565
Investigation.Type  0.000000
dtype: float64
```

```
columns_drop=['Schedule','Accident.Number','Air.carrier','FAR.Descript
ion','Aircraft.Category','Longitude','Latitude','Airport.Code','Airpor
t.Name','Registration.Number',
'Event.Id']
```

```
df=df.drop(columns=columns_drop)
```

```
df.columns
```

```
Index(['Investigation.Type', 'Event.Date', 'Location', 'Country',
       'Injury.Severity', 'Aircraft.damage', 'Make', 'Model',
       'Amateur.Built',
       'Number.ofEngines', 'Engine.Type', 'Purpose.of.flight',
       'Total.Fatal.Injuries', 'Total.Serious.Injuries',
       'Total.Minor.Injuries', 'Total.Uninjured', 'Weather.Condition',
       'Broad.phase.of.flight', 'Report.Status', 'Publication.Date'],
      dtype='object')
```

```
df.isnull().mean().sort_values(ascending=False)*100
```

```
Broad.phase.of.flight    30.614447
Publication.Date          17.198004
Total.Serious.Injuries   14.140381
Total.Minor.Injuries     13.491760
Total.Fatal.Injuries     12.893725
Engine.Type              8.033004
Report.Status            7.250613
Purpose.of.flight        7.038153
Number.ofEngines         6.916747
Total.Uninjured          6.723398
Weather.Condition        5.127139
Aircraft.damage          3.668023
Injury.Severity          1.201691
Country                  0.331617
Amateur.Built            0.192226
Model                    0.180984
Make                     0.148385
Location                 0.136019
Event.Date               0.077565
Investigation.Type       0.000000
dtype: float64
```

```
df.describe()
```

|       | Number.of.Engines | Total.Fatal.Injuries | Total.Serious.Injuries |
|-------|-------------------|----------------------|------------------------|
| \     |                   |                      |                        |
| count | 82805.000000      | 77488.000000         | 76379.000000           |
| mean  | 1.146585          | 0.647855             | 0.279881               |
| std   | 0.446510          | 5.485960             | 1.544084               |
| min   | 0.000000          | 0.000000             | 0.000000               |
| 25%   | 1.000000          | 0.000000             | 0.000000               |
| 50%   | 1.000000          | 0.000000             | 0.000000               |
| 75%   | 1.000000          | 0.000000             | 0.000000               |
| max   | 8.000000          | 349.000000           | 161.000000             |

|       | Total.Minor.Injuries | Total.Uninjured |
|-------|----------------------|-----------------|
| count | 76956.000000         | 82977.000000    |
| mean  | 0.357061             | 5.325440        |
| std   | 2.235625             | 27.913634       |
| min   | 0.000000             | 0.000000        |
| 25%   | 0.000000             | 0.000000        |
| 50%   | 0.000000             | 1.000000        |
| 75%   | 0.000000             | 2.000000        |
| max   | 380.000000           | 699.000000      |

df.info()

<class 'pandas.core.frame.DataFrame'>

Int64Index: 88958 entries, 0 to 90347

Data columns (total 20 columns):

| #   | Column                 | Non-Null | Count    | Dtype   |
|-----|------------------------|----------|----------|---------|
| --- | -----                  | -----    | -----    | -----   |
| 0   | Investigation.Type     | 88958    | non-null | object  |
| 1   | Event.Date             | 88889    | non-null | object  |
| 2   | Location               | 88837    | non-null | object  |
| 3   | Country                | 88663    | non-null | object  |
| 4   | Injury.Severity        | 87889    | non-null | object  |
| 5   | Aircraft.damage        | 85695    | non-null | object  |
| 6   | Make                   | 88826    | non-null | object  |
| 7   | Model                  | 88797    | non-null | object  |
| 8   | Amateur.Built          | 88787    | non-null | object  |
| 9   | Number.of.Engines      | 82805    | non-null | float64 |
| 10  | Engine.Type            | 81812    | non-null | object  |
| 11  | Purpose.of.flight      | 82697    | non-null | object  |
| 12  | Total.Fatal.Injuries   | 77488    | non-null | float64 |
| 13  | Total.Serious.Injuries | 76379    | non-null | float64 |
| 14  | Total.Minor.Injuries   | 76956    | non-null | float64 |



|    |                       |       |          |         |
|----|-----------------------|-------|----------|---------|
| 15 | Total.Uninjured       | 82977 | non-null | float64 |
| 16 | Weather.Condition     | 84397 | non-null | object  |
| 17 | Broad.phase.of.flight | 61724 | non-null | object  |
| 18 | Report.Status         | 82508 | non-null | object  |
| 19 | Publication.Date      | 73659 | non-null | object  |

dtypes: float64(5), object(15)

memory usage: 14.3+ MB

```
continous_df=df.select_dtypes(include=['number']) #this is to select the continous data type
```

```
continous_df.columns
```

```
Index(['Number.ofEngines', 'Total.Fatal.Injuries',
      'Total.Serious.Injuries',
      'Total.Minor.Injuries', 'Total.Uninjured'],
      dtype='object')
```

```
cartegorical_df=df.select_dtypes(include=['object']) ##this is to select the cartegorical data type
```

```
cartegorical_df.columns
```

```
Index(['Investigation.Type', 'Event.Date', 'Location', 'Country',
      'Injury.Severity', 'Aircraft.damage', 'Make', 'Model',
      'Amateur.Built',
      'Engine.Type', 'Purpose.of.flight', 'Weather.Condition',
      'Broad.phase.of.flight', 'Report.Status', 'Publication.Date'],
      dtype='object')
```

```
continous_col=['Number.ofEngines', 'Total.Fatal.Injuries',
              'Total.Serious.Injuries',
              'Total.Minor.Injuries', 'Total.Uninjured']
```

```
df[continous_col]=df[continous_col].fillna(df[continous_col].mean())
```

```
df.isnull().sum()
```

|                        |      |
|------------------------|------|
| Investigation.Type     | 0    |
| Event.Date             | 69   |
| Location               | 121  |
| Country                | 295  |
| Injury.Severity        | 1069 |
| Aircraft.damage        | 3263 |
| Make                   | 132  |
| Model                  | 161  |
| Amateur.Built          | 171  |
| Number.ofEngines       | 0    |
| Engine.Type            | 7146 |
| Purpose.of.flight      | 6261 |
| Total.Fatal.Injuries   | 0    |
| Total.Serious.Injuries | 0    |
| Total.Minor.Injuries   | 0    |

```

Total.Uninjured          0
Weather.Condition        4561
Broad.phase.of.flight    27234
Report.Status            6450
Publication.Date         15299
dtype: int64

categorical_col=['Investigation.Type', 'Event.Date', 'Location',
                 'Country', 'Injury.Severity', 'Aircraft.damage', 'Make',
                 'Model',
                 'Amateur.Built', 'Engine.Type', 'Purpose.of.flight',
                 'Weather.Condition', 'Broad.phase.of.flight', 'Report.Status',
                 'Publication.Date']

df=df.dropna()

df.isnull().sum()

Investigation.Type      0
Event.Date              0
Location                0
Country                 0
Injury.Severity         0
Aircraft.damage         0
Make                    0
Model                   0
Amateur.Built           0
Number.of.Engines       0
Engine.Type             0
Purpose.of.flight       0
Total.Fatal.Injuries    0
Total.Serious.Injuries  0
Total.Minor.Injuries    0
Total.Uninjured         0
Weather.Condition       0
Broad.phase.of.flight   0
Report.Status           0
Publication.Date        0
dtype: int64

```

## Data Analysis

In this section we ,plotted a heat map,pie charts,bar charts,and a line chart.

```

df.corr()

```

|                        | Number.of.Engines | Total.Fatal.Injuries | \ |
|------------------------|-------------------|----------------------|---|
| Number.of.Engines      | 1.000000          | 0.107369             |   |
| Total.Fatal.Injuries   | 0.107369          | 1.000000             |   |
| Total.Serious.Injuries | 0.046869          | 0.249504             |   |

|                        |          |           |
|------------------------|----------|-----------|
| Total.Minor.Injuries   | 0.072331 | 0.096988  |
| Total.Uninjured        | 0.345813 | -0.009425 |
| Total.Serious.Injuries |          |           |
| Total.Minor.Injuries \ |          |           |
| Number.of.Engines      | 0.046869 | 0.072331  |
| Total.Fatal.Injuries   | 0.249504 | 0.096988  |
| Total.Serious.Injuries | 1.000000 | 0.393828  |
| Total.Minor.Injuries   | 0.393828 | 1.000000  |
| Total.Uninjured        | 0.023362 | 0.166887  |

|                        |           |
|------------------------|-----------|
| Total.Uninjured        |           |
| Number.of.Engines      | 0.345813  |
| Total.Fatal.Injuries   | -0.009425 |
| Total.Serious.Injuries | 0.023362  |
| Total.Minor.Injuries   | 0.166887  |
| Total.Uninjured        | 1.000000  |

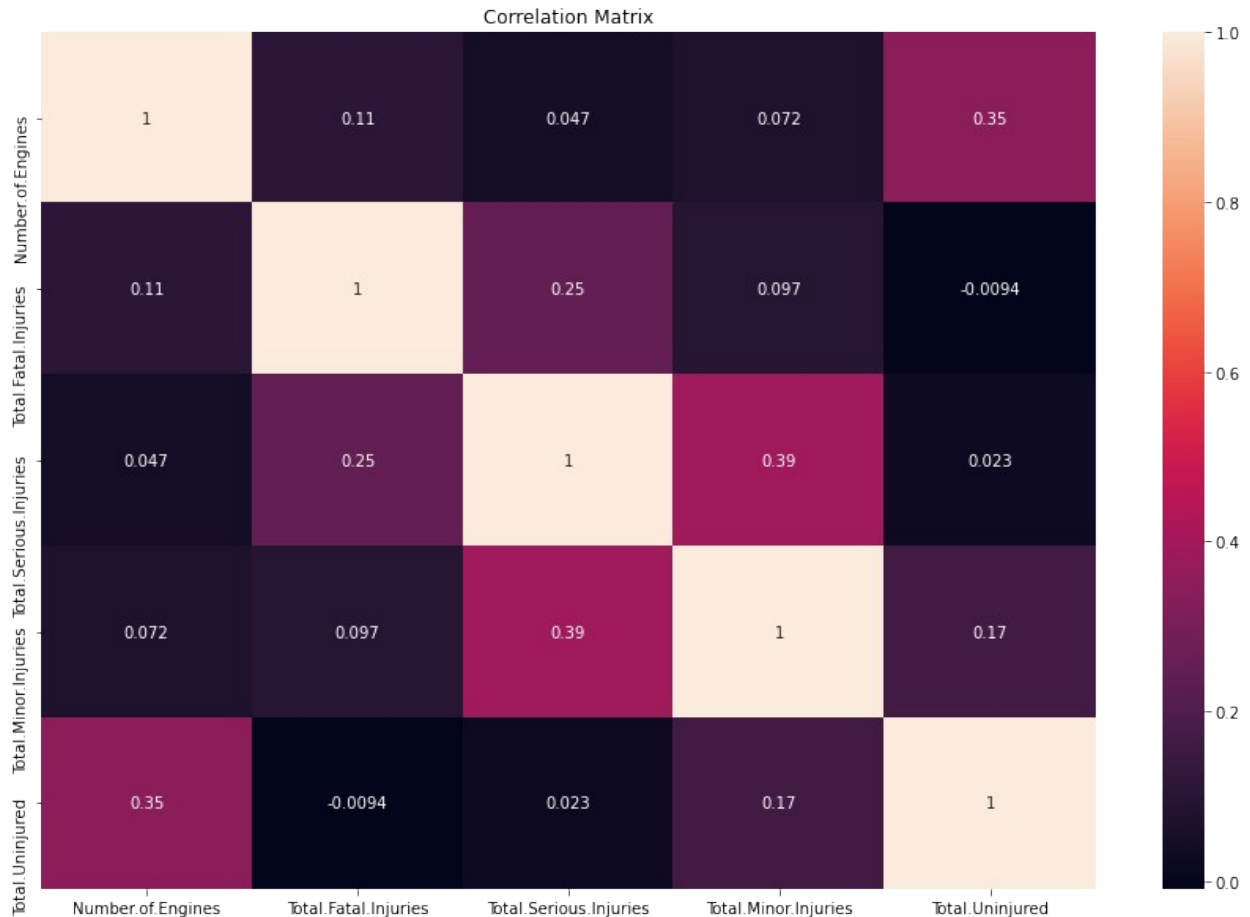
```
matrix=df.corr()

plt.figure(figsize=(15,10))

sns.heatmap(matrix, annot=True)

plt.title('Correlation Matrix')

plt.show()
```



From the above heat map, there is low correlation between the continuous variables, meaning they are not related. All the correlation values are below 0.4

```
cartegorical_col_counts=df[cartegorical_col].nunique()
# checking unique values in columns with cartegorical data
```

```
cartegorical_col_counts
```

```
Investigation.Type      2
Event.Date             8197
Location               14102
Country                 38
Injury.Severity         42
Aircraft.damage         3
Make                   3959
Model                  6937
Amateur.Built           2
Engine.Type             6
Purpose.of.flight       22
Weather.Condition        3
Broad.phase.of.flight   12
Report.Status           2
```

```
Publication.Date          1439
dtype: int64

investigation_count=df['Investigation.Type'].value_counts()

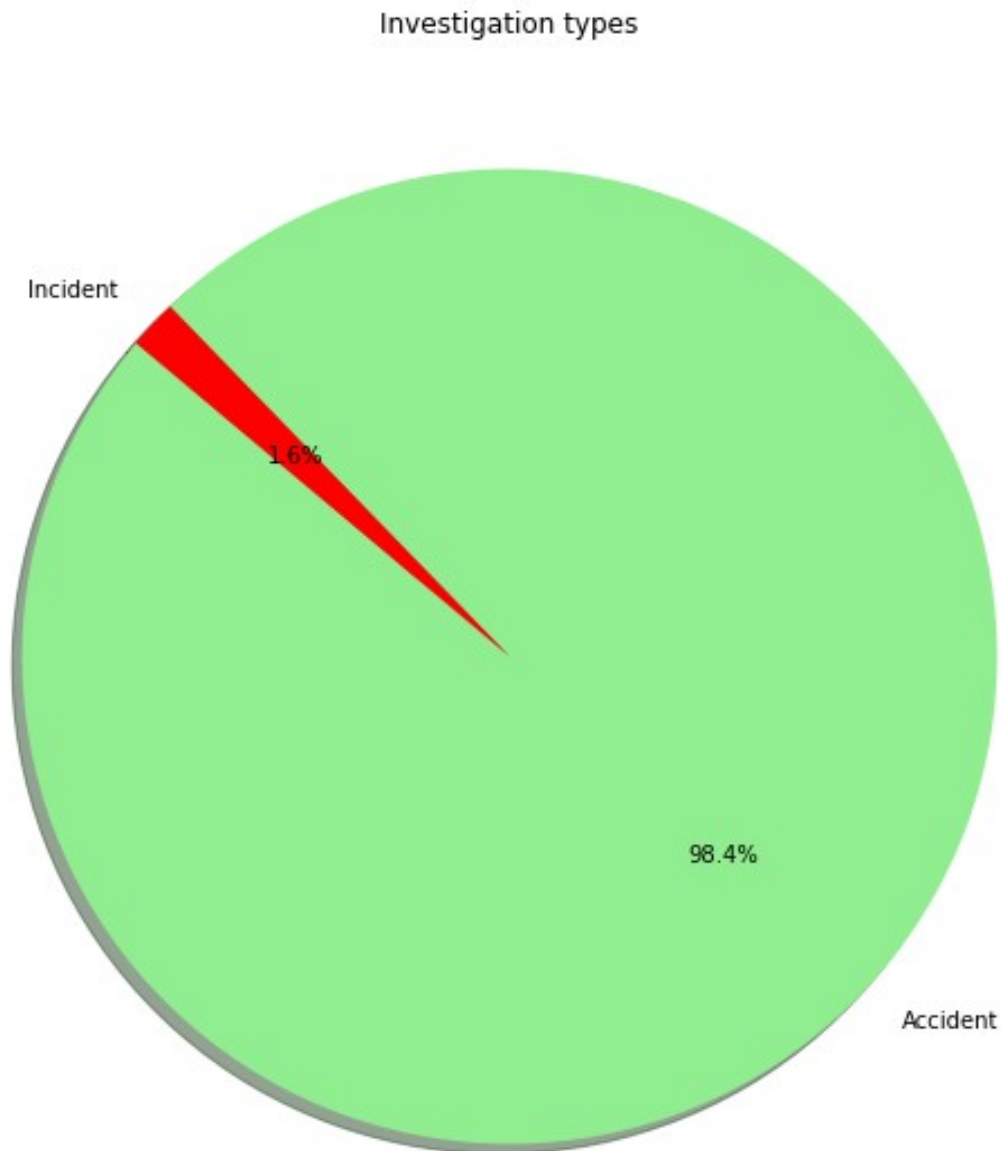
labels=investigation_count.index
values=investigation_count.values
colour=['lightgreen','red']

plt.figure(figsize=(15,10),facecolor='white')

plt.pie(values,labels=labels,colors=colour,autopct='%1.1f%%',shadow=True,startangle=140)

plt.title('Investigation types')

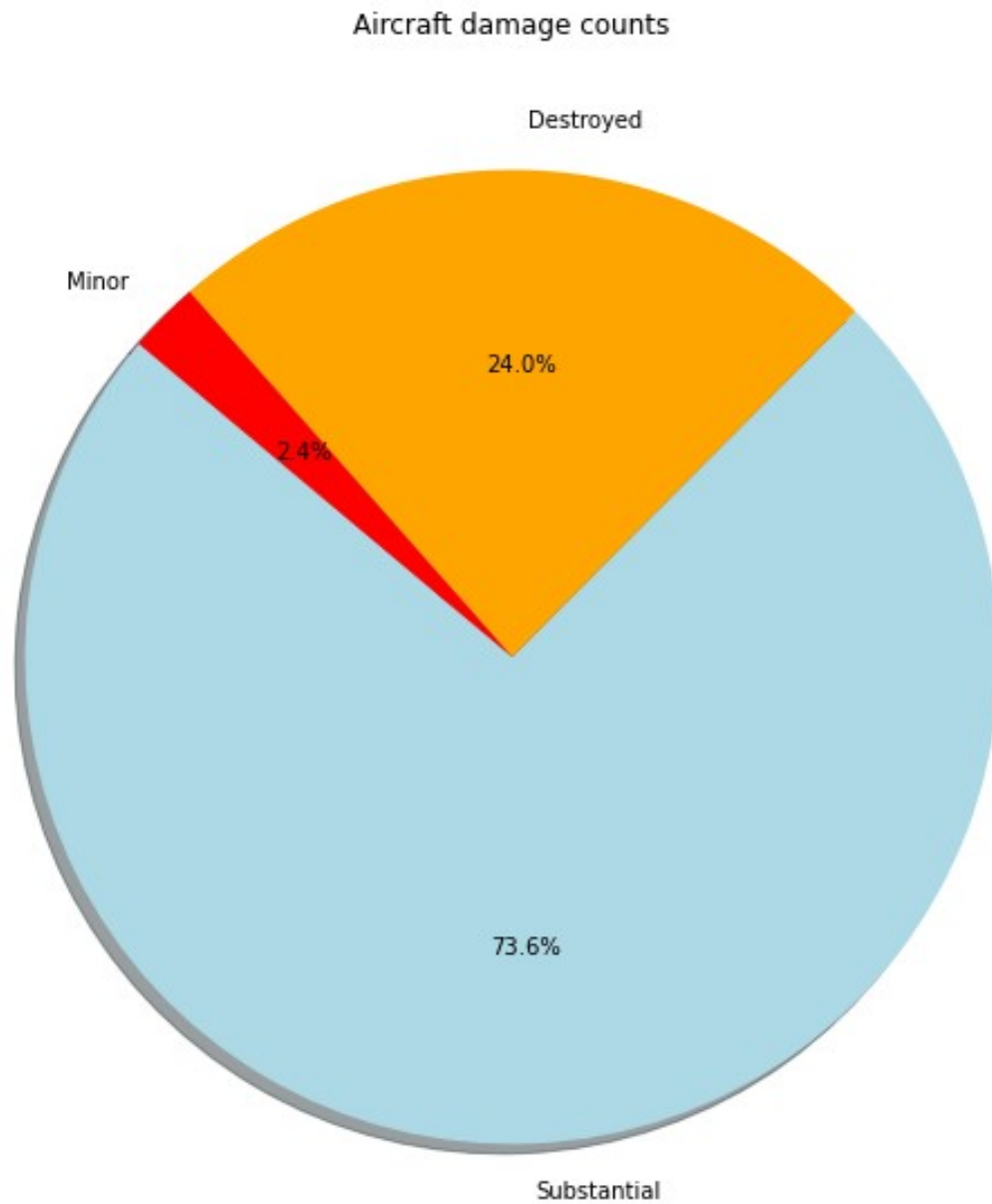
plt.show()
```



The Accidents had the highest investigation at 98.4%, while Incidents had a percentage of 1.6%

```
Aircraft_damage_count=df['Aircraft.damage'].value_counts()  
labels=Aircraft_damage_count.index  
values=Aircraft_damage_count.values  
colour=['lightblue','orange','red']  
  
plt.figure(figsize=(15,10),facecolor='white')  
plt.pie(values,labels=labels,colors=colour,autopct='%1.1f%
```

```
%',shadow=True,startangle=140)  
plt.title('Aircraft damage counts')  
plt.show()
```



In the accident scenario, the accident was highly substantial at 73.6%, while damaged as 24% and minor 2.4%

```
Amateur_Built_count=df['Amateur.Built'].value_counts()
```

```
labels=Amateur_Built_count.index
values=Amateur_Built_count.values
colour=['orange','blue']

plt.figure(figsize=(15,10),facecolor='white')

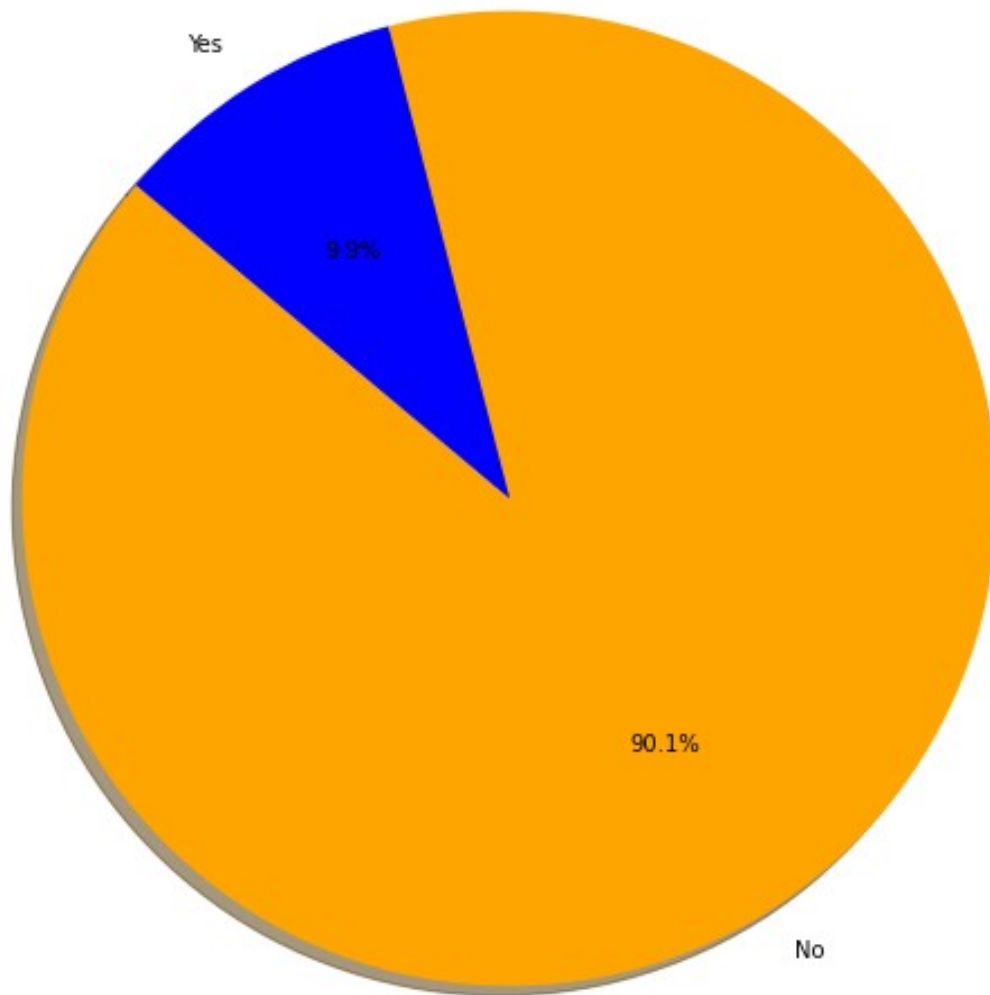
plt.pie(values,labels=labels,colors=colour,autopct='%1.1f%%',shadow=True,startangle=140)

plt.title('Amateur Built counts')

plt.show()
```



Amateur Built counts



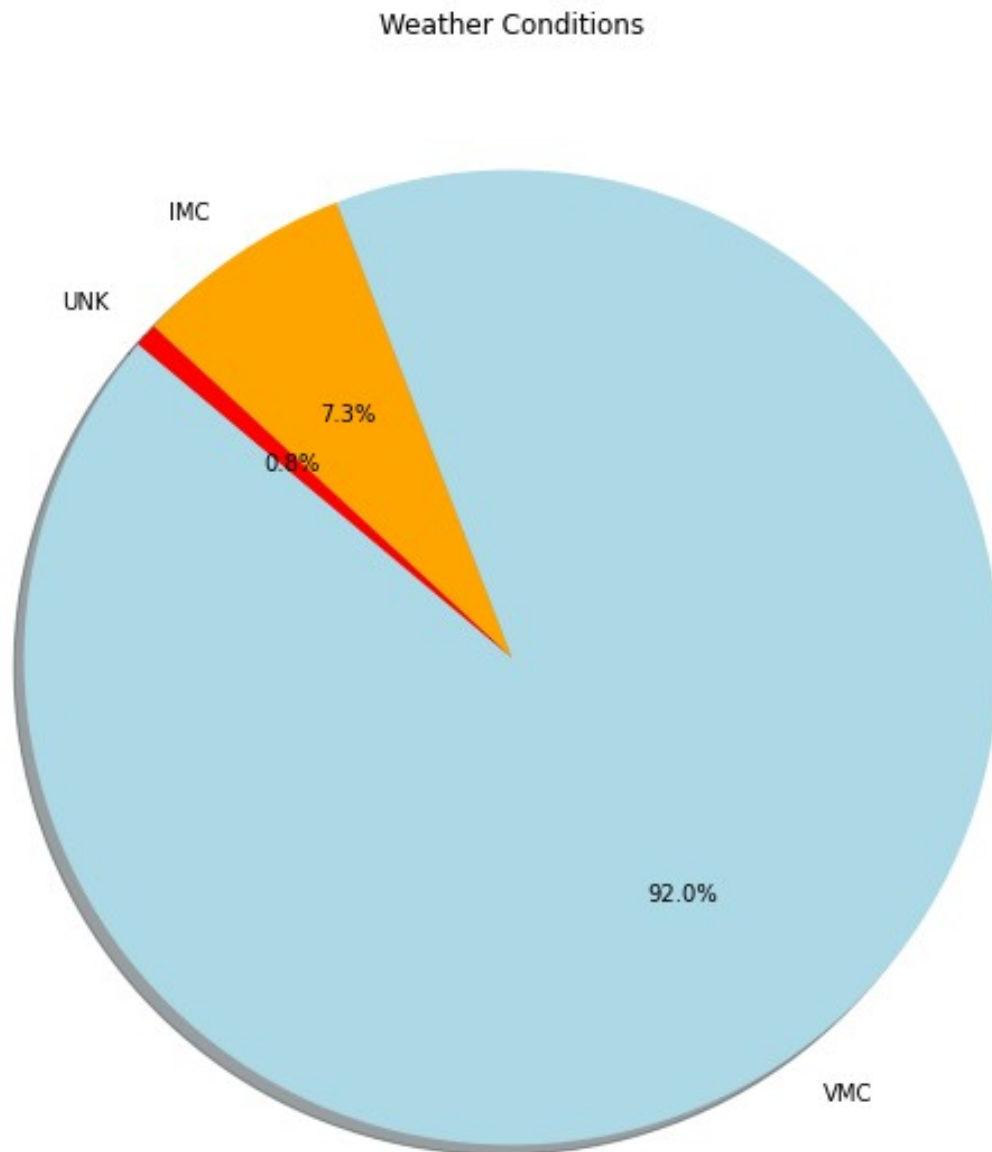
The planes that were not amateur built had highest accidents at 90.1%, while amateur built accidents were at 9.9 %

```
Weather_Condition_count=df['Weather.Condition'].value_counts()  
  
labels=Weather_Condition_count.index  
values=Weather_Condition_count.values  
colour=['lightblue','orange','red']  
  
plt.figure(figsize=(15,10),facecolor='white')
```

```
plt.pie(values,labels=labels,colors=colour,autopct='%1.1f%%',shadow=True,startangle=140)

plt.title('Weather Conditions')

plt.show()
```



when it comes to weather conditions VMC(Visual Metrological Condition this are weather conditions with sufficient visibilty and cloud clearance for pilot fly) at 92% the accidents were highest,with the IMC(Intsrument Metrological conditions,this are the weather condition below the minimum criteria for flying by visual reference requiring pilots to use instrument),here the

accident percentage was 7.3%,lastly the UNK(Unkown weather indicating information is unavailable or not determined,requiring critical safety concern)the accident percentage was 0.8%

```
df['Engine.Type'].value_counts()

Reciprocating      41499
Turbo Shaft        1808
Turbo Prop         1625
Unknown            797
Turbo Fan          661
Turbo Jet          292
Name: Engine.Type, dtype: int64

df=df.drop(df[df['Engine.Type'] == 'Unknown'].index) # deleting
unknown cartegory

engine_value_counts=df['Engine.Type'].value_counts()
x=engine_value_counts.index
y=engine_value_counts.values

plt.figure(figsize=(15,10))

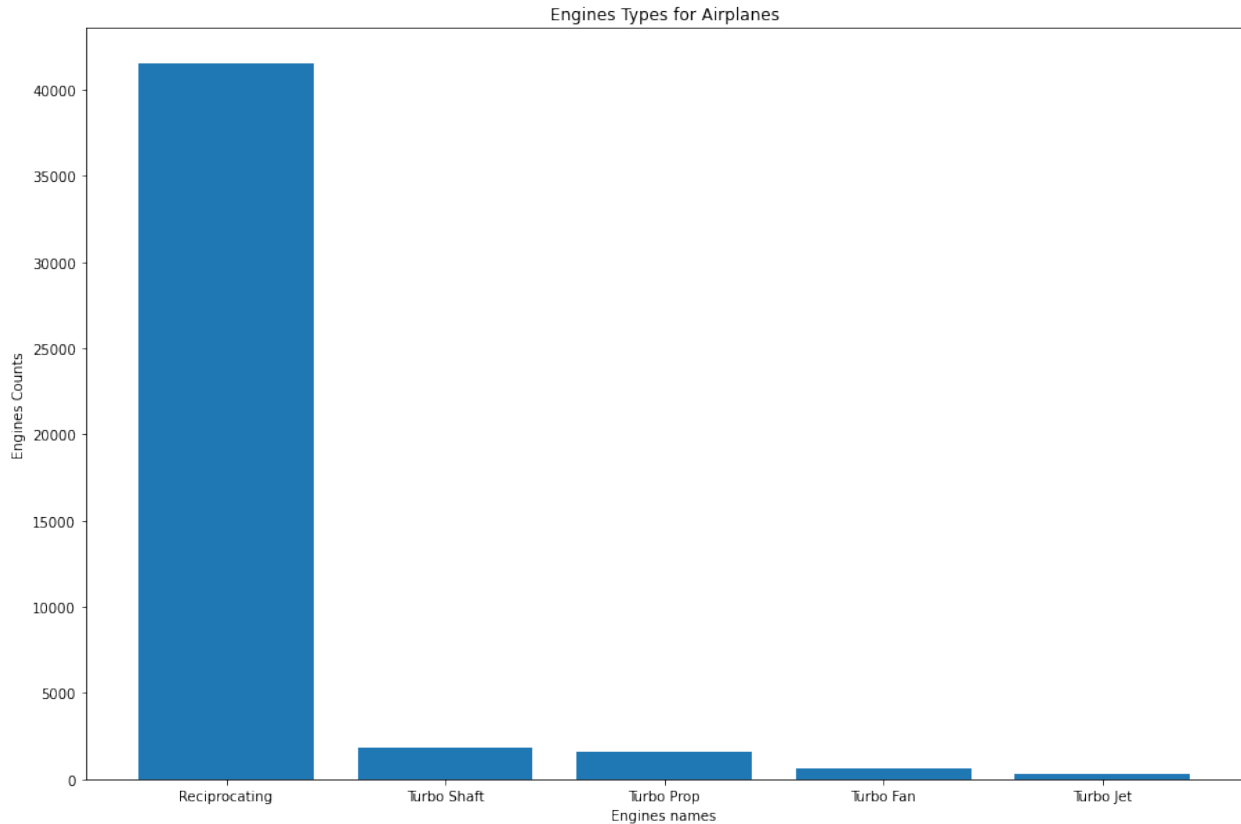
plt.bar(x,y)

plt.title('Engines Types for Airplanes')

plt.xlabel('Engines names')

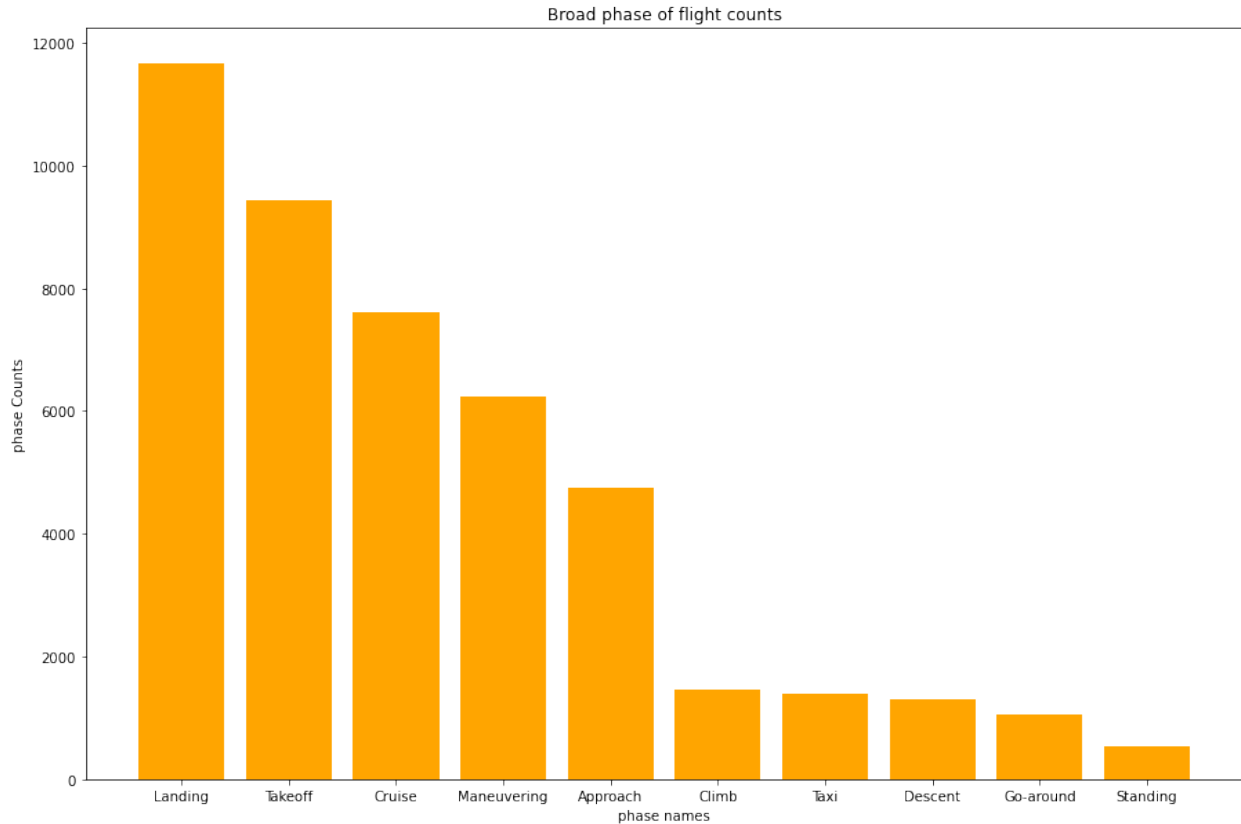
plt.ylabel('Engines Counts')

plt.show()
```



The aircraft with reciprocating Engine type contributed to the highest number of accident ,which was significantly higher than all other engine types.

```
df = df.drop(df[df['Broad.phase.of.flight'].isin(['Unknown',  
'Other'])].index)  
# we are dropping Unkown and Other cartegories  
  
Broad_phase_of_flight_counts=df['Broad.phase.of.flight'].value_counts(  
)  
x=Broad_phase_of_flight_counts.index  
y=Broad_phase_of_flight_counts.values  
  
plt.figure(figsize=(15,10))  
plt.bar(x,y,color=['orange'])  
plt.title('Broad phase of flight counts')  
plt.xlabel('phase names')  
plt.ylabel('phase Counts')  
plt.show()
```



Most accidents took place while landing, takeoff, maneuvering and while approaching.

```
df['Event.Date']=pd.to_datetime(df['Event.Date']) # converting to
datetime from text
df['Event.Date'].dt.month #converting to the months
```

|       |    |
|-------|----|
| 2     | 8  |
| 3     | 6  |
| 7     | 1  |
| 8     | 1  |
| 9     | 1  |
| ...   |    |
| 63906 | 12 |
| 63908 | 12 |
| 63909 | 12 |
| 63910 | 12 |
| 63911 | 12 |

```
Name: Event.Date, Length: 45442, dtype: int64

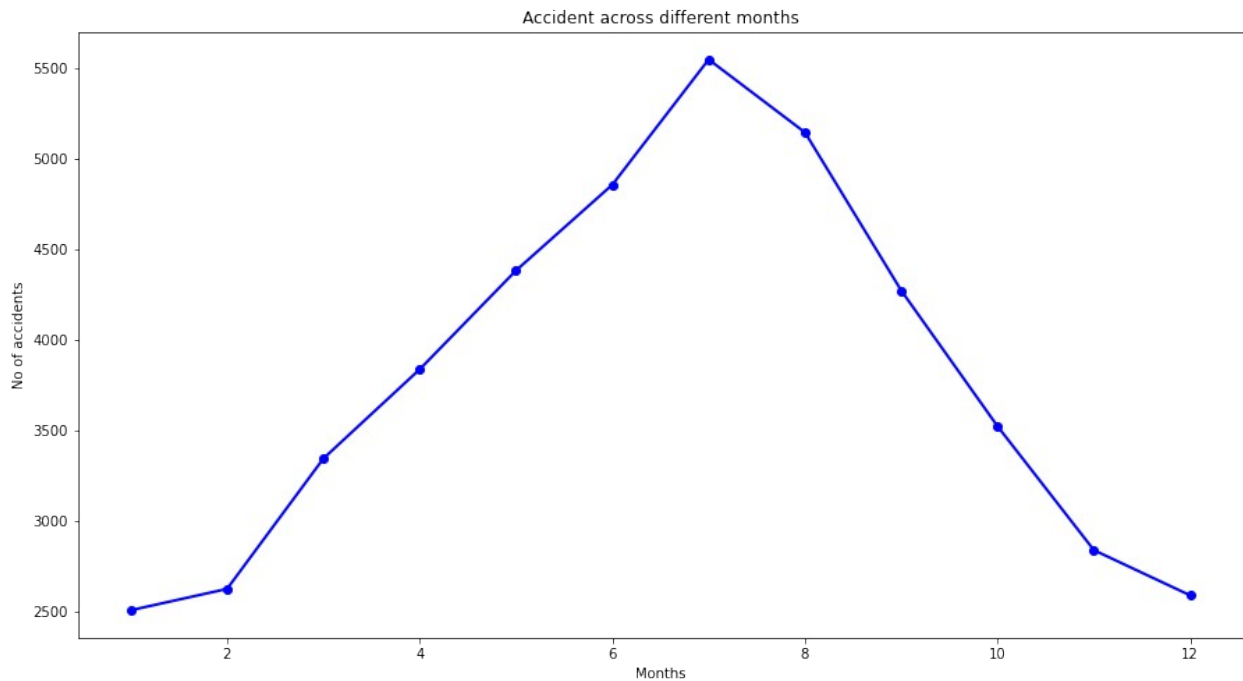
monthly_accidents=df['Event.Date'].dt.month.value_counts().sort_index(
)

x=monthly_accidents.index
y=monthly_accidents.values
```

```
plt.figure(figsize=(15,8))

plt.plot(x,y, marker='o',color='blue',linewidth=2)
plt.title('Accident across different months')
plt.xlabel('Months')
plt.ylabel('No of accidents')

plt.show()
```



Most accidents occurred between April and October, with July having the most.

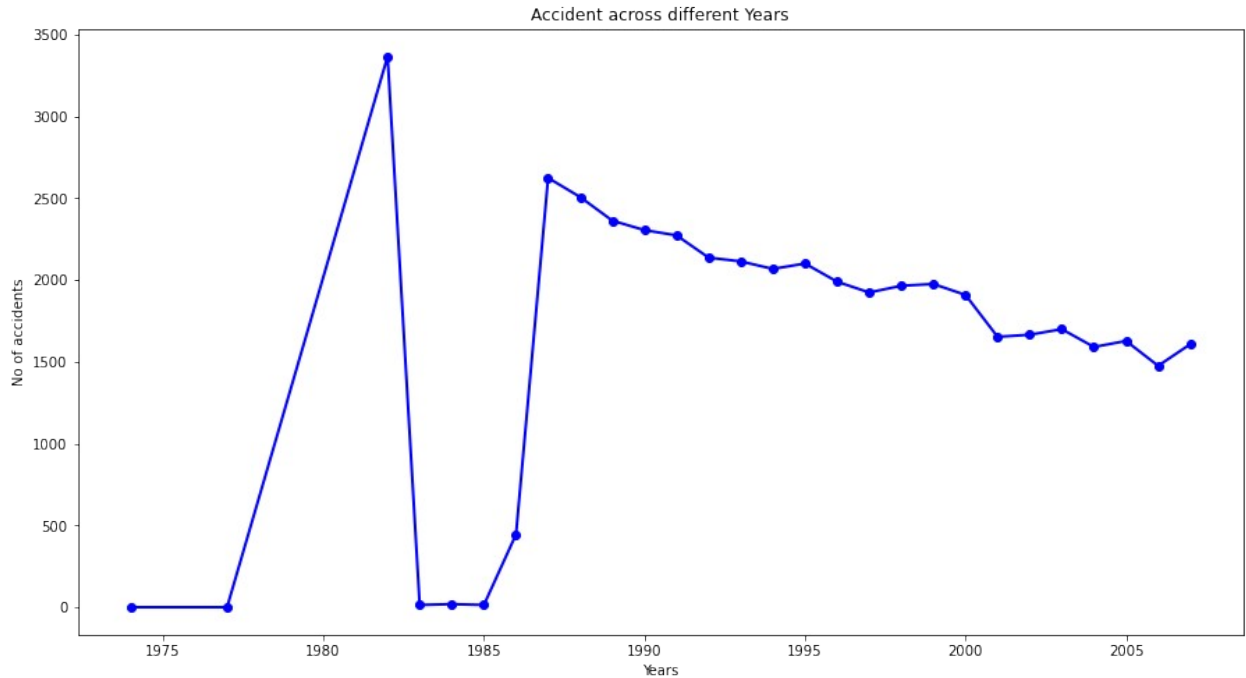
```
yearly_accidents=df['Event.Date'].dt.year.value_counts().sort_index()

x=yearly_accidents.index
y=yearly_accidents.values

plt.figure(figsize=(15,8))

plt.plot(x,y, marker='o',color='blue',linewidth=2)
plt.title('Accident across different Years')
plt.xlabel('Years')
plt.ylabel('No of accidents')

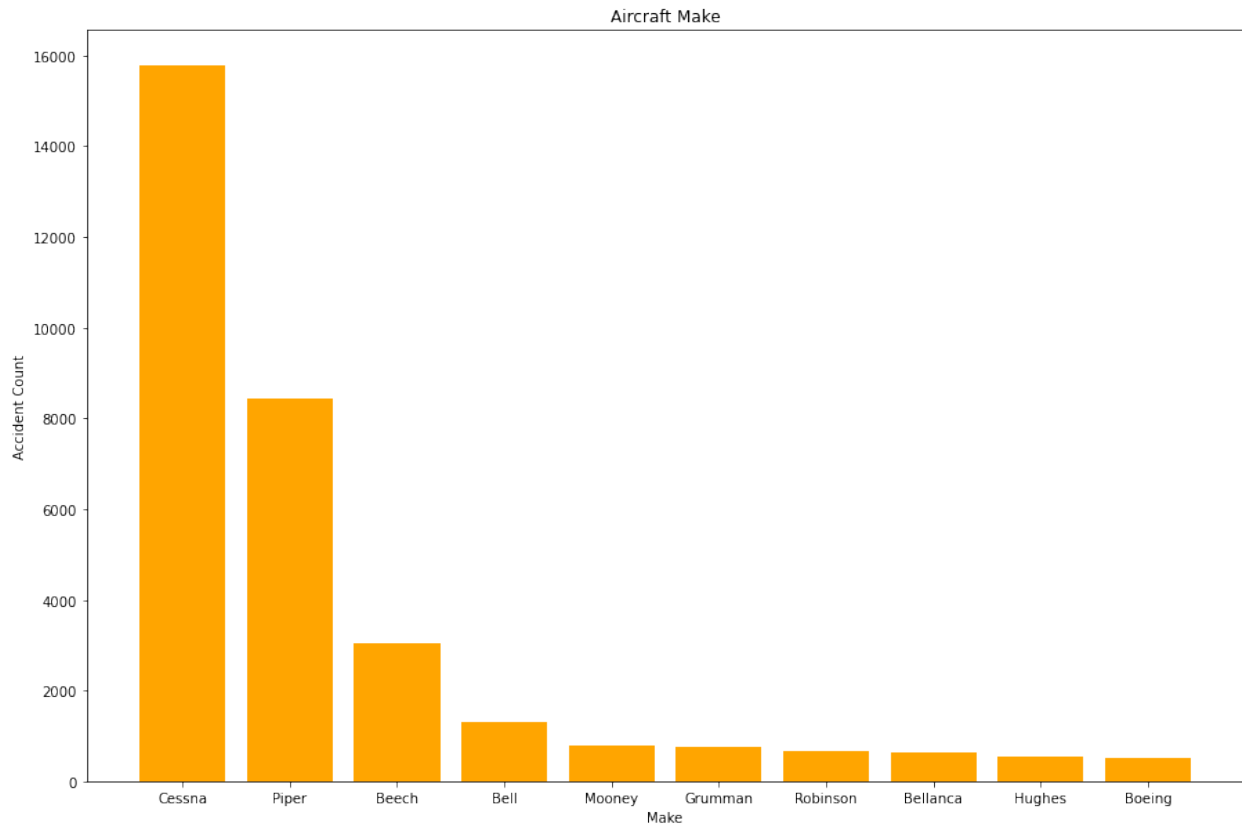
plt.show()
```



Over the years the number of accidents the accidents have declined

```
Make_counts=df['Make'].value_counts()
top_10=Make_counts.head(10)

plt.figure(figsize=(15,10))
plt.bar(top_10.index,top_10.values,color=['orange'])
plt.title('Aircraft Make')
plt.xlabel('Make')
plt.ylabel('Accident Count')
plt.show()
```

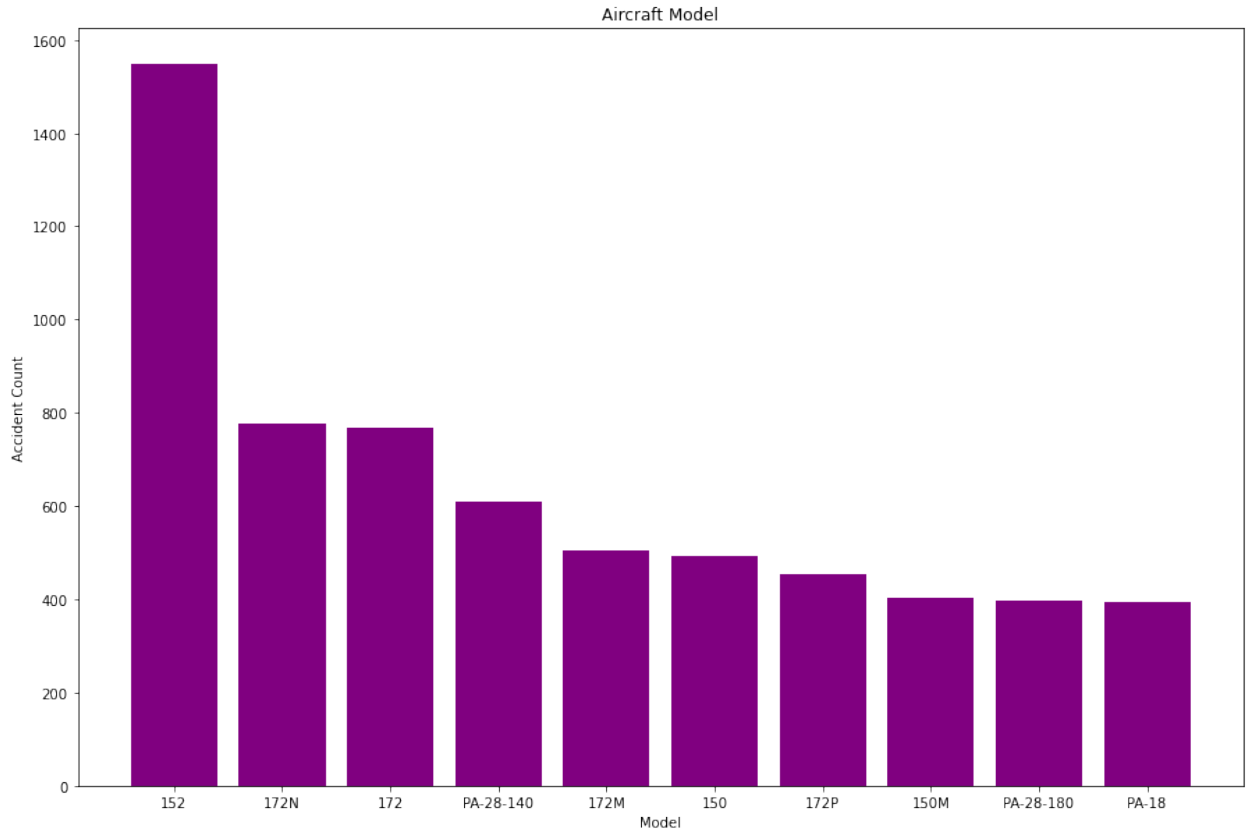


Cessna,Piper,Beech and Bell had the highest number of recorded accidents.

```
Model_counts=df['Model'].value_counts()
top_10=Model_counts.head(10)

plt.figure(figsize=(15,10))
plt.bar(top_10.index,top_10.values,color=['purple'])
plt.title('Aircraft Model')
plt.xlabel('Model')
plt.ylabel('Accident Count')
plt.show()
```





The model that had the highest accidents was model-152, while models such as 150M, PA-28-180, PA-18 had the least number of accidents among the top ten aircraft models

```
df=df.drop(df[df['Purpose.of.flight'] == 'Unknown'].index)# dropping
unknown cartegory

purpose_counts=df['Purpose.of.flight'].value_counts()

plt.figure(figsize=(15,10))

plt.bar(purpose_counts.index,purpose_counts.values,color=['pink'])

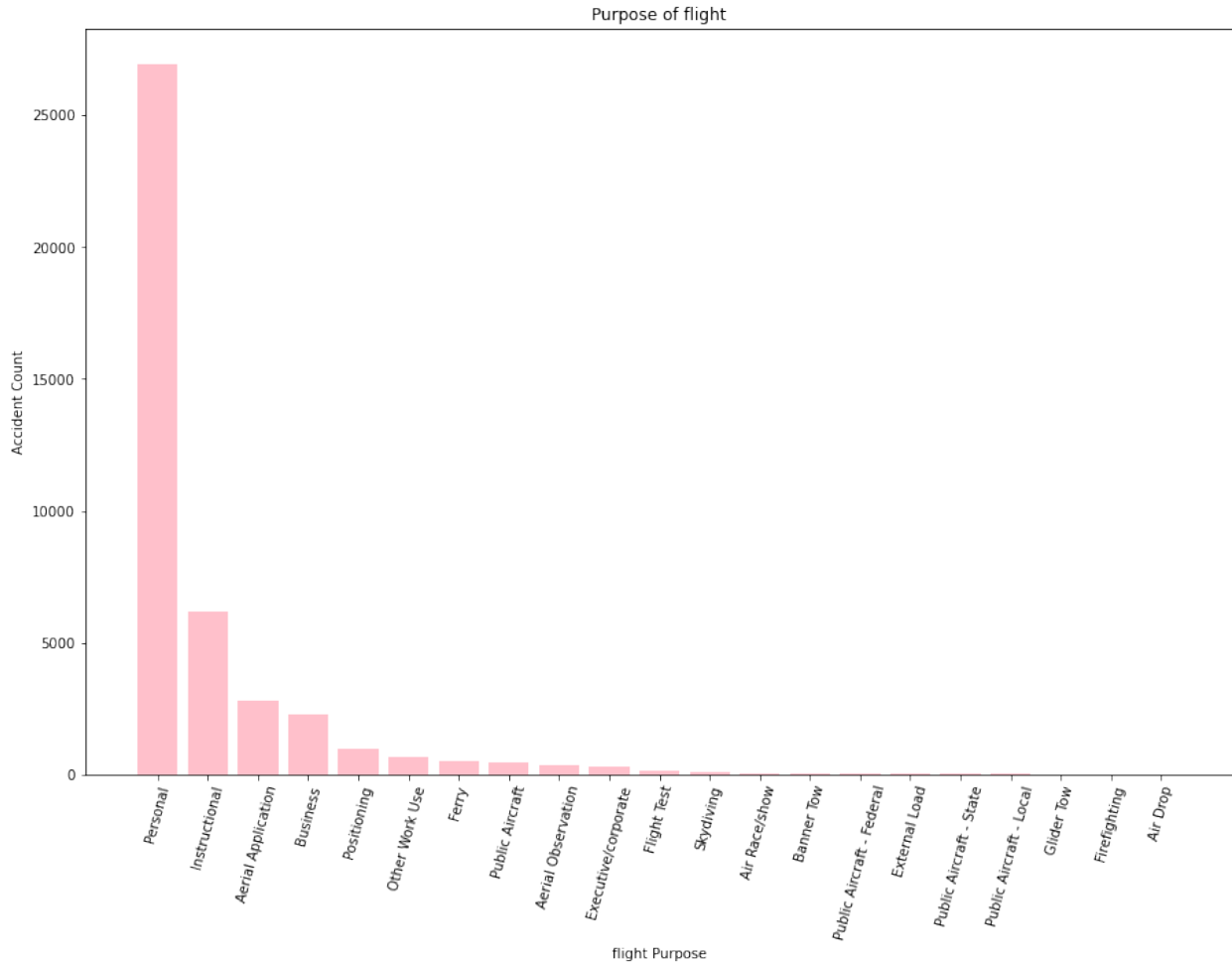
plt.title('Purpose of flight')

plt.xlabel('flight Purpose')

plt.xticks(rotation=75)

plt.ylabel('Accident Count')

plt.show()
```



Personal flights has a significant number of accidents, followed by instructional and aerial Application.

```
# df.to_csv("cleaned_data.csv", index=False)
df.head()
```

|   | Investigation.Type | Event.Date | Location         | Country \     |
|---|--------------------|------------|------------------|---------------|
| 2 | Accident           | 1974-08-30 | Saltville, VA    | United States |
| 3 | Accident           | 1977-06-19 | EUREKA, CA       | United States |
| 7 | Accident           | 1982-01-01 | PULLMAN, WA      | United States |
| 8 | Accident           | 1982-01-01 | EAST HANOVER, NJ | United States |
| 9 | Accident           | 1982-01-01 | JACKSONVILLE, FL | United States |

|                 | Injury.Severity | Aircraft.damage | Make     | Model |
|-----------------|-----------------|-----------------|----------|-------|
| Amateur.Built \ |                 |                 |          |       |
| 2               | Fatal(3)        | Destroyed       | Cessna   | 172M  |
| No              |                 |                 |          |       |
| 3               | Fatal(2)        | Destroyed       | Rockwell | 112   |
| No              |                 |                 |          |       |
| 7               | Non-Fatal       | Substantial     | Cessna   | 140   |

|                      |                        |                       |                   |              |
|----------------------|------------------------|-----------------------|-------------------|--------------|
| No                   |                        |                       |                   |              |
| 8                    | Non-Fatal              | Substantial           | Cessna            | 401B         |
| No                   |                        |                       |                   |              |
| 9                    | Non-Fatal              | Substantial           | North American    | NAVION L-17B |
| No                   |                        |                       |                   |              |
|                      | Number.of.Engines      | Engine.Type           | Purpose.of.flight |              |
| Total.Fatal.Injuries | \                      |                       |                   |              |
| 2                    | 1.0                    | Reciprocating         | Personal          |              |
| 3.0                  |                        |                       |                   |              |
| 3                    | 1.0                    | Reciprocating         | Personal          |              |
| 2.0                  |                        |                       |                   |              |
| 7                    | 1.0                    | Reciprocating         | Personal          |              |
| 0.0                  |                        |                       |                   |              |
| 8                    | 2.0                    | Reciprocating         | Business          |              |
| 0.0                  |                        |                       |                   |              |
| 9                    | 1.0                    | Reciprocating         | Personal          |              |
| 0.0                  |                        |                       |                   |              |
|                      | Total.Serious.Injuries | Total.Minor.Injuries  | Total.Uninjured   | \            |
| 2                    | 0.279881               | 0.357061              | 5.32544           |              |
| 3                    | 0.000000               | 0.000000              | 0.000000          |              |
| 7                    | 0.000000               | 0.000000              | 2.000000          |              |
| 8                    | 0.000000               | 0.000000              | 2.000000          |              |
| 9                    | 0.000000               | 3.000000              | 0.000000          |              |
|                      |                        |                       |                   |              |
|                      | Weather.Condition      | Broad.phase.of.flight | Report.Status     |              |
| Publication.Date     |                        |                       |                   |              |
| 2                    | IMC                    | Cruise                | Probable Cause    | 26-02-       |
| 2007                 |                        |                       |                   |              |
| 3                    | IMC                    | Cruise                | Probable Cause    | 12-09-       |
| 2000                 |                        |                       |                   |              |
| 7                    | VMC                    | Takeoff               | Probable Cause    | 01-01-       |
| 1982                 |                        |                       |                   |              |
| 8                    | IMC                    | Landing               | Probable Cause    | 01-01-       |
| 1982                 |                        |                       |                   |              |
| 9                    | IMC                    | Cruise                | Probable Cause    | 01-01-       |
| 1982                 |                        |                       |                   |              |

## Recommendations

1. Based on the Airplane make i recommend Boeing,Hughes,Bellanca,Robinson,Grumman,Mooney,for purchase as they have lower number of accidents,compared to models like cessna,that had a very high number of accidents.
2. Based on the Airplane Model,i recommend,the models: 150M, PA-28-180, PA-18 as they had the least number of accident in a sapmpe of the top ten models with the highest accidents.

3. Based on the phase of flight,most accidents occurred during,Landing,takeoff,maneuvering,cruise and approach,thus i recommend regular training for pilots to mitigate this accidents.
4. Based on the weather conditions,most accidents occurred when the weather was VMC(Visual Metrological Condition this are weather conditions with sufficient visibilty and cloud clearance for pilot fly),which is basically good weather to fly under,I recommend that pilots are cautioned to be extra careful at all times.