AVIATION BUSINESS

The Main Objective of this Jupyter Notebook is to find an aircraft which is the Lowest Risk for the company to start the business.

THE FIRST STEP

The First thing i did was to import all relevant libraries into the jupyter notebook such as Matplotlib,Pandas and numpy.I imported the abstract base class Number from the numbers module using the method "from numbers import Number".import warnings warnings.filterwarnings('ignore') is a way to suppress all warnings.

THE SECOND STEP

The second thing i needed to do was to read a CSV file in Python using the pandas library, you can use the pd.read_csv() function. encoding='latin-1' tells Python to interpret the file you're reading or writing using the Latin-1 (ISO-8859-1) character encoding.low_memory=False ensures accurate type inference when reading large CSV files.

Third Step

This method shows the rows and columns of the dataset.

The Fourth Step

df.dropna method removes rows and columns that contains NULL values.

The Fifth Step

for the variable aircraft_model I used Value_Counts() to determnine the Aircraft Category frequency distribution of elements within the the series of the dataframe.

The Sixth Step

For the variable aircraft_type I used Value_Counts() to determnine the Make frequency distribution of elements within the the series of the dataframe.

The Seventh Step

For the variable yearly_accidents I used Value_Counts() to determnine the Event Date frequency distribution of elements within the the series of the dataframe.

Eight Step

I have Constructed a Bar chart to Show the Top 5 Aircraft Categories in Aviation accidents in which Airplane's, Helicopter's, Glider's, Ballon's and Gyrocraft's were the most commonly used. I named a variable named data in which i created a dictionary that contained the Aircraft category and the category number's in which inside i created a list that contains the aircrafts and the number values respectively.I then used plt.subplots() function to specify the size of the figure in inches. I used to the variable name aircraft_count in which i used .plot function to specify the kind of chart i want, axes and the color of the chart i want. plt.title is to specify the title i want for the chart which is 'Top 5 Aircraft Categories in Aviation Accidents'. plt.xlabel is the x axis title which is aircraft category. plt.ylabel is the y axis title which is the category numbers. plt.tight_layout() is a method used to automatically adjust the spacing between subplots to prevent overlapping of labels, titles, or other elements in a figure. It ensures that all elements fit neatly within the figure boundaries. plt show() function is used to display a plot created.

Ninth Step

I have Constructed a line chart to Show the Top 7 most used Aircrafts Make in which Cessna','Piper','Beech','Bell','Boeing','Grumman','Mooney were the most commonly used. plt.plot takes x and y coordinates as arguments to create a line plot connecting these points . plt.title is to specify the title i want for the chart which is 'Top 7 Most Used Aircrafts Make'. plt.xlabel is the x axis title which is make. plt.ylabel is the y axis title which is No_of_Aircraft_make. plt show() function is used to display a plot created.

The last Step

I have constructed a scatter plot which focuses on the plane crashes over a year in which focuses on the Event_dates and No_of_accidents. plt.subplots is used to create the scatter plot figure and one or more subplot axes in a single call. df.sort_values is is a method used to sort a DataFrame by one or more columns in this instance it is the event.date. plt.scatter is used to create a scatter plot, which displays individual data points on a two-dimensional plane plt.title is to specify the title i want for the chart which is 'Plane Crashes Over a Year'. plt.xlabel() is the x axis title which is Event_Date. plt.ylabel() is the y axis title which is Plane Crashes Over a Year. plt.legend is used to display a legend. A legend is the label box that explains what each line, point, or bar represents in the graph — useful when plotting multiple data series. plt show() function is used to display a plot created.

In [49]:

```
#Import all relevant Libraries
import pandas as pd
import numpy as np
from numbers import Number
import warnings
warnings.filterwarnings('ignore')
import matplotlib
import matplotlib.pyplot as plt
```

In [235…

```
#Read the relevant Csv File
df = pd.read_csv('./AviationData.csv', encoding='latin1', low_memory=False)
df.head()
```

Out[235…

	Event.Id	Investigation.Type	Accident.Number	Event.Date	Loca
0	20001218X45444	Accident	SEA87LA080	24/10/1948	MC CREE
1	20001218X45447	Accident	LAX94LA336	19/07/1962	BRIDGEP
2	20061025X01555	Accident	NYC07LA005	30/08/1974	Saltvill
3	20001218X45448	Accident	LAX96LA321	19/06/1977	EUREK
4	20041105X01764	Accident	CHI79FA064	02/08/1979	Cantor

5 rows × 31 columns

In [33]:

```
#Reveal Rows and Columns of the Dataset
print("Dataset shape", df.shape)
df.head()
```

```
Dataset shape (88889, 31)
```

Out[33]:

	Event.Id	Investigation.Type	Accident.Number	Event.Date	Loca
0	20001218X45444	Accident	SEA87LA080	24/10/1948	MC CREE
1	20001218X45447	Accident	LAX94LA336	19/07/1962	BRIDGEP
2	20061025X01555	Accident	NYC07LA005	30/08/1974	Saltvill
3	20001218X45448	Accident	LAX96LA321	19/06/1977	EUREK
4	20041105X01764	Accident	CHI79FA064	02/08/1979	Cantor

5 rows × 31 columns

In [24]:

```
#Remove Null Values
new_df=df.dropna(subset=['Latitude','Longitude','Airport.Code','Airport.Name
new_df
```

Out[24]:

	Event.Id	Investigation.Type	Accident.Number	Event.Date	
45404	20001211X11573	Accident	LAX99FA051	1998-12-17	AN
47988	20001212X20664	Incident	LAX00IA127	2000-03-13	F
49192	20001212X21754	Accident	LAX00FA310	2000-08-25	
49969	20010223X00505	Accident	ANC01LA031	2001-01-19	KOI

4 rows × 31 columns

In [225…

```
#Value_Counts() to determnine the Aircraft Category frequency distribution c
aircraft_model=df['Aircraft.Category'].value_counts().head(7)
print(aircraft_model)
```

```
Aircraft.Category
Airplane             27617
Helicopter            3440
Glider                 508
Balloon                231
Gyrocraft              173
Weight-Shift           161
Powered Parachute       91
Name: count, dtype: int64
```

In [237…

```
#Value_Counts() to determnine the Aircraft Category frequency distribution c
Aircraft_type=df['Make'].value_counts().head(10)
print(Aircraft_type)
```

```
Make
Cessna     22227
Piper      12029
CESSNA      4922
Beech       4330
PIPER       2841
Bell        2134
Boeing      1594
BOEING      1151
Grumman     1094
Mooney      1092
Name: count, dtype: int64
```

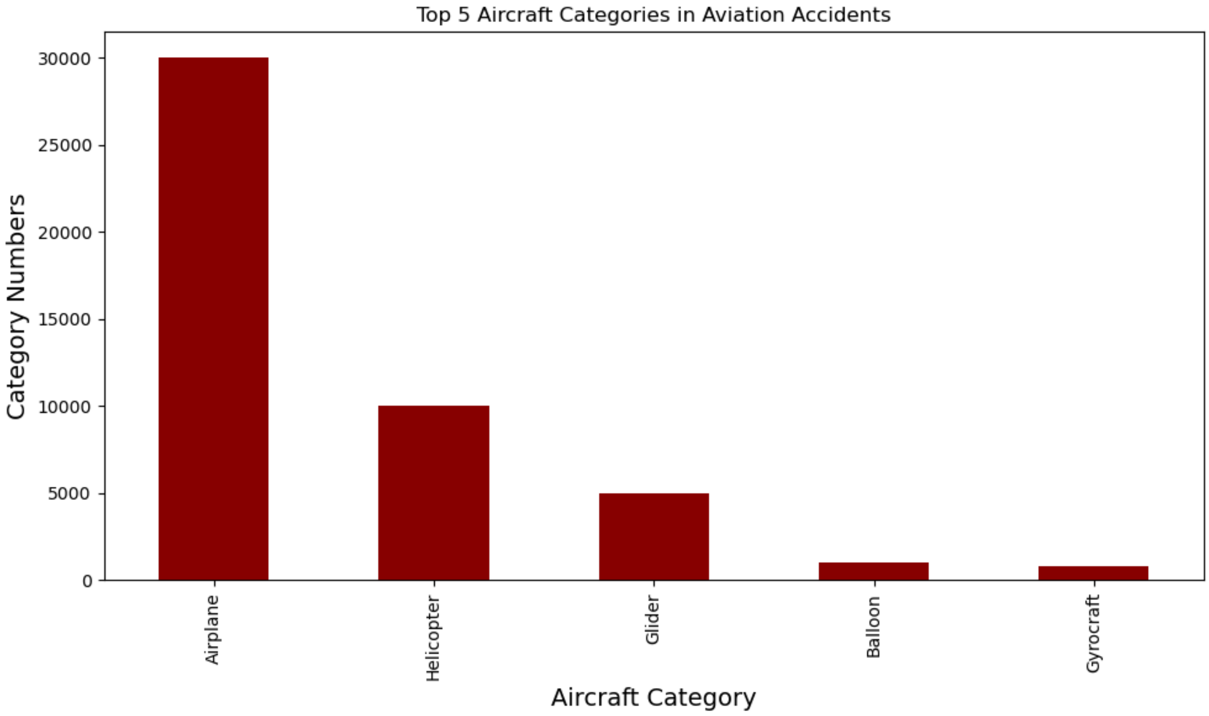
In [228…

```
#Value_Counts() to determnine the Aircraft Category frequency distribution c
yearly_accidents= df['Event.Date'].value_counts().head()
print(yearly_accidents)
```

```
Event.Date
30/06/1984    25
16/05/1982    25
08/07/2000    25
05/08/1983    24
25/08/1984    24
17/05/1986    24
05/06/1983    24
Name: count, dtype: int64
```

In [227…

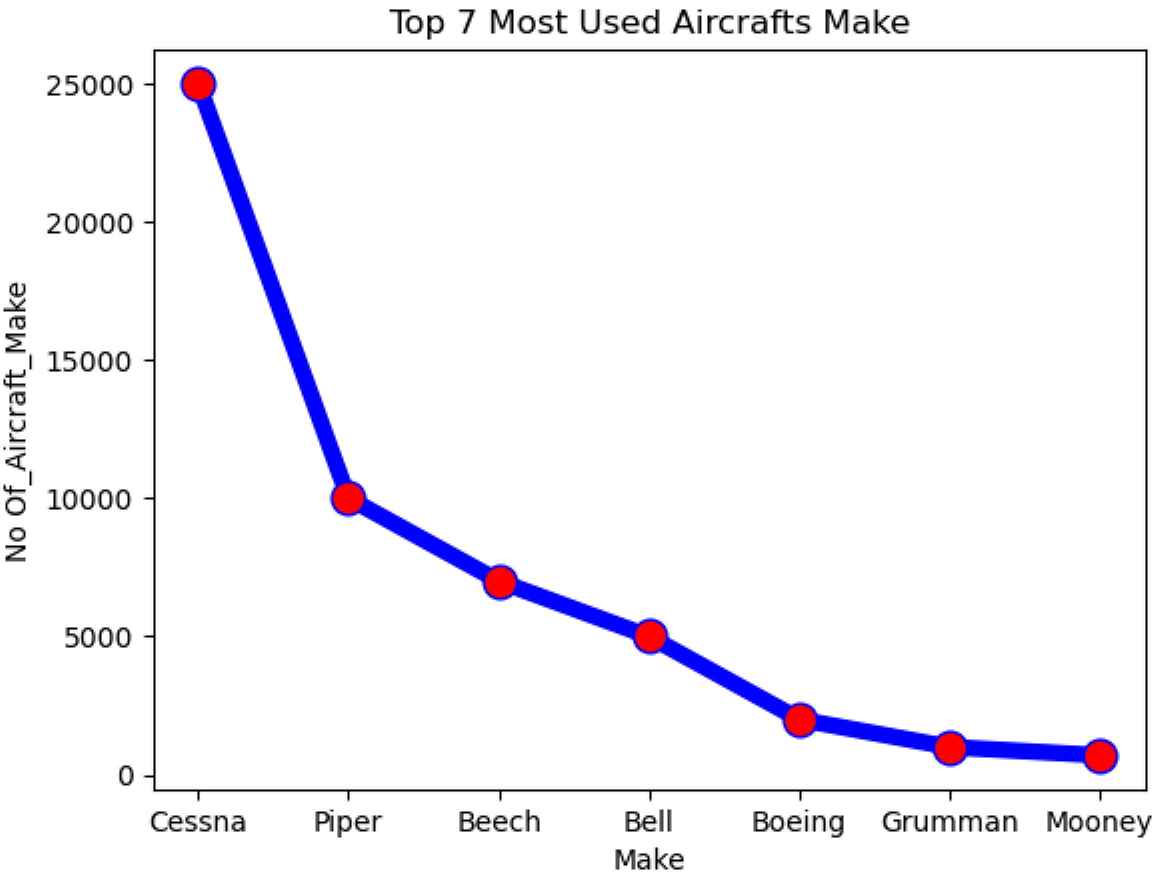
```
#Plot a Bar chart
data={'Aircraft Category': ['Airplane', 'Helicopter', 'Glider', 'Balloon', '
'Category Numbers': [30000,10000,5000,1000,800]}
fig, ax = plt.subplots(figsize=(10,6))
aircraft_Count.plot(kind='bar', ax=ax, color='darkred')
plt.title('Top 5 Aircraft Categories in Aviation Accidents')
plt.xlabel('Aircraft Category',fontsize=14)
plt.ylabel('Category Numbers',fontsize=14)
plt.tight_layout()
plt.show()
```



In [234…

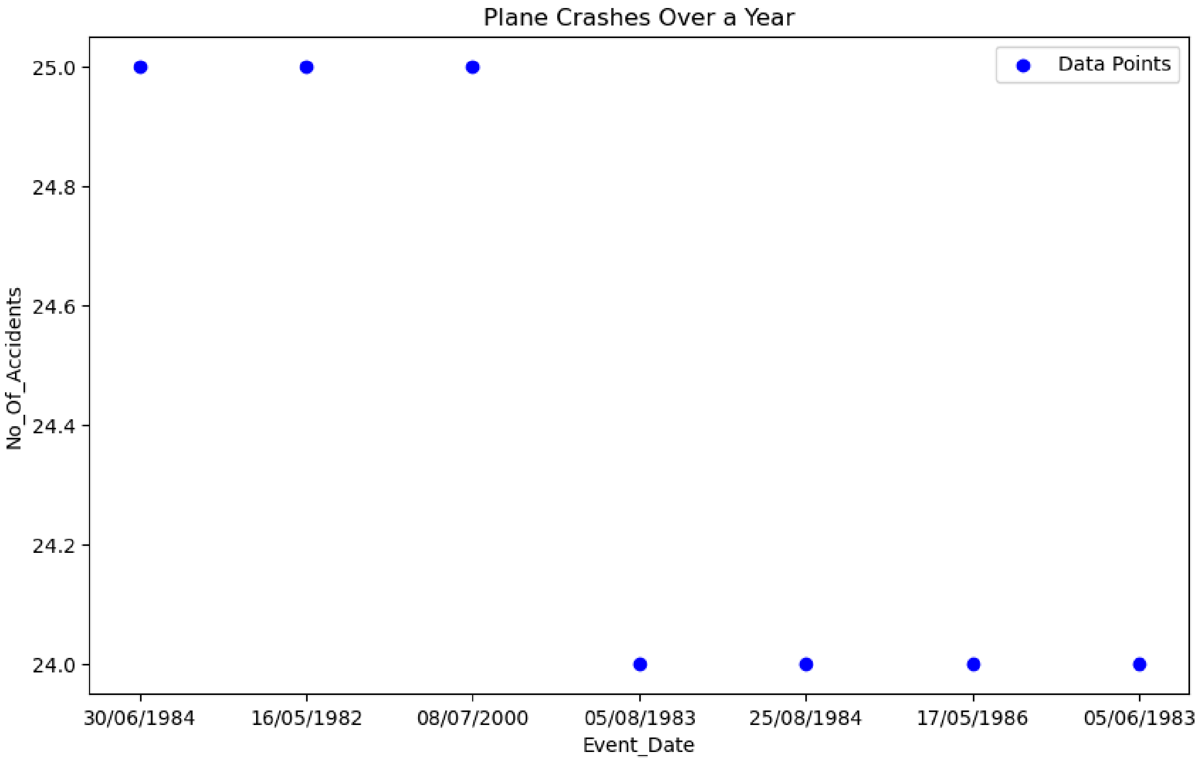
```
#Plot a Line chart
Make=['Cessna','Piper','Beech','Bell','Boeing','Grumman','Mooney',]
No_of_Aircraft_Make=[25000,10000,7000,5000,2000,1000,700]

plt.plot(Make, No_of_Aircraft_Make, marker='o', linewidth=6, color='blue', m
plt.title('Top 7 Most Used Aircrafts Make', fontsize=12)
plt.xlabel('Make')
plt.ylabel('No Of_Aircraft_Make')
plt.show()
```



In [232…

```
#Plot a scatter plot
Event_Date=['30/06/1984','16/05/1982','08/07/2000','05/08/1983','25/08/1984'
No_Of_Accidents=[25,25,25,24,24,24,24]
df_sorted=df.sort_values(by='Event.Date')
Accident_figure, ax = plt.subplots(figsize=(10, 6))
plt.scatter(Event_Date, No_Of_Accidents, color='blue', marker='o', label='Da
plt.xlabel('Event_Date')
plt.ylabel ('No_Of_Accidents')
plt.title('Plane Crashes Over a Year')
plt.legend()
plt.show()
```


Plane Crashes Over a Year
● Data Points
No_Of_Accidents
25.0
24.8
24.6
24.4
24.2
24.0
30/06/1984
16/05/1982
08/07/2000
05/08/1983
25/08/1984
17/05/1986
05/06/1983
Event_Date

```
In [ ]:
```