

In []: BUSINESS UNDERSTANDING

The company **is** planning to venture into the aviation industry by buying **and** operating

Aircraft accidents can lead to major financial losses, legal liabilities, damaged reputations,

This analysis aims to review historical aviation accident data to identify types of aircraft accidents.

PROJECT OBJECTIVES

1. To analyse aircraft accident trends over time.

2. To identify aircraft types **with** the highest accident rates.

3. To evaluate which type of damage level has the most fatalities.

4. To analyze which operators have the highest accidents.

5. To generate insights **and** recommendations that management can use to guide aircraft safety policies.

DATA UNDERSTANDING

In [1]: *#importing necessary Libraries*

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

In [2]: *#Loading dataset*

```
df = pd.read_csv('flight.csv')
df
```

Out[2]:

	Unnamed: 0	acc.date	type	reg	operator	fat	location	dmg
0	0	3 Jan 2022	British Aerospace 4121 Jetstream 41	ZS-NRJ	SA Airlink	0	near Venetia Mine Airport	sub
1	1	4 Jan 2022	British Aerospace 3101 Jetstream 31	HR-AYY	LANHSA - Línea Aérea Nacional de Honduras S.A	0	Roaatán-Juan Manuel Gálvez International Airpor...	sub
2	2	5 Jan 2022	Boeing 737-4H6	EP-CAP	Caspian Airlines	0	Isfahan- Shahid Beheshti Airport (IFN)	sub
3	3	8 Jan 2022	Tupolev Tu-204- 100C	RA- 64032	Cainiao, opb Aviastar-TU	0	Hangzhou Xiaoshan International Airport (GHG)	w/o
4	4	12 Jan 2022	Beechcraft 200 Super King Air	NaN	private	0	Machakilha, Toledo District, Graham Creek area	w/o
...
2495	1245	20 Dec 2018	Cessna 560 Citation V	N188CW	Chen Aircrafts LLC	4	2 km NE of Atlanta- Fulton County Airport, GA (...)	w/o
2496	1246	22 Dec 2018	PZL-Mielec M28 Skytruck	GNB- 96107	Guardia Nacional Bolivariana de Venezuela - GNBV	0	Kamarata Airport (KTV)	sub
2497	1247	24 Dec 2018	Antonov An-26B	9T-TAB	Air Force of the Democratic Republic of the Congo	0	Beni Airport (BNC)	w/o
2498	1248	31 Dec 2018	Boeing 757-2B7 (WL)	N938UW	American Airlines	0	Charlotte- Douglas International Airport, NC (C...	sub

	Unnamed: 0	acc.date	type	reg	operator	fat	location	dmg
2499	1249	unk. date 2018	Rockwell Sabreliner 80	N337KL	private	0	Eugene Airport, OR (EUG)	sub

2500 rows × 8 columns

In [3]: `# Looking at datatypes and missing values
df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2500 entries, 0 to 2499
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype  
 ---  --          --          --    
 0   Unnamed: 0   2500 non-null   int64  
 1   acc.date    2500 non-null   object  
 2   type        2500 non-null   object  
 3   reg         2408 non-null   object  
 4   operator    2486 non-null   object  
 5   fat         2488 non-null   object  
 6   location    2500 non-null   object  
 7   dmg         2500 non-null   object  
dtypes: int64(1), object(7)
memory usage: 156.4+ KB
```

In [4]: `# checking missing value
df.isna().sum()`

```
Out[4]: Unnamed: 0      0
acc.date      0
type         0
reg          92
operator     14
fat          12
location     0
dmg          0
dtype: int64
```

df.columns

In [5]: `#Checking for rows and columns
df.shape`

Out[5]: (2500, 8)

In [6]: `# checking for duplicates
df.duplicated().value_counts()`

```
Out[6]: False    1250
True     1250
Name: count, dtype: int64
```

DATA PREPARATION

```
In [7]: #Sorting duplicates
df[df.duplicated(keep=False)].sort_values(by='type')
df
```

Out[7]:

	Unnamed: 0	acc.date	type	reg	operator	fat	location	dmg
0	0	3 Jan 2022	British Aerospace 4121 Jetstream 41	ZS-NRJ	SA Airlink	0	near Venetia Mine Airport	sub
1	1	4 Jan 2022	British Aerospace 3101 Jetstream 31	HR-AYY	LANHSA - Línea Aérea Nacional de Honduras S.A	0	Roaatán-Juan Manuel Gálvez International Airpor...	sub
2	2	5 Jan 2022	Boeing 737-4H6	EP-CAP	Caspian Airlines	0	Isfahan- Shahid Beheshti Airport (IFN)	sub
3	3	8 Jan 2022	Tupolev Tu-204- 100C	RA- 64032	Cainiao, opb Aviastar-TU	0	Hangzhou Xiaoshan International Airport (GHG)	w/o
4	4	12 Jan 2022	Beechcraft 200 Super King Air	NaN	private	0	Machakilha, Toledo District, Graham Creek area	w/o
...
2495	1245	20 Dec 2018	Cessna 560 Citation V	N188CW	Chen Aircrafts LLC	4	2 km NE of Atlanta- Fulton County Airport, GA (...)	w/o
2496	1246	22 Dec 2018	PZL-Mielec M28 Skytruck	GNB- 96107	Guardia Nacional Bolivariana de Venezuela - GNBV	0	Kamarata Airport (KTV)	sub
2497	1247	24 Dec 2018	Antonov An-26B	9T-TAB	Air Force of the Democratic Republic of the Congo	0	Beni Airport (BNC)	w/o
2498	1248	31 Dec 2018	Boeing 757-2B7 (WL)	N938UW	American Airlines	0	Charlotte- Douglas International Airport, NC (C...	sub

	Unnamed: 0	acc.date	type	reg	operator	fat	location	dmg
2499	1249	unk. date 2018	Rockwell Sabreliner 80	N337KL	private	0	Eugene Airport, OR (EUG)	sub

2500 rows × 8 columns

```
In [8]: #Dropping duplicates
df = df.drop_duplicates()
df
```

Out[8]:

	Unnamed: 0	acc.date	type	reg	operator	fat	location	dmg
0	0	3 Jan 2022	British Aerospace 4121 Jetstream 41	ZS-NRJ	SA Airlink	0	near Venetia Mine Airport	sub
1	1	4 Jan 2022	British Aerospace 3101 Jetstream 31	HR-AYY	LANHSA - Línea Aérea Nacional de Honduras S.A	0	Roaatán-Juan Manuel Gálvez International Airpor...	sub
2	2	5 Jan 2022	Boeing 737-4H6	EP-CAP	Caspian Airlines	0	Isfahan- Shahid Beheshti Airport (IFN)	sub
3	3	8 Jan 2022	Tupolev Tu-204- 100C	RA- 64032	Cainiao, opb Aviastar-TU	0	Hangzhou Xiaoshan International Airport (GHG)	w/o
4	4	12 Jan 2022	Beechcraft 200 Super King Air	NaN	private	0	Machakilha, Toledo District, Graham Creek area	w/o
...
1245	1245	20 Dec 2018	Cessna 560 Citation V	N188CW	Chen Aircrafts LLC	4	2 km NE of Atlanta- Fulton County Airport, GA (...)	w/o
1246	1246	22 Dec 2018	PZL-Mielec M28 Skytruck	GNB- 96107	Guardia Nacional Bolivariana de Venezuela - GNBV	0	Kamarata Airport (KTV)	sub
1247	1247	24 Dec 2018	Antonov An-26B	9T-TAB	Air Force of the Democratic Republic of the Congo	0	Beni Airport (BNC)	w/o
1248	1248	31 Dec 2018	Boeing 757-2B7 (WL)	N938UW	American Airlines	0	Charlotte- Douglas International Airport, NC (C...	sub

	Unnamed: 0	acc.date	type	reg	operator	fat	location	dmg
1249	1249	unk. date 2018	Rockwell Sabreliner 80	N337KL	private	0	Eugene Airport, OR (EUG)	sub

1250 rows × 8 columns

```
In [9]: #Remove unnecessary index column 'Unnamed: 0'  
if 'Unnamed: 0' in df.columns:  
    df = df.drop(columns=['Unnamed: 0'])  
df
```

Out[9]:		acc.date	type	reg	operator	fat	location	dmg
	0	3 Jan 2022	British Aerospace 4121 Jetstream 41	ZS-NRJ	SA Airlink	0	near Venetia Mine Airport	sub
	1	4 Jan 2022	British Aerospace 3101 Jetstream 31	HR-AYY	LANHSA - Línea Aérea Nacional de Honduras S.A	0	Roatán-Juan Manuel Gálvez International Airport...	sub
	2	5 Jan 2022	Boeing 737-4H6	EP-CAP	Caspian Airlines	0	Isfahan-Shahid Beheshti Airport (IFN)	sub
	3	8 Jan 2022	Tupolev Tu-204-100C	RA-64032	Cainiao, opb Aviastar-TU	0	Hangzhou Xiaoshan International Airport (GHG)	w/o
	4	12 Jan 2022	Beechcraft 200 Super King Air	NaN	private	0	Machakilha, Toledo District, Graham Creek area	w/o

	1245	20 Dec 2018	Cessna 560 Citation V	N188CW	Chen Aircrafts LLC	4	2 km NE of Atlanta-Fulton County Airport, GA (...)	w/o
	1246	22 Dec 2018	PZL-Mielec M28 Skytruck	GNB-96107	Guardia Nacional Bolivariana de Venezuela - GNBV	0	Kamarata Airport (KTV)	sub
	1247	24 Dec 2018	Antonov An-26B	9T-TAB	Air Force of the Democratic Republic of the Congo	0	Beni Airport (BNC)	w/o
	1248	31 Dec 2018	Boeing 757-2B7 (WL)	N938UW	American Airlines	0	Charlotte-Douglas International Airport, NC (C...	sub
	1249	unk. date 2018	Rockwell Sabreliner 80	N337KL	private	0	Eugene Airport, OR (EUG)	sub

1250 rows × 7 columns

In [10]: # Selecting relevant columns

```
rel_columns = ['acc.date', 'type', 'operator', 'fat', 'dmg']
df = df[rel_columns]
df.head()
```

	acc.date	type	operator	fat	dmg
0	3 Jan 2022	British Aerospace 4121 Jetstream 41	SA Airlink	0	sub
1	4 Jan 2022	British Aerospace 3101 Jetstream 31	LANHSA - Línea Aérea Nacional de Honduras S.A	0	sub
2	5 Jan 2022	Boeing 737-4H6	Caspian Airlines	0	sub
3	8 Jan 2022	Tupolev Tu-204-100C	Cainiao, opb Aviastar-TU	0	w/o
4	12 Jan 2022	Beechcraft 200 Super King Air	private	0	w/o

```
In [11]: #naming the columns properly
df.columns = ['acc.date', 'aircraft_type', 'operator', 'fatalities', 'damage']
df.head()
```

	acc.date	aircraft_type	operator	fatalities	damage
0	3 Jan 2022	British Aerospace 4121 Jetstream 41	SA Airlink	0	sub
1	4 Jan 2022	British Aerospace 3101 Jetstream 31	LANHSA - Línea Aérea Nacional de Honduras S.A	0	sub
2	5 Jan 2022	Boeing 737-4H6	Caspian Airlines	0	sub
3	8 Jan 2022	Tupolev Tu-204-100C	Cainiao, opb Aviastar-TU	0	w/o
4	12 Jan 2022	Beechcraft 200 Super King Air	private	0	w/o

```
In [12]: # Remove rows where 'acc.date' is empty after cleaning
df = df.copy()
#Ensuring the date is in a string
df['acc.date'] = df['acc.date'].astype(str)
#Removing 'xx' and '' and replacing with False
df['acc.date'] = df['acc.date'].str.replace('xx', '', regex=False)
#Removing unk.date, and replacing with false
df['acc.date'] = df['acc.date'].str.replace('unk. date', '', regex=False)
# Removing spaces in our string
df['acc.date'] = df['acc.date'].str.strip()
# Convert to datetime
df['acc.date'] = pd.to_datetime(df['acc.date'], errors='coerce')
#Converting acc.date from object to datetime
# Drop rows where conversion failed (NaT)
df = df.dropna(subset=['acc.date'])
df['acc.date']
```

```
Out[12]: 0    2022-01-03
         1    2022-01-04
         2    2022-01-05
         3    2022-01-08
         4    2022-01-12
         ...
        1244  2018-12-20
        1245  2018-12-20
        1246  2018-12-22
        1247  2018-12-24
        1248  2018-12-31
Name: acc.date, Length: 1247, dtype: datetime64[ns]
```

```
In [13]: # Extract year from 'acc.date' and create a new column
df['year'] = df['acc.date'].dt.year
df
```

	acc.date	aircraft_type	operator	fatalities	damage	year
0	2022-01-03	British Aerospace 4121 Jetstream 41	SA Airlink	0	sub	2022
1	2022-01-04	British Aerospace 3101 Jetstream 31	LANHSA - Línea Aérea Nacional de Honduras S.A	0	sub	2022
2	2022-01-05	Boeing 737-4H6	Caspian Airlines	0	sub	2022
3	2022-01-08	Tupolev Tu-204-100C	Cainiao, opb Aviastar-TU	0	w/o	2022
4	2022-01-12	Beechcraft 200 Super King Air	private	0	w/o	2022
...
1244	2018-12-20	Antonov An-26B	Gomair	7	w/o	2018
1245	2018-12-20	Cessna 560 Citation V	Chen Aircrafts LLC	4	w/o	2018
1246	2018-12-22	PZL-Mielec M28 Skytruck	Guardia Nacional Bolivariana de Venezuela - GNBV	0	sub	2018
1247	2018-12-24	Antonov An-26B	Air Force of the Democratic Republic of the Congo	0	w/o	2018
1248	2018-12-31	Boeing 757-2B7 (WL)	American Airlines	0	sub	2018

1247 rows × 6 columns

```
In [14]: #Remove spaces
df['aircraft_type'] = df['aircraft_type'].str.strip()
# Standardize Capitalization
df['aircraft_type'] = df['aircraft_type'].str.title()
df['aircraft_type']
```

```
Out[14]: 0      British Aerospace 4121 Jetstream 41
1      British Aerospace 3101 Jetstream 31
2                  Boeing 737-4H6
3                  Tupolev Tu-204-100C
4      Beechcraft 200 Super King Air
...
1244          Antonov An-26B
1245          Cessna 560 Citation V
1246          Pzl-Mielec M28 Skytruck
1247          Antonov An-26B
1248          Boeing 757-2B7 (W1)
Name: aircraft_type, Length: 1247, dtype: object
```

```
In [15]: #Cleaning Operator column
#Remove spaces
df['operator'] = df['operator'].str.strip()
# Standardize Capitalization
df['operator'] = df['operator'].str.title()
df['operator'] = df['operator'].fillna('unknown')
df['operator']
```

```
Out[15]: 0                      Sa Airlink
1      Lanhsa - Línea Aérea Nacional De Honduras S.A
2                  Caspian Airlines
3      Cainiao, Opb Aviastar-Tu
4                  Private
...
1244          Gomair
1245          Chen Aircrafts Llc
1246      Guardia Nacional Bolivariana De Venezuela - Gnbv
1247      Air Force Of The Democratic Republic Of The Congo
1248          American Airlines
Name: operator, Length: 1247, dtype: object
```

```
In [16]: # Cleaning Fatalities column
#Converting from a string to numeric, filling missing values and making it an integer
df['fatalities'] = pd.to_numeric(df['fatalities'], errors='coerce').fillna(0).astype(int)
df['fatalities']
```

```
Out[16]: 0      0
         1      0
         2      0
         3      0
         4      0
         ..
        1244    7
        1245    4
        1246    0
        1247    0
        1248    0
Name: fatalities, Length: 1247, dtype: int32
```

```
In [35]: # Strip whitespace and Lowercase
df['damage'] = df['damage'].str.strip().str.lower()

# Replace shorthand codes with full, standardized Labels
df['damage'] = df['damage'].replace({'sub': 'Substantial', 'w/o': 'Written Off', 'non': 'Non-Damaging' })

# Fill any missing values
df['damage'] = df['damage'].fillna('Unknown')
df['damage']
```

```
Out[35]: 0      substantial
         1      substantial
         2      substantial
         3      written off
         4      written off
         ...
        1244    written off
        1245    written off
        1246    substantial
        1247    written off
        1248    substantial
Name: damage, Length: 1247, dtype: object
```

```
In [18]: df
```

Out[18]:

	acc.date	aircraft_type	operator	fatalities	damage	year
0	2022-01-03	British Aerospace 4121 Jetstream 41	Sa Airlink	0	Substantial	2022
1	2022-01-04	British Aerospace 3101 Jetstream 31	Lanhsa - Línea Aérea Nacional De Honduras S.A	0	Substantial	2022
2	2022-01-05	Boeing 737-4H6	Caspian Airlines	0	Substantial	2022
3	2022-01-08	Tupolev Tu-204-100C	Cainiao, Opb Aviastar-Tu	0	Written Off	2022
4	2022-01-12	Beechcraft 200 Super King Air	Private	0	Written Off	2022
...
1244	2018-12-20	Antonov An-26B	Gomair	7	Written Off	2018
1245	2018-12-20	Cessna 560 Citation V	Chen Aircrafts Llc	4	Written Off	2018
1246	2018-12-22	Pzl-Mielec M28 Skytruck	Guardia Nacional Bolivariana De Venezuela - Gnbv	0	Substantial	2018
1247	2018-12-24	Antonov An-26B	Air Force Of The Democratic Republic Of The Congo	0	Written Off	2018
1248	2018-12-31	Boeing 757-2B7 (WL)	American Airlines	0	Substantial	2018

1247 rows × 6 columns

DATA ANALYSIS

In [19]:

```
#Grouping the data by year
accidents_per_year = df.groupby(df['acc.date'].dt.year).size()
accidents_per_year.head()
```

Out[19]: acc.date

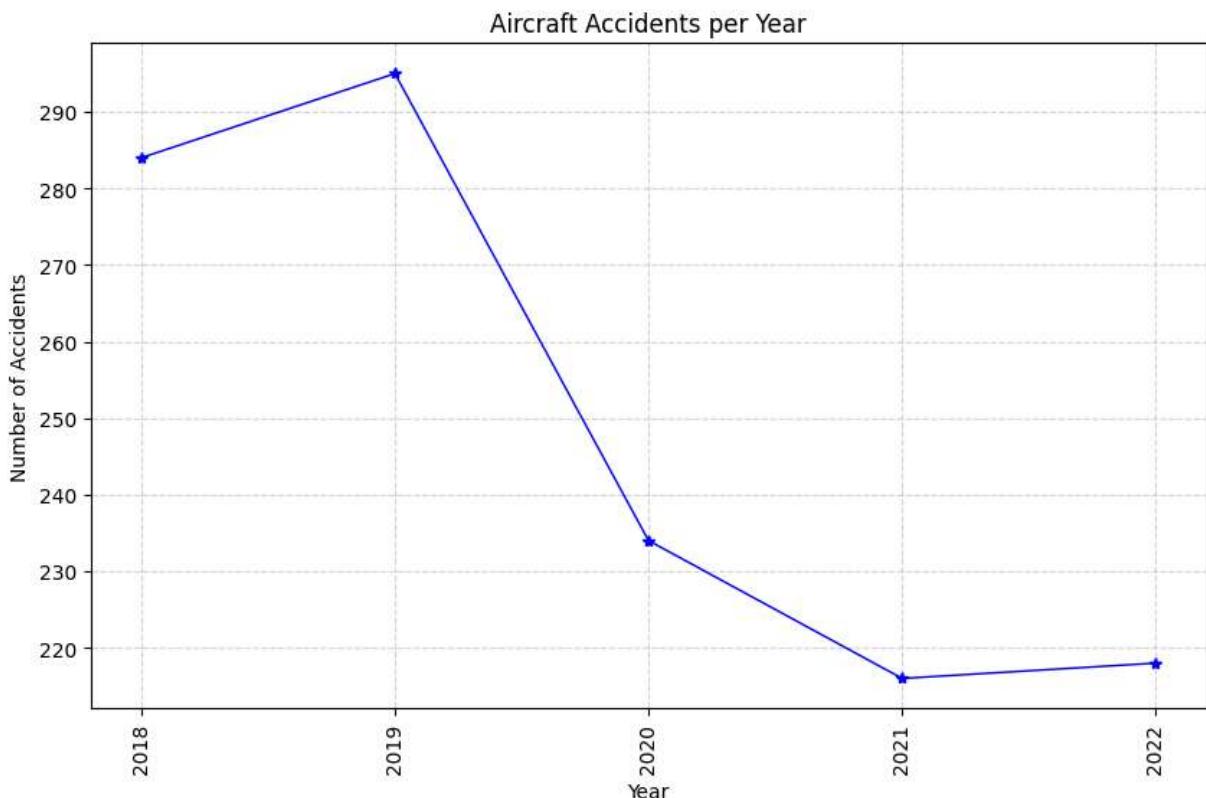
2018	284
2019	295
2020	234
2021	216
2022	218

dtype: int64

In [20]:

```
# Plotting how many accidents were there each year to see the trend
plt.figure(figsize=(10,6))
# Line with markers
plt.plot(accidents_per_year.index, accidents_per_year.values, marker='*', color='blue')
```

```
# Labels and title
plt.xlabel('Year')
plt.ylabel('Number of Accidents')
plt.title('Aircraft Accidents per Year')
# Make x-axis show all years
plt.xticks(accidents_per_year.index, rotation=90)
# grid
plt.grid(True, linestyle='--', alpha=0.5)
plt.show()
```



In [44]: `accidents_per_type = df['aircraft_type'].value_counts()
accidents_per_type`

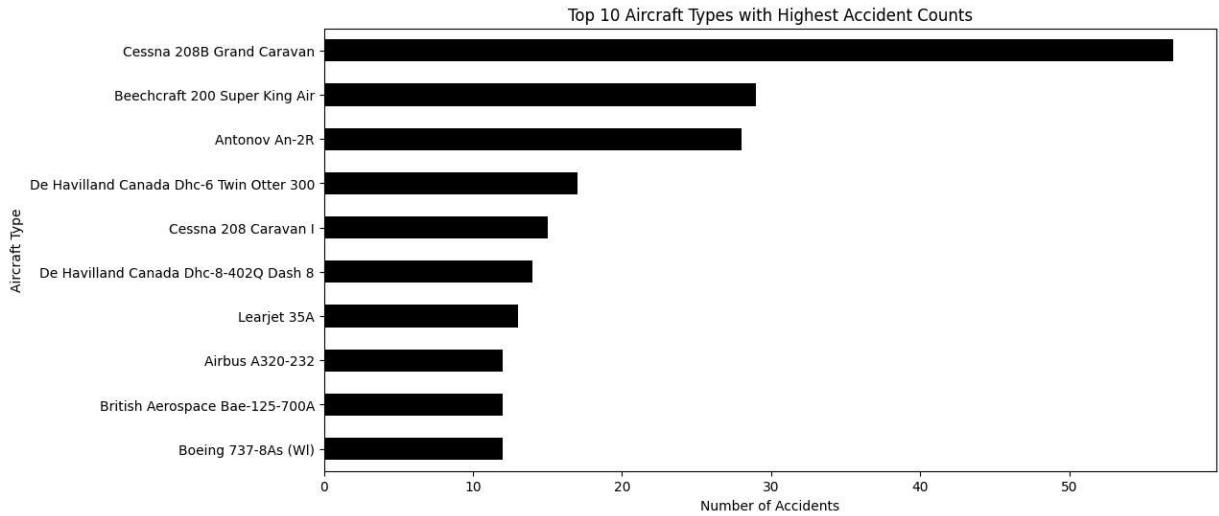
Out[44]: `aircraft_type`

Cessna 208B Grand Caravan	57
Beechcraft 200 Super King Air	29
Antonov An-2R	28
De Havilland Canada Dhc-6 Twin Otter 300	17
Cessna 208 Caravan I	15
..	
Boeing 767-375Er	1
Boeing 747-412 (Bcf)	1
Boeing 747-412F (Scd)	1
Boeing 737-76N (Wl)	1
Boeing 757-2B7 (Wl)	1

`Name: count, Length: 522, dtype: int64`

In [29]: `# Identify aircraft types with the highest accidents rates .
#Count how many accidents each aircraft type has
accidents_per_type = df['aircraft_type'].value_counts()
#Select the top 10 aircraft types with the highest accident counts`

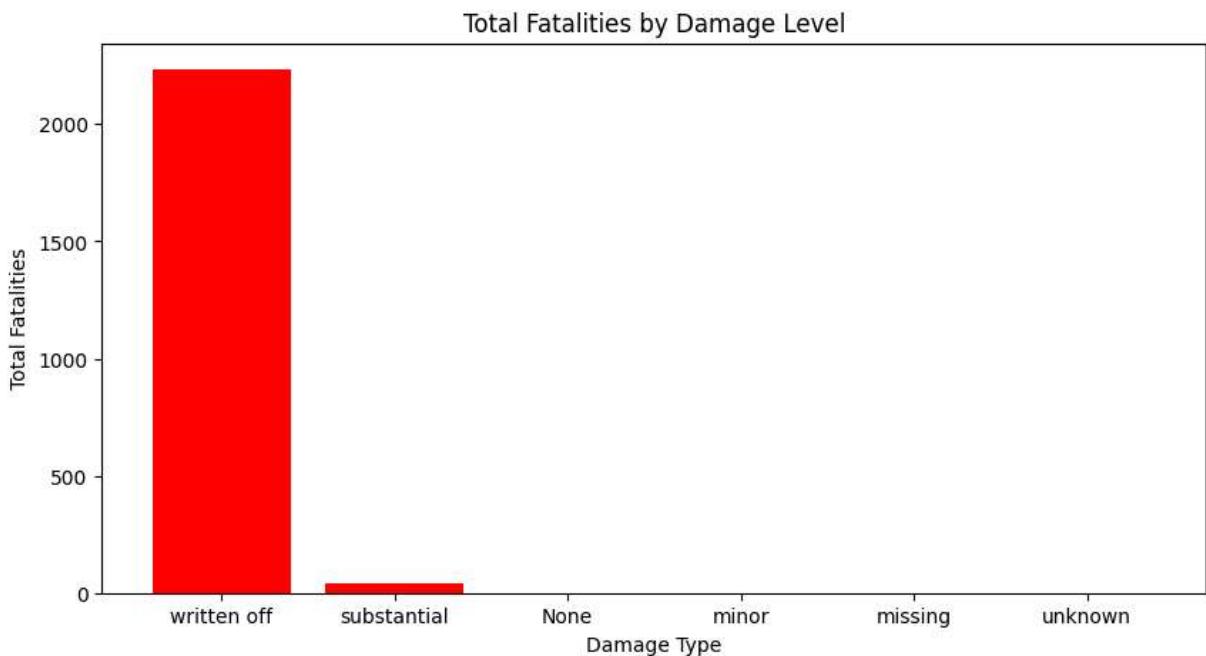
```
top_10_aircraft = accidents_per_type.head(10)
# Plot a horizontal bar chart
plt.figure(figsize=(12,6))
top_10_aircraft.sort_values().plot(kind='barh', color='black')
plt.title('Top 10 Aircraft Types with Highest Accident Counts')
plt.xlabel('Number of Accidents')
plt.ylabel('Aircraft Type')
plt.show()
```



In [42]:

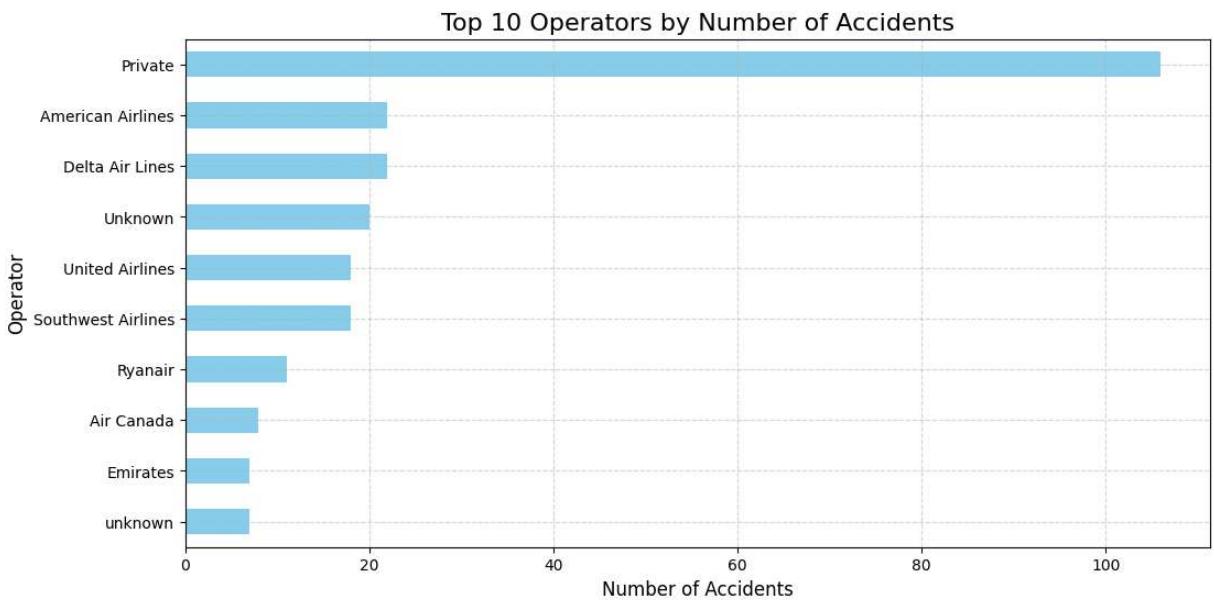
```
#evaluate the fatalities with the type of damage level
fatalities = df['fatalities'].value_counts()
#Total fatalities per aircraft type
fatalities_by_damage = df.groupby('damage')['fatalities'].sum().sort_values(ascending=True)
# Total accidents by damage type
damage_summary = df['damage'].value_counts()

# Visualize
plt.figure(figsize=(10,5))
plt.bar(x=fatalities_by_damage.index, height=fatalities_by_damage.values, color='red')
plt.title("Total Fatalities by Damage Level")
plt.xlabel('Damage Type')
plt.ylabel("Total Fatalities")
plt.show()
```



```
In [25]: # which operators had the highest accidents
```

```
accidents_per_operator = df['operator'].value_counts()
top_10_operators = accidents_per_operator.head(10)
plt.figure(figsize=(12,6))
top_10_operators.sort_values().plot(kind='barh', color='skyblue')
plt.title('Top 10 Operators by Number of Accidents', fontsize=16)
plt.xlabel('Number of Accidents', fontsize=12)
plt.ylabel('Operator', fontsize=12)
plt.show()
```



FINDINGS

- 1.The highest number of accidents were in 2019, followed by 2018.Accidents dropped significantly in 2020-2021,due to COVID-19 lockdowns, with a slight increase in 2022 as

aviation operations resumed.

2.The Cessna 208B Grand Caravan experienced most accidents, followed by Beechcraft 200 Super King Air, Antonov An-2R, De Havilland Canada DHC-6 Twin Otter 300, and Cessna 208 Caravan I.

3.Aircraft classified as 'written off' show the highest fatality rates, while accidents with 'substantial' damage resulted in relatively few fatalities.This indicates that more severe damage generally corresponds to higher fatalities.

4.Private operators experienced the highest number of accidents, highlighting a potential area for strict safety oversight. 15

Recommendations

1.After COVID-19, more safety measures should be taken as operations resume to normalcy and also standards to be raised to curb COVID-19 like testing our clients for it to prevent the spread.

2.Purchases should be made on flights with the lowest accidents rates.If need be to purchase flights with high risk extra caution on safety should be taken.

3.Prevention of accidents and impact mitigation to prevent written off aircraft which causes high fatalities should be taken example better weather monitoring methods.

4.Strict measures on private aircraft should be taken like regular audits ,maintenance and pilot certification since it has the highest accidents.