Airplane Accident Data Analysis

Name Catherine Sheila
Project: Phase 1 Data Analysis
Tools Used: Python (Pandas), Tableau

Objective

The goal of this project is to analyze global aircraft accident data to identify patterns, trends, and risk factors associated with different types of flights, weather conditions, and aircraft categories.

We aim to answer the following key business questions:

1. Which aircraft category is associated with injuries or fatalities?
2. Are commercial or private aircraft more prone to accidents?
3. Which countries experience the most aircraft incidents?
4. Injuries both Fatal and Severe in the aircraft industry
5. Which aircraft makes and models are most frequently involved in accidents?
6. Which flight phases (takeoff, landing, etc.) have the highest number of accidents?

In [89]:
```
import pandas as pd
import numpy as np

df = pd.read_csv(r"C:\Users\USER\Desktop\Aviation_Data.csv")
print(df.head())
```

```
c:\Users\USER\anaconda3\envs\learn-env\lib\site-packages\IPython\core\interac
tiveshell.py:3145: DtypeWarning: Columns (6,7,28) have mixed types.Specify dt
ype 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  ... Purpose.of.flight Air.carrier Total.Fatal.Injuries  \
0          NaN  ...          Personal         NaN                  2.0
1          NaN  ...          Personal         NaN                  4.0
2          NaN  ...          Personal         NaN                  3.0
3          NaN  ...          Personal         NaN                  2.0
4          NaN  ...          Personal         NaN                  1.0

  Total.Serious.Injuries Total.Minor.Injuries Total.Uninjured  \
0                    0.0                  0.0             0.0
1                    0.0                  0.0             0.0
2                    NaN                  NaN             NaN
3                    0.0                  0.0             0.0
4                    2.0                  NaN             0.0

  Weather.Condition  Broad.phase.of.flight   Report.Status Publication.Date
0               UNK                 Cruise  Probable Cause             NaN
1               UNK               Unknown  Probable Cause      19-09-1996
2               IMC                 Cruise  Probable Cause      26-02-2007
3               IMC                 Cruise  Probable Cause      12-09-2000
4               VMC              Approach  Probable Cause      16-04-1980

[5 rows x 31 columns]
```

In [90]:
```
#df.info()
#print(df)
```

Data cleaning and preparation. Identify total number of null data in the data set

In [91]:
```
df.isnull().sum()
print(df)
```

```
             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
...               ...                ...             ...         ...
90343  20221227106491           Accident      ERA23LA093  2022-12-26
90344  20221227106494           Accident      ERA23LA095  2022-12-26
90345  20221227106497           Accident      WPR23LA075  2022-12-26
90346  20221227106498           Accident      WPR23LA076  2022-12-26
90347  20221230106513           Accident      ERA23LA097  2022-12-29

              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
...                ...            ...      ...       ...          ...
90343    Annapolis, MD  United States      NaN       NaN          NaN
90344      Hampton, NH  United States      NaN       NaN          NaN
90345       Payson, AZ  United States  341525N  1112021W          PAN
90346       Morgan, UT  United States      NaN       NaN          NaN
90347       Athens, GA  United States      NaN       NaN          NaN

      Airport.Name  ... Purpose.of.flight         Air.carrier  \
0              NaN  ...          Personal                 NaN
1              NaN  ...          Personal                 NaN
2              NaN  ...          Personal                 NaN
3              NaN  ...          Personal                 NaN
4              NaN  ...          Personal                 NaN
...            ...  ...               ...                 ...
90343          NaN  ...          Personal                 NaN
90344          NaN  ...               NaN                 NaN
90345       PAYSON  ...          Personal                 NaN
90346          NaN  ...          Personal  MC CESSNA 210N LLC
90347          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
...                    ...                    ...                   ...
90343                  0.0                    1.0                   0.0
90344                  0.0                    0.0                   0.0
90345                  0.0                    0.0                   0.0
90346                  0.0                    0.0                   0.0
90347                  0.0                    1.0                   0.0

      Total.Uninjured Weather.Condition  Broad.phase.of.flight  \
0                 0.0               UNK                 Cruise
1                 0.0               UNK                Unknown
2                 NaN               IMC                 Cruise
3                 0.0               IMC                 Cruise
```

```
4                0.0              VMC               Approach
...              ...              ...                    ...
90343            0.0              NaN                    NaN
90344            0.0              NaN                    NaN
90345            1.0              VMC                    NaN
90346            0.0              NaN                    NaN
90347            1.0              NaN                    NaN

        Report.Status Publication.Date
0      Probable Cause              NaN
1      Probable Cause       19-09-1996
2      Probable Cause       26-02-2007
3      Probable Cause       12-09-2000
4      Probable Cause       16-04-1980
...              ...              ...
90343            NaN       29-12-2022
90344            NaN              NaN
90345            NaN       27-12-2022
90346            NaN              NaN
90347            NaN       30-12-2022

[90348 rows x 31 columns]
```

In [92]:
```
df.columns
print(df)
```

```
             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
...               ...                ...             ...         ...
90343  20221227106491           Accident      ERA23LA093  2022-12-26
90344  20221227106494           Accident      ERA23LA095  2022-12-26
90345  20221227106497           Accident      WPR23LA075  2022-12-26
90346  20221227106498           Accident      WPR23LA076  2022-12-26
90347  20221230106513           Accident      ERA23LA097  2022-12-29

               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
...                ...            ...      ...       ...          ...
90343    Annapolis, MD  United States      NaN       NaN          NaN
90344      Hampton, NH  United States      NaN       NaN          NaN
90345       Payson, AZ  United States  341525N  1112021W          PAN
90346       Morgan, UT  United States      NaN       NaN          NaN
90347       Athens, GA  United States      NaN       NaN          NaN

      Airport.Name  ... Purpose.of.flight         Air.carrier  \
0              NaN  ...          Personal                 NaN
1              NaN  ...          Personal                 NaN
2              NaN  ...          Personal                 NaN
3              NaN  ...          Personal                 NaN
4              NaN  ...          Personal                 NaN
...            ...  ...               ...                 ...
90343          NaN  ...          Personal                 NaN
90344          NaN  ...               NaN                 NaN
90345       PAYSON  ...          Personal                 NaN
90346          NaN  ...          Personal  MC CESSNA 210N LLC
90347          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
...                    ...                    ...                  ...
90343                  0.0                    1.0                  0.0
90344                  0.0                    0.0                  0.0
90345                  0.0                    0.0                  0.0
90346                  0.0                    0.0                  0.0
90347                  0.0                    1.0                  0.0

      Total.Uninjured Weather.Condition  Broad.phase.of.flight  \
0                 0.0               UNK                 Cruise
1                 0.0               UNK                Unknown
2                 NaN               IMC                 Cruise
3                 0.0               IMC                 Cruise
```

```
4                0.0              VMC               Approach
...              ...              ...                    ...
90343            0.0              NaN                    NaN
90344            0.0              NaN                    NaN
90345            1.0              VMC                    NaN
90346            0.0              NaN                    NaN
90347            1.0              NaN                    NaN

        Report.Status Publication.Date
0      Probable Cause              NaN
1      Probable Cause       19-09-1996
2      Probable Cause       26-02-2007
3      Probable Cause       12-09-2000
4      Probable Cause       16-04-1980
...               ...              ...
90343             NaN       29-12-2022
90344             NaN              NaN
90345             NaN       27-12-2022
90346             NaN              NaN
90347             NaN       30-12-2022

[90348 rows x 31 columns]
```

In [ ]:

In [93]:
```
#df.shape
#print(df)
```

In [94]:
```
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
```

In [95]:
```
#pd.set_option('display.max_columns', None)
#df.info()

print(df)
```

```
             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
...               ...                ...             ...         ...
90343  20221227106491           Accident      ERA23LA093  2022-12-26
90344  20221227106494           Accident      ERA23LA095  2022-12-26
90345  20221227106497           Accident      WPR23LA075  2022-12-26
90346  20221227106498           Accident      WPR23LA076  2022-12-26
90347  20221230106513           Accident      ERA23LA097  2022-12-29

              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
...                ...            ...      ...       ...          ...
90343    Annapolis, MD  United States      NaN       NaN          NaN
90344      Hampton, NH  United States      NaN       NaN          NaN
90345       Payson, AZ  United States  341525N  1112021W          PAN
90346       Morgan, UT  United States      NaN       NaN          NaN
90347       Athens, GA  United States      NaN       NaN          NaN

      Airport.Name  ... Purpose.of.flight         Air.carrier  \
0              NaN  ...          Personal                 NaN
1              NaN  ...          Personal                 NaN
2              NaN  ...          Personal                 NaN
3              NaN  ...          Personal                 NaN
4              NaN  ...          Personal                 NaN
...            ...  ...               ...                 ...
90343          NaN  ...          Personal                 NaN
90344          NaN  ...               NaN                 NaN
90345       PAYSON  ...          Personal                 NaN
90346          NaN  ...          Personal  MC CESSNA 210N LLC
90347          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
...                    ...                    ...                   ...
90343                  0.0                    1.0                   0.0
90344                  0.0                    0.0                   0.0
90345                  0.0                    0.0                   0.0
90346                  0.0                    0.0                   0.0
90347                  0.0                    1.0                   0.0

      Total.Uninjured Weather.Condition  Broad.phase.of.flight  \
0                 0.0               UNK                 Cruise
1                 0.0               UNK                Unknown
2                 NaN               IMC                 Cruise
3                 0.0               IMC                 Cruise
```

```
4                0.0              VMC               Approach
...              ...              ...                    ...
90343            0.0              NaN                    NaN
90344            0.0              NaN                    NaN
90345            1.0              VMC                    NaN
90346            0.0              NaN                    NaN
90347            1.0              NaN                    NaN

        Report.Status Publication.Date
0      Probable Cause              NaN
1      Probable Cause       19-09-1996
2      Probable Cause       26-02-2007
3      Probable Cause       12-09-2000
4      Probable Cause       16-04-1980
...               ...              ...
90343             NaN       29-12-2022
90344             NaN              NaN
90345             NaN       27-12-2022
90346             NaN              NaN
90347             NaN       30-12-2022

[90348 rows x 31 columns]
```

In [96]:
```
#print(df.head())
```

we clean and prepare the dataset for analysis by droping columns with less importance to our objective. e.g publication date which will only give us insight on when the accident event was published. Then save DataFrame to cleaned Data Frame for analysis and maintain the original data

In [97]:
```
cleaned_df = df.drop('Publication.Date', axis = 1)
cleaned_df
```

Out[97]:

	Event.Id	Investigation.Type	Accident.Number	Event.Date	Location	Country
0	20001218X45444	Accident	SEA87LA080	1948-10-24	MOOSE CREEK, ID	United States
1	20001218X45447	Accident	LAX94LA336	1962-07-19	BRIDGEPORT, CA	United States
2	20061025X01555	Accident	NYC07LA005	1974-08-30	Saltville, VA	United States
3	20001218X45448	Accident	LAX96LA321	1977-06-19	EUREKA, CA	United States
4	20041105X01764	Accident	CHI79FA064	1979-08-02	Canton, OH	United States
...	...	...	...	...	...	...
90343	20221227106491	Accident	ERA23LA093	2022-12-26	Annapolis, MD	United States
90344	20221227106494	Accident	ERA23LA095	2022-12-26	Hampton, NH	United States
90345	20221227106497	Accident	WPR23LA075	2022-12-26	Payson, AZ	United States
90346	20221227106498	Accident	WPR23LA076	2022-12-26	Morgan, UT	United States
90347	20221230106513	Accident	ERA23LA097	2022-12-29	Athens, GA	United States

90348 rows × 30 columns

In [98]:
```
#cleaned_df.info()
#cleaned_df
```

In [99]:
```
#cleaned_df.info()
#cleaned_df
```

we clean and prepare the dataset for analysis by droping columns with less importance to our objective.

In [100]:
```
cleaned_df = df.drop([
    'Event.Id',
    'Latitude',
    'Longitude',
    'Airport.Code',
    'Airport.Name',
    'Accident.Number',
    'Registration.Number',
    'FAR.Description',
    'Schedule',
    'Air.carrier',
    'Report.Status',
    'Publication.Date'
], axis=1)
cleaned_df
```

Out[100]:

	Investigation.Type	Event.Date	Location	Country	Injury.Severity	Aircraft.damage	Ai
0	Accident	1948-10-24	MOOSE CREEK, ID	United States	Fatal(2)	Destroyed	
1	Accident	1962-07-19	BRIDGEPORT, CA	United States	Fatal(4)	Destroyed	
2	Accident	1974-08-30	Saltville, VA	United States	Fatal(3)	Destroyed	
3	Accident	1977-06-19	EUREKA, CA	United States	Fatal(2)	Destroyed	
4	Accident	1979-08-02	Canton, OH	United States	Fatal(1)	Destroyed	
...	...	...	...	...	...	...	
90343	Accident	2022-12-26	Annapolis, MD	United States	Minor	NaN	
90344	Accident	2022-12-26	Hampton, NH	United States	NaN	NaN	
90345	Accident	2022-12-26	Payson, AZ	United States	Non-Fatal	Substantial	
90346	Accident	2022-12-26	Morgan, UT	United States	NaN	NaN	
90347	Accident	2022-12-29	Athens, GA	United States	Minor	NaN	

90348 rows × 19 columns

In [101]:
```
cleaned_df.info()
cleaned_df
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 90348 entries, 0 to 90347
Data columns (total 19 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   Investigation.Type      90348 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   Aircraft.Category       32287 non-null  object 
 7   Make                    88826 non-null  object 
 8   Model                   88797 non-null  object 
 9   Amateur.Built           88787 non-null  object 
 10  Number.of.Engines       82805 non-null  float64
 11  Engine.Type             81812 non-null  object 
 12  Purpose.of.flight       82697 non-null  object 
 13  Total.Fatal.Injuries    77488 non-null  float64
 14  Total.Serious.Injuries  76379 non-null  float64
 15  Total.Minor.Injuries    76956 non-null  float64
 16  Total.Uninjured         82977 non-null  float64
 17  Weather.Condition       84397 non-null  object 
 18  Broad.phase.of.flight   61724 non-null  object 
dtypes: float64(5), object(14)
memory usage: 13.1+ MB
```

Out[101]:

	Investigation.Type	Event.Date	Location	Country	Injury.Severity	Aircraft.damage	Ai
0	Accident	1948-10-24	MOOSE CREEK, ID	United States	Fatal(2)	Destroyed	
1	Accident	1962-07-19	BRIDGEPORT, CA	United States	Fatal(4)	Destroyed	
2	Accident	1974-08-30	Saltville, VA	United States	Fatal(3)	Destroyed	
3	Accident	1977-06-19	EUREKA, CA	United States	Fatal(2)	Destroyed	
4	Accident	1979-08-02	Canton, OH	United States	Fatal(1)	Destroyed	
...	...	...	...	...	...	...	
90343	Accident	2022-12-26	Annapolis, MD	United States	Minor	NaN	
90344	Accident	2022-12-26	Hampton, NH	United States	NaN	NaN	
90345	Accident	2022-12-26	Payson, AZ	United States	Non-Fatal	Substantial	
90346	Accident	2022-12-26	Morgan, UT	United States	NaN	NaN	
90347	Accident	2022-12-29	Athens, GA	United States	Minor	NaN	

90348 rows × 19 columns

Identify columns with zero or missing values

In [102]:
```
cleaned_df= cleaned_df.isna().sum()
cleaned_df
```

Out[102]:
```
Investigation.Type           0
Event.Date                1459
Location                  1511
Country                   1685
Injury.Severity           2459
Aircraft.damage           4653
Aircraft.Category        58061
Make                      1522
Model                     1551
Amateur.Built             1561
Number.of.Engines         7543
Engine.Type               8536
Purpose.of.flight         7651
Total.Fatal.Injuries     12860
Total.Serious.Injuries   13969
Total.Minor.Injuries     13392
Total.Uninjured           7371
Weather.Condition         5951
Broad.phase.of.flight    28624
dtype: int64
```

In [103]:
```
print(cleaned_df)
```

```
Investigation.Type           0
Event.Date                1459
Location                  1511
Country                   1685
Injury.Severity           2459
Aircraft.damage           4653
Aircraft.Category        58061
Make                      1522
Model                     1551
Amateur.Built             1561
Number.of.Engines         7543
Engine.Type               8536
Purpose.of.flight         7651
Total.Fatal.Injuries     12860
Total.Serious.Injuries   13969
Total.Minor.Injuries     13392
Total.Uninjured           7371
Weather.Condition         5951
Broad.phase.of.flight    28624
dtype: int64
```

Create a new column combining all the injury columns to identify total injuries caused by the accidents

In [104]:
```
cleaned_df['Total.injuries'] = cleaned_df['Total.Serious.Injuries'] + cleaned_
df['Total.Minor.Injuries']
```

Replace missing values in the new column column with zeros since the values are unknown.

In [105]:
```
cleaned_df['Total.injuries'] = cleaned_df['Total.injuries'].fillna(0)
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
<ipython-input-105-ff5abb979d76> in <module>
----> 1 cleaned_df['Total.injuries'] = cleaned_df['Total.injuries'].fillna(0)

AttributeError: 'numpy.int64' object has no attribute 'fillna'
```

Replace missing values in the relevant numerical columns with zero by grouping all the numerial columns to a string in order to assignment zero value.

In [ ]:
```
zero_cols = [
    'Number.of.Engines',
    'Total.Fatal.Injuries',
    'Total.Serious.Injuries',
    'Total.Minor.Injuries',
    'Total.Uninjured'


]

df[zero_cols] = df[zero_cols].fillna(0)
```

Replace missing values in the categorical columns with unknown by grouping all the categorical columns to a string in order to assignment unknown value.

In [ ]:
```
unknown_cols = ['Investigation.Type','Location','Country',
    'Injury.Severity','Aircraft.damage','Aircraft.Category',
    'Make','Model','Amateur.Built','Engine.Type',
    'Purpose.of.flight','Weather.Condition',
    'Broad.phase.of.flight','Event.Date'
]
df[unknown_cols] = df[unknown_cols].fillna('Unknown')
```

In [ ]:
```
cleaned_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 90348 entries, 0 to 90347
Data columns (total 20 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   Investigation.Type      90348 non-null  object 
 1   Event.Date              90348 non-null  object 
 2   Location                90348 non-null  object 
 3   Country                 90348 non-null  object 
 4   Injury.Severity         90348 non-null  object 
 5   Aircraft.damage         90348 non-null  object 
 6   Aircraft.Category       90348 non-null  object 
 7   Make                    90348 non-null  object 
 8   Model                   90348 non-null  object 
 9   Amateur.Built           90348 non-null  object 
 10  Number.of.Engines       90348 non-null  float64
 11  Engine.Type             90348 non-null  object 
 12  Purpose.of.flight       90348 non-null  object 
 13  Total.Fatal.Injuries    90348 non-null  float64
 14  Total.Serious.Injuries  90348 non-null  float64
 15  Total.Minor.Injuries    90348 non-null  float64
 16  Total.Uninjured         90348 non-null  float64
 17  Weather.Condition       90348 non-null  object 
 18  Broad.phase.of.flight   90348 non-null  object 
 19  Total.injuries          90348 non-null  float64
dtypes: float64(6), object(14)
memory usage: 13.8+ MB
```

In [ ]:
```

```

Cleaned data with relevant columns and filled missing values

In [ ]:
```
cleaned_df
```

Out[ ]:

	Investigation.Type	Event.Date	Location	Country	Injury.Severity	Aircraft.damage	Ai
0	Accident	1948-10-24	MOOSE CREEK, ID	United States	Fatal(2)	Destroyed	
1	Accident	1962-07-19	BRIDGEPORT, CA	United States	Fatal(4)	Destroyed	
2	Accident	1974-08-30	Saltville, VA	United States	Fatal(3)	Destroyed	
3	Accident	1977-06-19	EUREKA, CA	United States	Fatal(2)	Destroyed	
4	Accident	1979-08-02	Canton, OH	United States	Fatal(1)	Destroyed	
...	...	...	...	...	...	...	
90343	Accident	2022-12-26	Annapolis, MD	United States	Minor	Unknown	
90344	Accident	2022-12-26	Hampton, NH	United States	Unknown	Unknown	
90345	Accident	2022-12-26	Payson, AZ	United States	Non-Fatal	Substantial	
90346	Accident	2022-12-26	Morgan, UT	United States	Unknown	Unknown	
90347	Accident	2022-12-29	Athens, GA	United States	Minor	Unknown	

90348 rows × 20 columns

Exported csv file for Analysis in tableau.

In [ ]:
```
cleaned_df.to_csv("C:/Users/USER/Documents/cleaned_aircraft_data.csv", index=F
alse)
print(cleaned_df)
```

```
      Investigation.Type  Event.Date         Location        Country  \
0               Accident  1948-10-24  MOOSE CREEK, ID  United States
1               Accident  1962-07-19   BRIDGEPORT, CA  United States
2               Accident  1974-08-30    Saltville, VA  United States
3               Accident  1977-06-19       EUREKA, CA  United States
4               Accident  1979-08-02       Canton, OH  United States
...                  ...         ...              ...           ...
90343           Accident  2022-12-26    Annapolis, MD  United States
90344           Accident  2022-12-26      Hampton, NH  United States
90345           Accident  2022-12-26       Payson, AZ  United States
90346           Accident  2022-12-26       Morgan, UT  United States
90347           Accident  2022-12-29       Athens, GA  United States

      Injury.Severity Aircraft.damage Aircraft.Category  \
0            Fatal(2)       Destroyed           Unknown
1            Fatal(4)       Destroyed           Unknown
2            Fatal(3)       Destroyed           Unknown
3            Fatal(2)       Destroyed           Unknown
4            Fatal(1)       Destroyed           Unknown
...               ...             ...               ...
90343           Minor         Unknown           Unknown
90344         Unknown         Unknown           Unknown
90345      Non-Fatal     Substantial          Airplane
90346         Unknown         Unknown           Unknown
90347           Minor         Unknown           Unknown

                           Make      Model Amateur.Built  Number.of.Engines
\
0                        Stinson      108-3            No                1.0
1                          Piper   PA24-180            No                1.0
2                         Cessna       172M            No                1.0
3                        Rockwell        112            No                1.0
4                         Cessna        501            No                0.0
...                          ...        ...           ...               ...
90343                     PIPER  PA-28-151            No                0.0
90344                   BELLANCA       7ECA            No                0.0
90345  AMERICAN CHAMPION AIRCRAFT      8GCBC            No                1.0
90346                     CESSNA       210N            No                0.0
90347                     PIPER  PA-24-260            No                0.0

        Engine.Type Purpose.of.flight  Total.Fatal.Injuries  \
0      Reciprocating          Personal                   2.0
1      Reciprocating          Personal                   4.0
2      Reciprocating          Personal                   3.0
3      Reciprocating          Personal                   2.0
4            Unknown          Personal                   1.0
...              ...               ...                   ...
90343        Unknown          Personal                   0.0
90344        Unknown           Unknown                   0.0
90345        Unknown          Personal                   0.0
90346        Unknown          Personal                   0.0
90347        Unknown          Personal                   0.0

       Total.Serious.Injuries  Total.Minor.Injuries  Total.Uninjured  \
0                         0.0                   0.0              0.0
1                         0.0                   0.0              0.0
2                         0.0                   0.0              0.0
```

3	0.0	0.0	0.0
4	2.0	0.0	0.0
...	...	...	...
90343	1.0	0.0	0.0
90344	0.0	0.0	0.0
90345	0.0	0.0	1.0
90346	0.0	0.0	0.0
90347	1.0	0.0	1.0

	Weather.Condition	Broad.phase.of.flight	Total.injuries
0	UNK	Cruise	0.0
1	UNK	Unknown	0.0
2	IMC	Cruise	0.0
3	IMC	Cruise	0.0
4	VMC	Approach	2.0
...	...	...	...
90343	Unknown	Unknown	1.0
90344	Unknown	Unknown	0.0
90345	VMC	Unknown	0.0
90346	Unknown	Unknown	0.0
90347	Unknown	Unknown	1.0

[90348 rows x 20 columns]