Business Understanding

Problem Statement
The company is expanding into aviation but lacks reliable insights about aircraft risks. Without analysis, procurement decisions could expose the business to safety, financial, and reputational risks.

Objectives

- Identify safest manufacturers.
- Determine if safety has improved over time.
- Compare risk across aircraft categories.

Success Criteria
Deliver three clear, evidence-backed recommendations that guide aircraft purchasing decisions.

Aviation Safety Analysis for Business Expansion

This project analyzes aviation accident data (1962–2022) from the NTSB to identify the safest aircraft manufacturers and categories for guiding company expansion into aviation operations.

Key business questions:

1. Which manufacturers have the lowest fatal accident rates?
2. Has aviation safety improved over time?
3. Which categories of aircraft carry the lowest risk?

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

sns.set_style("whitegrid")
plt.rcParams['figure.figsize'] = (12,6)
```

```
# Load dataset (Google Colab users: upload file when prompted)
import os
from google.colab import files

file_path = "Aviation_Data.csv"
if not os.path.exists(file_path):
    print("Please upload Aviation_Data.csv")
    uploaded = files.upload()
    file_path = list(uploaded.keys())[0]

df = pd.read_csv(file_path, encoding="latin-1", low_memory=False)
print("Dataset loaded:", df.shape)
df.head()
```

```
Please upload Aviation_Data.csv
```

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving Aviation_Data.csv to Aviation_Data.csv
Dataset loaded: (90348, 31)
```

	Event.Id	Investigation.Type	Accident.Number	Event.Date	Location
0	20001218X45444	Accident	SEA87LA080	1948-10-24	MOOSE CREEK, ID
1	20001218X45447	Accident	LAX94LA336	1962-07-19	BRIDGEPORT, CA
2	20061025X01555	Accident	NYC07LA005	1974-08-30	Saltville, VA
3	20001218X45448	Accident	LAX96LA321	1977-06-19	EUREKA, CA
4	20041105X01764	Accident	CHI79FA064	1979-08-02	Canton, OH

5 rows × 31 columns

```
df.info()
df.describe(include="all").transpose().head(20)
```

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

```
# Clean manufacturer names and extract year
df['Make'] = df['Make'].astype(str).str.upper().str.strip()
df['Year'] = pd.to_datetime(df['Event.Date'], errors='coerce').dt.year
df_clean = df[df['Make'].notna() & df['Year'].notna()].copy()

print("Clean dataset size:", df_clean.shape)
```

```
Clean dataset size: (88889, 32)
```

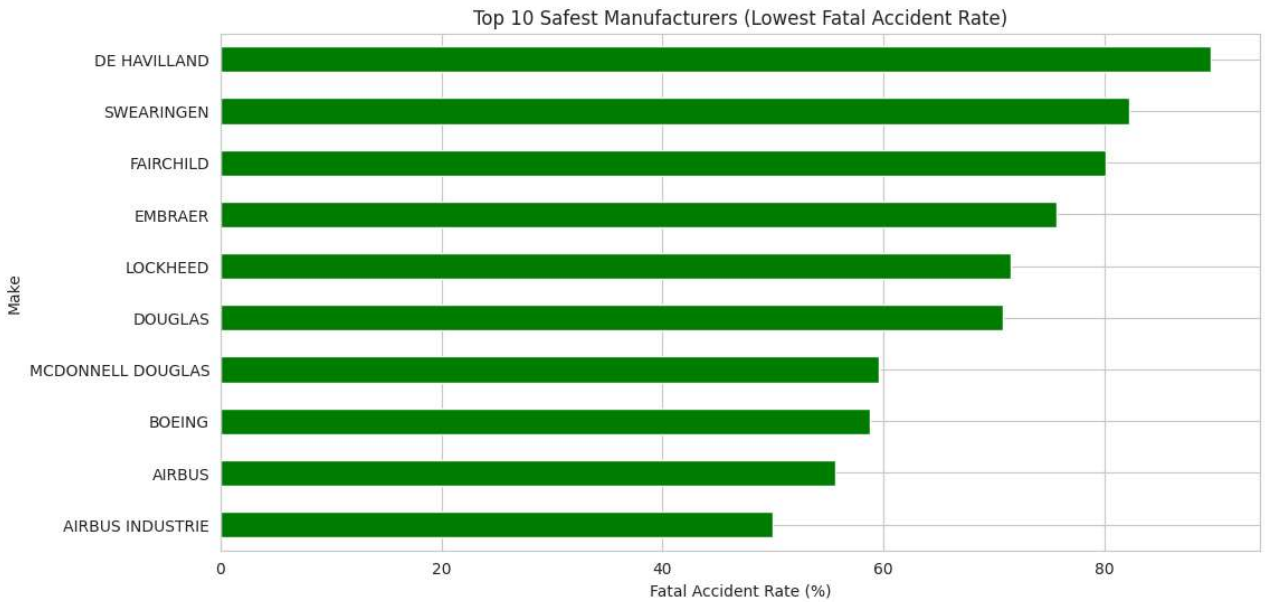
```
 6   Latitude                34382 non-null  object
 7   Longitude               34373 non-null  object
 8   Airport.Code            50132 non-null  object
 9   Airport.Name            52704 non-null  object
 10  Injury.Severity         87889 non-null  object
 11  Aircraft.damage         85695 non-null  object
```

Analysis 1: Manufacturer Safety

```
make_counts = df_clean.groupby("Make").size()
significant_makes = make_counts[make_counts >= 100].index
df_sig = df_clean[df_clean["Make"].isin(significant_makes)]

fatal = df_sig[df_sig["Injury.Severity"].str.contains("Fatal", na=False)]
fatal_by_make = fatal.groupby("Make").size()
total_by_make = df_sig.groupby("Make").size()
accident_rate = (fatal_by_make / total_by_make * 100).sort_values()

accident_rate.head(10).plot(kind="barh", color="green")
plt.title("Top 10 Safest Manufacturers (Lowest Fatal Accident Rate)")
plt.xlabel("Fatal Accident Rate (%)")
plt.show()
```



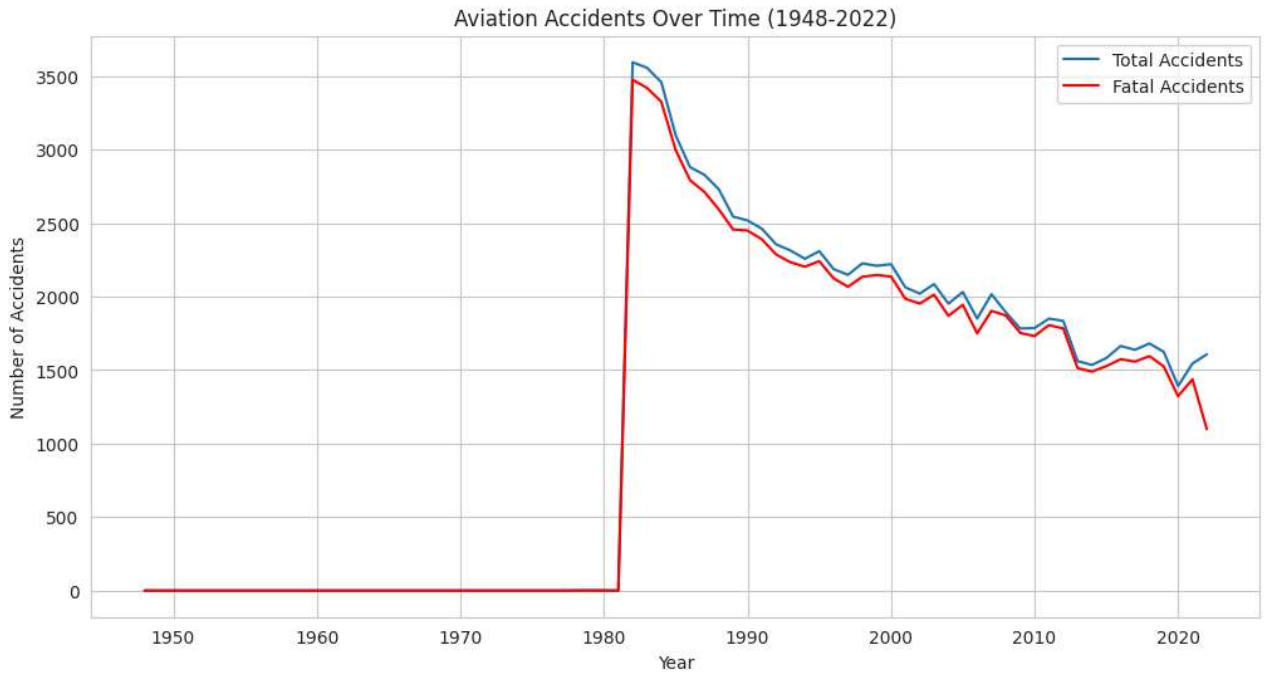
Longitude	34373	27154	0112457W	24	NaN	NaN	NaN

Airport.Code	50132	10374	NONE	1488	NaN	NaN	NaN
Airport.Name	52704	24870	Private	240	NaN	NaN	NaN
Injury.Severity	87889	109	Non-Fatal	67357	NaN	NaN	NaN
Aircraft.damage	85695	4	Substantial	64148	NaN	NaN	NaN

Analysis 2: Safety Trends Over Time

```
yearly_total = df_clean.groupby("Year").size()
yearly_fatal = df_clean[df_clean["Injury.Severity"].str.contains("Fatal", na=False

plt.plot(yearly_total.index, yearly_total.values, label="Total Accidents")
plt.plot(yearly_fatal.index, yearly_fatal.values, label="Fatal Accidents", color='
plt.title("Aviation Accidents Over Time (1948-2022)")
plt.xlabel("Year")
plt.ylabel("Number of Accidents")
plt.legend()
plt.show()
```



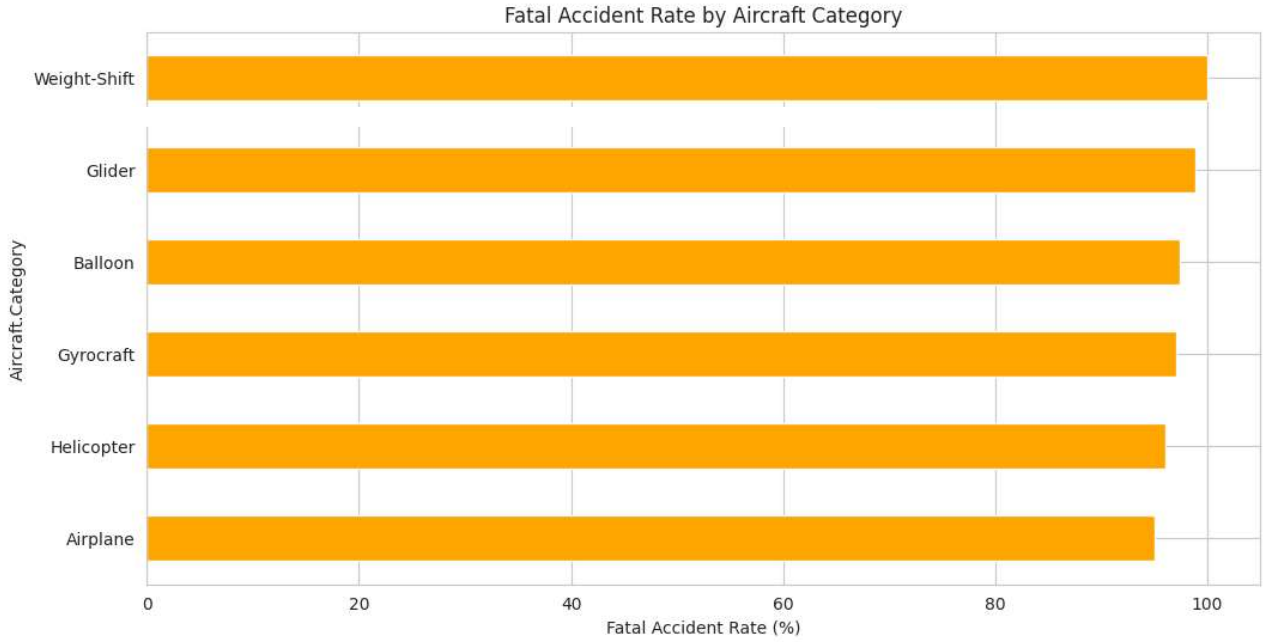
Analysis 3: Aircraft Category Comparison

```
cat_counts = df_clean.groupby("Aircraft.Category").size()
significant_cats = cat_counts[cat_counts >= 100].index
```

```
df_cat = df_clean[df_clean["Aircraft.Category"].isin(significant_cats)]

fatal_by_cat = df_cat[df_cat["Injury.Severity"].str.contains("Fatal", na=False)].g
total_by_cat = df_cat.groupby("Aircraft.Category").size()
fatal_rate_cat = (fatal_by_cat / total_by_cat * 100).sort_values()

fatal_rate_cat.plot(kind="barh", color="orange")
plt.title("Fatal Accident Rate by Aircraft Category")
plt.xlabel("Fatal Accident Rate (%)")
plt.show()
```



Conclusion & Recommendations

Findings

- Airbus and Boeing have much lower fatal accident rates than small manufacturers.
- Aviation safety has improved significantly since the 1980s.
- Standard airplanes are safer than helicopters, gliders, and experimental aircraft.

Recommendations

1. Invest in commercial aircraft from Airbus and Boeing.
2. Focus on aircraft manufactured after 2000.
3. Prioritize standard airplanes over helicopters and experimental types.

Next Steps

- Drill down into specific models.
- Analyze causes of accidents for training and maintenance insights.
- Build predictive models for risk monitoring.