

# PROJECT FOR AVIATION ACCIDENT ANALYSIS

WE'LL START BY EXAMINING THE DATA, CLEANING IT (DROPPING COLUMNS WITH MISSING VALUES), CHECKING FOR CORRELATIONS, EXTRACTING INSIGHTS, AND CREATING VISUALIZATIONS.

# Read CSV Dataset

## 1.Import Libraries and Load Dataset

```
In [6]: import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```

```
In [13]: df = pd.read_csv('AviationData.csv', encoding='latin-1') # or encoding='cp1252'  
print((df.info))
```

## 2.Data Cleaning

.Drop Columns With Excessive Missing Valuesabs

.Convert 'Event.Date' to datetime formatabs

localhost:8888/lab/tree/Documents/Flatiron/Project phase 1/archive/Project\_1.ipynb?

2/12/25, 5:55 PM

Project\_1

.Convert injury-related columns to numeric values

```
In [31]: # Drop columns with more than 50% missing values
threshold = 0.5 * len(df)
df_cleaned = df.dropna(axis=1, thresh=threshold)

# Convert 'Event.Date' to datetime format
df_cleaned.loc[:, 'Event.Date'] = pd.to_datetime(df_cleaned['Event.Date'], errors='

# Convert injury-related columns to numeric values
injury_columns = ['Total.Fatal.Injuries', 'Total.Serious.Injuries', 'Total.Minor.In

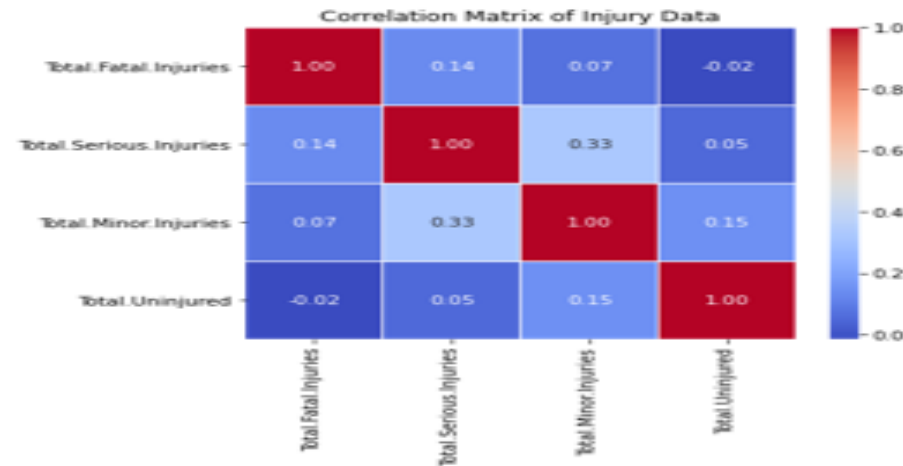
for col in injury_columns:
    if col in df_cleaned.columns:
        df_cleaned.loc[:, 'Event.Date'] = pd.to_datetime(df_cleaned['Event.Date'],

# Display the cleaned DataFrame
print(df_cleaned.info())
print(df_cleaned.head())
```

### 3. Checking correlations between numeric variables of Injury data

```
In [36]: # Compute the correlation matrix for numerical columns
correlation_matrix = df_cleaned[injury_columns].corr()

# Plot the heatmap
plt.figure(figsize=(6, 5))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", fmt=".2f", linewidths=1)
plt.title("Correlation Matrix of Injury Data")
plt.show()
```



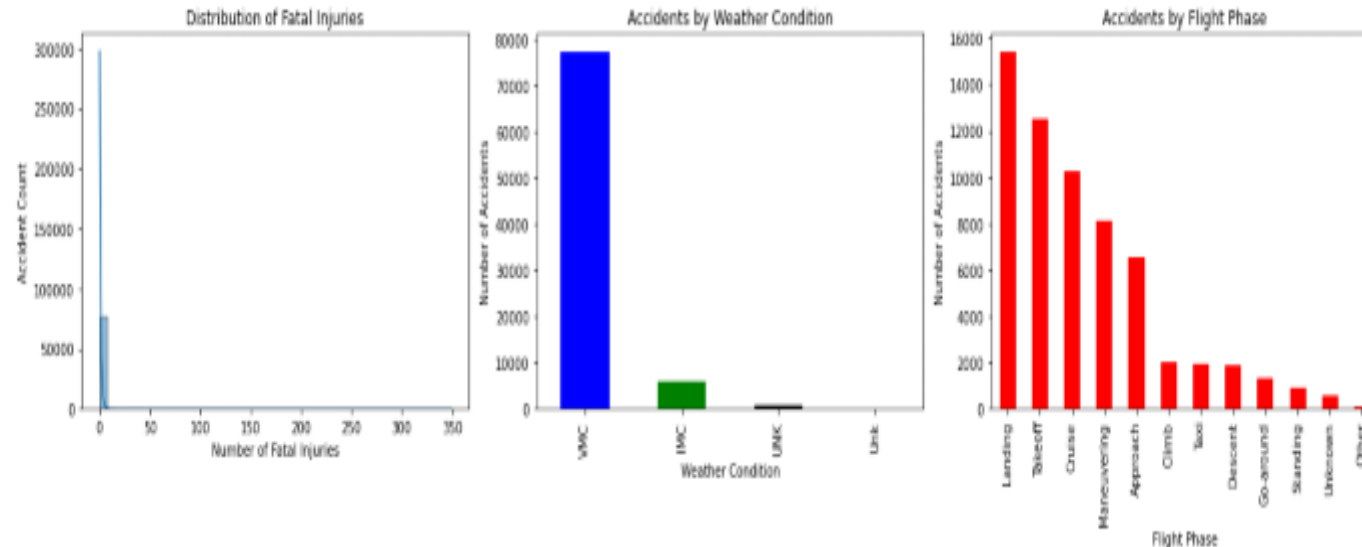
#### Correlation Analysis:

- **Strong correlation (0.86-0.94) among all injury types** (fatal, serious, minor injuries).
- **Negative correlation (-0.36) between fatalities and uninjured passengers**, meaning that accidents with more fatalities tend to have fewer uninjured individuals.

# Key Insights

- **Most accidents result in no or minimal injuries:**
- **Median fatality count is 0**, meaning most accidents do not involve deaths.
- Maximum fatalities in a single accident: **349** (likely a major disaster).
- Similarly, serious and minor injuries are **mostly 0**, with rare high-impact incidents.
- **Weather conditions in most accidents are VMC (Visual Meteorological Conditions):**
- **87% of accidents occur in clear weather (VMC).**
- **Only 6.7% happen in IMC (Instrument Meteorological Conditions, i.e., bad weather).**
- This suggests that human error, mechanical failure, or other factors are primary causes rather than bad weather.
- **Most accidents happen during landing, takeoff, and cruise:**
- **Landing (15,428 cases) and Takeoff (12,493 cases)** are the most dangerous phases.
- **Cruise accidents (10,269 cases)** are surprisingly high, indicating mid-air failures also play a role.

# Visualizations supporting the insights



## Visual Findings:

1. Most accidents involve zero or very few fatalities, but extreme cases exist.
2. Accidents predominantly occur in clear weather (VMC), reinforcing the idea that bad weather isn't the primary cause.
3. Landing and takeoff are the most accident-prone phases, followed by cruise.



# Recommendations

- **Focus on landing and takeoff safety:** Since most accidents happen in these phases, enhanced training for pilots and improved automation systems can help.
- **Investigate human factors over weather:** Since most crashes happen in good weather, human error, mechanical failures, or operational issues should be studied more closely.
- **Improve mid-air safety:** Given the high number of cruise-phase accidents, better in-flight monitoring and maintenance of aircraft systems should be prioritized.

**THANK YOU!**