A Comprehensive Study of Aviation Accidents: Trends, Patterns, and Safety Insights Across Time and Space

Masters Project

Submitted by

Bipropod Das Shubro

WM-ASDS (11th Batch)

Section-B

Roll-20231185

Supervised by

Farhana Afrin Duty

Assistant Professor

Department of Statistics

Jahangirnagar University

Contents

Introduction

- Aviation is considered one of the safest transportation modes. Despite advancements, accidents still occur.
- This study analyzes historical accident data to identify trends, factors, and vulnerabilities across military, commercial, and private aviation.
- It evaluates the impact of technological innovations and regulatory changes on safety improvements.
- Findings offer actionable insights for enhancing future aviation safety

Objective of the Study

- Examine the historical evolution of aviation accidents and fatalities in relation to technological advancements and safety protocols.
- Analyze the differences in accident rates and severity among military, commercial, and private aviation operators.
- Identify geographic regions and aircraft types associated with higher accident frequencies and fatality rates.
- Investigate the influence of time factors (e.g., time of day) on accident outcomes.
- Assess the impact of major aviation innovations and regulations on the frequency and severity of accidents over time. (Time Series Analysis)

Data Set Information

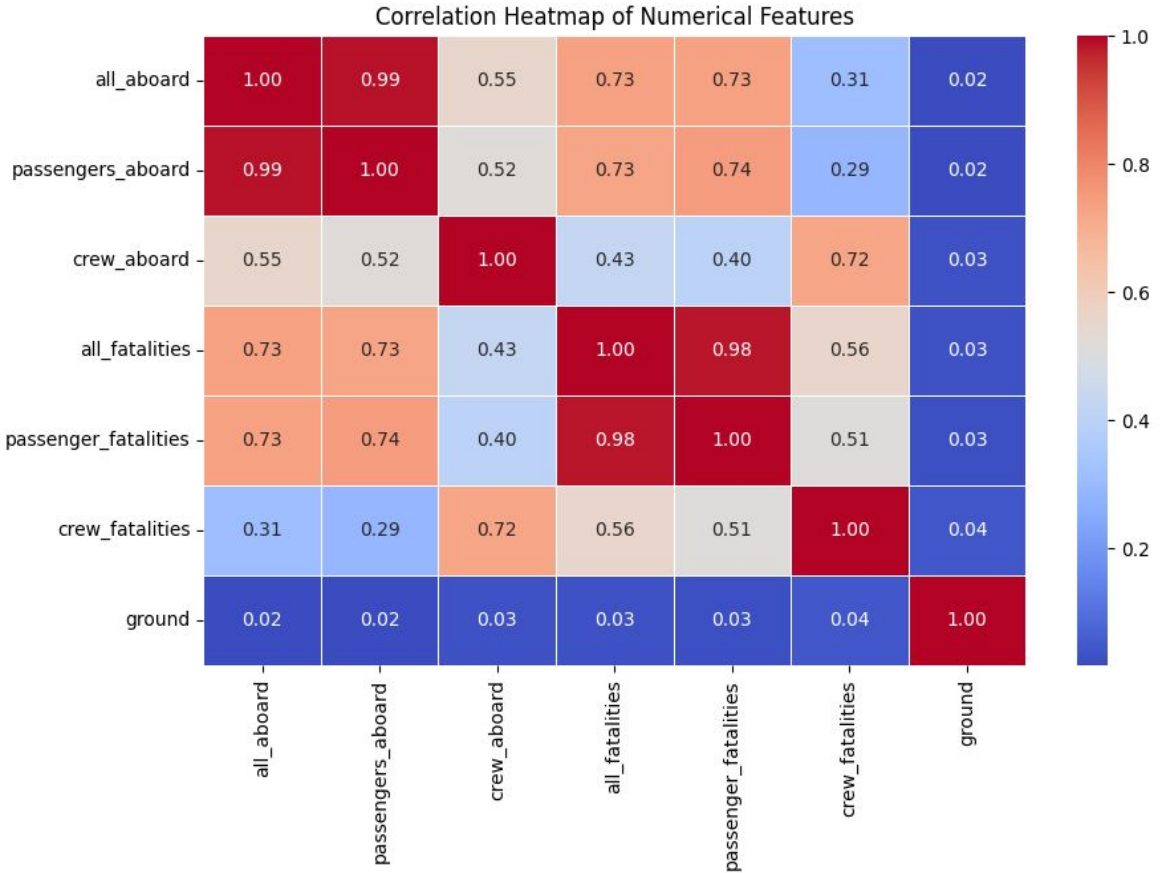
- Data source: https://www.planecrashinfo.com/database.htm
- Web scraping techniques used for data extraction.
- Dataset has 5052 entries and total 18 columns.
- Data includes accident date, location, aircraft type, fatalities, and Summary.
- Technologies used: BeautifulSoup, Requests, CSV Module, Pandas.

Exploratory Data Analysis

- Data includes accident date, time, location, operator, flight no, route, aircraft type, registration, fatalities, and Summary.
- Created heatmaps to show correlations between variables.
- Used bar charts to analyze fatalities over time
- Developed maps for geospatial accident analysis.
- Plotted time-series trends of accident occurrences.

Correlation analysis for numerical column

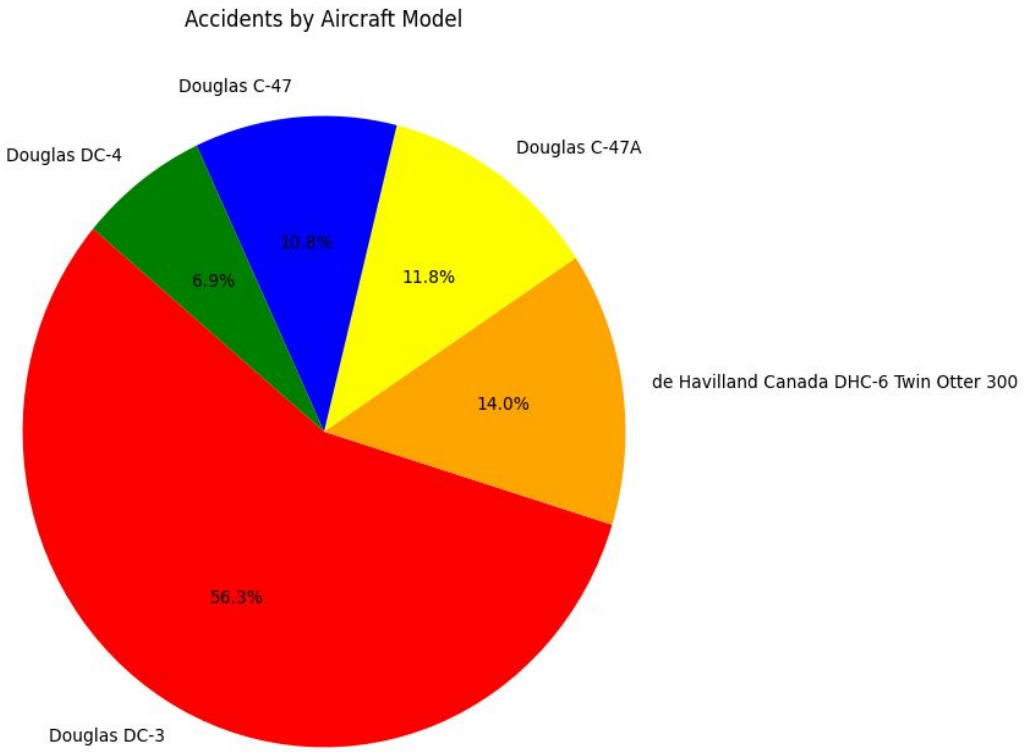
1. **Highly Correlated Variables**:
   - ⇨ all_aboard and passengers_aboard have a nearly perfect correlation (0.99)
   - ⇨ all_fatalities and passenger_fatalities are highly correlated (0.98)



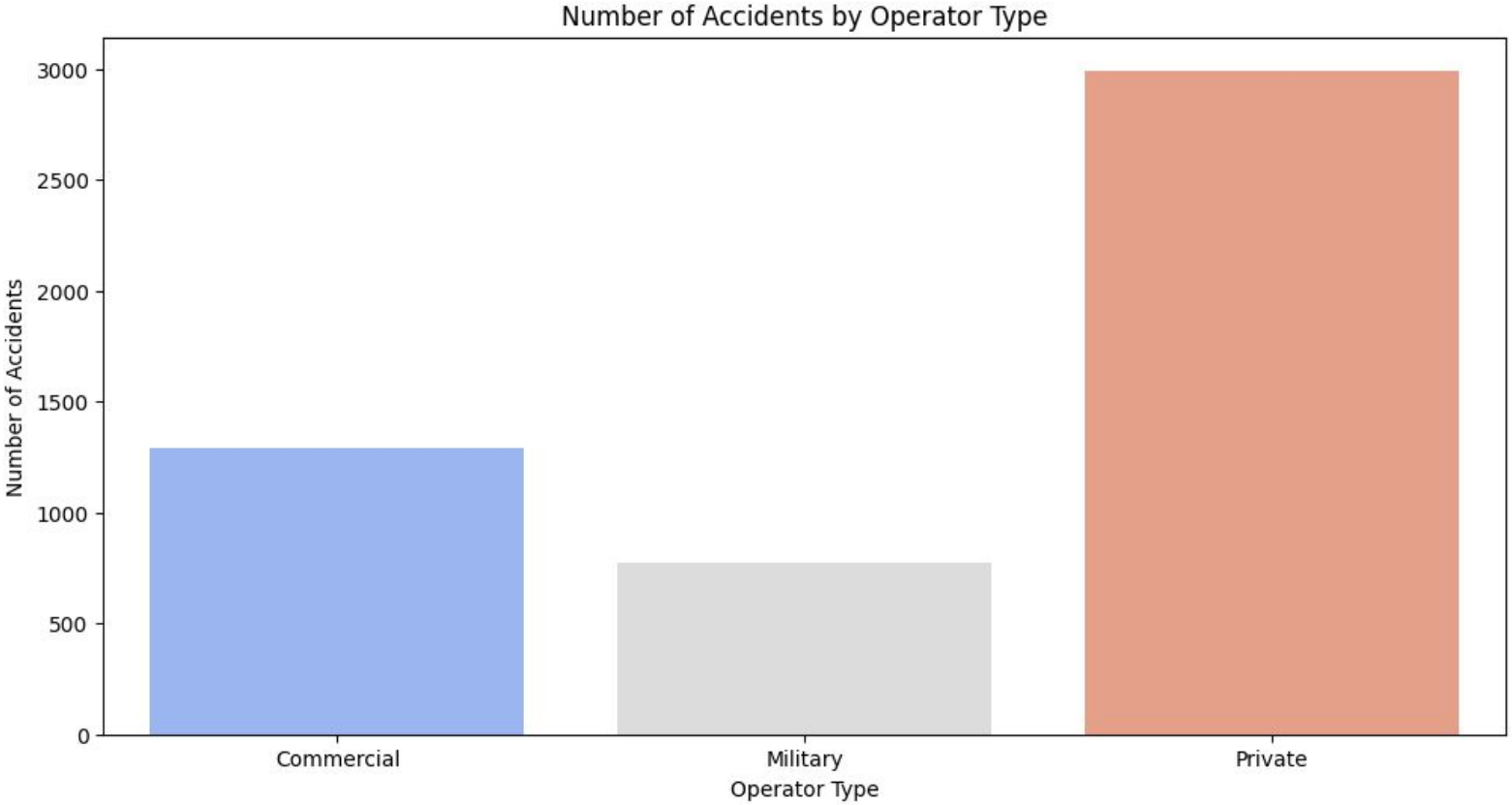
Accidents by aircraft model

Airplane models have caused the most crashes:

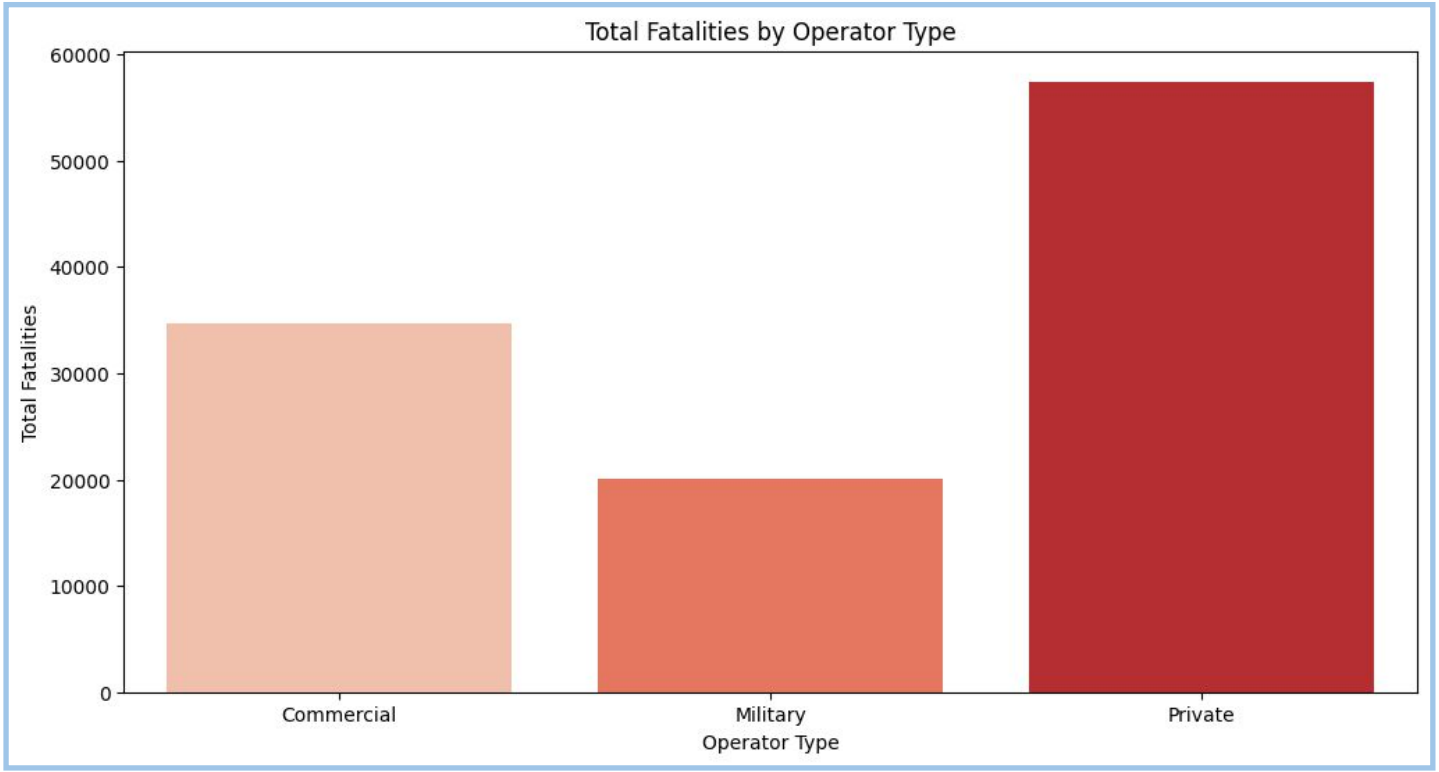
Douglas DC-3	331
de Havilland Canada DHC-6 Twin Otter 300	83
Douglas C-47A	70



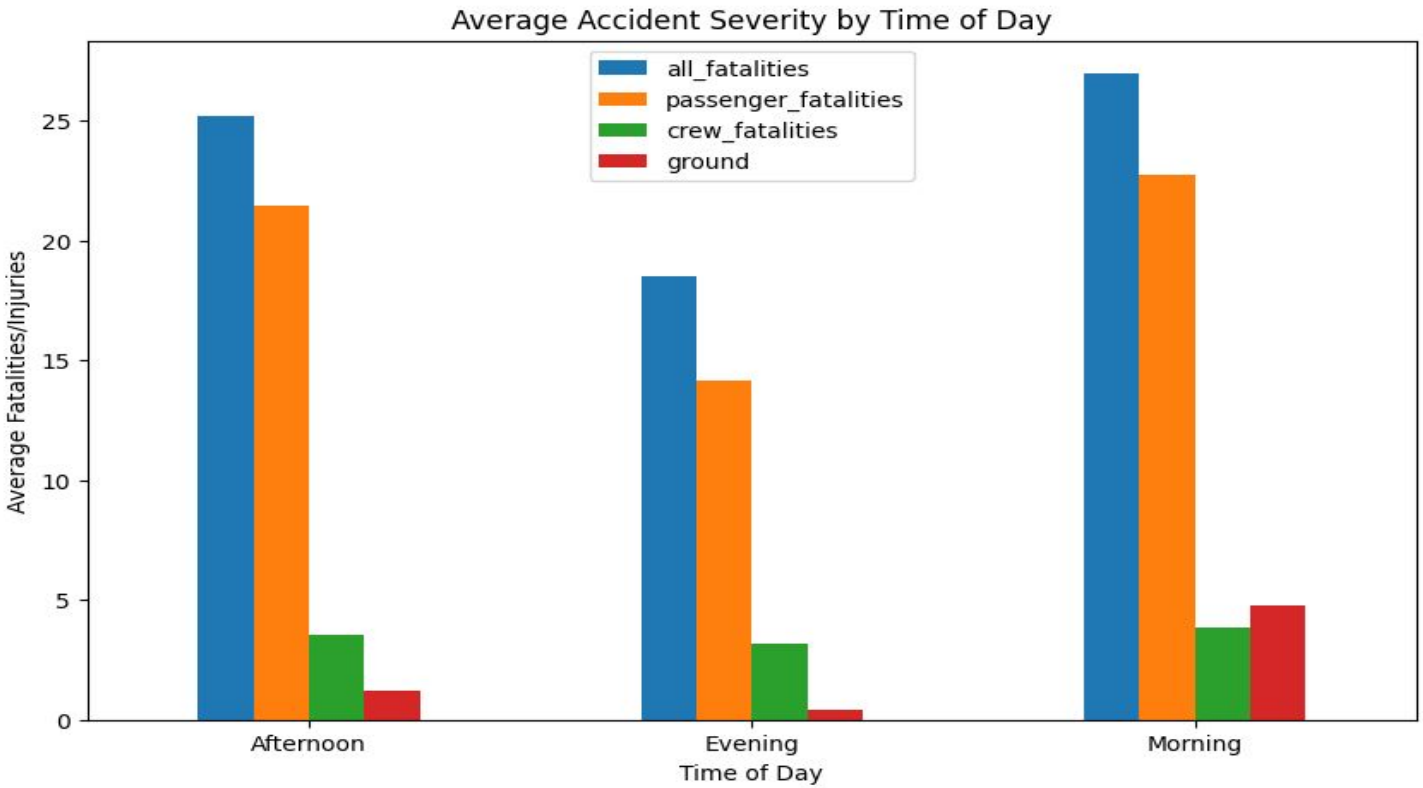
Number Accidents by Operator Type



Total Fatalities by Operator Type



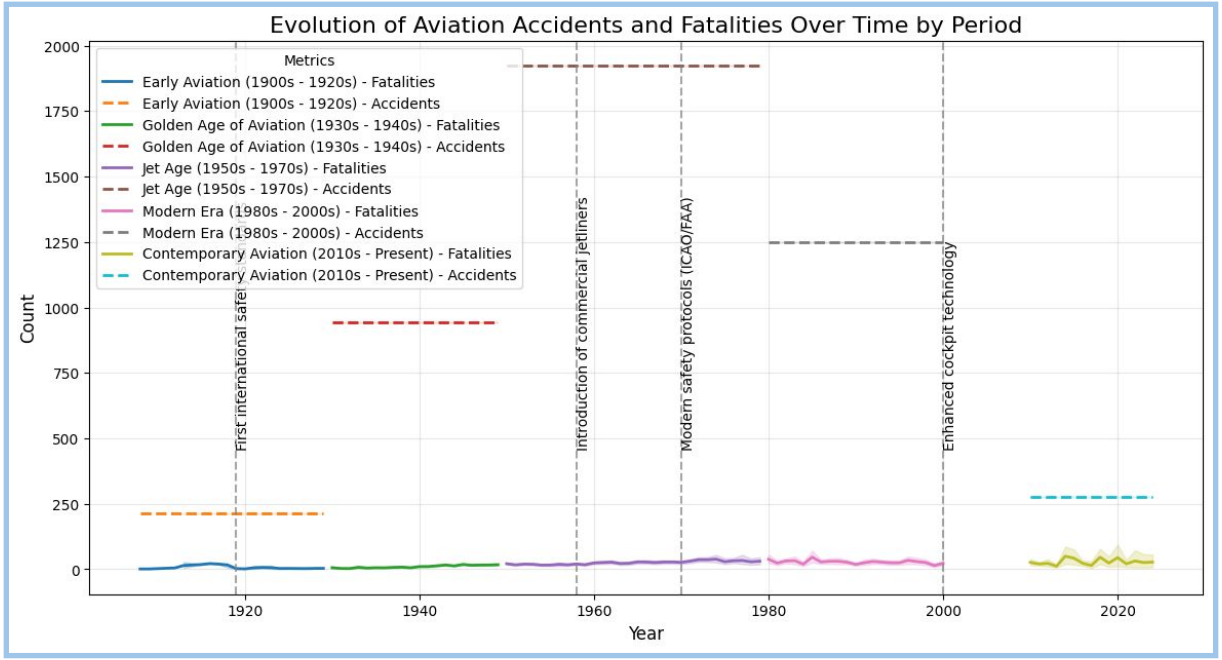
Average Accident by Time of Day



Historical Accident Analysis

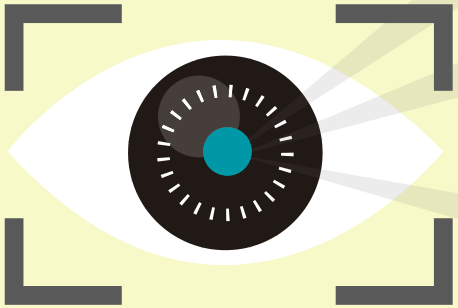
	Fatalities	Accidents
Early Aviation (1900s - 1920s)	1064.0	215.0
Golden Age of Aviation (1930s - 1940s)	10962.0	942.0
Jet Age (1950s - 1970s)	48560.0	1923.0
Modern Era (1980s - 2000s)	34278.0	1250.0
Contemporary Aviation (2010s - Present)	7632.0	278.0

Historical Accident Analysis



Operator Analysis

Operator Type



- Private
- Military
- Commercial

Operator Analysis

Aviation accidents categorized into Military, Commercial, and Private.

Private aviation has the highest number of accidents.

Most crashes: Douglas DC-3 (331 crashes), DHC-6 Twin Otter 300 (83 crashes)

Least crashes: Boeing 737-8ASWL (1 crash), Vickers Viscount 759D (1 crash)

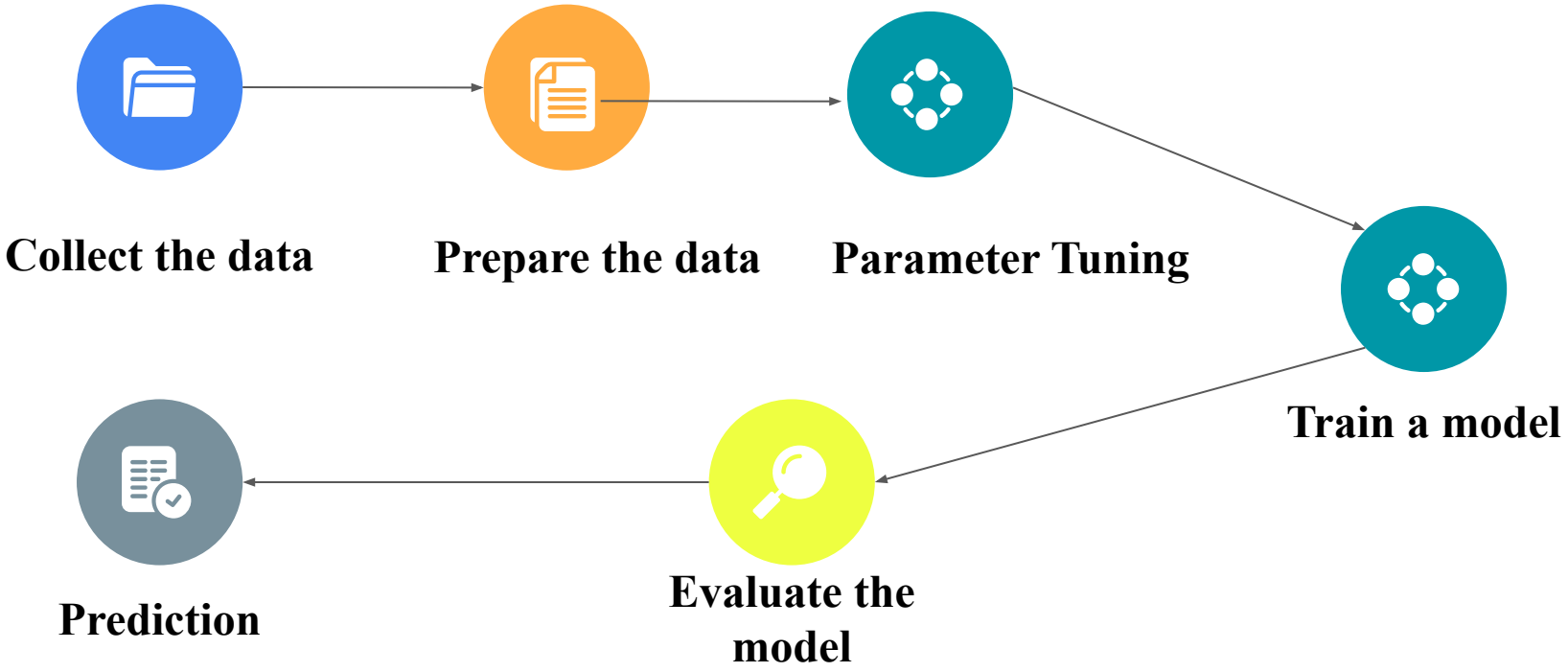
Commercial aviation has stricter safety regulations.

Military aviation has fewer accidents but higher fatality rates per crash.

Operator Analysis

Operator Type	Total Fatalities	Avg Fatalities	Total Passenger Fatalities	Avg Passenger Fatalities	Total Crew Fatalities	Avg Crew Fatalities	Total Ground Fatalities	Avg Ground Fatalities	Total Accidents
Commercial	34,680	26.88	29,944	23.21	4,412	3.42	3,224	2.50	1,290
Military	20,110	26.12	13,500	17.53	3,717	4.83	1,227	1.59	770
Private	57,453	19.20	47,655	15.93	9,104	3.04	4,298	1.44	2,992

Predicting Operator Type – Machine Learning



Machine Learning Algorithms

01

Logistic Regression

It is used to calculate or predict the probability of a binary (yes/no) event occurring.

02

Random Forest

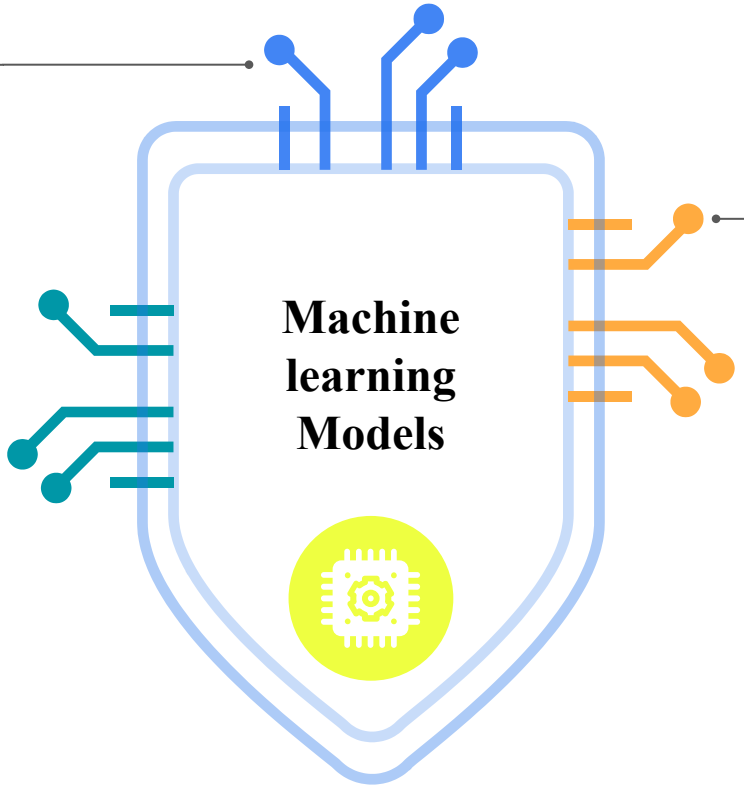
The algorithm searches for the best feature among a random subset of features, instead of searching for the most important feature

04

KNN

Stores all the available data and classifies a new data point based on the similarity.

Machine Learning Models



01

Random Forest

Accuracy: 62.22%

02

KNN

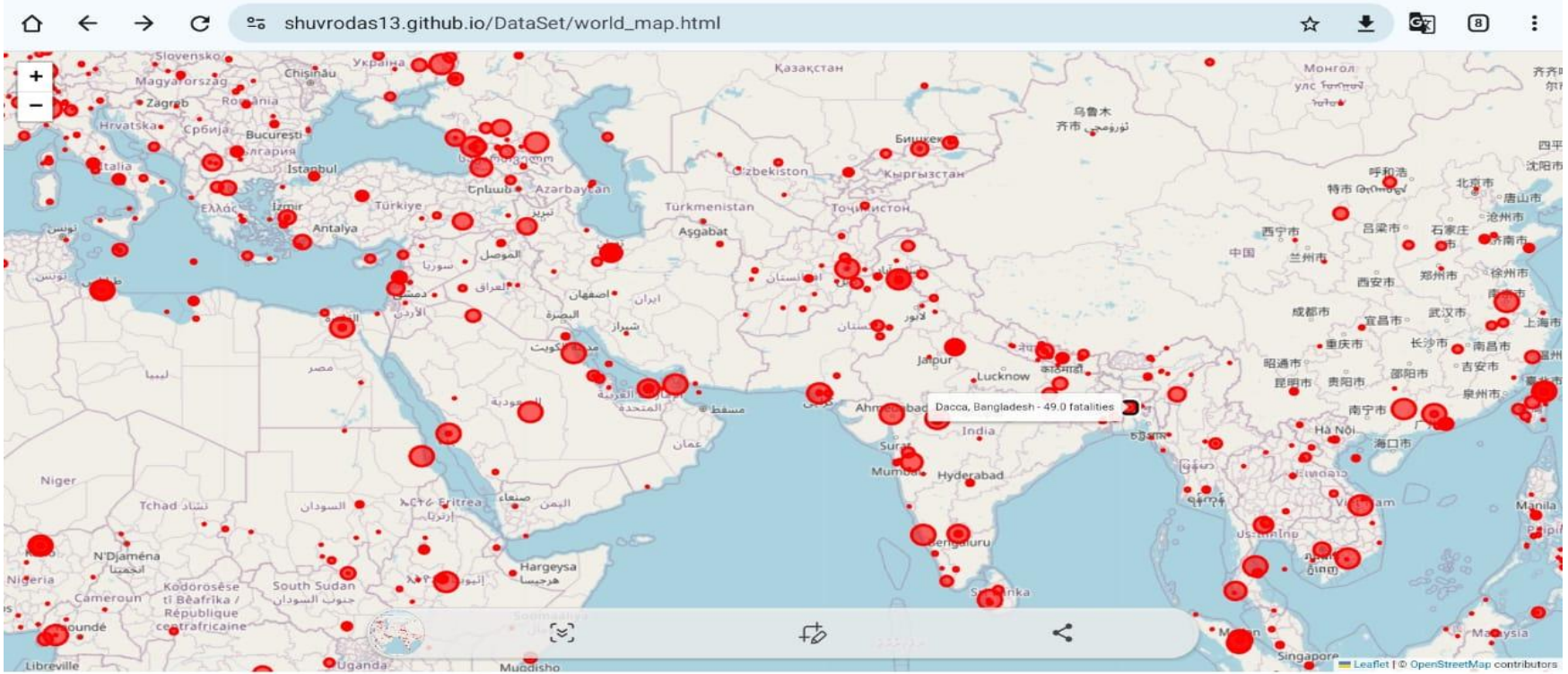
Accuracy: 55.59%

03

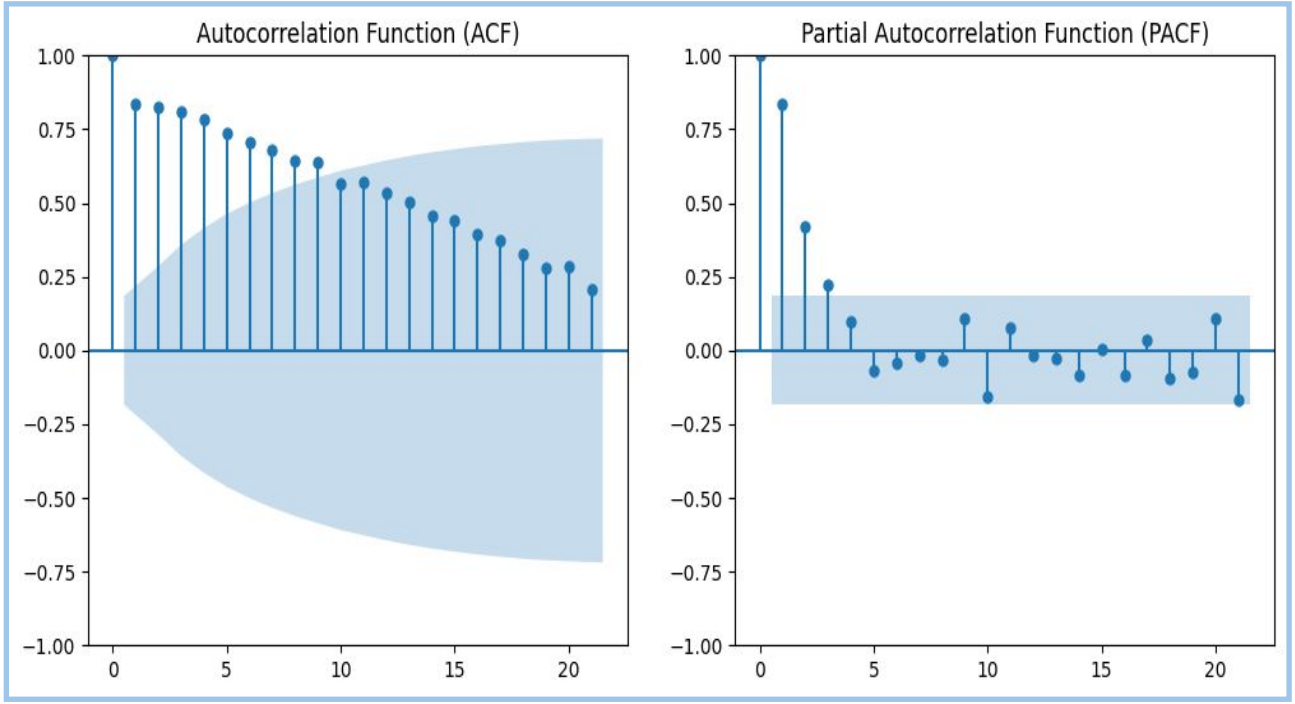
Logistic Regression

Accuracy: 62.02%

Geospatial analysis



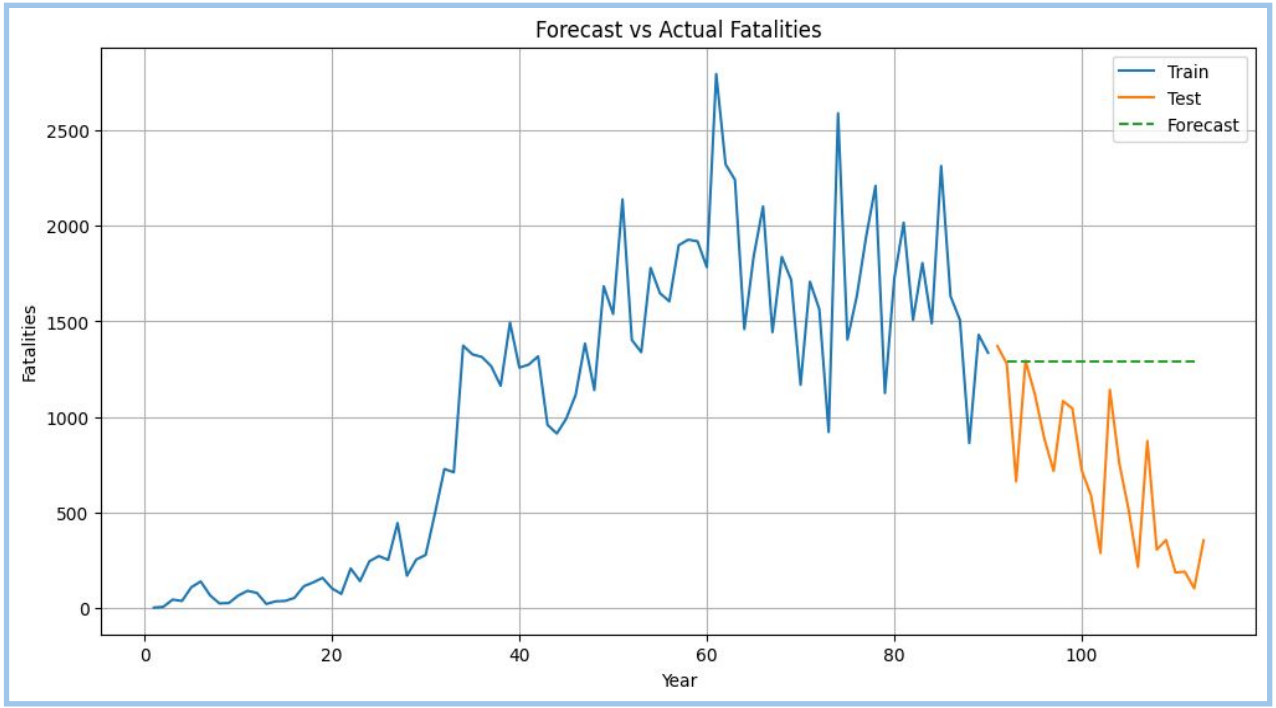
Time Series Analysis



Time Series Analysis

- Aviation fatalities analyzed over 113 years.
- Used ARIMA (0,1,2) for forecasting.
- Trends indicate a steady decline in accidents.
- Forecasting helps predict future accident rates.

Time Series Analysis



Time Series Analysis

- ARIMA model predicts a stable trend in fatalities.
- Test set shows some deviations.
- Continued monitoring needed for safety improvements.
- Unexpected events could still impact future trends.

Key Findings

- Aviation accidents have decreased significantly over time.
- Private aviation has the highest accident count.
- Military crashes are fewer but more severe.
- Technological advancements & regulations improved safety.

Challenges & Limitations

- Data limitations: Some missing or incomplete records.
- Class imbalance affected machine learning predictions.
- Geospatial data issues for certain accident locations.
- Unexpected factors (e.g., terrorism, human error) not fully modeled.

Conclusion

- Aviation safety has evolved with fewer fatalities today.
- Private aviation needs better regulations and safety measures.
- Machine learning models can assist in accident risk assessment.
- Future research should explore human error factors in accidents.

Thank you