

KawiraE / Civil_Aviation_Analysis

<> Code

Issues

Pull requests

Actions

Projects

Security

Insights

Settings

0 stars

0 forks

1 watching

Branches

Activity

Tags

Public repository

1 Branch

0 Tags

Go to file

t

Go to file

+

Add file

Code

KawiraE

Updated README with [the changes made in the file]93a3946 · 2 hours ago🕒

.gitignore	Initial commit	3 days ago
AviationData.csv	initial commit	3 days ago
README.md	Updated README with [the change...	2 hours ago
USState_Codes.csv	initial commit	3 days ago
cleaned_av_data.csv	added files	4 hours ago
index.ipynb	Updated README with [the change...	2 hours ago

README

Civil_Aviation_Analysis

Business Overview

The aviation industry is one of the most crucial sectors in global transportation, connecting businesses, individuals, and economies across vast distances. However, safety remains a critical concern, with aviation accidents posing significant risks to operations, passengers, and reputation. By analyzing data spanning over six decades (1962–2023), this project aims to provide actionable insights that can support organizations in mitigating risks and enhancing decision-making.

This analysis focuses on key factors such as accident trends over time, geographical accident hotspots, the impact of weather conditions, aircraft types, and flight phases associated with varying severity levels. These insights are intended to empower stakeholders in both commercial and private aviation sectors to make data-driven decisions, improve safety measures, and promote sustainable growth in their operations.

With safety at the core of this study, the findings presented here are designed to guide strategic actions, ensuring long-term resilience and reliability in a dynamic and rapidly evolving aviation landscape.

DATA PREPARATION

To ensure the aviation accident analysis is accurate, insightful, and ready for visualization, thorough data preparation was carried out. The following steps were undertaken:

1. Data Reading: The dataset was loaded into Python using the Pandas library and examined for a preliminary understanding of its structure and content.
2. Data cleaning : duplicate entries were eliminated, data types were adjusted, and outliers were filtered out to maintain accuracy.
3. Addressing Missing Data: Missing entries were managed by employing techniques like mean substitution for numerical values and categorizing missing categorical data as 'Unknown.'
4. Data Grouping: The data was organized by aircraft model, year, and accident causes to identify trends and glean important insights.

Data Reading Overview

The data reading phase focuses on importing and exploring the aviation accident dataset to understand its structure and contents. This involves:

- Importing the Data: Loading the dataset from sources like CSV, Excel, or databases into Python or Tableau for analysis.
- Initial Exploration: Inspecting the dataset's shape (number of rows and columns) to understand its size. Viewing the first few rows to familiarize yourself with the data structure.
- Summary Statistics: Generating descriptive statistics to understand numerical columns, such as averages and ranges for variables like Total_Injuries or Year.
- Missing and Inconsistent Data: Identifying missing values in critical columns like Phase_of_Flight or Weather. Checking for inconsistencies, such as unexpected values or incorrect data types (e.g., dates stored as strings).
- Understanding Relationships: Observing relationships between key variables to identify potential trends or insights worth visualizing.

DATA CLEANING

Data cleaning is a fundamental process in preparing the dataset for analysis. This step involves identifying and correcting any issues within our dataset to ensure its reliability and validity. The aviation accident dataset, spanning incidents from 1962 to 2023, will undergo various cleaning techniques to address missing values, inconsistencies, and other data quality issues.

Here's a detailed breakdown of the steps involved in this process:

1. Identifying and Handling Missing Values

Assessment: Each column was analyzed to determine the extent of missing values. Columns with excessive missing data, such as those with over 30-40% missing values, were reviewed to evaluate their importance to the analysis. Strategies Used:

- numerical columns (e.g., fatalities or injuries), missing values were replaced using the mean or median, depending on the distribution of the data. This ensures that outliers do not distort the imputation process.
- categorical columns (e.g., cause of accident, aircraft type), missing values were replaced with "Unknown" to maintain the context of the dataset without introducing bias.

2. Identifying and Removing Duplicate Records if any.

Duplicate rows can occur due to data entry errors or redundant records in the source dataset.

3. Standardizing Data Types

Ensuring correct data types for each column is essential for seamless processing and analysis: Dates: The "Date" column was converted to a datetime format, enabling accurate time-based analysis, such as accident trends over decades. Numerical Fields: Columns such as "Fatalities" and "Injuries" were confirmed as float or int types to allow calculations and aggregations. Text Columns: Categorical fields (e.g., "Aircraft Type" or "Cause of Accident") were standardized by normalizing case (e.g., converting all text to lowercase) and stripping unnecessary whitespace or punctuation.

4. Addressing Inconsistencies

Inconsistencies in the data, such as variations in formatting or naming conventions, will be identified and corrected: Geographical Data: The supplementary USState_Codes.csv file was utilized to map state abbreviations (e.g., "CA") to their full names (e.g., "California"), ensuring consistency in the "Location" column. Categorical Data: Variants of the same category (e.g., "Boeing 737" vs. "Boeing-737") were consolidated into a single, standardized format.

5. Detecting and Handling Outliers

Outliers—values that deviate significantly from the norm—can distort the analysis if not addressed: Numerical Columns: Columns like "Fatalities" and "Injuries" were assessed for extreme values. Outliers were either retained if they were deemed relevant (e.g., rare but valid incidents) or flagged for exclusion if they were likely errors. Approach: Boxplots and histograms were used to visualize distributions and identify potential outliers.

6. Final Validation

Once the data cleaning steps were completed, the dataset underwent thorough validation to ensure its readiness for analysis: Summary Statistics: The .describe() function was used to generate statistical summaries, confirming the accuracy and consistency of numerical data. Spot Checks: Random samples of the dataset were inspected to verify that missing values were appropriately handled, duplicates were removed, and data types were consistent. Shape Check: The .shape attribute was used to confirm the final number of rows and columns in the cleaned dataset, ensuring alignment with expectations.

7. Preparing for Visualization

To streamline the visualization process, the cleaned dataset will be aggregated where necessary. Key groupings include:

- Grouping by aircraft model to analyze accident rates.
- Summarizing by cause of accident to identify the leading factors.
- Aggregating by year to observe trends over time.

OBJECTIVES OF DATA CLEANING

- Identify missing values in a dataframe using built-in methods
- Explain why missing values are a problem in data science
- Determine and implement the optimal technique for dealing with missing, duplicate, and erroneous values in a given dataset.

Data Visualization Introduction

The data visualization phase transforms raw aviation accident data into meaningful insights through a range of compelling plots. By analyzing trends over time, geographical patterns, weather impacts, and flight phases, this section sheds light on key safety challenges. These visualizations aim to support stakeholders in identifying high-risk factors and formulating strategies to enhance aviation safety.

OVERVIEW

The visualization phase of this project was designed to uncover insights and trends within the aviation accident dataset, using a range of compelling plots to analyze various aspects of the data. Below is an overview of the visualizations created and their corresponding insights:

1. Number of Accidents Over Time:

- A line plot revealed temporal trends, showing fluctuations in accidents over decades and indicating potential improvements in aviation safety.

2. Distribution of Accident Severity:

- A bar plot highlighted the frequency of accidents categorized as minor, serious, or fatal, offering insights into the severity of aviation incidents.

3. Accidents by Day of the Week:

- A bar plot explored weekday patterns, identifying the days with the highest accident occurrences.

4. Most Common Aircraft Types:

- A horizontal bar plot showcased the top aircraft models involved in accidents, helping pinpoint frequently involved designs.

5. Geographical Distribution of Accidents:

- An interactive heatmap visually emphasized high-risk regions, providing a geographical perspective on accident density.

6. Types of Injuries vs. Phase of Flight:

- A stacked bar plot illustrated the distribution of injury severity (fatal, serious, minor) across flight phases, highlighting high-risk stages like takeoff and landing.

7. Accidents by Number of Engines:

- A bar plot analyzed accident proportions by engine type, shedding light on risks associated with single-engine, twin-engine, and multi-engine aircraft.

8. Impact of Weather Conditions:

- A scatter plot revealed the relationship between weather conditions and accident fatalities, identifying the weather scenarios with the highest risks.

9. Manufacturer vs. Number of Engines:

- A stacked bar plot uncovered trends across manufacturers and engine types, helping stakeholders assess which combinations are more prone to accidents.

Insights from the Visualization Phase

- Identified high-risk conditions and phases (e.g., landing and takeoff).
- Revealed geographical regions with frequent accidents and their possible contributing factors.
- Pinpointed aircraft types, manufacturers, and weather conditions associated with higher risk.
- Provided actionable insights to guide improvements in aviation safety protocols, training, and operations.

These visualizations played a key role in transforming raw data into meaningful insights, equipping stakeholders with a deeper understanding of aviation safety dynamics and areas for improvement.

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

● Jupyter Notebook 100.0%