Project Report

Project: Exploratory Analysis of Airplane Crashes and Fatalities Dataset

Dhiraj H. Gawhare
Roll Number: 22M0062
M. Tech (Aerodynamics)
IIT Bombay.

Project Report: Exploratory Analysis of Airplane Crashes and Fatalities Dataset

Introduction

Greetings from the Kaggle bot! This is an automatically-generated kernel with starter code demonstrating how to read in the data and begin exploring the "Airplane Crashes and Fatalities Since 1908" dataset. The purpose of this project is to conduct an exploratory analysis of historical airplane crashes and fatalities, utilizing data visualization techniques to uncover insights and trends.

Exploratory Analysis

The exploratory analysis is carried out in a step-by-step manner, using Python libraries such as **matplotlib**, **numpy**, **pandas**, and **seaborn** for data manipulation and visualization. The analysis includes the following key steps:

1. Importing Libraries and Defining Functions

The project begins by importing necessary libraries for data analysis and visualization. Custom functions are defined to facilitate the creation of distribution graphs, correlation matrices, and scatter/density plots. These functions will aid in presenting the data in a visual and comprehensible manner.

2. Dataset Overview and Access

The dataset under examination is the "Airplane Crashes and Fatalities Since 1908" dataset. By utilizing Python's file handling and **pd.read_csv()** function, the dataset is loaded into the analysis environment. The initial steps involve accessing the dataset, understanding its structure, and identifying the attributes it contains.

3. Data Visualization Functions

Several data visualization functions are introduced to enhance the analysis process. These functions are designed to provide insights into attribute distributions, correlations between attributes, and relationships between numerical columns.

4. Data Distribution Analysis

The analysis involves using the defined distribution plotting function to generate histograms and bar graphs. These visualizations offer insights into the frequency distribution of attributes such as crash

years, fatalities, and locations. The histograms help to identify patterns and trends related to historical airplane crashes.

5. Correlation Matrix

A correlation matrix is computed and visualized to assess the relationships between numerical attributes. The matrix provides insights into potential correlations between attributes, helping to identify attributes that may have a significant impact on each other.

6. Scatter and Density Plots

The analysis includes creating scatter plots and density plots to explore relationships between pairs of numerical attributes. These plots are particularly useful for uncovering trends, clusters, and potential outliers within the data.

Conclusion

In conclusion, the exploratory analysis of the "Airplane Crashes and Fatalities Since 1908" dataset has provided valuable insights into historical airplane crash and fatality data. The analysis involved importing, accessing, and visualizing the dataset's attributes using Python libraries and functions. The visualizations of attribute distributions, correlation matrix, and scatter/density plots have contributed to a deeper understanding of historical trends and patterns related to airplane crashes.

This exploratory analysis serves as a foundation for further research and analysis, potentially leading to more advanced studies using machine learning and predictive modeling techniques. The insights gained from this analysis can contribute to aviation safety initiatives and better understanding of historical trends in aviation accidents.

Acknowledgements

This analysis was conducted using Python programming and various data visualization libraries. The "Airplane Crashes and Fatalities Since 1908" dataset was sourced from Kaggle, and we acknowledge the contributions of the dataset creators and the Kaggle community for making this analysis possible.