Aviation Accident Analysis Using CRISP-DM Methodology

Data Mining Approach for Aviation Safety and Risk Assessment

Summary

This Dataset was collected due to increased Aircraft accidents. The goal of this dataset is to reduce the aircraft accidents. I've analysed this dataset by first getting rid of all the columns with over 50% null values, getting rid of all the rows without united sates since that is where most accidents took place. I analysed it to a point where I was able to know the aircraft models that caused most accidents in a certain year. I also realised that there are three Aircraft that don't have the the total fatal injuries and total serious injuries, they only have one minor injury.

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

The problem identified in this dataset is that there is a mechanical related issue whereby in some airplanes the number if engines are less, there are weather related problems and some aircraft models are at a very high risk.

Data Overview

The source of this data is Kaggle.

The dataset has a historical record of 1982 to 2023.

The purpose of this dataset is to find solutions for the increased aircraft accidents.

This dataset has 88889 rows and 31 accidents.

There are Numerical data, Categorical data, Date/Time data, Text data and Boolean where there are fatal and nonfatal.

Methods

I first checked the duplicates, the null values, then dropped the null values.

I later decided to drop all columns with over 50% missing values, dropped records where no accidents in united states.

Then added a day, month and year column, merge the same airport names together, merge same registration numbers, merge different capitalization of make together and split location in city and state.

Then categorize the amount of injuries as this is already in another column.

Then drew a line graph where x = years and y = Event.Id, a bar graph of x = number of months and y = years, another on with the number of accidents and then the last one is the model of the aircraft that caused accidents in a certain year.

Results

I realized that there are specific models that cause more accidents and there are three model that have zero Total Fatal Injuries, zero Total Serious Injuries and one Total Minor Injury.

The Model are

Model	Make
120A	Lindstrand Balloons Usa
RICHARDSON-75	Richardson Andrew P
AX8 105S2	Thunder & Colt

Conclusions

Present your conclusions about the project here. Can include business recommendations, project limitations, and/or future improvement ideas

Thank You!

Email: jacklinewaceke199@gmail.com
GitHub: @JacklineWaceke
LinkedIn: Jackline Waceke