PHASE 1 PROJECT.

AVIATION RISK AND ANALYSIS PROJECT.

OVERVIEW.

In this project it deals with a company that wants to go into a new business and start a new with different goals and expand there network. There main objective is to get into the aviation sector of the business world and expand there influence majorly on buying aeroplane and using them for purposes. Mainly Private, military and Commercial purposes. Featuring categorized data on aircraft types, causes, and outcomes, it provides insightful synopses and lessons for the start of the business. Thus due to the eagerness of the business to start they overlooked the risks accompanied by the airline business thus poor decision making. My project aims to look at the risks in the business and also easily search and filter information, supported by interactive visualization tools. This initiative enhances safety standards, serves as an educational resource, and facilitates research in aviation safety. By ensuring accurate data collection and ongoing maintenance, the project fosters a culture of safety and informed decision-making in the aviation industry.

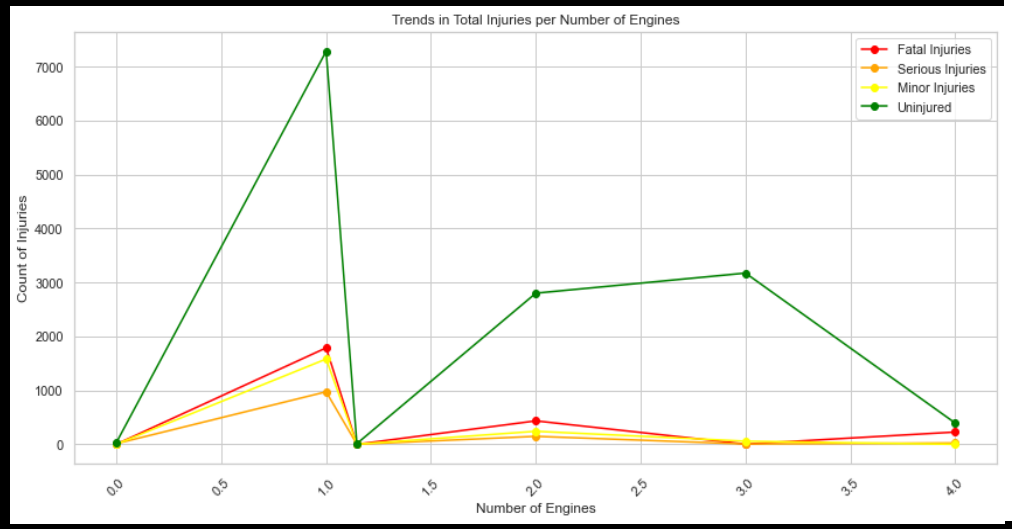
BUSINESS UNDERSTANDING.

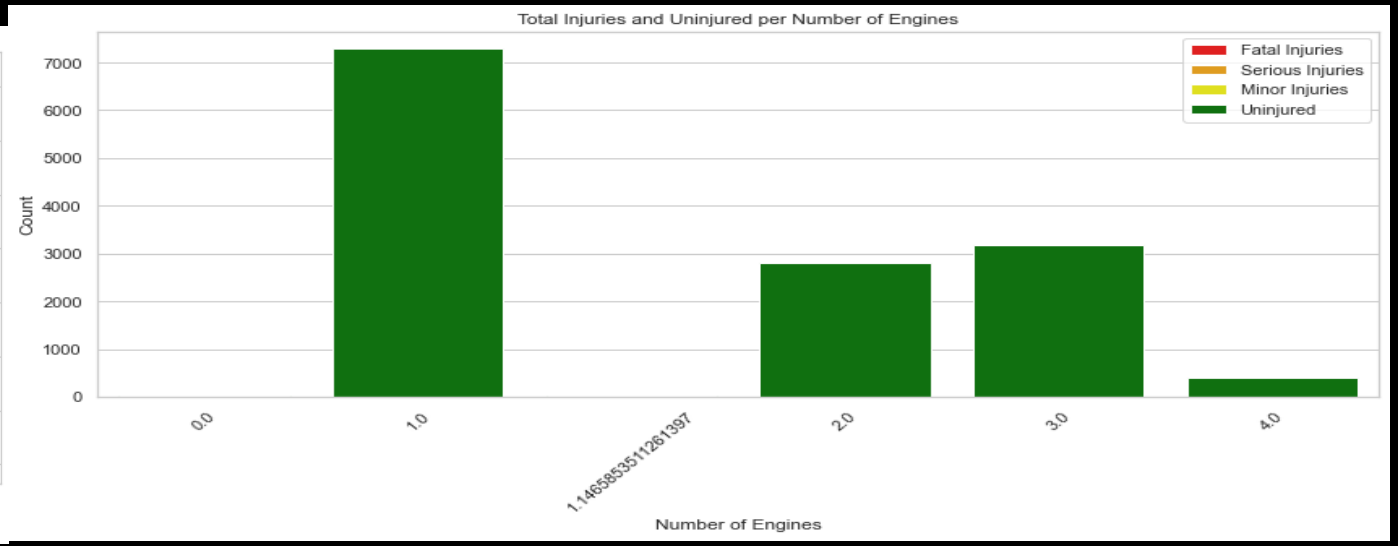
The Business at hand deals with airplanes by which is a large area of expertise and needs analysis. Targeting aviation professionals, researchers, regulatory bodies, and the general public, the database addresses the critical need for accessible information to identify trends and preventative measures. Its value lies in offering detailed synopses and categorized data that inform decision-making and improve safety protocols. Potential revenue streams include subscription services, partnerships with aviation organizations, and data licencing. The companies selling the airlines, the airports they will be operating in , the government of the country of base operations the crew workers and the passengers are all stakeholders in the business.

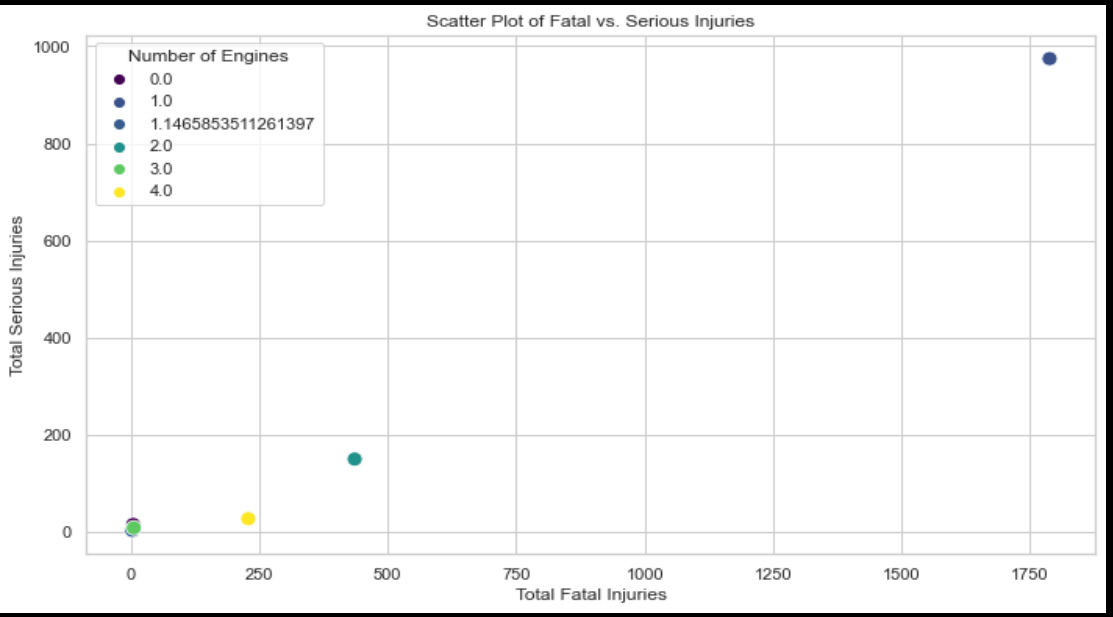
DATA UNDERSTANDING.

- For my project, I am working with a dataset on Aviation Accident Database & Synopses, up to 2023 obtained from Kaggle, where it is frequently updated and made accessible for research purposes. The data is stored in CSV format within a folder named "data.“ In my Github repository.
- It comprises 90,348 rows and 31 columns, where each column represents a distinct field and each row corresponds to an individual aircraft accident investigation entry. The dataset features a combination of continuous and categorical data, with 26 of the columns (approximately 84%) being categorical, stored as object types. The remaining 5 columns are continuous, represented in float form.
- This dataset focuses specifically on aircraft accidents, providing a rich foundation for analysis. I was able to find the sums of the missing values in the data and also the percentages of the missing values thus dropped some of the columns for efficiency and continuation on my project.
- After dropping the columns the shape of the data changed to 90,348 rows and 26 columns. Moving forward, I plan to explore the unique values and distributions of the categorical columns, analyze trends in the continuous data, and prepare for necessary data cleaning and exploratory data analysis to uncover valuable insights.

DATA ANALYSIS.







1) The line plot displays trends in total injuries across different engine counts over time or ordered categories. Each line represents a different injury type, allowing viewers to see how the counts change as the number of engines increases. This visualization is particularly effective for illustrating trends and comparisons, making it easy to identify which injury types are rising or falling in relation to the number of engines.

2) The bar plot displays the total counts of different types of injuries (Fatal, Serious, Minor, Uninjured) for each category of the number of engines. This visualization allows for straightforward comparisons across categories, making it easy to identify which engine counts are associated with higher or lower injury totals. The use of color differentiation enhances clarity, helping in quickly discern between the different injury types.

3) The scatter plot showcases the relationship between fatal and serious injuries, using points to represent data for different engine counts. This visualization allows one to observe trends and patterns in the relationship between these two variables. The color coding based on engine counts helps identify whether certain engine categories are associated with specific injury types, making it easier to spot clusters or trends.

RECOMENDATION

Increased Focus on Data Analysis: The aviation industry should increasingly rely on data analytics to predict and prevent potential hazards, using historical data to improve safety measures thus decreasing the airplane risks and increase efficiency.

And also should higher more skillful pilots and invest on good and efficient planes for safety measures.

CONCLUSION.

The conclusion to my project is that Human error remains a leading cause of aviation accidents, often linked to decision-making, fatigue, and inadequate training and to improve that to the stakeholders they should find adequate people for the jobs to come.

THANK YOU NOTE.

I wish to acknowledge all the individuals who assisted me in compiling the research material especially my Technical Mentor for giving me the required guidelines, my family members for their moral support and encouragement.

Technical Mentor – MaryAnn Mwikali.

NAME – Boniface Mutwiri Kimutai.

Linkedin Profile – Boniface Mutwiri.