PROJECT

Aviation Accident Analysis

Problem Definition

Over the years, the aviation industry has witnessed tragic accidents, claiming hundreds of lives due to various errors.
These are primarily caused by factors such as pilot error, weather conditions, mechanical failure, fuel mismanagement, and more. As a result, analysing air accident data
becomes crucial in addressing important questions like how safety measures can be improved, what technological advancements should airline companies consider while designing aircrafts, and which areas and situations can be categorized as high-risk.
Most importantly, this analysis plays a pivotal role in maintaining public trust in air travel.

Objective

The aim of this project is to reveal insights into the main causes of air crashes and other aviation accidents, with a focus on understanding how these can be prevented in future.
Here, we plan to perform analysis on various attributes within the data files by implementing a data pipeline in AWS. This pipeline will merge data files and transform all necessary rows and columns for final analysis.
Some of the columns we will focus on are accident location, number of injuries sustained, weather conditions, and aircraft design details like type, make, model, among other relevant factors.
Some key analytical points we will explore are:
● What areas and weather conditions are more prone to airline accidents?
● Which type of accidents and incidents are common and why?
● What are common technical faults that lead to such tragedies?
● With advancements in aviation technology, have the number of accidents reduced over the
years?
By diving into these attributes, we aim to discover patterns and identify correlations which have
the potential to inform safety and mitigation strategies for similar incidents in the future.

Data Source

The Aviation Accident Database contains information from 1962 and later about civil aviation accidents and selected incidents within the United States, its territories and possessions, and in international waters This dataset has more than 30 columns and 88,890 records providing air accident information over the past 40 years.
The attributes are both continuous and categorical in nature and since there are many missing values, we will be using different techniques to transform the data before performing analysis. Finally, the dataset will be divided into 2 or more files before loading in the pipeline.

Data Cleaning

Before implementing the dataset, data cleaning is perfomed for a few columns in Python to impute null values. The most frequently occurring value (mode) was used for imputation.

```
In [94]: threshold = 0.3
         missing_counts = df.isna().sum()
         total_row = len(df)
         drop_col = [i for i in df.columns if missing_counts[i] / total_row > threshold]
         df.drop(columns=drop_col, inplace=True)
         df.shape

Out[94]: (88889, 22)
```

Fig1. Sample of Column Data Cleaning

```
In [97]: subset_col = list(unique_values[unique_values['missing_values'] > 1000]['Column'])
         df.dropna(subset=subset_col, inplace=True)
         df.shape

Out[97]: (63325, 21)
```

Fig2. Sample of Row Data Cleaning

Tableau Sheets

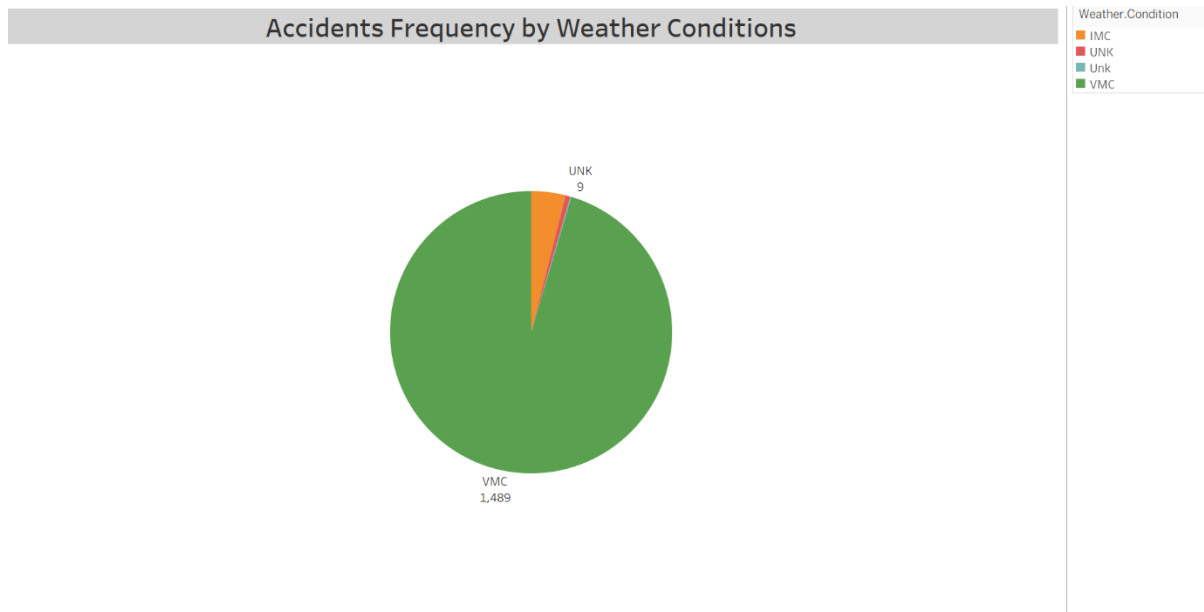This shows the accidents frequency by the weather conditions in the United States.

Fig3. Accidents Frequency by Weather Conditions

Aircraft makes that have the most Frequent Accidents. This data is filtered with the top 10 Aircraft Makes.
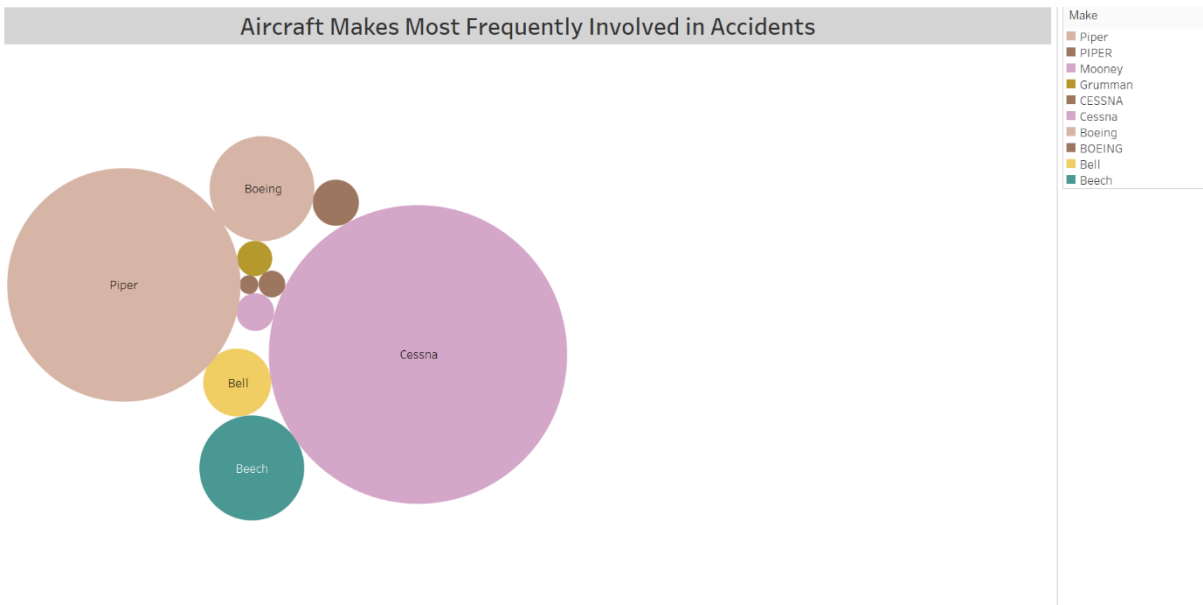

Fig4. Aircraft Makes Most Frequency Involved In Accidents

These are the Number of Fatal and non-Fatal Accidents and are Categorised according to the top 10 States.
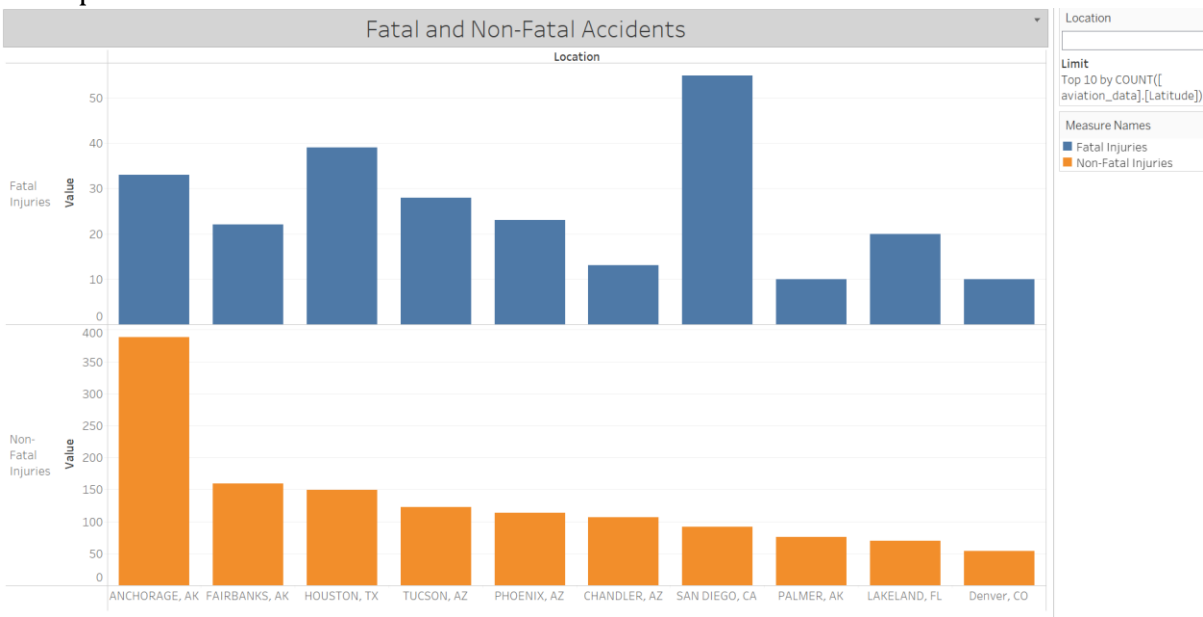

Fig5. Fatal and Non-Fatal Accidents from the top 10 States

This gives the count of number of accidents, Fatal and Non-Fatal Accidents that occur in different types of flight phases: i.e: Approach, Cilmb etc....
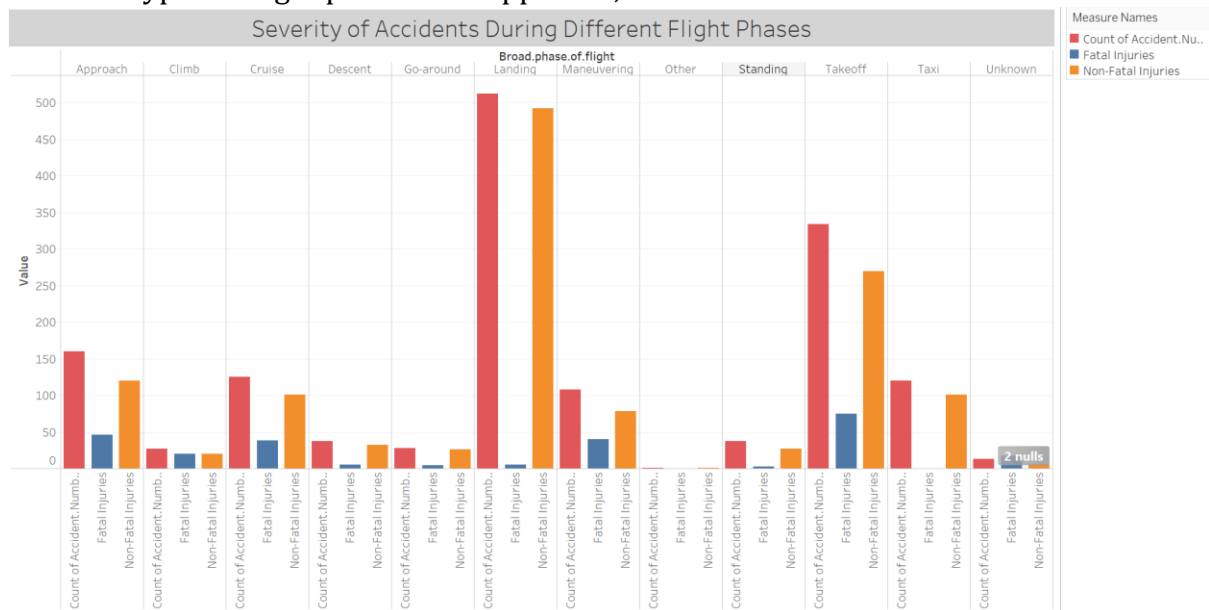

Fig6. Severity of Accidents During Different Flight Phases

The following shows Accidents across the years from the lowest to the highest.
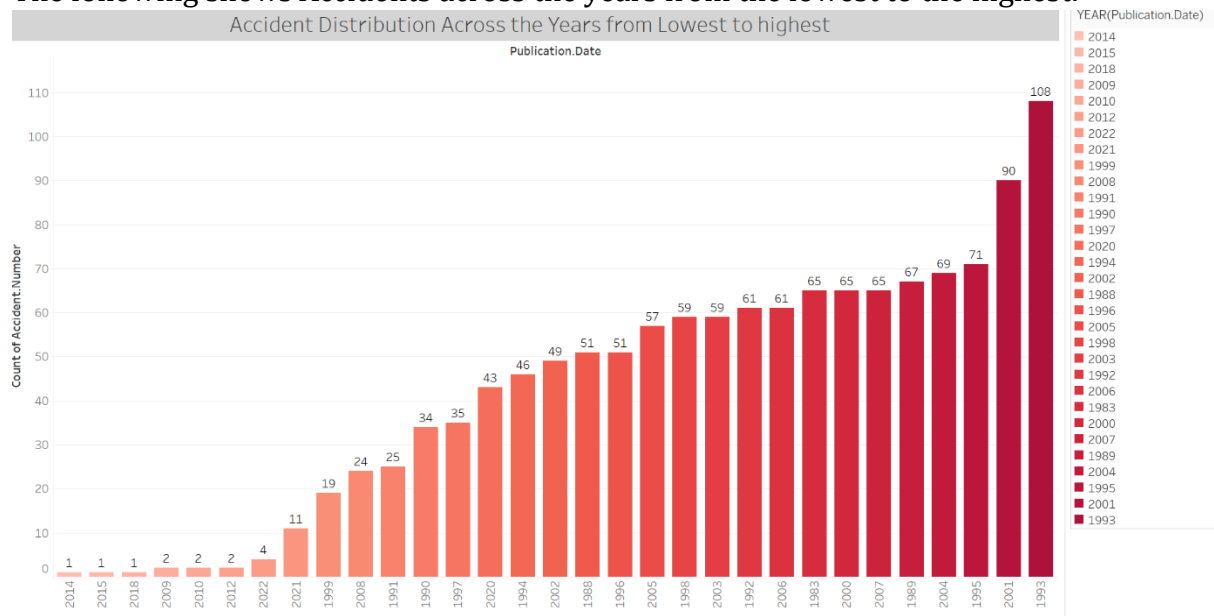

Fig7. Accident Distribution Across the Years form Lowest to the Highest

Tableau Dashboard

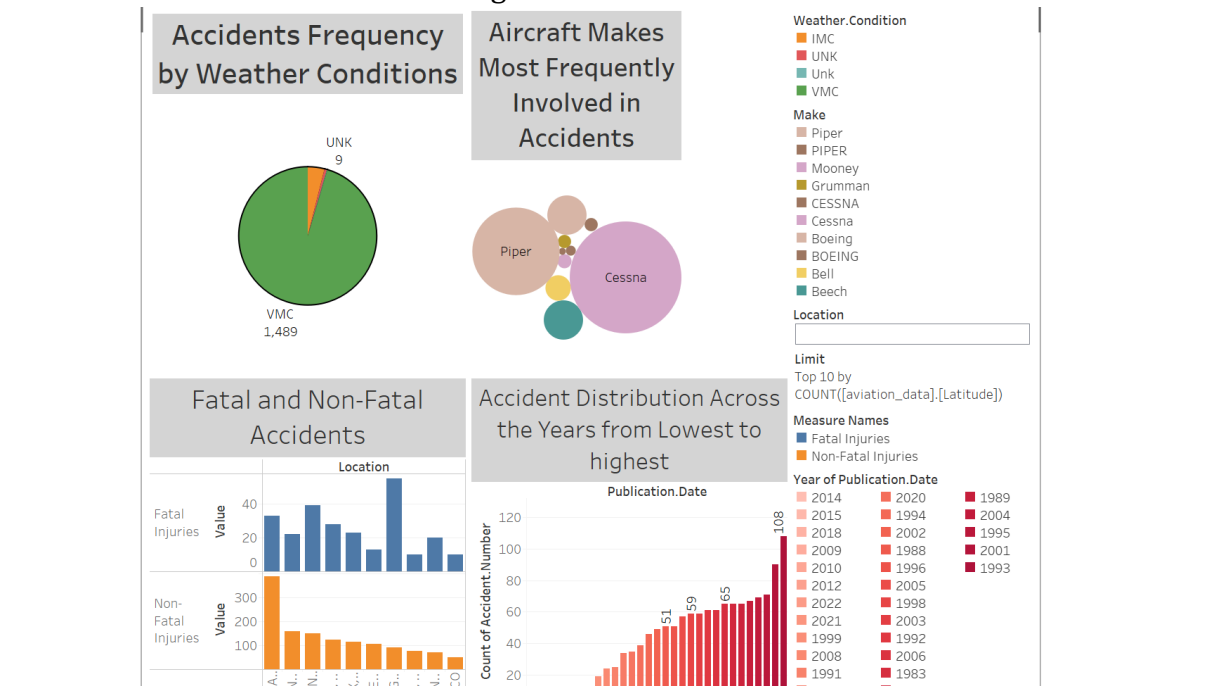This is the Tableau Dashboard that gives visualisation the Data.

Fig8. Tableau Dashboard