ckucewicz / aircraft_safety_project

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

1 star 0 forks 1 watching 1 Branch 0 Tags Activity

Public repository

master 1 Branch 0 Tags Go to file t Go to file Add file Code

ckucewicz Update README.md now

.gitignore	EDA Update	3 days ago
Accidents by Flight Phase.png	EDA Update	3 days ago
Non-fatal_rates.png	presentation visualizations upload	9 minutes ago
Notebook.ipynb	Final Notebook	22 minutes ago
Number of Engines.png	EDA Update	3 days ago
README.md	Update README.md	now
Type of Engines.png	EDA Update	3 days ago
aircraft_safety_cleaned.csv	EDA Update	3 days ago
aircraft_safety_dashboard.pdf	Uploaded Dashboard	13 minutes ago
death_toll.png	presentation visualizations upload	9 minutes ago
deaths_per_flight.png	presentation visualizations upload	9 minutes ago

README

Aircraft Safety Analysis

Author: Chris Kucewicz

Business Understanding

Background:

My company is planning to diversify its porfolio by extending to different markets. More specifically, my company is looking to invest in airplanes in order for their expansion to different markets to be successful.

Business Goals:

The primary focus of this data science project is to assess and mitigate the risks associated with various aircraft models. Specifically, we aim to identify and recommend aircraft types that carry the least potential risk. This information is crucial for making informed decisions regarding aircraft investments and ensuring the success of our company's expansion into new markets.

Business Success Criteria:

The success of this project will be measured by providing three well-supported recommendations on the least risky aircraft models for our company to invest in. For this project, the term least risky refers to types of aircrafts with the least amount of crashes, the least number of casualties, and the lowest fatality rate.

Data Understanding

The National Transportation Safety Board (NTSB) collects data on aviation accidents and incidents that occur in the United States (which include its territories) as well as international waters. This dataset includes 31 features and 88,889 observations or entries.

Each entry in the dataset represents an aircraft involved in an accident (or incident). For each aircraft there is a unique ID associated with the specific accident (or incident) the aircraft was involved in. Additional information is included about each entry, such as the accident (or incident) date, location, and number of injuries, as well as characteristics about the aircraft, such as the make, model, and number of engines.

Data Preparation

In data preparation, I created three functions (`display_df_information()` , `multiple_value_counts()` , and `examine_features()`) to assess data types, NaN values, and entry details. Columns with high NaN values or high cardinality with a uniform distribution were dropped: `Longitude` , `Latitude` , `FAR.Description` , `Schedule` , `Air.carrier` , `Airport.Code` , `Airport.Name` , `Publication.Date` , `Report.Status` , `Registration.Number` , `Investigation.Type` , `Purpose.of.flight` .

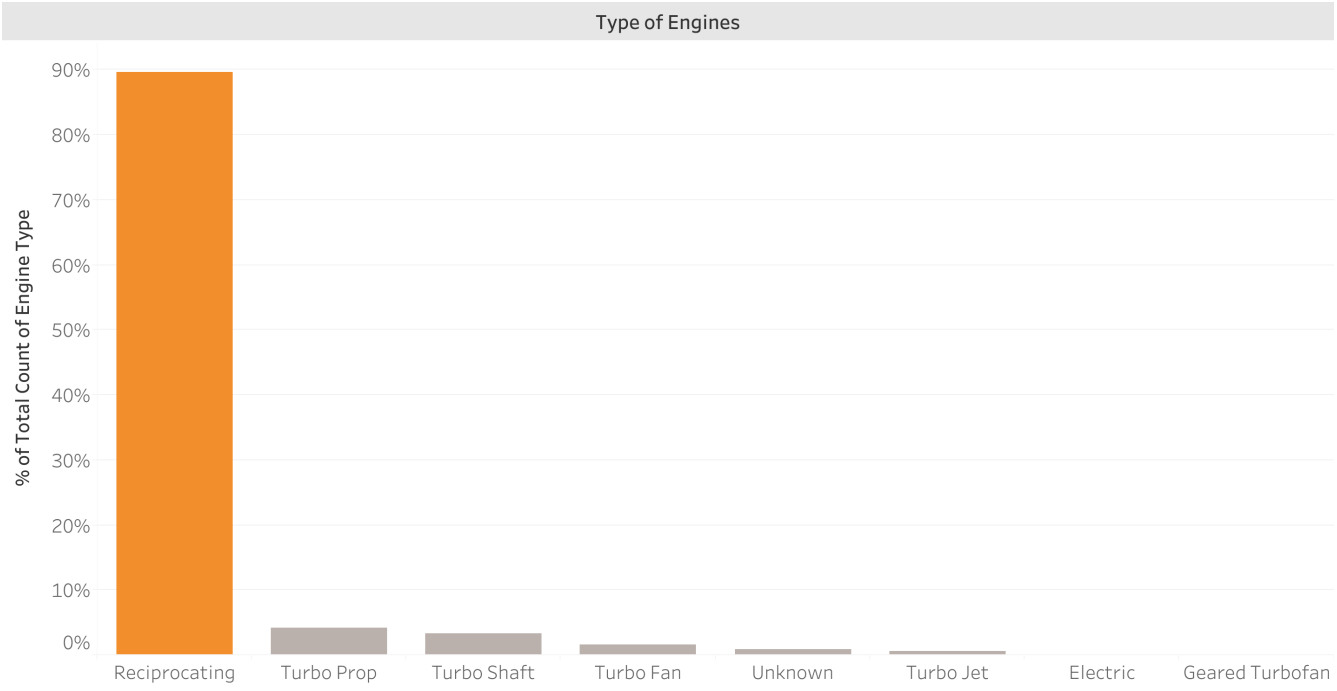
Focused on professionally built airplanes, I filtered rows where `Aircraft.Type` was 'airplane' and `Professionally.Built` was 'yes'. `Injury.severity` values were grouped into 'Non-Fatal', 'Fatal', and 'Unavailable', consolidating data from `Total.Fatal.Injuries` .

Using a lambda function, I standardized case sensitivity and removed duplicates in the Make column. Rows with a small percentage of NaN values were dropped using `.dropna()` .

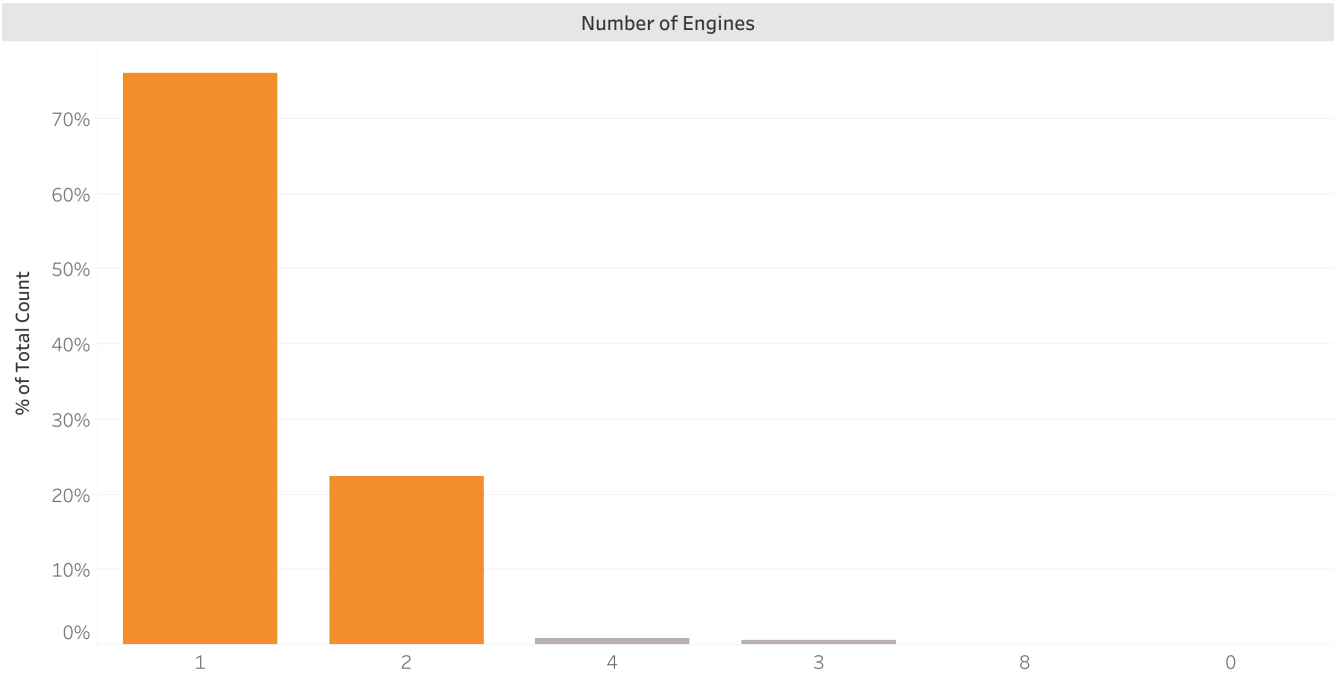
Remaining NaN values in `Total.Fatal.Injuries` , `Total.Serious.Injuries` , `Total.Minor.Injuries` , and `Total.Uninjured` were retained due to their significant count, avoiding noise addition. The cleaned dataframe now has 18 columns and over 65,000 entries.

Exploratory Data Analysis

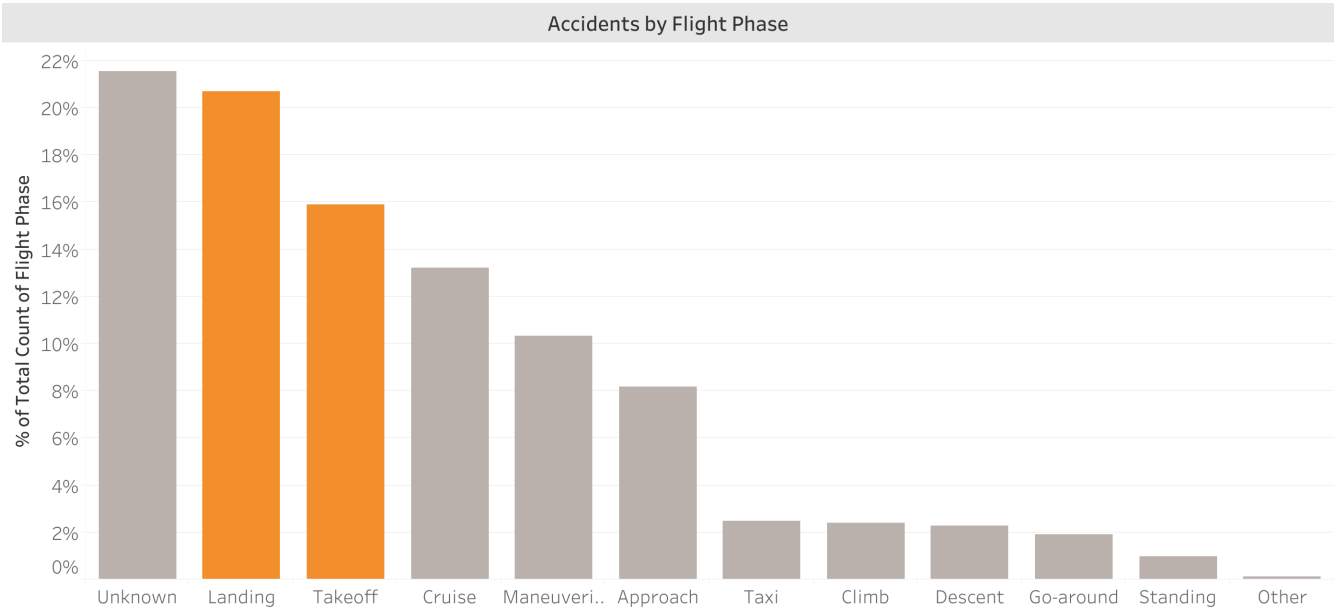
The following are findings from this analysis:



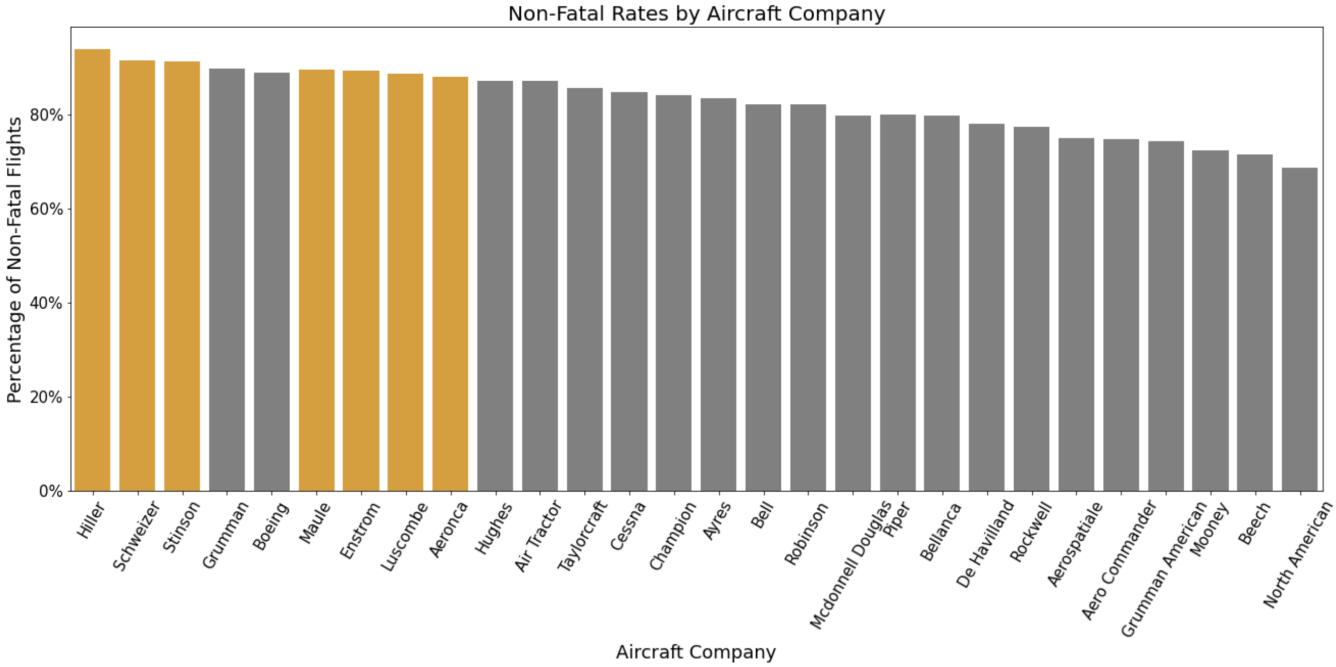
- Almost 90% of reported accidents involved aircraft with a Engine Type classified as 'Reciprocating'

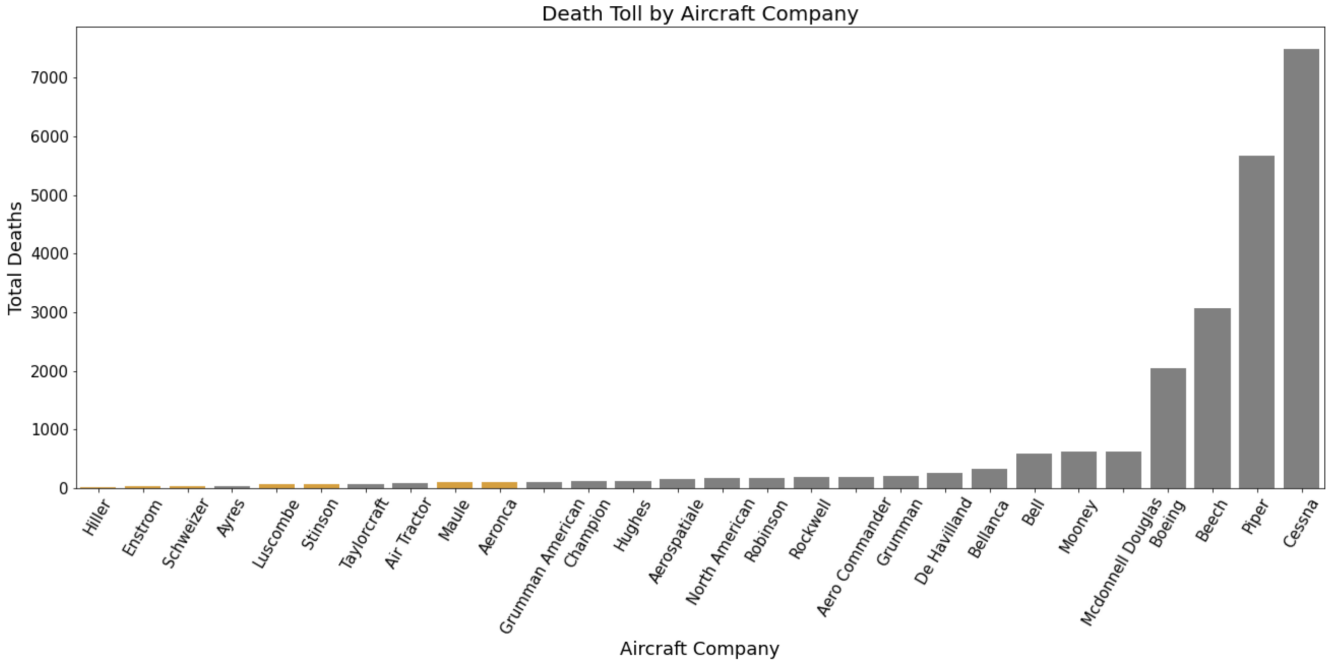


- ~98% of reported accidents involved aircraft with either 1 or 2 engines



- Almost one-third of accidents (~31%) occurred during either the 'Landing' or 'Takeoff' phases of the flight





- Based on metrics for lowest fatality rate, least number of deaths, and lowest number of deaths per flight, the safest brands of airplanes include:
  - Hiller

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%