Evaluation of Aircraft Safety for Commercial and Private Enterprises

Presenter: *Setare Hajarolasvadi*

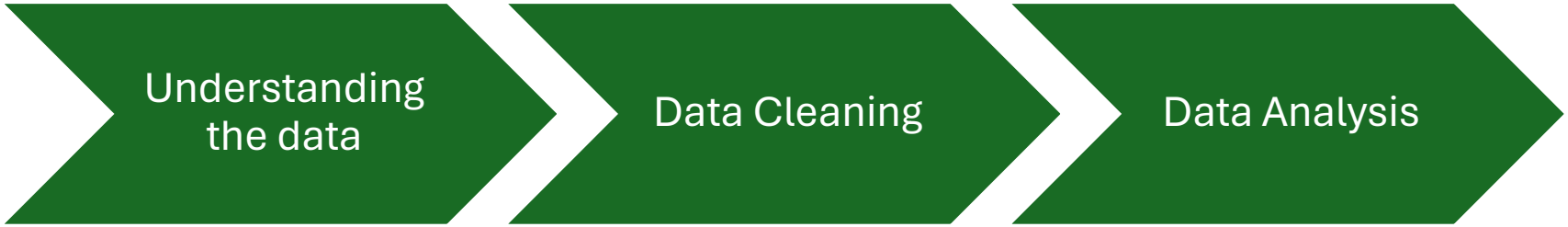
DS Flex Program: *Phase I Project*

Overview

- Objective
- Dataset
- Data Cleaning
- Data Analysis
- Results
- Conclusion

Objective

- This project aims to evaluate the potential risks of aircrafts and make recommendations for stakeholders interested in purchasing and operating airplanes for commercial and private enterprises.



Dataset

- The aviation accident dataset on Kaggle, which consists of aviation accident/incident records from 1948 to date. It contains information about civil aviation accidents and selected incidents within the United States, its territories and possessions, and in international waters.

Dataset – Cont'd

- **Investigation type:** incident vs. accident. Incidents refer to occurrences that do not result in signifcant damage to the aircraft.
- **Engine type:** includes reciprocating, turbo jet, etc. Engine type has been documented to have an effect with aircraft safety.
- **FAR description:** Represents descriptions or codes that specify which specific Federal Aviation Regulations are relevant to each accident.
- **Weather condition:** IMC (Instrument Meteorological Conditions) and VMC (Visual Meteorological Conditions).
- **Broad phase of flight:** indicates the phase of flight at which the accident or incident happened: "Cruise", "Taxi", etc.
- **Report status:** This item shows whether the report on the accident is at its final stage or it's developing.

```
RangeIndex: 90348 entries, 0 to 90347
Data columns (total 31 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Event.Id                88889 non-null  object
 1   Investigation.Type      90348 non-null  object
 2   Accident.Number         88889 non-null  object
 3   Event.Date              88889 non-null  object
 4   Location                88837 non-null  object
 5   Country                 88663 non-null  object
 6   Latitude                34382 non-null  object
 7   Longitude               34373 non-null  object
 8   Airport.Code            50132 non-null  object
 9   Airport.Name            52704 non-null  object
 10  Injury.Severity         87889 non-null  object
 11  Aircraft.damage         85695 non-null  object
 12  Aircraft.Category       32287 non-null  object
 13  Registration.Number     87507 non-null  object
 14  Make                    88826 non-null  object
 15  Model                   88797 non-null  object
 16  Amateur.Built           88787 non-null  object
 17  Number.of.Engines       82805 non-null  float64
 18  Engine.Type             81793 non-null  object
 19  FAR.Description         32023 non-null  object
 20  Schedule                12582 non-null  object
 21  Purpose.of.flight       82697 non-null  object
 22  Air.carrier             16648 non-null  object
 23  Total.Fatal.Injuries    77488 non-null  float64
 24  Total.Serious.Injuries  76379 non-null  float64
 25  Total.Minor.Injuries    76956 non-null  float64
 26  Total.Uninjured         82977 non-null  float64
 27  Weather.Condition       84397 non-null  object
 28  Broad.phase.of.flight   61724 non-null  object
 29  Report.Status           82505 non-null  object
 30  Publication.Date        73659 non-null  object
```

Dataset – Cont'd

Time and location:

- event date, location, country

Severity:

- injury severity, aircraft damage, total injuries and total fatal/serious/minor injuries, total uninjured

Aircraft properties:

- air carrier, aircraft category, make, model, number of engines, engine type,

Other:

- FAR description, purpose of flight, weather condition.

```
RangeIndex: 90348 entries, 0 to 90347
Data columns (total 31 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Event.Id                88889 non-null  object
 1   Investigation.Type      90348 non-null  object
 2   Accident.Number         88889 non-null  object
 3   Event.Date              88889 non-null  object
 4   Location                88837 non-null  object
 5   Country                 88663 non-null  object
 6   Latitude                34382 non-null  object
 7   Longitude               34373 non-null  object
 8   Airport.Code            50132 non-null  object
 9   Airport.Name            52704 non-null  object
 10  Injury.Severity         87889 non-null  object
 11  Aircraft.damage         85695 non-null  object
 12  Aircraft.Category       32287 non-null  object
 13  Registration.Number     87507 non-null  object
 14  Make                    88826 non-null  object
 15  Model                   88797 non-null  object
 16  Amateur.Built           88787 non-null  object
 17  Number.of.Engines       82805 non-null  float64
 18  Engine.Type             81793 non-null  object
 19  FAR.Description         32023 non-null  object
 20  Schedule                12582 non-null  object
 21  Purpose.of.flight       82697 non-null  object
 22  Air.carrier             16648 non-null  object
 23  Total.Fatal.Injuries    77488 non-null  float64
 24  Total.Serious.Injuries  76379 non-null  float64
 25  Total.Minor.Injuries    76956 non-null  float64
 26  Total.Uninjured         82977 non-null  float64
 27  Weather.Condition       84397 non-null  object
 28  Broad.phase.of.flight   61724 non-null  object
 29  Report.Status           82505 non-null  object
 30  Publication.Date        73659 non-null  object
```

Data Cleaning

- **Type Conversion:** We convert event date from an object to datetime.
- **Handling Missing Values:**
  - All columns in which the number of missing values exceeds or is equal to 30% of the total entries are removed.
  - “UNK”, “Unk” and “Unknown” in the dataset are replaced with Nans.
  - The rows with Nans for the remainder of the data frame are removed.

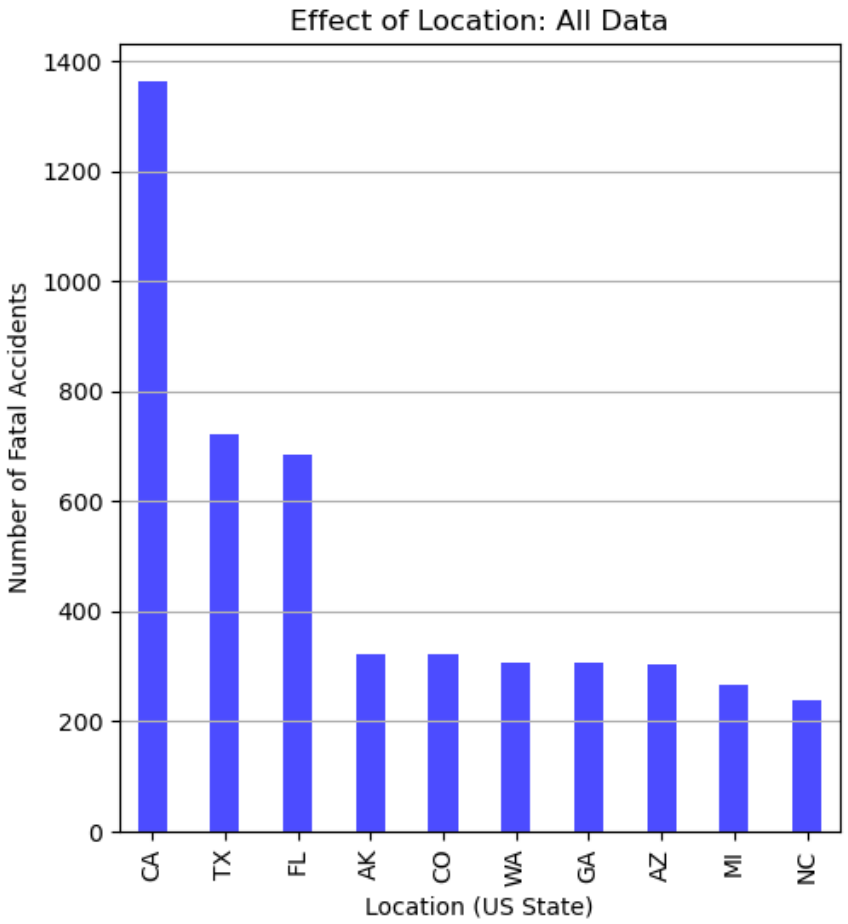
  64% of the data entries are preserved in this manner.
- **Filtering and Cleaning:**
  - Only data where the country is specified as United States is kept.
  - We change the location column such that it only contains the abbreviation for the states’ names.
  - We unify the capitalization of fields in the Make column.
  - The number of fatal injuries listed in parenthesis in front of the fields in the injury severity column is removed.

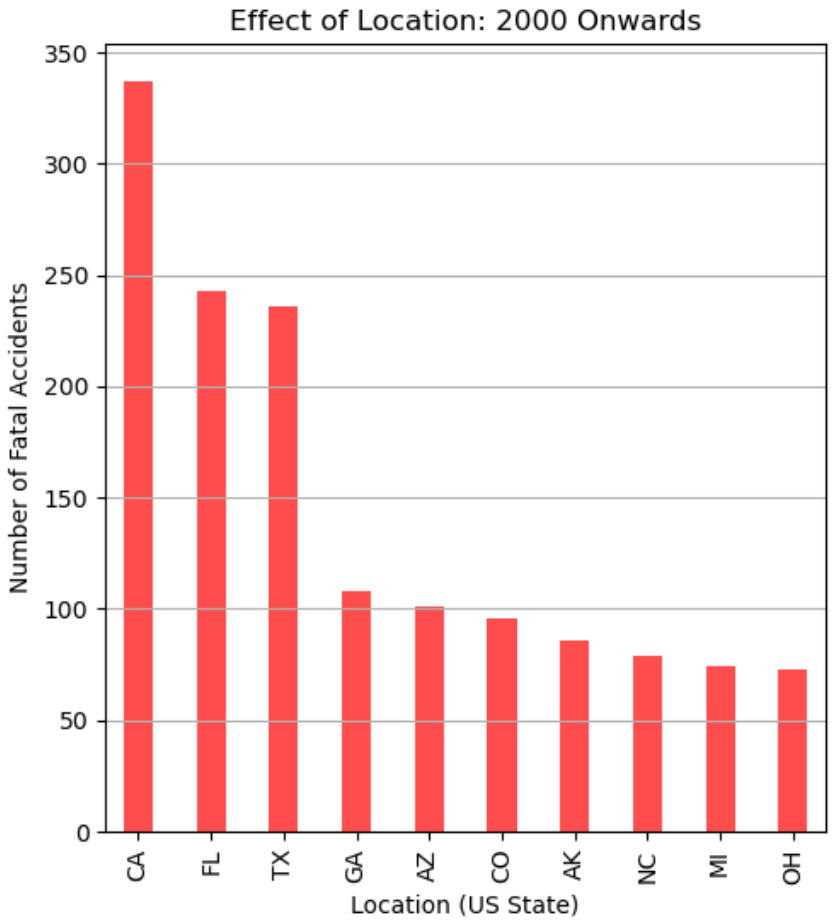
Data Analysis – Approach

- Now that the data is clean, we group the data based on factors we consider important for aircraft safety to do the analysis.
- In the absence of information on the total number of flights (including the ones not involving an accident), we choose the *total number of fatal accidents* as a metric for evaluating safety.

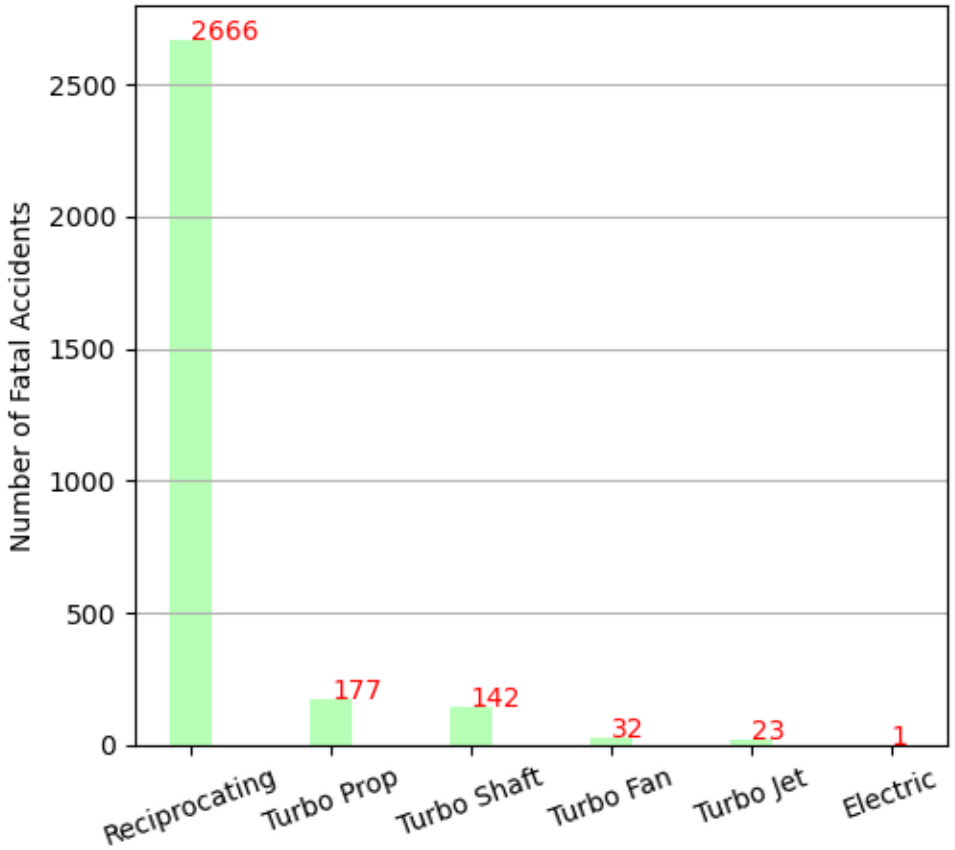
Event_date	Location	Country	Injury_severity	Make	Model	Number_of_engines	Engine_type	Weather_condition	Purpose_of_flight	Total_onboard	Aircraft_size
1977-06-19	CA	United States	Fatal	Rockwell	112	1.0	Reciprocating	IMC	Personal	2.0	very small aircraft
1981-08-01	MN	United States	Fatal	Cessna	180	1.0	Reciprocating	IMC	Personal	4.0	very small aircraft
1982-01-01	WA	United States	Non-Fatal	Cessna	140	1.0	Reciprocating	VMC	Personal	2.0	very small aircraft
1982-01-01	NJ	United States	Non-Fatal	Cessna	401B	2.0	Reciprocating	IMC	Business	2.0	very small aircraft
1982-01-01	FL	United States	Non-Fatal	North American	NAVION L-17B	1.0	Reciprocating	IMC	Personal	3.0	very small aircraft

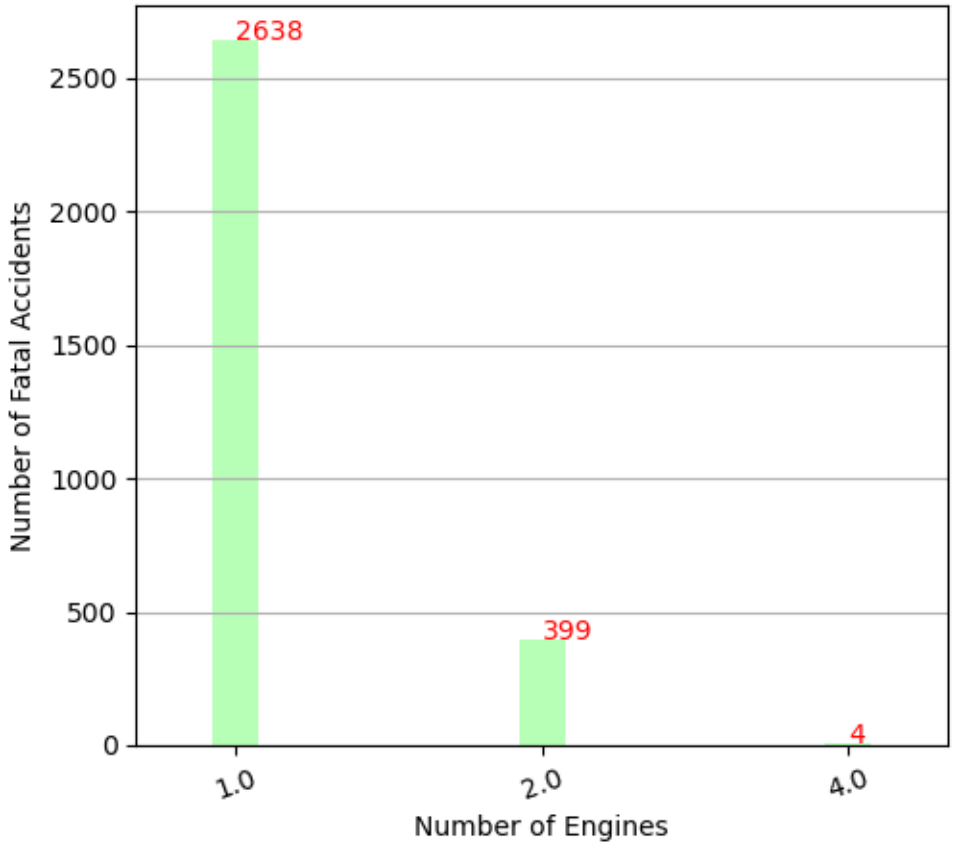
Results: Location



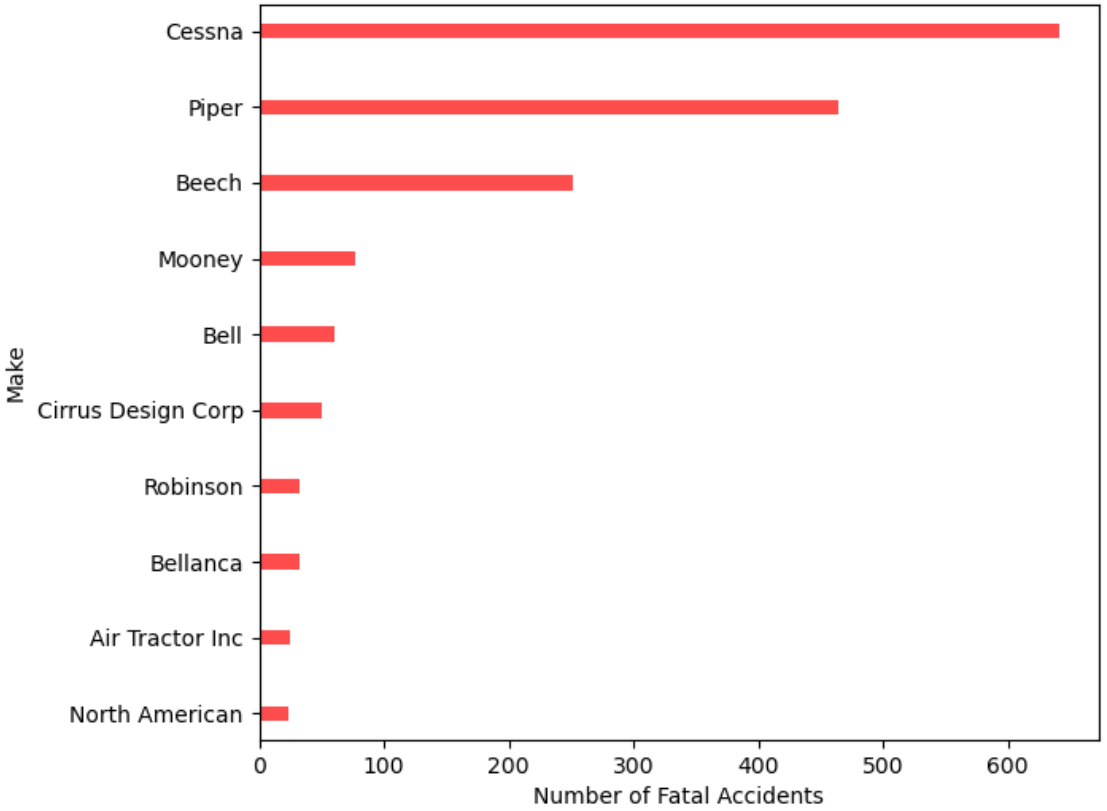


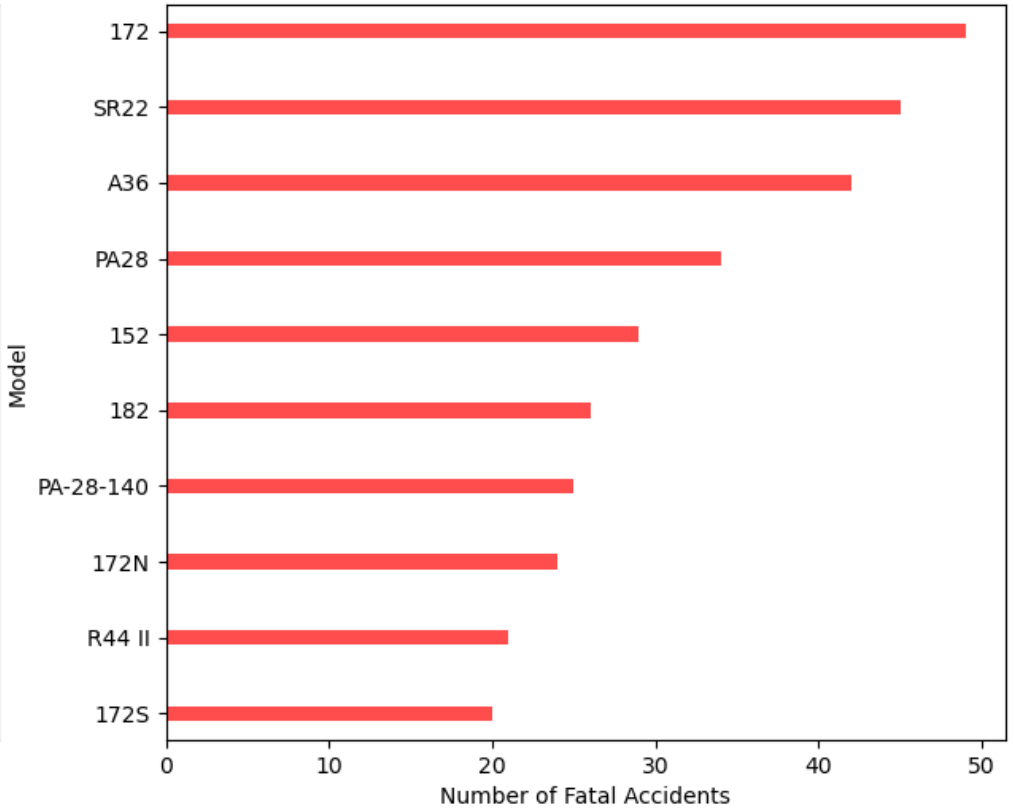
Results: Engine Type & Count



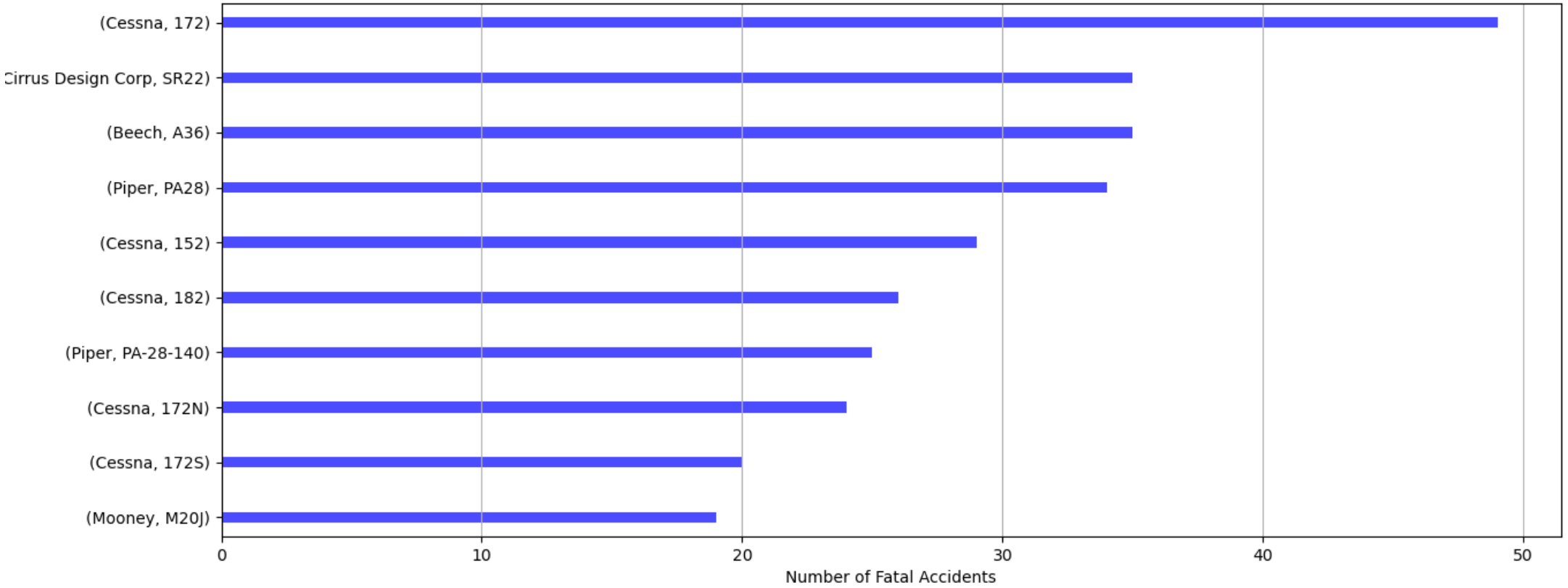


Results: Make & Model

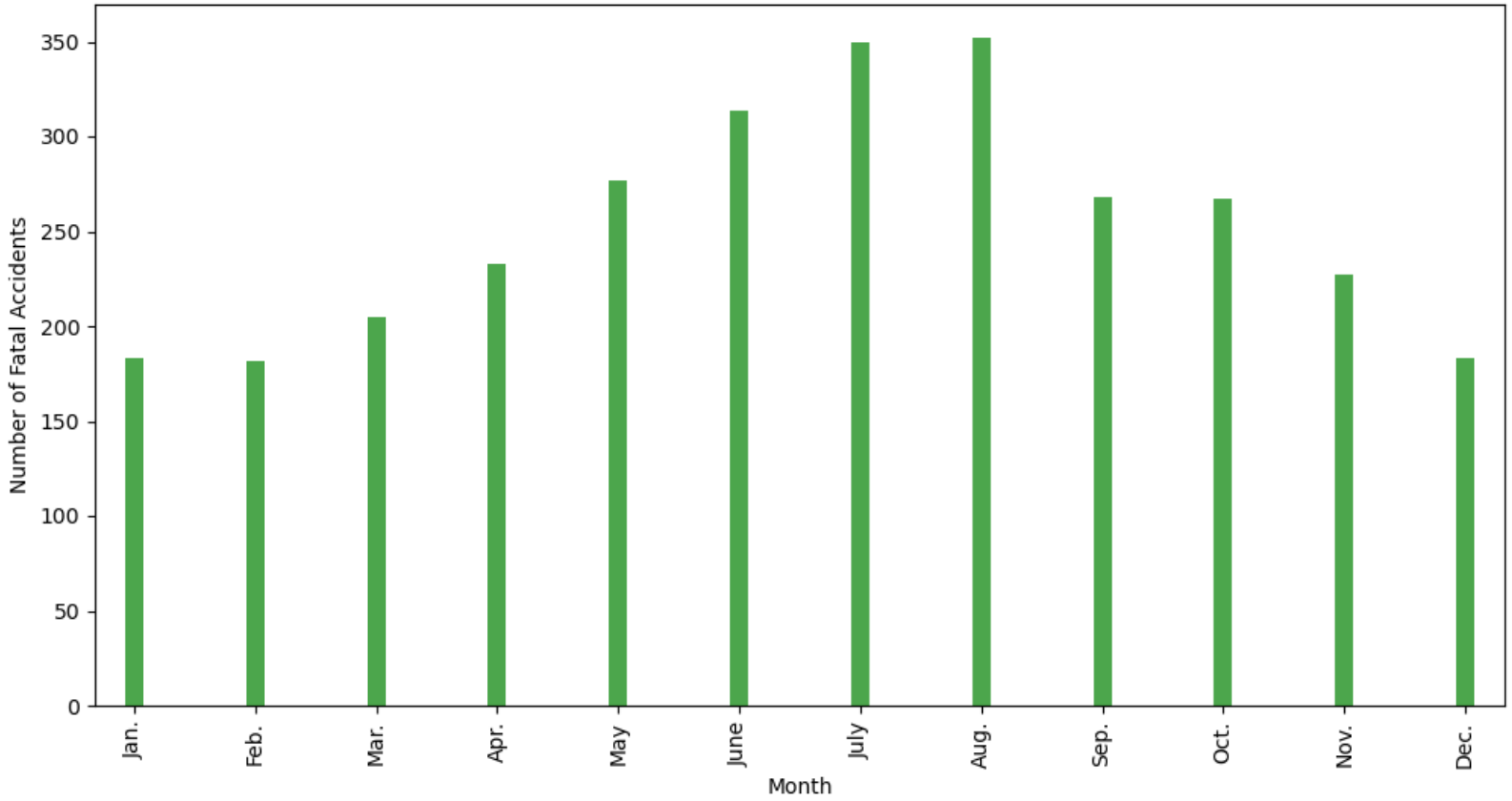




Results: Make & Model – Cont'd



Results: Time of Year



Recommendations

1. C**alifornia, Florida and Texas** are the top three states in terms of the number of fatal aviation accidents. Most fatal accidents in these states are associated with flights that are conducted for personal (recreational) or instructional purposes. Therefore, if we solely rely on the data available at hand, **it is not recommended to pursue business in these three states for recreational or instructional aircrafts**.

2. Aircrafts with reciprocating engines are highly involved in fatal aviation accidents. This doesn't necessarily mean that reciprocating engines are the cause of fatal accidents. However, it does indicate a correlation. Therefore, **it's not recommended to invest in aircrafts with reciprocating engines.**

3. Aircrafts with higher number of engines seem to be safer in general. This is reasonable since having more engines ensures that the aircraft can remain operational in case one fails. Therefore, **it is recommended to be cautious about investing in small single-engine aircrafts.**

Recommendations – Cont'd

4. **Cessna, Piper and Beech** are associated with the highest number of fatal accidents. However, when both make and model are considered, **Cirrus Design Corp, SR22** is also among aircrafts with the highest number of fatal accidents. **It is, therefore, recommended to avoid using these make and models.**

5. Unlike an initial preconception that there may be more fatal accidents in months with extreme weather, **July and August** turn out to be the months with the highest number of fatal accidents. This observation may have an underlying cause: there are simply more flights taking place in these months. Regardless, **it seems reasonable to recommend that stakeholders expect and plan better for aviation incidents and accidents during these months.**

Limitations

The above study is limited in the following ways:

1. The analysis is limited to the United States.
2. In the absence of information on the total number of flights (including those without accidents), we're relying on the absolute number of fatal accidents as a metric for our analysis. This (without normalization) may be misleading.