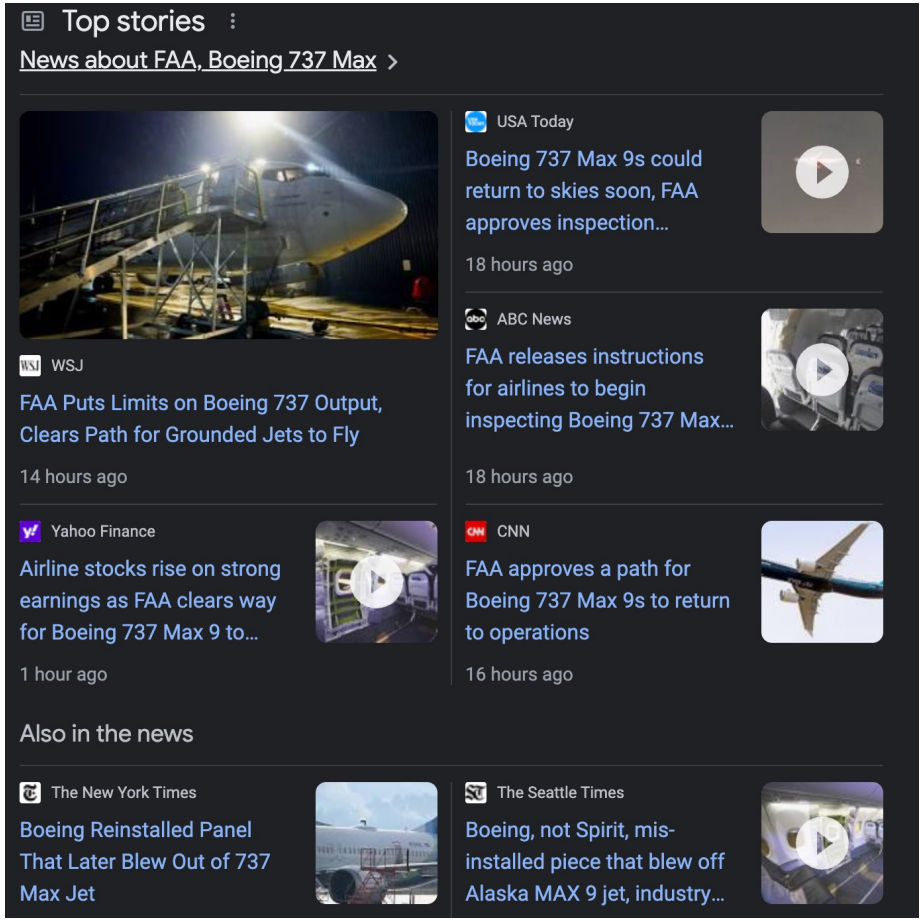
Data Science Flex 1 Project Presentation

Alexandra Yakovleva

Introduction.

- Air Transportation is relatively novel way of travelling (1920, Golden Age of Aviation) and transferring cargo (1960, Jet Era).
- Air Transportation is much safer (0.01 deaths/100 million miles) than any other form of transportation, however, it is as safe as the operator, the equipment, and the safety standards.
- The standards of the equipment maintenance have been rigorously updating over the decades by government safety.
- There is an increasing pressure to develop more efficient aircrafts, aircraft development is in the hands of private companies and is not governed by any standards.

Aircraft Safety in the News.



The news you don't want to be in.

Business Context.

- **Broad Task:** *Identify aircrafts with the lowest safety risks for commercial and private aviation enterprises.*
- **Break it down into subtasks**:

  - *What type of safety data is available for aviation transportation?*

    Explore sources of accident information for safety analysis.

What can be used to measure aircraft safety?

Explore and establish the safety risk measures for aircrafts in context of the data available.

  - *What are the potential commercial and private aviation enterprises?*

    Explore the most common flight purposes and compare safety ratings for the most common aircraft makes, models.

Safety Data.

- *Data Source:*

  For the data analysis, we'll be using the National Transportation Safety Board's Aviation Accidents Dataset. The dataset is collected by a government entity (trusted source) and has no copyright, meaning that it can be used freely for commercial purposes.

- *Safety:*

  The International Air Transport Association (IATA) defines aircraft safety rating as total fatal occurrences per 1.000.000 miles flown. We have no access for the total miles flown but we can use the aircraft's entry count in database as normalizing factor.

- *Possible aviation enterprises:*

  The dataset contains information about flight purposes to help us identify the most popular aviation enterprises.

Data Processing Steps.

1. *Identifying Informative Columns:*

   We identified key columns describing purpose of the flight, date of the accident, injuries, make and model.

   We identified columns that will be irrelevant for our analysis, as well as columns that will negatively impact the dataset quality.

2. *Dealing with Missing/Incomplete Information:*

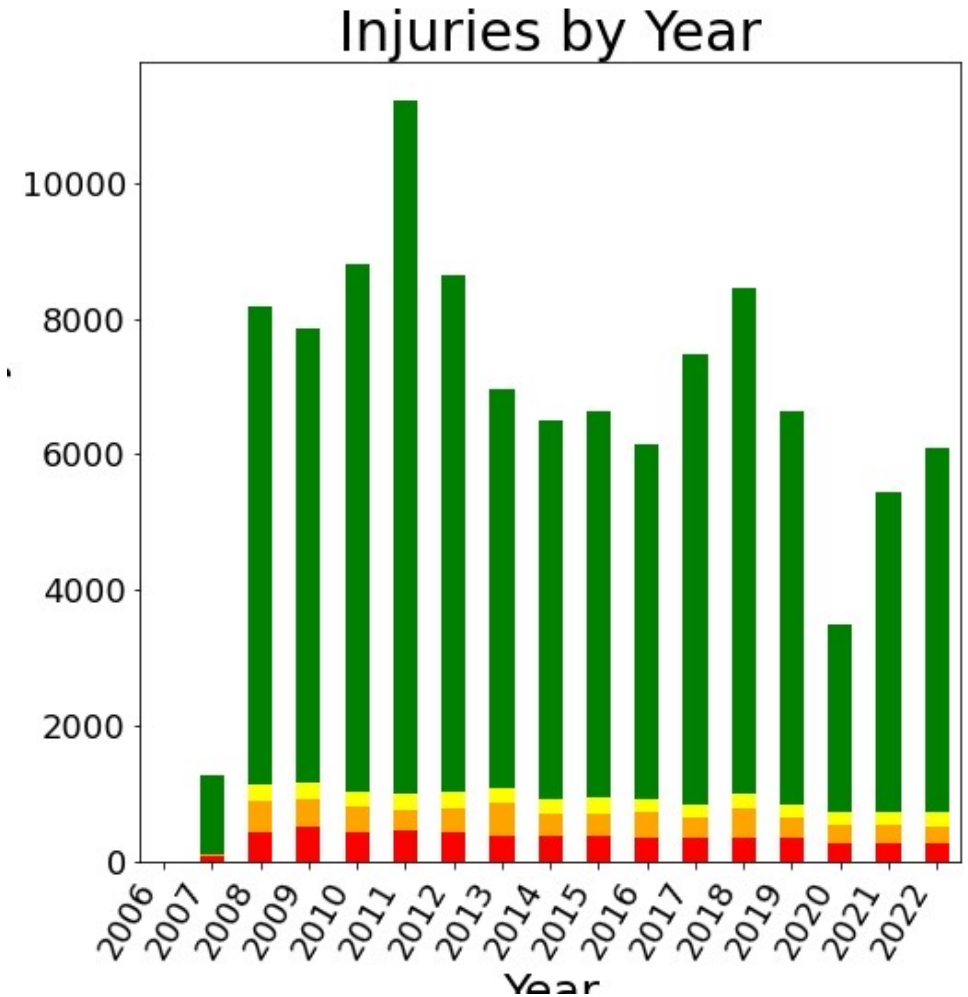
   We utilized all available information about the Flight Type in other columns to fill in the missing flight categories.

   We removed duplicate entries from Make and Model Column.

3. *Identifying Relevant Timeframe:*

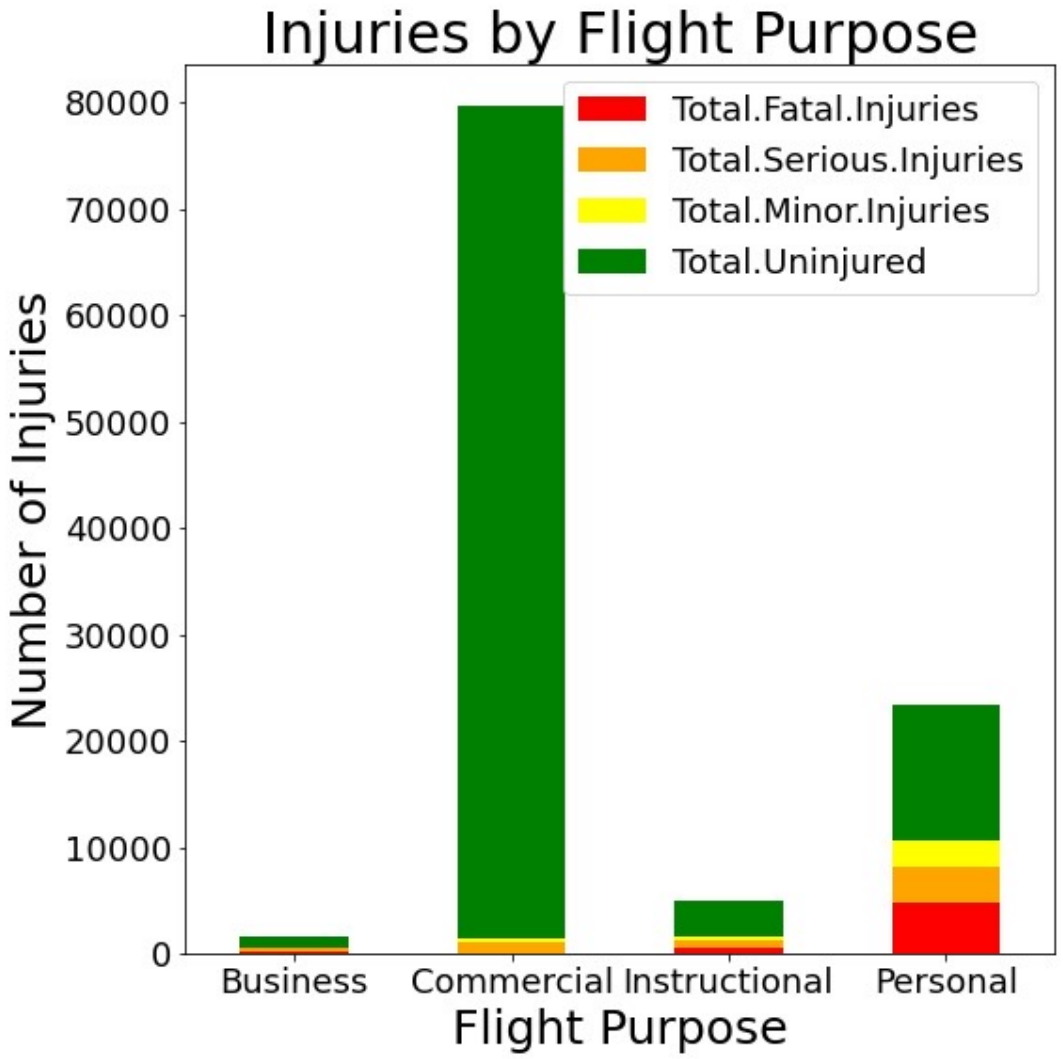
   We performed initial visualization of accidents over time and identified the most informative and relevant timeframe for the future analysis.

Exploratory Analysis: Flying Got Safer.



Identifying Four Most Common Flight Purpose Categories.

1. Commercial (least fatalities).
2. Personal (most fatalities and injuries).
3. Instructional.
4. Business.



Computing Safety Rankings.

- Identify most common aircrafts for each flying category and their count.
- Sum up all injuries for each aircraft make and model.
- Divide all injury values by the count.
- Order by the Fatality Rank (normalized fatalities) ascending.

Example: Most Common Personal Aircrafts and Fatality Rank.

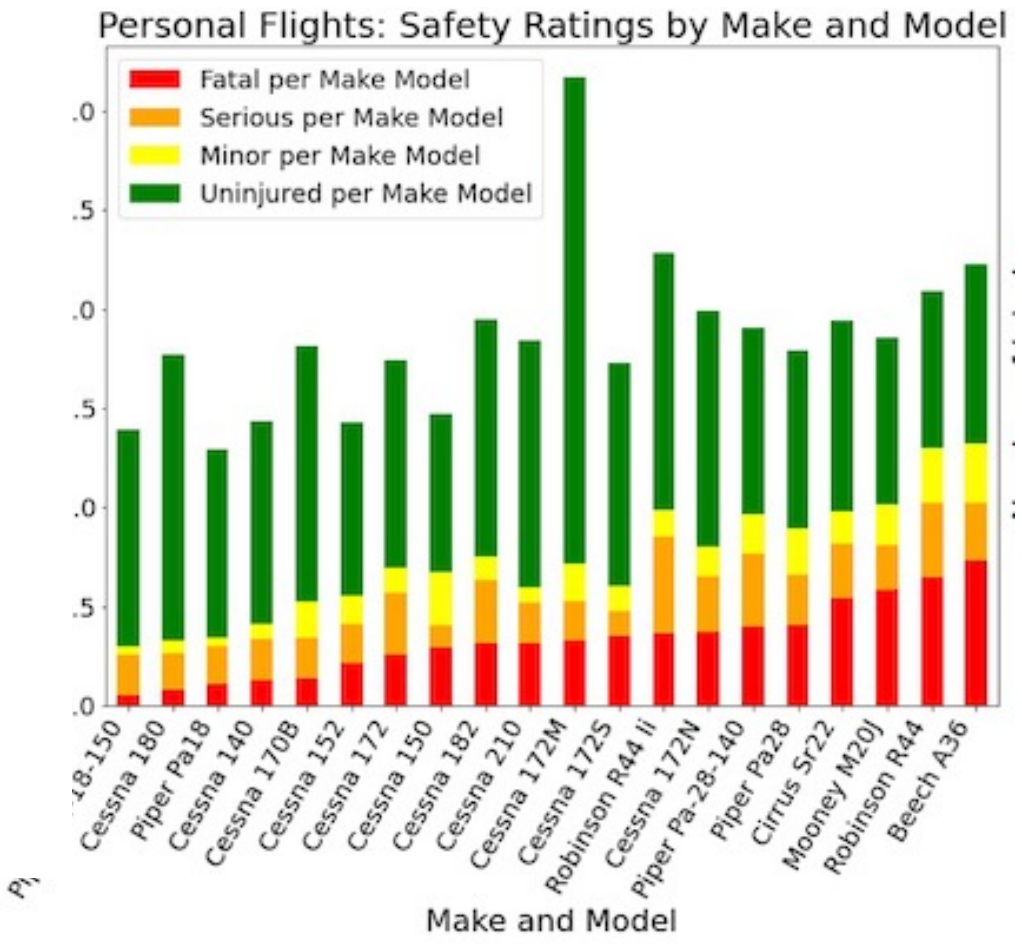
$$\textit{Normalized Value} = \frac{\textit{Total Value}}{\textit{Count}}$$

Make.Model	Total.Fatal.Injuries	Total.Serious.Injuries	Total.Minor.Injuries	Total.Uninjured	Count
Piper Pa-18-150	5.0	23.0	5.0	120.0	110
Piper Pa18	7.0	12.0	3.0	61.0	64
Cessna 170B	10.0	15.0	14.0	95.0	74
Cessna 140	11.0	18.0	7.0	89.0	87
Cessna 180	12.0	30.0	11.0	232.0	161
Cessna 152	19.0	17.0	13.0	77.0	88
Cessna 210	20.0	13.0	5.0	80.0	64
Robinson R44 Ii	27.0	36.0	10.0	96.0	74
Cessna 172S	27.0	10.0	10.0	88.0	78
Cessna 150	28.0	11.0	26.0	76.0	96
Cessna 172M	29.0	17.0	17.0	216.0	88
Piper Pa-28-140	34.0	32.0	17.0	81.0	86
Cessna 172N	37.0	29.0	15.0	120.0	101
Mooney M20J	40.0	16.0	14.0	58.0	69
Robinson R44	49.0	29.0	21.0	60.0	76
Cessna 182	54.0	57.0	21.0	209.0	175
Piper Pa28	59.0	37.0	34.0	130.0	145
Cessna 172	87.0	109.0	43.0	359.0	343
Beech A36	90.0	36.0	37.0	111.0	123
Cirrus Sr22	96.0	49.0	28.0	170.0	177

Example: Most Common Personal Aircrafts and Fatality Rank.

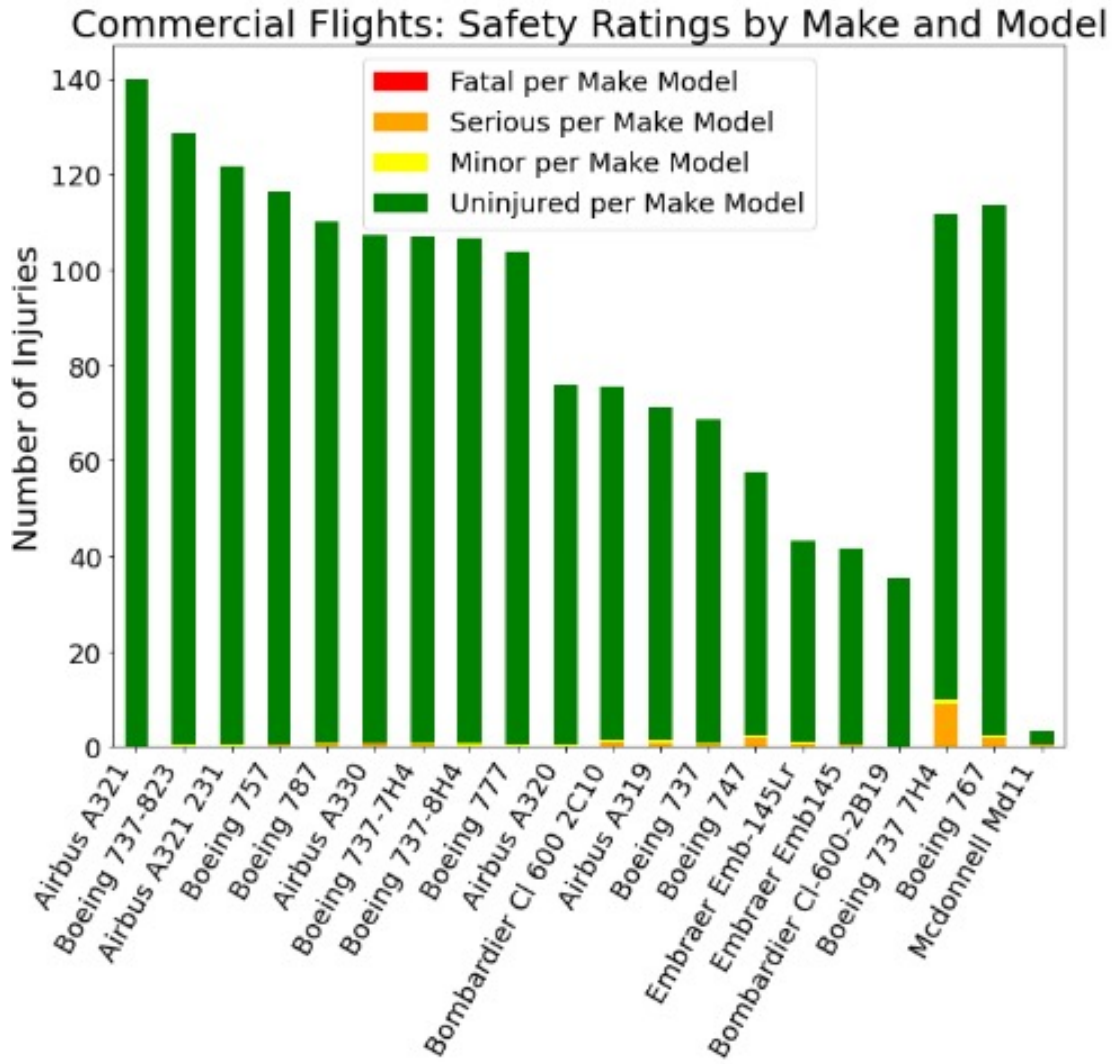
Make.Model	Fatal per Make Model	Serious per Make Model	Minor per Make Model	Uninjured per Make Model
Piper Pa-18-150	0.045455	0.209091	0.045455	1.090909
Cessna 180	0.074534	0.186335	0.068323	1.440994
Piper Pa18	0.109375	0.187500	0.046875	0.953125
Cessna 140	0.126437	0.206897	0.080460	1.022989
Cessna 170B	0.135135	0.202703	0.189189	1.283784
Cessna 152	0.215909	0.193182	0.147727	0.875000
Cessna 172	0.253644	0.317784	0.125364	1.046647
Cessna 150	0.291667	0.114583	0.270833	0.791667
Cessna 182	0.308571	0.325714	0.120000	1.194286
Cessna 210	0.312500	0.203125	0.078125	1.250000
Cessna 172M	0.329545	0.193182	0.193182	2.454545
Cessna 172S	0.346154	0.128205	0.128205	1.128205
Robinson R44 Ii	0.364865	0.486486	0.135135	1.297297
Cessna 172N	0.366337	0.287129	0.148515	1.188119
Piper Pa-28-140	0.395349	0.372093	0.197674	0.941860
Piper Pa28	0.406897	0.255172	0.234483	0.896552
Cirrus Sr22	0.542373	0.276836	0.158192	0.960452
Mooney M20J	0.579710	0.231884	0.202899	0.840580
Robinson R44	0.644737	0.381579	0.276316	0.789474
Beech A36	0.731707	0.292683	0.300813	0.902439

Entries in table are ordered by Fatality.



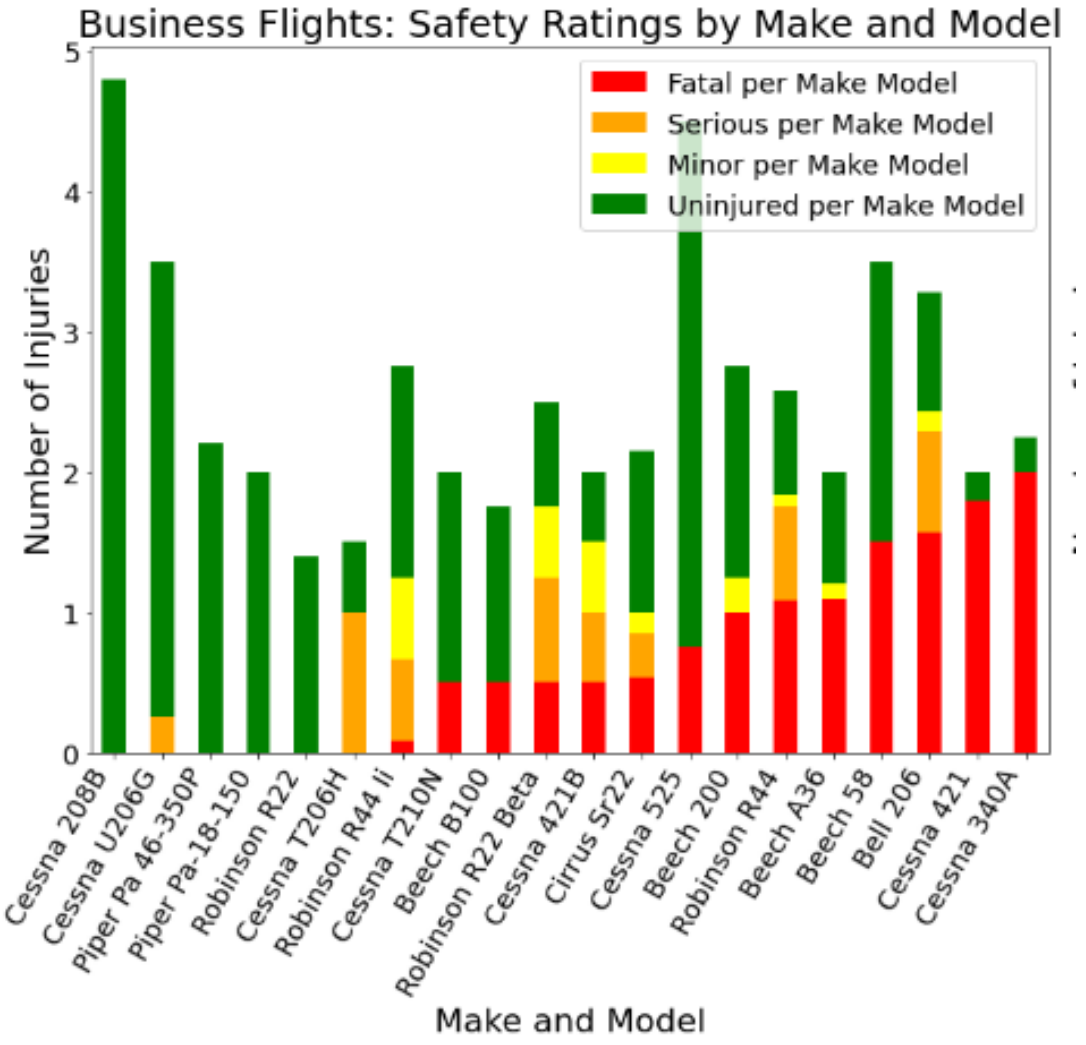
Identifying Safest Aircrafts For Commercial Purposes.

Aircraft	Fatality Rank
Airbus A321	0
Boeing 737	0
Boeing 757	0
Boeing 787	0
Airbus A320	0



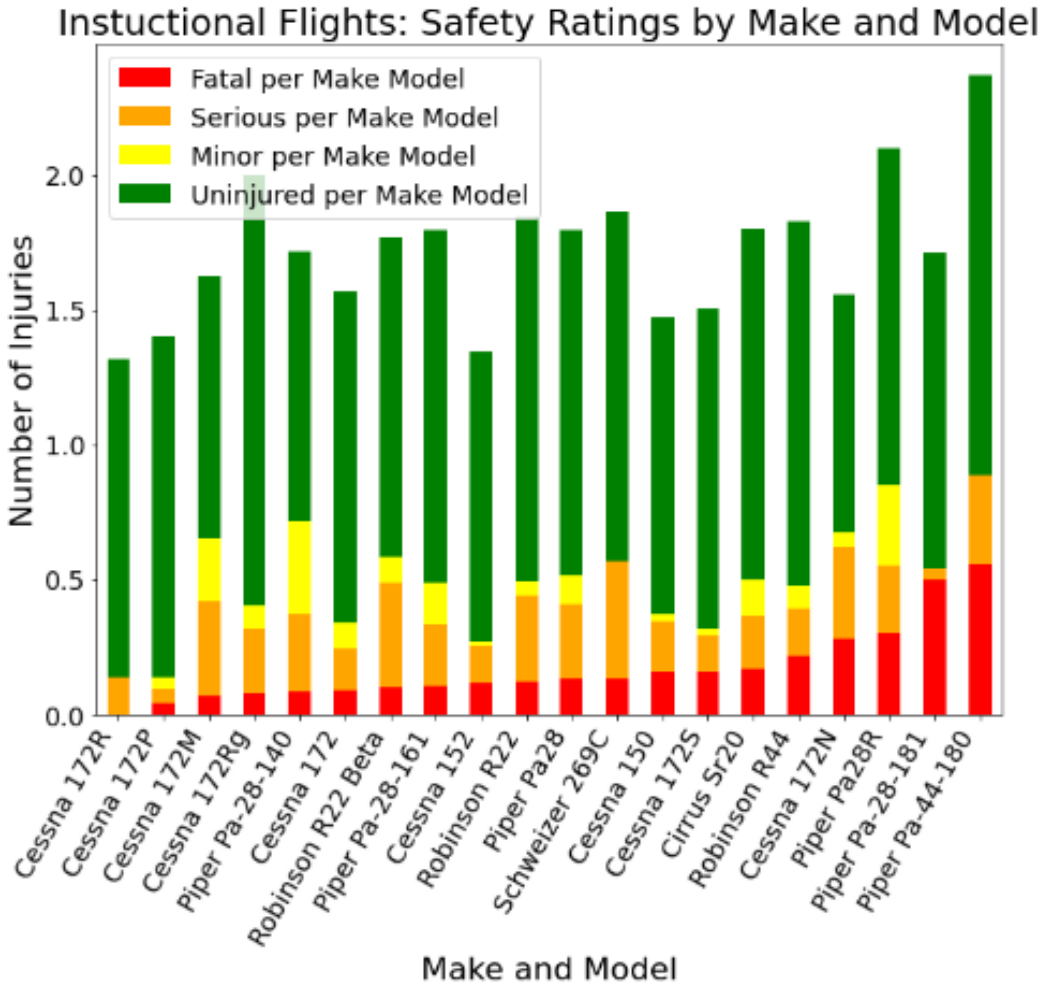
Identifying Safest Aircrafts For Business Purposes.

Aircraft	Fatality Rank
Cessna 208B	0
Cessna U206G	0
Piper Pa 46-350P	0
Piper Pa 18-150	0
Robinson R22	0



Identifying Safest Aircrafts For Instructional Purposes.

Aircraft	Fatality Rank
Cessna 172 R	0
Cessna 172 P	0.04
Cessna 172 M	0.07
Cessna 172 Rg	0.08
Piper Pa-28-140	0.085



Top Recommendations from Safety Perspective.

Commercial	Business	Instructional
Airbus A321	Cessna 208B	Cessna 172 R
Boeing 737	Cessna U206G	*Cessna 172 P**
Boeing 757	Piper Pa 46-350P	*Cessna 172 M**
Boeing 787	Piper Pa 18-150	*Cessna 172 Rg**
Airbus A320	Robinson R22	*Piper Pa-28-140**

**Have non-zero Fatality Rank*

Future Work.

- Align safety metrics with the IATA.
- Based on the accident description, perform text analysis to remove fatal accidents related to human error (number can be as high as 80%).
- Research aircraft Make and Model to further compact the accidents dataset (difference between Cessna 172 P and 172 N).

References.

- Fatality Rates on IATA.
- NTSB Aviation Accidents.

Contact Information.

- Reach me at leksea@gmail.com or linkedin:alex.yakovleva