AVIATION ACCIDENT ANALYSIS

INTRODUCTION

Our dataset is from kaggle.

It contains information from 1962 and later about civil aviation accidents in the US and its territories.

It has 31 features and about 88,000 rows

OBJECTIVES

My goal for this dataset is to determine :

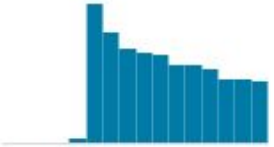
*which aircrafts have the lowest risk.

*how accident rates have changed over time.

*compare the accident rates on different aircraft manufacturers and models.

*see how the data in this dataset correlate

Understanding the data

Event.Id Event ID	Investigation.Type Investigation Type	Accident.Number NTSB Number	Event.Date Date of Event	Location Event Location	Country Country
87951 unique values	Accident 96% Incident 4%	**88863** unique values	 1948-10-24 2022-12-29	**27759** unique values	United States Brazil Other (6267)
20001218X45444	Accident	SEA87LA080	1948-10-24	MOOSE CREEK, ID	United States
20001218X45447	Accident	LAX94LA336	1962-07-19	BRIDGEPORT, CA	United States
20061025X01555	Accident	NYC07LA005	1974-08-30	Saltville, VA	United States
20001218X45448	Accident	LAX96LA321	1977-06-19	EUREKA, CA	United States
20041105X01764	Accident	CHI79FA064	1979-08-02	Canton, OH	United States
20170710X52551	Accident	NYC79AA106	1979-09-17	BOSTON, MA	United States

DATA CLEANING

As you've seen our data was kinda messy, ie it has missing values, it is inconsistent etc so it was really important that we clean it

We clean data by:

*filling missing values

*removing duplicates

*removing whitespaces

*ensuring all the data has a consistent format

Feature Engineering

This is creating new features from existing columns or separating columns to get a new feature

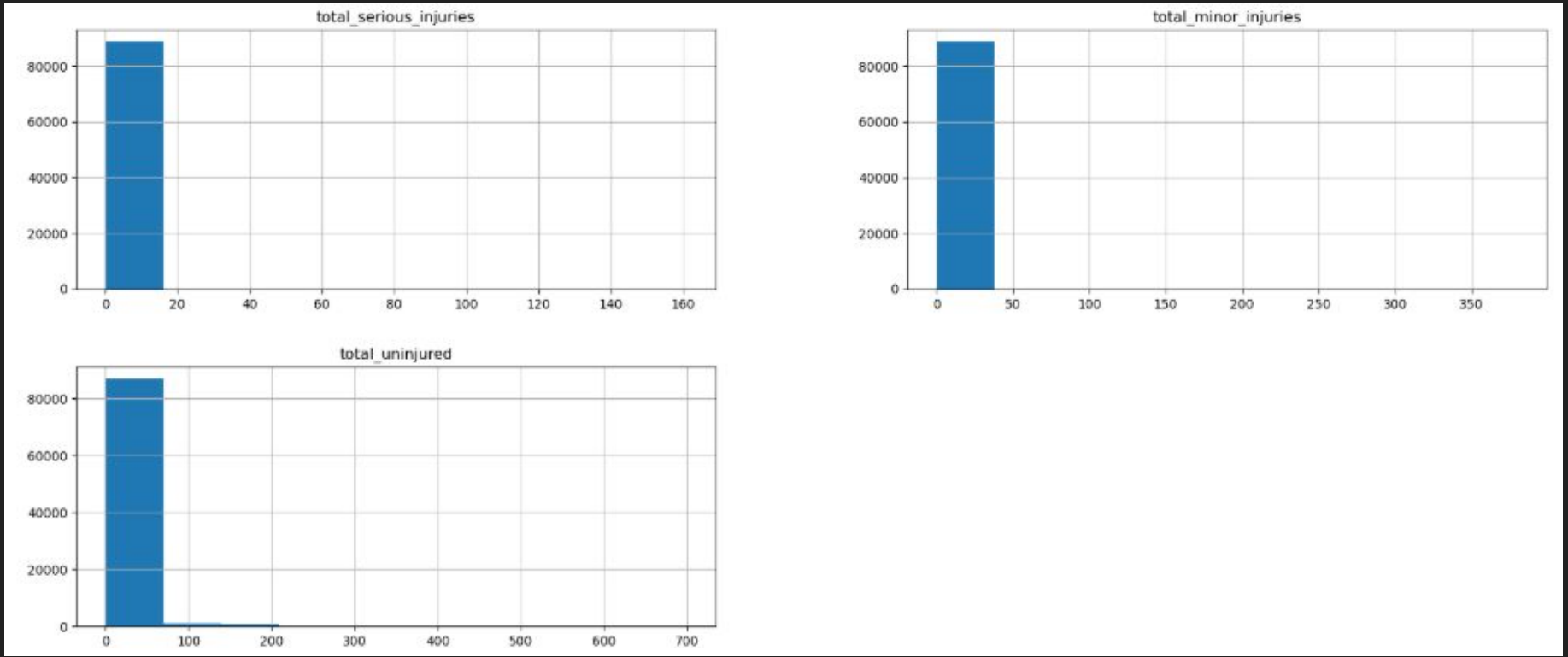
The new features we made are

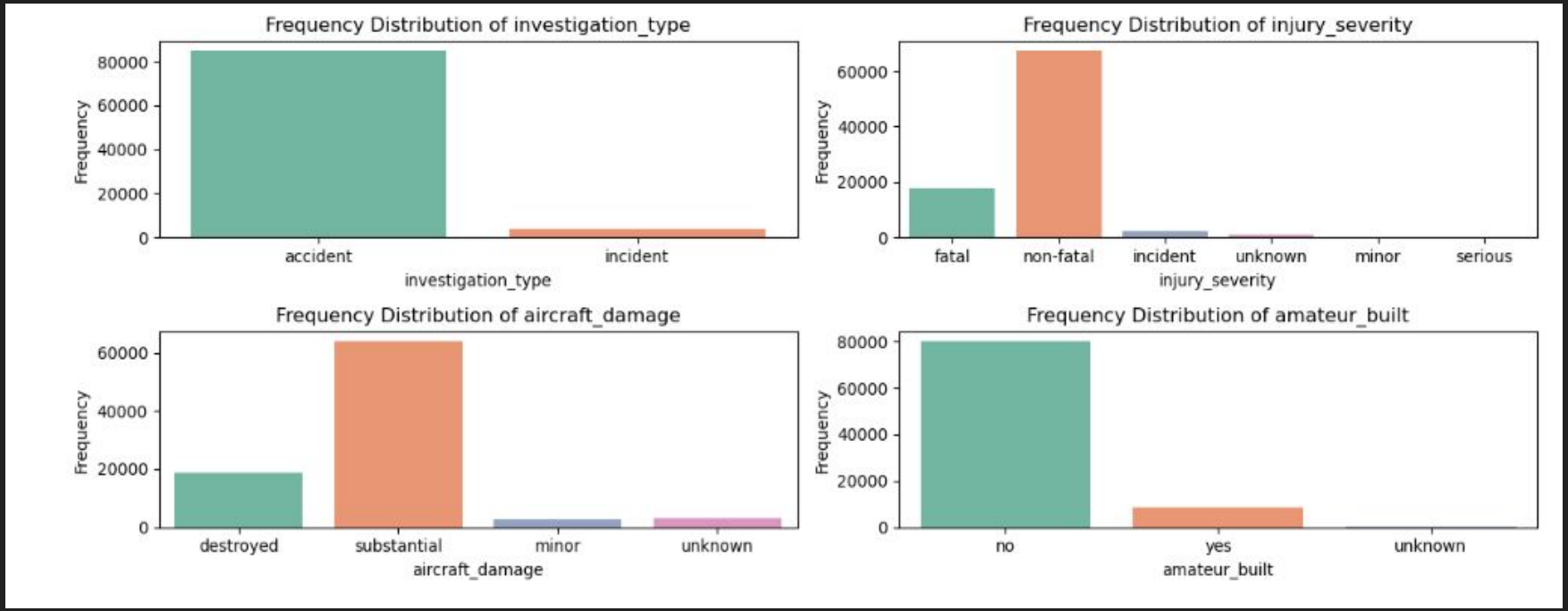
*the year column

*the region column

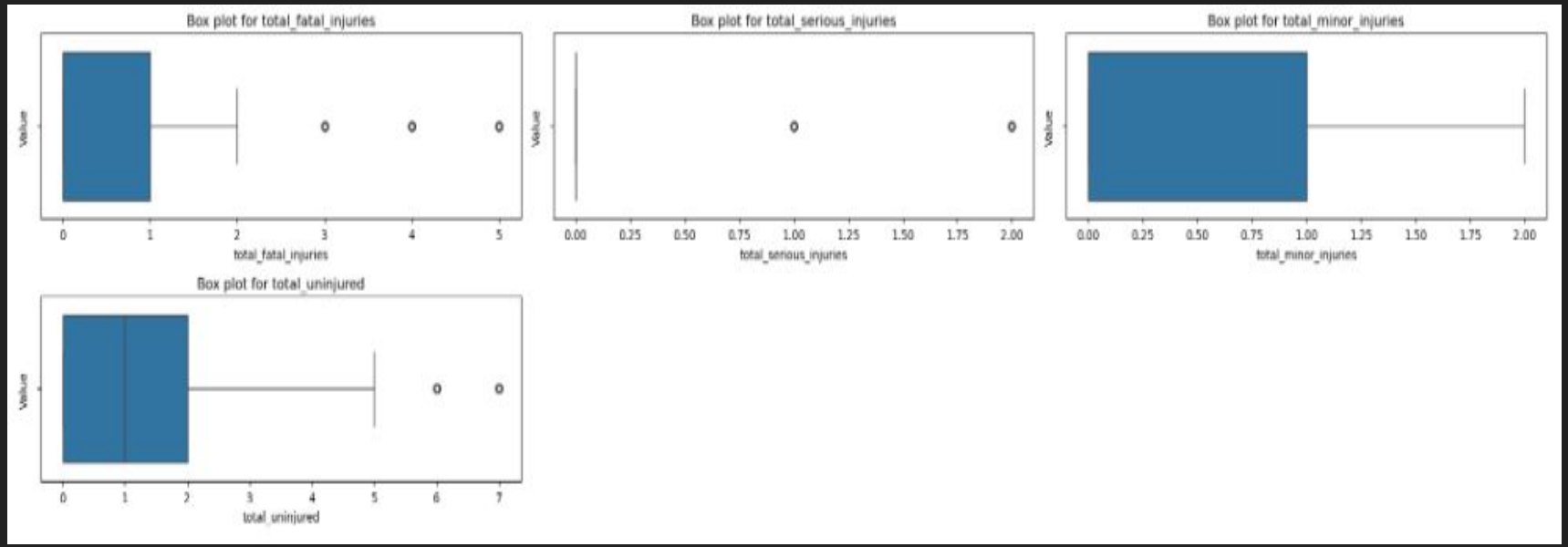
*total_injuries column

Data Distribution

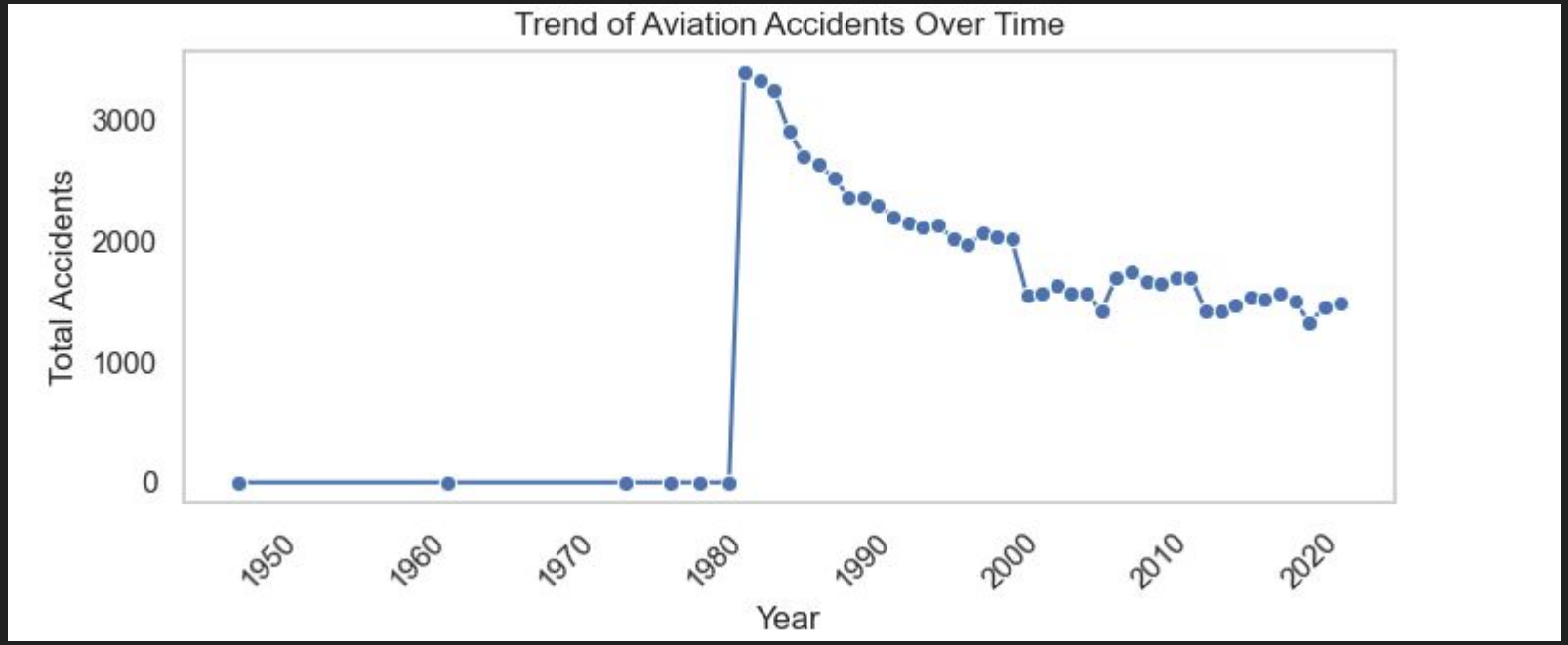



Frequency Distribution of investigation_type
Frequency
80000
60000
40000
20000
0
accident
incident
investigation_type
Frequency Distribution of injury_severity
Frequency
60000
40000
20000
0
fatal
non-fatal
incident
unknown
minor
serious
injury_severity
Frequency Distribution of aircraft_damage
Frequency
60000
40000
20000
0
destroyed
substantial
minor
unknown
aircraft_damage
Frequency Distribution of amateur_built
Frequency
80000
60000
40000
20000
0
no
yes
unknown
amateur_built

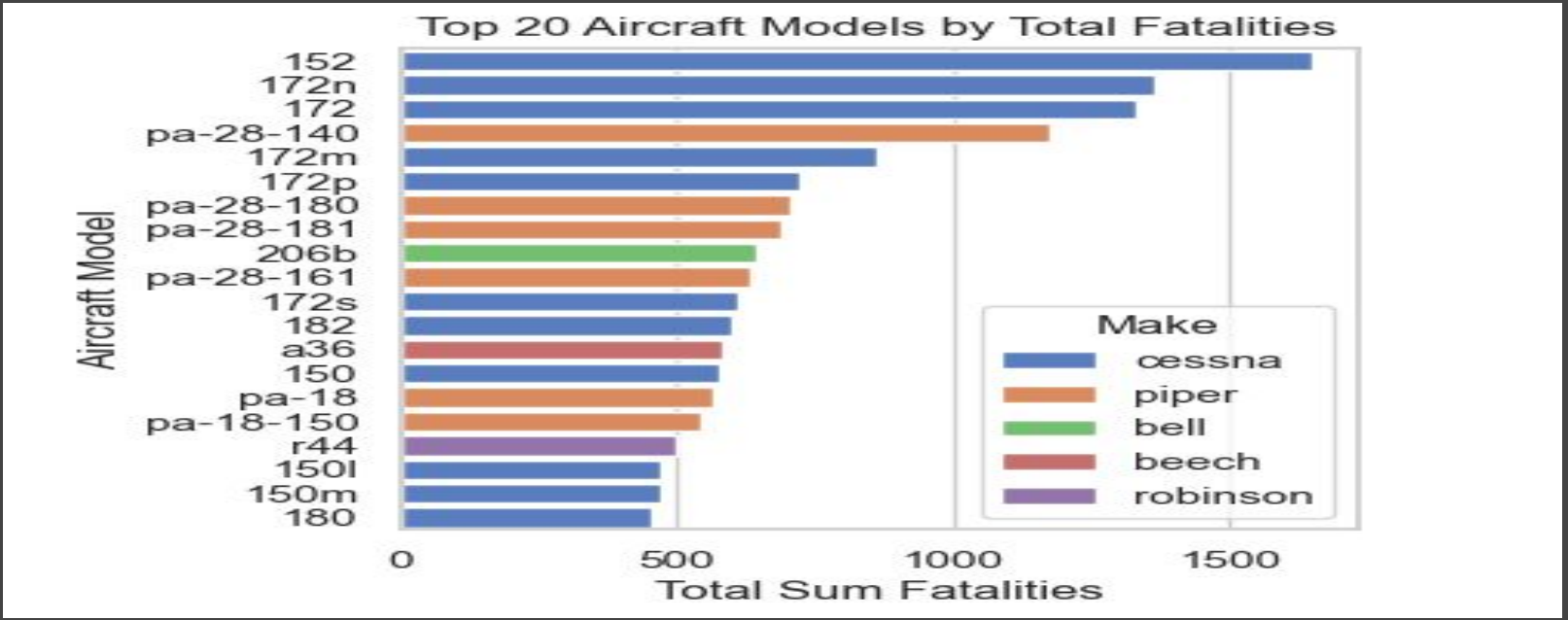
Outlier Detection

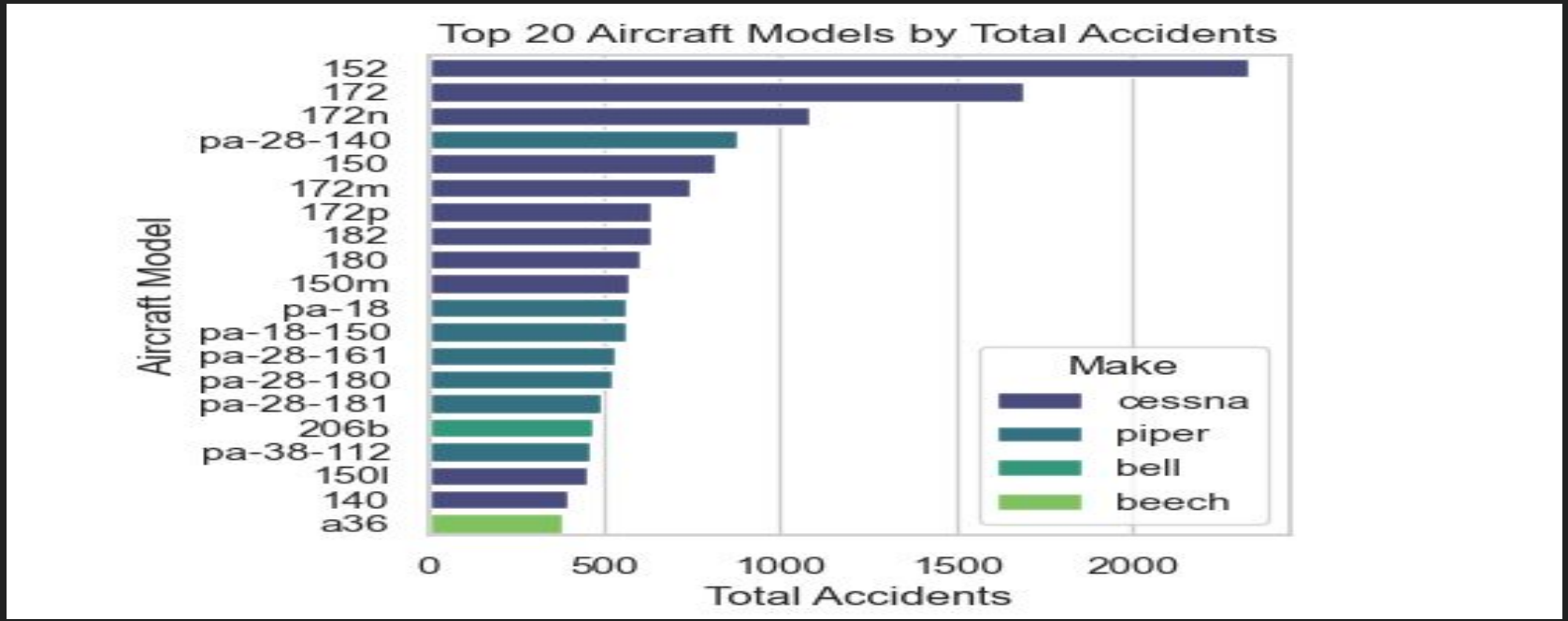


Accident over time analysis

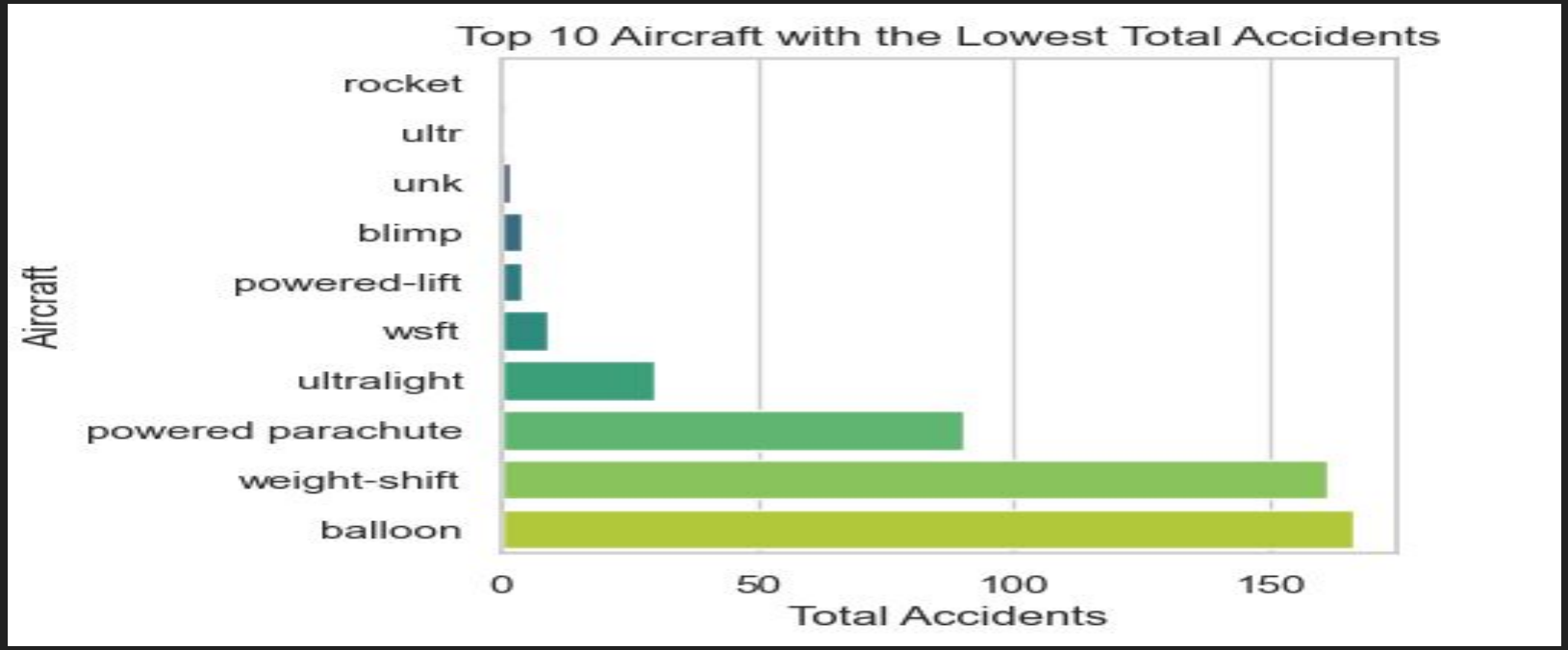


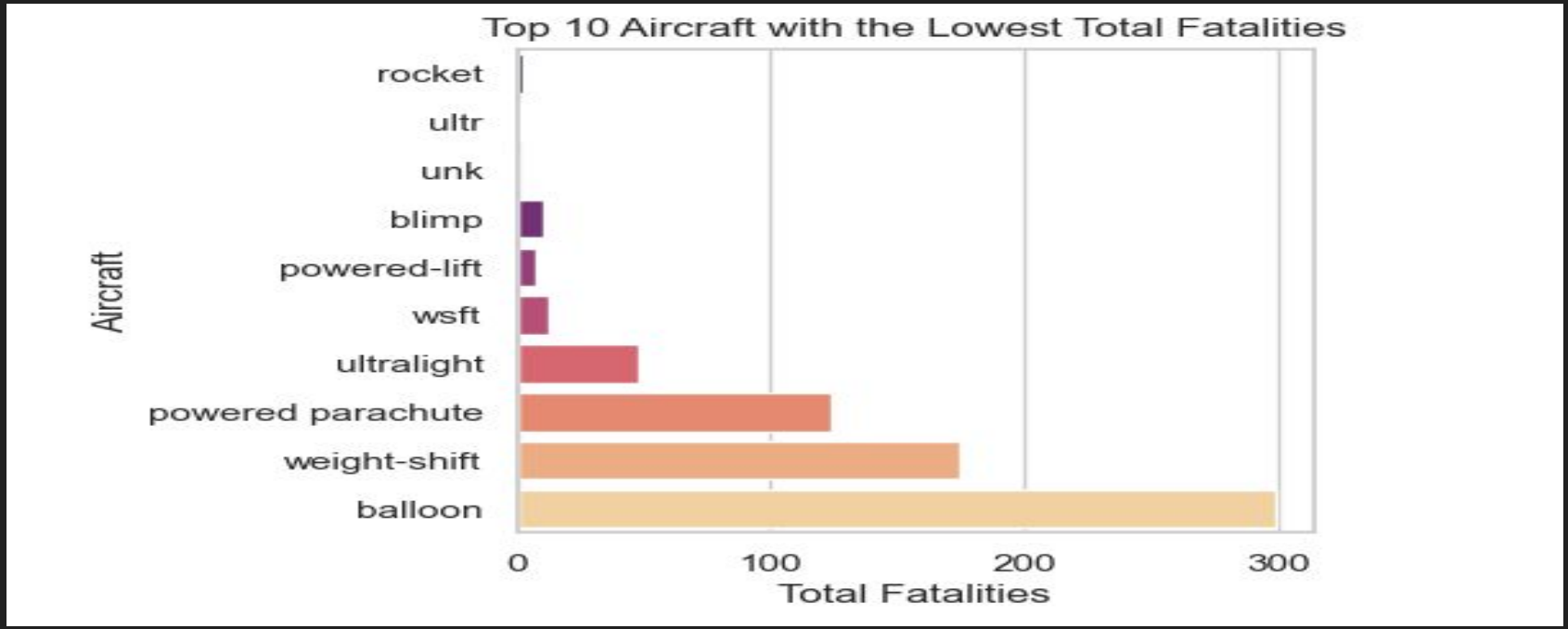
Accident rates on different makes and models

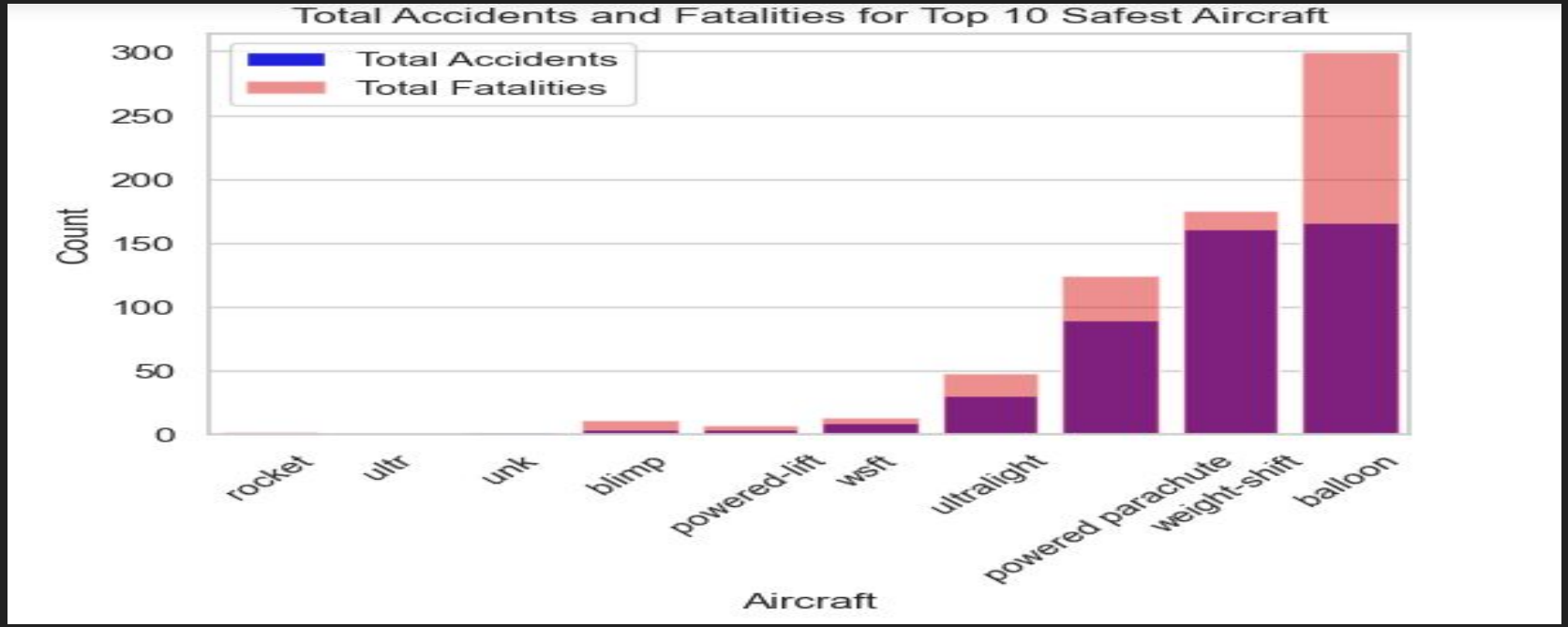



Top 20 Aircraft Models by Total Accidents
Aircraft Model
152
172
172n
pa-28-140
150
172m
172p
182
180
150m
pa-18
pa-18-150
pa-28-161
pa-28-180
pa-28-181
206b
pa-38-112
150l
140
a36
0
500
1000
1500
2000
Total Accidents
Make
cessna
piper
bell
beech

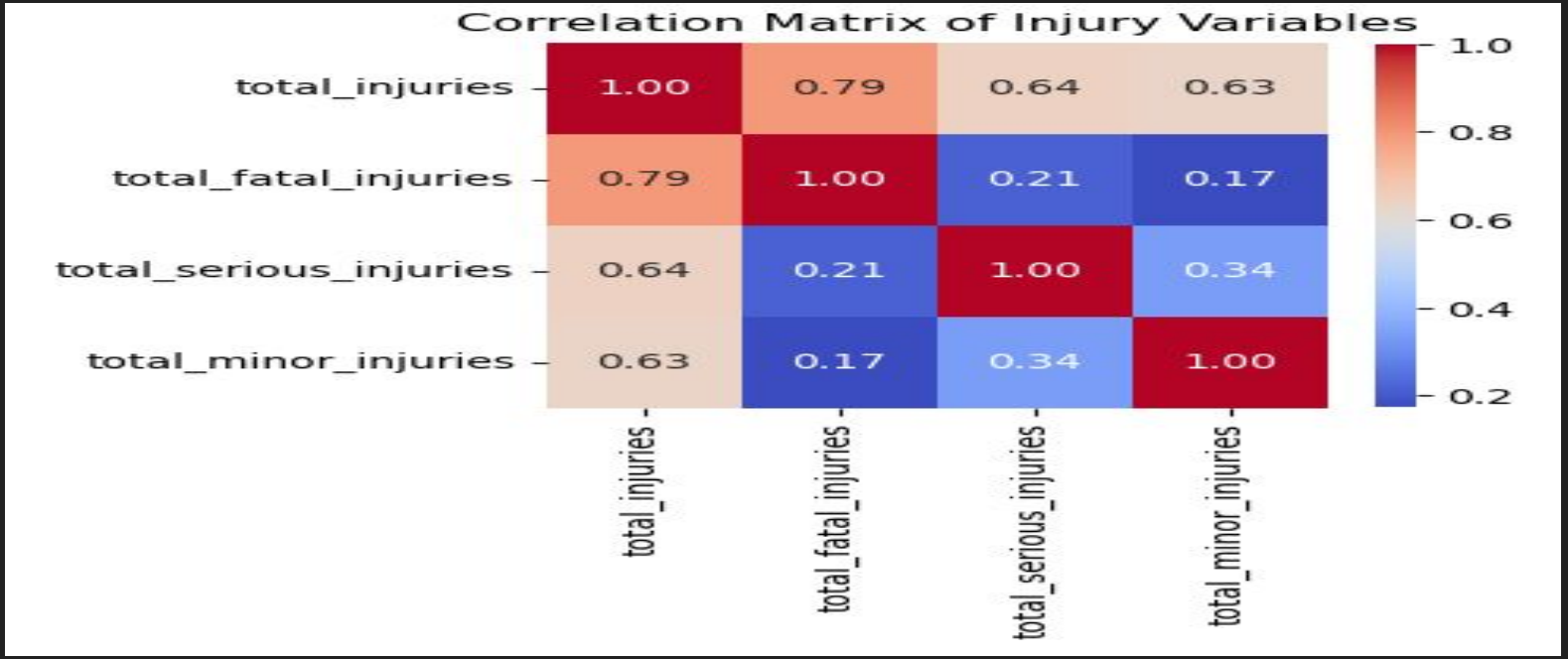
Risk assessment analysis




Top 10 Aircraft with the Lowest Total Fatalities
rocket
ultr
unk
blimp
powered-lift
wsft
ultralight
powered parachute
weight-shift
balloon
Aircraft
0
100
200
300
Total Fatalities


Total Accidents and Fatalities for Top 10 Safest Aircraft
Total Accidents
Total Fatalities
Count
300
250
200
150
100
50
0
rocket
ultr
unk
blimp
powered-lift
wsft
ultralight
powered parachute
weight-shift
balloon
Aircraft

Correlation analysis



Conclusions

From the analysis done:

- There was a large no of unknowns affecting the analysis
- There's high correlation between total_injuries and fatal injuries,minor injuries and lastly serious injuries
- Rate of accidents is decreasing over time
-

Recommendations

My recommendation are

- Find the unknown values then redo the analysis
- Continue following safety regulations to continue decreasing the accidents number