Risk Analysis for Air Travel

By Caroline Wanjiru Woto

Introduction

- Risk analysis tis the process of evaluation and managing the hazards in aviation operations. This risk can affect the safety , operation , security and general reputation of an aviation company
- Its very vital to identify the risks or hazard involved in air travel as most of the accidents results in fatal ,serious, minor and injured injuries

Definition of terms

- Fatal Injury: its an accident that result to death . According to the ICAO, a death that occurs within 30 days of the accident due to injuries sustained during the event is considered a fatal injury.
- Serious Injury; A serious injury is one that:

- Requires hospitalization for more than 48 hours, within 7 days of the accident.
- Causes severe hemorrhages, nerve, muscle, or tendon damage
- Results in a fracture of any bone (except simple fractures of fingers, toes, or nose).
- Involves second- or third-degree burns, or burns covering more than 5% of the body.
- Involves internal organ damage.

Introduction

- **Minor Injury;** A minor injury is one that does not qualify as serious but still requires medical attention. It may involve slight cuts, bruises, or other injuries that do not pose a significant threat to the person's life or long-term health.eg sprains, minor lacerations, or small burns that can be treated with basic medical care and do not require prolonged hospitalization.
- **Uninjured. This** category refers to individuals who do not sustain any injuries as a result of the accident. Despite being involved in the incident, these individuals remain physically unharmed.

.

Problem Statement

- Our company x- air is interested in in purchasing and operating airplanes for commercial and private enterprises, but do not know anything about the potential risks of aircraft. This project will focus on risk analysis in the aviation industry .
- In order to carried out risk analysis , we will use aviation accident data available on https://www.kaggle.com/datasets/khsamaha/aviation-accident-database-synopses\

Research Objectives

- The main objective of this study is to establish the risk in aviation industry .

- The specific objective of this study is

i. To determine airplane make that reported most accidents
ii. To assess why do the accidents happen
iii. To explore when do accidents happen. The establish if there is a relationship between weather and accidents,
iv. To understand the severity of the accident
v. To assess the trend (how) do accidents occur

Research Hypothesis

- $H_0$1: There is no relationship between the make /model of the plane and the accidents
- $H_0$2: Weather condition does not affect the number of accidents reported

Significance of the Study

- The study's findings will provide X-Air management with important information about aviation industry and the risks involved.
- X-Air Management will make informed decision on the model to pilot with in their portfolio expansion .
- Data science students will benefit from the study's findings, as well, because it will provide helpful insights on data cleaning , analysis & presentation

Limitation of the study

- There will be no consideration of factors that are out of the scope of The primary objective of this study .
- The researcher used secondary data and there was a lot of missing data which were substituted by the mean for numerical categories and unknown for categorical data .
- The data available for analysis was up to 2022, we missed current data that may have affected the outcome of the study due to advancement in technologies and improved security measures that has ben adopted in the recent past

Data Analysis

- In Oder to carry out data analysis , the data was cleaned using pandas
- Initial meta data showed the following

```
#metadata for df1
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 88889 entries, 0 to 88888
Data columns (total 16 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Event.Id                88889 non-null  object
 1   Investigation.Type      88889 non-null  object
 2   Location                88837 non-null  object
 3   Country                 88663 non-null  object
 4   Latitude                34382 non-null  object
 5   Longitude               34373 non-null  object
 6   Make                    88826 non-null  object
 7   Model                   88797 non-null  object
 8   Aircraft.damage         85695 non-null  object
 9   Weather.Condition       84397 non-null  object
 10  Broad.phase.of.flight   61724 non-null  object
 11  Total.Fatal.Injuries    77488 non-null  float64
 12  Total.Serious.Injuries  76379 non-null  float64
 13  Total.Minor.Injuries    76956 non-null  float64
 14  Total.Uninjured         82977 non-null  float64
 15  Event.Date              88889 non-null  object
dtypes: float64(4), object(12)
memory usage: 10.9+ MB
```

basic summary statistics for each column

```
#calculating  the basic summary statistics for each column

df2.describe()
```

	Total.Fatal.Injuries	Total.Serious.Injuries	Total.Minor.Injuries	Total.Uninjured	year	month	d
count	77488.000000	76379.000000	76956.000000	82977.000000	88889.000000	88889.000000	88889.0000
mean	0.647855	0.279881	0.357061	5.325440	1999.206662	6.591221	15.729(
std	5.485960	1.544084	2.235625	27.913634	11.888226	3.062868	8.8371
min	0.000000	0.000000	0.000000	0.000000	1948.000000	1.000000	1.000(
25%	0.000000	0.000000	0.000000	0.000000	1989.000000	4.000000	8.000(
50%	0.000000	0.000000	0.000000	1.000000	1998.000000	7.000000	16.000(
75%	0.000000	0.000000	0.000000	2.000000	2009.000000	9.000000	23.000(
max	349.000000	161.000000	380.000000	699.000000	2022.000000	12.000000	31.000(

Data Analysis

- For missing value in numerical categories , we replace with mean

In [21]:
```
#for missing value in numerical categories , we replace with mean

df2["Total.Fatal.Injuries"].fillna(df2["Total.Fatal.Injuries"].mean(),inplace=True)
df2["Total.Serious.Injuries"].fillna(df2["Total.Serious.Injuries"].mean(),inplace=True)
df2["Total.Minor.Injuries"].fillna(df2["Total.Minor.Injuries"].mean(),inplace=True)
df2["Total.Uninjured"].fillna(df2["Total.Uninjured"].mean(),inplace=True)

df2
```

Out[21]:

ıdition	Broad.phase.of.flight	Total.Fatal.Injuries	Total.Serious.Injuries	Total.Minor.Injuries	Total.Uninjured	year	month	day
UNK	Cruise	2.0	0.000000	0.000000	0.00000	1948	10	24
UNK	Unknown	4.0	0.000000	0.000000	0.00000	1962	7	19
IMC	Cruise	3.0	0.279881	0.357061	5.32544	1974	8	30
IMC	Cruise	2.0	0.000000	0.000000	0.00000	1977	6	19
VMC	Approach	1.0	2.000000	0.357061	0.00000	1979	8	2
...	...	...	...	...	...	...	...	...
NaN	NaN	0.0	1.000000	0.000000	0.00000	2022	12	26
NaN	NaN	0.0	0.000000	0.000000	0.00000	2022	12	26
VMC	NaN	0.0	0.000000	0.000000	1.00000	2022	12	26
NaN	NaN	0.0	0.000000	0.000000	0.00000	2022	12	26
NaN	NaN	0.0	1.000000	0.000000	1.00000	2022	12	29

- For categorical data, the researcher used unknown to replace missing data . This will not distort the data

In [22]:
```
#sorting columns with missing data
missing_data=df2.isnull().sum()
missing_data.sort_values(ascending= False)
```

Out[22]:
```
Broad.phase.of.flight     27165
Weather.Condition          4492
Aircraft.damage            3194
Country                     226
Model                        92
Make                         63
Location                     52
day                           0
month                         0
year                          0
Total.Uninjured               0
Total.Minor.Injuries          0
Total.Serious.Injuries        0
Total.Fatal.Injuries          0
Investigation.Type            0
Event.Id                      0
dtype: int64
```

Getting unique values and their distribution of values in the categorical columns.

In [23]:
```
df2['Broad.phase.of.flight'].unique()
```

Out[23]:
```
array(['Cruise', 'Unknown', 'Approach', 'Climb', 'Takeoff', 'Landing',
       'Taxi', 'Descent', 'Maneuvering', 'Standing', 'Go-around', 'Other',
       nan], dtype=object)
```

Results And Discussions

- Cleaned data was transferred to excel for further analysis in Tableau

```
In [30]:  #safe df to excel for analysis using Tableau
          df2.to_csv('AviationData1.csv',index= False)
```

- To understand the Risk in air industry, the Analyst sort was guided by 5 w 1h methodology.5w 1h examines
- **W**hat- The material(model/Make of the aircrafts)
- **W**hy – The probable root causes (relationship of weather condition and accidents
- **W**hen- The time the accidents occurred
- **W**here – Where did most accidents occurred
- **W**ho- Fatality levels and the people affected
- **H**ow- The pattern/when do most accidents occurred

Results And Discussions

- Cessna Aircraft and model 152 recorded the highest no of accidents i.e 22227 and 2367 accidents respectively
- Below models and makes topped the list with lowest accidents
- Make

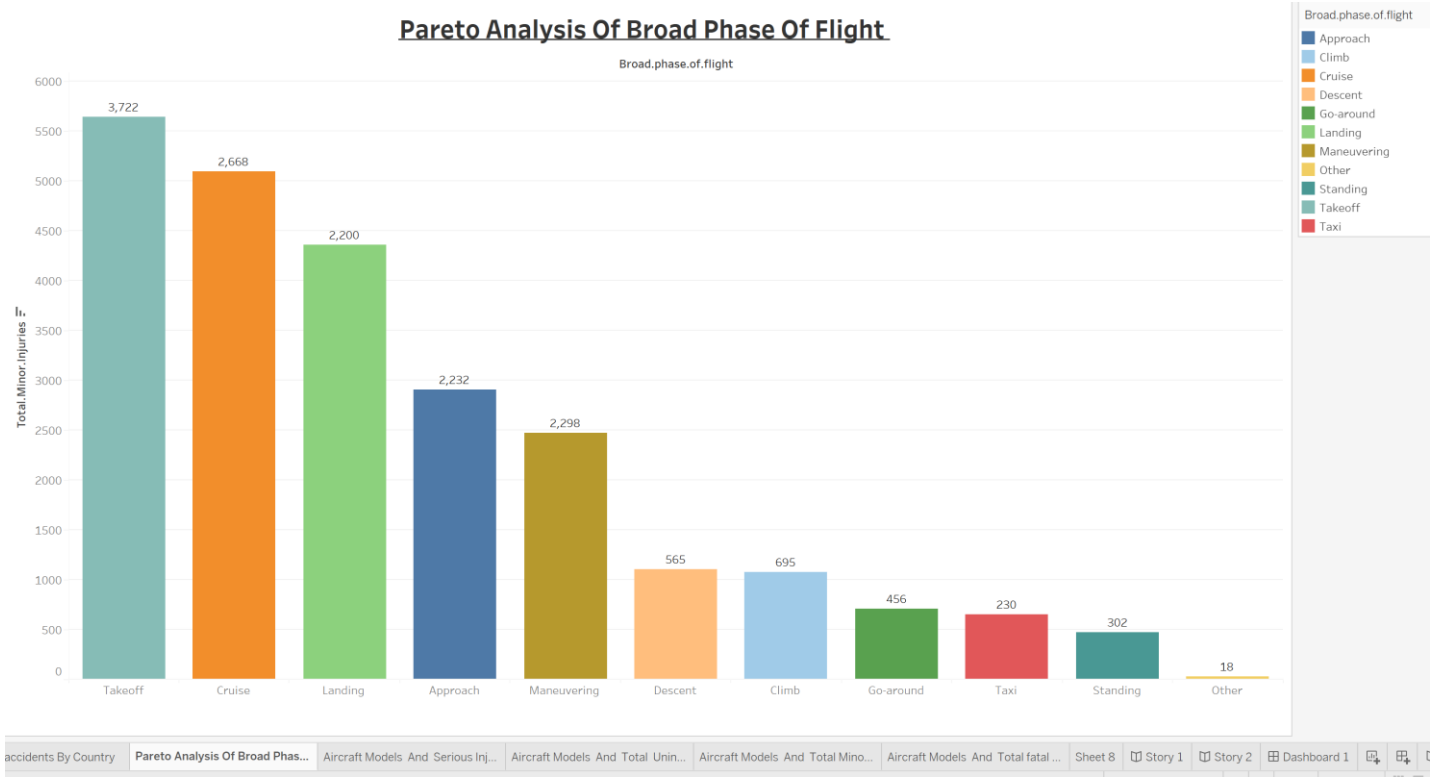
```
BOYD BRUCE             1
Casten                 1
SEACE DAVID A          1
POWERCHUTE             1
Gray Jim Robert        1
Name: Make, Length: 8237, dtype: int64
```

- Model

```
                      ...
DUNCAN/VARIEZE         1
B 206 SERIES 1         1
Airborne Edge-X        1
Renegade II            1
28                     1
Name: Model, Length: 12318, dtype: int64
```

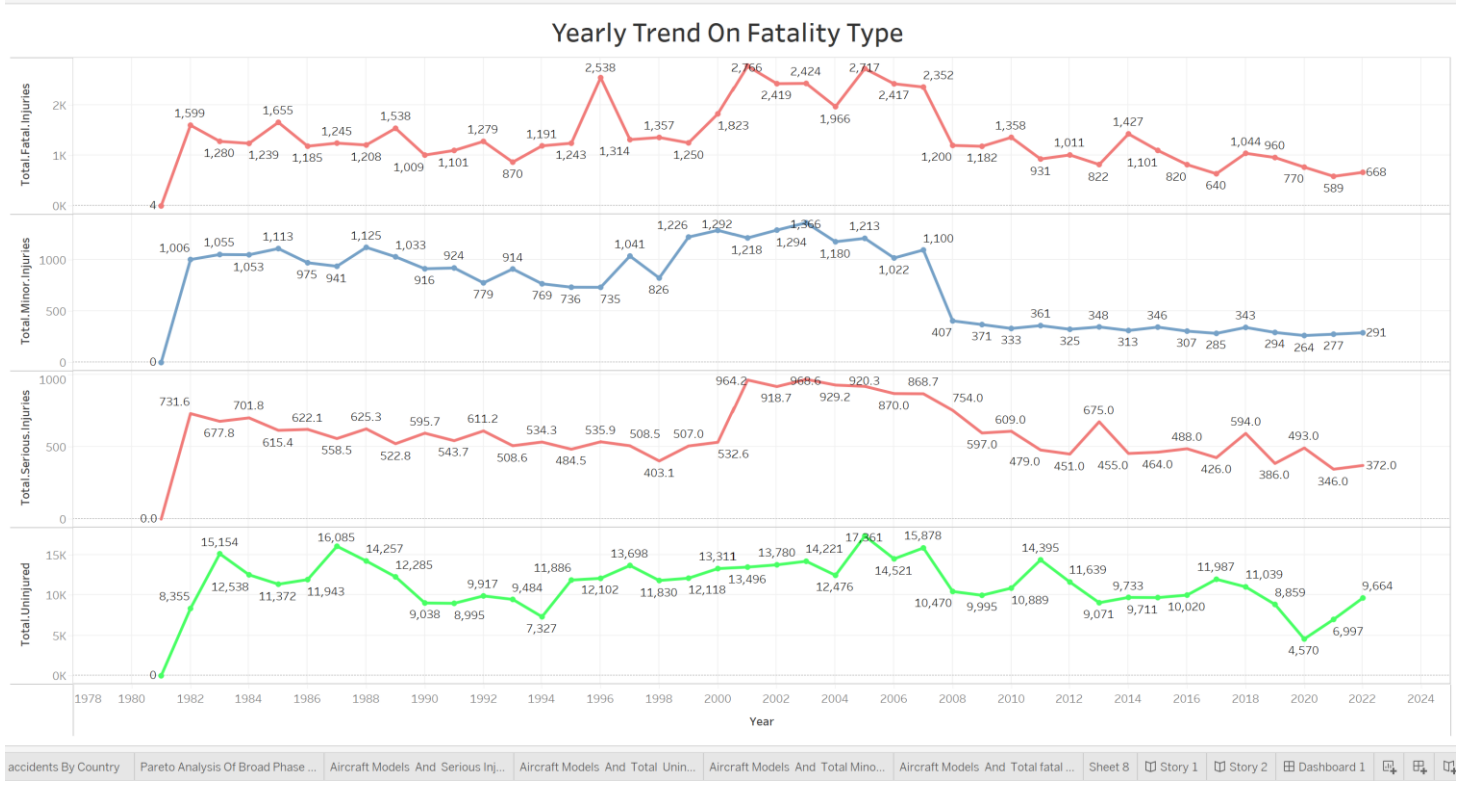
Results And Discussions

- The available data showed most accidents occurred while takeoff, cruising, landing, and maneuvering ..



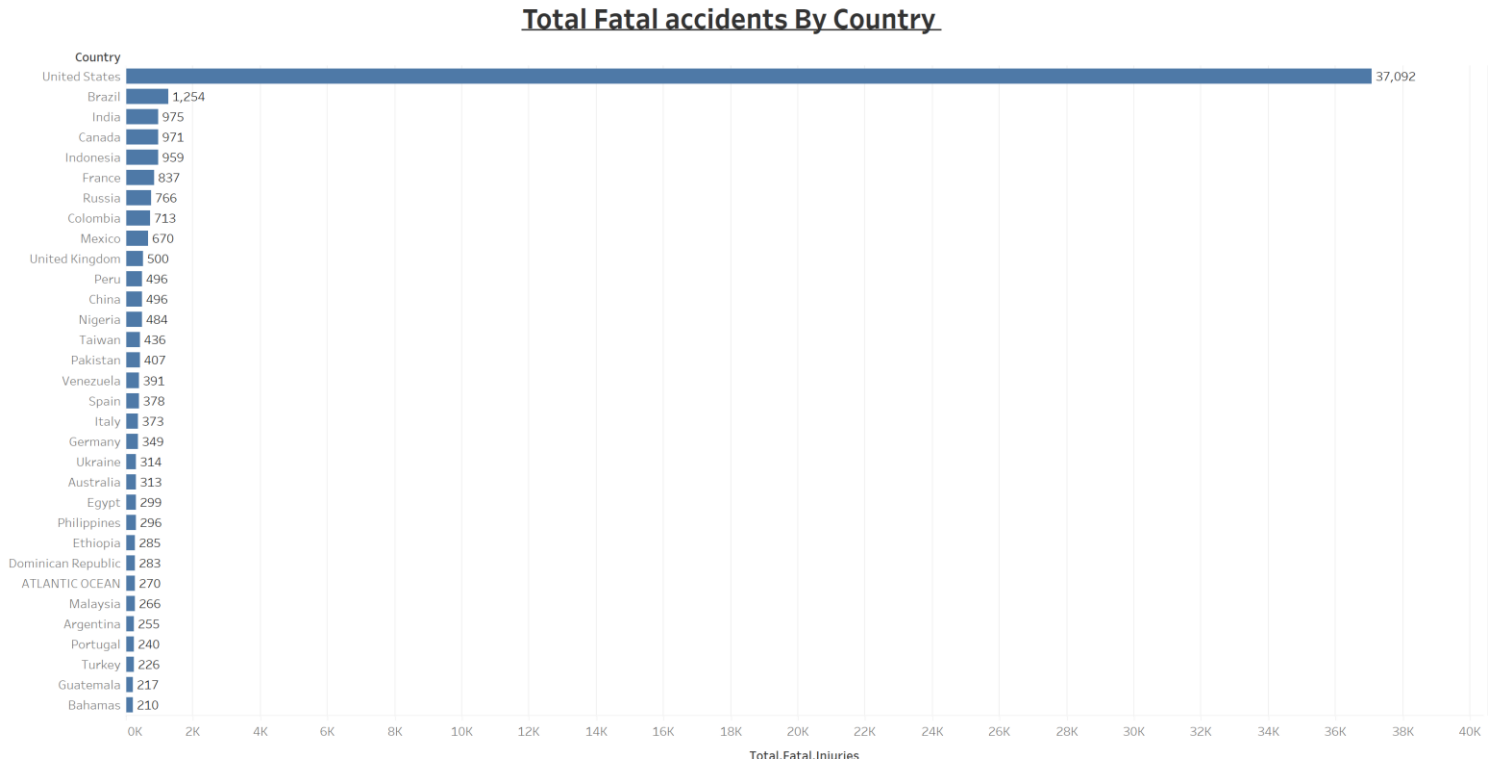
Results And Discussions

- In the recent past , the no of accidents and fatality rate is on the decline ,
- We have an upward trajectory on the total uninjured as shown below



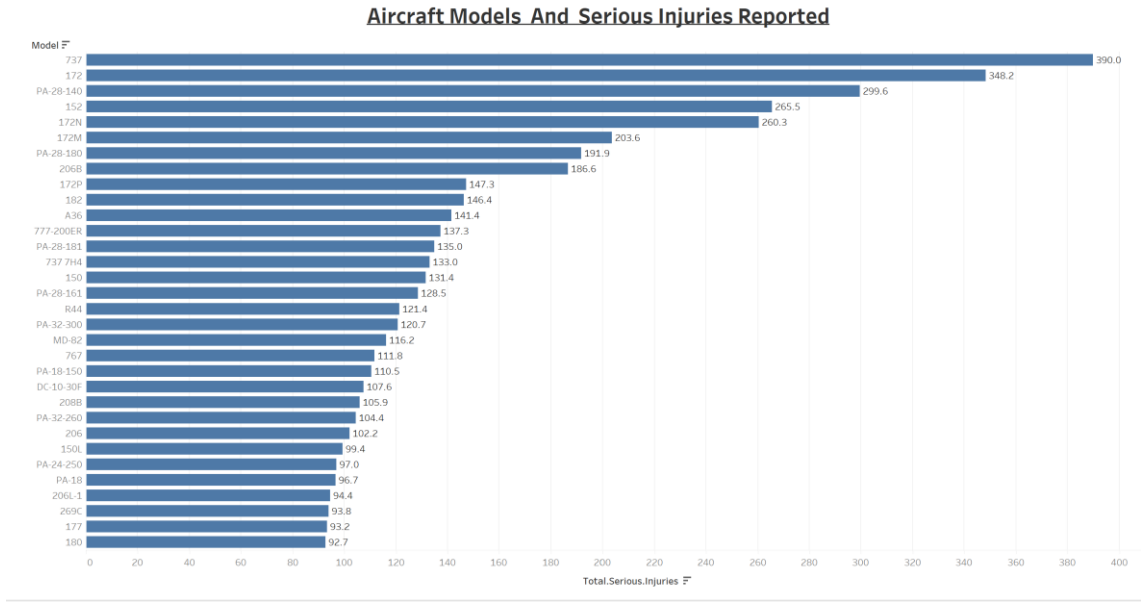
Results And Discussions

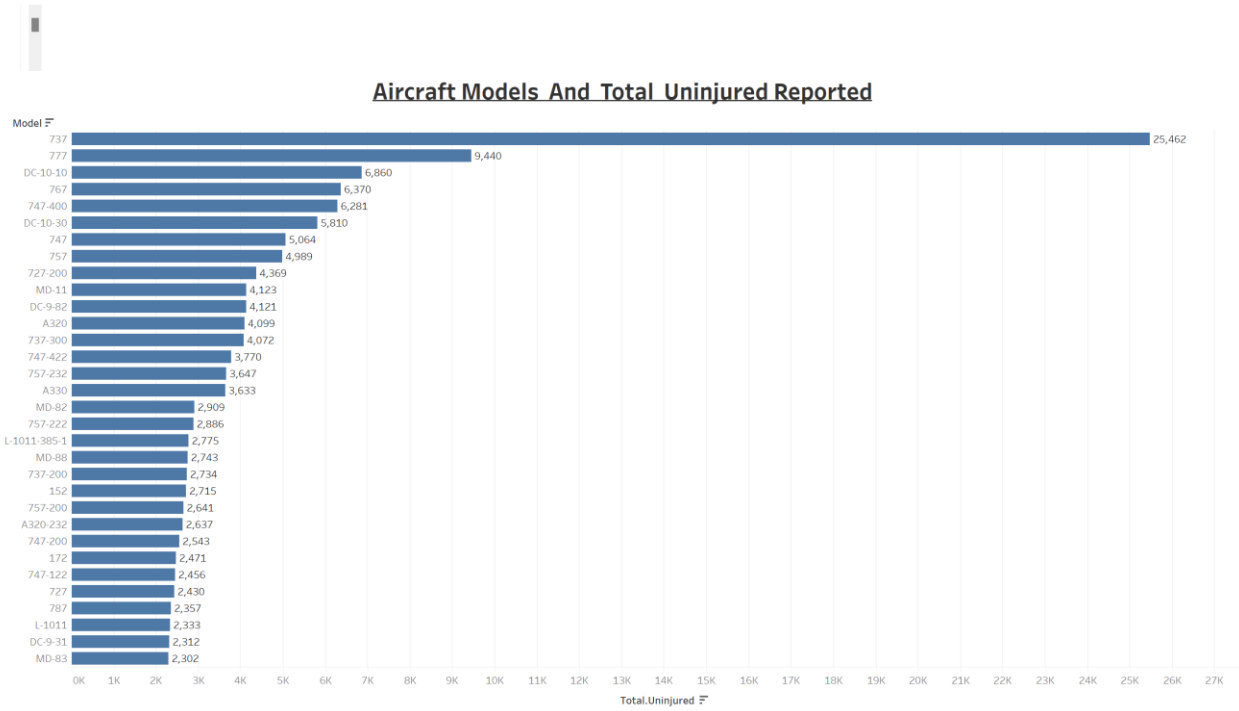
- USA recorded most accidents accounting to 92%



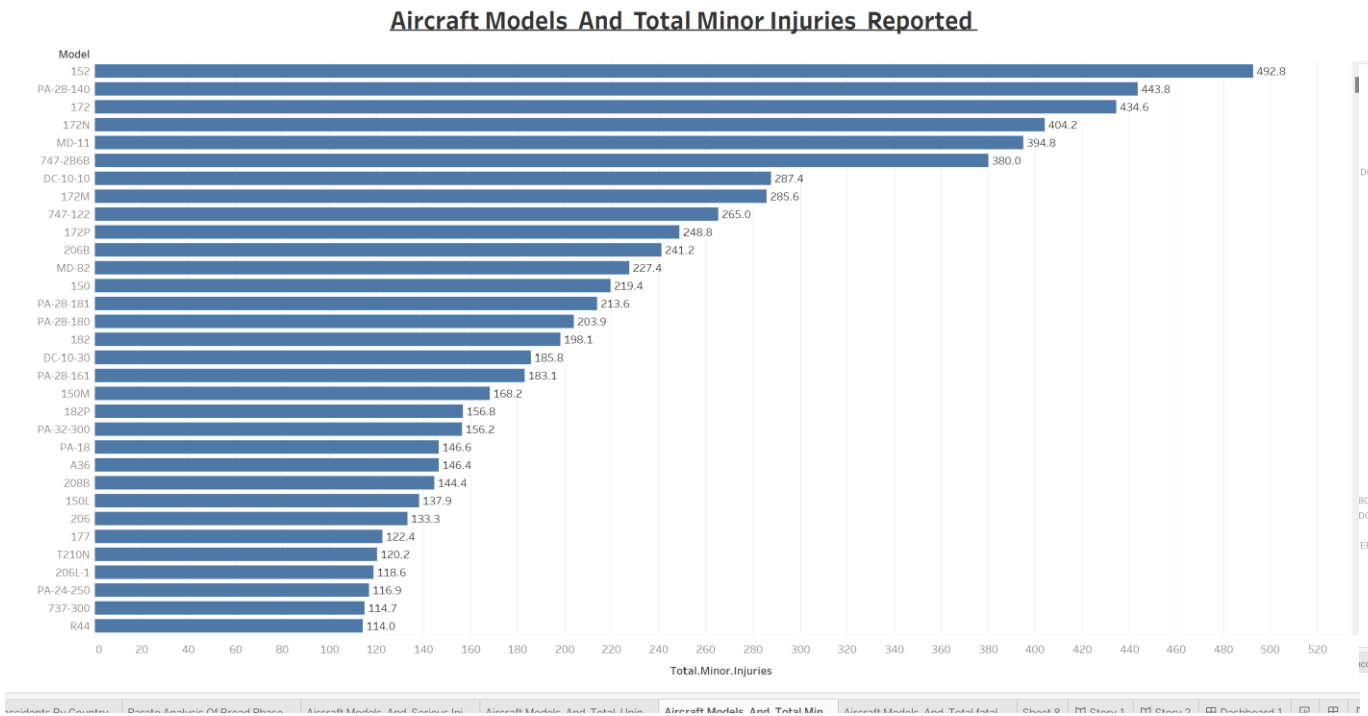
Results And Discussions

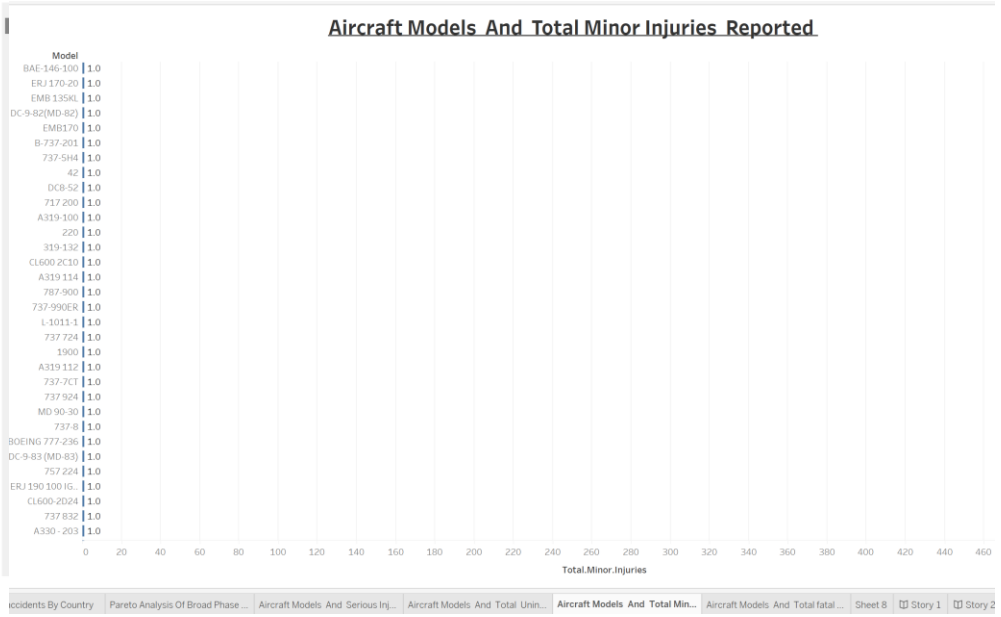
- Model 737 recorded the highest no of serious injuries and highest uninjured accidents



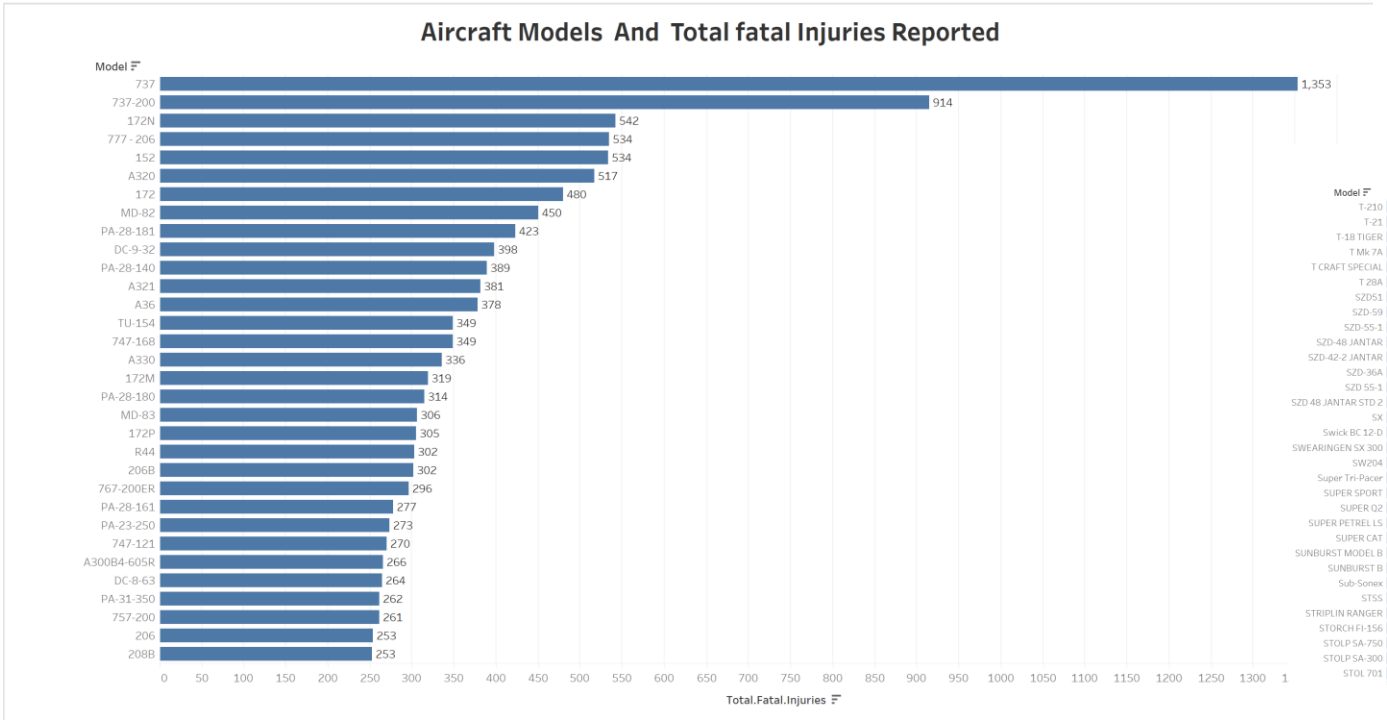


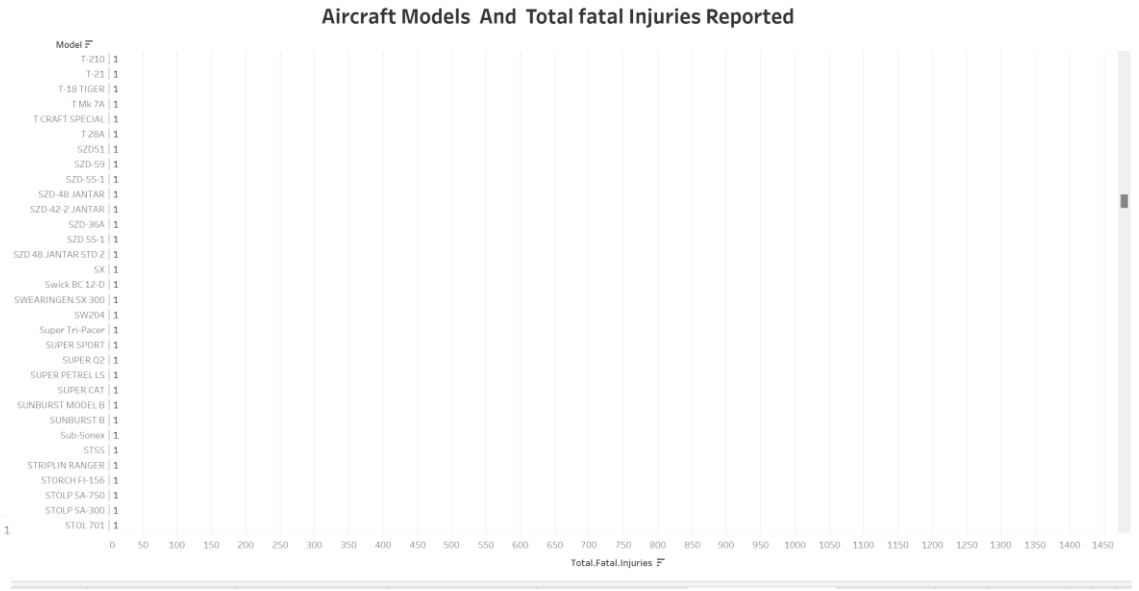
Results And Discussions- models with Minor Injuries



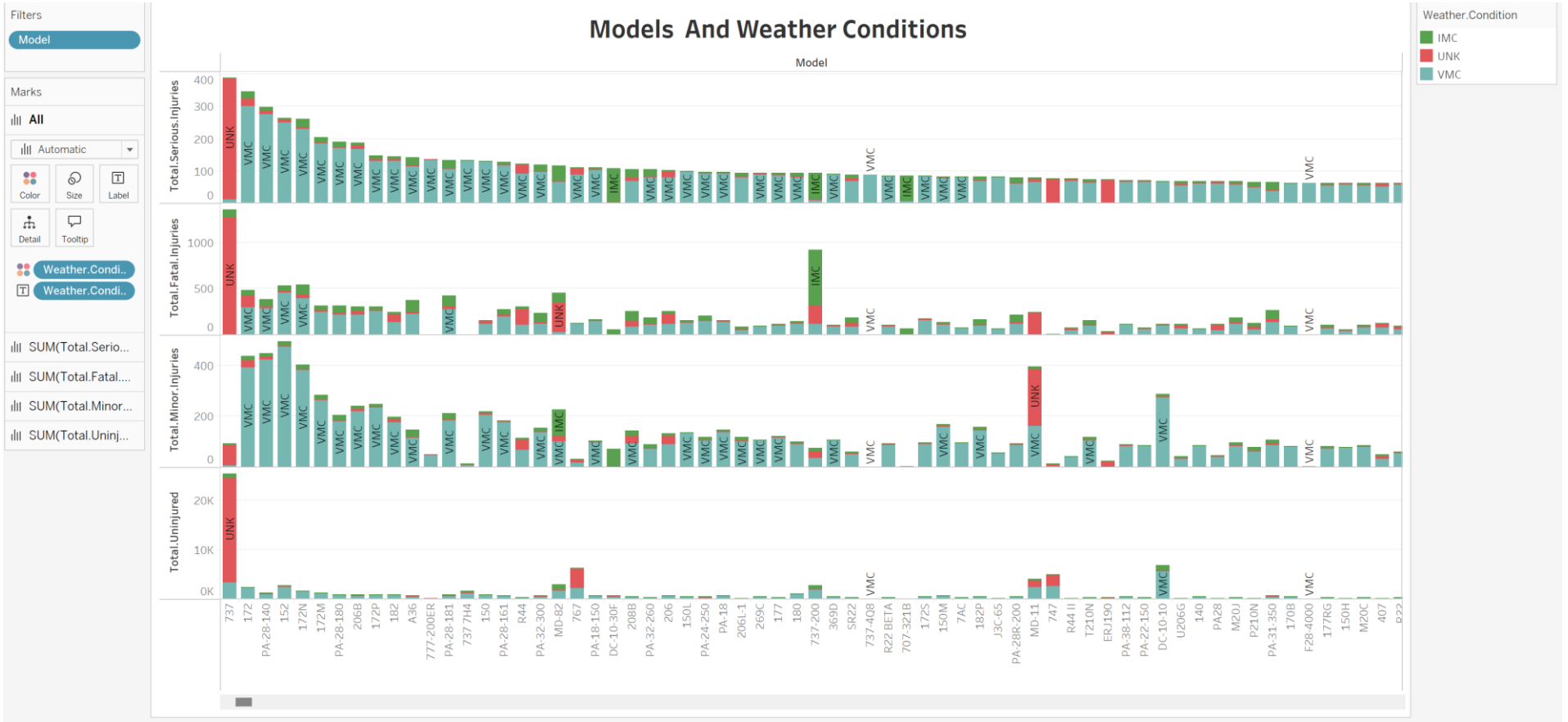


Results And Discussions- Models with fatal injuries





Results And Discussions- Models and Weather conditions



Conclussion and recommendations

- The is a high relationship between accident and Broad.phase.of.flight‘ as most accidents occurred during take off
- Most accidents were reported in the USA though accidents recorded in other countries had he highest number of fatality .
- Model 737 was the safest though lead in fatality and serious injuries as shown

Conclusion and recommendations

- X-Air can pilot on model 737 . Though the analyst recommend further understanding on operational costs and customer ratings.
- Further research on current safety improvements in the aviation industry with special emphasis on model 737 is also recommended
- Research hypothesis were nullified