# CRASHED AIRPLANES

# INTRODUCTION..

Over the last centuries, the world has seen several airplane crashes both in the sky and on the ground. The development of technology aims to decrease technological difficulties and human errors. However, fatalities and aircraft damages caused by crashes do not cease to exist.

Our results found in this project will benefit the ongoing investigations into this important topic. Understanding what factors cause airplane crashes helps aviation industries make continuous improvement in flight safety.

# OBJECTIVES:

- Our objective is to perform an Exploratory Data Analysis (EDA) to determine the common cause/reason of airplane crash, countries with maximum/minimum airplane crashes, fatalities vs survived ratio and any other interesting trend. In addition, to develop a Model determining the following;
  - NMF
  - LDA

# DESCRIPTION OF THE DATA:

- This dataset presents number of crashed airplanes since 1908, where we have 5268 record, Also we have 13 features in this dataset.

Out[7]:

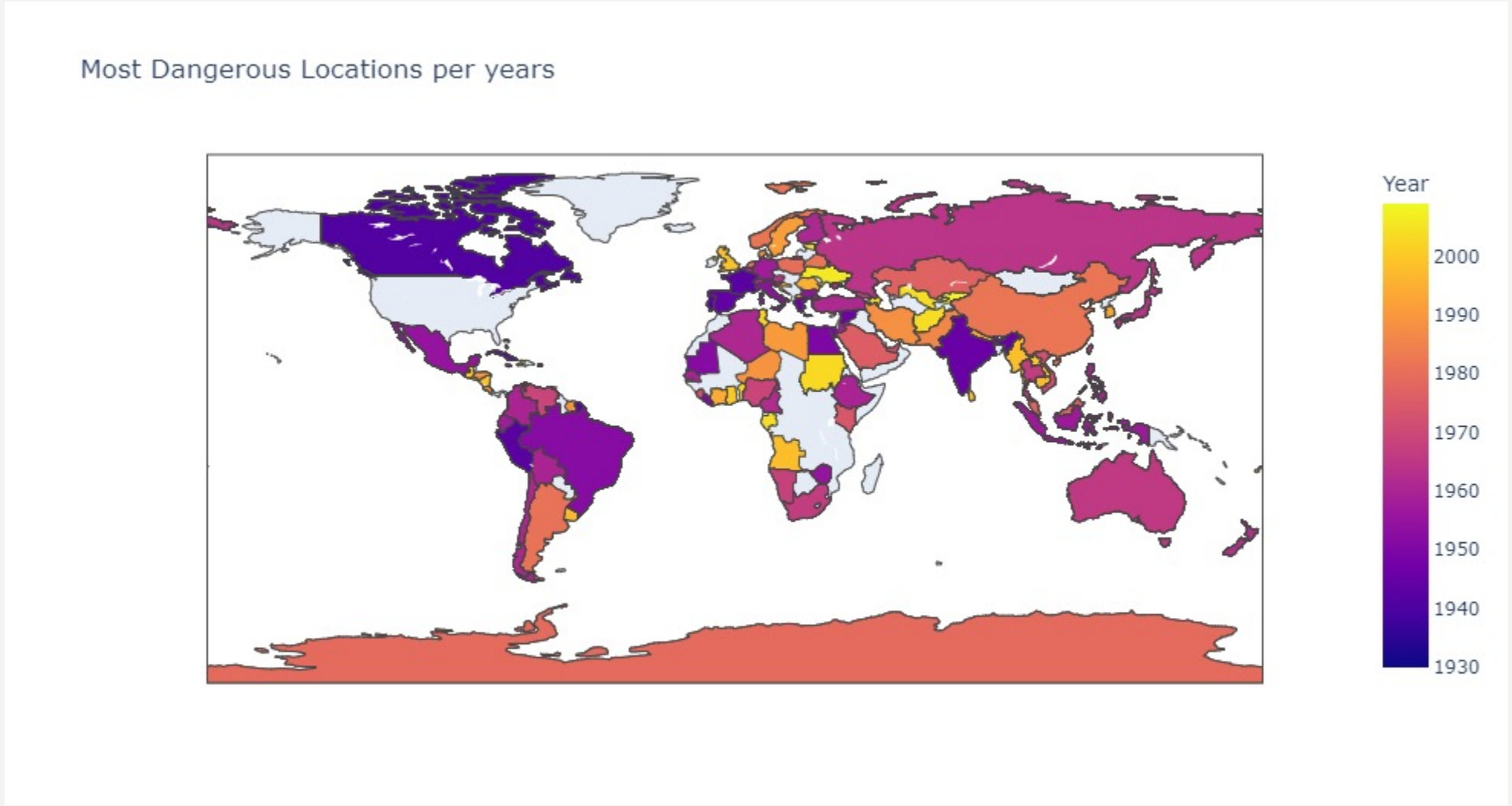
| | Date | Time | Location | Operator | Flight # | Route | Type | Registration | cn/ln | Aboard | Fatalities | Ground | Summary | Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **208** | 01/19/1930 | 18:23 | Oceanside, California | Maddux Airlines | 7 | Aqua Caliente, Mexico - Los Angeles | Ford 5-AT-C Tri Motor | NC9689 | 5-AT-046 | 16.0 | 16.0 | 0.0 | While en route to Los Angeles, the pilot, flyi... | 1930 |
| **236** | 03/31/1931 | 10:45 | Bazaar, Kansas | Trans Continental and Western Air | 599 | Kansas City - Wichita - Los Angeles | Fokker F10A Trimotor | NC-999 | 1063 | 8.0 | 8.0 | 0.0 | Shortly after taking off from Kansas City, one... | 1931 |
| **334** | 08/31/1934 | 23:42 | Amazonia, Missouri | Rapid Air Transport | 6 | Omaha - St. Joseph | Stinson SM-6000B | NC10809 | 5004 | 5.0 | 5.0 | 0.0 | The plane crashed about 11 miles from St. Jose... | 1934 |

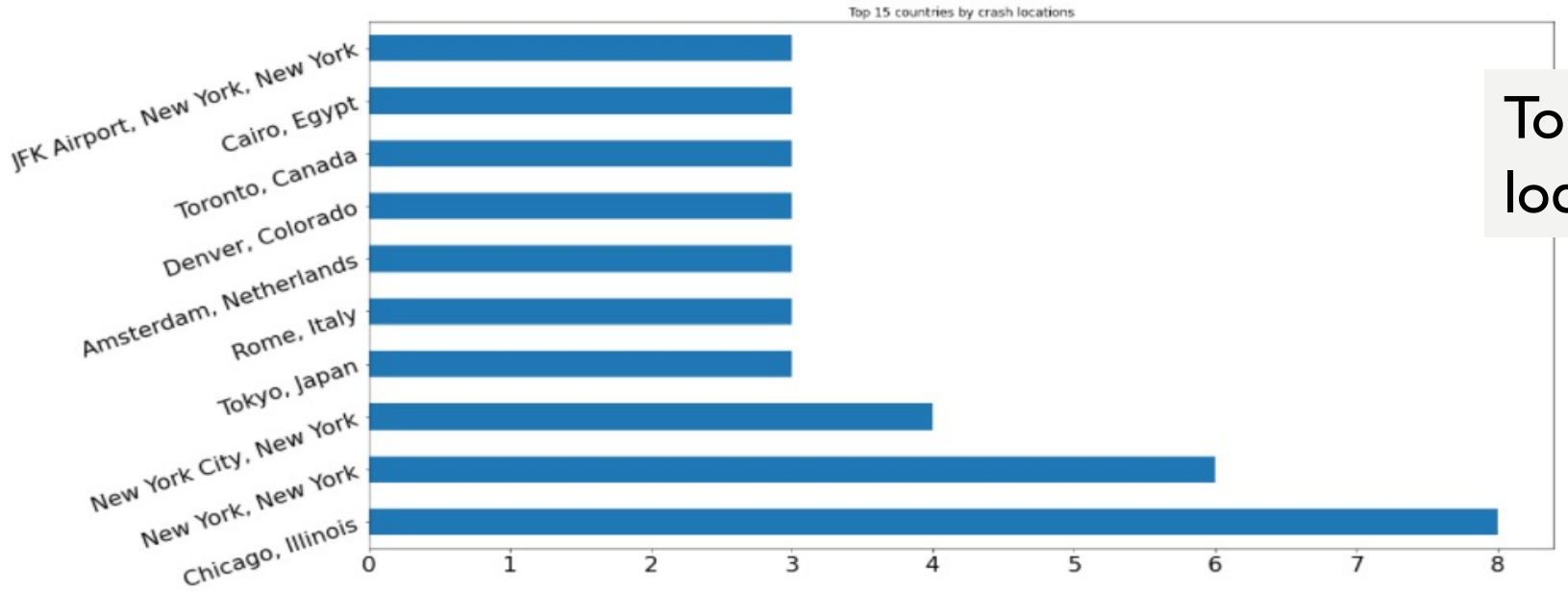
# DATA PREPARATION:

- First, we check for duplicated and null data and drop it
- Second, add year column to the dataset.
- Third add Survivor column by subtract number of Aboard and Fatalities on each airplane.
- Forth, add 'summary clean' column that contain summary after text cleaning using StopWords, Stemmer.

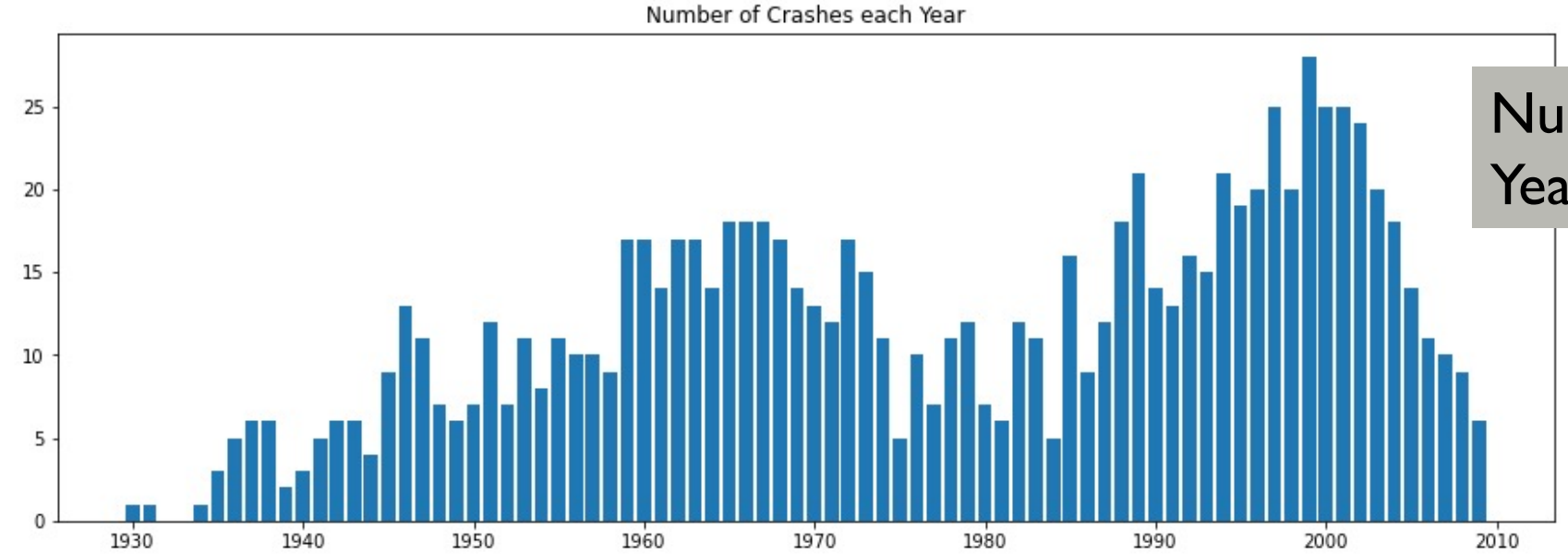
| | Date | Time | Location | Operator | Flight # | Route | Type | Registration | cn/ln | Aboard | Fatalities | Ground | Summary | Year | Survivor | Summary_clean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 01/19/1930 | 18:23 | Oceanside, California | Maddux Airlines | 7 | Aqua Caliente, Mexico - Los Angeles | Ford 5-AT-C Tri Motor | NC9689 | 5-AT-046 | 16.0 | 16.0 | 0.0 | While en route to Los Angeles, the pilot, flyi... | 1930 | 0.0 | en route los angeles pilot flying low altitude... |
| 5 | 03/31/1931 | 10:45 | Bazaar, Kansas | Trans Continental and Western Air | 599 | Kansas City - Wichita - Los Angeles | Fokker F10A Trimotor | NC-999 | 1063 | 8.0 | 8.0 | 0.0 | Shortly after taking off from Kansas City, one... | 1931 | 0.0 | shortly taking kansas city one aircrafts wings... |

# RESULT:

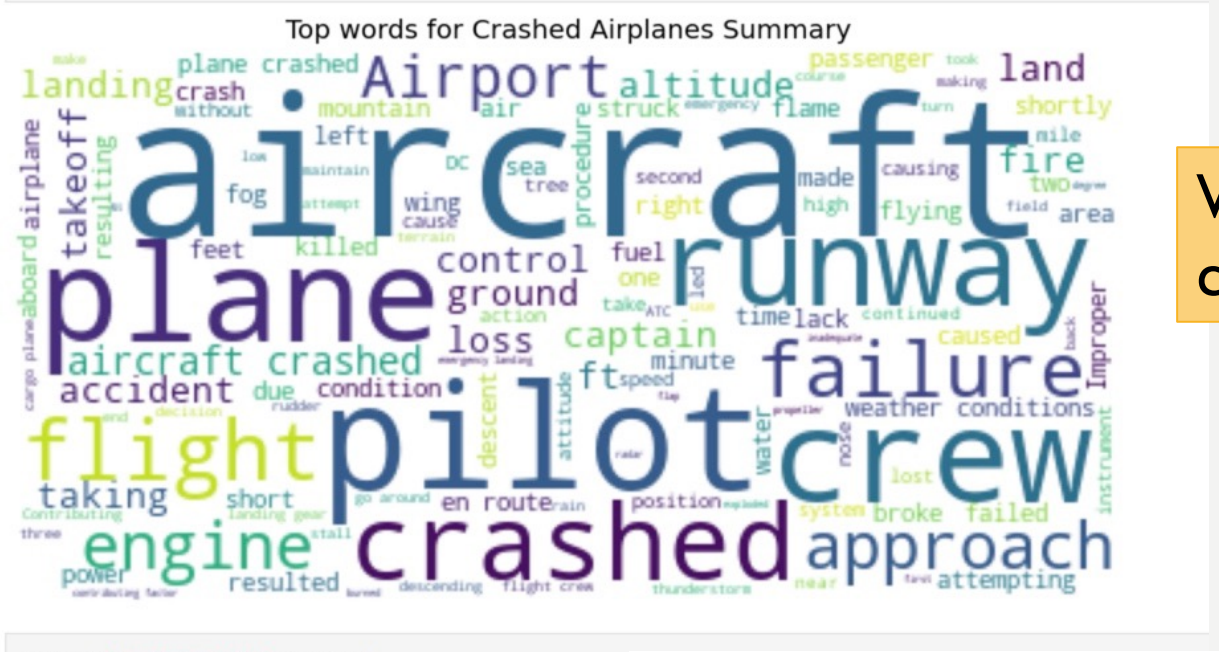

Most Dangerous Locations per years
Year
2000
1990
1980
1970
1960
1950
1940
1930




Top 15 countries by crash locations..

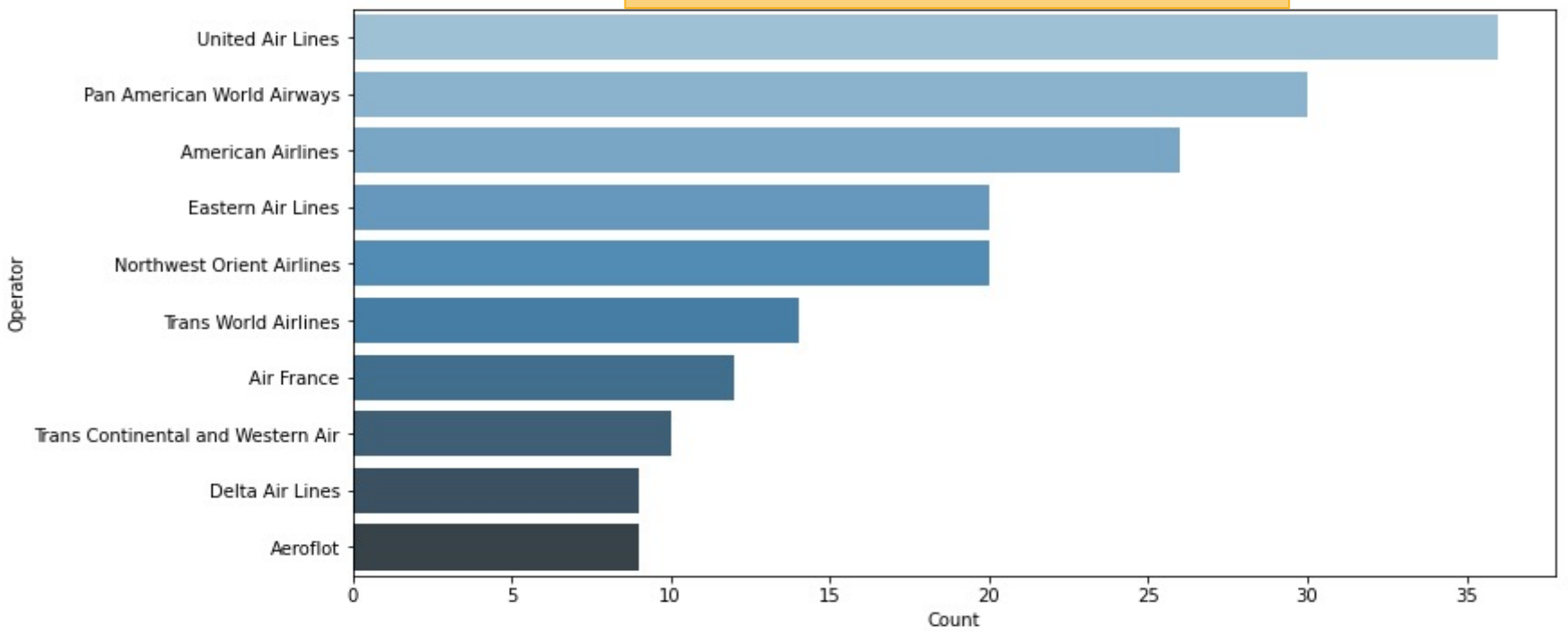



Number of Crashes each Year

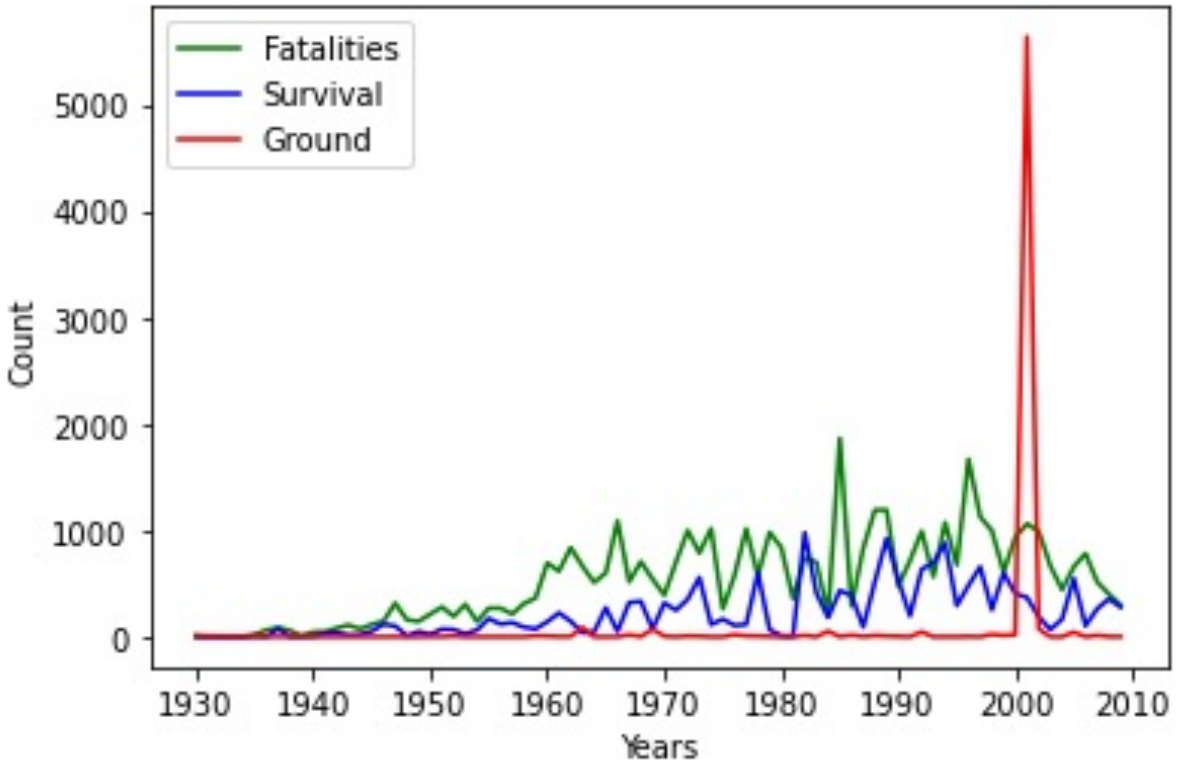



Word cloud for the most common terms

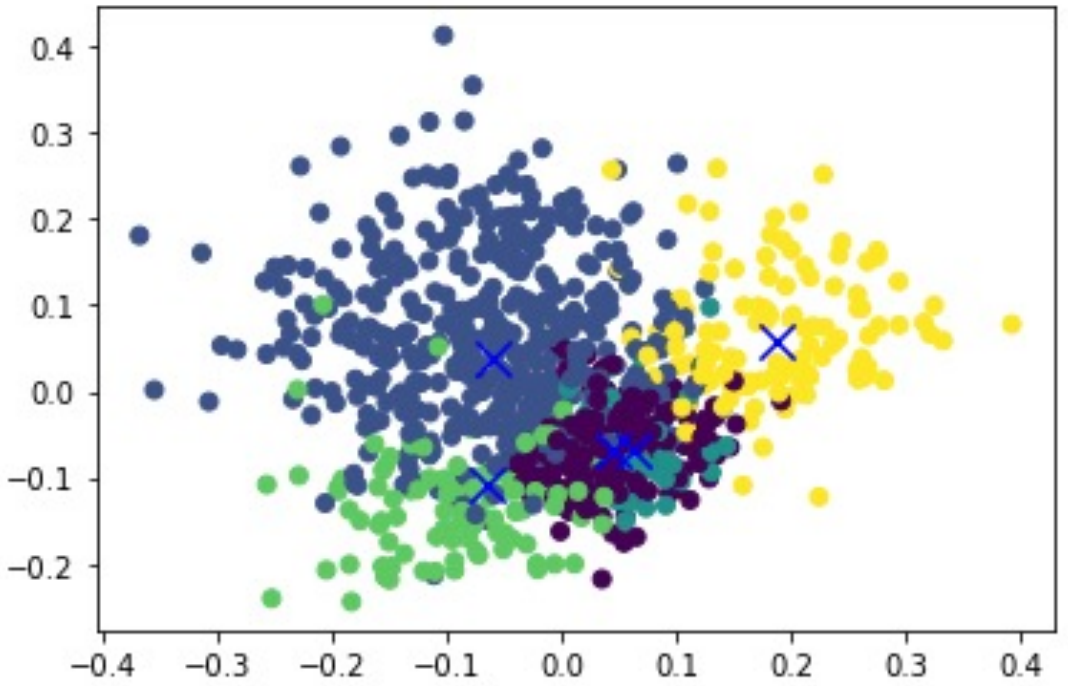
Total Count by Operator



Fatalities, Survival and on Ground per year:




Clustering the most common terms:



# NMF - Model

| | Summary_clean | Operator | Location |
|---|---|---|---|
| 208 | en rout los angel pilot fli low altitud due poor weather condit tri turn back left turn left wing struck hill crash burn advers weather condit plane flew sudden squall caus pilot lose control | Maddux Airlines | Oceanside, California |
| 236 | short take kansa citi one aircraft wing separ flight penetr thunderstorm experienc strong turbul ice plane crash wheat field wingaileron flutter brought moistur leak wing interior weaken glue bond wooden spar notr dame footbal coach knute rockn kill | Trans Continental and Western Air | Bazaar, Kansas |

[54]:

| | Weather | Fire | Bomb | Failur | Mountain | Crash | dominant_topic |
|---|---|---|---|---|---|---|---|
| **Doc0** | 0.04 | 0.01 | 0.00 | 0.03 | 0.11 | 0.00 | 4 |
| **Doc1** | 0.03 | 0.01 | 0.08 | 0.03 | 0.01 | 0.10 | 5 |
| **Doc2** | 0.00 | 0.00 | 0.00 | 0.28 | 0.00 | 0.00 | 3 |
| **Doc3** | 0.00 | 0.15 | 0.00 | 0.03 | 0.00 | 0.00 | 1 |
| **Doc4** | 0.15 | 0.00 | 0.04 | 0.01 | 0.01 | 0.01 | 0 |
| **...** | ... | ... | ... | ... | ... | ... | ... |
| **Doc655** | 0.00 | 0.00 | 0.04 | 0.08 | 0.11 | 0.05 | 4 |
| **Doc656** | 0.02 | 0.01 | 0.23 | 0.00 | 0.03 | 0.03 | 2 |
| **Doc657** | 0.00 | 0.00 | 0.00 | 0.05 | 0.11 | 0.00 | 4 |
| **Doc658** | 0.06 | 0.00 | 0.03 | 0.04 | 0.04 | 0.06 | 0 |
| **Doc659** | 0.00 | 0.24 | 0.00 | 0.06 | 0.03 | 0.00 | 1 |

660 rows × 7 columns

[55]:

# LDA-Model

```
In [55]: X_train.head()
```

Out[55]:

| | Summary_clean | Operator | Location |
|---|---|---|---|
| 1121 | plane crash north slope santa ynez peak mile northwest santa barbara poor weather failur crew maintain minimum altitud rout flown reason unknown | Southwest Airways | Refugio Pass, near Santa Barbara, California |
| 3529 | aircraft crash return airport report sever vibrat attempt land aircraft crash field bounc slid mobil home ground handler forgot close air start access door crew failur control monitor flight path air speed aircraft detect vibrat | Galaxy Airlines | Reno, Nevada |
| 4773 | aircraft explod caught fire minut schedul takeoff five member cabin crew peopl aboard plane time wit said heard explos flame erupt aboard aircraft ntsb investig report center fuel tank explod follow right tank minut later caus explos unclear center fuel tank locat near air condit pack generat heat run nonstop prior explos | Thai Airways | Bangkok, Thailand |
| 833 | crash loos power engin due fuel starvat accumul carburetor ice follow loss power engin result fuel starvat reason fuel starvat could determin | American Airlines | Michigan City, Indiana |
| 938 | execut second instrument approach shannon airport aircraft struck ground feet approach end intend runway broke apart burn continu instrument approach altitud insuffici clear terrain failur fluoresc light may contribut caus plane name clipper empress sky | Pan American World Airways | Shannon, Ireland |

```
In [56]: Wtest = clf.transform(vectorizer.transform(X_test.Summary_clean[:10]))
```

| | Weather | Fire | Crash | dominant_topic |
|---|---|---|---|---|
| Doc0 | 0.06 | 0.06 | 0.87 | 2 |
| Doc1 | 0.06 | 0.06 | 0.89 | 2 |
| Doc2 | 0.06 | 0.05 | 0.89 | 2 |
| Doc3 | 0.08 | 0.08 | 0.83 | 2 |
| Doc4 | 0.06 | 0.05 | 0.89 | 2 |
| ... | ... | ... | ... | ... |
| Doc655 | 0.07 | 0.07 | 0.86 | 2 |
| Doc656 | 0.06 | 0.06 | 0.89 | 2 |
| Doc657 | 0.07 | 0.07 | 0.86 | 2 |
| Doc658 | 0.06 | 0.06 | 0.87 | 2 |
| Doc659 | 0.08 | 0.08 | 0.84 | 2 |

# CONCLUSION..

The main objective of our project is to raise awareness of flight safety and better understand its problems and progress, so that aviation industries can continue to improve. We hope that more information understanding will lead to industry changes that save lives.

# THANK YOU