#### Flight Risk Assessment Analysis

My name Humera Yasmeen, completed my Bachelors of Engineering in Electronics and Communication Engineering, Anna University. I have experience in software industry for over 10 years. Front-end developer, worked across various domains like banking & insurance, healthcare, hospitality, automation and aviation sectors. Capable of handling backend development and maintainence.

Working as Senior Technical Lead in software development under Agile project in aviation sector for flight safety management.

**Problem Statement:**

Working in the project which ensures the safety management for the flight and risk assessment. It is a general practice in the aviation sector, to collect the data from the crew on their own health conditions, sleep duration, flight conditions, weather conditions, flight operational status to ensure the safety. In case of the accident or any failures, there will be a report submitted with Reporting Program Tool(RPT). There are around 11 steps involved in this and inputs will be provided by different people and directed to concerned department and risk matrix will be mapped to it, which requires lot of tedious work.

With the initial data provided by concerned professionals, collecting them and predicting the potential risk was a monotonous process. Also, reading the comments entered, were to be read manually and concerned actions need to be taken, which would delay the analysis process.

**Problem Solution:**

* To solve this business problem, I followed the technique of collecting the aircraft accident data for last 10 years for US flights and the form filled before the flight takeoff.
* Did initial code clean up, removed unwanted columns, formatted date and understood the distribution of the flight accidents with time series forecast.
* With Natural Language Processing with the comments on findings, was able to categorize the accidents as environmental issues, technical problems with the flights, pilot’s action towards an issue.
* Furthermore, was able to screw down how the age of the flight has implications on the flight crash.
* With respect to the pilot, was able to drill down the intial data to check if the psycological patterns have impact the decision making skills.

**Techniques Used:**

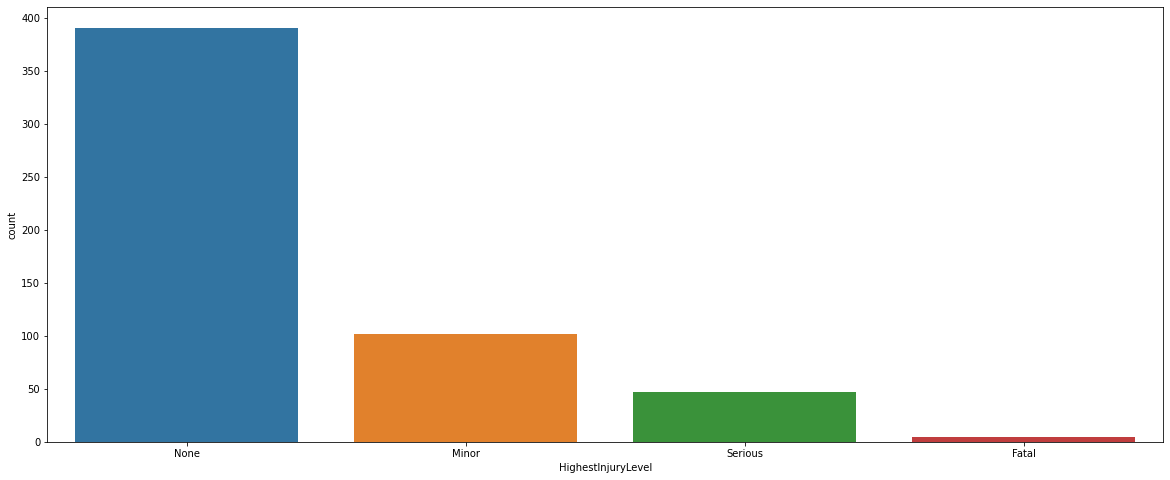
* Python, numpy.
* Supervised learning technique(Logistic Regression, Random Forest, Multinomial Naive Bayes, Support Vector Classifier, Decision Tree Classifier, K Nearest Neighbour, Gaussian Naive Bayes).
* Natural Language Processing with count vectorizer(word cloud, lemmatization).
* Flask for model deployment in local machine .

**Organisation Impact:**

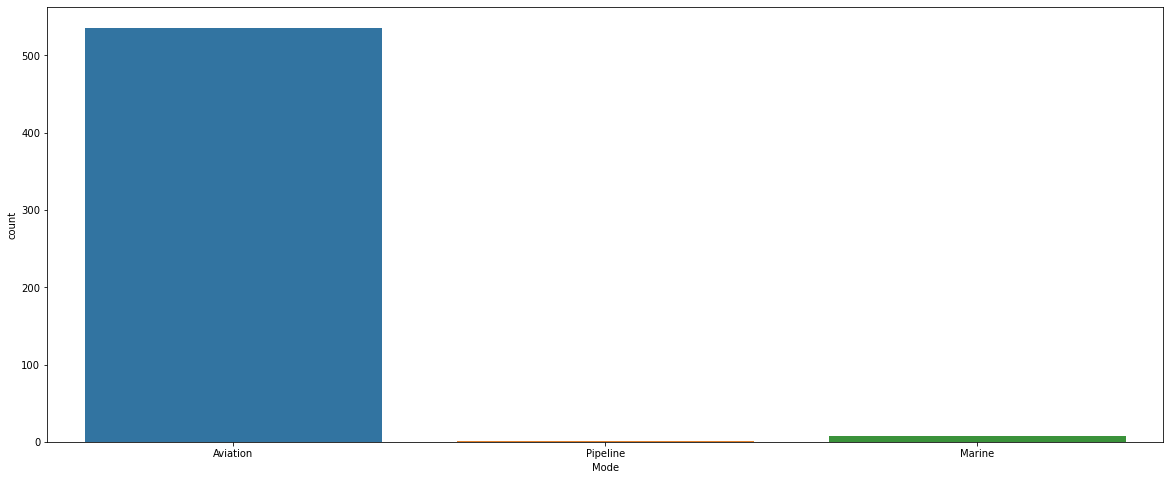
With the above techniques, was able to generate a model. With the data collected before the flight take off, was able to predict the potential risk associated with the journey. With the comments provided by different personnels, was able to do natural language process and categorize into issues if any and redirect to the concerned department automatically. Earlier, initial takeoff process would take around 3 hours of time for completion. With automation, it has been reduced to 2 hours. Also, in case of any flight accidents and first information report, root cause analysis could be determined and re-routed to concerned departments for futher actions. Before, this proposal, needed human interference to close the ticket.

With this prototype, human interaction for risk assessment has been reduced. Also, it has reduced the work hours of take off process completion.

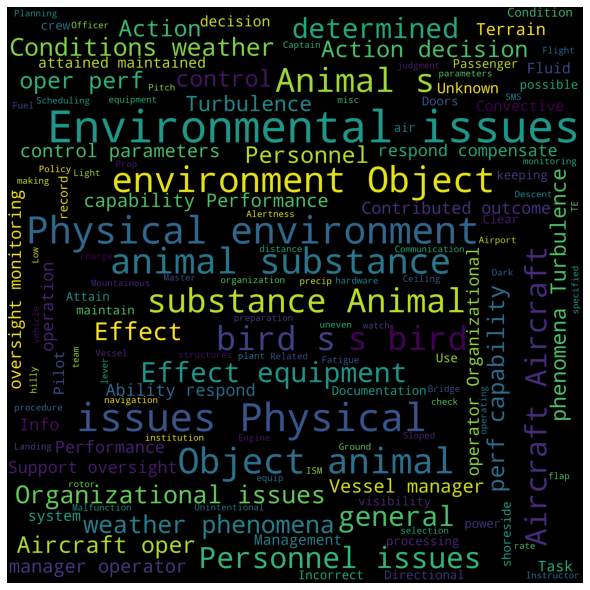
**Data Insights:**



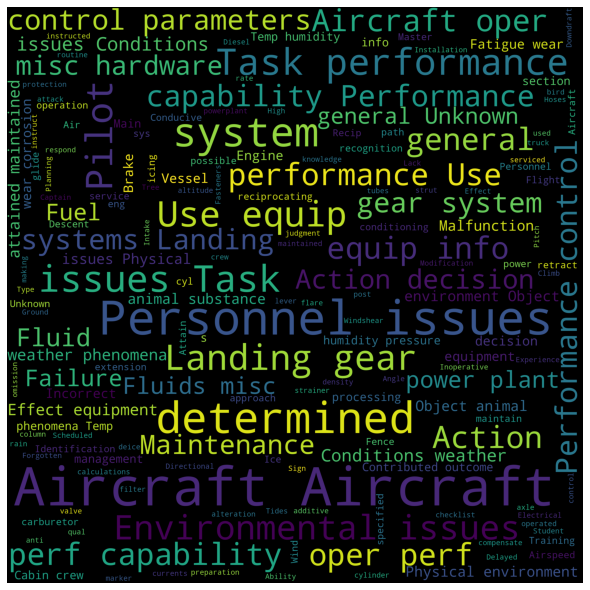
***Figure: Highest Injury Level in Flight Crashes***



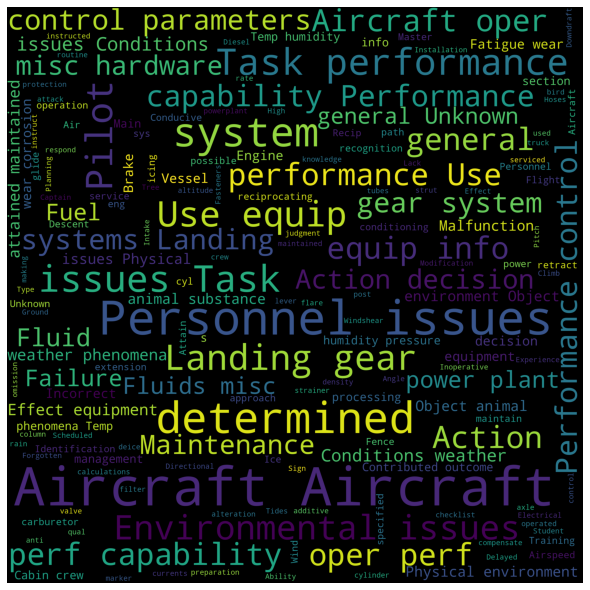
***Figure: Modes where Flight Crashes Occured***



***Figure: Word Cloud for environmental related accidents***



***Figure: Word Cloud for technical related accidents***



***Figure: Word Cloud for pilot related accidents***

**Major Takeaways:**

* Understood the difficulties in the data collection. How the data has to be approached like cleaning the intial data, grouping the data, understanding the data with intial analysis.
* Generating arount 10 models and comparing them with their residuals.
* With NLP, was able to understand the intial data preparation, tokenization technique, segregating the data based on the inputs provided
* Generated the front-end screen to collect the data from the end user and redirecting to the output screen.
* Model deployment with Flask, where I was able to successfully deploy a model as a pkl and with the new data provided by the end user, predict the risk accordingly.

It was indeed a memory refresher to apply all the data science skills covered in the course till now.

Thanks Great Learning !!!