**Introduction**

* Problem statement

The problem statement for airline safety trend analysis is to identify and analyze the trends in airline safety over time, in order to improve the safety of air travel and reduce the likelihood of accidents and incidents. This involves collecting and analyzing data on various safety metrics such as the number of accidents, incidents, fatalities, and injuries, as well as the causes and contributing factors of these events. The analysis should also consider factors such as changes in technology, regulations, and airline operations that may impact safety trends. The ultimate goal of the analysis is to identify areas where improvements can be made to enhance airline safety, and to develop strategies and interventions that can help reduce the risk of accidents and incidents in the future.

* Explain why the problem is important/interesting

The problem statement for airline safety trend analysis is important and interesting for several reasons:

1. Ensuring the safety of passengers and crew is of paramount importance for the airline industry, and understanding safety trends can help identify areas where improvements can be made to enhance safety and reduce the likelihood of accidents and incidents.
2. By analyzing safety trends, stakeholders in the airline industry, including airlines, regulatory bodies, and other organizations, can gain valuable insights into the factors that contribute to accidents and incidents, and develop targeted interventions to address these issues.
3. Understanding safety trends is also important for passengers, who want to feel confident that they are flying with a safe airline. By monitoring safety trends and holding airlines accountable for their safety records, passengers can make informed decisions about which airlines to fly with.
4. Safety trends in the airline industry can also have broader implications for public safety and the economy. Accidents and incidents can lead to loss of life, property damage, and other negative impacts, and understanding safety trends can help prevent these outcomes.

* Who would be interested in solving this problem, i.e., who would you be trying to sell this project to?

The project of analyzing airline safety trends would be of interest to a wide range of stakeholders, including airlines, regulatory bodies, safety organizations, investors, insurance companies, and the public.

* Where did you get your data?

Data related to Airline industry will be used from Kaggle and related to the price trend. Since the project is related Airline industry and analyzing travelling pattern to predict the future airline travel price trend, I’m attaching the datasets that I will use to build the models for the predictive analytics.  The project is to develop accurate demand forecasting model to control the availability in Airline industry.

* Why is this data useful to solve the problem?

**Methods/Results**

* What did you find out by exploring the data?
* Are there any visualizations that help tell a story with your data?

I’ve used various visualization to explore and understand the dataset. Below are few of them and the reason I used them for the project.

1. Time series plot: A time series plot can show the trend of airline accidents over time, and highlight any seasonality or trends in the data. It can be useful for identifying changes in the frequency and severity of accidents over time.
2. Bar chart: A bar chart can be used to show the frequency of accidents by different categories such as aircraft type, airline, or cause of the accident. This can be useful for identifying the most common types of accidents and for comparing accident rates across different categories.
3. Scatter plot: A scatter plot can be used to show the relationship between two variables, such as the age of the aircraft and the frequency of accidents. It can be useful for identifying patterns and relationships in the data and for identifying potential risk factors.
4. Heat map: A heat map can be used to visualize the frequency of accidents by time of day, day of the week, or other time periods. This can be useful for identifying patterns and trends in the data and for identifying times of day or week when accidents are more likely to occur.

Top of Form

Bottom of Form

* What steps did you perform to prepare the data?

1. Data cleaning: Once the data is collected, this had to be cleaned and preprocessed to remove any missing or irrelevant data, duplicates, and inconsistencies.
2. Feature selection: The next step is to select the relevant features that are important in predicting airline accidents. These features could include the type of aircraft, weather conditions, pilot experience, maintenance records, and other factors that could contribute to accidents.
3. Feature engineering: After feature selection, it is essential to engineer or transform the features to make them more informative and suitable for predictive modeling. This step may involve scaling, normalization, or encoding categorical variables.
4. Splitting the data: The data is then split into a training set and a test set. The training set is used to build the predictive model, while the test set is used to evaluate the model's performance.
5. Model building: Several predictive modeling techniques can be used to build a model to predict airline accidents, such as logistic regression, decision trees, random forests, and neural networks.

* What type of types of modeling are you using on your data?

**Forecast model**

Since Forecast models are one of the most prominent predictive model types. They predict future values based on historical data. In addition, these models manage metric value predictions by estimating the numeric value for new data based on learnings from historical data.

### Time series model

A time series model is used to predict future events based on past data ordered in a sequence. It is an econometric technique used to predict future values based on past values. A time series model uses the trends, seasonality and cyclicality of a system, as well as other factors to forecast future behavior. I’ve used SARIMAX time series forecasting model that can be used to predict airline accidents.

* What metric(s) are you using to measure your results?

R-squared: it indicates how many variables compared to the total variables the model predicted. R-squared does not take into consideration any biases that might be present in the data.

Average error: it is the numerical difference between the predicted value and the actual value.

Median error: the average of all difference between the predicted and the actual values.

Median absolute error: represents the average of the absolute differences between prediction and actual observation. All individual differences have equal weight, and big outliers can therefore affect the final evaluation of the model.

* Why did you choose the metric(s) you chose?

**Conclusion**

* What did you learn?

The dataset had to go through lot of cleaning process and still the dataset is not a static one. and because of that the model outcome for RMS is pretty high . I believe we need to take other factors into account while building the model and instead of the taking the actual numbers for the year, we should have more breakdown of the numbers based on the monthly or daily data where we could have calculated the moving average and standard deviation and that could have given a better result.

* What recommendations would you make based off your analysis?

Based on an analysis of airline accident data trends, here are some possible recommendations that could be made to improve aviation safety:

Analyze data: Continue to collect and analyze data on airline accidents to identify trends, patterns, and potential risk factors, and use this information to inform safety policies and practices.

Training and education: Provide adequate training and education to pilots, air traffic controllers, maintenance personnel, and other aviation professionals to ensure that they are up-to-date on the latest safety procedures and best practices.

* Is your model ready for deployment?

I believe the model cannot be deployed at the moment because of the following issues.

Data issue: There is not enough current data and the source cannot be relied upon that makes the model a work in progress.

Scalability: The model is tested only with limited dataset and need to be flexible enough to a scalable model.

Use different reliable source of dataset as well as a proper environment is required to deploy the model

* What do you need to consider ethically regarding the data, your model and the presentation of results?

Here are a few key ethical considerations that I kept while building the model.

1. Data privacy: Ensured that the data used in the analysis is anonymized and de-identified to protect the privacy of individuals involved in accidents.
2. Transparency and accountability: I’m transparent about the data sources, methodology, and assumptions used in the analysis and provide clear explanations of the model's predictions and limitations. This can help ensure that stakeholders understand the results and can hold the analyst accountable for the accuracy and ethical use of the data.

**References**

Dataset: <https://www.kaggle.com/code/ruslankl/airplane-crashes-data-visualization/data>

<https://www.kaggle.com/datasets/saurograndi/airplane-crashes-since-1908>

Ref: <https://analyticsindiamag.com/complete-guide-to-sarimax-in-python-for-time-series-modeling/>

<https://www.statsmodels.org/stable/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html>

<https://www.jadsmkbdatalab.nl/forecasting-with-sarimax-models/>