INTEGRATED VISUALIZATION AND PREDICTIVE ANALYSIS OF FLIGHT SCHEDULE DISRUPTION

Megha Hole, Pujan Rijal

Abstract—The aviation sector must manage flight schedules effectively to maintain timely and safe operations. The performance of airports and airlines can be greatly impacted by flight interruptions, such as cancellations and delays, which can result in lost revenue and disgruntled passengers. Technology that can anticipate flight delays in advance and illustrate the specifics of the flight schedule to help with decision-making is needed to solve this problem. Predictive analytics and visualizations are combined in the suggested method to give a complete picture of disruptions to flight schedules. To forecast the risk of flight disruptions, the predictive analytics model will analyze historical flight data, weather data, and other pertinent elements. The output of the prediction algorithm will be displayed through visualizations that offer details about the disruptions and timetables of flights. Any possible disruptions will also be highlighted by the visualizations, along with suggestions for how to lessen their effects. Using an integrated strategy, airlines and airports will be able to control flight disruptions and lessen the effects they have on business operations. The suggested method will enhance customer satisfaction while increasing the effectiveness and dependability of flight scheduling. Airlines and airports will be able to respond promptly and efficiently to any possible interruptions since it will give a data-driven approach to managing flight disruptions.

Index Terms— Visualization, Flight Delay, Flight Cancellation, Diverted Flights, Weather delay

INTRODUCTION

The dataset used for this project was extracted from Kaggle and includes four years' worth of US aircraft schedule data. The goal is to use this data to visualize the frequency of flight arrivals and departures and to build a dashboard that makes it simple to identify flight delays and cancellations based on airport IDs and location. The dataset's month-by-month data will allow for the investigation of flight cancellations or delays during particular months. The dataset also contains airline arrival and departure times for each US airport. The causes of flight delays and cancellations will be determined using a weather dataset.

We have prepared the dashboard as shown below comprising the visualizations via Tableau. We have created six visualizations using the data from these datasets. The first visualization examines the arrival and departure delay per airline. The second visualization indicates the cancellation count per airline. The third visualization looks at the arrival delays in the month of January. The fourth visualization concentrates on most diverted flights, while the fifth visualization helps determine the number of cancellations per day. Similarly, the sixth visualization helps in identifying the past 10 years count of delay due to weather conditions.

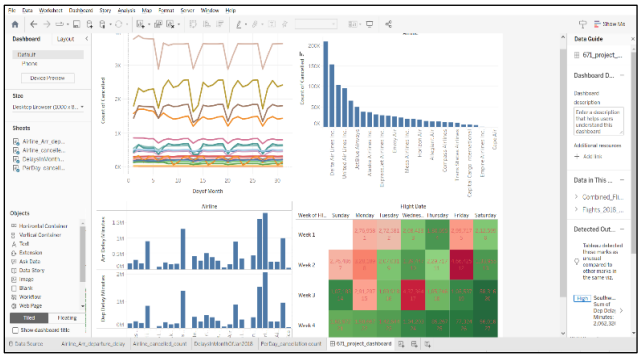

Fig 1. Dashboard

.

1 RELATED WORK

Several studies have formerly been carried out by the domain experts to analyse the flight status. One research paper aimed at predicting flight delays using machine learning algorithms which could have a significant impact on the economy, see Flight Delay Prediction: Vishrut Raj, Viran Raj, Satyam Singh, Adityanath Mishra, Rajkumar Goel Institute of Technology/AKTU, June 2021| IJIRT | Volume 8 Issue 1 | ISSN: 2349-6002. Another paper demonstrated different approaches used to build the flight prediction models. It primarily emphasized on the complexity of air transportation system, the importance of accurate prediction alongside the increasing use of machine learning methods, see L. Carvalho, A. Sternberg, L. Maia Gonc¸alves, A. Beatriz Cruz, J.A. Soares, D. Brandaˆo, D. Carvalho, e E. Ogasawara, 2020, On the relevance of data science for flight delay research: a systematic review, Transport Reviews. A different research group proposed a Deep Learning Model for predicting flight delays by extracting important features from data using a stack denoising autoencoder, applying Levenberg-algorithm to find weight and bias values, and optimizing the output for high accuracy, see Flight delay prediction based on deep learning and Levenberg-Marquart algorithm: Maryam Farshchian Yazdi, Seyed Reza Kamel, Seyyed Javad Mahdavi Chabok, and Maryam Kheirabadi).

2 TARGET AUDIENCE FOR THE STUDY

This study's target audience is not just decision-makers and industry professionals in the aviation sector. Additionally, it comprises tourists and companies who are interested in comprehending and keeping track of changes to flight schedules.

The research findings and dashboards that were created can offer travelers vital information on airline cancellations and delays, allowing them to better plan their travels and prepare in case of delays. By having a better grasp of flight schedules and delays, businesses that depend on air travel can also profit from the research. This can help manage logistics and save operational costs.

In addition, this study may be useful to academics and researchers working in the domains of data analytics, aviation, and transportation. This is because the research's methodology and findings may be used as a helpful resource for future work in these fields.

3 Visualizations

We have prepared the following visualizations:

3.1 Arrival and Departure Delay Per Airline

To facilitate this study, we have considered the attributes relating to the airline and the arrival and departure delay in minutes. The visualization shows that Southwest Airlines had the most arrival and departure delays in 2018.

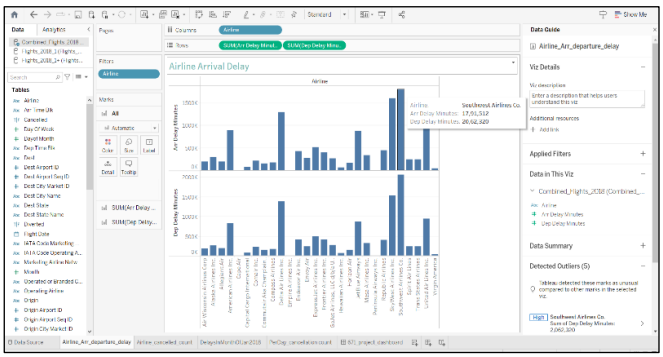

Fig 2 - Arrival and Departure Delay Per Airline

To discover which airlines, experience a higher rate of arrival delays and implement corrective actions, decision-makers in the aviation business can use this graphic, which gives information on the arrival delay time for each airline. This visualization can also be used by businesses and tourists to steer clear of airlines with a history of delayed flights.

3.2 Cancellation Count Per Airline

This section of the report is based on the visualization of the cancelled flights per airline in 2018. According to the visualization, Southwest Airlines had the most cancellations i.e 209,750.

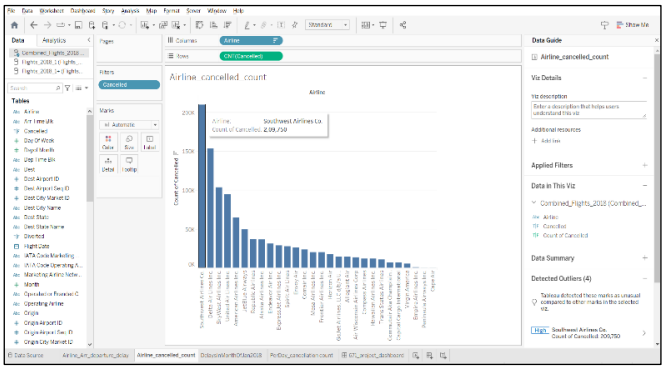

Fig 3 - Cancellation Count Per Airline er Airline

Decision-makers in the aviation business can use this visualization to identify which airlines have a higher rate of cancellations and take corrective action by using the data on the number of cancellations by each airline provided. Avoiding airlines with a higher cancellation risk might be beneficial for both travelers and businesses.

3.3 Arrival Delays in the Month of January

In this section, we visualized the arrival delays in the month of January 2018 by taking into consideration the attributes relating to Flight Date and the arrival delay in minutes. The visualization depicts that the most arrival delays occurred on the Friday of Week 2 i.e., 466,425.

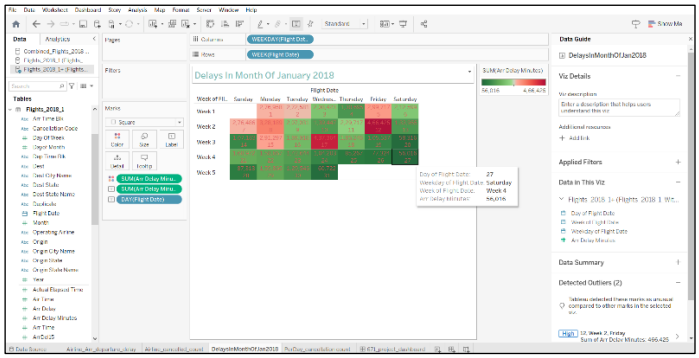

Fig 4 - Arrival Delays in the Month of January

This visualization aids in determining whether any certain months or seasons are more likely to have arrival delays than others. This data can be used by decision-makers in the aviation sector to create informed plans and implement preventative steps to lessen arrival delays during particular months. Using this depiction, travelers and organizations can avoid booking flights during periods when arrival delays are more frequent.

3.4 Percentage of Diverted Flights

We analyzed the airlines w.r.t the percentage of diverted flights in 2018. Based on analysis, we noticed that Peninsula Airways has the most % of diverted flights in 2018.

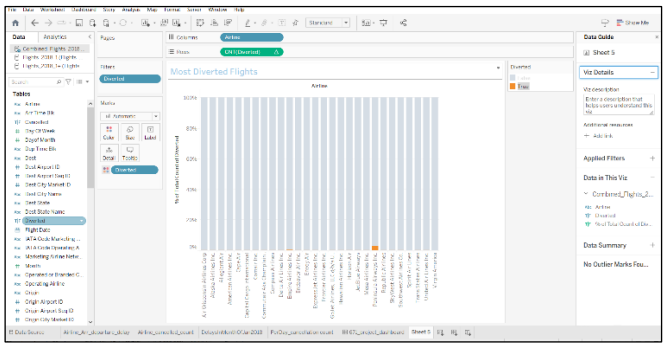

Fig 5 - Percentage of Diverted Flights

To help decision-makers in the aviation sector discover possible areas for development and make data-driven decisions to lessen the number of diverted flights, this graphic provides information on the

airports with the largest number of diverted flights. Avoiding airports with a high likelihood of flight delays can also be beneficial for travelers and businesses.

3.5 Number of Cancellations Per Day

The attributes we chose for this study were the count of cancelled flights and the Days of Month for January 2018. As per the visualization, we noticed that Southwest Airlines had the highest number of cancellations per day in the month of January.

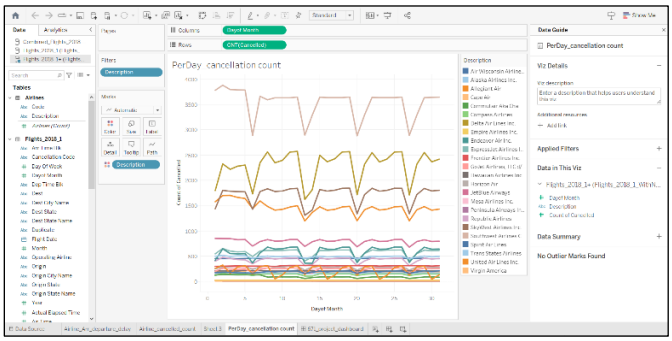

Fig 5 - Number of Cancellations Per Day Flights

This visualization makes it easier to pinpoint the weeks or months of the year that see the most flight cancellations. It can help decision-makers in the aviation sector develop informed plans and implement preventative steps to lessen the frequency of canceled flights. Travelers and organizations can also benefit from using this visualization to avoid booking flights on days when there is a high probability of cancellation.

3.6 Past 10 Years Count of Delay due to Weather Conditions

For the purpose of this visualization, we combined 10 years of data files with the fields 'Year' and 'Weather delay'. As per the visualization, we noticed that in 2013, the count of flights delayed due to weather was 16,069. We discovered the highest count of delay due to weather to be 20,932. And similarly, in the year 2022 it was 20,338.

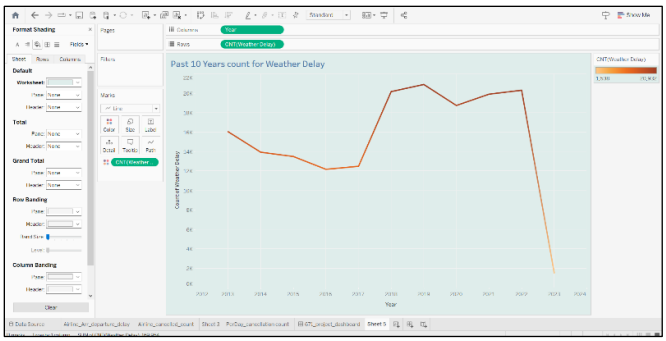

Fig 6 - Number Past 10 Years Count of Delay due to Weather Conditions Flights

This visualization is important for the study since it offers data on the frequency and severity of weather delays over the previous ten years. Decision-makers in the aviation sector can use this image to better understand weather patterns and how they affect flight schedules. The industry may enhance its response to weather-related delays and lessen their influence on flight schedules by examining this data.

Due to its insights into how weather-related delays effect flight schedules, this graphic can be helpful for both tourists and enterprises. Making informed choices about travel arrangements can help, such as avoiding specific airports or airlines during periods of intense weather disruption.

4 Discussion

There are some restrictions and observations that can be made about this study, including:

Data Restrictions: The usefulness and quality of the datasets utilized in this investigation are critically important. The correctness and completeness of the datasets can have an impact on the visualizations' accuracy and the conclusions drawn from them. The outcomes could potentially be impacted by missing data or inconsistencies.

Weather Information: To determine the causes of aircraft delays or cancellations, the study makes use of an extra weather dataset. Although meteorological data can be complicated and unpredictable, there may also be other variables, such as technical difficulties, air traffic jams, or security concerns, that cause delays and cancellations.

Generalizability: The study only considers domestic flight schedules, therefore it may not be generalizable to other nations or areas. Weather patterns, airline policies, and infrastructure are just a few of the variables that might change depending on where you are in the world.

Interpretation: This study's visualizations were produced using data analysis, but the interpretation of the findings might be arbitrary. Depending on their background, experience, and context, various audiences may perceive the visuals in different ways.

Causality: The visualizations and conclusions drawn from them may show correlation, but not always cause and effect. The visualizations can show trends and patterns, but it's crucial to do more research to understand their underlying reasons.

Overall, this study offers insightful information about aircraft schedules and delays, however there are some considerations for limits. The study can act as a springboard for additional investigation and analysis to boost the accuracy and dependability of airline schedules.

5 Conclusion

In summary, the goal of this study was to visualize airline schedule data and pinpoint patterns associated with flight delays and cancellations. This study was able to produce visuals that emphasize airline-specific arrival delays and cancellation counts, trends linked to arrival delays in January, the most diverted aircraft and so on by utilizing datasets from Kaggle and weather sources.

Decision-makers in the aviation sector can use these visualizations as a useful tool to pinpoint areas for development and make data-driven choices that will increase the effectiveness and dependability of flight schedules. These visualizations can also help travelers and businesses by improving their comprehension of flight delays and helping them to make the required preparations.

Overall, this work establishes a basis for future research and development in this area and highlights the benefits of combining data analytics and visualization approaches to get insights on flight schedules and disruptions.

Acknowledgments

The authors wish to thank Professor Jonathan Leidig for his resourcefulness, encouragement and cooperation that aided in the successful completion of the project.

References

[1] Flight Delay Prediction: Vishrut Raj, Viran Raj, Satyam Singh, Adityanath Mishra, Rajkumar Goel Institute of Technology/AKTU, June 2021| IJIRT | Volume 8 Issue 1 | ISSN: 2349-6002

[2] L. Carvalho, A. Sternberg, L. Maia Goncalves, A. Beatriz Cruz, J.A. Soares, D. Branda˜o, D. Car- valho, e E. Ogasawara, 2020, On the relevance of data science for flight delay research: a systematic review, Transport Reviews

[3] Flight delay prediction based on deep learning and Levenberg-Marquart algorithm: Maryam Farshchian Yazdi, Seyed Reza Kamel, Seyyed Javad Mahdavi Chabok, and Maryam Kheirabadi).