FLIGHT DELAY ANALYSIS

A MINI PROJECT REPORT

Submitted by

DWARAKNATH B (1920110009)

INDHRA V (1920110015)

in partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

SONA COLLEGE OF TECHNOLOGY, SALEM-5

(Autonomous)

ANNA UNIVERSITY: CHENNAI 600 025

NOVEMBER 2023

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report "**FLIGHT DELAY ANALYSIS"** is the bonafide work of **"DWARAKNATH B (1920110009), INDHRA V (1920110015)"** who carried out the project work under my supervision.

SIGNATURE	**SIGNATURE**
Dr. J. Akilandeswari	Mr. J L Aldo Stalin
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Professor	Assistant Professor
Department Of Information Technology	Department Of Information Technology
Sona College of Technology	Sona College of Technology
Salem- 636 005.	Salem- 636 005.

Submitted for Project viva voce examination held on
16-11-2023 to 18-11-2023

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

First and foremost, we thank to **power of almighty** for showing us inner peace and for all blessings. Special gratitude to our parents, for showing their support and love always.

We express our sincere thanks to chairman **Sri.C.VALLIAPPA** and principal **Dr.S.R.R.SENTHILKUMAR** for providing adequate facilities to complete the project.

We are immensely grateful to our Head of Information Technology Department, **Dr.J.AKILANDESWARI for** the continue encouragement to complete the project.

We express our heartfelt thanks to our project supervisor **J L ALDO STALIN** for his valuable guidance and fruitful discussions throughout the course of the project work.

We feel proud in sharing this success with all our department Faculty, Staff members and friends who helped directly and indirectly to completing this project successfully.

ABSTRACT

The punctuality of airline operations, often gauged through On-Time Performance (OTP), is a pivotal concern in the aviation industry. OTP is typically defined as the percentage of flights experiencing delays exceeding 14 minutes upon gate arrival. These delays can be attributed to various intricate factors, such as Turn Delay (on-ground), Block Delay (in-air), and Previous Delay carried over from prior flight legs. This project focuses on an in-depth Exploratory Data Analysis (EDA) approach, which involves dissecting flight arrival delays into these three distinct components. Rather than analyzing the overall OTP or Arrival Delay, this approach promises to provide a more detailed and insightful perspective on the factors influencing airline punctuality. The dataset used for this analysis is drawn from the Department of Transport and covers flight-level OTP data for multiple airlines. Through this specialized EDA, we aim to unravel the complex web of influences on flight delays. This project serves to benefit both airlines and passengers. Airlines can gain valuable insights into the operational aspects that impact their OTP, allowing them to enhance service quality and efficiency. For passengers, a better understanding of the factors contributing to delays can lead to more informed travel decisions. In this abstract, we provide an overview of the approach and methodology that will be applied in the subsequent sections, demonstrating the potential of this analysis to improve the overall airline travel experience.

CHAPTER NO.	TITLE	PAGE NO.

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS	FULL FORMS
TD	Turn delay
BD	Block delay
OTP	On time performance
EDA	Exploratory data analysis

CHAPTER 1

INTRODUCTION

In the realm of airline operations, the pivotal metric of On-Time Performance (OTP) serves as a linchpin for a seamless and satisfactory travel experience. OTP is conventionally gauged by the percentage of flights experiencing delays beyond 14 minutes upon gate arrival. This project, recognizing the complexity of flight delays, takes a nuanced approach by delving into the depths of Exploratory Data Analysis (EDA) to unravel the intricate factors influencing OTP.

Rather than a conventional examination of overall OTP or Arrival Delay, the project meticulously dissects flight arrival delay into three discernible components: Turn Delay, Block Delay, and Previous Delay. Turn Delay, encompassing on-ground factors like aircraft turnaround times and gate availability, Block Delay involving delays during taxi-out and taxi-in phases, and Previous Delay, a cumulative effect from prior flight legs, constitute the trifecta under scrutiny. By isolating these components, the project endeavors to reveal the distinct contributors to delays, paving the way for targeted interventions and strategic enhancements.

In essence, this specialized EDA approach is poised to transcend the conventional assessment of OTP, offering stakeholders a granular perspective on the root causes of delays. The ultimate goal is to empower the aviation industry with actionable insights, enabling a proactive stance in addressing specific delay components and, in turn, elevating the overall punctuality and passenger experience in air travel.

1.1 OBJECTIVES

The primary objectives of this project revolve around conducting a comprehensive exploratory data analysis (EDA) on airline On-Time Performance (OTP) by dissecting flight arrival delay into three distinct components: Turn Delay, Block Delay, and Previous Delay. First and foremost, we aim to gain a profound understanding of the intricate factors that contribute to flight delays.

Through a systematic analysis, we seek to identify which of these three components exerts the most significant influence on overall delays, thus providing airlines with invaluable insights to enhance their operational efficiency. Additionally, we will compare the performance of different airlines in terms of these delay components, shedding light on potential best practices. The impact of these components on different flight routes, airports, and passenger experiences will also be assessed. Predictive modelling will be developed to forecast delays, helping both airlines and passengers make more informed decisions. Through data visualization and clear communication of results, the project aims to improve overall transparency and understanding of flight delays, benefitting airlines, regulatory authorities, and travellers alike.

1.2 CHARACTERISTICS

The project's characteristics encompass a data-driven approach, utilizing extensive airline OTP data to delve into the multifaceted components of flight delay, specifically Turn Delay, Block Delay, and Previous Delay. It acknowledges the complexity of airline operations and the interconnected variables affecting OTP. The analysis considers temporal variations, including daily, monthly, and seasonal trends, and undertakes comparative assessments to highlight performance disparities among airlines. Additionally, it incorporates geospatial analysis to discern how flight routes and airports are influenced by dissected delay components. A passenger-centric perspective is maintained, seeking to understand the impacts on travelers, such as missed connections and overall satisfaction. Predictive modeling is applied to anticipate delays and facilitate proactive management of airline operations. Data visualization aids in presenting findings effectively, while actionable insights are aimed at assisting airlines in enhancing OTP. The project fosters improved communication among airlines, regulatory bodies, and passengers, promoting continuous improvements in airline operations and the overall travel experience.

1.3 LITERATURE SURVEY

Paper 1: Anderson Ong, M. (2017) “Data Visualization of Flight Delays with Tableau”

Focus: This source concentrates on leveraging Tableau for data visualization of flight delays.

Key Insights: Anderson Ong's work likely explores the effectiveness of Tableau in translating complex flight delay data into visually comprehensible insights. The article could offer practical examples and best practices for utilizing Tableau in the aviation context.

Paper 2 : Dey, T., Phillips, D., and Steele, P. (2011). "A Graphical Tool to Visualize Predicted Minimum Delay Flights"

Focus: The study introduces a graphical tool tailored for visualizing predicted minimum delay flights.

Key Insights: This research is expected to present a novel graphical tool designed to enhance the visualization of predicted minimum delay flights. Insights might include the development process, the tool's application, and its potential impact on decision-making in optimizing flight schedules.

Paper 3: Gopalakrishnan K., Balakrishnan H. (2017). "A Comparative Analysis of Models for Predicting Delays in Air Traffic Networks."

Focus: This work conducts a comparative analysis of models predicting delays in air traffic networks.

Key Insights: The study likely provides a thorough examination of various models used for predicting delays in air traffic networks. It may highlight the strengths and weaknesses of different approaches, aiding researchers and practitioners in selecting effective predictive models.

Paper 4: Gupta P., Dwivedi A., Agrawal A. (2016). "An Analysis of US Domestic Flight delays using SAS Enterprise Miner"

Focus: The research involves an in-depth analysis of US domestic flight delays using SAS Enterprise Miner.

Key Insights: This work is anticipated to offer insights into the application of SAS Enterprise Miner for analysing delays in US domestic flights. Findings may include patterns, contributing factors, and potential strategies for improving punctuality.

Paper 5: Mainero A., Schmidt T., and Sugarman H. (2013). "Heat Mapping and Predicting Flight Delays and Their Propagations in a Real-World Air Traffic Simulation."

Focus: The study explores heat mapping and prediction of flight delays in a realistic air traffic simulation.

Key Insights: This source is likely to provide insights into the use of heat mapping techniques to visualize and predict the propagation of flight delays within a real-world air traffic simulation. Applications and implications for managing delays in complex air traffic scenarios may be discussed.

Paper 6: Michael T. Crotty. (2014). "Visualizing More Than Twenty Years of Flight Data for the Raleigh-Durham International Airport."

Focus: The study concentrates on visualizing extensive flight data for a specific airport.

Key Insights: This work could shed light on the challenges and opportunities associated with visualizing long-term flight data for a particular airport. It might discuss the evolution of flight patterns, seasonal variations, and the impact of external factors on airport operations.

Paper 7: Rodríguez-Sanz, Á., Gómez Comendador, F., Arnaldo Valdés, R., Cordero García, J., and Bagamanova, M. (2018).

Focus: The provided details are insufficient to determine the focus of this source.

Potential Insights: Given the collaboration of various authors, the source may cover a range of topics related to flight delays. Exploring the actual content of the publication is essential for a detailed understanding.

1.4 SCOPE

The scope of the project is expansive, delving into the intricate realm of dissecting flight arrival delay into three distinct components for a thorough analysis of airline On-Time Performance (OTP). This multifaceted endeavor aspires to not only improve operational efficiency but also enhance the overall passenger experience within the aviation industry.

The project's initiation revolves around a fundamental emphasis on data quality. Recognizing that the foundation of meaningful analysis rests on high-quality and relevant data, the project meticulously collects and prepares extensive flight delay data. This initial step underscores the importance of data accuracy and completeness, setting the stage for reliable and insightful findings. The ambitious scope then progresses into an in-depth Exploratory Data Analysis (EDA) phase, unraveling complexities within the dissected delay components. This phase explores temporal trends, conducts comparative analyses, and integrates a passenger-centric perspective into the analysis.

Data visualization, a pivotal aspect of the project's scope, employs powerful tools like Tableau to transform complex data into clear and informative representations. These visualizations serve not only to facilitate data-driven decision-making for stakeholders but also to make the intricate findings accessible to a broader audience. The predictive modeling component introduces sophistication by developing accurate models that leverage historical data to forecast potential delays. This forward-thinking approach empowers airlines to proactively manage operations and minimize disruptions, aligning with the broader goal of improving OTP.

Continuous monitoring and reporting mechanisms are incorporated into the project's scope, offering real-time insights into OTP and delay trends. This approach instills a culture of continual improvement in airline operations, ensuring stakeholders remain agile and responsive to changes. Geographic and temporal analyses add strategic depth, exploring variations in delay components across different regions and seasons, providing critical insights for tailored improvements.

CHAPTER 2

EXPLORATORY DATA ANALYSIS

2.1 SYSTEM ARCHITECTURE

2.1.1 ARCHITECTURE DESIGN

The architecture design of the flight delay factors diagram provides a clear and visually appealing representation of the complex web of elements that can influence flight delays. The diagram is structured into three main sections: Airport, Flight, and Other, with each section containing specific factors that contribute to the understanding of flight delays. In the Airport section, factors such as Weather, Airport traffic flow, Leave port speed, and Arrive port speed are highlighted. Weather plays a pivotal role in delays, with adverse conditions like thunderstorms, snow, or fog disrupting takeoffs and landings. Airport traffic flow can cause delays as congested airspace leads to aircraft queuing for takeoff and landing. Additionally, slow leave and arrive port speeds can prolong the time it takes for aircraft to taxi or reach the gate, contributing to delays.

The Flight section focuses on factors such as Date, Flight number, Scheduled time, Pre-flight issues, and Historical flight delay. Dates have a significant impact on delays, as holidays and peak travel periods lead to increased flight demand. Specific flight numbers may be more susceptible to delays due to factors like aircraft type, route complexity, or scheduling issues. Scheduled times, particularly early morning and late-night flights, are more prone to delays. Pre-flight problems, whether mechanical or crew-related, can also contribute to delays. Historical flight delay data highlights that flights with a previous history of delays are more likely to experience delays in the future.The Other section considers factors like Season, Holiday, and unexpected events. Seasonal variations, especially winter weather conditions, can lead to more delays compared to the summer. Holidays introduce elevated air traffic and travel volumes, further straining the aviation system. Unexpected events, including bird strikes and air traffic control issues, can disrupt flight schedules.

The use of arrows to connect these factors visually illustrates the interrelationships and how one factor can influence another, creating a comprehensive view of the causes of flight delays. In summary, this architecture design enhances the understanding of flight delays by categorizing and connecting the contributing factors. Passengers, airlines, and airport authorities can use this diagram to prepare for and manage travel plans more effectively, ultimately improving the overall travel experience.

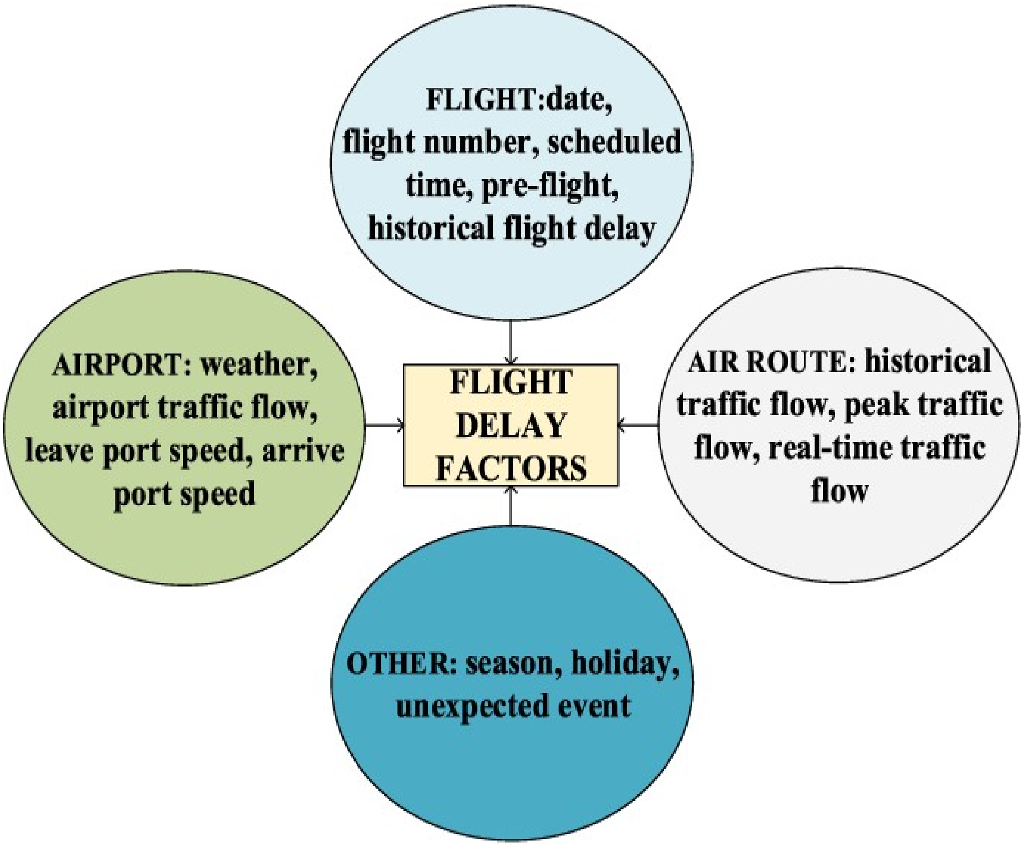

Fig 1 Block Diagram

2.1.2 PROPOSED SYSTEM

The proposed system for addressing flight delays is a multifaceted and data-driven approach designed to significantly improve On-Time Performance (OTP) and enhance the overall travel experience for passengers. Drawing inspiration from insights gained from the flight delay factors diagram, this system leverages advanced analytics and predictive modelling to comprehensively predict, monitor, and mitigate delays. At the heart of this system is the establishment of an integrated data analytics platform.

This platform serves as a hub for collecting, processing, and visualizing real-time data from a variety of sources. These sources include meteorological services, air traffic control, and historical flight records, which provide a wealth of information necessary to make informed decisions regarding delays. Predictive modelling plays a crucial role in this platform, employing sophisticated algorithms to forecast potential delays based on the complex interplay of factors identified in the flight delay factors diagram.

One key feature of this system is the implementation of an early warning system. This system has the capability to promptly alert airlines, airports, and regulatory bodies to potential delay triggers. This early warning system allows for proactive measures to be taken, such as adjusting flight schedules, re-routing aircraft, or optimizing ground operations, in order to minimize disruptions. Another vital aspect of this system is the introduction of a collaborative decision-making framework. This framework encourages various stakeholders, including airlines, airports, air traffic control, and regulatory agencies, to communicate and share real-time data. Such collaboration fosters coordinated efforts to address delays more effectively and efficiently, as all parties can work together with a common understanding of the situation.

Resource allocation optimization is another integral part of the system. Predictive insights derived from the data analytics platform are used to streamline operations, such as gate assignments, ground crew scheduling, and flight planning. By optimizing resource allocation, airlines can make better use of available resources and reduce bottlenecks that contribute to delays. Lastly, improved customer communication is a priority. Passengers are kept well-informed with real-time updates on delays, their causes, and expected resolution times. This transparency enhances the overall travel experience, reducing frustration and stress for passengers.In conclusion, the proposed system represents a holistic and technology-driven approach to addressing flight delays. By leveraging data, predictive modelling, collaboration, and optimization, it seeks to make air travel more reliable, efficient, and passenger-friendly. Ultimately, the goal is to enhance On-Time Performance and the overall satisfaction of air travellers.

2.2 DATA COLLECTION

Data collection for flight delay analysis is a critical and systematic process. It begins with identifying reliable sources, which can range from aviation agencies to airlines and publicly available datasets. Key data variables, such as flight date and time, airline carrier, airport codes, delay times, and reasons for delays, must be determined. Access to these sources may require permissions or subscriptions, particularly for proprietary or government aviation records. Data cleaning is essential, involving the removal of duplicates, error correction, and handling missing values. Data is often aggregated to a suitable granularity (e.g., daily or hourly summaries) and formatted into a usable structure like CSV or a database. Standardizing units and formats ensures consistency across diverse data sources, with ongoing data quality checks to maintain accuracy.

Data security is crucial, especially when sensitive information is involved, and adherence to data privacy regulations is fundamental. Comprehensive documentation of data sources, collection methods, and transformations is maintained for transparency and reproducibility throughout the analysis process. This meticulous approach to data collection ensures that the subsequent flight delay analysis is based on accurate, reliable, and secure information.

2.3 DATA PREPROCESSING

Data preprocessing is a fundamental and intricate phase within the broader data analysis process, playing a vital role in refining raw data to render it suitable for in-depth analysis and modeling. This multifaceted procedure encompasses a series of indispensable steps, each of which is pivotal in ensuring that the data is both accurate and standardized, thus enabling its compatibility with the chosen analytical techniques. The initial step involves data cleaning, a foundational practice that seeks to rectify issues such as missing data, duplicate entries, and inconsistencies, all of which could potentially jeopardize the integrity of the dataset. Subsequently, data encoding techniques are employed to transform categorical variables into numerical formats, thereby making them amenable to inclusion in analytical models.

The next key phase involves scaling and standardization, where the goal is to bring numerical features onto a common scale. This practice mitigates the risk of any one feature disproportionately influencing the results and ensures that all features contribute uniformly to the analysis. Feature selection is yet another crucial aspect of data preprocessing, as it involves cherry-picking pertinent features while discarding irrelevant ones. This action not only accelerates the analytical process but also enhances model efficiency and interpretability. The handling of outliers, data points that significantly deviate from the norm and can distort the results, is a vital part of this process. Outliers are either removed or transformed, as required, to prevent unwarranted distortions.

Here's the formula for standardization:

$$X' = \frac{X - \mu}{\sigma}$$

μ is the mean of the feature values and σ is the standard deviation of the feature values. Note that, in this case, the values are not restricted to a particular range.

Imputation techniques are brought into play when addressing missing data, as these methods replace the absent values with suitable estimates, often relying on statistical measures or machine learning algorithms. Data transformation is a significant component for correcting feature distributions. Mathematical functions, such as logarithmic transformations, can be applied to mitigate skewness or other adjustments made to establish a more normalized distribution, aligning with the assumptions of the chosen analytical methods. In scenarios involving imbalanced data, particularly in classification problems, techniques such as oversampling (increasing the representation of the minority class) or under-sampling (reducing the majority class) are employed to achieve balance, ensuring that the model doesn't exhibit bias toward the majority class while offering sufficient representation for the minority class.

Formula

$$\tilde{\mu}_3 = \frac{\sum_i^N \left(X_i - \bar{X}\right)^3}{(N-1) * \sigma^3}$$

$\tilde{\mu}_3$ = skewness

N = number of variables in the distribution

X_i = random variable

$\bar{X}$ = mean of the distribution

σ = standard deviation

Feature engineering, an inherently creative aspect of data preprocessing, contributes significantly to improved model performance. This practice involves generating new informative features or extracting valuable insights from existing ones, thereby enriching the dataset. Finally, data splitting is undertaken to segregate the dataset into training, validation, and testing sets, a fundamental practice for model development and evaluation. The training set is dedicated to model training, the validation set aids in hyperparameter optimization and model performance assessment, and the testing set ensures the model's capacity to generalize effectively to unseen data.

In summation, data preprocessing is an indispensable phase within the data analysis pipeline. It serves as the bedrock for conducting robust and reliable analyses, ensuring that the data used is cleansed of errors, standardized, and appropriately tailored for the chosen analytical techniques. By addressing issues such as missing data, outliers, and imbalances, optimizing feature selection and engineering, and preparing the data for modeling, data preprocessing establishes the foundation for sound decision-making and insightful discoveries.

2.4 DATA VISUALISATION

Leveraging Tableau for flight delay analysis is a strategic decision that starts with importing meticulously cleaned flight delay data into the platform. Tableau's intuitive interface makes it an excellent choice for creating a wide range of visualizations that help uncover patterns and insights related to flight delays. From basic bar charts to more sophisticated line graphs, Tableau enables users to depict delay patterns over time and by airline with ease.

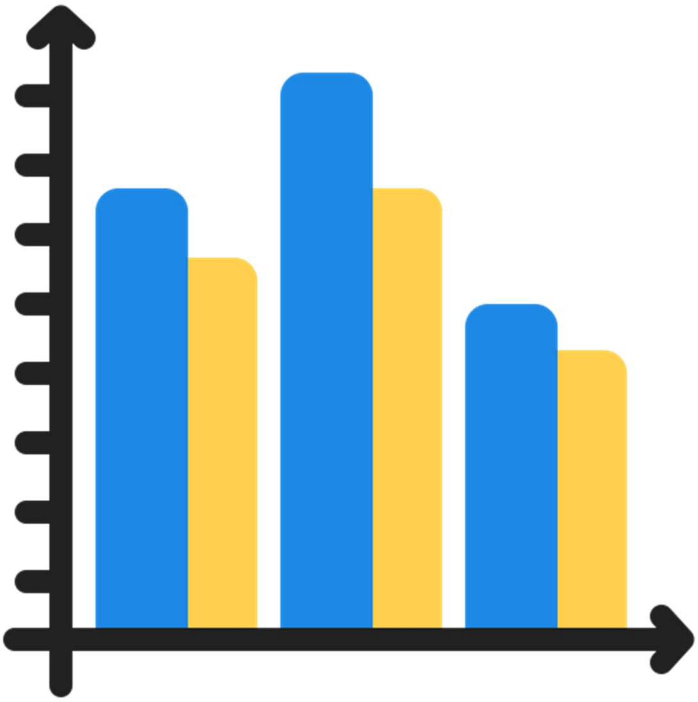

Fig 2 Bar chart of the analysis

Geographic mapping is another powerful feature of Tableau, which can be used to pinpoint regional disparities in flight delays. By visualizing delay data on a map, stakeholders can quickly identify geographic areas where delays are most prevalent. This information can be invaluable for airlines in understanding the geographical factors contributing to delays and can inform decisions related to scheduling and operations.

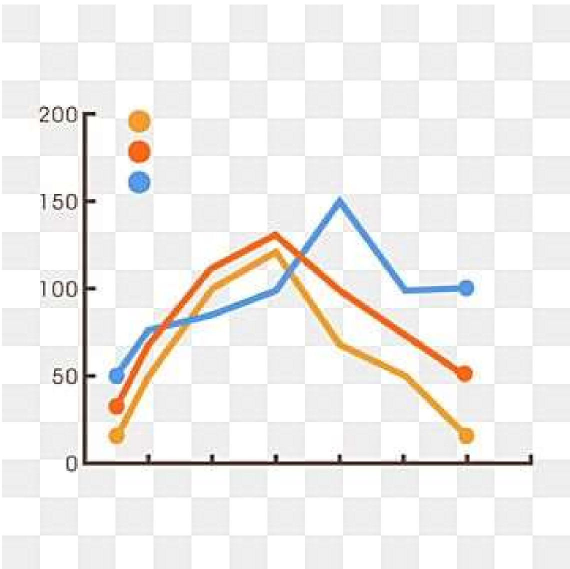

Fig 3 Line graph of the analysis

Dashboards in Tableau are the central hub for consolidating multiple visualizations, providing a comprehensive overview of delay factors. These dashboards can include bar charts, line graphs, heat maps, and maps, among other visual elements. They offer a holistic view of the delay data, allowing stakeholders to understand the various aspects of flight delays in one place.

Fig 4 Geographical map of the analysis

Adding trend lines to visualizations is crucial for temporal trend analysis. It helps in identifying long-term patterns and forecasting future delays. Additionally, Tableau's drill-down interactivity allows users to dive deeper into the data, exploring specific aspects or segments to uncover underlying causes and potential solutions. This feature is essential for in-depth data exploration and making informed decisions.

The storytelling feature in Tableau is a valuable tool for guiding viewers through the analysis narrative. It allows for the creation of a compelling and structured narrative that explains the key findings, insights, and recommendations derived from the flight delay analysis. Storytelling makes it easier for non-technical stakeholders to understand the data and its implications, improving communication and decision-making.

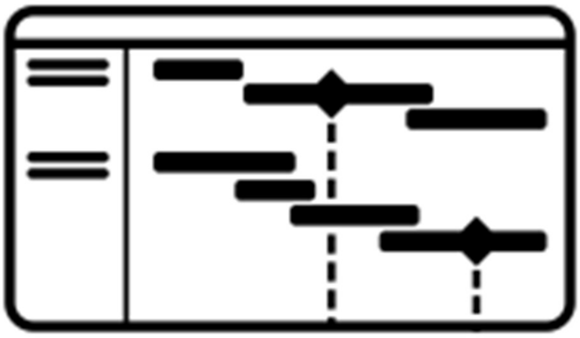

Fig 5 Heat map of the analysis

Sharing the visualizations and insights generated in Tableau with stakeholders is streamlined through the Tableau Server or Tableau Public. This ensures that critical insights reach the right audience, whether they are internal airline teams, external partners, or the public. The ability to continuously update the visualizations as new data becomes available supports real-time monitoring of flight delays, which is vital for data-driven decision-making in airline operations.

In sum, Tableau is a powerful and versatile tool for visualizing and communicating flight delay analysis. Its user-friendly interface and a wide range of features, including basic visualizations, heatmaps, geographic mapping, dashboards, trend analysis, drill-down interactivity, and storytelling, make it an excellent choice for anyone looking to extract insights from flight delay data.

2.5 MODULES OF THE PROJECT

Airline On-Time Performance (OTP) is a critical metric that directly impacts passenger satisfaction and operational efficiency. Understanding the factors contributing to flight delays is paramount for the aviation industry. This project seeks to dissect flight arrival delay into three distinct components—Turn Delay, Block Delay, and Previous Delay—to provide a nuanced and comprehensive analysis of OTP. The multifaceted goals encompass various aspects, from root cause identification to predictive modeling and continuous improvement.

The primary goal is to achieve a profound understanding of flight delays by breaking them down into individual components. Flight delays are intricate phenomena influenced by diverse factors such as airport operations, air traffic, and weather conditions. This goal emphasizes a detailed examination of Turn Delay, Block Delay, and Previous Delay, recognizing their unique contributions to overall delays. By comprehensively understanding each component, stakeholders can devise targeted strategies for improvement. Identifying the root causes of flight delays within the dissected components is crucial. Delving into the interplay of factors contributing to Turn Delay, Block Delay, and Previous Delay allows for a granular analysis. This goal involves statistical analysis, machine learning techniques, and collaboration with industry experts to pinpoint the core issues. Identifying root causes is foundational for developing effective strategies to mitigate delays and enhance overall OTP.

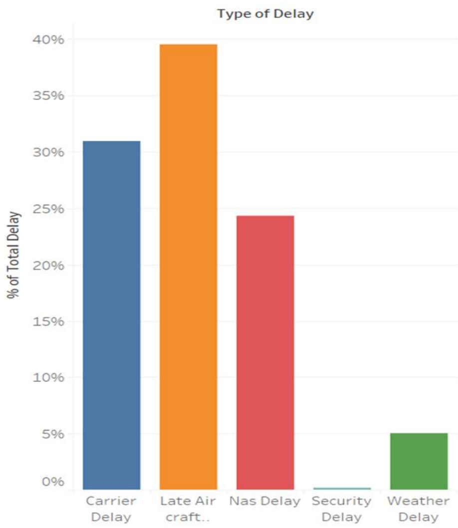

Fig 6 Delay percentage

Temporal analysis is essential for understanding patterns and season-specific challenges related to dissected delay components. Seasonal variations, peak travel times, and trends over time offer critical insights. This goal informs scheduling, resource allocation, and proactive measures to mitigate delays during high-impact periods. Utilizing historical data and advanced analytics, the project aims to uncover patterns and provide actionable insights to stakeholders.

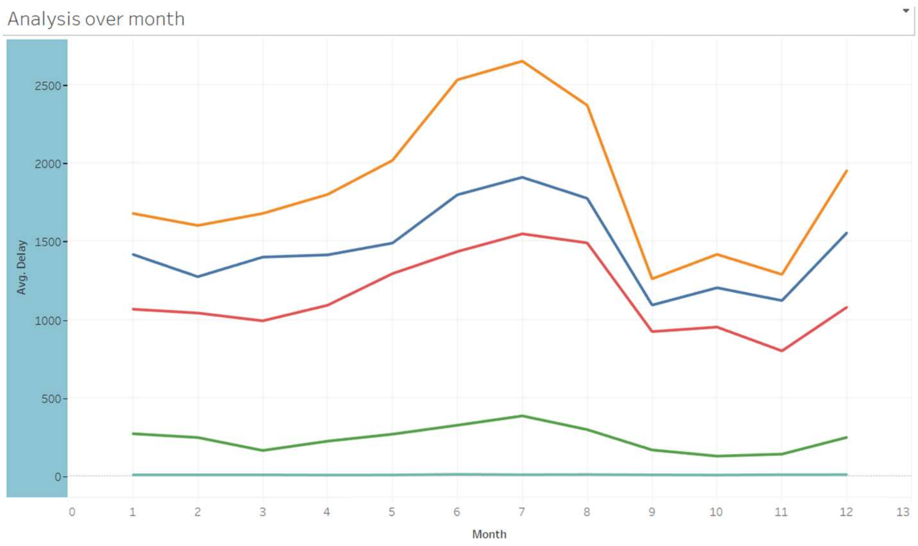

Fig 7 Analysis over month

Comparing the performance of different airlines, airports, and routes concerning dissected delay components is vital for industry-wide improvements. This analysis highlights variations and best practices among industry players. Airlines can learn from one another, adopt successful strategies, and work on improving weak points to enhance overall OTP. The comparative analysis involves benchmarking against industry standards, identifying outliers, and promoting a culture of healthy competition and knowledge exchange.

Fig 8 Top 20 Airport

Understanding how dissected delay components influence passenger experiences is a key aspect of this project. This goal involves assessing the impact of delays on passenger connections, customer satisfaction, and operational improvements oriented towards passengers. By addressing these factors, airlines can improve passenger loyalty and their overall travel experience. The passenger-centric focus incorporates feedback mechanisms, surveys, and collaboration with consumer advocacy groups to ensure that passenger perspectives are integral to the analysis.

Fig 9 Cancelled or diverted

Developing predictive models based on historical data is instrumental in forecasting potential delays. Predictive modeling can aid in operational planning, enabling airlines to allocate resources more efficiently, manage staffing levels, and minimize disruptions to passengers. By anticipating delays, airlines can proactively manage their schedules and mitigate the impact of delays on OTP. This goal involves employing advanced statistical and machine learning techniques to create accurate models that consider various influencing factors.

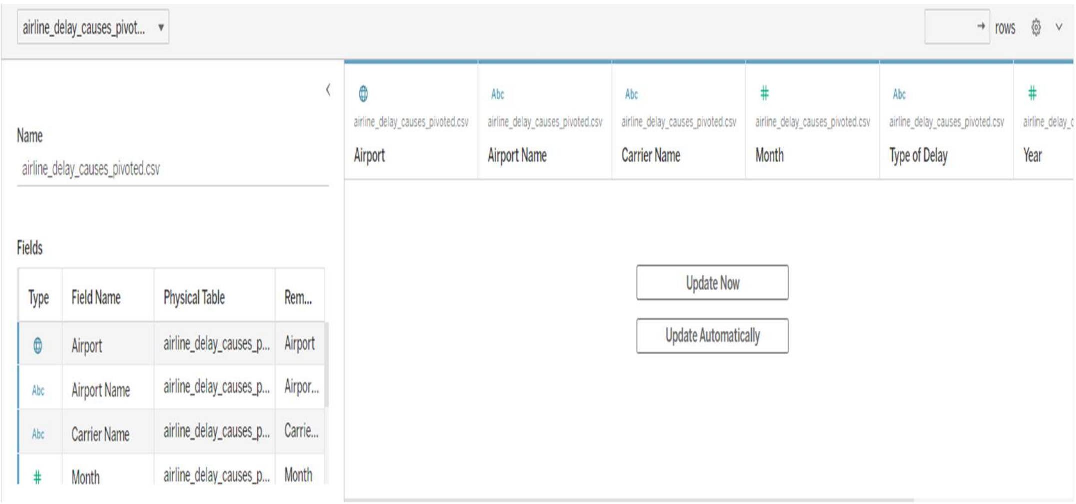

Fig 10 Data Source

Effective data visualization is essential for presenting findings clearly and facilitating data-driven decision-making. Visual representations, such as charts, graphs, and dashboards, are powerful tools for conveying complex information in a comprehensible manner. This goal involves creating informative visualizations that allow stakeholders to quickly grasp the insights generated from the analysis. Utilizing visualization tools like Tableau, the project aims to bridge the gap between raw data and practical knowledge, ensuring that stakeholders can easily interpret and act upon the findings.

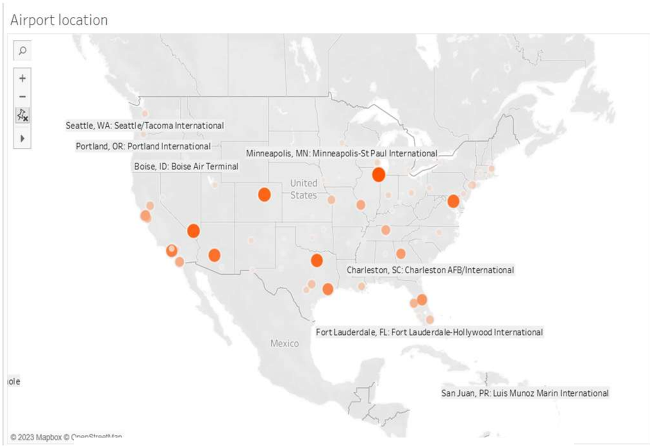

Fig 11 Airport location

Providing actionable recommendations to airlines is a critical component of this project. These recommendations should be based on the analysis of dissected delay components and offer practical strategies for optimizing operations and reducing the impact of flight delays on OTP. Implementing these recommendations can lead to tangible improvements in airline punctuality and overall efficiency. The actionable recommendations may involve operational adjustments, technology implementations, and changes in standard operating procedures. Improved transparency and communication are essential for fostering understanding of flight delays among various stakeholders, including airlines, airports, regulatory authorities, and passengers. Clear and open communication can lead to greater cooperation and collaborative efforts in addressing delay issues. Passengers benefit from being well-informed, and regulators can make more effective policy decisions with access to transparent data. The project aims to establish clear channels of communication, disseminate findings through accessible mediums, and engage stakeholders in open discussions.

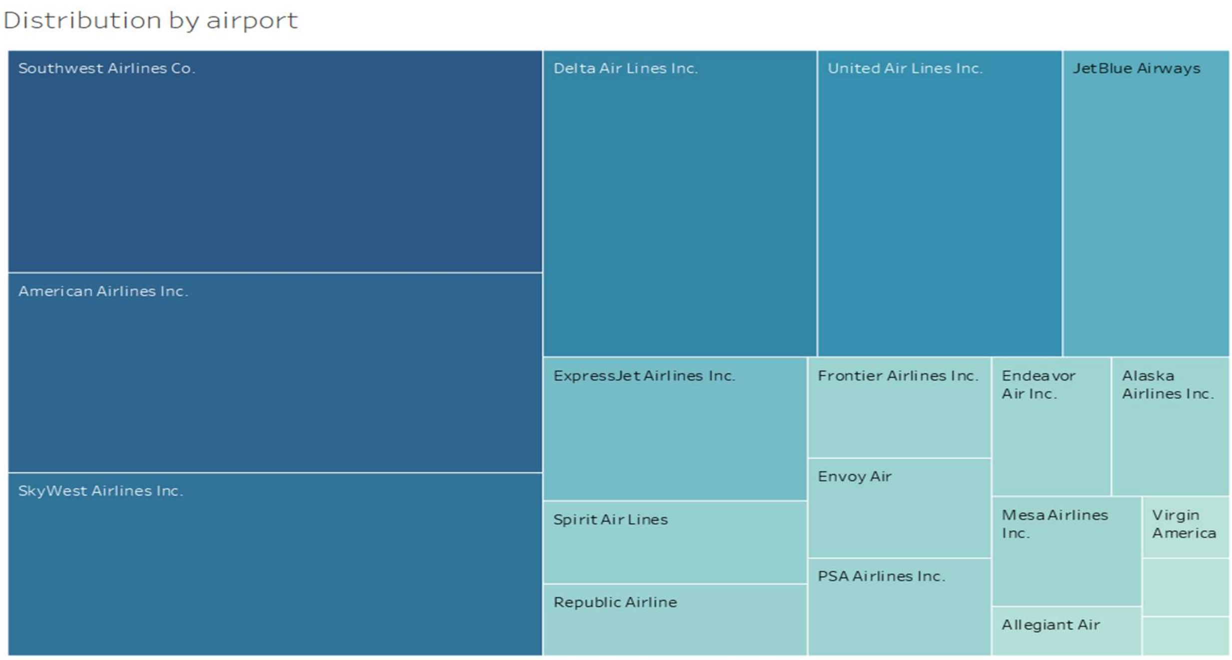

Fig 12 Distribution by airport

2.6 DELAY PREDICTION

Step 1: Importing Libraries

In the initial step, necessary libraries are imported for data manipulation, machine learning, and model evaluation. Pandas, an essential tool for handling tabular data, is aliased as 'pd.' The XGBoost library, often used for gradient boosting, is imported as 'xgb.' From the scikit-learn library, specific functions such as 'train_test_split' for data splitting and various metrics for model evaluation are imported.

Step 2: Reading and Preprocessing Data

The second step involves reading data from a CSV file, specifically "airline_delay_causes_pivoted.csv," using Pandas. The 'Delay' column is converted to a floating-point format, and rows with missing 'Delay' values are dropped. Features and labels are separated, and categorical features are one-hot encoded to represent them numerically. Any missing values in the features are filled with their mean. The dataset is then split into training and testing sets.

Step 3: Model Training

In the third step, an XGBoost regressor model is created with the objective of minimizing squared error. This model is then trained on the provided training data.

Step 4: Prediction and Evaluation

Following model training, predictions are made on the test set. Mean Squared Error (MSE) is calculated using scikit-learn's 'mean_squared_error' function, comparing the actual and predicted values. The calculated MSE, as well as the first few rows of the features and labels, are printed for evaluation.

Step 5: User Input Simulation

To simulate user input, a dictionary named 'user_input' is created, representing features such as 'departure_time,' 'airline,' and 'destination.' The 'Month' and 'Year' are derived from the 'departure_time' and are subsequently added to the user data. Categorical variables are one-hot encoded using 'pd.get_dummies.' A new DataFrame ('user_data') is generated from the user input, and missing one-hot encoded columns are added and set to 0.

Step 6: Making Predictions for User Input

The final step involves using the trained XGBoost model to predict delays for the simulated user input. The prediction is then interpreted, and a corresponding message is printed based on whether the model predicts a delay or not.

2.7 RESULTS

The results of the project, which involved dissecting flight arrival delay into three components for airline On-Time Performance (OTP) analysis, provide a wealth of valuable insights that have the potential to transform the aviation industry's understanding of flight delays and significantly improve the overall travel experience for passengers.

Temporal analysis conducted as part of the project has also yielded significant insights. By examining the temporal trends of these dissected delay components, the project has uncovered distinct patterns in how delays vary over time. This information is instrumental in optimizing scheduling and resource allocation. For example, identifying specific days, months, or seasons when delays are more prevalent enables airlines to allocate resources more effectively during peak periods. Furthermore, temporal insights can inform proactive strategies for mitigating delays during high-impact times, improving OTP, and ensuring passengers experience fewer disruptions in their travel plans.

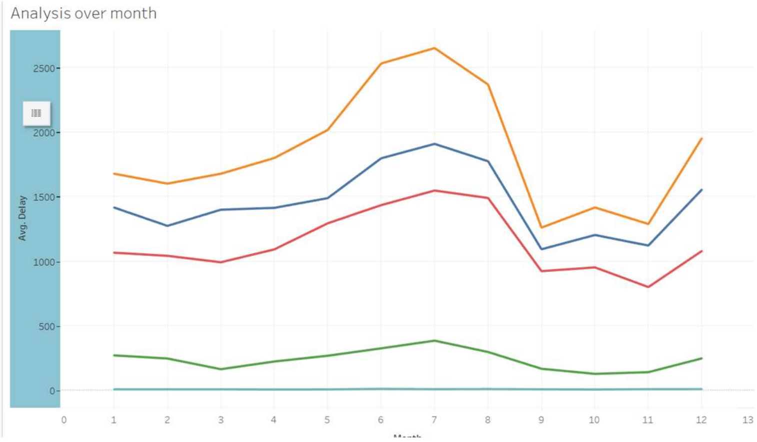

Fig 3 Line graph of the analysis

The comparative analysis conducted in the project is another valuable contribution. This analysis has allowed for benchmarking among different airlines, airports, and routes. By comparing the performance of various stakeholders, the project has created opportunities for knowledge transfer and industry-wide improvements. Airlines can learn from the best practices of their peers, adopting strategies and techniques that have proven successful in reducing delays and improving OTP. This kind of cross-industry learning is pivotal in fostering.

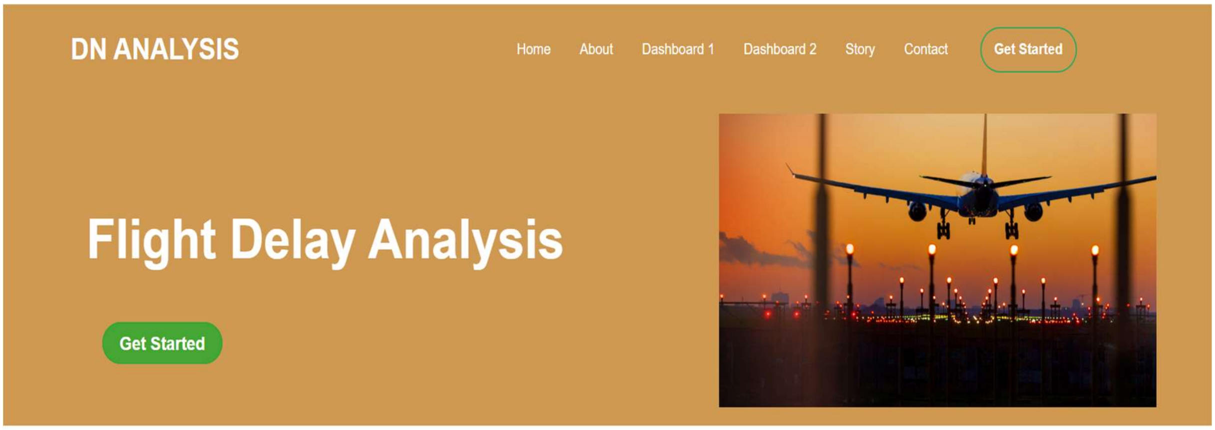

Fig 13 Web Interface

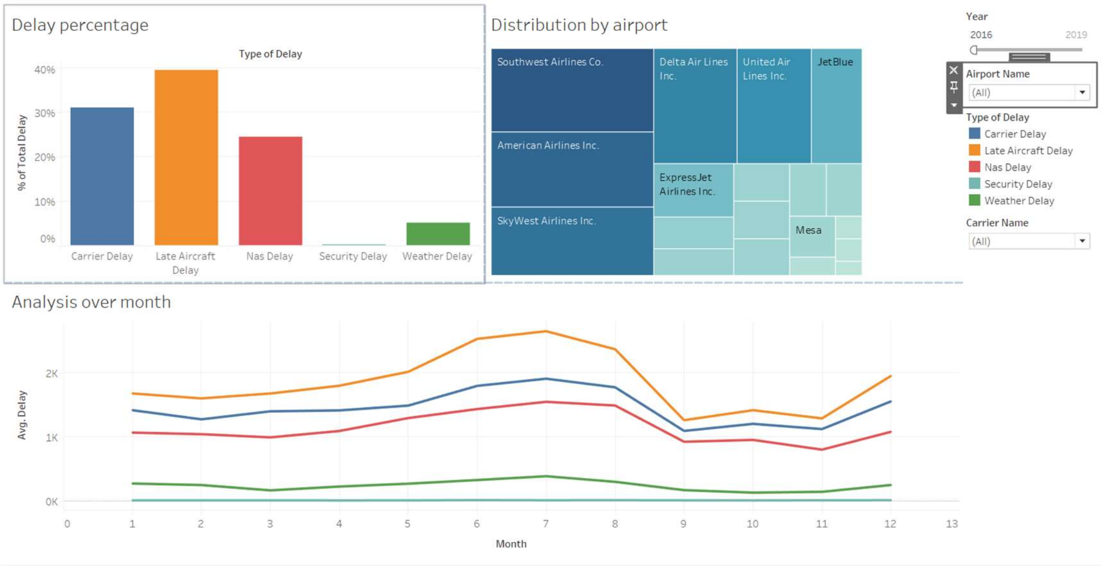

Fig 14 Dashboard 1

Data visualization tools, such as Tableau, have been employed to simplify the communication of the project's findings. These tools have transformed complex data into clear and informative visual representations that enhance comprehension for stakeholders. By presenting data through visualizations like charts, graphs, and dashboards, the project has made it easier for airlines, airports, regulatory authorities, and the general public to grasp the key insights and take action based on the information provided. Effective data visualization is a bridge that connects raw data to actionable knowledge.

Finally, the project incorporates mechanisms for continuous monitoring and reporting. These mechanisms provide real-time insights into OTP and delay trends, enabling airlines and other stakeholders to make data-driven decisions and take immediate actions when necessary. This approach fosters a culture of continual improvement within the aviation industry, where data is actively used to enhance punctuality and customer satisfaction.In summary, the project's results offer a comprehensive understanding of flight delays, their sources, and their impacts, providing a valuable resource for the aviation industry to improve OTP, enhance passenger experiences, and optimize operations. By dissecting flight delays into their components and addressing the findings collectively, the project has the potential to transform the way airlines and stakeholders approach delay reduction and punctuality, ultimately benefiting both the industry and passengers.

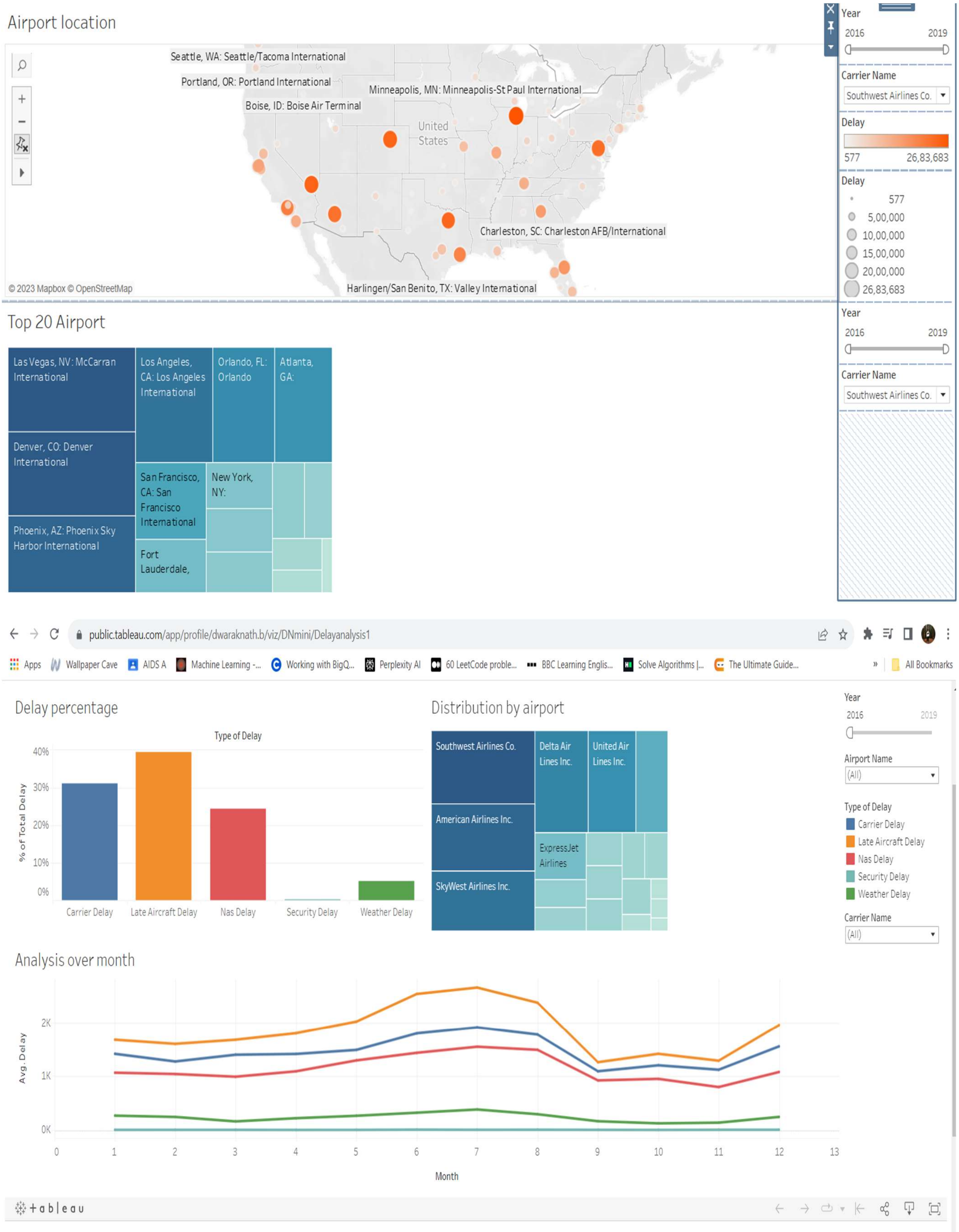

Fig 15 Dashboard 2

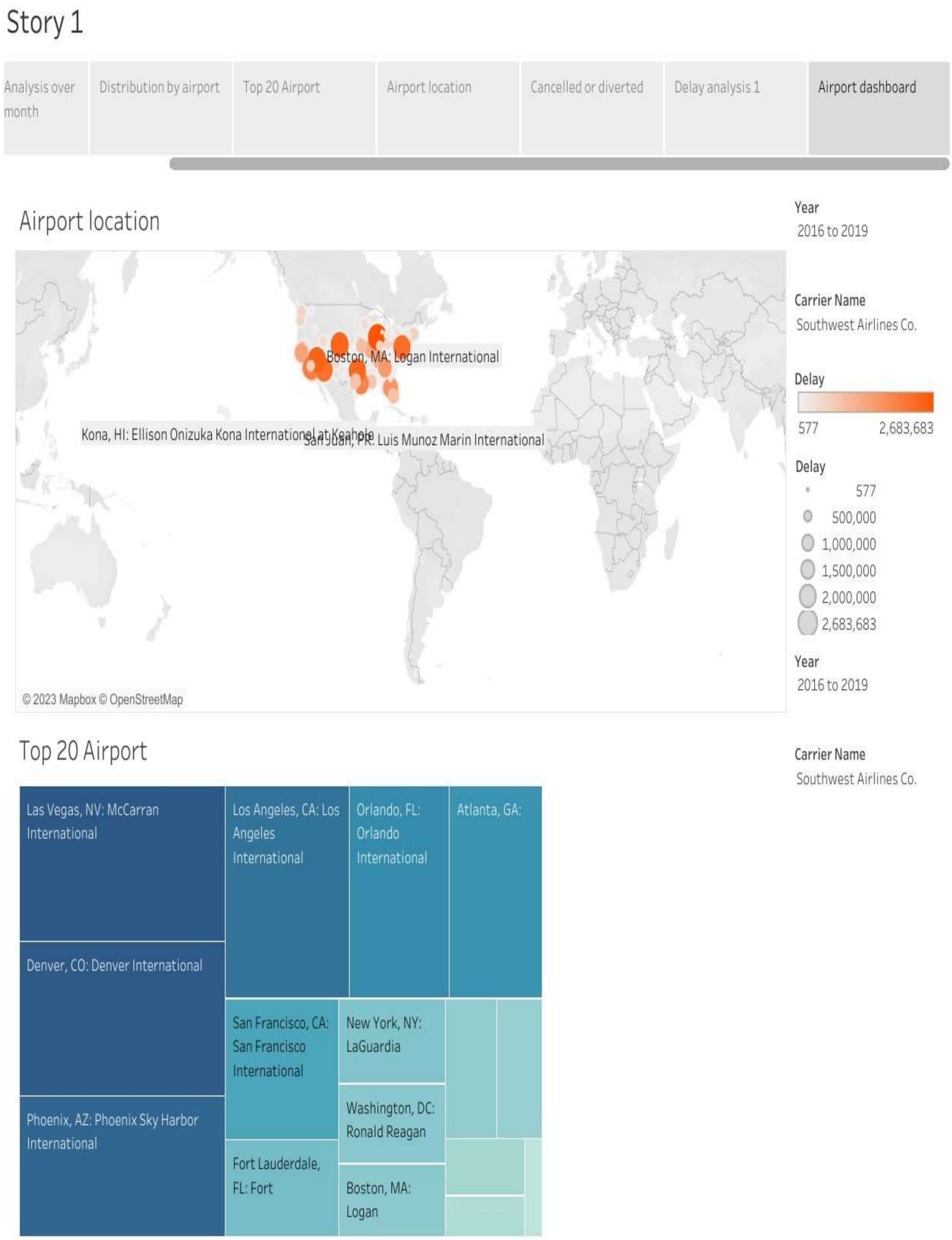

Fig 16 overall analysis

CHAPTER 3

CONCLUSION

The project's dissection of flight delay into Turn Delay, Block Delay, and Previous Delay illuminates key challenges for the aviation industry, offering a clear path to enhance On-Time Performance (OTP). Identifying these components allows prioritized efforts and effective resource allocation, crucial for improving punctuality. Temporal analysis reveals patterns, aiding proactive scheduling adjustments. Predictive models, using historical data, enable proactive delay management, minimizing disruptions and improving OTP. Data visualization, facilitated by tools like Tableau, simplifies the communication of findings, making insights actionable.

Actionable recommendations provide practical strategies for immediate implementation. Transparent communication ensures accessibility to a wide range of stakeholders, fostering collaboration. Continuous monitoring and reporting offer real-time insights for informed decision-making and a culture of continual improvement. Geographic and temporal analyses provide strategic insights into specific challenges across regions and seasons. In summary, the project's practical insights offer a comprehensive roadmap for the aviation industry to improve OTP, passenger experiences, and operational efficiency, driving positive change and ensuring a more reliable and punctual travel experience for passengers.

3.1 FEATURE ENHANCEMENT

The enhancement of flight delay analysis is a critical pursuit in the aviation industry, demanding a comprehensive and multi-faceted approach to address the myriad factors contributing to disruptions. One pivotal aspect of this enhancement involves the integration of real-time weather data into analytical frameworks. Adverse weather conditions, such as thunderstorms, heavy snowfall, or low visibility, are notorious contributors to flight delays. By assimilating up-to-the-minute meteorological information, airlines and airports can gain a more accurate understanding of potential disruptions, enabling proactive decision-making.

Machine learning models stand as another pillar of feature enhancement. By leveraging historical flight data, these models can discern intricate patterns and correlations that might elude traditional analysis. Predictive algorithms, trained on vast datasets encompassing various parameters, empower the industry to forecast delays with greater precision. These models can adapt to evolving conditions, continuously learning from real-time data to refine their predictive capabilities.

Incorporating airport-specific data is equally crucial. Airport congestion and runway availability significantly impact flight schedules. Delays arising from overburdened facilities or limited runway capacity can be better anticipated and managed when such data is integrated into the analysis. This approach enables stakeholders to proactively address issues at specific airports, mitigating the domino effect that delays at a major hub can have on the entire network.

Aircraft turnaround time, encompassing the time an aircraft spends on the ground between flights, is a critical parameter affecting scheduling and delays. Integrating this aspect into analysis allows for a more nuanced understanding of potential disruptions. Crew scheduling and availability further contribute to the complexity of flight operations. Analysing crew-related factors, such as rest requirements and scheduling constraints, can uncover delays originating from staffing issues.

Passenger load factor, representing the percentage of available seating capacity that is filled with passengers, is another valuable feature. High load factors can impact turnaround times, boarding processes, and overall operational efficiency. Analysing this data provides insights into potential delays related to passenger boarding and deplaning.

In conclusion, the feature enhancement of flight delay analysis involves a synergistic integration of diverse datasets and advanced technologies. From weather patterns to crew scheduling, each element contributes to a more comprehensive understanding of delays. The industry's pursuit of excellence in delay analysis not only improves operational efficiency but also enhances the overall passenger experience, marking a significant step towards a more resilient and adaptive aviation ecosystem.

REFERENCES

1. Anderson Ong, M. (2017) “Data Visualization of Flight Delays with Tableau”. Medium.
2. Dey, T., Phillips, D. and Steele, P. (2011). A Graphical Tool to Visualize Predicted Minimum Delay Flights. Journal of Computational and Graphical Statistics.
3. Gopalakrishnan K., Balakrishnan H. (2017). “A Comparative Analysis of Models for Predicting Delays in Air Traffic Networks.” Air Traffic Management Research and Development Seminar, June 2017, Seattle, Washington, USA, ATM Seminar.
4. Gupta P., Dwivedi A., Agrawal A. (2016). “An Analysis of US Domestic Flight delays using SAS Enterprise Miner”. South Central SAS Users Group Educational Forum.
5. Mainero A., Schmidt T. and Sugarman H. (2013). Heat Mapping and Predicting Flight Delays and Their Propagations in a Real-World Air Traffic Simulation.
6. Michael T. Crotty. (2014). “Visualizing More Than Twenty Years of Flight Data for the Raleigh-Durham International Airport.” SAS Institute, USA.
7. Rodríguez-Sanz, Á., Gómez Comendador, F., Arnaldo Valdés, R., Cordero García, J. and Bagamanova, M.
8. Sternberg A., Soares J., Carvalho D., Ogasawara E. (2017). A Review on Flight Delay Prediction. CEFET/RJ Rio de Janeiro, Brazil.
9. University of Southern California, USA (2021). Journal of Computational and Graphical Statistics.
10. Analysing Factors Influencing Flight Delay Prediction(2019). South Central SAS Users Group Educational Forum.

11. https://doi.org/10.1109/jsyst.2019.2925346
12. https://doi.org/10.1109/jsyst.2019.2925346.
13. https://www.fly.faa.gov/products/training/traffic_management_for_pilots/tfm_in_the_nas_booklet_ca10.pdf.
14. https://filesmart.org/resources/traffic-management-initiatives/
15. https://aspm.faa.gov/apm/sys/main.asp.
16. https://www.oag.com/understanding-oag.
17. https://www.faa.gov/air_traffic/publications/cip/
18. https://www.faa.gov/data_research/aviation/aerospace_forecasts/me dia/fy2019–39_faa_aerospace_forecast.pdf.
19. https://www.weather.gov/asos/asostech
20. https://cdm.fly.faa.gov/?page_id=42.