International Journal of Engineering Applied Sciences and Technology, 2025
Vol. 10, Issue 01, ISSN No. 2455-2143, Pages 82-90
Published Online May 2025 in IJEAST (http://www.ijeast.com)

COMPREHENSIVE DATA ANALYTICS STUDY OF THE AIRLINE INDUSTRY FROM 2012–2020 USING TABLE AU

K.Nihanth, S.Harsh, A.Sai Kiran
U G Scholar, Dept. of Artificial Intelligence and Data Science, Department of IT
Methodist College of Engineering and Technology,
Hyderabad, Telangana, India

Dr. Diana Moses
Professor, Department of Computer Science and Engineering,
Methodist College of Engineering and Technology,
Hyderabad, Telangana, India

Abstract— **The airline industry is increasingly leveraging data-driven techniques to improve customer satisfaction, operational efficiency, and sustainability. Several studies focus on utilizing text mining, sentiment analysis, and machine learning to analyze customer feedback from social media platforms, such as Twitter, to better understand passenger experiences. Hybrid models combining lexicon-based methods with deep learning have been proposed to enhance sentiment accuracy, while methodologies for analyzing flight delays, cancellations, and safety reports have been introduced to identify operational inefficiencies. Furthermore, sustainability practices in the airline industry are explored through machine learning clustering and bibliometric analysis, providing valuable insights for airlines to improve their environmental performance. This study also highlights the importance of integrating structured flight operation data with customer sentiment data for more comprehensive insights. Additionally, the use of Tableau facilitates real-time visual storytelling, allowing stakeholders to make faster and more informed decisions. Overall, the integration of advanced data analysis methods offers significant potential to enhance airline operations and customer service while addressing sustainability challenges.**

Keywords— **Airline Industry, Data Analytics, Business Intelligence, Operational Efficiency**

I. INTRODUCTION

LITERATURE SURVEY:

Aljedaani et al., Similar to other service industries, the airline industry utilizes such feedback for determining customers' satisfaction levels and improving the quality of experience where needed. This, of course, requires accurate sentiments from the user tweets. Existing sentiment analysis models suffer from low accuracy on account of the contradictions found in the tweet text and the assigned label. From this perspective, this study proposes a hybrid sentiment analysis approach where the lexicon-based methods are used with deep learning models to improve sentiment accuracy [1].

Miyamoto et al., With the growth in commercial aviation traffic and the need for improved environmental performance, strategies to lower emissions that can be implemented in the near term are necessary. Since novel technology takes time to enter the market, operational improvements that employ existing aircraft and require no new infrastructure are fit for this goal. While quantified data collected throughout aviation, such as arrival/departure statistics and flight data, have been well-utilized, text data collected through safety reports have not been leveraged to their full extent. In this paper, a methodology is presented that can use aviation text data to identify high-level causes of flight delays and cancellations, using delays as a metric of operational inefficiency [2].

Kumar et al., Customer's experience is one of the important concerns for airline industries. Twitter is one of the popular social media platforms where fight travellers share their feedback in the form of tweets. This study presents a machine learning approach to analyse the tweets to improve the customer's experience. Features were extracted from the tweets using word embedding with Glove dictionary approach and n-gram approach. Further, SVM (support vector machine) and several ANN (artificial neural network) architectures were considered to develop a classification model that maps the tweet into positive and negative category [3].

Tian et al., From the start, the airline industry has remarkably connected countries all over the world through rapid long-distance transportation, helping people overcome geographic barriers. Consequently, this has ushered in substantial economic growth, both nationally and internationally. The

airline industry produces vast amounts of data, capturing a diverse set of information about their operations, including data related to passengers, freight, flights, and much more. Analysing air travel data can advance the understanding of airline market dynamics, allowing companies to provide customized, efficient, and safe transportation services [4]. Yenkikar et al., Airlines operate in a competitive marketplace and must upgrade their services to meet customer safety and comfort. Post-pandemic, the government and airlines resumed flights with many restrictions, the impact which is unexplored. An increasing number of customers use social media to leave reviews and in this age of Machine Learning (ML), if a model is available to automatically polarize flyer sentiments, it can help airlines upscale. In this work, a custom dataset is scraped from Twitter by including online reviews of five Indian airlines [5].

Rose et al., The complexity of commercial aviation operations has grown substantially in recent years, together with a diversification of techniques for collecting and analysing flight data. As a result, data-driven frameworks for enhancing flight safety have grown in popularity. Data-driven techniques offer efficient and repeatable exploration of patterns and anomalies in large datasets. Text-based flight safety data presents a unique challenge in its subjectivity, and relies on natural language processing tools to extract underlying trends from narratives. In this paper, a methodology is presented for the analysis of aviation safety narratives based on text-based accounts of in-flight events and categorical metadata parameters which accompany them [6].

Şapaloğlu et al., This study aims to comprehensively explore sustainability practices, guidelines, and emerging trends in the airline industry through an in-depth analysis of their sustainability reviews. A thorough bibliometric analysis of airline sustainability was conducted using the Scopus Database. Additionally, employing text-mining techniques, a meticulous analysis focused on the sustainability reports of the leading 20 airlines compared to the subsequent 20, considered as followers, in the textual examination of sustainability reports [7].

Walton et al., One often hears of the dangers of transporting lithium batteries by air, but there are many other potentially dangerous goods (DG) that are transported on aircraft, both cargo and passenger. If not properly packaged and handled these dangerous goods can be a risk to passengers and crew or to overall flight safety and have led to the loss of life and aircraft in the past. The purpose of this study was to examine the types of dangerous goods (DG) incidents on aircraft as reported to the U.S. Aviation Safety Reporting System (ASRS) database from January 2010 to January 2020. This research was conducted using the SAS Text Miner software package. The Text Miner software provides an understanding of how the narrative text of data relates to each other [8].

Hasib et al., One of the most important things that managers need is feedback from customers about the quality of the products and services their firms offer. Getting feedback from customers is now typically done via online channels including social media, instant messaging, and review websites. However, due to the massive volume, a powerful tool is required for processing the data received from various sources. Sentiment analysis is a form of NLP that can instantly determine if a user is feeling positive, negative, or neutral towards a given topic [9].

Arachchi et al., This analysis delves into sustainability within the aviation sector using machine learning and clustering. It uncovers distinct airline clusters based on sustainability focus. The study was conducted utilizing both the Random Forest algorithm and the K-means clustering algorithm. Despite uncovering trends, the analysis concentrates on 16 out of 17 United Nations sustainability goals, overlooking one aspect. Future research could benefit from better data collection and advanced models to improve sustainability analyses in aviation and similar industries [10].

II. MATERIALS & METHODS :

Tableau :

Tableau is an advanced data visualization software for converting raw data into actionable insights using interactive dashboards, charts, and reports. It finds extensive applications in business intelligence (BI), data analysis, and decision-making because it can process large volumes of data and offer real-time data analysis capabilities. Thanks to its drag-and-drop interface, users can create sophisticated visualizations without needing advanced coding skills. Tableau also connects easily with numerous data sources, such as Excel, SQL databases, and cloud-based systems. Its sophisticated capabilities, including geospatial analysis, predictive analytics, and AI-powered insights, make it a go-to for organizations seeking to improve data-driven decision-making.

Features :

Our platform provides an easy-to-use drag-and-drop interface, which simplifies creating beautiful visualizations with no coding required. It provides multi-data source support, including Excel, SQL, and cloud databases, for easy integration of data. With interactive dashboards, users can analyze in real time and get better insights immediately. With mapping and geospatial analysis, there is even more accuracy in location-based insights, further improving decision-making. Also, our system offers automated refreshing and sharing of data, maintaining reports as up-to-date and readily available as needed for collaborative purposes.

Overview Of Dataset:

This flight data was taken from 2012 to 2020 regarding a set of flights across various airlines. The dataset contains 401 recorded flights, each with 12 attributes: Flight ID, Airline, Flight Date, Departure Time, Arrival Time, Departure Location, Arrival Location, Departure City, Arrival City, Flight Duration Minutes, Passenger Count, and Ticket

International Journal of Engineering Applied Sciences and Technology, 2025
Vol. 10, Issue 01, ISSN No. 2455-2143, Pages 82-90
Published Online May 2025 in IJEAST (http://www.ijeast.com)

Revenue. The data provides insights into various airlines such as American Airlines, United Airlines, and Delta Airlines, with flights operating between major cities like New York, Los Angeles, Chicago, and San Francisco. All flights have a consistent duration of 150 minutes. Passenger counts range from 180 to 250 per flight, and ticket revenues vary, with an average of around $51,800. This dataset can be analysed to understand airline performance, flight scheduling, and revenue generation across different regions. The data also offers valuable information on the operational efficiency of airlines over a period of 8 years.

Stakeholders Benefits From This Dataset:

This dataset offers a wide range of benefits to various stakeholders in the airline industry. For operational teams, it provides insights into flight frequency, timing, and route efficiency, supporting better schedule planning and resource allocation. Revenue management teams can analyze passenger counts and ticket revenue to optimize pricing and improve yield management. Route planners and network analysts can evaluate the performance of different city pairs to inform strategic expansion or reduction decisions. Financial analysts and executive leadership gain a clear view of per-flight profitability, enabling informed investment and budgeting decisions. Marketing departments may use the data to identify popular routes and tailor promotional efforts accordingly, while airport authorities can plan for traffic volumes and infrastructure needs. Additionally, customer experience teams may draw on timing and route consistency data to enhance service quality and satisfaction. Overall, the dataset empowers stakeholders to make data-driven decisions that improve operational efficiency, financial performance, and customer engagement.

III. VISUALIZAIONS:

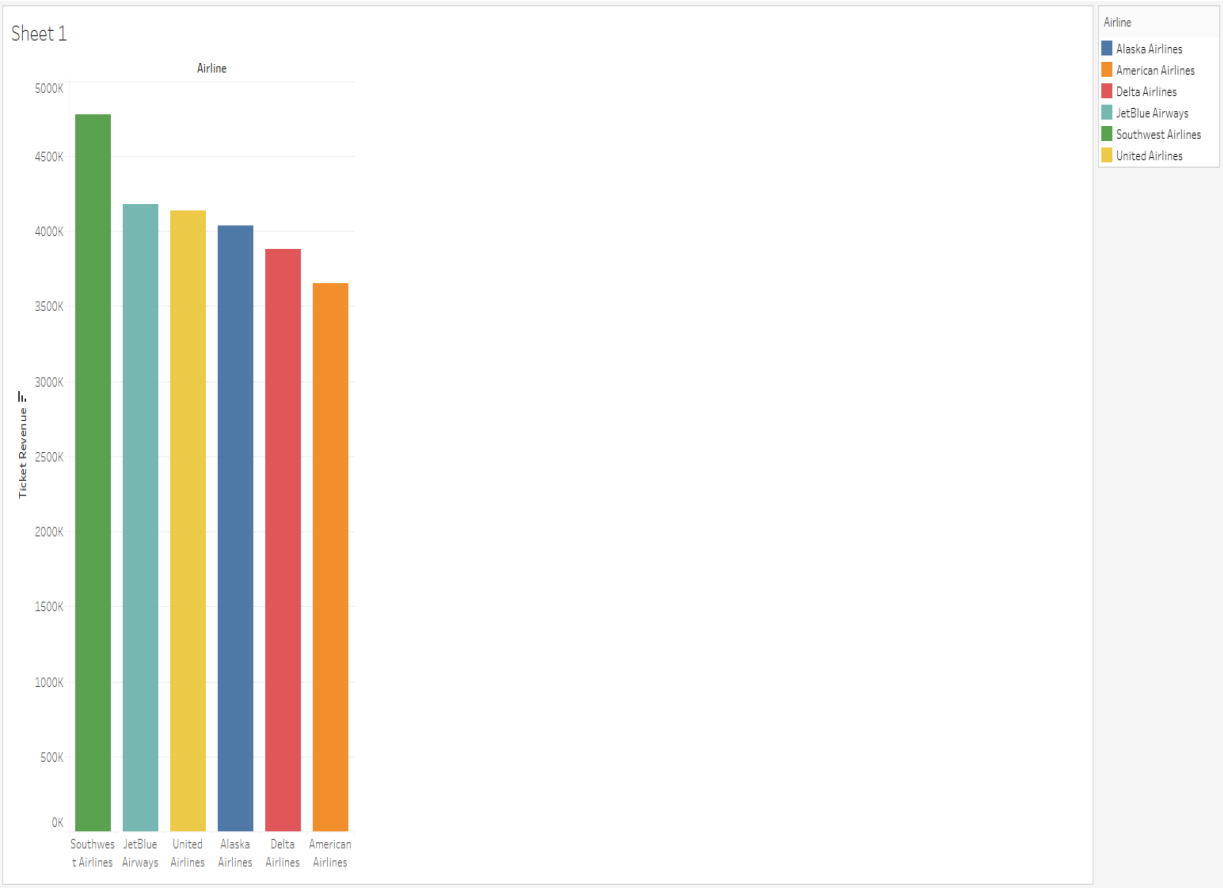

FIG.1

This graph shows the sum of ticket revenue for each Airline. Color shows details about Airline. The view is filtered on Airline, which keeps Southwest Airlines, Jetblue Airways, United Airways, Alaska Airlines, Delta Airlines & American Airlines. the Southwest Airlines has the highest ticket revenue among all the airlines.

International Journal of Engineering Applied Sciences and Technology, 2025
Vol. 10, Issue 01, ISSN No. 2455-2143, Pages 82-90
Published Online May 2025 in IJEAST (http://www.ijeast.com)

FIG.2

This graph shows the count of passengers for each Airline. Color shows details about Airline. The view is filtered on Airline, which keeps Southwest Airlines,Jetblue Airways, United Airways, Alaska Airlines, Delta Airlines & American Airlines. the Southwest Airlines has the most number of passengers among all the airlines.

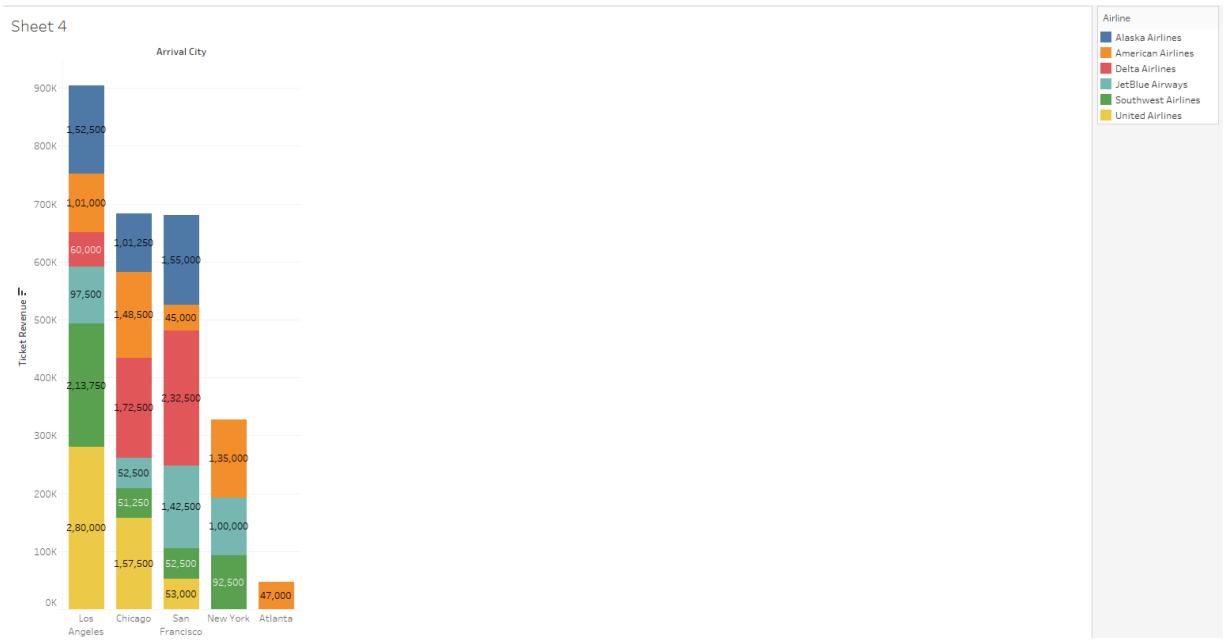

FIG.3

This chart represents the ticket revenue generated by various airlines in different arrival cities. The data suggests that Los Angeles is a major hub for airline revenue, with United Airlines leading in that city. Other cities show varied contributions across airlines, highlighting regional differences in airline dominance.

International Journal of Engineering Applied Sciences and Technology, 2025
Vol. 10, Issue 01, ISSN No. 2455-2143, Pages 82-90
Published Online May 2025 in IJEAST (http://www.ijeast.com)

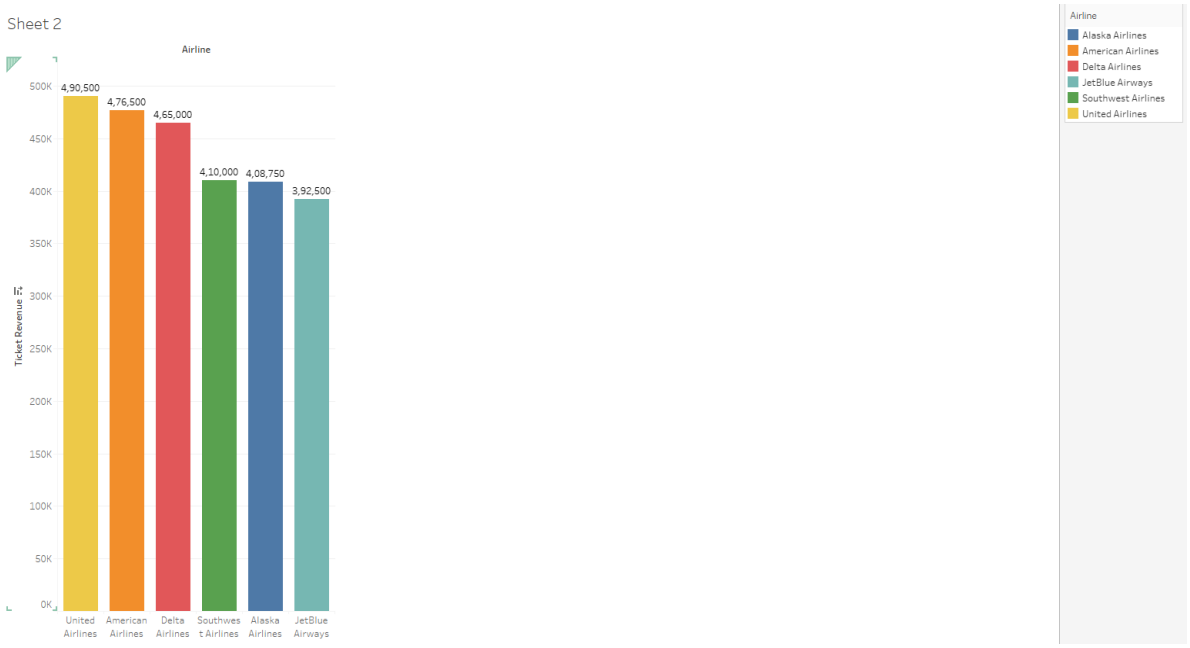

Fig.4

This chart compares the ticket revenue of different airlines. The chart indicates a relatively balanced distribution of revenue among the airlines, with United Airlines standing out as the top performer.

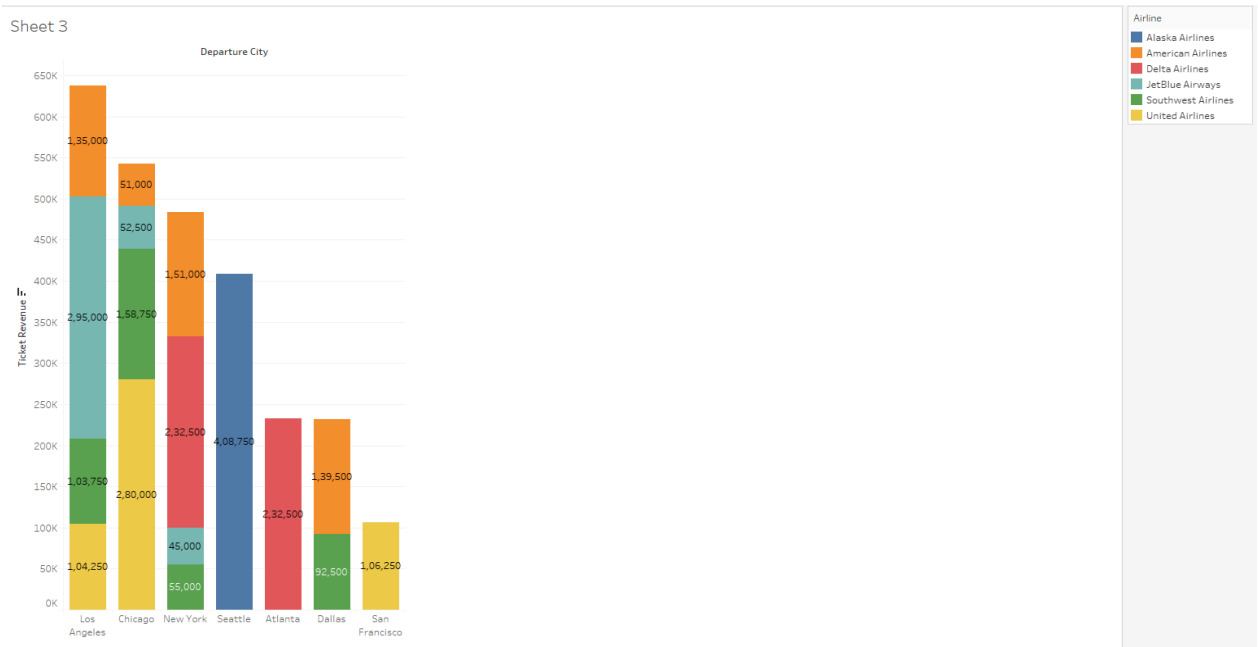

FIG.5

This chart shows the ticket revenue generated by different airlines across various departure cities. The data indicates that major hubs like Los Angeles, Chicago, and New York are critical for airline ticket revenues, with specific airlines having a dominant presence in certain cities.

International Journal of Engineering Applied Sciences and Technology, 2025
Vol. 10, Issue 01, ISSN No. 2455-2143, Pages 82-90
Published Online May 2025 in IJEAST (http://www.ijeast.com)

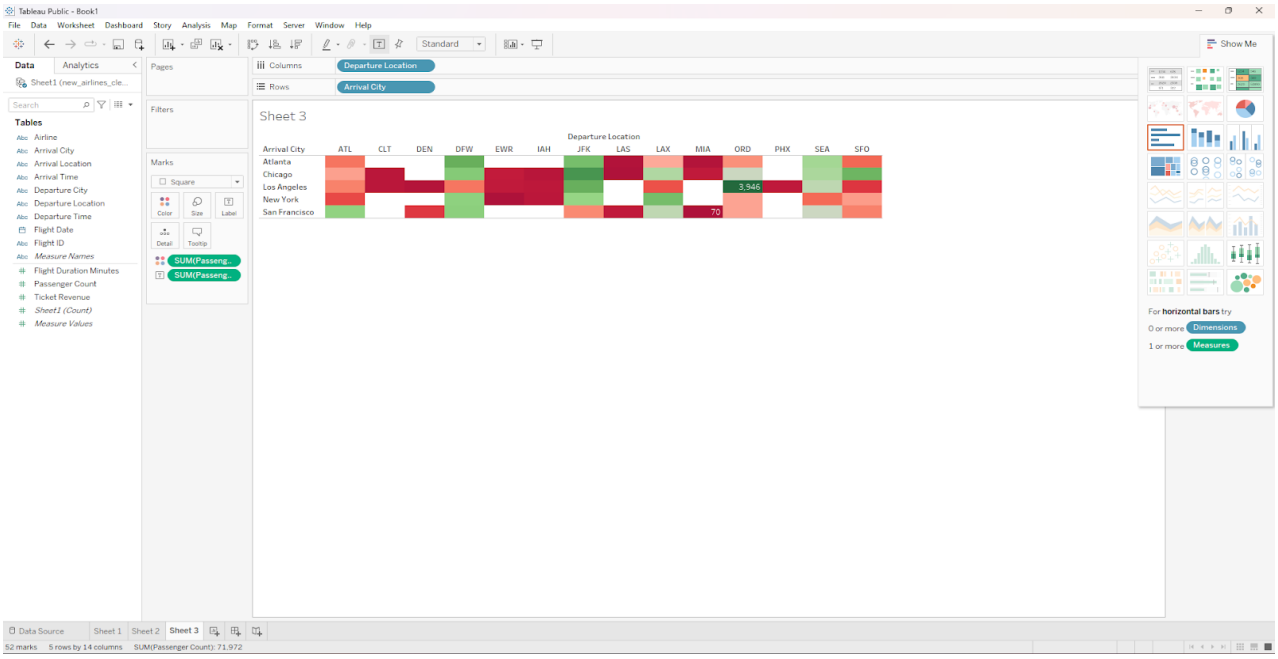

FIG.6

This visualization represents a heat map showing the distribution of passengers across various departure and arrival locations.

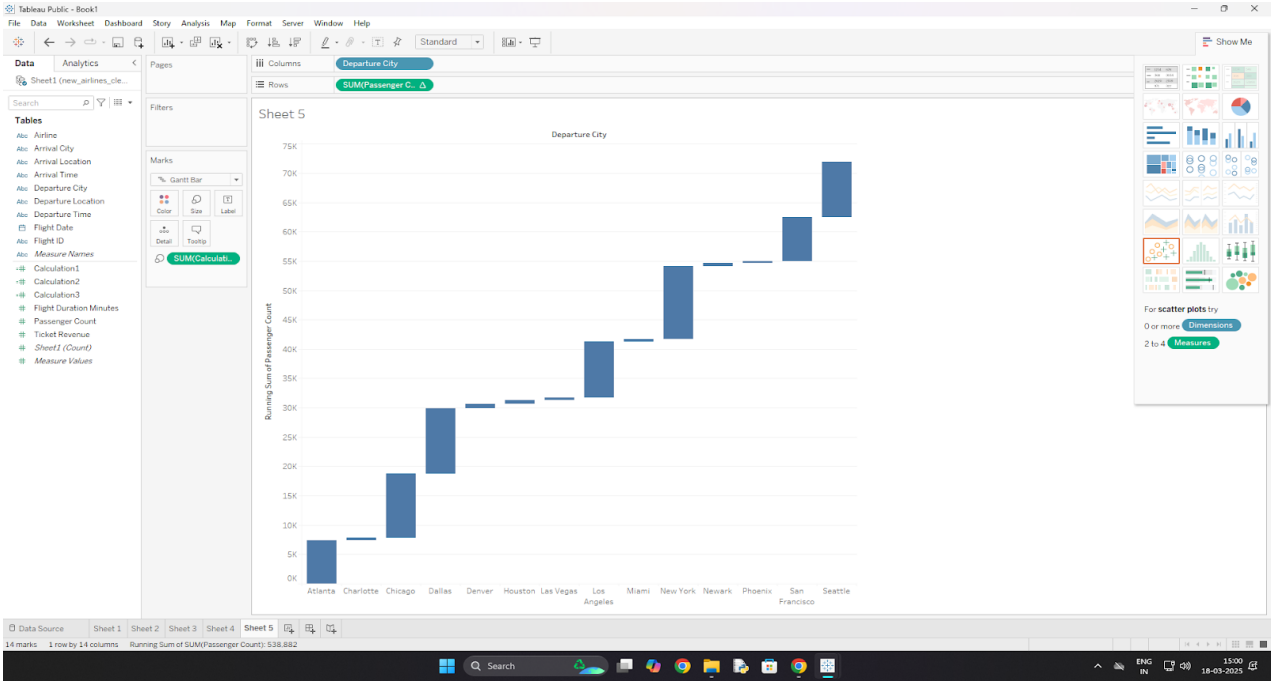

FIG.7

This Gantt Bar chart shows the running cumulative passenger count across departure cities. Each bar represents the total passengers up to that city, with Seattle contributing the highest cumulative count. It highlights how individual cities add to the overall total.

International Journal of Engineering Applied Sciences and Technology, 2025
Vol. 10, Issue 01, ISSN No. 2455-2143, Pages 82-90
Published Online May 2025 in IJEAST (http://www.ijeast.com)

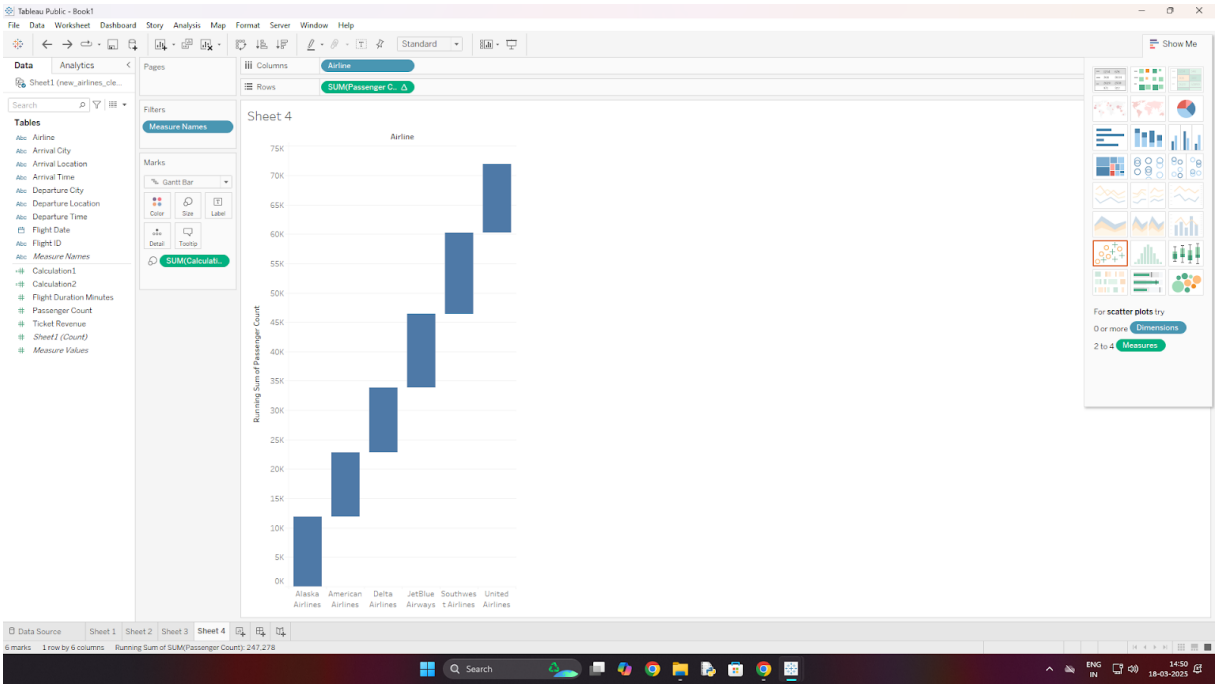

FIG.8

This Gantt Bar chart illustrates the running cumulative passenger count for different airlines. Each bar represents the total passengers up to that airline, with United Airlines contributing the highest cumulative count. It shows how each airline adds to the overall passenger total.

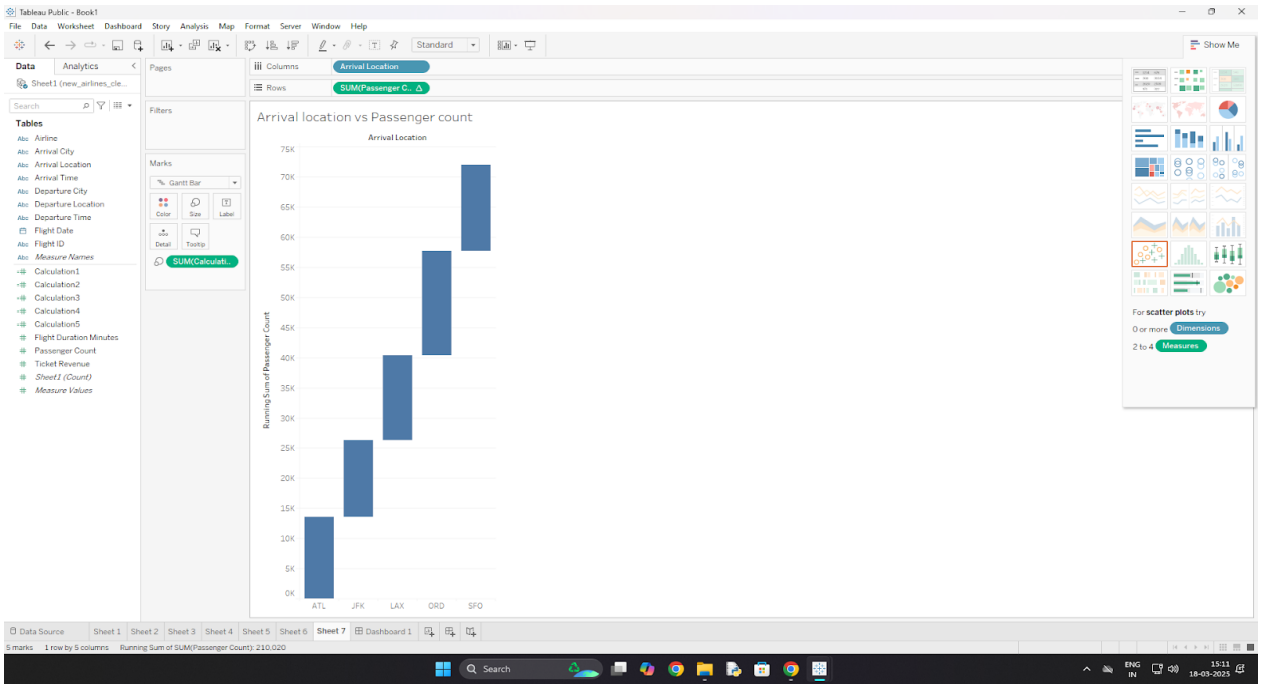

FIG.9

This Gantt Bar chart visualizes the running cumulative passenger count for various arrival locations. Each bar represents the aggregated total passenger count up to that location, with SFO (San Francisco) contributing the highest cumulative value. It demonstrates how different arrival locations cumulatively add to the overall passenger count.

International Journal of Engineering Applied Sciences and Technology, 2025
Vol. 10, Issue 01, ISSN No. 2455-2143, Pages 82-90
Published Online May 2025 in IJEAST (http://www.ijeast.com)

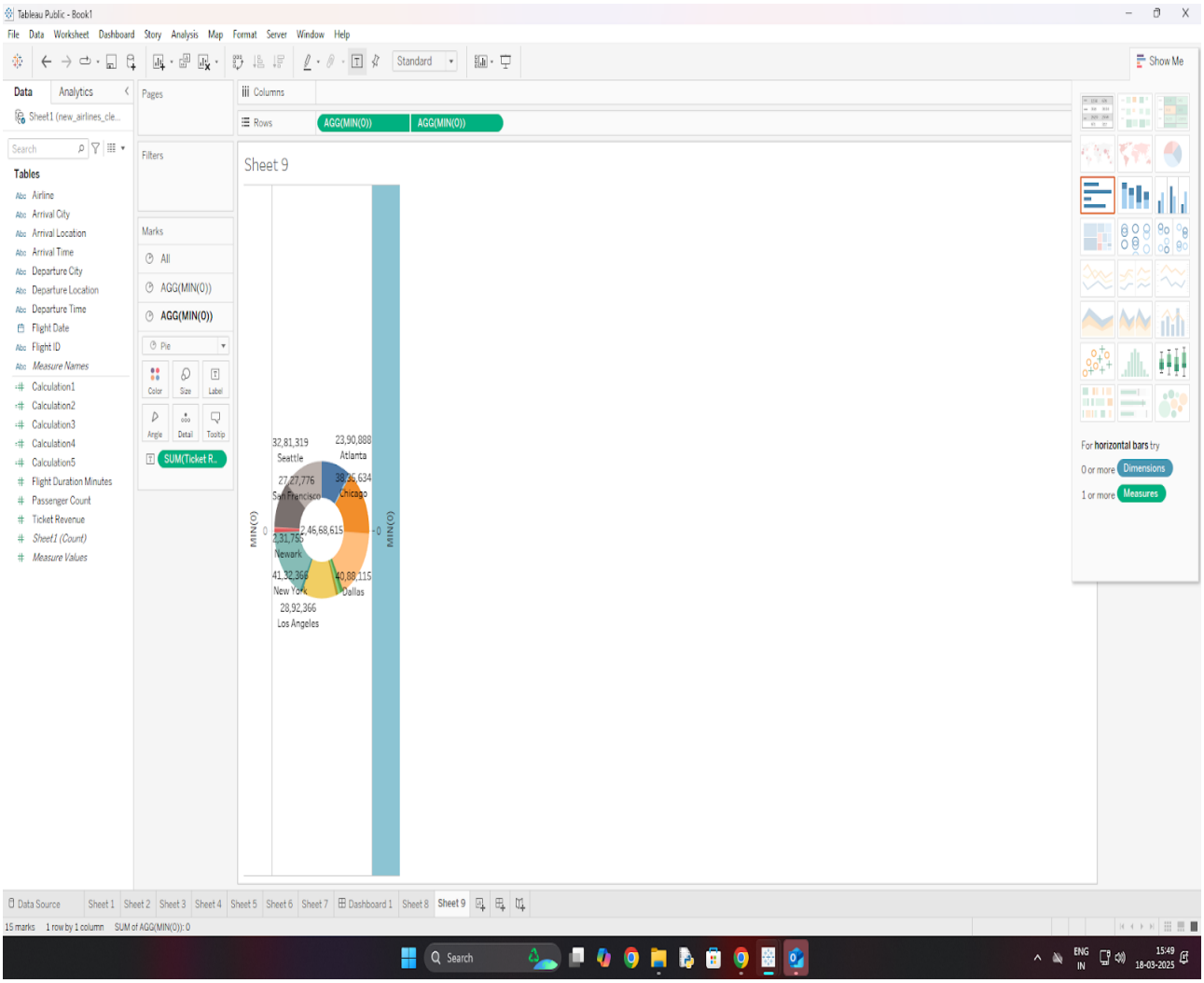

FIG.10

The pie chart shows ticket revenue for various cities, with New York having the highest at 41.32 million. Other cities include Seattle (32.81 million), Los Angeles (28.92 million), and Dallas (10.88 million). The revenues range from around 10 million to 41 million.

IV. RESULTS & DISCUSSION

This dataset contains information on 402 flights across 6 airlines, focusing on flight durations, passenger counts, and ticket revenues. American Airlines has the highest number of entries, and all flights have a consistent duration of 150 minutes. Passenger counts range from 180 to 250, with ticket revenues averaging $51,800 per flight.

Visualizations show that major cities like Los Angeles, New York, and Chicago are key hubs for airline operations, with significant ticket revenue generated in these locations. United Airlines is the top performer, particularly in Los Angeles, highlighting its dominance in specific markets.

Further analysis using Gantt charts demonstrates the cumulative passenger count across different cities, with Seattle and San Francisco having the highest totals. Additionally, the heatmap provides insights into passenger distribution, showing that certain departure and arrival locations contribute more to overall traffic than others.

This dataset is valuable for stakeholders such as airline operators, revenue managers, and route planners. It helps them optimize schedules, improve pricing strategies, and identify profitable routes. Overall, the data supports data-driven decisions to enhance operational efficiency, financial performance, and customer satisfaction.

V. CONCLUSION

The study reinforces the growing significance of data analytics in transforming the airline industry through enhanced customer experience, operational optimization, and sustainable practices. The use of sentiment analysis and machine learning techniques on social media data helps identify customer pain

points and allows airlines to respond more effectively. Operational inefficiencies such as delays and cancellations can be diagnosed using textual safety reports combined with structured flight data. **Notably, this study demonstrates that integrating customer sentiment data with structured operational datasets enables a 360-degree view of airline performance—bridging the gap between what passengers feel and how airlines operate. Furthermore, the application of Tableau proves essential in transforming complex datasets into intuitive, real-time visual dashboards that support faster, more accurate decision-making for all stakeholders involved.** These insights validate the value of combining multiple data sources and visualization tools to drive continuous improvement in the airline industry

VI. REFERENCE

[1]. Aljedaani, Wajdi, et al. "Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry." Knowledge-Based Systems 255 (2022): 109780.

[2]. Miyamoto, Ayaka, Mayank V. Bendarkar, and Dimitri N. Mavris. "Natural language processing of aviation safety reports to identify inefficient operational patterns." Aerospace 9.8 (2022): 450.

[3]. Kumar, Sachin, and Mikhail Zymbler. "A machine learning approach to analyze customer satisfaction from airline tweets." Journal of Big Data 6.1 (2019): 1-16.

[4]. Tian, Haiman, et al. "Data analytics for air travel data: a survey and new perspectives." ACM Computing Surveys (CSUR) 54.8 (2021): 1-35.

[5]. Yenkikar, Anuradha, and C. Narendra Babu. "AirBERT: A fine-tuned language representation model for airlines tweet sentiment analysis." Intelligent Decision Technologies 17.2 (2023): 435-455.

[6]. Rose, Rodrigo L., Tejas G. Puranik, and Dimitri N. Mavris. "Natural language processing based method for clustering and analysis of aviation safety narratives." Aerospace 7.10 (2020): 143.

[7]. Şapaloğlu, İbrahim. "Text mining on sustainability reports of top 40 airlines and bibliometric analysis of airline's sustainability." Environmental Research and Technology 7.2 (2024): 186-193.

[8]. Walton, Robert O., and Jim W. Marion. "A textual analysis of dangerous goods incidents on aircraft." Transportation research procedia 51 (2020): 152-159.

[9]. Hasib, Khan Md, et al. "A systematic review on airlines industries based on sentiment analysis and topic modeling." (2023).

[10]. Arachchi, Thamindu Gambheera, Mahekha Dahanayaka, and H. Niles Perera. "Analyzing Sustainability Initiatives of the Airline Industry Through Random Forest Classification and K-Means Clustering Techniques." 2024 New Trends in Civil Aviation (NTCA). IEEE, 2024.

[11]. Yousefzadeh Aghdam, Mahdi, et al. "Ontology generation for flight safety messages in air traffic management." Journal of big data 8.1 (2021): 61.

[12]. Sezgen, Eren, Keith J. Mason, and Robert Mayer. "Voice of airline passenger: A text mining approach to understand customer satisfaction." Journal of Air Transport Management 77 (2019): 65-74.

[13]. Clarke, Stephen SB, et al. "Natural Language Understanding and Extraction of Flight Constraints Recorded in Letters of Agreement." AIAA Aviation Forum. 2022.

[14]. Gates, Dominic. "Flawed analysis, failed oversight: How Boeing, FAA certified the suspect 737 MAX flight control system." The Seattle Times 17 (2019).

[15]. Kwon, Hye-Jin, et al. "Topic modeling and sentiment analysis of online review for airlines." Information 12.2 (2021): 78.