|  |
| --- |
| **INFO 584: DATA VISUALIZATION**        **Done by:  Dharshana Rajasekar**  **Thanuja Sekar** |
| ACKNOWLEDGEMENT  We would like to express our utmost gratitude to our professor, **Eyyub Kibis**, academic course assistant **Xinrong Lei** and the Feliciano School of Business at Montclair State University, for allowing us to work on this interesting project on the topic "Airlines Delay". It has helped us to hone our visualization skills as a team and prepared us for the corporate world.      ABSTRACT    Our study focuses on the causes of flight delays in the United States of America. Our analysis intrigued us to ponder  the reasons for the significant delays in multiple airlines. Hence, our project intended to learn the reasons  contributing to the delay of flights by the airline and focussed more towards the top 10 airlines that had reported delays  in US. An interactive dashboard was developed as a result of this objective. The dashboard serves as a filter to  demonstrate the airline delay across the different states of US. Overall, the project has been specific, measurable, achievable, relevant, and time-bound by the smart goals. The result concluded in categorizing the delay caused into one of the two types; man-made or natural. INTRODUCTION **Why Airlines Delay Analysis?**  The next question arises in one’s mind is *“Why Airlines Delay?”* In modern times, the aviation industry plays a vital role in the world's transportation sector and a lot of businesses rely on various airlines to connect them with other parts of the world. But, due to extreme weather conditions, technical/manual problems, carrier problems etc., may directly affect the airline services leading to flight delays. **Project Overview** We analyzed a sample dataset with 318,018 records related to airlines delay to find the root cause for the delay and thus analyze which airport has the most delays. As the phrase quotes *“time is invaluable*,*”* so in order to save more time by studying delay, this analysis was thus performed. **Objective of the Study** To analyze the causes of flight delays in the US airlines across the different states/airports using Tableau visualizations.  **Questions Interested -**   1. What are the reasons for the delays? Is it natural or man-made? 2. Are the delays specific to any airplane or airport? 3. Find the top 10 airplane carriers that face the most delays and compare it with the least delayed carriers.   **SMART Goal:**  The goal is to, within 30 days, identify and mitigate airline delays by a 10% reduction; achieved by interacting with airlines, airport authorities and performing a forecast analysis based on the available historical data, and thus minimizing revenue loss incurred on a monthly basis caused due to delays to the airlines.  **DATA SOURCE:**   * The dataset was obtained from **Kaggle** (online community platform for datasets). * The dataset consists of 21 columns and 3,28,207 rows of records.      * **Data Preprocessing** was performed **using python.** * Percentage of missing values were less than 1%.      * As the of the missing values were < 1%, the records were dropped to improve the efficiency of the analysis. * Using **the tableau prep –**  1. The airport name column was split into separate columns as City, State 2. Makedate() function was used to concatenate the year and month variable and form date.   **Information on Attributes used in the analysis** –   |  |  |  | | --- | --- | --- | | **Variables** | **Type of variable** | **Short description** | | year | Numerical | The attribute mentions the year in which the delay of the flight occurred. | | month | Numerical | Month attribute is the numeric representation of the month in which the flight delay occurred. | | carrier | String | Carrier is an alpha-numeric/numeric variable that represents the carrier id of the airline. | | carrier\_name | String | Carrier Name is an alpha variable that represents the carrier's name of the airline. | | airport | String | Airport variable represents the airport’s code. | | airport\_name | String | The attribute stores the airport’s name. | | airport\_city\_name | String | Name of the city in which the airport is located. | | airport\_state | String | The attribute represents the states code where the airport is located. | | arr\_flights | Numerical | Variable represents the Number of flights arriving at airport. | | arr\_del15 | Numerical | Variable represents the Number of flights arriving at airport more than 15 minutes late. | | carrier\_ct | Numerical | Variable represents the Number of flights delayed due to air carrier. | | weather\_ct | Numerical | Variable represents the Number of flights delay due to weather. | | nas\_ct | Numerical | Variable represents Number of flights delayed due to National Aviation System (e.g. heavy air traffic). | | security\_ct | Numerical | Variable represents Number of flights canceled due to a security breach. | | late\_aircraft\_ct | Numerical | Variable represents Number of flights delayed as a result of another flight on the same aircraft delayed | | arr\_cancelled | Numerical | Variable represents the Number of cancelled flights. | | arr\_diverted | Numerical | Variable represents the Number of flights that were diverted. | | arr\_delay | Numerical | Variable represents the Total time (minutes) of delayed flight. | | carrier\_delay | Numerical | Variable represents the Total time (minutes) of delay due to air carrier | | weather\_delay | Numerical | Variable represents the Total time (minutes) of delay due to inclement weather. | | nas\_delay | Numerical | Variable represents the Total time (minutes) of delay due to National Aviation System. | | security\_delay | Numerical | Variable represents the Total time (minutes) of delay as a result of a security issue. | | late\_aircraft\_delay | Numerical | Variable represents the Total time (minutes) of delay flights as a result of a previous flight on the same airplane being late. |   **DELAY FACTORS:**  The airlines report the causes of delays in five broad categories:   * Air Carrier: The cause of the cancellation or delay was due to circumstances within the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fuelling, etc.). * Extreme Weather: Significant meteorological conditions (actual or forecasted) that, in the judgment of the carrier, delays or prevents the operation of a flight such as tornado, blizzard or hurricane. * National Aviation System (NAS): Delays and cancellations attributable to the national aviation system that refer to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control. * Late-arriving aircraft: A previous flight with same aircraft arrived late, causing the present flight to depart late. * Security: Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.   **ANALYSIS:**   * As the initial part of the analysis, we performed data preprocessing using python to study the quality of the data and drop the redundant records. * The pre-processed data was analyzed in Tableau through different visualizations. * The context filter was used to drill down the visualization and to answer the following questions:   a) What are the most and the least delayed US carriers that flies within the different US states?  b) Which factors contribute to the delays the most?   * Filled map charts in tableau were used to visualize the average delay across the state for the year 2022.ill     The above map differentiates the delays of the states with the help of different diverging colors.   * The most delayed carriers in the country / state is visualized with the help of outer rounded bar chart.     **RESULTS:**   * As a result of the analysis, the percentage of the delay factors is obtained. * The findings of the analysis suggest that 40.03% delay is usually due to the carrier while 37.35 % due to the Aircraft delay ,13.75% due to the Nas delay and 5.33% Weather aircraft delay. We can interpret that most of the delays are man-made in the US.      * Further, we were able to drill down the findings to selected states. The below dashboard shows the details of Texas state.   **DRAWBACKS:**  1)The dataset for many years were found to be missing.  2)We were unable to forecast the results of the year 2023 due the missing values.  Future scope for Research:  Since the most of delays in the US are man-made, research should be performed to find ways to avoid/mitigate the delays. This would increase the Airlines profit.  **REFERENCES:**   * <https://www.kaggle.com/datasets/ryanjt/airline-delay-cause?select=Airline_Delay_Cause.csv> * <https://towardsdatascience.com/a-practical-guide-for-exploratory-data-analysis-flight-delays-f8a713ef7121> |