Data related to flights and various airline carriers holds significant potential for in-depth analysis, uncovering market trends and hidden patterns. This analysis enables companies to understand customer preferences, enhancing their ability to provide better support and services. Additionally, it offers valuable insights into the workings of the airline industry, contributing to improved safety and reliability of air travel. By leveraging this data, airlines can make informed decisions that enhance operational efficiency, customer satisfaction, and overall market competitiveness.

About the dataset:
The data set was available from the US Transport and Statistics Bureau.
https://www.transtats.bts.gov/tables.asp?gnoyr_VQ=FGJ&flf_gnoyr_anzr=g_bagVZR_eRcbegVaT
The dataset includes on time data of flights operated by different airlines. It includes details like:

- Scheduled and actual departure and arrival times
- Canceled and diverted flights
- Time spent taxiing out and taxiing in
- Reasons for delays and cancellations
- Total time spent in the air
- Distance covered for non-stop flights

The visualization is based on three csv files:
1. Airlines.csv - Consists of airline code and corresponding airline names.
2. Airports.csv - Consists of airport id and corresponding airport names.
3. Flights.csv - This is the main dataset containing parameters like arrival time, departure time, number of flights, etc.
These datasets are connected to one another using inner joins. The main dataset only contains the unique airport ID and airline code for each airport and airline respectively. This does not provide geographical information on which airport and airline is being referred. To mitigate this, two separate datasets containing airport and airline names corresponding to their codes are being used. These datasets are being correlated with the main dataset using inner joins on both.

Made use of the ETL pipeline: Extracted data from US Transport and Statistics Bureau website; performed Transformations on data for analysis ; loaded the data into Tableau workspace (data source);

Key Performance Indicators:

1. On time performance : On-time performance measures the proportion of flights that arrive on schedule. With an overall on-time rate of 80.60%, this KPI is crucial for assessing the punctuality of airlines and their adherence to flight schedules. Visualizations like the treemap of airline carriers' delays and the busiest airports by flight numbers highlight the operational efficiency and reliability of airlines and airports, emphasizing those that consistently perform well in terms of timely arrivals.
2. Cancellation Rate - The cancellation rate indicates the percentage of flights that are canceled. With 36,462 canceled flights out of 2,000,000 scheduled flights, this translates

to a cancellation rate of 1.82%. A lower cancellation rate reflects an airline's ability to maintain its schedule and manage unforeseen disruptions. The popularity and activity levels of airlines and airports, as seen in the visualizations of the top 10 popular airlines and busiest airports, provide context for understanding which operators maintain higher operational stability and lower cancellation rates.

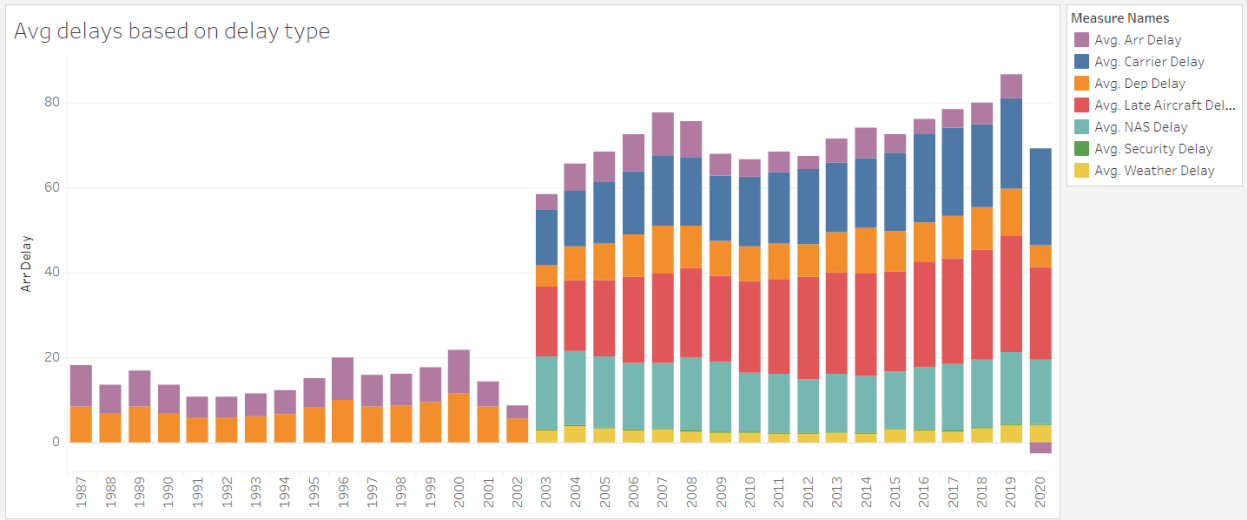
3. Average delay time - Average delay time measures the extent of delays experienced by passengers, providing insight into the performance and efficiency of airlines. The plot showing various types of average delays over the years and the treemap of airline carriers' delays offer a comprehensive view of how delay times have evolved and which carriers experience the most significant delays.
4. Diversion rate - The diversion rate tracks the frequency of flights that are diverted to alternative airports. This KPI is essential for understanding how often airlines and airports encounter severe operational disruptions. The treemap of delays and the visualization of busy airports can help identify patterns and factors contributing to higher diversion rates.
5. Distance covered - Distance covered is useful for analyzing route performance and fuel efficiency. By examining the routes filtered by origin city and carrier airline, this KPI can highlight which airlines operate the longest routes and how efficiently they manage their networks. The visualizations of popular airlines and busiest airports also provide context for understanding overall route coverage and performance.
6. Reasons for delays - Identifying the root causes of delays and cancellations helps in making targeted improvements. The plot illustrating average delays categorized by delay type over the years is instrumental in this analysis. Understanding these reasons can guide airlines and airports in addressing specific issues, improving overall operational efficiency, and reducing the frequency of delays and cancellations.
7. On time arrival percentage - Indicates the reliability and punctuality of the airline and airport operations
8. On time departure percentage - Reflects the efficiency of airport ground operations and airline scheduling

Visualization using Tableau:

1. What is the cause of delays over the years 1987-2020?

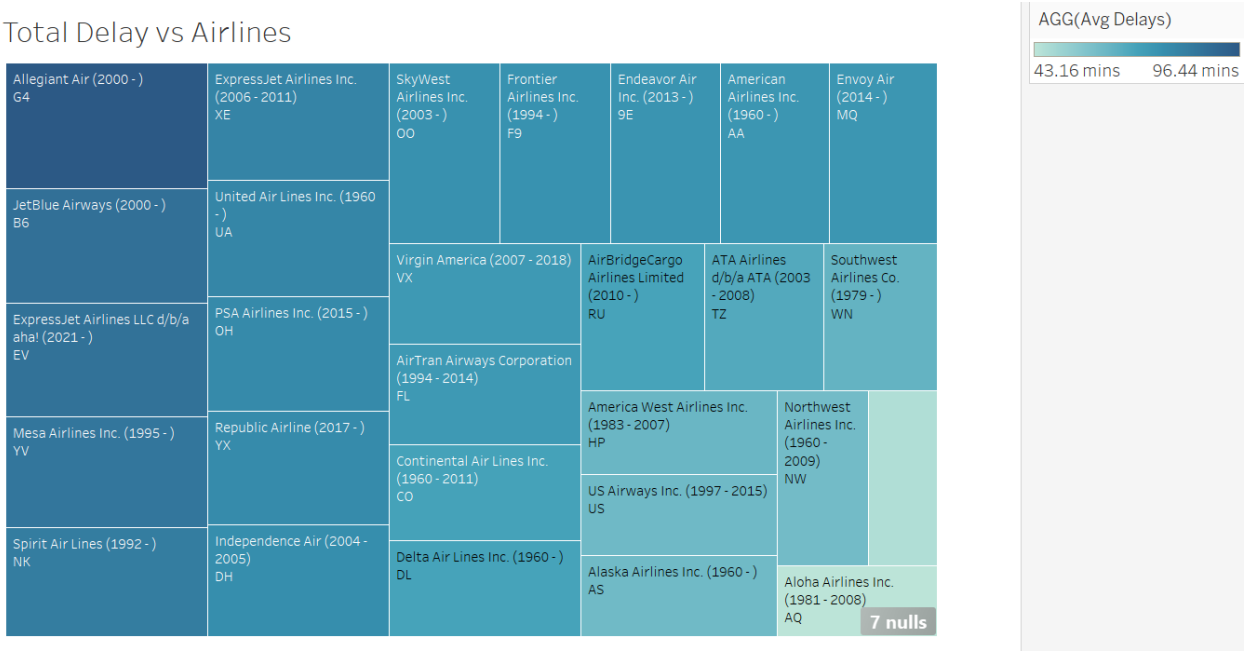
The following plot illustrates various types of average delays for each year from 1987 to 2020. Notably, only arrival and departure delays were recorded in the dataset prior to 2003. The plot

also highlights the average delays categorized by delay type for each given year.



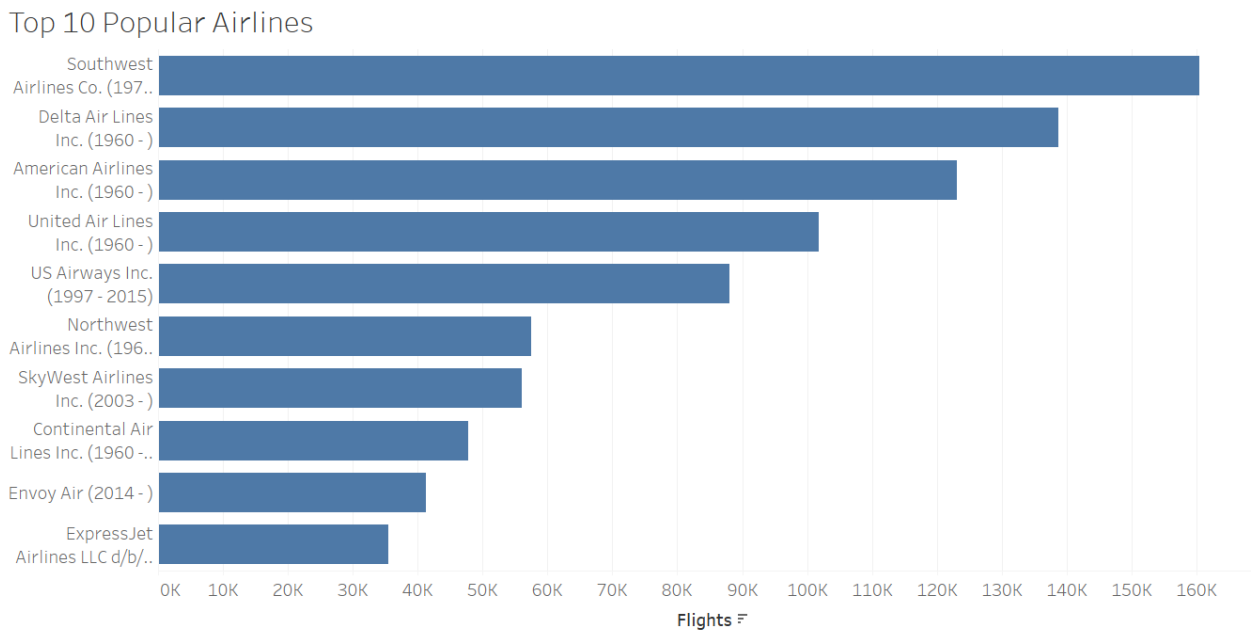
2. Which airline carriers have the most average delay?

This treemap displays airline carriers based on the frequency of their delays(in minutes). The color coding ranges from dark blue, representing the longest average delays in minutes, to light blue, indicating the shortest delays.



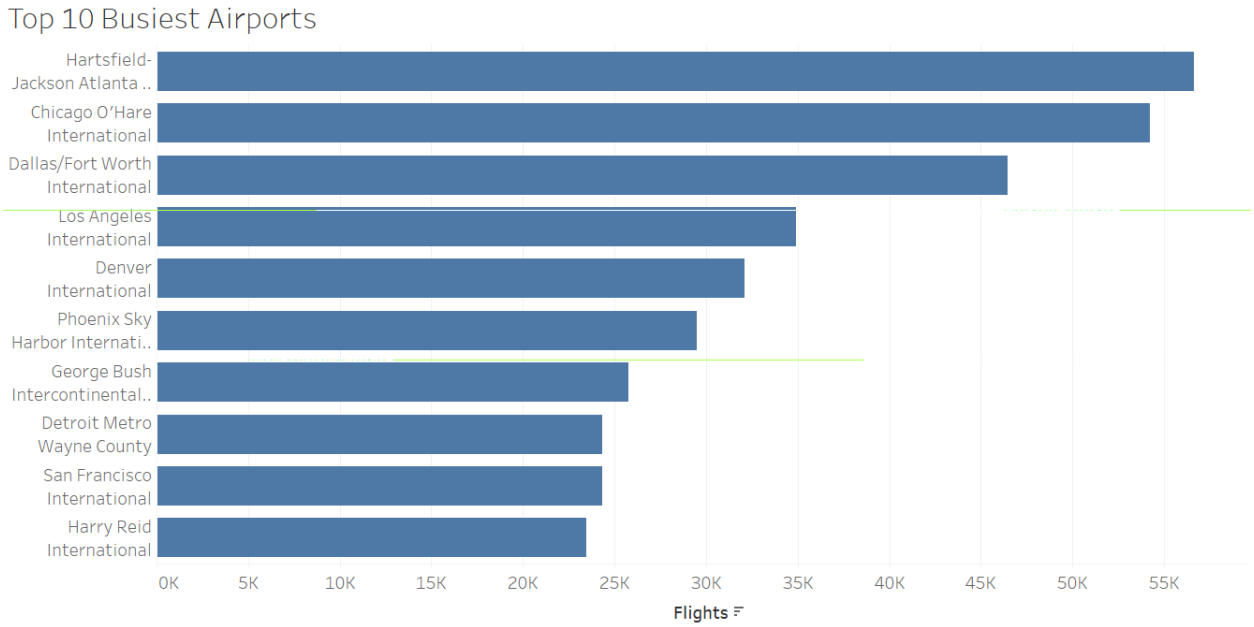
3. Top 10 popular airlines by no of flights

The popularity of airlines can be ranked based on the total number of flights they operate, considering all origin and destination cities. This ranking reflects the airline's operational scale and market presence, indicating which airlines offer the most extensive flight networks and services across various routes.



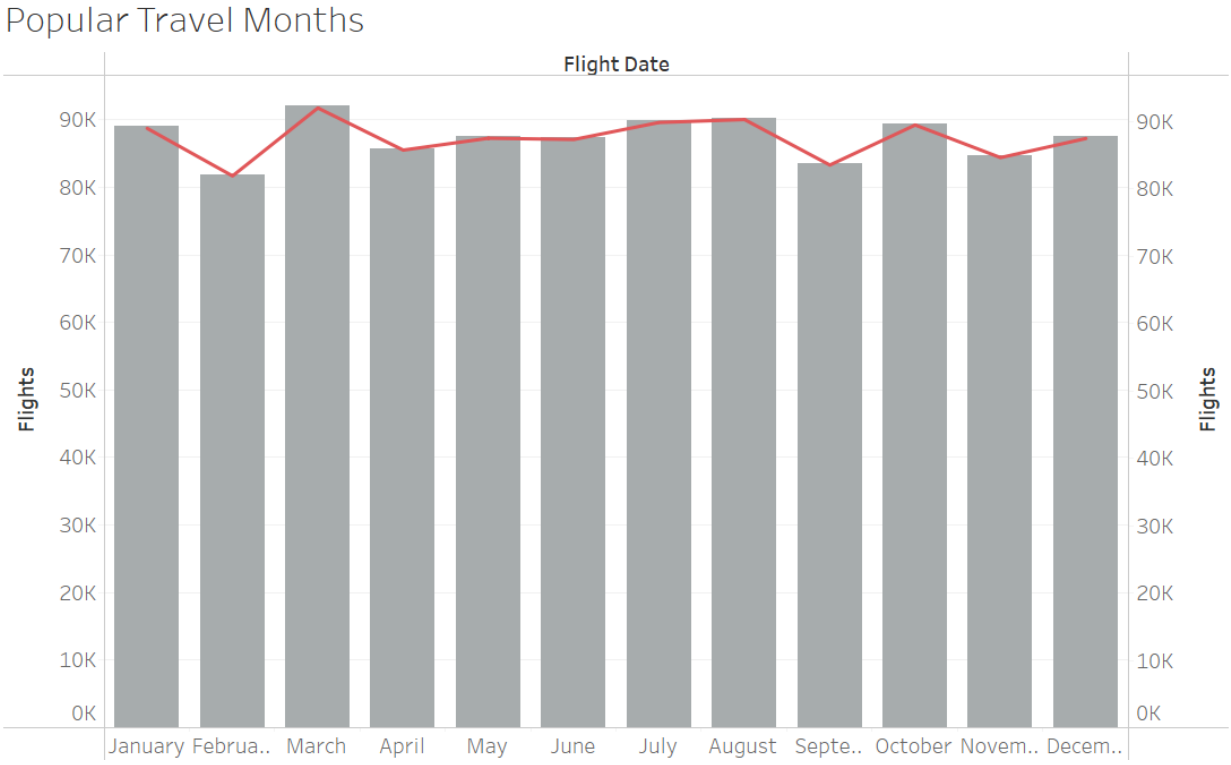
4. Top 10 Busiest Airports by no of flights

The busiest airports can be ranked based on the total number of flights they handle, considering all origin and destination cities. This ranking highlights the airport's activity levels and their significance as major hubs. A higher number of flights indicates greater connectivity and passenger traffic, showcasing the airport's role in facilitating extensive travel networks.



5. Most popular travel months

The number of flights taken each month over all years indicates customer travel preferences. A higher number of passengers suggests a greater preference for traveling in that month. It is observed that the number of flights per month remains relatively consistent.



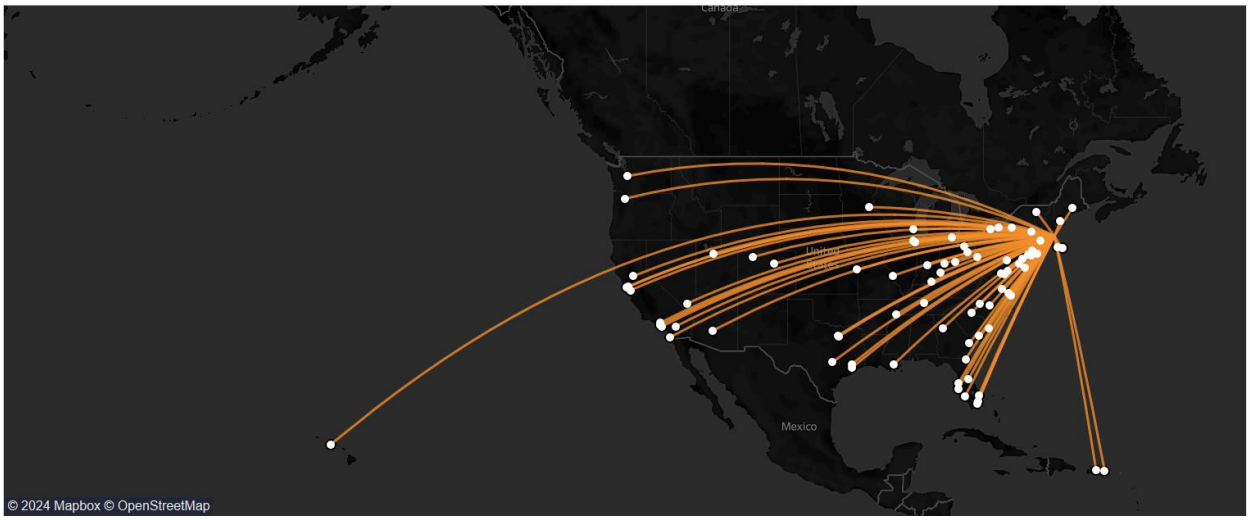
6. Routes filtered by origin city and carrier airline

This dashboard illustrates the routes from a selected origin city to all destinations in the dataset, offering a visual representation of the destination locations. The top filters enable users to refine the routes by specific airline carriers and origin cities.

Route mapping from Source city to Dest city

Choose Airline | Airline: (All)

Choose Origin City | Origin City Name: Boston, MA



7. Month-wise Statistics for all years

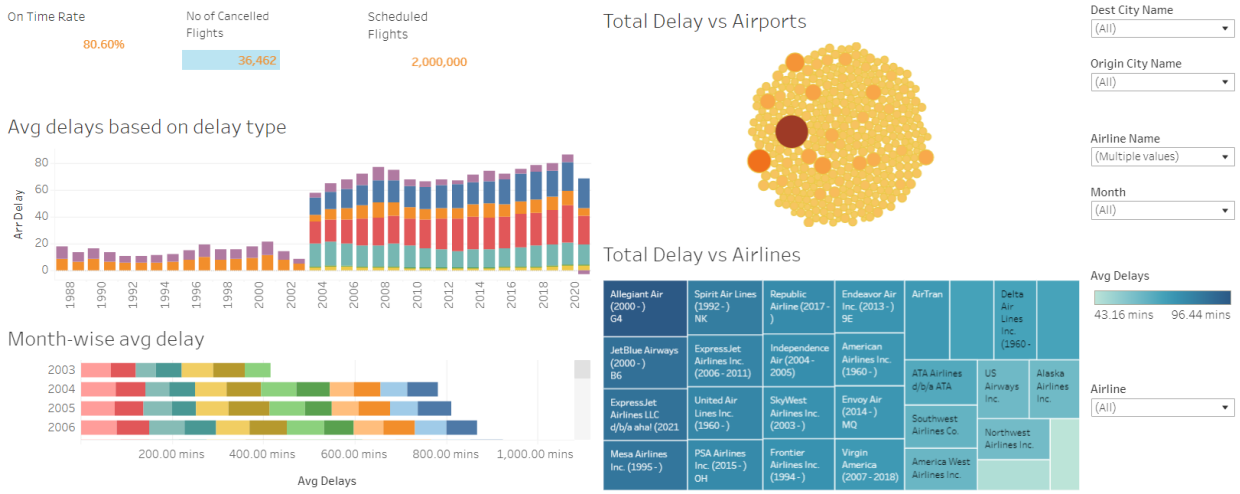
The following table showcases several statistics for airline performance such as canceled flights, total number of flights and diverted flights.

Month-wise Analysis

Month of Flight Date	% of Total Cancelled along Table (Down)	Cancelled	% of Total Diverted along Table (Down)	Diverted	Flights
January	13.77%	5,021	9.87%	453	169,561
February	11.46%	4,178	7.69%	353	155,629
March	12.00%	4,375	8.61%	395	175,296
April	5.86%	2,138	7.12%	327	163,016
May	5.78%	2,106	7.80%	358	166,763
June	7.25%	2,644	10.48%	481	166,203
July	7.50%	2,733	10.83%	497	171,679
August	7.33%	2,674	9.30%	427	171,990
September	9.22%	3,361	6.84%	314	160,375
October	5.36%	1,956	5.88%	270	170,617
November	4.95%	1,804	6.51%	299	161,952
December	9.52%	3,472	9.06%	416	166,919

Overall findings from the data:

This is one of the dashboards generated to provide an overview of important metrics.



Future Scope:

With more data related to flight model, fueling information, passenger information, etc we may be able to uncover more details such as airplane performance.