Introduction

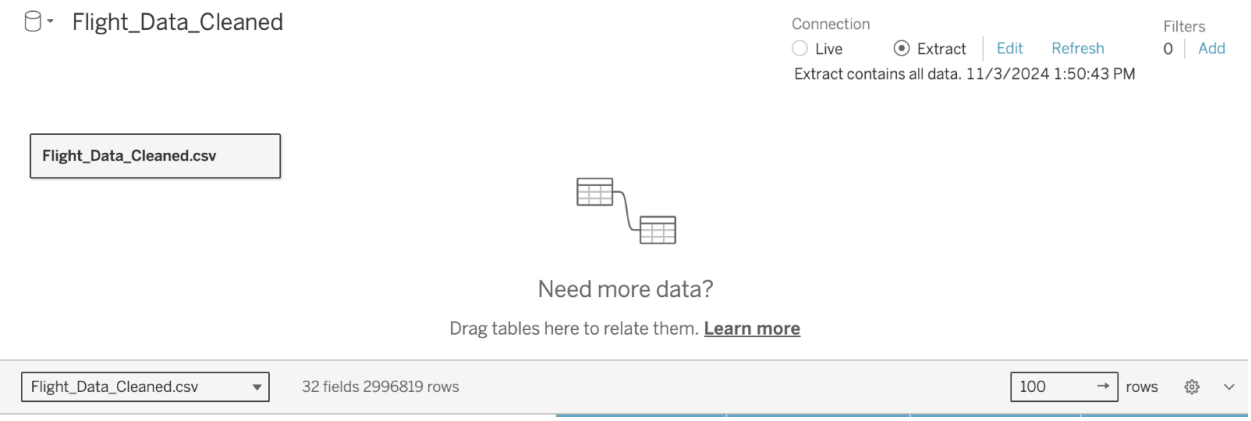
Over the past couple of years, data storytelling has proved to be a crucial tool for decision-making based on data across industries, in operationally intensive sectors such as airline management. This assignment requires the use of Tableau in developing a thorough dashboard, including flight delays, cancellations, and diversions by airlines in the U.S. from 2019 to 2023. The dataset used is obtained from the U.S. Department of Transportation, provides a highly informative insight into airlines' performance measures, where one can inculcate flight delays, reasons for cancellation, and lastly, diversions. As indicated, the three important intents of this assignment are as follows:

1. Improve advanced data storytelling by structuring insights into a form that best resonates with the end users.
2. Explore and apply the more advanced Tableau skills for data visualization, like custom calculations, filters, and geo-spatial mapping.
3. Design a data-driven dashboard with an insight-driven interface to enable airline managers and operational analysts to make better decisions to optimize schedules, reduce delays, and hence improve the overall customer experience.

These objectives, in turn, have organized the project into three major tasks, each meeting different learning outcomes and assessment criteria. Task 1 focuses on the creation of an engaging data narrative for intended audiences; Task 2 embraces the rich exploration of data, putting into practice advanced functionality in Tableau; and Task 3 covers building an interactive dashboard that would allow real-time, data-driven decision-making.

Data Connection and Background

The "Flight Delay and Cancellation Dataset" was preprocessed using Python in Google Colab. This preprocessing consisted of handling missing values, formatting columns, and making the data consistent throughout for the readiness of analysis. After cleaning the data, it was then imported to Tableau as an extract for fast refresh and smooth performance which permits efficient real-time exploration with flexible filtering across the data.



This dataset covers the years from **2019 through 2023** and contains the following key fields:

- **Flight Date:** It contains the exact date on which the flight took place for time-series analysis over different years.
- **Airline:** The carrier that operated each flight; this would allow cross-comparison of airline performance.
- **Delay Metrics:** It consists of arrival and departure delays that could give sufficient evidence of punctuality and bottlenecks.
- **Cancellation and Diversion Indicators:** It records whether the flight is cancelled or diverted, and it has associated metrics that determine reliability.
- **Origin and Destination City:** This field captures the origin and destination cities of each flight, providing valuable geographic information.

Primary Objectives and Analytical Techniques

This assignment extends Tableau competencies through hands-on practice in the following exercises:

- **Calculations and Parameters:** This work creates custom fields for the average delays, percentage of cancellations, and diversions by airline to accurately benchmark.
- **Advanced Visualizations:** Stacked bar charts, geo-spatial maps, and heat maps further tell the story with data and intuitively communicate complex information.
- **Interactivity:** It is ensured by filters, parameters, and drill-downs that make the dashboards dynamic, which then allows users to interact with the dashboard to drill down on specific metrics relevant to them.

A structured, insightful, and interactive dashboard that captures both operational patterns and areas for potential improvement will finally lead the stakeholders to make data-driven and impactful decisions.

Task 1: Elevating Data Storytelling Skills (LO1)

Introduction to Task 1

Task 1 identifies effective data storytelling for operational insights. Consequently, this would be a visualization of the dataset in the most appealing, accessible, and understandable way possible for managers and business analysts concerned with airlines. Each one of these stakeholders has his or her needs related to the visualizations on delay trends, cancellation rates, and diversions since each of them focuses on airline efficiency.

Defining the Target Audience

This dashboard's target audience would be **airline operation managers** and **business analysts** who promote efficiency while minimizing delays. The visualizations have been customized to show operational pain points, such as high-delay periods, frequently affected airlines, and patterns that show avenues for optimization. These stakeholders' priorities, in turn, then informed the selection of fields and the design approach desiring clear, actionable insights with respect to variables such as delay duration and cancellation metrics.

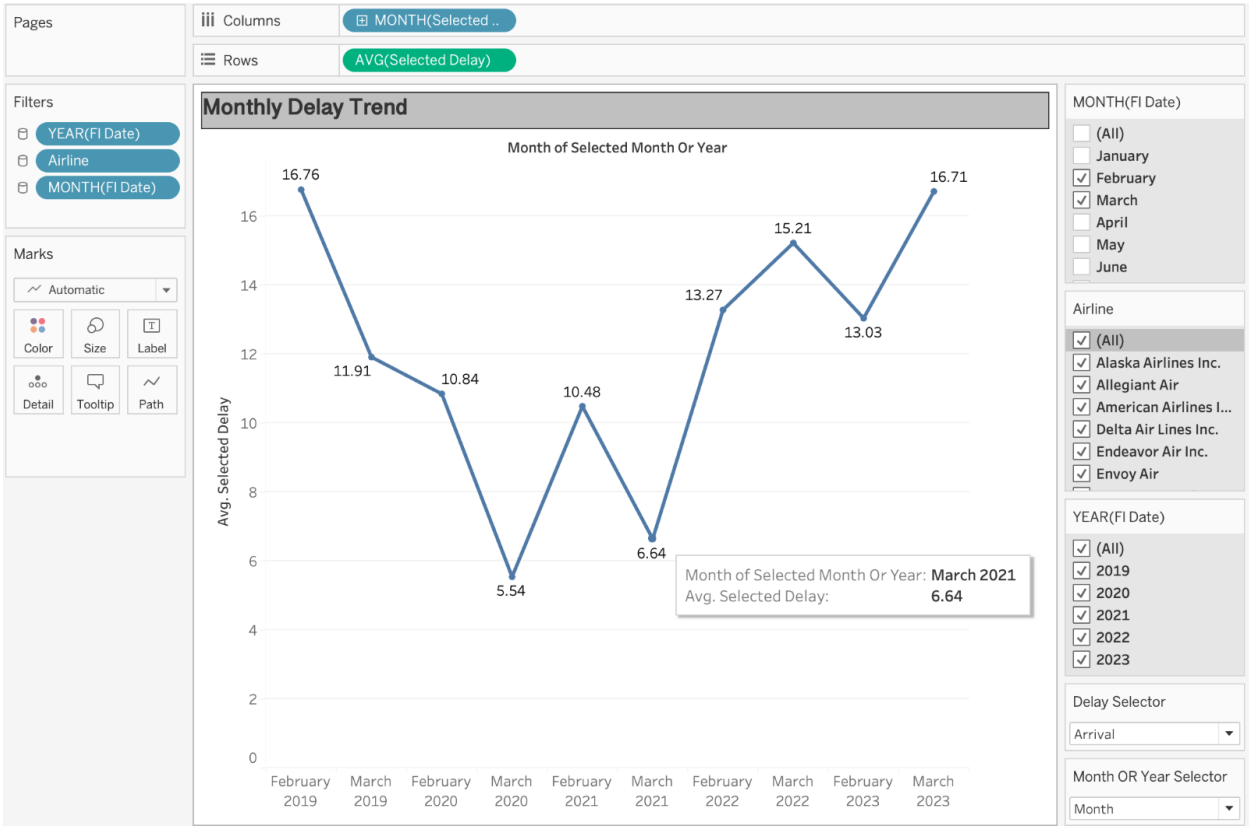
Crafting an Engaging Introduction

Framing it initially as a historical record of flight delays and cancellations over a five-year period sets a certain context for the data. The introduction underlines how this would help reach useful understanding regarding customer satisfaction, operational efficiency, and cost management. Setting such a framework helps establish the data as part of some big story of trends within industries; thus, the relevance of material presented on this dashboard becomes immediately apparent to stakeholders.

Creating Visualizations for Key Insights

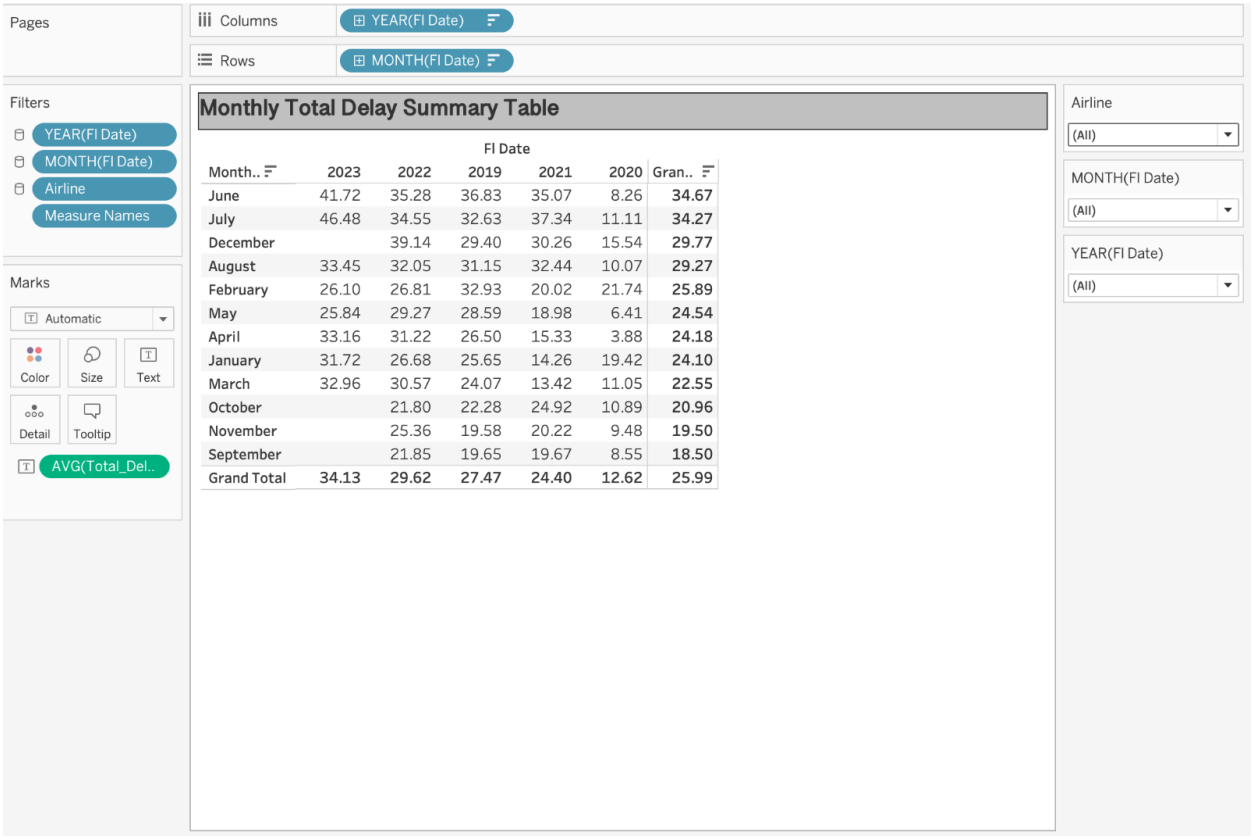
To create a structured narrative, the following visualizations were developed:

1. **Monthly Delay Trend**:



   a. **Description:** This is a line chart showing month-over-month trends in average arrival and departure delays. It had a parameter that would toggle the chart between arrival and departure views, depending on one's analytical needs.
   b. **Insights:** Peaks in delays are seen in **July** and **December**, obviously due to seasonal spikes in travel demand and due to weather-related disruptions. A notable decline in 2020 coincides with reduced travel during the pandemic.
   c. **Implications:** This suggests that airlines should staff and schedule people to minimize delays during peak travel months.

2. **Total Delay Summary**:



a. **Description:** The table above summarizes the monthly average-of-year delays and aggregates to show the total delays throughout the year.
b. **Insights: June, July,** and **December** have been consistent with recording the highest average delays, while **September** and **October** record the least delays.
c. **Implications:** Airlines could use this to plan and make necessary preparations during the holiday and summer periods that record the highest delays.

3. **Airline Delay Summary**:

a. **Description:** The following table presents the average delay comparisons of each airline; it includes average selected delay, and average total delay.
b. **Insights:** On-time performance exhibits the poorest results for JetBlue Airways, Frontier Airlines, and Allegiant Air. This is because these airlines have some of the longest average delay times; this could be due to operational

challenges such as inefficient scheduling or high-volume traffic at hub airports. Conversely, some of the airlines with lower average delays include Horizon Air and Hawaiian Airlines due to good operational practices or favorable route conditions.

c. **Implications:** The airlines with higher delays, such as JetBlue and Frontier, would do well to analyze their operational inefficiencies and take steps toward improvement. Those with the lowest consistent delays, Horizon and Hawaiian Airlines, might provide the industry benchmarks regarding best practices in the management of delays and efficient scheduling.

Interpretation and Analysis of Data Patterns

Data indicate the following key trends, among others:

- **Seasonal Peaks:** Higher delays during July and December are months that have more travelers, hence more resources.
- **Pandemic Impact:** The decrease in delay minutes in 2020 offers a fair comparison between normal operations and extraordinary circumstances.
- **Airline-specific Trends:** Consistently high delays from the same airlines, such as JetBlue, need operational adjustments.

Compelling Conclusion for Effective Communication

The above visualizations show that seasonal planning and benchmarking play a significant role in this airline industry. For instance, carriers experiencing elevated levels of sustained delay can take certain measures like **improved scheduling** and **resource rearrangement** during peak periods. It can also deploy best practices of low-delay airlines to high-delay airlines to achieve operational efficiencies.

Conclusion of Task 1

Task 1 set the foundation of a storyline for transforming raw data into something meaningful to an airline operator and analysts. By focusing on critical metrics and presenting them in a clear and visual format, the dashboard enables informed decisions.

Task 2: Mastery of Data Exploration and Visualization (LO2)

Introduction to Task 2

Task 2 will center on deepening the insights derivable from the data set using more advanced features of Tableau. In this project, use of custom calculations, advanced chart types, and geo-spatial mapping to communicate a richer understanding of airline performance data will be used. The main goals are to enhance their data exploration competency and build a set of interactive tools stakeholders can use for more focused analysis.

Selection of Complex Dataset

The following dataset covers performance flight records between 2019 and 2023, including delays, cancellations, and diversions of U.S. airlines. The provided tableau extract is only one rich dataset in terms of critical metrics with respect to the following: **departure delays, arrival delays, cancellations, diversions**, and various other operational details that might be leveraged for exploring granular levels of trends, airline comparisons, and delay patterns.

Advanced Tableau Techniques for Visualization

Task 2 focuses on enhanced visualization techniques, achieved by applying calculated fields, parameters, and interactive filters. While the user is central to the process of creating a dashboard, these techniques support stakeholders by facilitating multiple data view switching and concentrating on metrics relevant to their goals.

Parameters

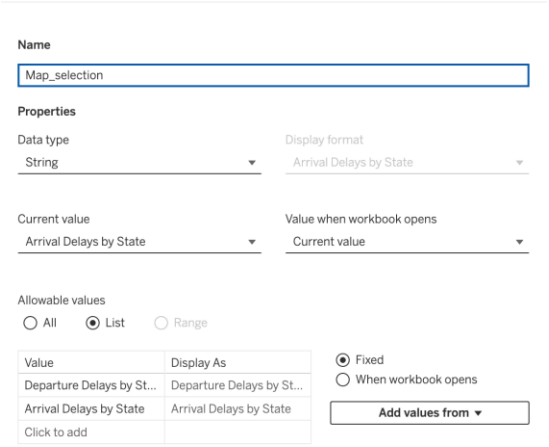
1. **Delay Selector**:

   It enables a parameter that toggles users between the "Arrival" and "Departure" delay views within the Monthly Delay Trend chart, making it much easier for users to focus on certain aspects of the delay performance that best suit the analytical requirement.

2. **Month or Year Selector**:

   This parameter will enable the toggling of monthly and yearly delay analysis by the user for flexible explorations of trends in these different periods.

3. **Parameters for View Switching**
   Other parameters are placed to allow the switching of views for arrival versus departure delays, which allows users to toggle between the two metrics within one interactive visualization.

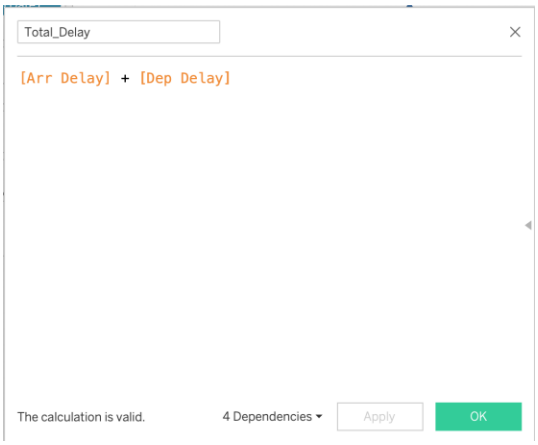


Custom Calculations

To enhance the analysis, custom calculations that provided deeper insights into flight performance are created:

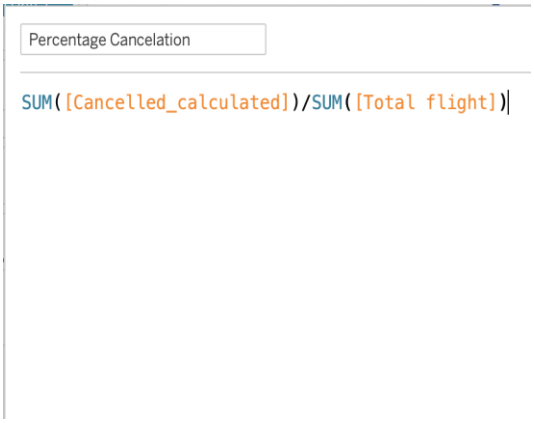
1. **Total Delay Calculation**:

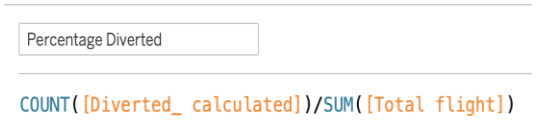
   This is calculated by adding average arrival and departure delays. This field was used to capture the holistic view of delay impacts, which was important in creating the Total Delay Summary Table and Airline Delay Summary chart.



2. **Cancellation and Diversion Rates**:

   These fields provide the cancellation and diversion percentages by dividing the number of cancellations or diversions by the total flights for each airline. The output in these fields can provide the likelihood of cancellation/diversion and allow for effective comparison across airlines.





3. **Cancelled Calculated Field**
   a. **Formula:** IF [Cancelled] = TRUE THEN 1 ELSE NULL END
   b. **Purpose:** This field assigns 1 to canceled flights and NULL to others. By summing this field and dividing it by the total flights per airline, we obtain the **cancellation percentage**, which offers a quick view of each airline's reliability.

Cancelled_calculated

```
IF [Cancelled] = TRUE THEN 1
ELSE NULL
END
```

4. **Diverted Calculated Field**
    a. **Formula:** IF [Diverted] = TRUE THEN 1 ELSE NULL END
    b. **Purpose:** This field assigns diverted flights with 1 and others as NULL. Summing these values and dividing it by total flights provides the **diversion percentage**, highlighting airlines with higher rates of in-flight diversions.

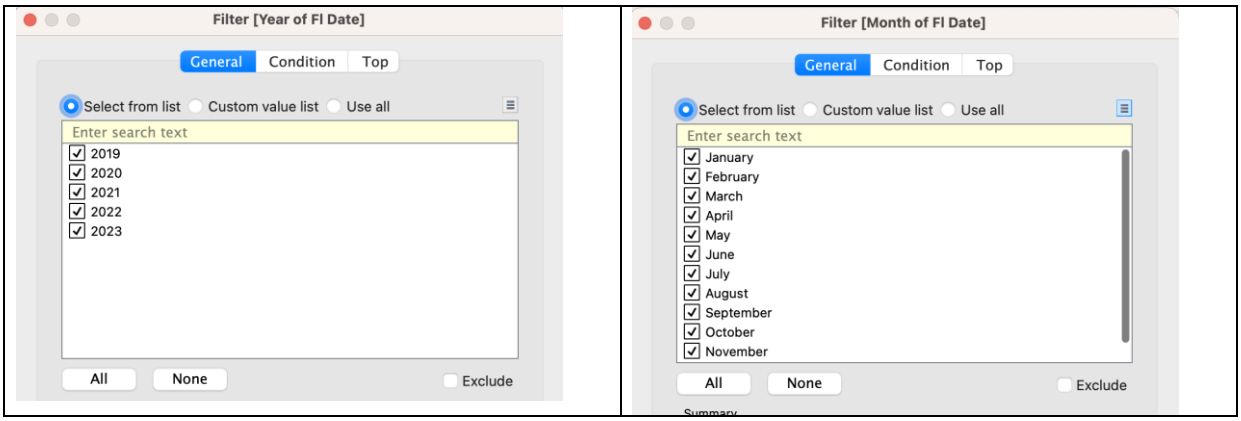
Diverted_ calculated

```
IF [Diverted] = TRUE THEN 1
ELSE NULL
END
```

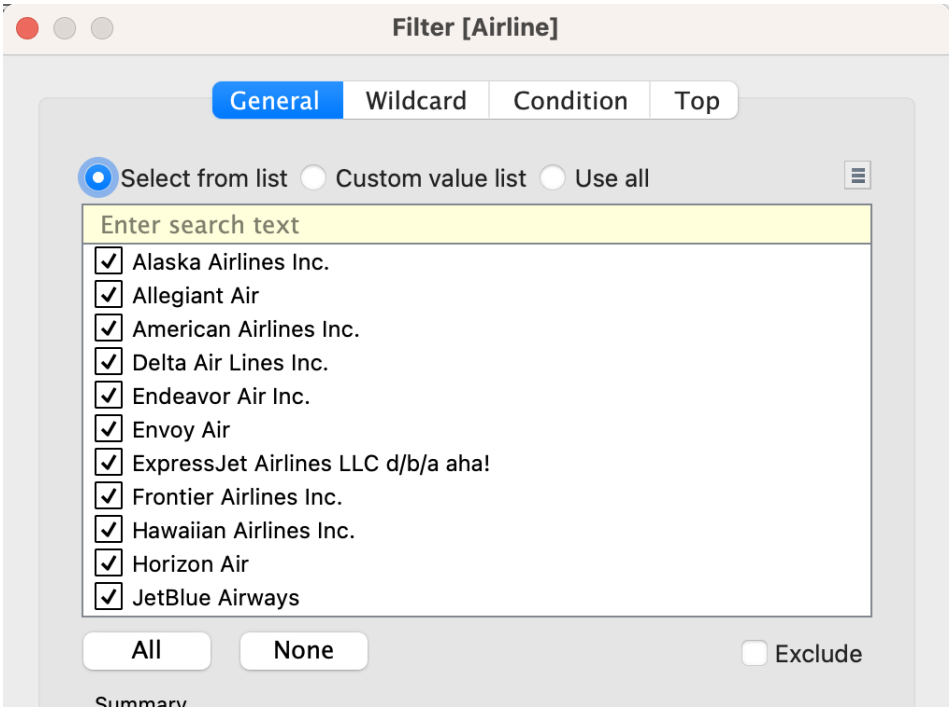
Filters

1. **Date Filters**

    YEAR(FI Date) and MONTH(FI Date) are used to filter the data for specific years and months of the year. These filters allow end users to focus on areas and time periods that matter most and see how trends change over time.



2. **Airline Filter**
   This filter allows the user to isolate data for a specific airline and, therefore, view each individual carrier's performance in greater detail within the chart of Airline Delay Summary.



Utilizing Advanced Chart Types

More complex chart types were devised to represent complex insights intuitively.

1. **Heat Maps for Cancellations and Diversions**

The *Airline Cancellations and Diversions Heatmap* was constructed to illustrate how cancellation/diversion counts for the reporting period concentrated on a given airline. The color gradient emphasizes which airlines have the highest rates so users can immediately tell which carriers are most prone to operational disruption.

Airlines like ExpressJet and Allegiant display higher cancellation rates, either due to operational or route-based issues.

2. **Stacked Bar Chart for Delays and Diversions**:

   Another stacked bar chart was used to present both the cancellations and diversions per airline in one view, enabling their comparison side by side.

   This chart shows which airlines are having more difficulties with cancellations or diversions, helping with resource support and operational planning.

Geo-spatial Analysis

Using Tableau's geo-spatial capabilities, an interactive map of average arrival delays by state has been created. Spatial analytics here show which regions the delays are most concentrated, giving a geographic dimension to some of the operational challenges.

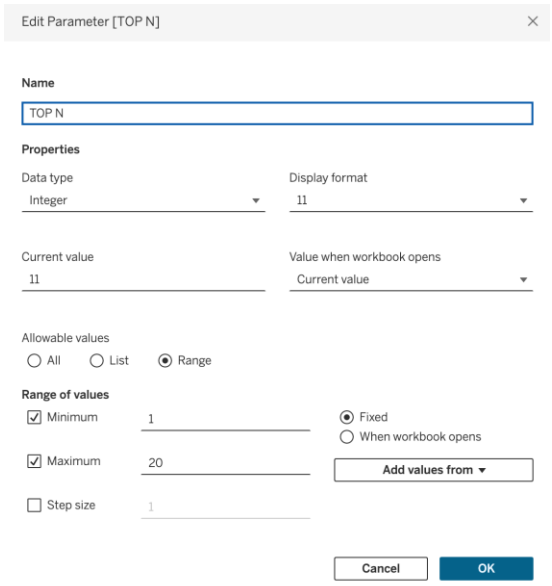
1. **Arrival Delay by State Map**:

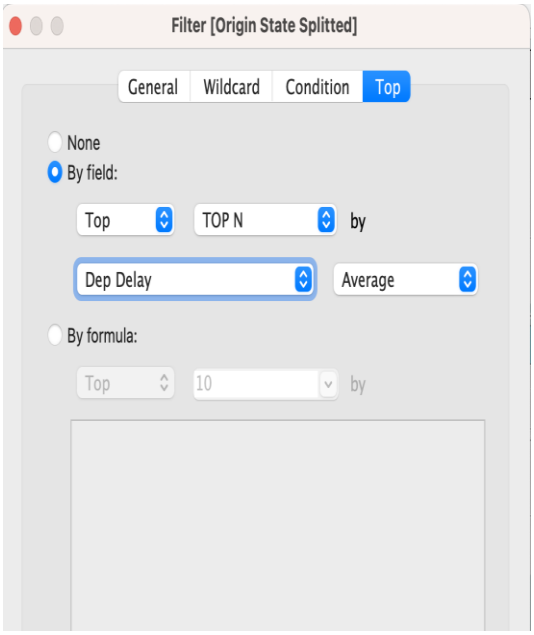
   Allows for the highlighting of the states that have high average arrival/departure delays so that stakeholders can determine delay-prone areas.

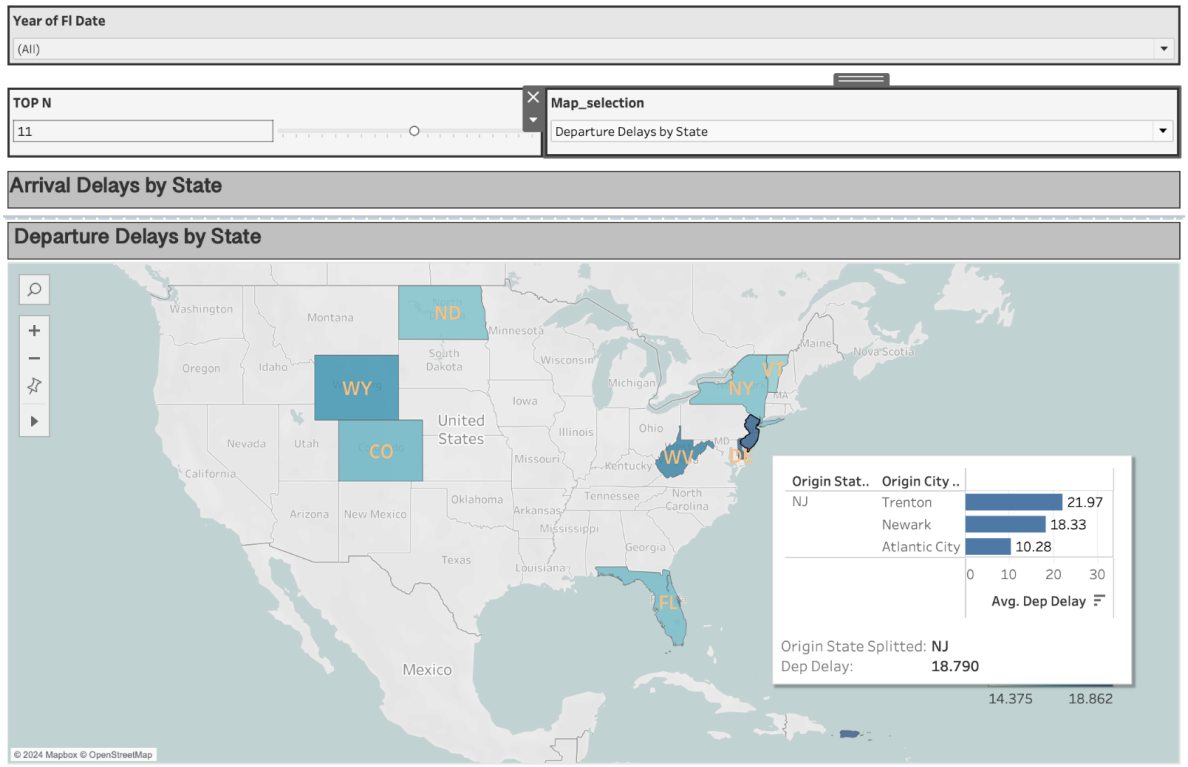
   States like New Jersey and New York always tend to have higher delays; this could be due to the congestion in big airports like Newark and JFK.

2. **Map Features and Interactivity**:

a. **Top N Filter:** Slider that sets a limit on the view to show only the top N states with the most delay, hence making it easy to spot major delay regions.
b. **Detailed Tooltips:** Displays the delay data for each major airport within a given state - hence, gives further granularity of city-level delays.







Dashboard Interactions for Exploration

To enhance user engagement, the following interactive elements were added to the dashboard:

1. **Toggle Between Arrival and Departure Views**
   Users can toggle between arrival and departure delays on the *Monthly Delay Trend*, so they can focus on the type of delay most relevant for their analysis.

2. **Switch Between Heatmap and Stacked Bar Chart**:
   The cancellation and diversion analysis allows the user to toggle between the heatmap and the stacked bar chart to provide two alternate perspectives of disruption rates.

3. **Filters and Top N Selection:**

   The "Top N" slider allows users to filter the data so that the view only includes the states with the greatest delays; it is easy to identify what areas are most prone to delays.

Interpretation of Visualized Insights

Advanced insight from these visualizations was derived as follows:

- **Seasonal Delay Patterns:** The Monthly Delay Trend chart depicted seasonal peaks in delay, such as in December and July, which are months characterized by heavy travel due to holidays or summer congestion.
- **High Cancellation and Diversion Rates for Specific Airlines:** Airlines with significant cancellation and diversion rates were identified via the use of a heatmap and stacked bar chart, including ExpressJet and Allegiant, indicating possible operational issues.

- **Regional Delay Hotspots:** The geo-spatial analysis identified states such as New Jersey and New York as hotspots for delays, indicative of probable congestion or airport-specific issues.

Conclusion of Task 2

Task 2 demonstrated the application of advanced Tableau techniques for deep exploration and visualization. Integrating custom calculations, complex chart types, and geo-spatial mapping to let the dashboard represent comprehensive and interactive views of data that facilitate users in understanding operational trends, thus supporting optimization of performance, and making strategic decisions.

Task 3: Crafting an Advanced Data-Driven Report (LO3)

Introduction to Task 3

Task 3 is focusing on Designing an improved data-driven dashboard to cater to the operational needs of managers and analysts in airlines. Design a dashboard that will show, through clear and actionable insights, delay trends, cancellation rates, and the performance of airlines to decision-makers.

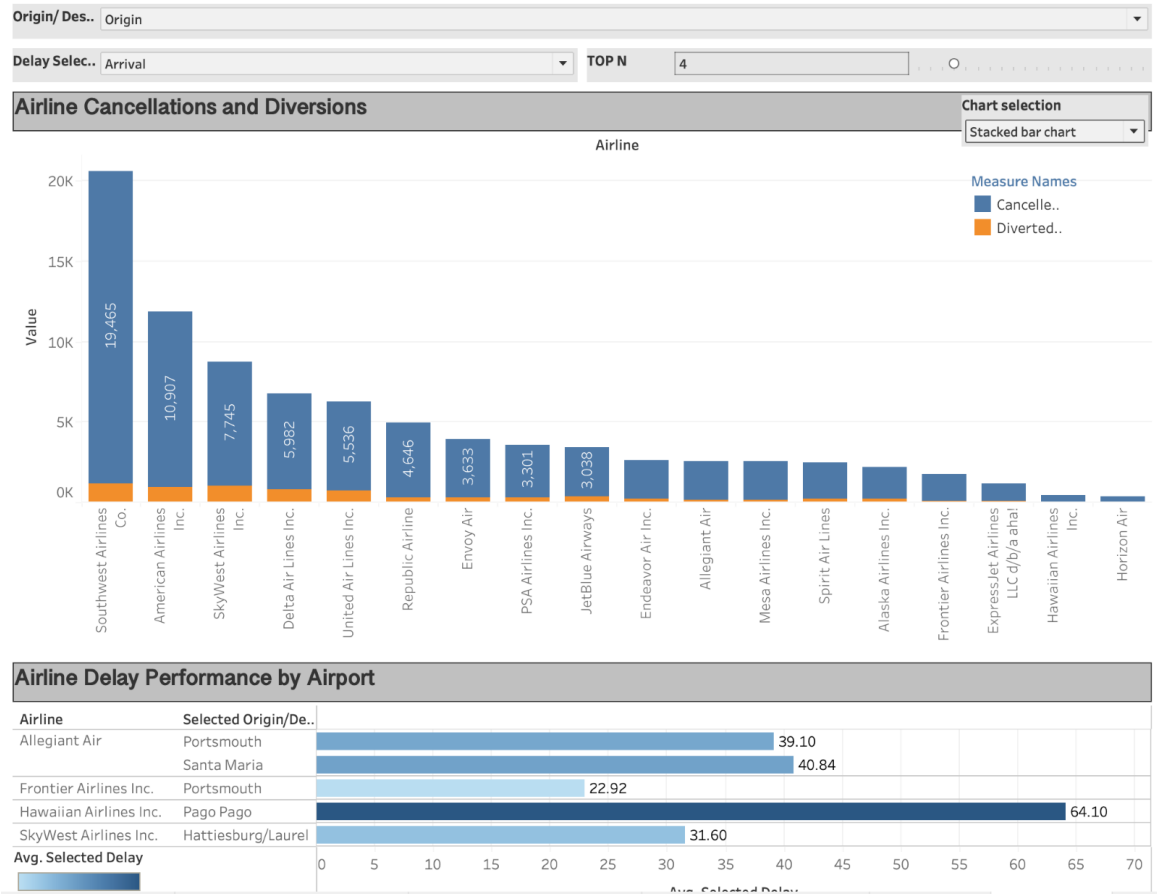
Identification of Business or Analytical Use Case

Key use cases for this dashboard include performance analytics for airlines. The main aim is to minimize delays, reduce flight cancellations, and increase efficiency within operations. This will apply in work processes for operations managers and analysts for airlines tracking metrics on performance and acting upon insights to improve service and punctuality.

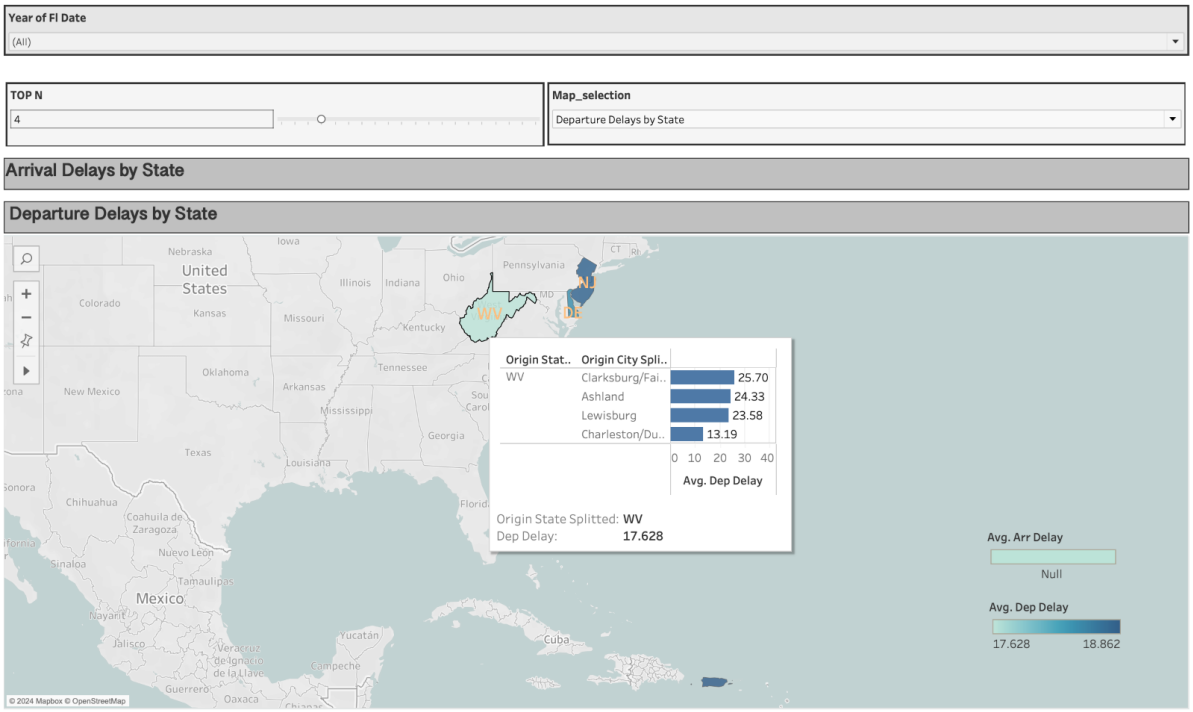
Designing an Advanced Tableau Report or Dashboard

The dashboard was designed with many advanced elements to make it suitable for both an analytical and practical purpose. Key design choices include:

- **Switchable Views for Cancellations and Delays**: It also includes switchable views, switching can be highlighted with the following visualization: Airline Cancellations and Diversions Heatmap and Stacked Bar Chart, which reveals comparable insights regarding airlines' performance in dealing with disruptions.



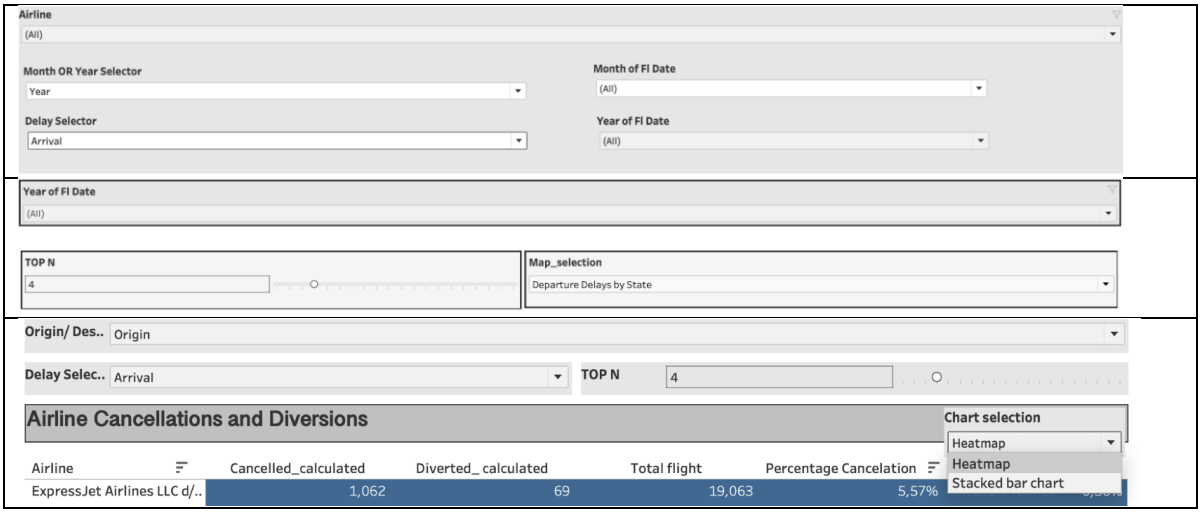
- **Interactive Geo-Spatial Map**: Using Tableau's geo-spatial capabilities, delays upon arrival and departure by state are visualized on a map in an interactive way to show the regional concentration of delays. For example, the view titled "Departure Delays by State" enables the operations manager to easily see which high-delay regions could receive higher priority in making improvements. Tooltips have been added to the map so that users can get more information at the city level for drilling into specific airports within each state.



- **Custom Calculations for Detailed Metrics**: It has calculated fields with custom code to carry out calculations on the dashboard for things like cancellation and diversion percentages. This will enable the stakeholder to visualize and compare different airlines according to their reliability caused by cancellations and diversions in performance benchmarking shown in the Airline Cancellations and Diversions Heatmap.

Cancelled_calculated	Diverted_ calculated	Percentage Cancelation	Percentage Diverted
1,062	69	5,57%	0,36%
2,383	175	4,52%	0,33%
2,373	162	3,65%	0,25%
19,465	1,135	3,38%	0,20%
4,646	313	3,25%	0,22%
3,301	265	3,09%	0,25%
3,633	311	3,00%	0,26%
10,907	980	2,85%	0,26%
3,038	356	2,70%	0,32%
1,666	88	2,59%	0,14%
2,312	198	2,42%	0,21%
7,745	1,006	2,25%	0,29%
5,536	698	2,18%	0,27%
2,394	220	2,13%	0,20%
1,934	239	1,93%	0,24%
374	23	1,82%	0,11%
5,982	782	1,52%	0,20%
388	28	1,21%	0,09%

- **Filters and Parameters for Enhanced Exploration**: Filters and parameters for detailed analysis include the following:
    - **Year and Month:** These filters enable focusing into seasons or year-over-year analysis.
    - **Delay Selector Parameter:** A way for the user to toggle between arrival and departure delays in the Monthly Delay Trend chart, for flexibility in the type of delay viewed.
    - **Top N Filter:** This will enable users to focus on the top delay-prone states so that it can be easier for them to identify and address the most critical areas.



These interactive features support a personalized experience for the users in ways that let each user steer through and analyze the data based on unique priorities of the stakeholder.

Facilitating Data-Driven Decisions

This interactive dashboard is designed with different tools that enhance the ability of various stakeholders to make informed decisions through data. The identification of trends related to delays allows stakeholders to optimize resources in those peak travel periods. Comparison of cancellation and diversion rates across various airlines informs decisions related to scheduling, staffing, and operational planning.

Conclusion of Task 3

This forms the user-centered and data-driven report that Task 3 will culminate into-to support airline managers and analysts through informed decisions. In fact, by designing the dashboard in line with business goals, a report could help improve operating efficiency and customer satisfaction by actionable insight.

Conclusion

Summary of Key Achievements and Learnings

In this assignment, I developed my skills in data storytelling and furthered advanced functionalities within Tableau and prepared a user-centered dashboard. The key highlights of this assignment are the effective use of geo-spatial mapping to locate areas prone to delays and custom calculations for minute metrics, such as cancellation percentages.

Reflection on Advancements in Data Storytelling

This project highlighted the importance of structuring the data narrative to suit audience needs using parameters, filters, and insights in an engaging visual manner that showed ways of turning complex insights into accessible ones to facilitate strategic decision-making.

Implications for Effective Data Communication

The abilities acquired underscore how data visualization helps connect raw data to useful insights. Clear communication through visual storytelling plays a key role to enable smart and meaningful choices in any business that relies on data.

Google Colab Link:
https://colab.research.google.com/drive/1PQTNpHvj9KL5kqrg9m8DrsHJR0-noWEo?usp=sharing