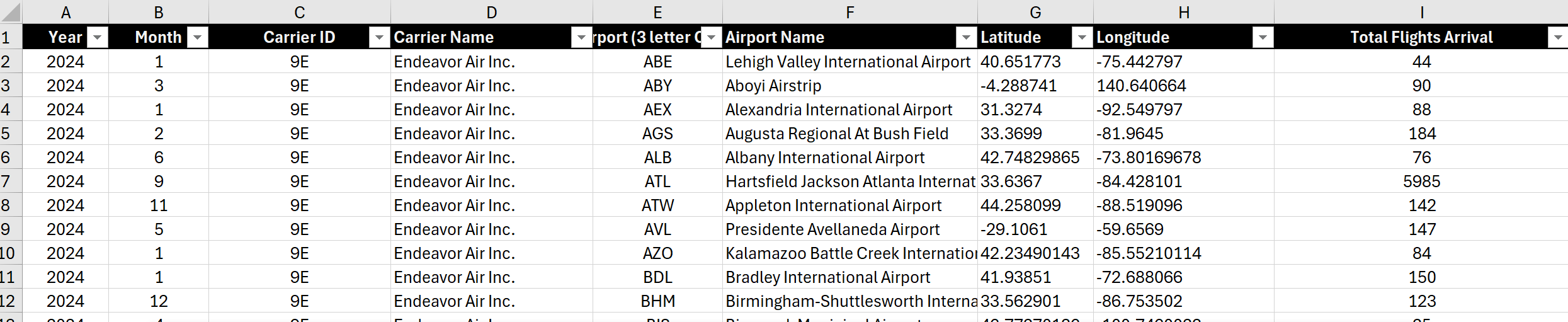
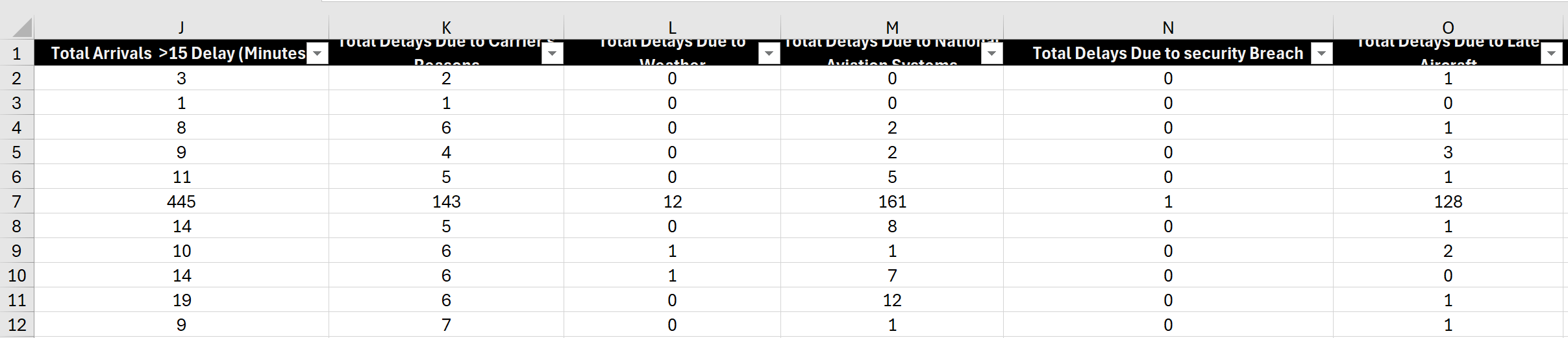
**INTRODUCTION**

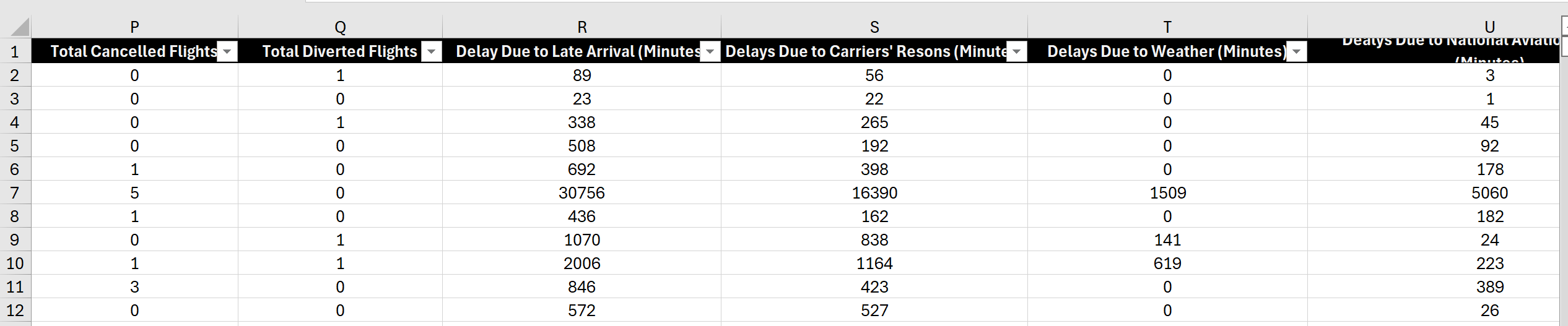
Flight delays are a significant challenge in the aviation industry, affecting airline operations, passenger experience, and overall efficiency. This dataset provides a comprehensive analysis of flight delays, capturing key factors such as departure and arrival times, airline performance, airport congestion, weather conditions, and air traffic control delays. By exploring these variables, the dataset aims to identify patterns and potential causes of delays, helping airlines, researchers, and policymakers develop strategies to improve on-time performance. Whether used for predictive modeling, operational analysis, or customer experience enhancement, this dataset serves as a valuable resource for understanding and mitigating flight disruptions.

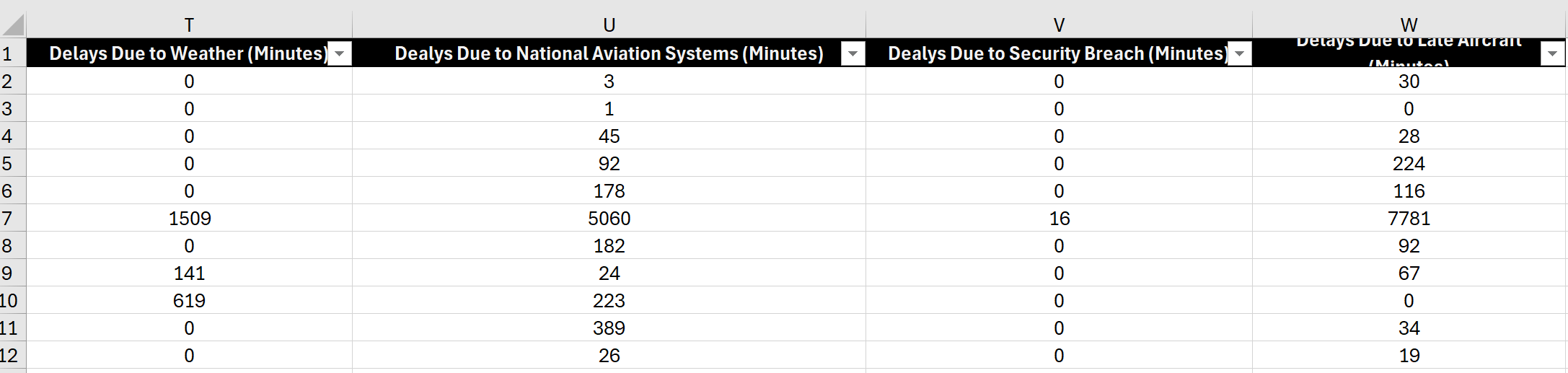
**DATASET**

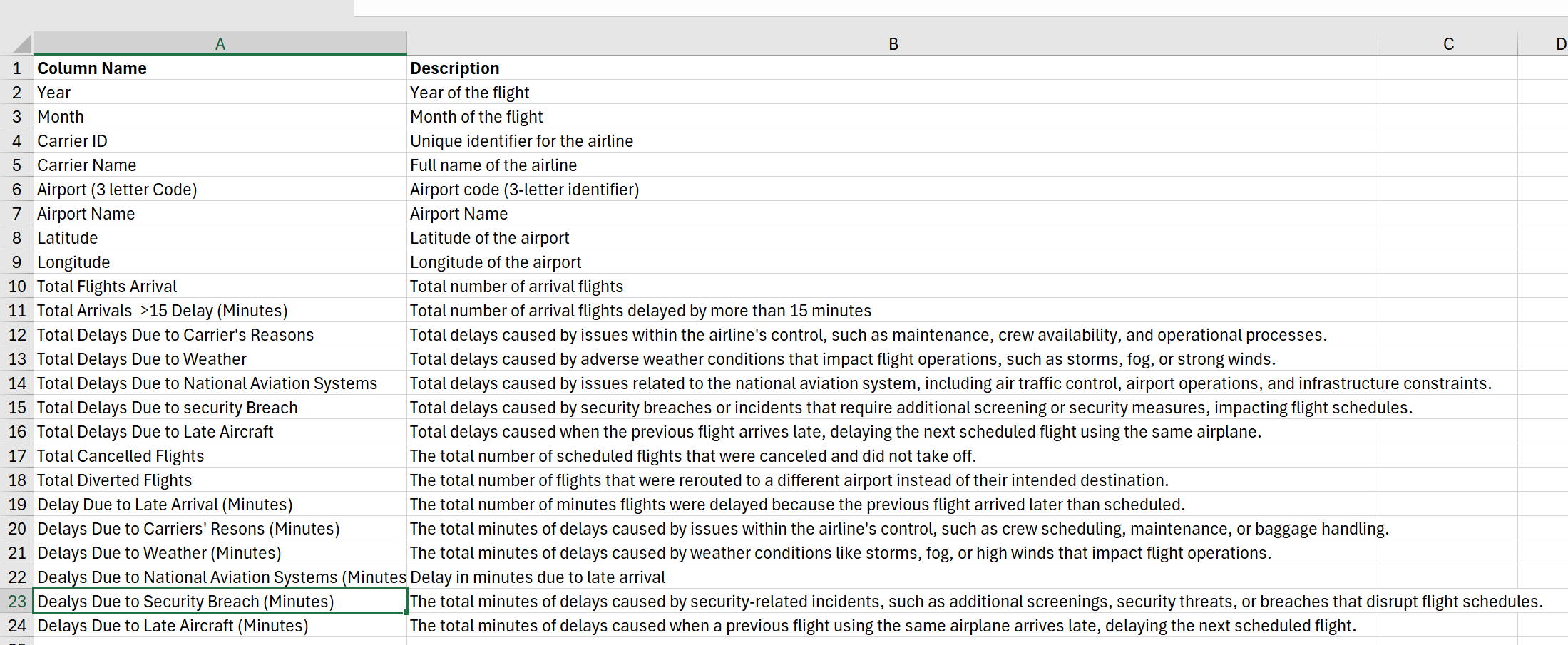
The dataset was gotten from FP20 ANALYTICS – which a data challenge site. Upon getting the data I tried to understand the data before answering the business question(s) being asked. From the exploratory data analysis I carried out, I was able to gather that this was a flight performance data from the year 2023 – 2024. The first few columns show the total flights that arrived at that particular airport by an airline and from this value which flight was late as a result of NAS (National Aviation Systems), carrier reasons, late aircraft, security breach, and weather. And then the last few columns tell me how long each of those delays took (in minutes). To explain better, let’s take the first record which is for carrier ID Endeavor Air Inc. (9E) for January, 2024. It arrived at the Lehigh Valley International Airport (ABE) with longitude (-75.442797) and latitude (40.651773) shows a total flights arrival of 44 and of this 44, 3 were delayed for more than 15 minutes. So that means the total number of flights delayed for >15 min was three for this particular record, 2 of which were delayed due to carrier’s reasons (caused by issues within the airline's control, such as maintenance, crew availability, and operational processes), 0 delays due to weather, National Aviation Systems (NAS), security breach, and 1 delay as a result of late aircraft. There were no flights scheduled for airport ABE that were canceled and did not take off and 1 flight was rerouted (diverted) to a different airport instead of their intended destination - Lehigh Valley International Airport (ABE). These delays resulted in a total of 89-minutes wait time for the passengers because the previous flight arrived later than scheduled. Of this 89-minute delay, 56-minutes were as a result of the carrier’s reasons, none was caused by weather and security breach, 3-minutes delay caused by NAS, and late aircraft accounted for a 30-minutes delay.





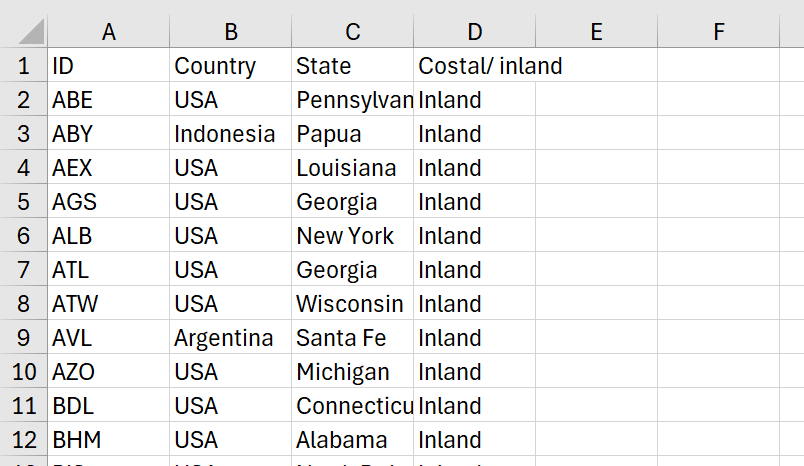






**DATA CLEANING**

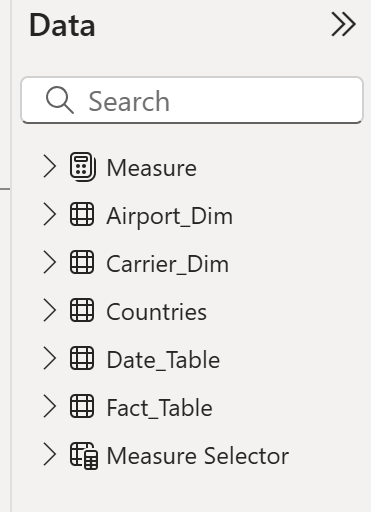
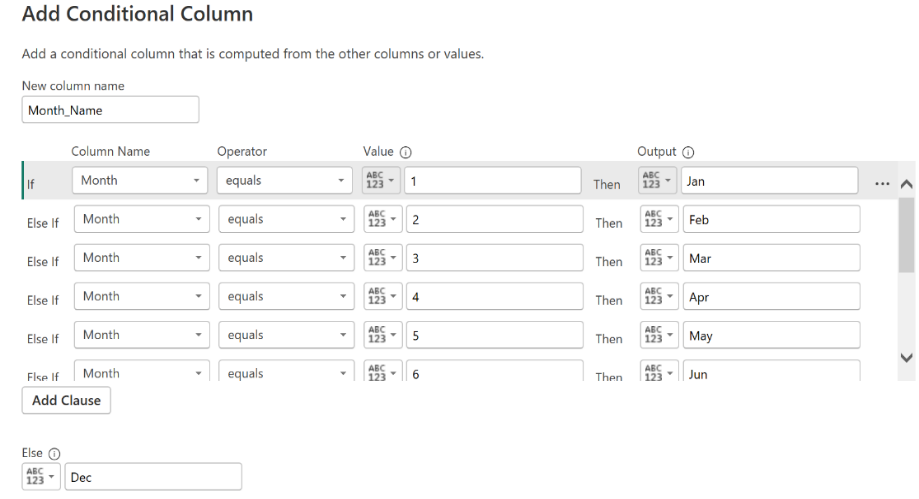
Once I understood the dataset, I then focused on understanding the business questions being asked and how I may tweak my data to provide the best possible insights. This made me realize that part of the business question being asked required me to get the countries in which the 360 (total number of unique airports in this dataset) airports are located, region, and whether they are located in inland or coastal areas. So, I created the “Countries” table as shown below:

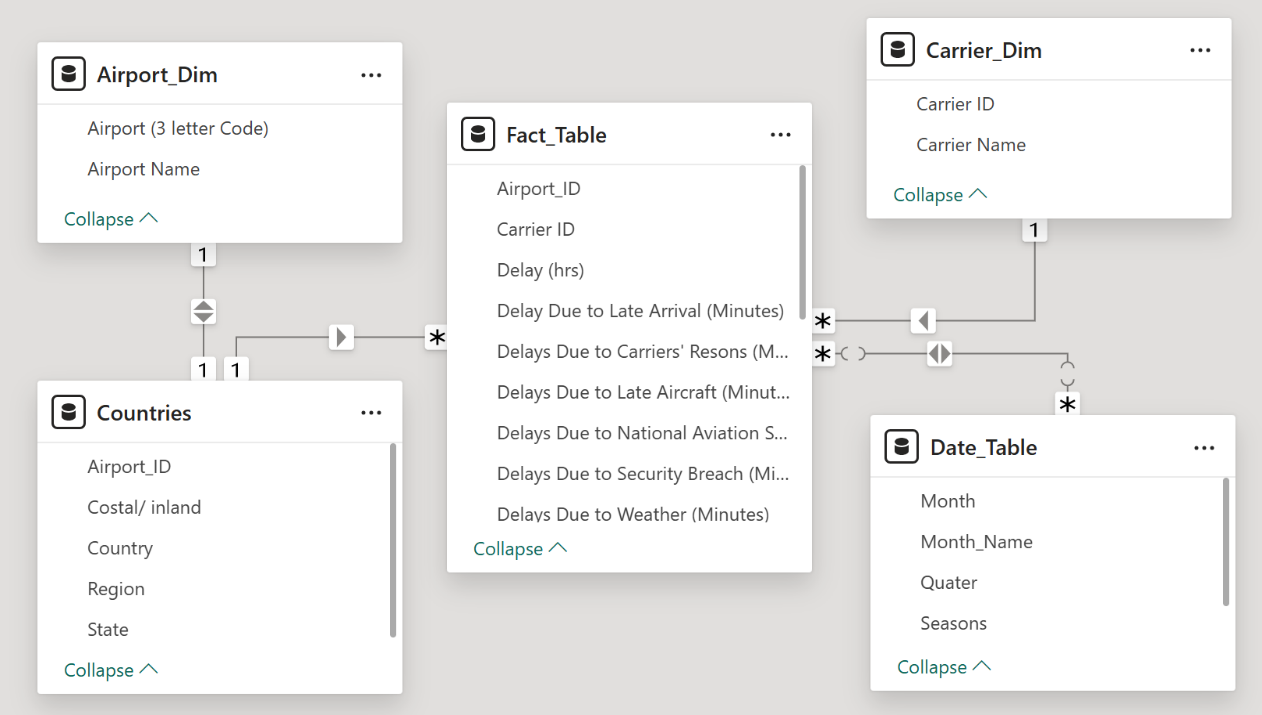


Upon understanding the business question, I then went on to tidy up the data by cross checking the data types to see if they are in the correct format. This was a pretty clean data so there wasn’t much of cleaning to be done.

**DATA MODELING**

Modeling in it simple terms, simply means showing relationships between data and how they interact with one another – how primary keys interact with foreign keys. Modelling this data, I was able to use a simple star schema to show how each table interact with one another and the fact table. I was able to create the “Date\_Table” using conditional formatting in power query for the month (pictured below), seasons, and quarter columns. I then created a many to one relationship between the table and the “Fact\_Table”. The “Airport\_Dim” shows the airport table alongside it unique identifiers, “Carrier\_Dim” shows the carrier table. All of which are connected to the “Fact\_Table”.





**DATA VISUALIZATION**

Before I could visualize the data, I had to write DAX measures that answers the business questions I was tasked with. Here are some of the DAX measures I wrote:

* **OTP by Airline** = CALCULATE([OnTime Performance %],

            ALLEXCEPT(Carrier\_Dim,Carrier\_Dim[Carrier Name])) – *this calculates if aircrafts (airline) arrived on time to their scheduled location.*

* **Overall Offpeak Delay Rate** = CALCULATE(

    DIVIDE(SUM(Fact\_Table[Total Arrivals  >15 Delay (Minutes)]), [Total Flight Volume], 0),

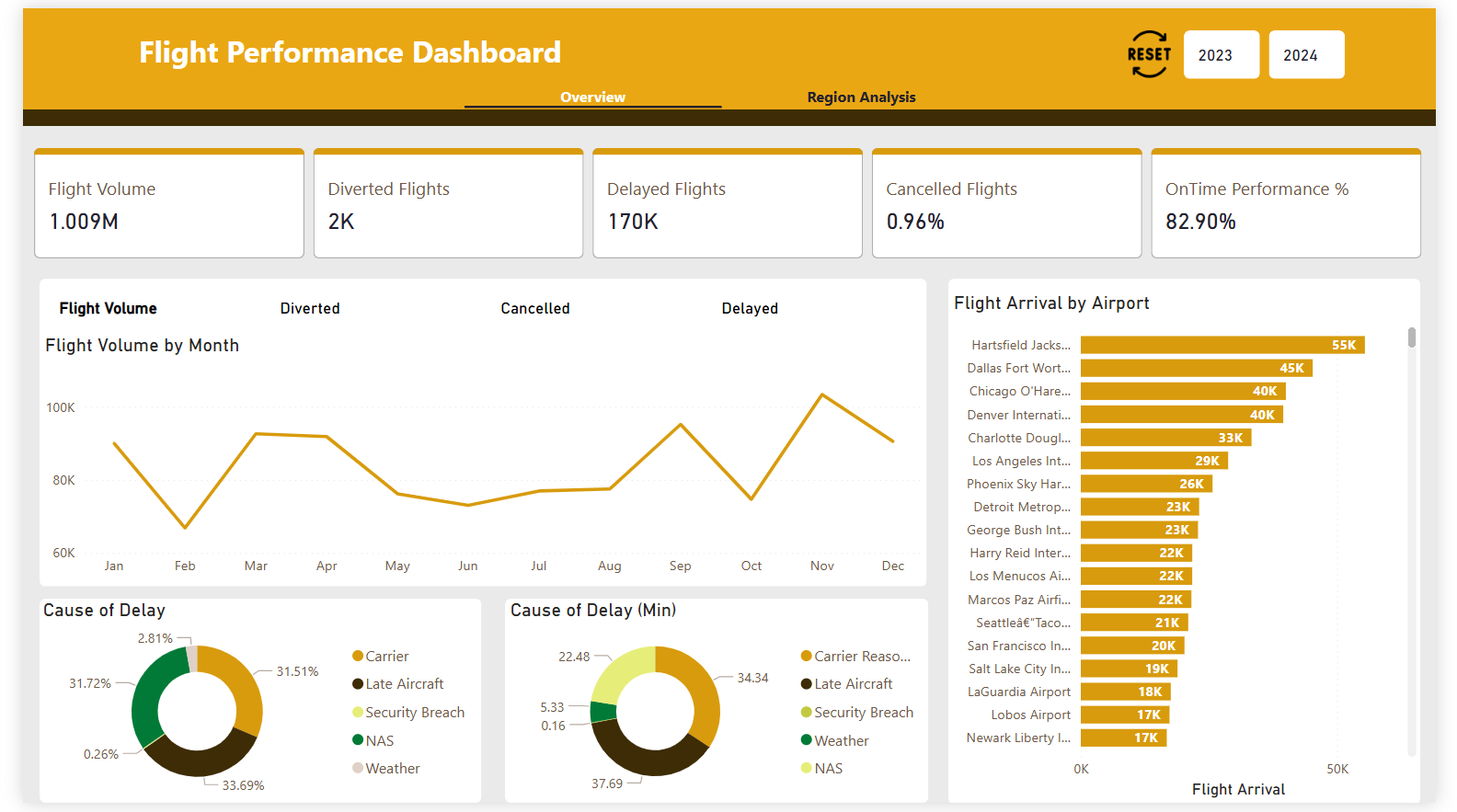
    Date\_Table[Month] IN {2,5,6,7,8,10}

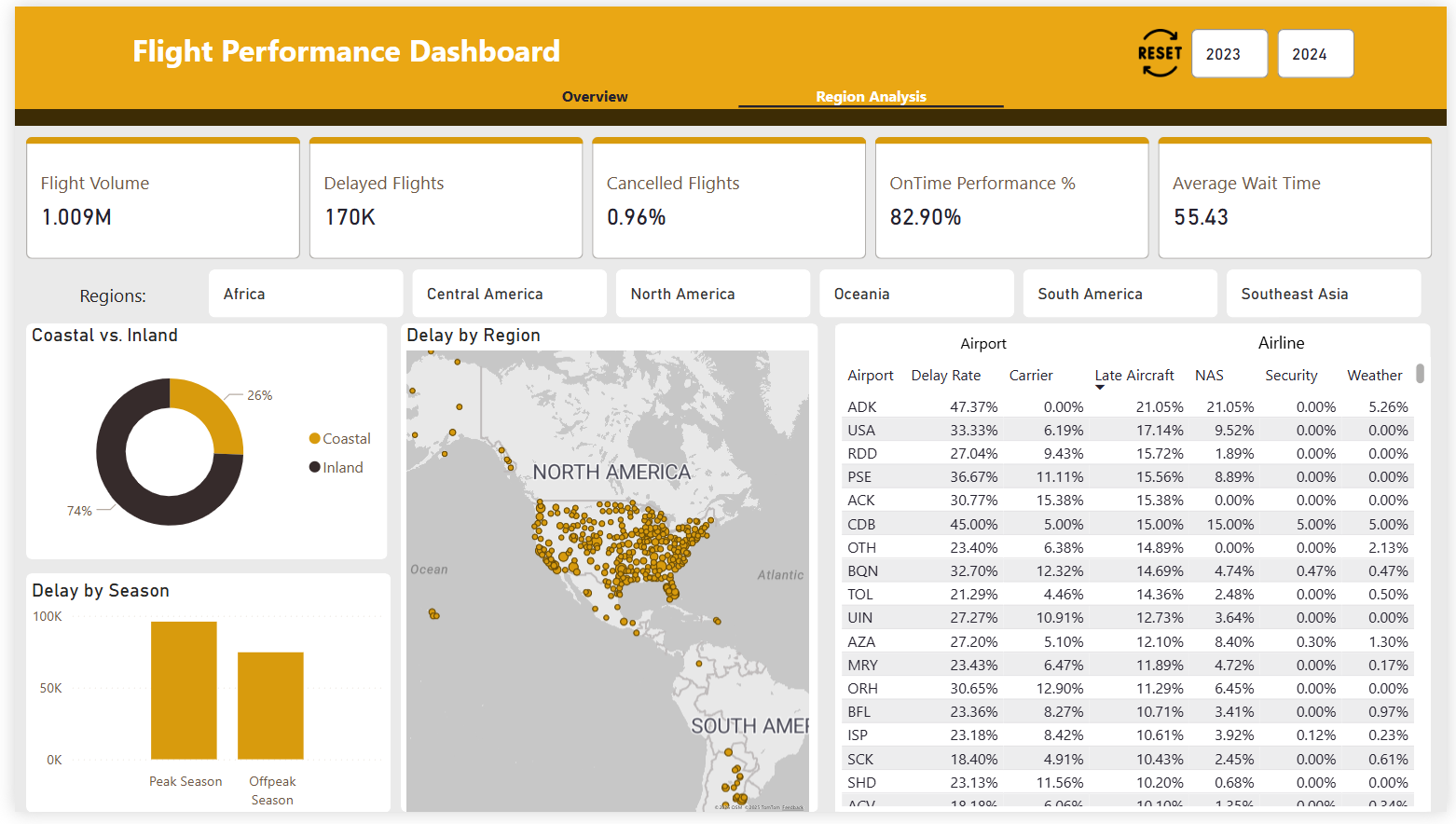
) – *this calculates the delay rate for Feb, May, Jun, Jul, Aug, and Oct (which for this dataset are known as the offpeak seasons that is when flight bookings are low compared to other months)*

* **Total Flight Volume** = SUM(Fact\_Table[Total Flights Arrival]) + SUM(Fact\_Table[Total Diverted Flights]) + SUM(Fact\_Table[Total Cancelled Flights]) – *this calculates the total number of flights scheduled*
* **Average Min** = DIVIDE(AVERAGE(Fact\_Table[Delay Due to Late Arrival (Minutes)]), 60,0) – *calculates the average amount of time a delay take i.e. how long do passengers have to wait on average when there is a delay.*
* **Cause of Delay (carrier)** = DIVIDE(SUM(Fact\_Table[Total Delays Due to Carrier's Reasons]), SUM(Fact\_Table[Total Arrivals  >15 Delay (Minutes)]), 0) – *to get what percentage of the total delay was caused by carrier reasons (this was computed for all the other delay and the result can be seen in the donut chart – cause of delay’ in the visual below)*
* **Count of Inland airport** = CALCULATE(COUNTA(Countries[Costal/ inland]), Countries[Costal/ inland] = "Inland") – *this counts total amount airport from the “countries” table that is inland*

Choosing the right chart to visualize the above-mentioned DAX measures is of outmost importance and can determine how your data is understood by the end users. As you will see in the pictures below, I employed the best possible chart option to convey my findings.

Using parameters was able to create “Measure selector” which allows me to apply my DAX measures to visuals as a filter. In the chart below, I used the DAX measures for total flight volume, diverted flights, cancelled flights, and delayed flights as a filter in my line chart.





**KEY INSIGHTS**

1. There was a total of 1.009 million flights of which 170k was delayed for various reasons (NAS, weather, security breach, late aircraft, and carrier’s reasons), 9643 flights were cancelled and 1925 flights were diverted and 838 flights arrived on time. The average wait time (the amount of time passengers had to wait as a result of a delay factor) was 55.43 mins across board.
2. The OTP (On Time Performance) which measures how often flights arrive at their scheduled gate times, a flight therefore is considered on time if it arrives within 15 minutes of its scheduled arrival. In this case the OTP was 82.90% which shows that the airlines in this case study were efficient and reliable.
3. The main cause of delay based on the number of flights that were delayed was as a result of late aircraft, which accounted for 33.69% of the total delay amount. NAS (National Aviation Systems) accounted for 31.72%, Carrier Reasons (31.51%), weather (2.81%), and security breach (0.26%).
4. 74% of the airports in this dataset are located in inland areas and 26% of the airports are located in coastal areas. While there is more flight volume (644.094k) at airports within the inland areas, coastal area airport had a lower delay as a result of weather (0.39%).
5. There are a total of 360 airports, of which 92 are located in coastal areas while 268 are inland. Dallas Fort Worth International airport which is an inland airport had the most diverted (158) flights and the greatest number of delays (8448), and Lobos Airport (Coastal) had the most cancelled flight.

**CONCLUSION**

Transportation of people and goods has always been an integral part of human evolution and globalization as a whole. And the quicker we can get from one point the other, the more our capacity – to think, to innovate, to expand – is increased. In this dataset I analyzed how delays in the air mode of transportation affects the operational efficiency of an airline, passengers, and can result in financial losses. The reasons for these delays were: Carrier’s reason, weather, National Aviation Systems (NAS), Security breach, and Late aircraft. Upon cleaning and visualizing the data I found that the major cause of delays are Late aircraft, NAS, and carrier’s reasons. They are responsible for more than 60% of the delays according to the data. By analyzing these patterns, stakeholders can develop effective strategies to improve scheduling, optimize operations, and enhance overall flight reliability. Leveraging data-driven decision-making will be crucial in minimizing disruptions and ensuring a smoother travel experience for passengers. With continued advancements in technology and operational planning, airlines can work towards reducing delays and improving efficiency across the aviation sector.

**RECOMMENDATIONS**

Airlines require a high level of labor, capital and process. When not properly managed or misused can lead to increased financial cost, bad customer experience (in this case delays) and an overall inefficient business process. One of the possible solutions to combat this would be for airlines to implement what is called “the lean approach” to their operations. This is a process aimed at reducing in the amount of waste (cost) and increasing the revenue using their already existing resources. Another solution would be to pre-service the aircraft before flight to prevent it from being towed to a different spot that makes that difficult seeing as servicing cant be done until a departure gate is available. All the all, the airline should look into standardizing their process and very often should always look for areas that may encourage waste. Late aircraft can be reduced by improving the scheduling process which allows for a realistic scheduling time that avoids tight turnarounds of aircrafts. Overall, to minimize flight delays and enhance operational efficiency, airlines should leverage predictive analytics to anticipate potential disruptions and proactively adjust schedules. Implementing real-time air traffic and weather monitoring systems can help optimize flight routes – even though in this dataset weather was responsible for the least number of delays and reduce delays caused by congestion or adverse conditions. Additionally, improving ground operations – such as faster turnaround times, efficient crew scheduling, and better coordination with air traffic control – can significantly enhance on-time performance. By utilizing insights from this dataset, airlines, airport authorities, and policymakers can develop data-driven strategies to improve punctuality, reduce passenger inconvenience, and optimize overall air travel reliability.