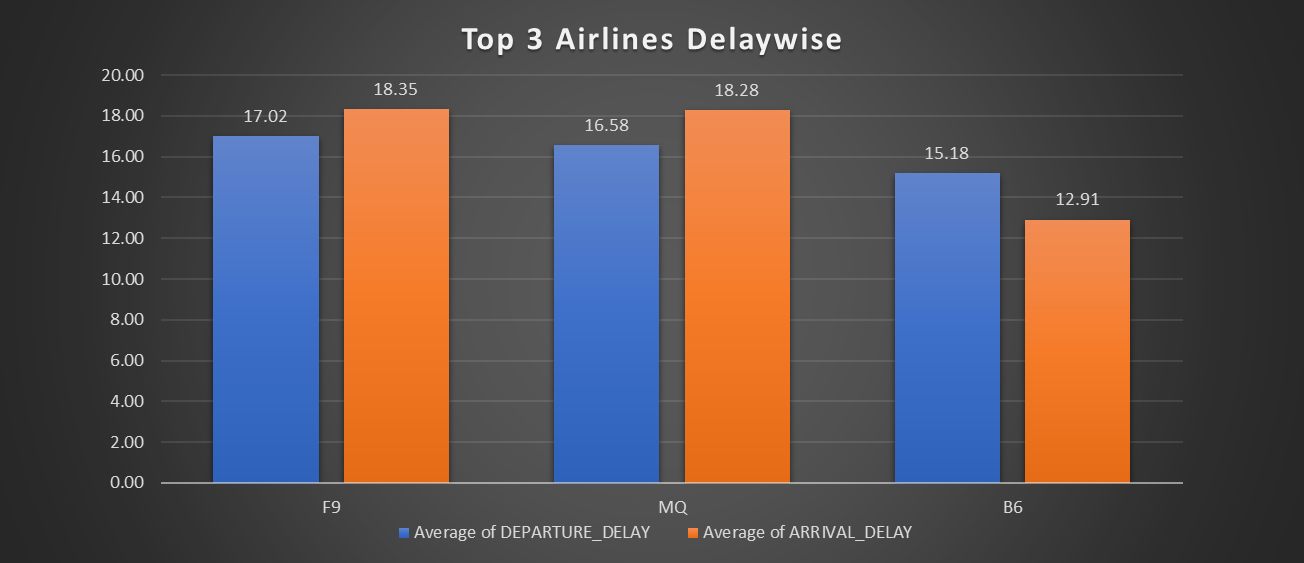
# **Travel & Hospitality**

### **Sky Analytics: Navigating the Complexities of Airline and Airport Operations**

### **Part 1: Excel Data Analysis: Manipulation, Formulas and Functions**

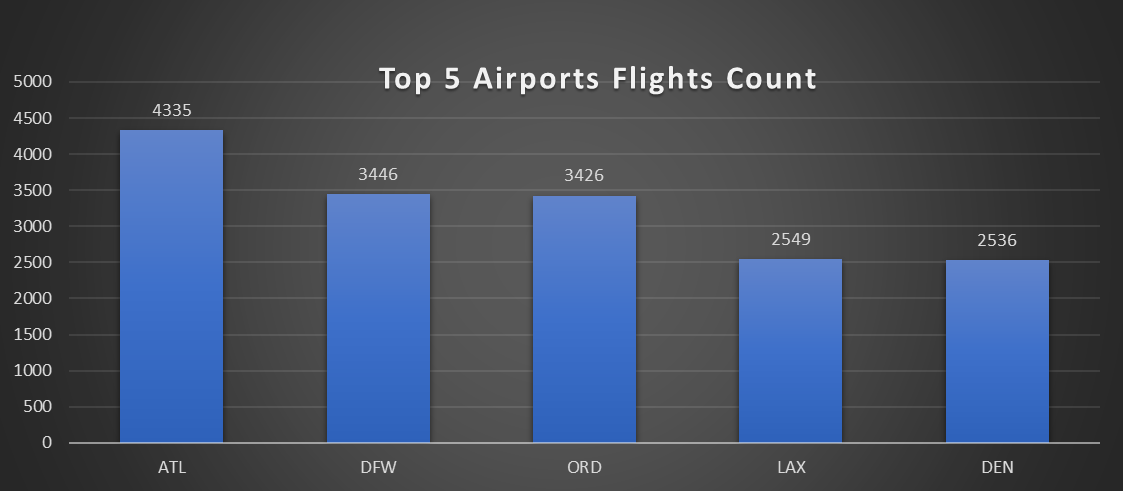
1. **Identify and address missing data in the movie's dataset. Are there any patterns in the missing data that can be noted?**
2. **Determine the average flight delay per airline. What are the top 3 airlines with the highest average delays?**

|  |  |  |
| --- | --- | --- |
| **Airlines** | **Average of DEPARTURE\_DELAY** | **Average of ARRIVAL\_DELAY** |
| F9 | 17.02 | 18.35 |
| MQ | 16.58 | 18.28 |
| B6 | 15.18 | 12.91 |



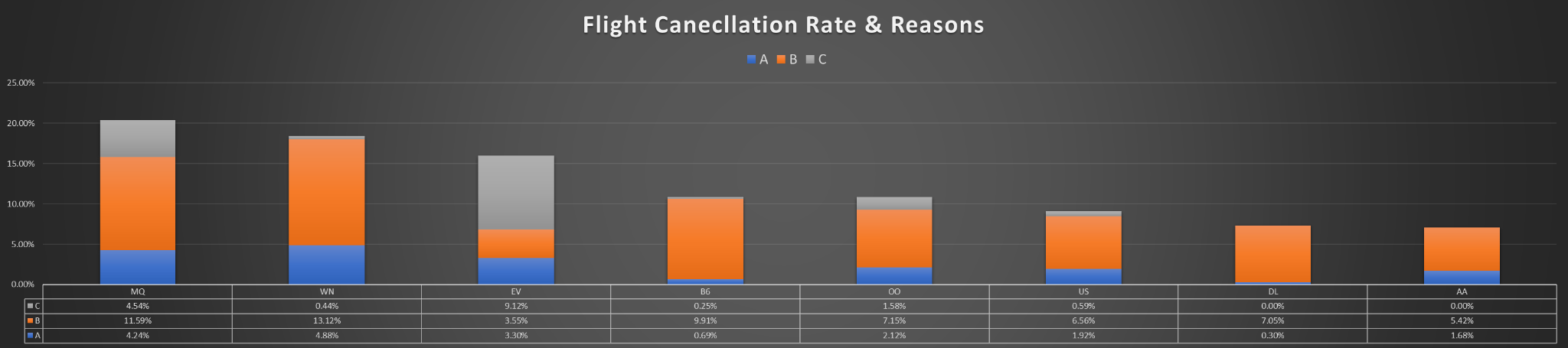
1. **Identify the top 5 busiest airports based on the number of incoming and outgoing flights.**

|  |  |
| --- | --- |
| **Row Labels** | **Count of AIRPORT** |
| ATL | 4335 |
| DEN | 2536 |
| DFW | 3446 |
| LAX | 2549 |
| ORD | 3426 |

****

1. **Analyse the flight cancellations: Which airline has the highest cancellation rate, and what are the most common reasons for cancellations?**

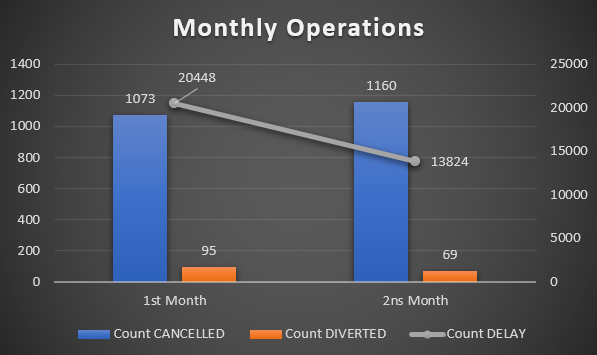
|  |  |  |  |
| --- | --- | --- | --- |
| **Airports** | **A** | **B** | **C** |
| **MQ** | **4.24%** | **11.59%** | **4.54%** |
| **WN** | **4.88%** | **13.12%** | **0.44%** |
| **EV** | **3.30%** | **3.55%** | **9.12%** |
| **B6** | **0.69%** | **9.91%** | **0.25%** |



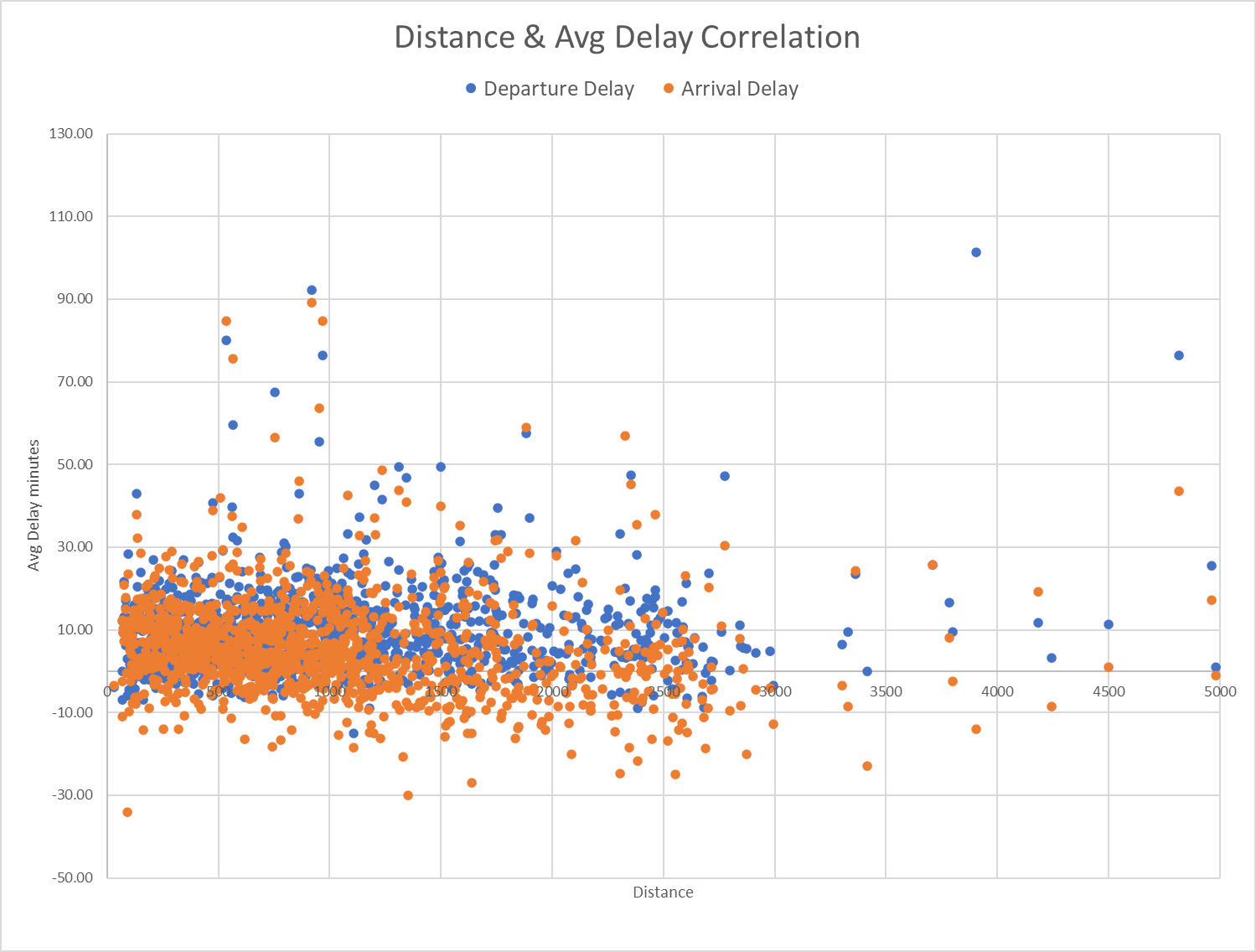
1. **Examine if there are seasonal patterns in flight operations. Are certain months more prone to delays or cancellations?**

|  |  |  |  |
| --- | --- | --- | --- |
| **Month** | **Count CANCELLED** | **Count DIVERTED** | **Count DELAY** |
| 1st Month | 1073 | 95 | 20448 |
| 2ns Month | 1160 | 69 | 13824 |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

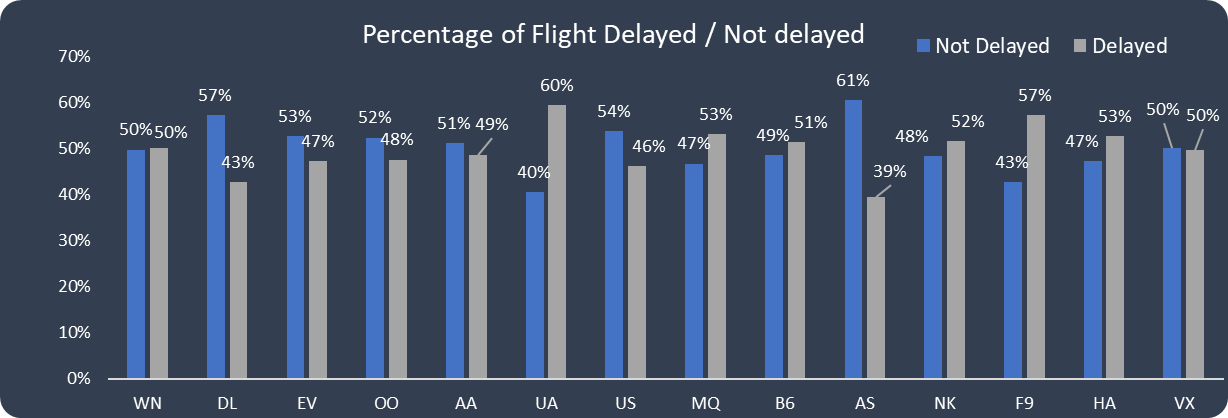
****

1. **Investigate if there's a correlation between the distance of the flight and the length of delays. Use scatter plots for visualisation.**

****

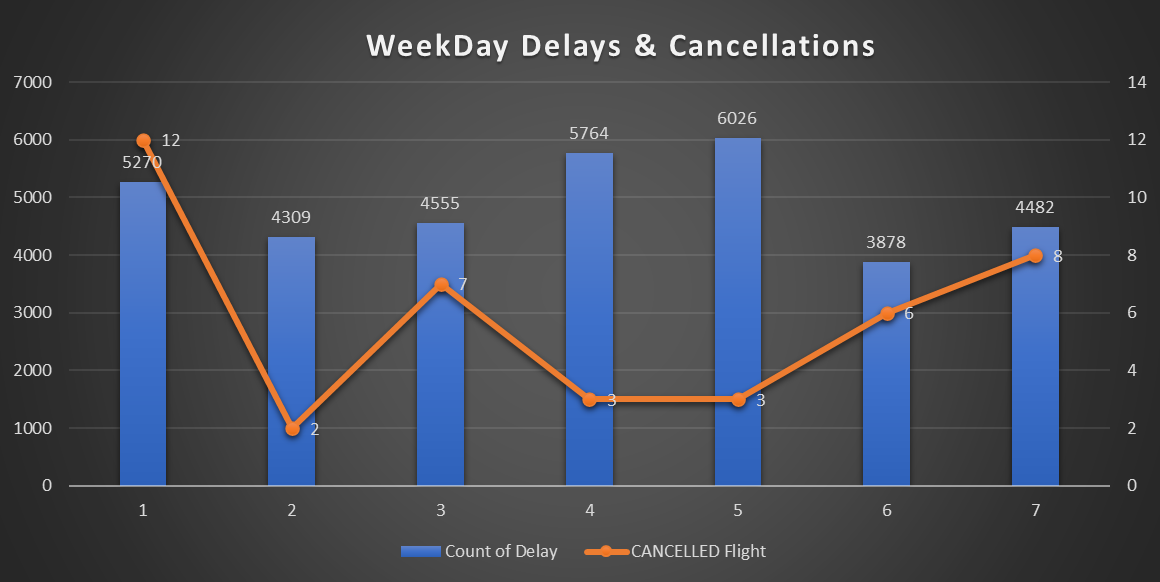
1. **Calculate the on-time performance (percentage of flights that are not delayed) for airlines. Rank them based on this metric.**

|  |  |  |
| --- | --- | --- |
| **Row Labels** | **Not Delayed** | **Delayed** |
| **WN** | **50%** | **50%** |
| **DL** | **57%** | **43%** |
| **EV** | **53%** | **47%** |
| **OO** | **52%** | **48%** |
| **AA** | **51%** | **49%** |
| **UA** | **40%** | **60%** |
| **US** | **54%** | **46%** |
| **MQ** | **47%** | **53%** |
| **B6** | **49%** | **51%** |
| **AS** | **61%** | **39%** |
| **NK** | **48%** | **52%** |
| **F9** | **43%** | **57%** |
| **HA** | **47%** | **53%** |
| **VX** | **50%** | **50%** |

****

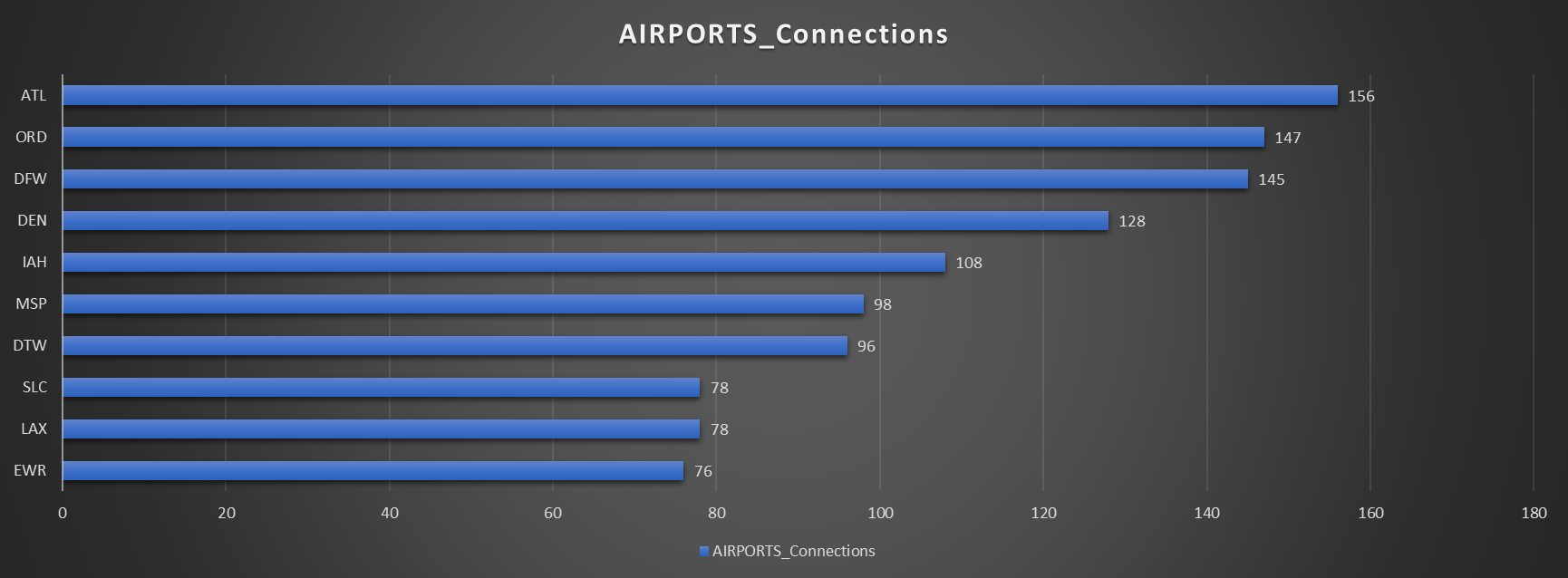
1. **Assess how flight operations (delays, cancellations) vary by the day of the week.**

|  |  |  |
| --- | --- | --- |
| **Weekday** | **Count of Delay** | **CANCELLED Flight** |
| **1** | **5270** | **12** |
| **2** | **4309** | **2** |
| **3** | **4555** | **7** |
| **4** | **5764** | **3** |
| **5** | **6026** | **3** |
| **6** | **3878** | **6** |
| **7** | **4482** | **8** |

****

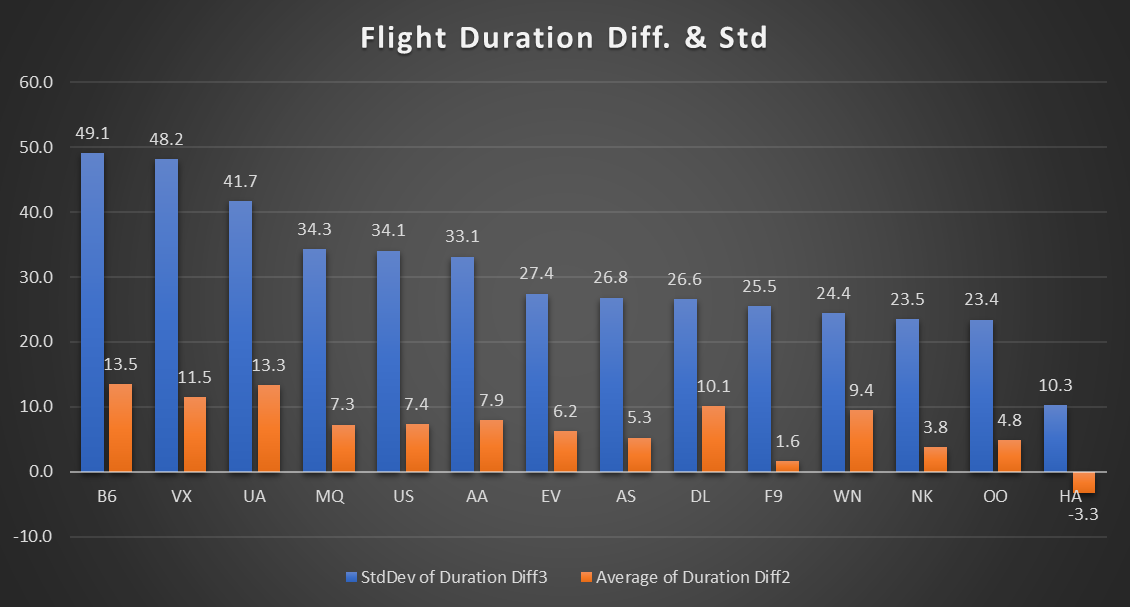
1. **Which airports serve as the most significant hubs in terms of connectivity (most destinations served)?**

|  |  |
| --- | --- |
| **Airports** | **AIRPORTS\_Connections** |
| **EWR** | **76** |
| **LAX** | **78** |
| **SLC** | **78** |
| **DTW** | **96** |
| **MSP** | **98** |
| **IAH** | **108** |
| **DEN** | **128** |
| **DFW** | **145** |
| **ORD** | **147** |
| **ATL** | **156** |

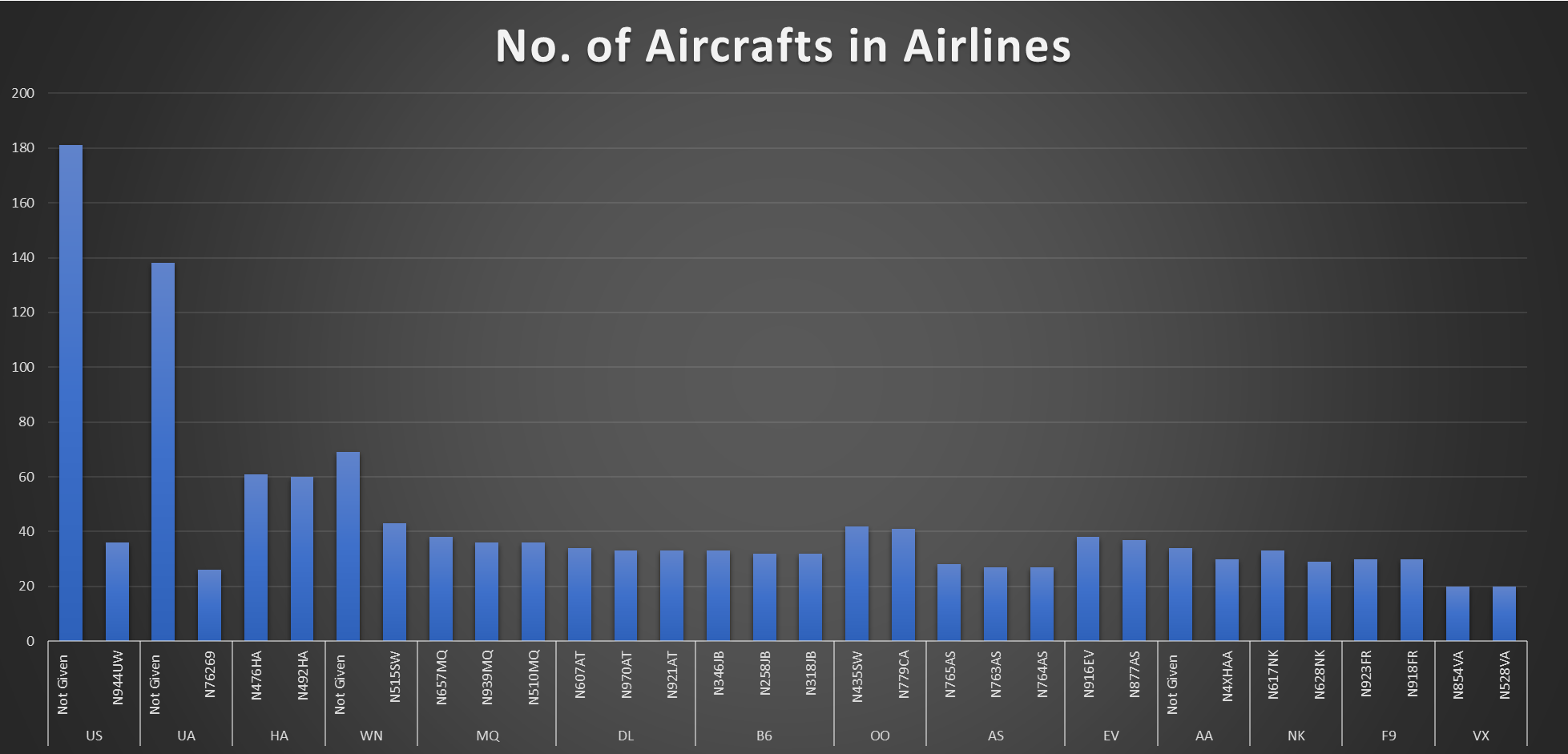
****

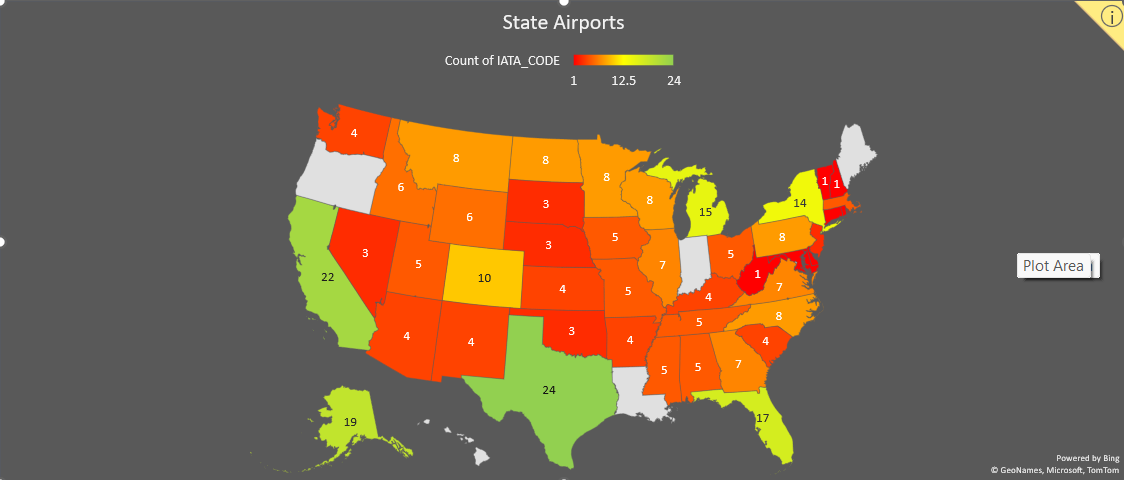
1. **Compare the scheduled flight duration versus the actual flight duration. Which airlines have the most and least deviation?**

|  |  |  |
| --- | --- | --- |
| **Airlines** | **Average of Duration Diff.** | **StdDev of Duration Diff.** |
| **B6** | **13.5** | **49.1** |
| **VX** | **11.5** | **48.2** |
| **UA** | **13.3** | **41.7** |
| **MQ** | **7.3** | **34.3** |
| **US** | **7.4** | **34.1** |
| **AA** | **7.9** | **33.1** |
| **EV** | **6.2** | **27.4** |
| **AS** | **5.3** | **26.8** |
| **DL** | **10.1** | **26.6** |
| **F9** | **1.6** | **25.5** |
| **WN** | **9.4** | **24.4** |
| **NK** | **3.8** | **23.5** |
| **OO** | **4.8** | **23.4** |

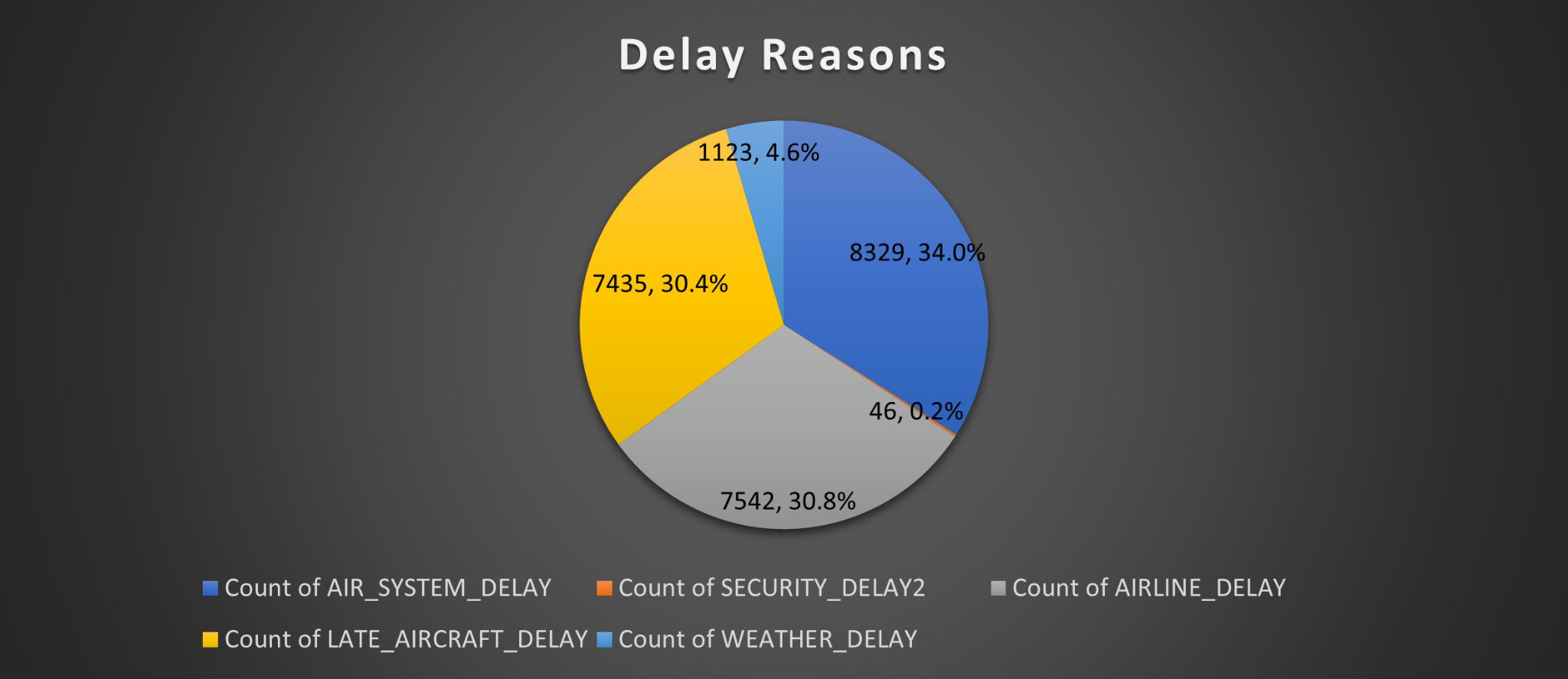
****

1. **Based on the tail numbers, determine which airline has the highest number of flights per aircraft, indicating fleet utilisation.**

****

1. **Using latitude and longitude data, analyse the geographical distribution of airports. Which states or regions have the highest concentration of airports?**
2. **For flights that are delayed, break down the delay types (airline, weather, security, etc.) and analyse their proportions.**

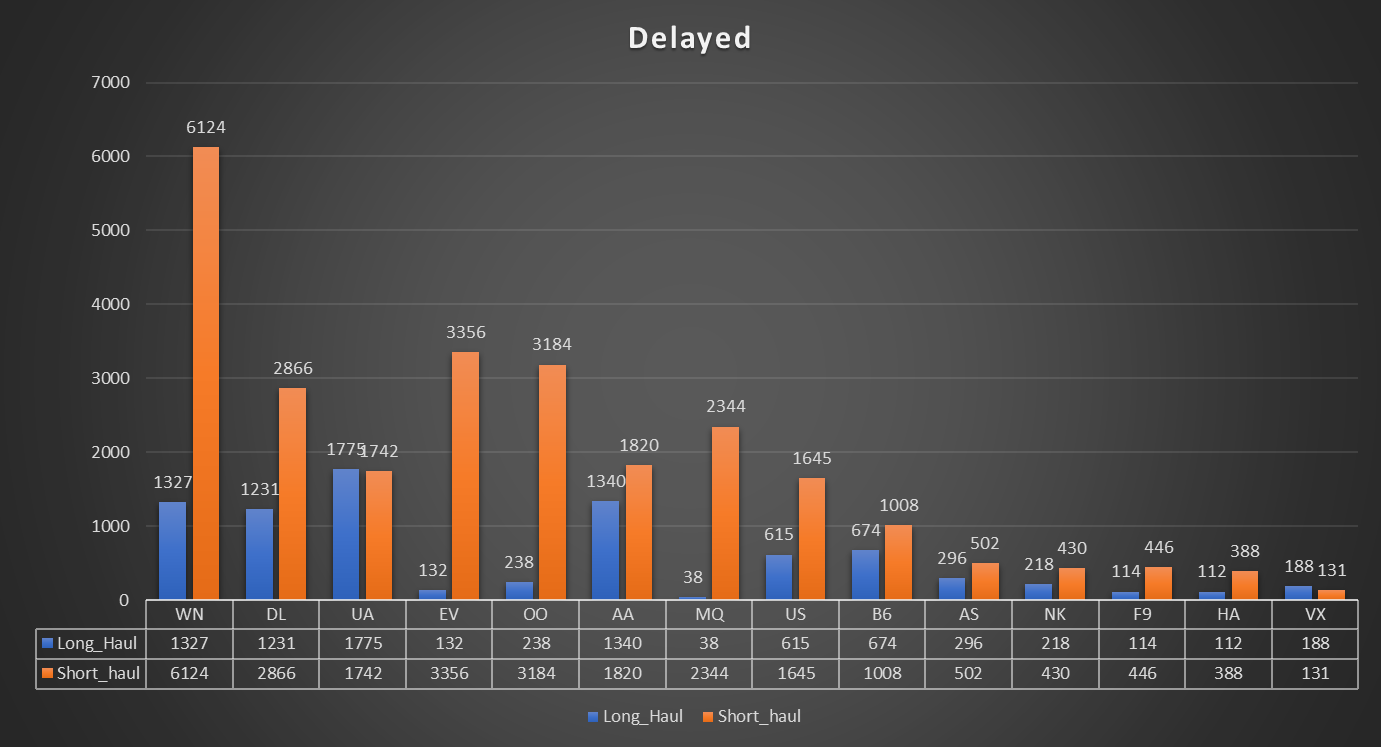
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Count of AIR\_SYSTEM\_DELAY** | **Count of SECURITY\_DELAY2** | **Count of AIRLINE\_DELAY** | **Count of LATE\_AIRCRAFT\_DELAY** | **Count of WEATHER\_DELAY** |
| **8329** | **46** | **7542** | **7435** | **1123** |
| **34.0** | **0.2** | **30.8** | **30.4** | **4.6** |

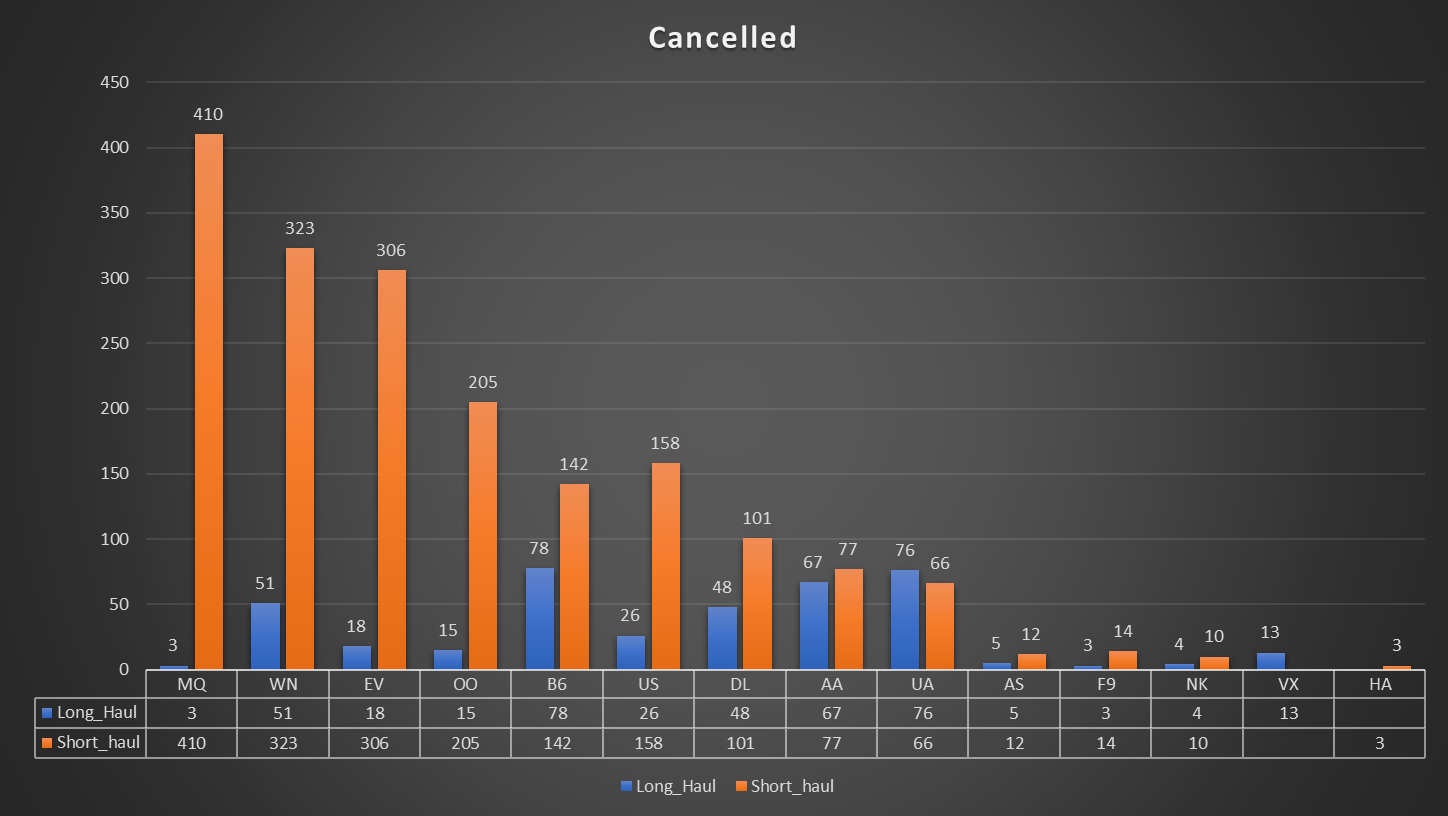
****

1. **Compare the operational metrics (delays, cancellations) between long-haul and short-haul flights for different airlines.**

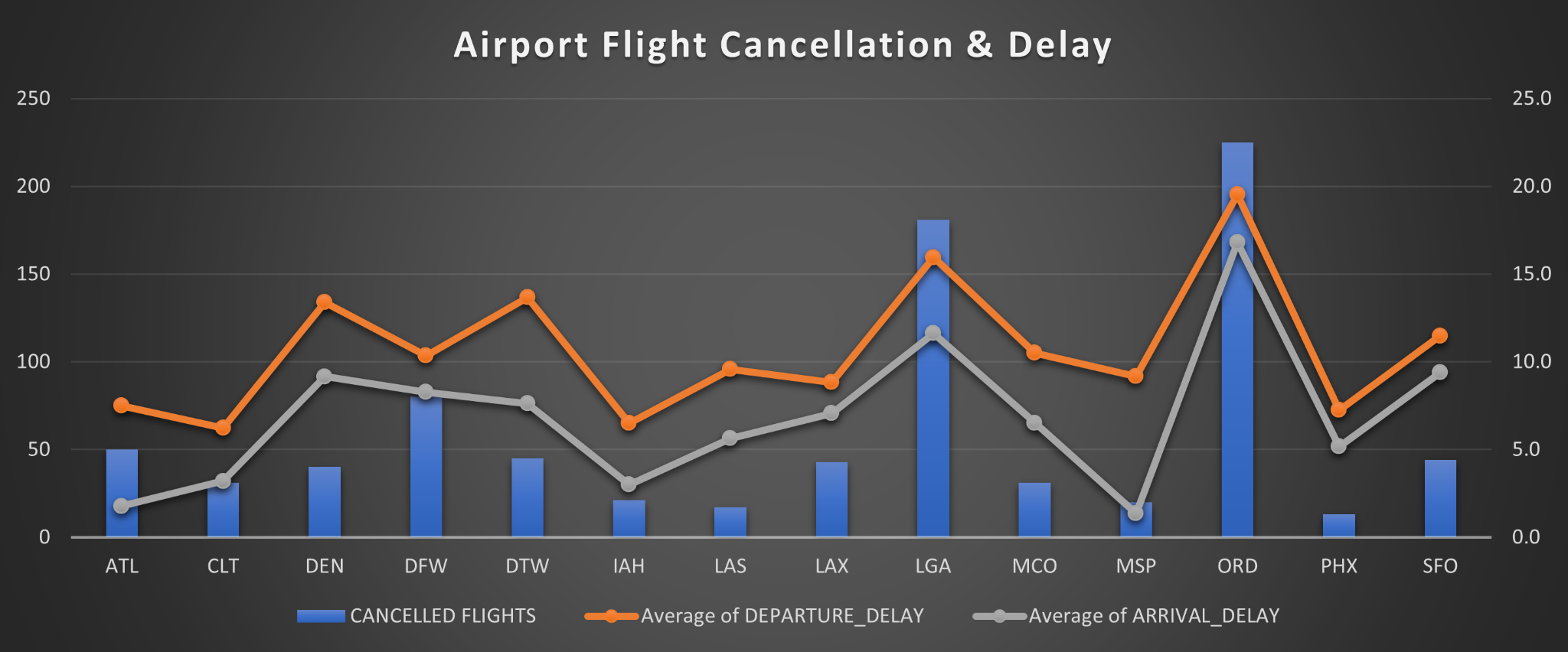
**Short haul flights < 3 hours**

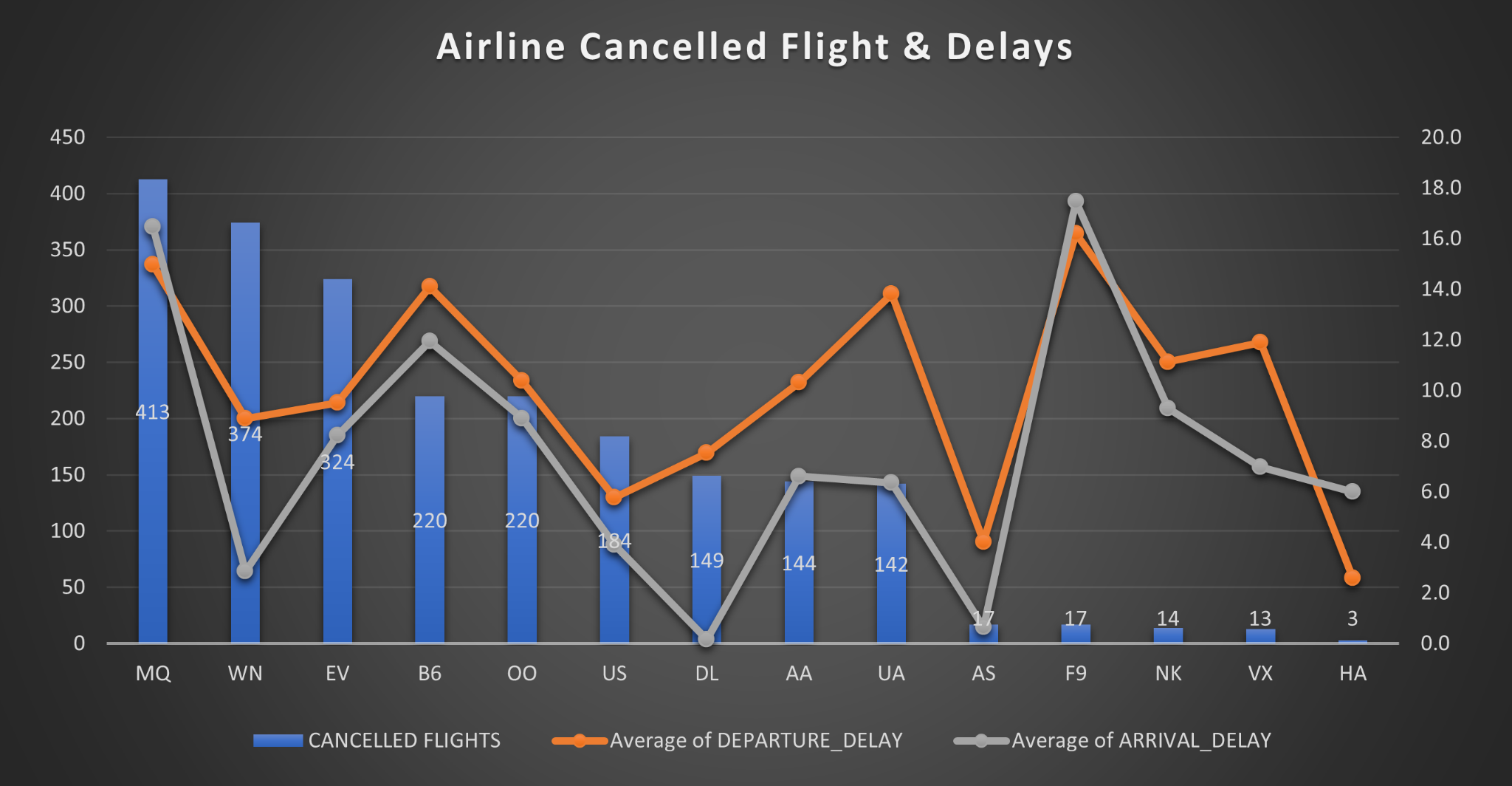
**Long Haul flights >3 hours**

****

****

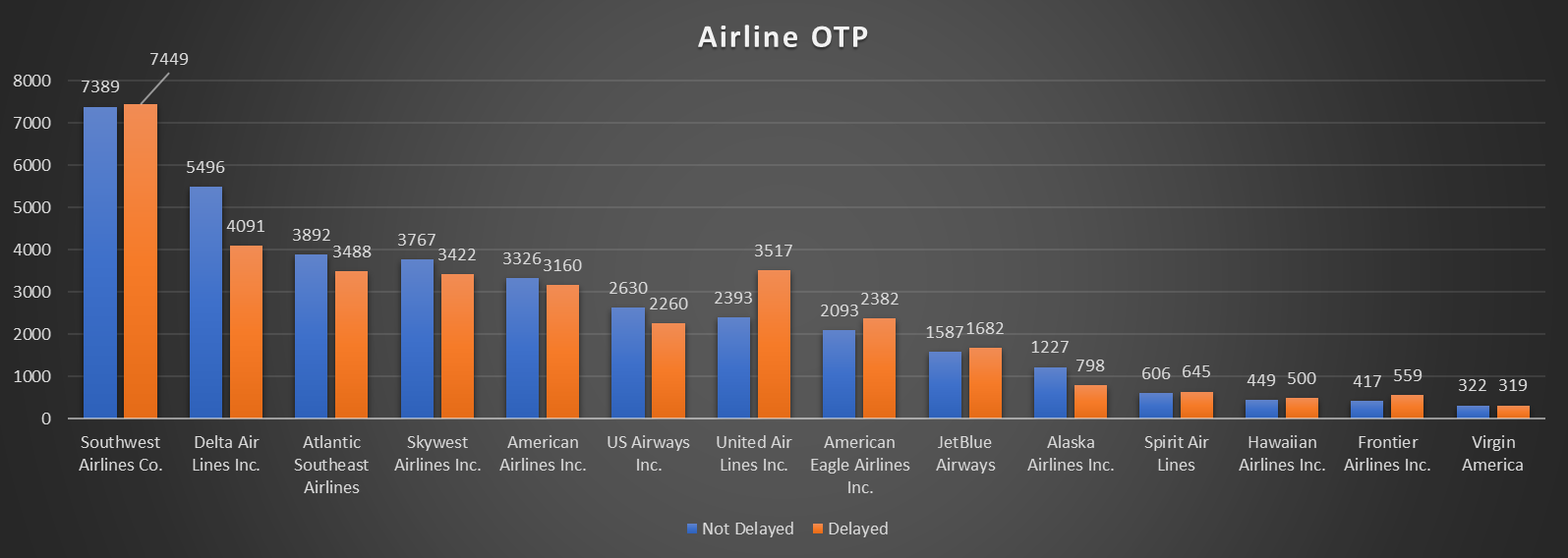
1. **Use pivot tables to summarise key operational metrics (like average delay, number of flights, cancellations) by airline and airport.**

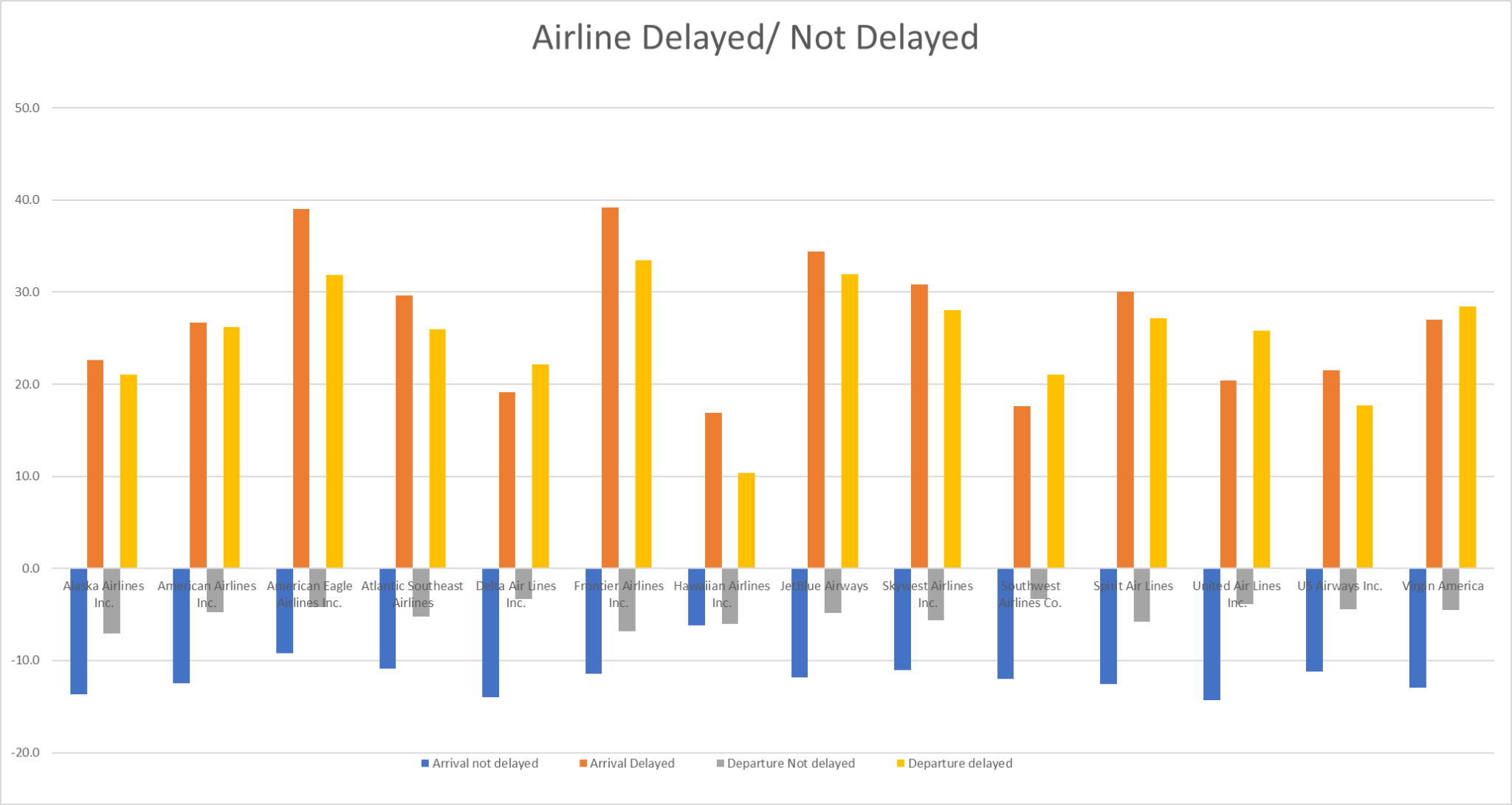
****

****

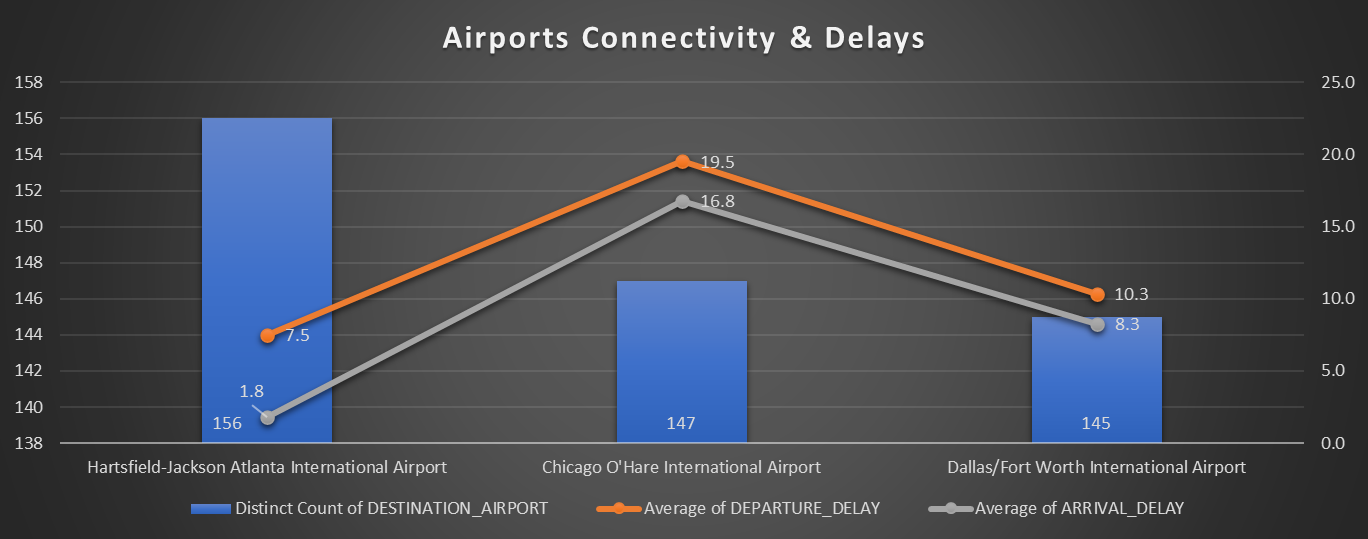
1. **Merge data from the "airlines.csv" and "flights.csv" to provide enhanced insights, such as correlating airline names with operational metrics. Analyse the merged data to determine the overall on-time performance of each airline, considering both arrival and departure delays.**

**On-time performance**

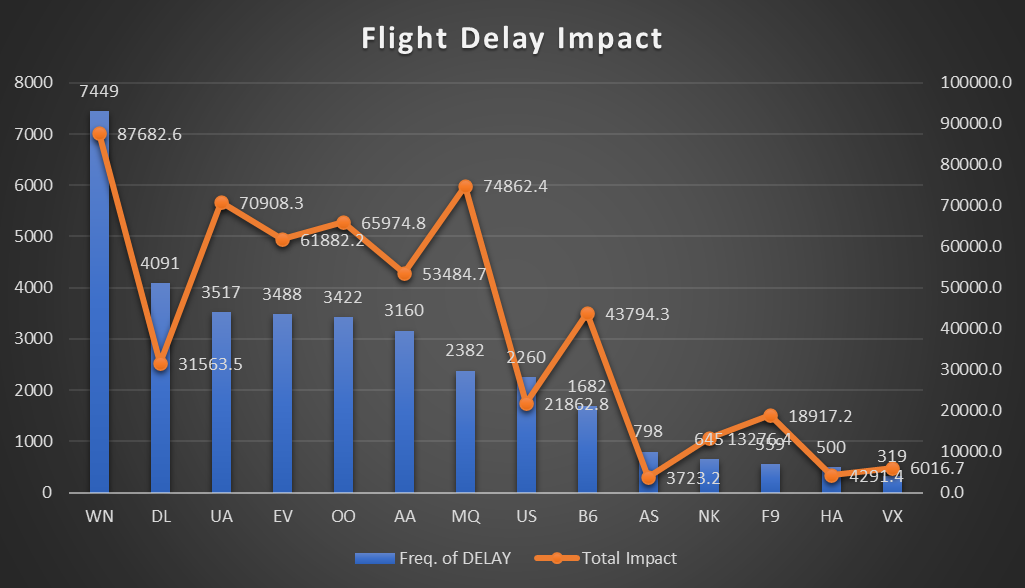
****

****

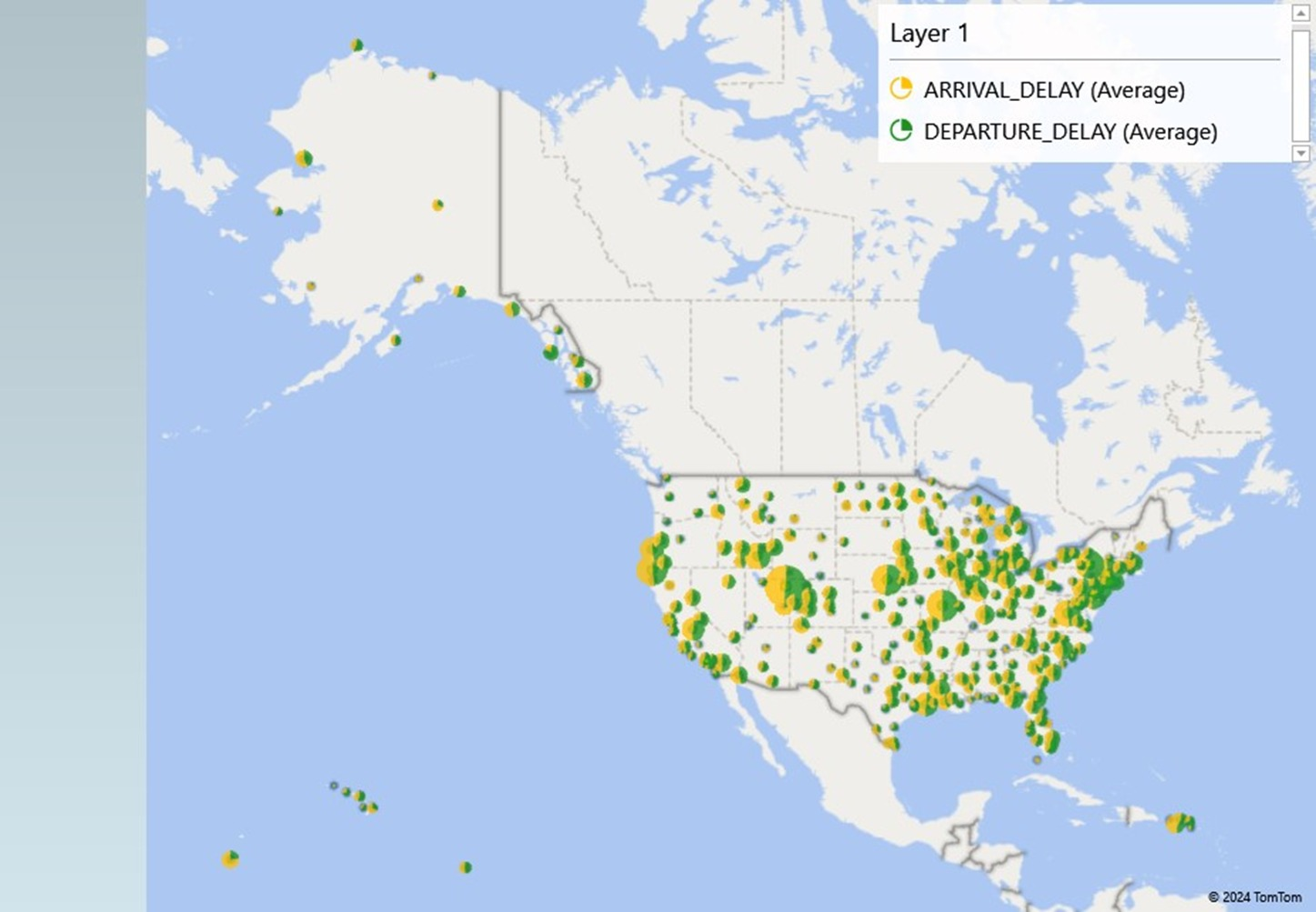
1. **Combine the "flights.csv" dataset with the "airports.csv" based on airport codes. Use the merged dataset to identify the top 3 airports in terms of flight connectivity (number of unique destinations served) and analyze their average delay times (both departure and arrival).**

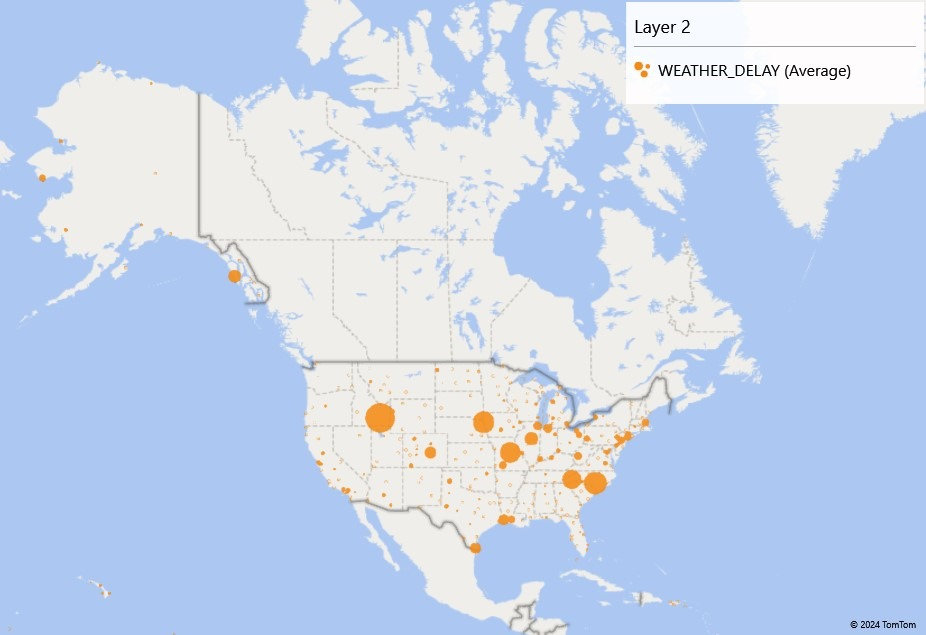
****

1. **Calculate the cumulative impact of delays for each airline. Consider both the frequency of delays and the average delay time. How do these factors combine to affect overall airline performance?**

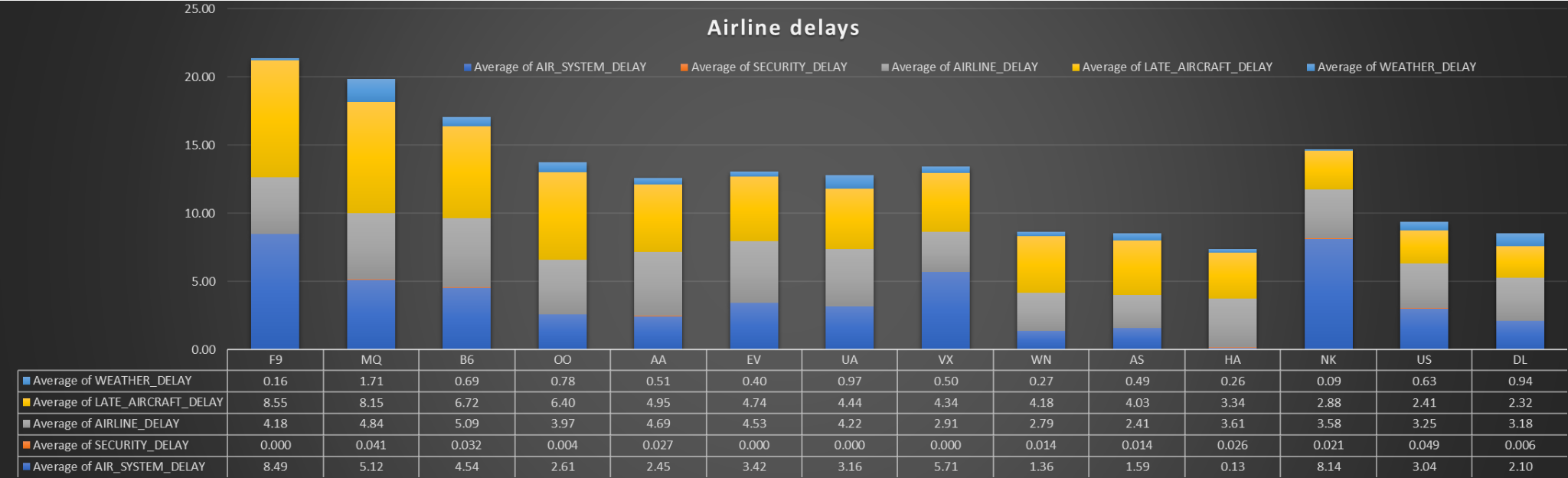
****

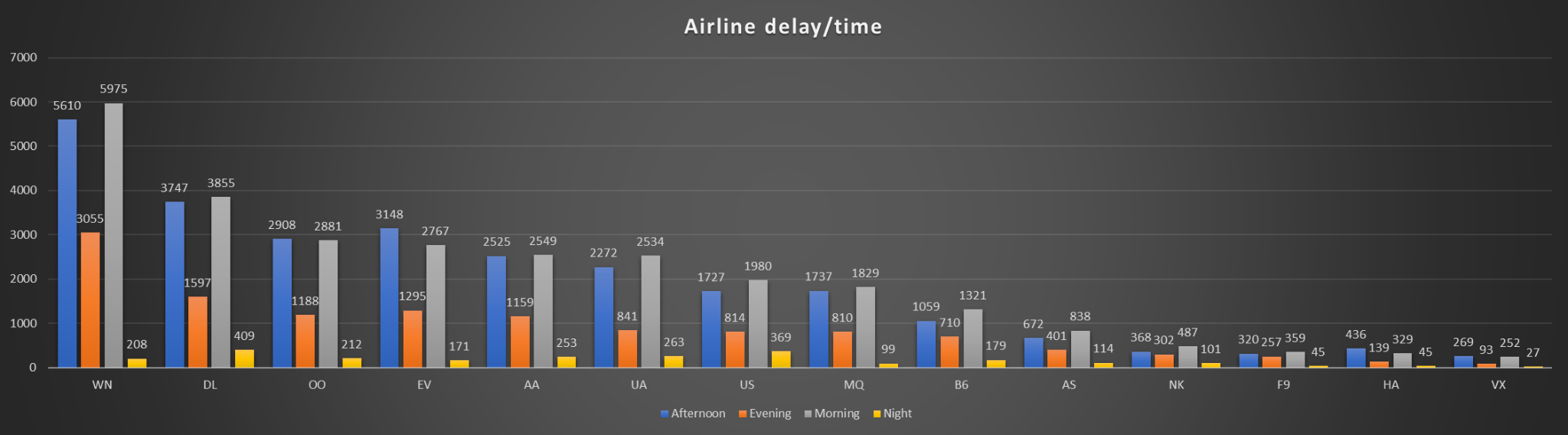
1. **Investigate if there's a correlation between airport performance (in terms of delays) and environmental factors like location (latitude and longitude from "airports.csv") and time of year (seasonal weather conditions). Are certain airports more prone to delays due to their geographical location and the associated weather patterns?**

****

****

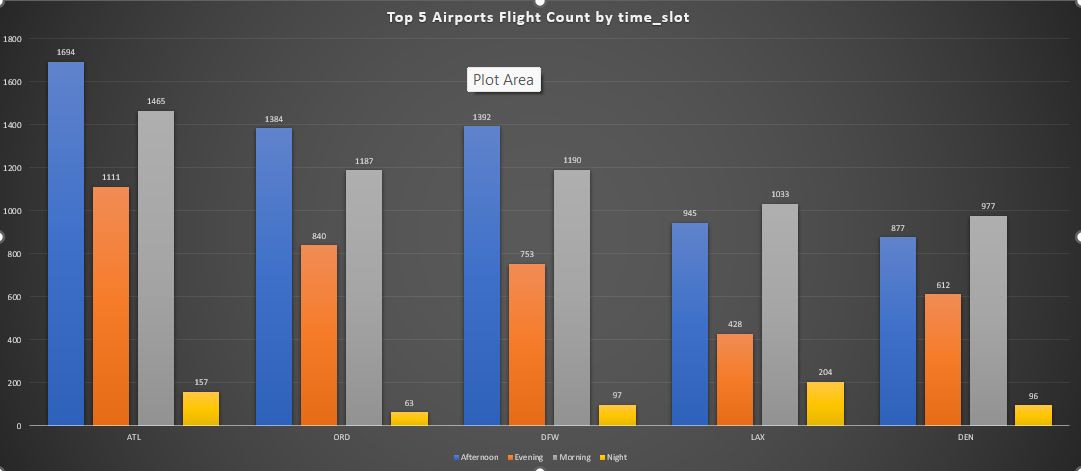
1. **Use nested functions to analyse the primary cause of delays for each airline. Determine if the predominant cause of delay (like airline delay, weather delay, security delay) varies by airline and time of day.**

****

****

1. **Analyse the flight frequencies to determine the peak operating hours for major airports. (Use a combination of Excel functions to categorise flights into different time slots (e.g., morning, afternoon, evening, night) and calculate the number of flights in each slot for the top 5 busiest airports.)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Airports** | **Afternoon** | **Evening** | **Morning** | **Night** |
| **ATL** | **1694** | **1111** | **1465** | **157** |
| **ORD** | **1384** | **840** | **1187** | **63** |
| **DFW** | **1392** | **753** | **1190** | **97** |
| **LAX** | **945** | **428** | **1033** | **204** |
| **DEN** | **877** | **612** | **977** | **96** |

****

### **Part 2: Building an Excel Dashboard**

**Create an interactive and dynamic Excel dashboard that provides a visual and analytical representation of key aspects of the aviation datasets. The dashboard should offer insights into flight operations, airline performance, and airport activity.**

### **Key Components to Include in the Dashboard:**

1. **Flight Operations Overview:**
   * **Incorporate a summary view showing total number of flights, average delays, and flight cancellations. Include filters to view data by specific airlines or airports.**
2. **Airline Performance Analysis:**
   * **Visualize performance metrics for each airline, such as average delay time, cancellation, and frequency of flights. Use charts like bar graphs or line charts for comparison.**
3. **Airport Traffic Visualization:**
   * **Display data on the busiest airports in terms of incoming and outgoing flights. Include interactive elements like a map for geographical representation and pie charts for traffic distribution.**
4. **Delay Cause Breakdown:**
   * **Provide a detailed analysis of the reasons for flight delays (weather, security, airline delays, etc.). Use a combination of pivot tables and stacked bar charts for this analysis.**
5. **Time-based Flight Trends:**
   * **Show how flight patterns and delay trends vary over different times of the day and across months. Implement a timeline or slider feature to allow users to explore temporal changes.**
6. **User Interaction Features:**
   * **Add slicers, dropdown menus, and timeline controls for users to interactively filter and explore the data across various dimensions like time periods, airlines, and airport codes.**
7. **Dashboard Aesthetics and Usability:**
   * **Ensure the dashboard is not only informative but also visually appealing and easy to navigate. Maintain a consistent color scheme and clear labels.**
8. **Advanced Analysis (Optional):**
   * **If feasible, include a predictive analysis section where users can forecast future trends in airline performance or airport traffic based on historical data.**