

# Dataset exploration

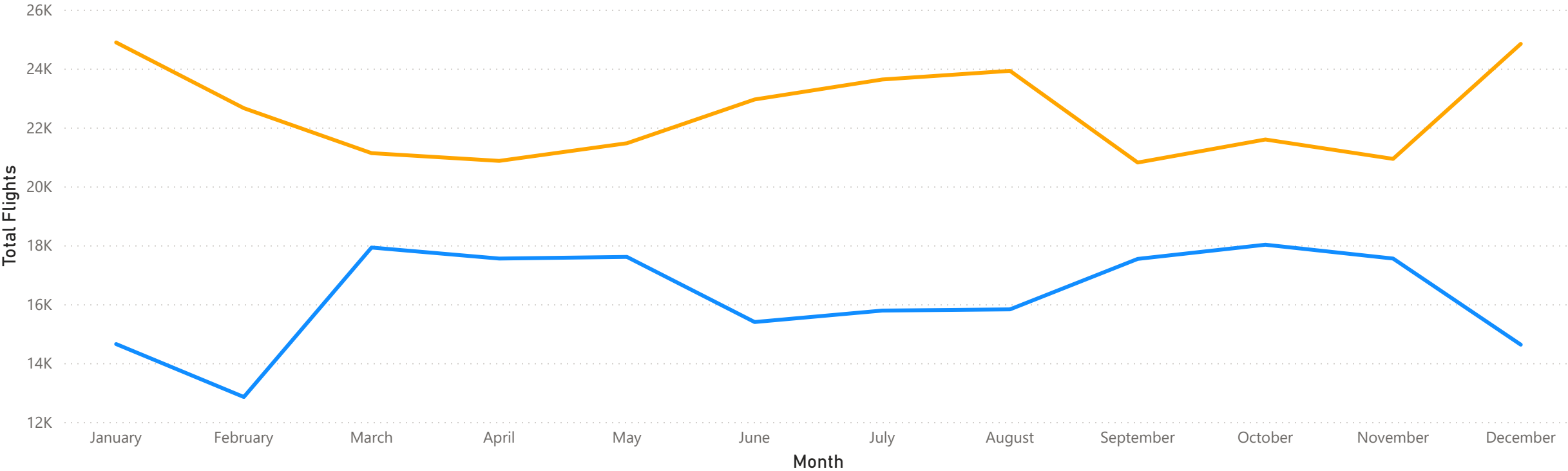
20.78

Avg Take-Off Delay (Mins)

5.62

Avg Passengers Satisfaction score (out of 10)

All flights during 2021  
Flight Status — Delayed — On time

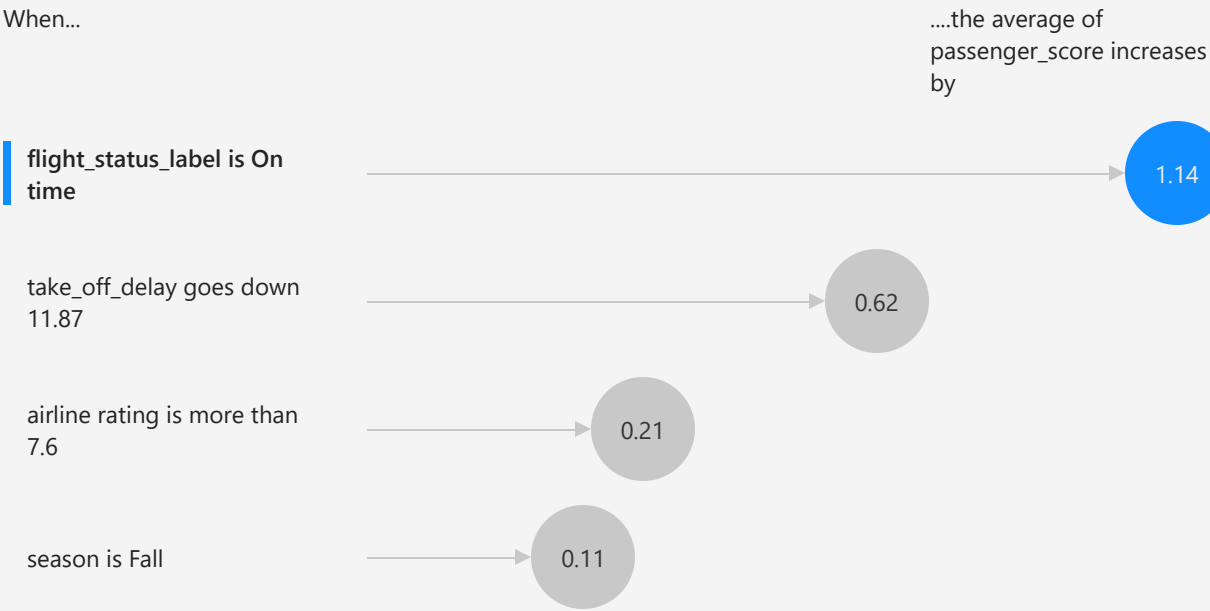


# Key Influencers

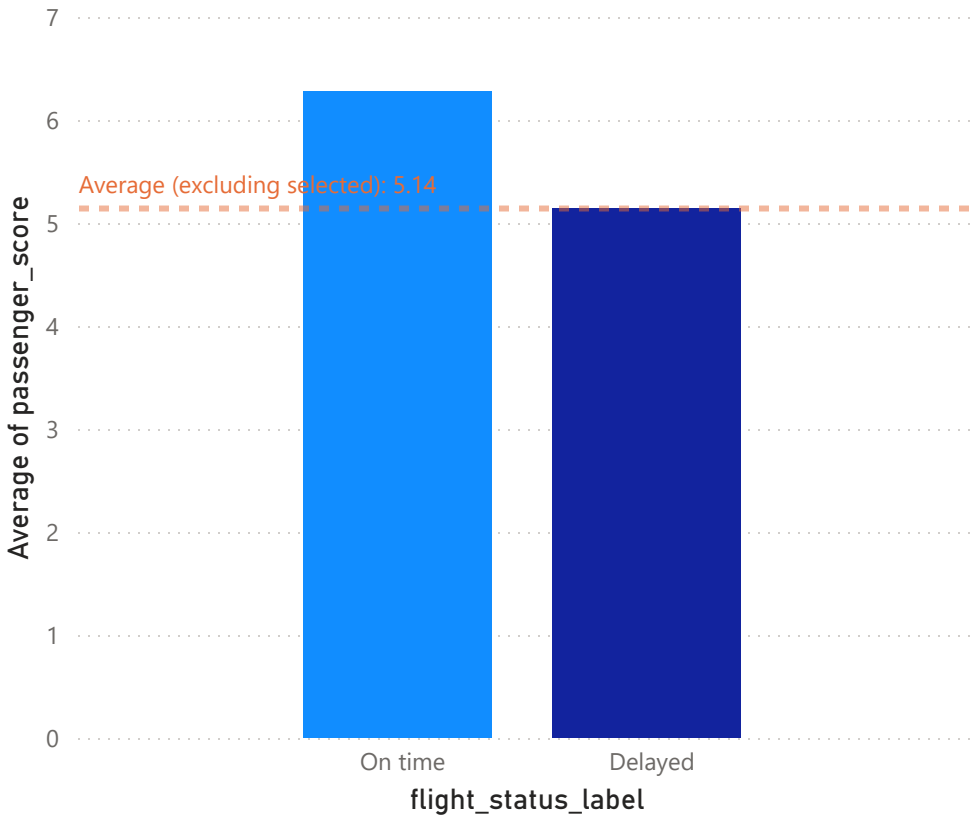
Key influencers   Top segments



What influences passenger\_score to Increase ?



← passenger\_score is more likely to increase when flight\_status\_label is On time than otherwise (on average).



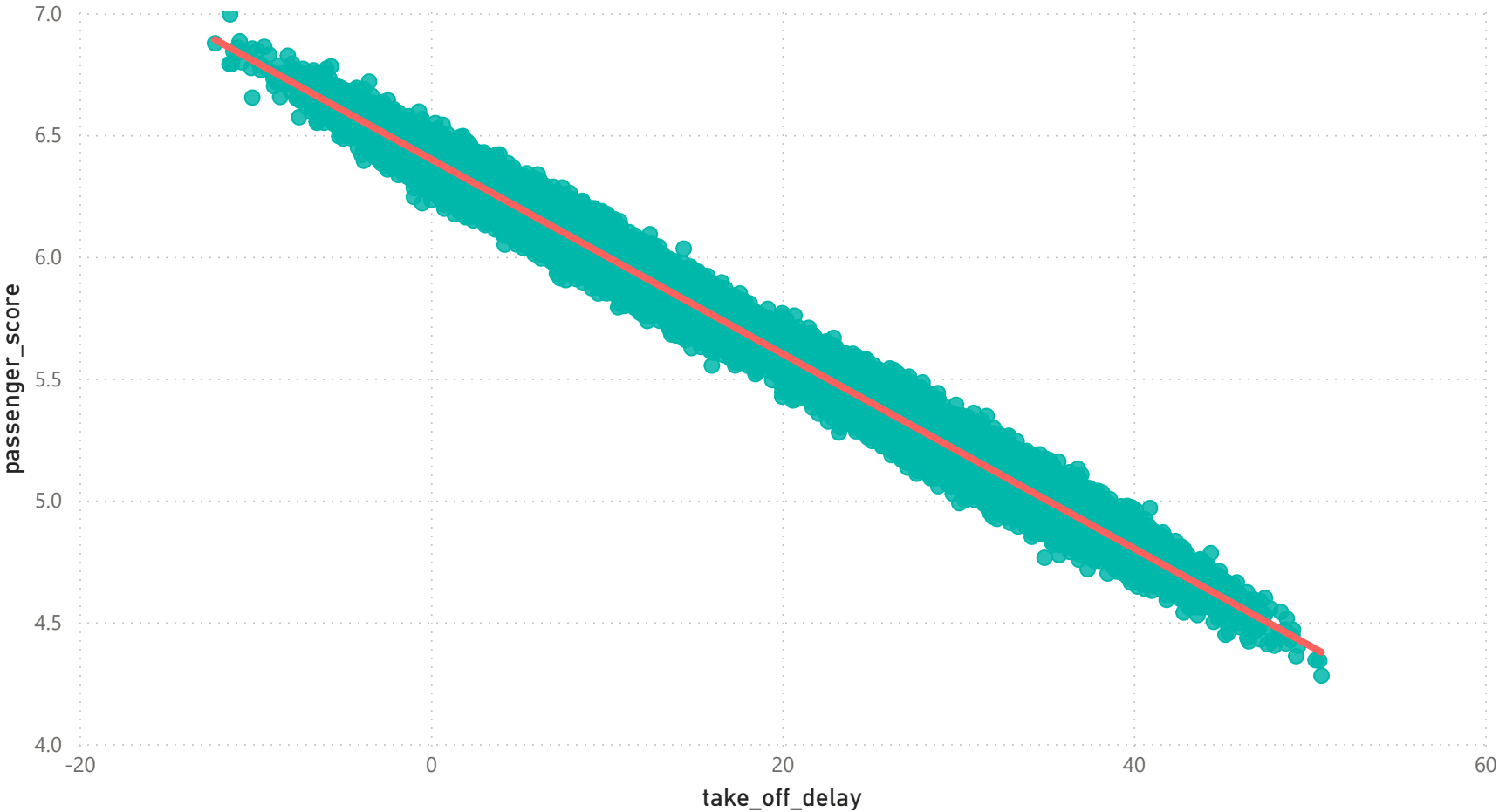
☐ Only show values that are influencers

# Passenger score and Take-off Delay

Passenger Score (y) = -0.04 \* Takeoff delay (x) + 6.40

Linear Regression Formula

Takeoff delay vs. Passenger score



Takeoff delay duration



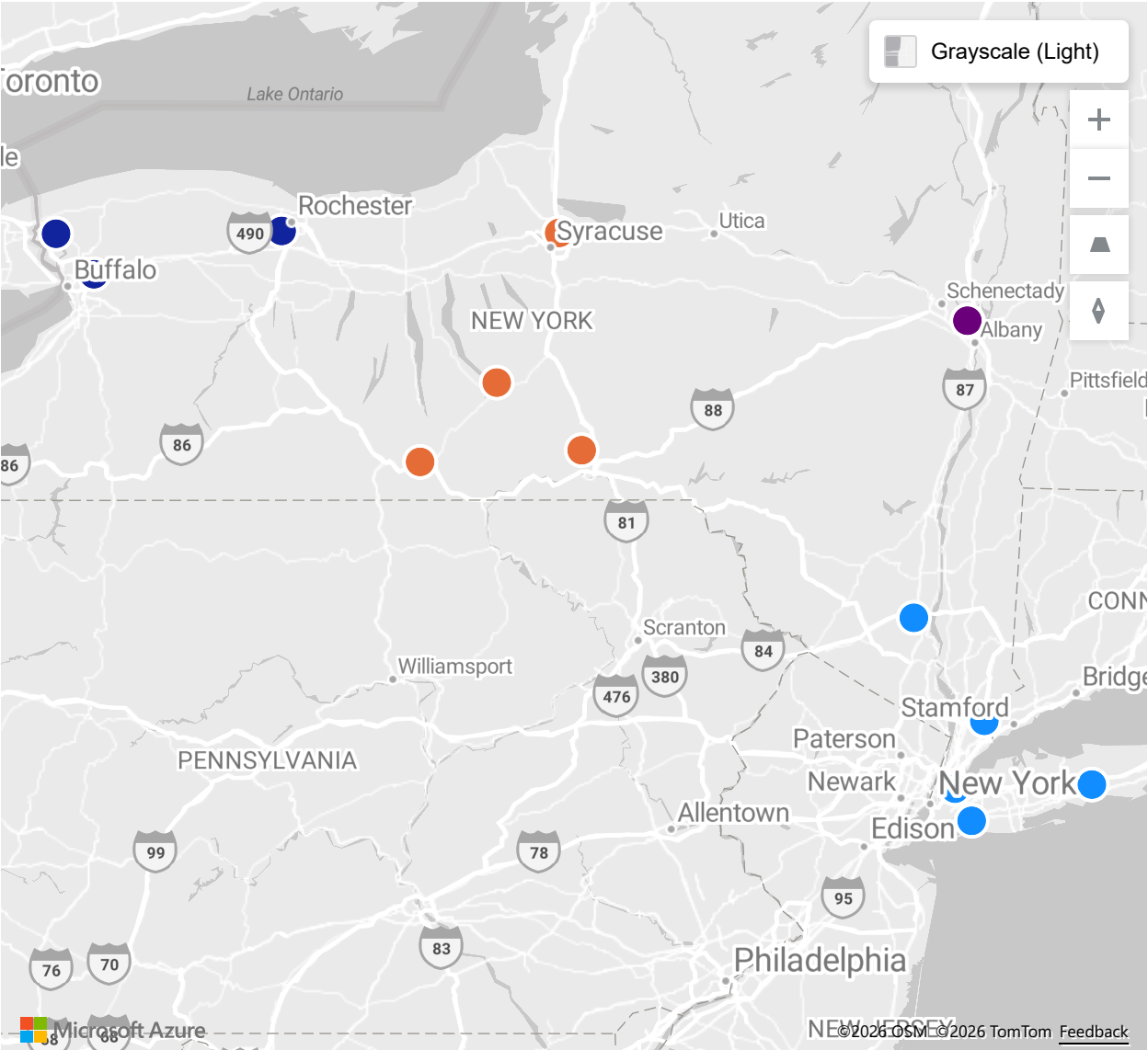
Predicted Passenger Score (y\*) = 6.12

Linear Regression Prediction

# Clusters for Airport Management

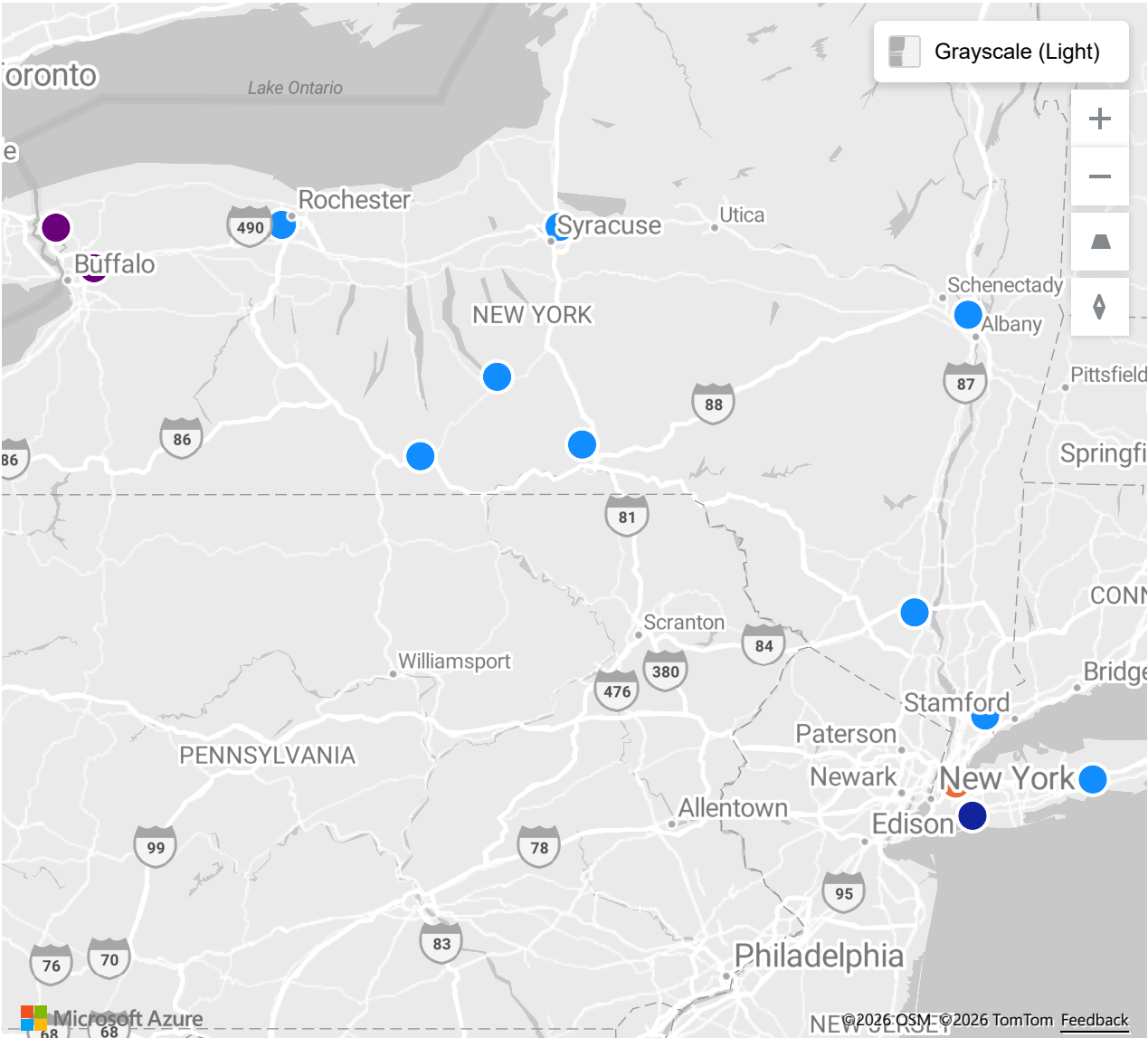
Geographic-based Clusters

Location Clusters    ● Cluster1    ● Cluster2    ● Cluster3    ● Cluster4

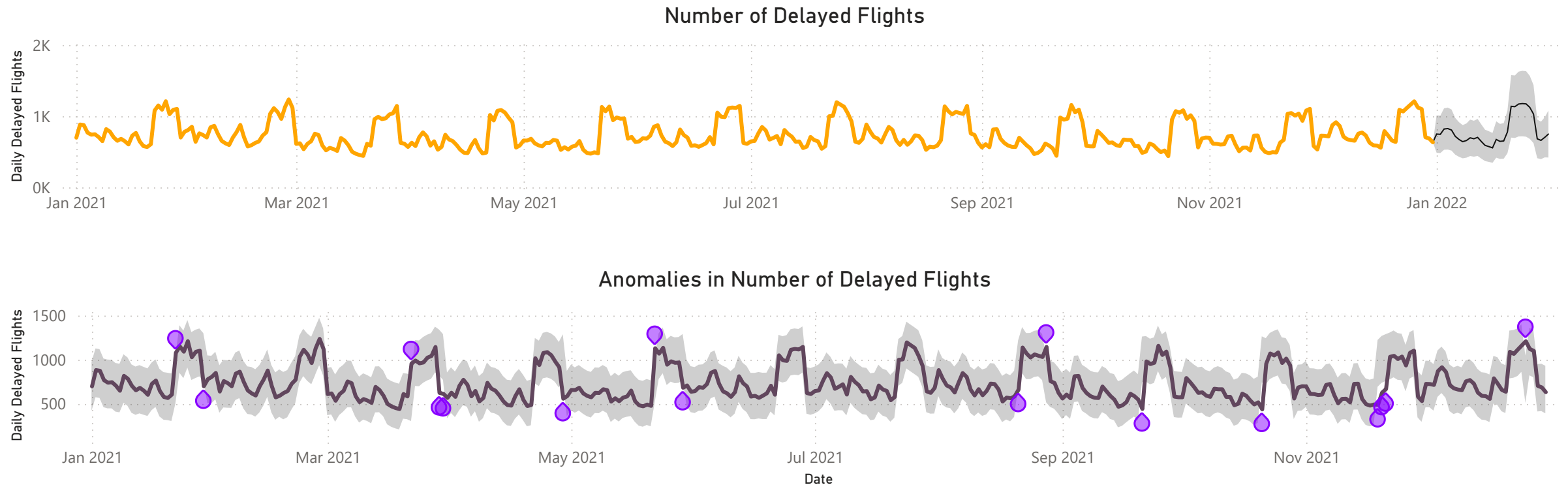


Volume-based Clusters

Volume Clusters    ● Cluster1    ● Cluster2    ● Cluster3    ● Cluster4



## Seasonality forecasting



Over the course of the year, daily delayed flights saw an overall downward trend, decreasing by 9.31% between Jan 1 and Dec 31.

However, the anomaly detection model identified several extreme fluctuations throughout the year.

The most significant outliers occurred on Dec 26, 2021, which saw a massive spike of 1,209 delayed flights and low unusually low delay on Sep 21, 2021 and Oct 21, 2021 with a range of 430 - 440 delayed flights.

# Flight Prediction

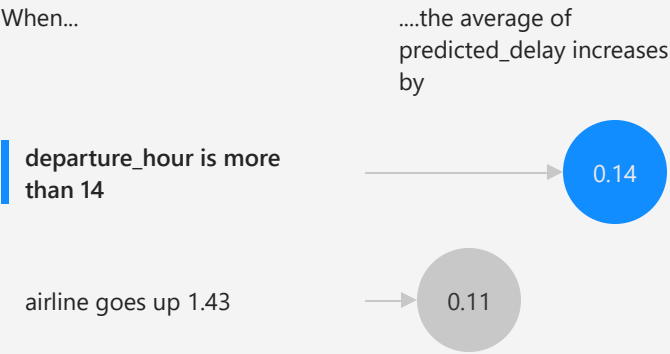
Key influencers

Top segments

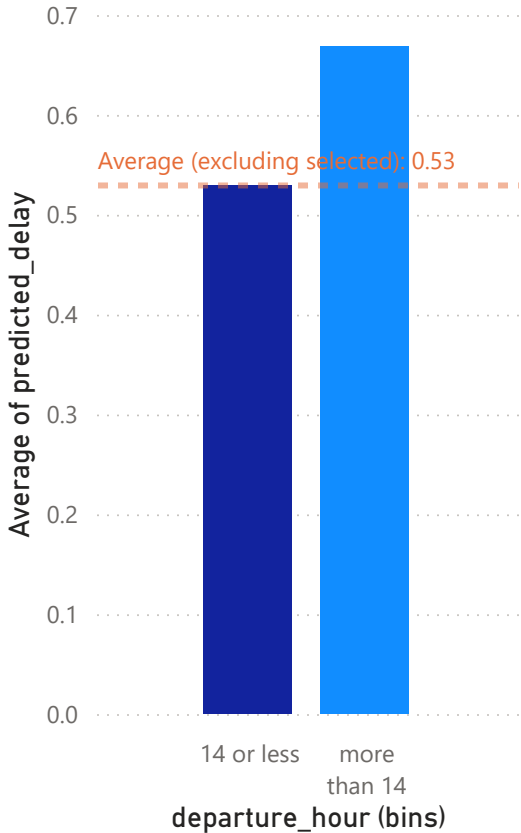
What influences predicted\_delay to 

Increase

 ?



← predicted\_delay is more likely to increase when departure\_hour is more than 14 than otherwise (on average).



☐ Only show values that are influencers

| airline | airport | predicted_delay | departure_hour | airline                    |
|---------|---------|-----------------|----------------|----------------------------|
| 0       | JFK     | 0.37            | 0              | Unique Air Lines Inc.      |
| 2       | ISP     | 0.37            | 0              | United States Airways Inc. |
| 2       | JFK     | 0.37            | 0              | United States Airways Inc. |
| 0       | ISP     | 0.38            | 1              | Unique Air Lines Inc.      |
| 0       | JFK     | 0.38            | 1              | Unique Air Lines Inc.      |
| 2       | ISP     | 0.38            | 1              | United States Airways Inc. |
| 0       | ISP     | 0.39            | 2              | Unique Air Lines Inc.      |
| 0       | JFK     | 0.39            | 2              | Unique Air Lines Inc.      |
| 2       | LGA     | 0.39            | 2              | United States Airways Inc. |
| 0       | ISP     | 0.39            | 3              | Unique Air Lines Inc.      |
| 0       | JFK     | 0.39            | 3              | Unique Air Lines Inc.      |
| 2       | ISP     | 0.39            | 3              | United States Airways Inc. |
| 2       | JFK     | 0.39            | 3              | United States Airways Inc. |
| 2       | LGA     | 0.39            | 3              | United States Airways Inc. |
| 2       | ISP     | 0.40            | 4              | United States Airways Inc. |
| 2       | JFK     | 0.40            | 4              | United States Airways Inc. |
| 2       | LGA     | 0.40            | 4              | United States Airways Inc. |
| 0       | JFK     | 0.41            | 5              | Unique Air Lines Inc.      |
| 2       | ISP     | 0.41            | 5              | United States Airways Inc. |
| 2       | JFK     | 0.41            | 5              | United States Airways Inc. |
| 2       | ISP     | 0.42            | 6              | United States Airways Inc. |
| 2       | JFK     | 0.42            | 6              | United States Airways Inc. |
| 0       | JFK     | 0.42            | 7              | Unique Air Lines Inc.      |
| 0       | LGA     | 0.42            | 7              | Unique Air Lines Inc.      |
| 2       | ISP     | 0.42            | 7              | United States Airways Inc. |
| 0       | JFK     | 0.44            | 9              | Unique Air Lines Inc.      |
| 0       | JFK     | 0.45            | 10             | Unique Air Lines Inc.      |
| 0       | LGA     | 0.45            | 10             | Unique Air Lines Inc.      |
| 2       | ISP     | 0.45            | 10             | United States Airways Inc. |

# Conclusions

---

**1. Key Influencers and Linear Regression: Which variable affects passenger score most? What would be the needed value of the independent variable to get a passenger score higher than 6.1 in the season that you chose to analyze?**

The most impactful variable overall is **Flight Status** (delayed flights drop the score by 1.14), while **Take-off delay** is the strongest numerical factor. To achieve a satisfaction score higher than 6.1 during the Summer season, take-off delays must be kept under **7.5 minutes**. The regression model shows a 7-minute delay yields a passing 6.12 score, which drops to a failing 6.08 at the 8-minute mark.

**2. Clustering: Given the two scenarios, what would be the best management strategy, geographic-based clustering or size-based clustering?**

I recommend moving forward with **Location-based strategy** since it satisfies the constraint of 5 maximum airport per office unlike the other strategy. Also, it keeps the airports close meaning managers can make their weekly on-site visits without driving across the entire state.

**3. Forecasting: What is the pattern shown in the time-series? How does the forecast look for the next month?**

The historical time-series reveals a highly volatile, **repeating weekly cyclical pattern** throughout the year, despite a slight overall downward trend. Looking ahead to the next month, the forecast predicts **this regular volatility will continue**. The model anticipates an initial spike in flight delays in early January, followed by a return to the established weekly rhythm, remaining safely within historical confidence intervals.

**4. Flight Prediction: What flights would you recommend the passenger to take during those dates?**

Based on the prediction model, the probability of a flight delay is most heavily influenced by late departures, specifically flights departing after 14:00. To minimize the risk of delay, I recommend booking one of the following flights, which all share the lowest predicted delay probability of 37%:

- **0:00** (Midnight) departure from **JFK** via Unique Air Lines Inc. (**Airline 0**).
- **0:00** (Midnight) departure from **ISP** via United States Airways (**Airline 2**).
- **0:00** (Midnight) departure from **JFK** via United States Airways. (**Airline 2**).