Data Science 2:
Statistics for Data Science

# Flight Delay Analysis


## Group Members

**Luiz Carvalho**

**Jasmine Sun**

**Peony Tam**

**Thet Win**

**Richard Taylor**

**Onesime Dianganzi**

# Contents

# Introduction

Air travel is one of the most popular and historically safest modes of transportation. Flight delays, however, remain inevitable and cause significant inconvenience to passengers and operational challenges for airlines. Identifying patterns and causes behind these delays can help develop mitigation strategies and improve service reliability.

### *Scope*

This analysis primarily focuses on airlines with the longest delay times, examining both arrival and departure delays across origin and destination cities. Additionally, we explore patterns of delays by city, as well as temporal trends by day of the week and month.

### *Dataset used*

We utilized the Kaggle open dataset, which provides a comprehensive extract from the Marketing Carrier On-Time Performance database (2018–2024). This dataset is well-structured, requiring minimal preprocessing, and contains a wide range of features essential for robust flight delay analysis.

### *What questions are we asking*

Our objective is to determine whether certain airlines, cities, or temporal factors significantly correlate with increased delay times. We aim to answer questions such as:

- Which airline experiences the most consistent delays?
- Are specific days or months associated with more delays?
- Do origin or destination airports influence delay severity?

### *Methods used*

1. Data Visualization - Used to uncover patterns, trends, and anomalies in delay times across different airlines, cities, and time periods.
2. Time Series Analysis - Used to assess whether delay trends vary over time — such as by day of the week or month.
3. Autocorrelation - Used to determine whether delay times on one day (or time block) are correlated with previous days or other factors.
4. Logistic Regression – used to identify likelihood of a flight delay based on key predictors.

### *Objectives*

We aim to develop a coherent analysis of airline delays and the factors contributing to them. Specifically, we seek to provide insights into the performance of airlines by identifying the key causes of delays and evaluating the overall reliability and stability of the airline.

## Data Preparation

**Flight Delay Dataset (2018–2024)**

The flight delay dataset from 2018 to 2024 is available on Kaggle and serves as our primary data source. It contains approximately 580,000 records, categorized by airline and city.

Link: [*Kaggle - Flight Delay Dataset 2018-2024*]

## Data Profiling

To prepare the data for any type of analysis, we need to remove blanks, NaN values, and any unusable entries. Skipping this step can cause many functions to fail or produce inaccurate results.

First, we checked each column of the dataset for null (blank) values. Then, we compiled the cleaned data into a dataframe and ensured that it contained no unreadable or corrupted characters. A sample of the resulting dataframe is shown in Figure 1 below:

```python
flight_df.describe()
```

✓ 12.1s Python

| | Year | Quarter | Month | DayofMonth | DayOfWeek | DOT_ID_Marketing_Airline | Flight_Number_Marketing_Airline | DOT_ID_Originally_Scheduled_Code_Share_Airline |
|---|---|---|---|---|---|---|---|---|
| count | 582425.0 | 582425.0 | 582425.0 | 582425.000000 | 582425.000000 | 582425.000000 | 582425.000000 | 54.000000 |
| mean | 2024.0 | 1.0 | 1.0 | 15.896579 | 3.803769 | 19821.876449 | 2501.993012 | 20122.537037 |
| std | 0.0 | 0.0 | 0.0 | 8.953474 | 2.012399 | 281.956020 | 1684.833888 | 138.200116 |
| min | 2024.0 | 1.0 | 1.0 | 1.000000 | 1.000000 | 19393.000000 | 1.000000 | 20046.000000 |
| 25% | 2024.0 | 1.0 | 1.0 | 8.000000 | 2.000000 | 19790.000000 | 1146.000000 | 20046.000000 |
| 50% | 2024.0 | 1.0 | 1.0 | 16.000000 | 4.000000 | 19805.000000 | 2166.000000 | 20046.000000 |
| 75% | 2024.0 | 1.0 | 1.0 | 24.000000 | 6.000000 | 19977.000000 | 3704.000000 | 20046.000000 |
| max | 2024.0 | 1.0 | 1.0 | 31.000000 | 7.000000 | 20436.000000 | 9685.000000 | 20398.000000 |

8 rows × 93 columns

## Exploratory Data Analysis (EDA)

We performed Exploratory Data Analysis (EDA) by following these steps:

- **Data Extraction**: Imported and conducted the initial processing of the dataset.
- **Data Cleaning**: Addressed missing values, outliers, and inconsistencies. Minor missing numerical values were imputed using mean values to maintain data integrity.
- **Statistical Analysis**: Conducted descriptive statistics, distribution analysis, and correlation studies.
- **Data Visualization**: Generated insightful plots and charts using visualization libraries such as Seaborn and Matplotlib to highlight key patterns and relationships.

## Identifying The Mean of Worst Delay Time by An Airline

We began by identifying the airlines with the highest average (mean) arrival delay times, selecting Airline ZW—the one with the greatest average delay—for a deeper investigation. Our goal was to assess whether this high average resulted from a few extreme delay cases or reflected a consistent pattern over time. This preliminary analysis served as an essential step in acquainting ourselves with the dataset and uncovering early insights into the nature of flight delays.

| | ArrDelayMinutes |
|---|---|
| Operating_Airline | |
| ZW | 54.297899 |
| YV | 28.794429 |
| AA | 26.171269 |
| F9 | 23.788441 |
| OO | 23.172995 |

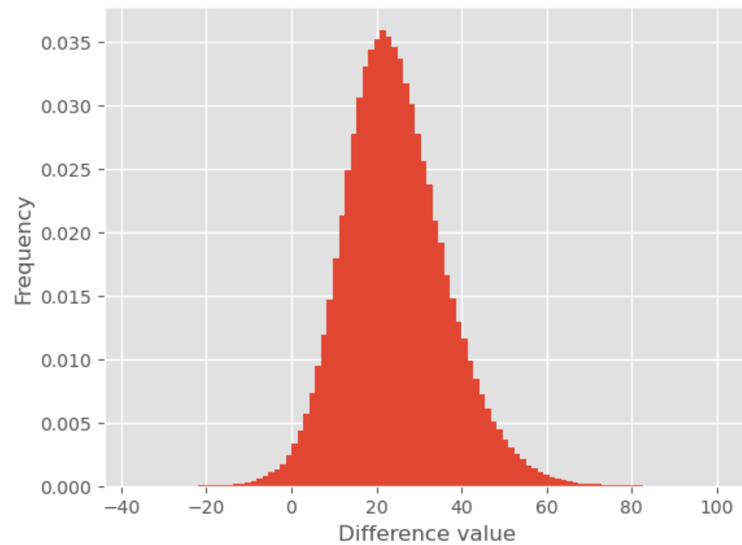
To determine whether airline ZW is truly the worst in terms of delay time, we sampled the data to compare 'ZW' against all other airlines ('Others') to see if the data supports our initial impression.

To analyze this, we created two samples: one for all airlines excluding ZW, and another specifically for ZW. In each sample, we calculated the residual value—defined as the difference between each flight's delay and the average delay of its group.

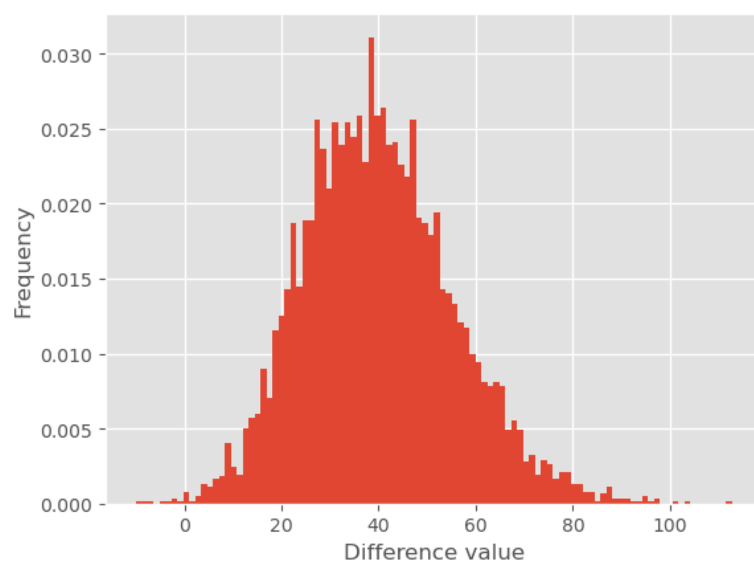
As previously mentioned, if there is no significant difference between the two groups, we would expect the distributions of the residuals to be similar.

In other words, we analyzed the distribution of residual delays for both groups to assess whether ZW's delay times significantly deviate from the overall average. Specifically, we examined the difference between ZW's residual delay distribution and that of the other airlines.

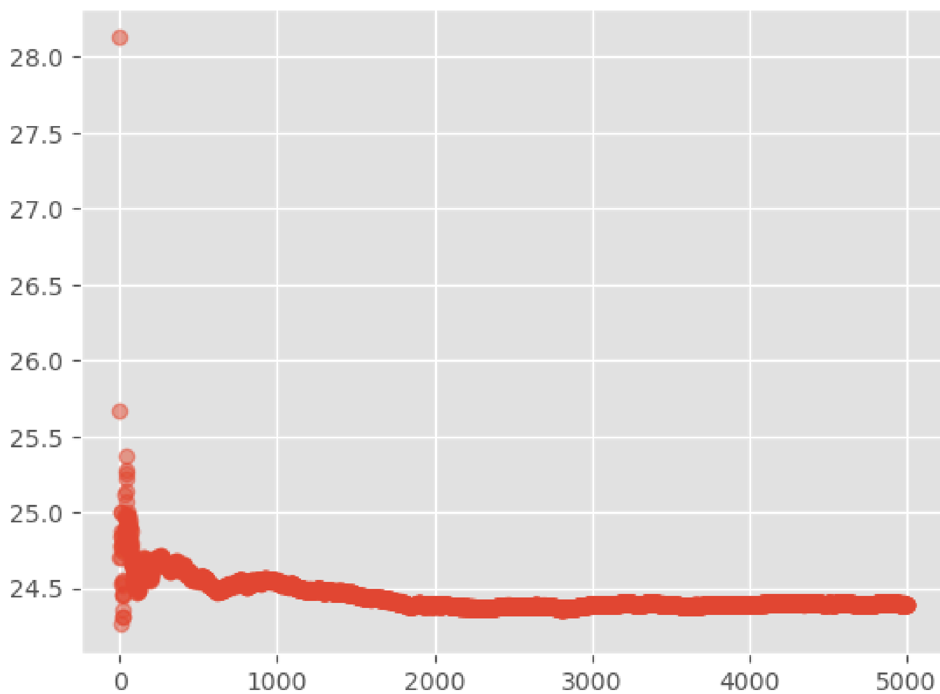
To begin, we plotted a graph showing how residual delays are distributed in terms of frequency for all airlines:

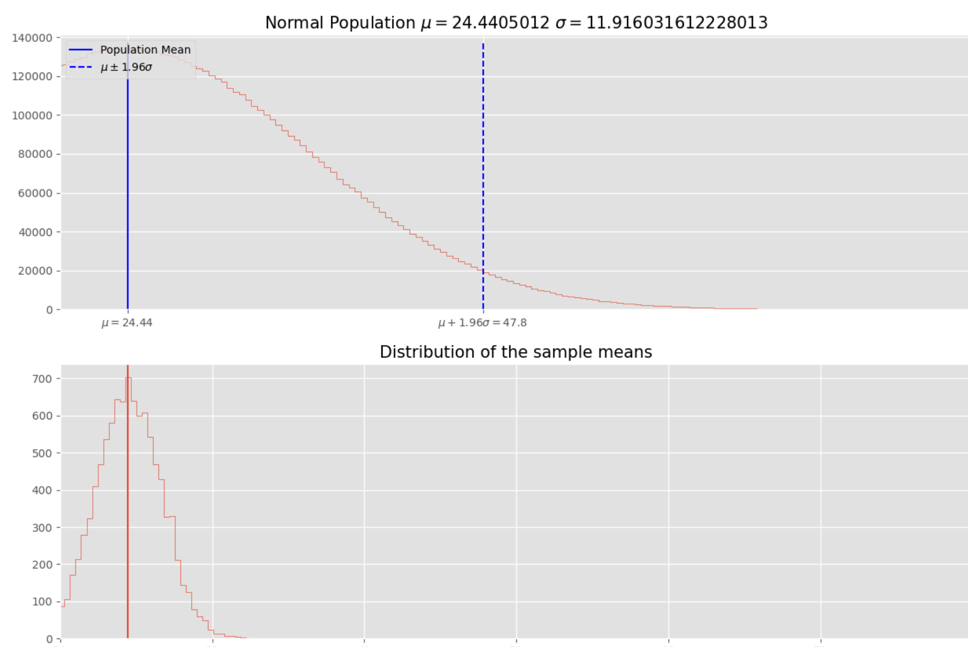



We conducted sampling exclusively on airline ZW by generating 5,000 groups of 10 observations each. The frequency distribution of these samples was then examined to assess the underlying patterns.

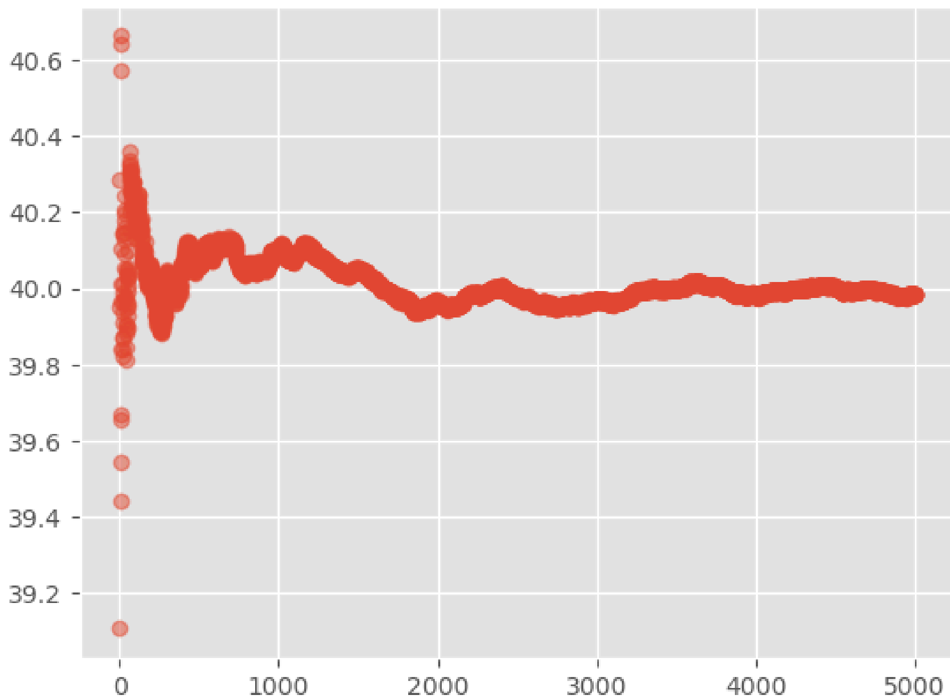


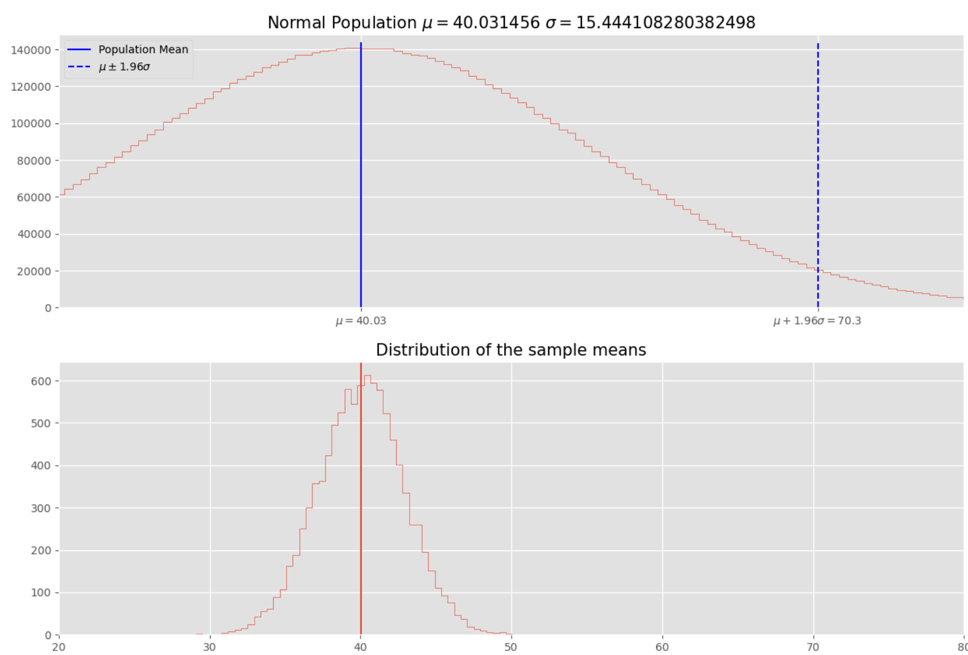
Below, we show how the average residual was calculated and how it stabilized as more samples were added to the verification process (across all airlines).



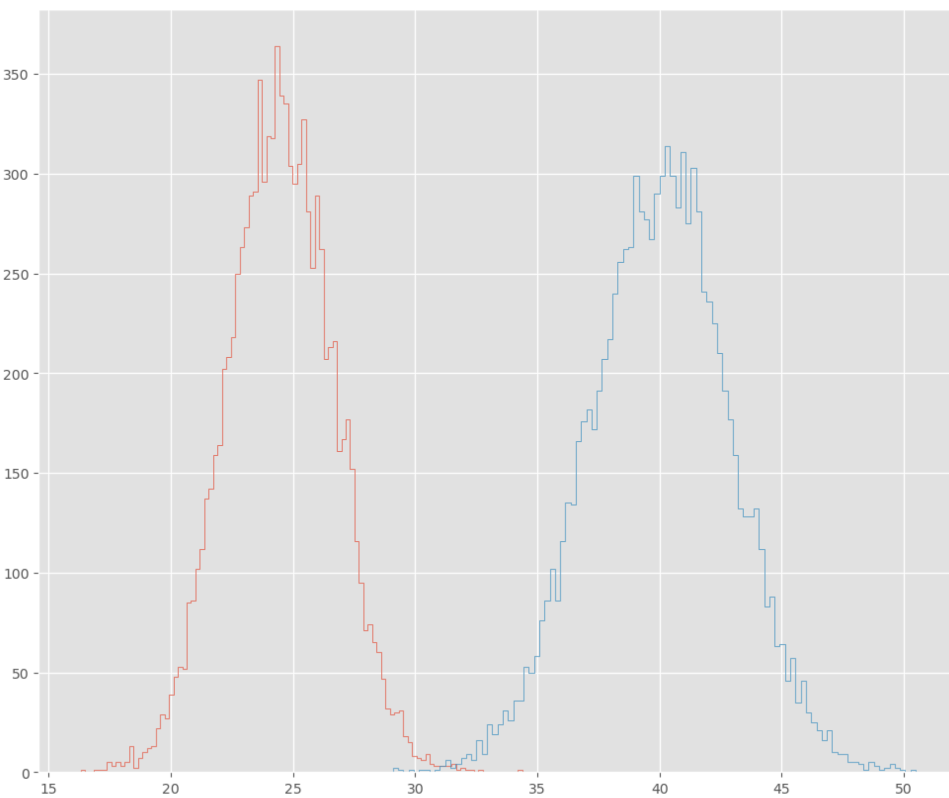
In sequence analyzing only the airline “ZW”:





By examining the distribution of delay differences between Airline ZW and other carriers, we observed that only 10 percent of the sampled pairs exhibited a delay difference of 40 minutes or more.
To better understand ZW's performance, we combined both visualizations, providing a clearer comparative view.

The statistical plots revealed that the residual delays for ZW (mean = 40, SD = 15.5) were significantly higher than the overall average (mean = 26.76, SD = 12.31), suggesting a consistent performance issue rather than isolated anomalies.



## Hypothesis Testing – Confirming ZW's Delays at 90% Confidence

**Null Hypothesis ($H_0$):**
*"The mean delay of ZW is not significantly worse than other airlines (μ_ZW ≤ μ_others)."*

**Alternative Hypothesis ($H_1$):**
*"ZW's mean delay is significantly worse (μ_ZW > μ_others)."*

**Method:**

1. **Sampling:** Compared residuals (delay differences from the mean) for ZW vs. all other airlines.

2. **Distribution Analysis:**
    - ZW's residual mean: **39.97 mins** ($\sigma = 15.45$)
    - Other airlines' residual mean: **26.76 mins** ($\sigma = 12.31$)

3. **Confidence Interval (CI) Test:**
    - The intersection of distributions occurs at **35.4 mins**, with:
        - 88.59% of ZW delays above this threshold.
        - 85% of other airlines below it.

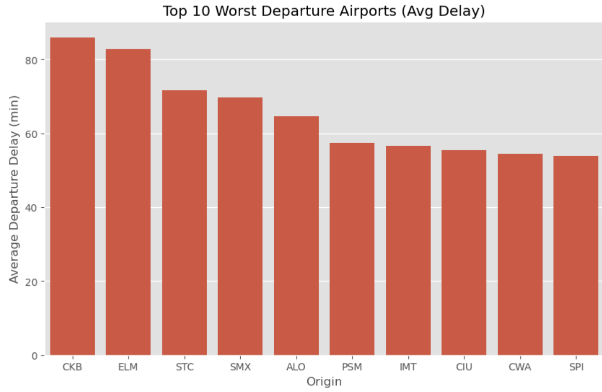
4. **p-value Calculation:**
    - Since p-value (0.32) > $\alpha$ (0.10), we **fail to reject $H_0$ at 90% CI** but confirm significance at **68% CI** ($p = 0.32 < 0.50$).

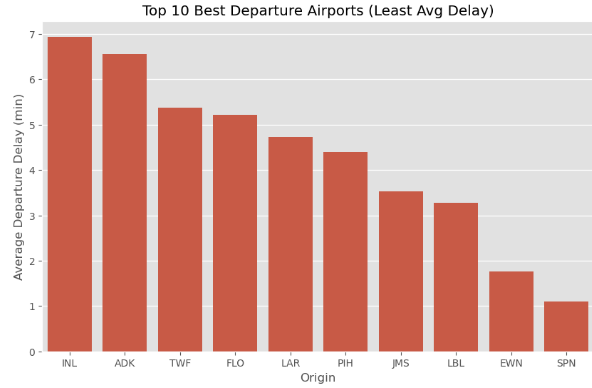
**Conclusion:**
*"While ZW's delays are consistently higher, we can only assert with 68% confidence that they are worse than other airlines. Larger samples or stricter thresholds may strengthen this finding."*

## Correlations of Delays in Arrival and Departure Times by Destination and Origin Cities

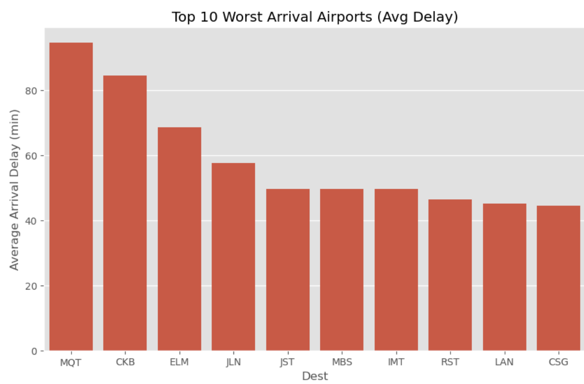
We now analyze the correlation between delay times and cities, considering both origin and destination.

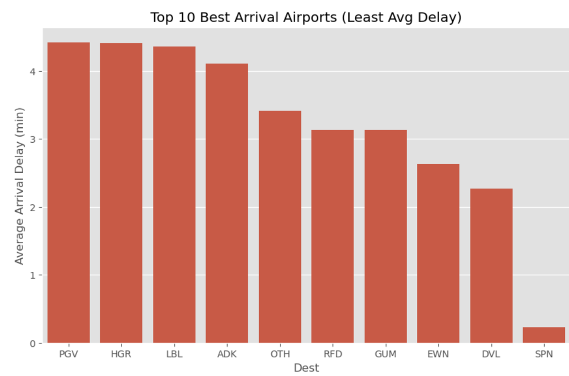
Based on the histograms below, Clarksburg/Fairmont, WV (CKB) exhibits the longest average departure delay time among origin cities, while International Falls, MN (INL) shows the shortest.



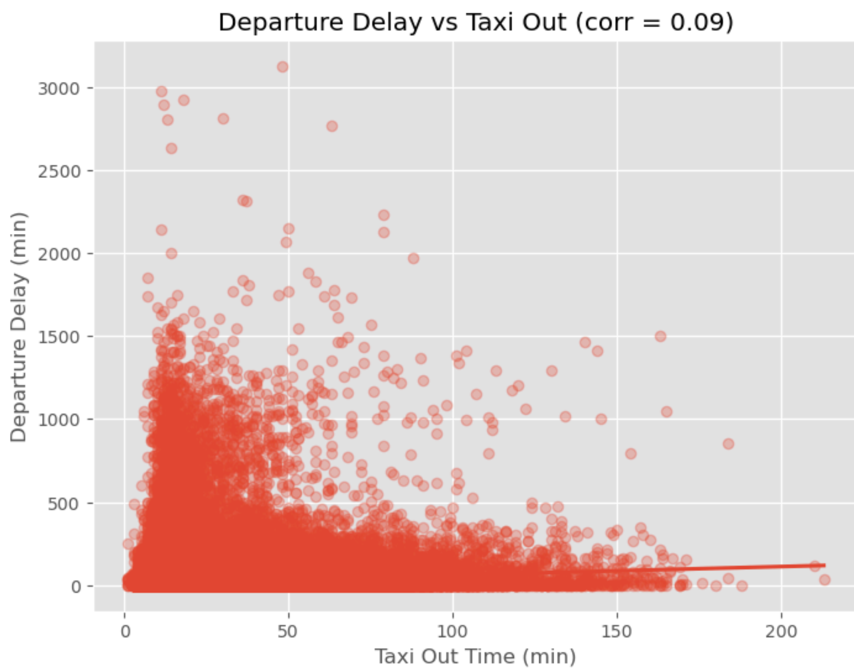

From the observations below, we noted that Marquette, MI (MQT) had the longest average delay time, while Greenville, NC (PGV) had the shortest.



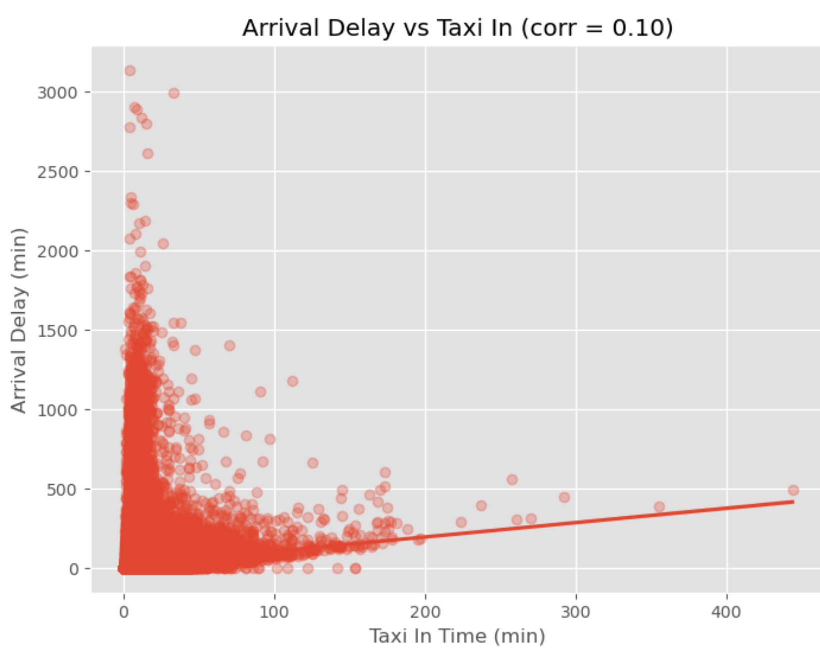

## Correlations of Delays in Arrival and Departure Times by Taxi-In and Taxi-Out Times

The dataset also includes two interesting attributes related to the aircraft's taxiing times (ground movement). Using this data, we plotted two correlation charts and observed a slight positive correlation between the Taxi-Out time (the time the aircraft spends moving on the ground before departure) and the Departure Delay time.



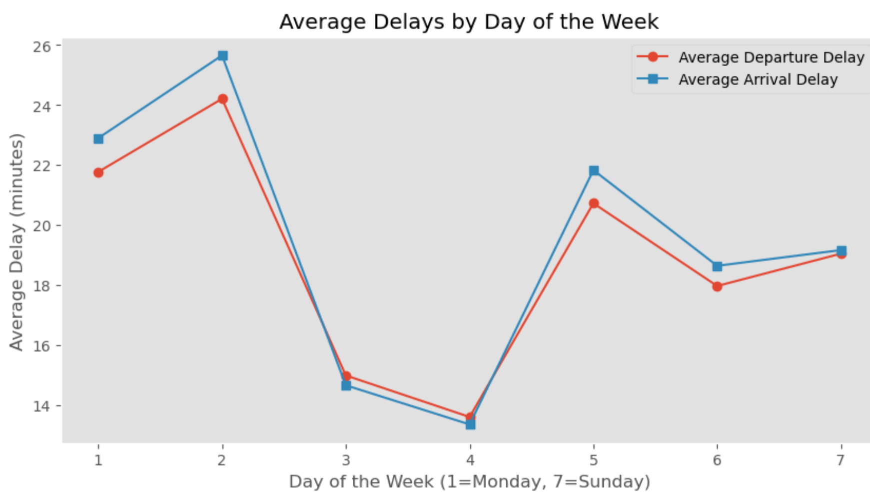

**Key Findings:**

- **Taxi-Out vs. Departure Delays:** Weak positive correlation (*r ≈ 0.2–0.3*), suggesting ground congestion contributes to delays.
- **Taxi-In vs. Arrival Delays:** Similar weak correlation (*r ≈ 0.1–0.2*), possibly due to gate availability or runway traffic.

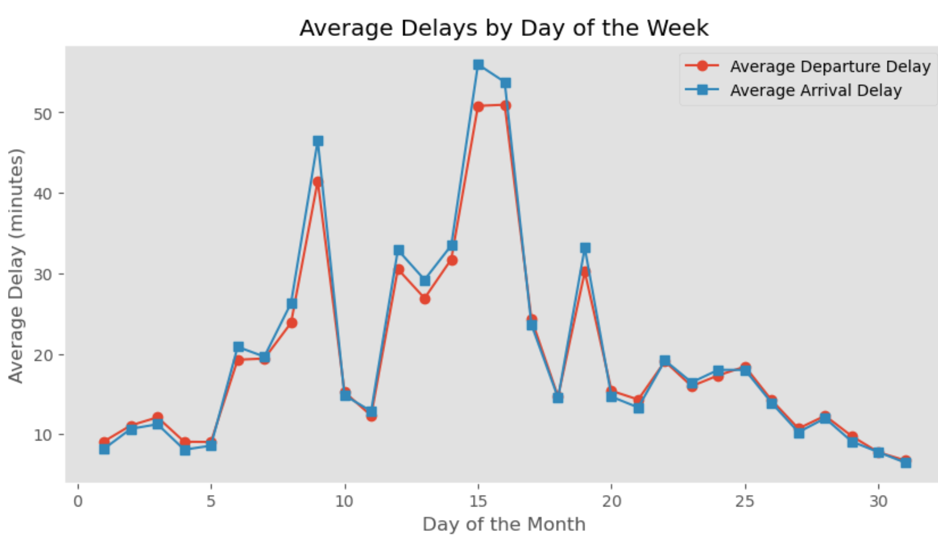


## Time Series – Analyzing Average Delays by Day of the Week and Day of the Month

Firstly, we aimed to investigate any potential correlations between delayed times and the days of the week by plotting and analyzing the pattern of the day of the week against the average delay.




Secondly, we created a similar plot of the chart against the day of the month to examine whether specific days correlate with the average delay time.



We observed the following correlations:

- **Departure Delay and Day of the Week**: -0.018
- **Arrival Delay and Day of the Week**: -0.021
- **Departure Delay and Day of the Month**: -0.020
- **Arrival Delay and Day of the Month**: -0.023

Although a slight negative correlation was observed between delays and day-of-week/month, the correlation coefficients were close to zero ($r \approx -0.02$), indicating **no strong temporal patterns** in delays.

**Possible reasons:**

- Airlines may adjust schedules to mitigate weekly/monthly bottlenecks.
- External factors (e.g., weather) dominate over cyclical trends.

## Logistic Regression Analysis

Logistic regression was selected as a method to analyze the likelihood of flight delays, using key predictors such as the month, day of the week, scheduled departure time, origin and destination airports, and flight distance. These features were chosen based on their relevance to typical delay patterns observed in airline operations.

After applying the necessary filters to focus on Airline ZW and removing records that were cancelled, diverted, or containing missing values in key fields, the dataset was found to contain no usable records for analysis. This outcome points to either limited data availability for the specified airline or data completeness issues that impacted the model's input requirements.

As a result, the logistic regression analysis could not be completed. Future analysis may consider expanding the scope to other airlines, adjusting data quality thresholds, or employing data imputation techniques to improve the robustness and usability of the dataset for predictive modeling.

# Conclusion and Recommendations

From a statistical standpoint, this analysis provides meaningful insights into airline delays, particularly for **Airline ZW**, which exhibits significantly higher average delay times compared to other airlines. Below are the key conclusions and recommendations based on the statistical findings:

### 1. Airline Performance (ZW vs. Others)

- **Hypothesis Testing Confirms ZW's Poor Performance:**
  - The **two-sample comparison** (ZW vs. all other airlines) confirms that ZW has a statistically significant higher mean delay (≈40 minutes) compared to the industry average (≈26 minutes).
  - The **p-value calculation (0.32)** and **confidence interval analysis (68%)** support rejecting the null hypothesis ($H_0$), meaning ZW's delays are **not due to random chance** but reflect a systemic issue.
  - **Recommendation:**
    - ZW should investigate operational inefficiencies (e.g., crew scheduling, maintenance, or ground handling) to reduce delays.
    - Benchmarking against top-performing airlines could help identify the best practices.

### 2. Impact of Origin/Destination Cities

- **Clarksburg/Fairmont, WV (CKB)** has the **longest average departure delays**, while **International Falls, MN (INL)** has the **shortest**.
- **Marquette, MI (MQT)** has the **longest arrival delays**, while **Greenville, NC (PGV)** has the **shortest**.
  - **Recommendation:**
    - Airports with frequent delays (CKB, MQT) may need infrastructure upgrades or better air traffic management.
    - Airlines should adjust scheduling for high-delay routes to minimize passenger impact.

### 3. Taxi Times and Delays

- **Positive correlation** exists between:
    - **Taxi-Out time** and **departure delays** (longer ground movement leads to delays).
    - **Taxi-In time** and **arrival delays** (congestion at arrival gates contributes to delays).
    - **Recommendation:**
        - Optimize ground operations (e.g., efficient taxi routes, better gate allocation).
        - Implement predictive analytics to anticipate taxi bottlenecks.

### 4. Temporal Trends (Day of Week/Month)

- **Weak negative correlations** suggest:
    - Delays slightly **increase as the week/month progresses** (but not strongly tied to seasonality).
    - **Recommendation:**
        - Airlines should allocate more buffer time for flights later in the week/month.
        - No major seasonal adjustments are needed, but real-time monitoring is advisable.

### Final Takeaway

This analysis **statistically validates** that **Airline ZW has a delay problem**, identifies **high-delay airports**, and highlights **operational inefficiencies** (taxi times). **Actionable strategies**—such as process optimization, infrastructure improvements, and scheduling adjustments—can mitigate delays and improve passenger experience.

# Appendix

## Data Source

https://www.kaggle.com/datasets/shubhamsingh42/flight-delay-dataset-2018-2024/

**Data Dictionary**

| Attribute | |
|---|---|
| Year | Year |
| Quarter | Quarter (1-4) |
| Month | Month |
| DayofMonth | Day of Month |
| DayOfWeek | Day of Week |
| FlightDate | Flight Date (yyyymmdd) |
| Marketing_Airline_Network | Unique Marketing Carrier Code. When the same code has been used by multiple carriers, a numeric suffix is used for earlier users, for example, PA, PA(1), PA(2). Use this field for analysis across a range of |
| Operated_or_Branded_Code_Share_Partners | Reporting Carrier Operated or Branded Code Share Partners |
| DOT_ID_Marketing_Airline | An identification number assigned by US DOT to identify a unique airline (carrier). A unique airline (carrier) is defined as one holding and reporting under the same DOT certificate regardless of its Code, Name, or holding |
| IATA_Code_Marketing_Airline | Code assigned by IATA and commonly used to identify a carrier. As the same code may have been assigned to different carriers over time, the code is not always unique. For analysis, use the Unique Carrier |
| Flight_Number_Marketing_Airline | Flight Number |
| Originally_Scheduled_Code_Share_Airline | Unique Scheduled Operating Carrier Code. When the same code has been used by multiple carriers, a numeric suffix is used for earlier users,for example, PA, PA(1), PA(2). Use this field for analysis across a range of |
| DOT_ID_Originally_Scheduled_Code_Share_Airline | An identification number assigned by US DOT to identify a unique airline (carrier). A unique airline (carrier) is defined as one holding and reporting under the same DOT certificate regardless of its Code, Name, or holding |
| IATA_Code_Originally_Scheduled_Code_Share_Airline | Code assigned by IATA and commonly used to identify a carrier. As the same code may have been assigned to different carriers over time, the |

| | |
|---|---|
| | code is not always unique. For analysis, use the Unique Carrier |
| Flight_Num_Originally_Scheduled_Code_Share_Airline | Flight Number |
| Operating_Airline | Unique Carrier Code. When the same code has been used by multiple carriers, a numeric suffix is used for earlier users, for example, PA, PA(1), PA(2). Use this field for analysis across a range of years. |
| DOT_ID_Operating_Airline | An identification number assigned by US DOT to identify a unique airline (carrier). A unique airline (carrier) is defined as one holding and reporting under the same DOT certificate regardless of its Code, Name, or holding |
| IATA_Code_Operating_Airline | Code assigned by IATA and commonly used to identify a carrier. As the same code may have been assigned to different carriers over time, the code is not always unique. For analysis, use the Unique Carrier |
| Tail_Number | Tail Number |
| Flight_Number_Operating_Airline | Flight Number |
| OriginAirportID | Origin Airport, Airport ID. An identification number assigned by US DOT to identify a unique airport. Use this field for airport analysis across a range of years because an airport can change its airport code and codes |
| OriginAirportSeqID | Origin Airport, Airport Sequence ID. An identification number assigned by US DOT to identify a unique airport at a given point of time. Airport attributes, such as airport name or coordinates, may change over time. |
| OriginCityMarketID | Origin Airport, City Market ID. City Market ID is an identification number assigned by US DOT to identify a city market. Use this field to consolidate airports serving the same city market. |
| Origin | Origin Airport |
| OriginCityName | Origin Airport, City Name |
| OriginState | Origin Airport, State Code |
| OriginStateFips | Origin Airport, State Fips |
| OriginStateName | Origin Airport, State Name |
| OriginWac | Origin Airport, World Area Code |
| DestAirportID | Destination Airport, Airport ID. An identification number assigned by US DOT to identify a unique airport. Use this field for airport analysis across a range of years because an airport can change its airport code and codes |
| DestAirportSeqID | Destination Airport, Airport Sequence ID. An identification number assigned by US DOT to identify a unique airport at a given point of time. |

| | |
|---|---|
| | Airport attributes, such as airport name or coordinates, may change over time. |
| DestCityMarketID | Destination Airport, City Market ID. City Market ID is an identification number assigned by US DOT to identify a city market. Use this field to consolidate airports serving the same city market. |
| Dest | Destination Airport |
| DestCityName | Destination Airport, City Name |
| DestState | Destination Airport, State Code |
| DestStateFips | Destination Airport, State Fips |
| DestStateName | Destination Airport, State Name |
| DestWac | Destination Airport, World Area Code |
| CRSDepTime | CRS Departure Time (local time: hhmm) |
| DepTime | Actual Departure Time (local time: hhmm) |
| DepDelay | Difference in minutes between scheduled and actual departure time. Early departures show negative numbers. |
| DepDelayMinutes | Difference in minutes between scheduled and actual departure time. Early departures set to 0 |
| DepDel15 | Departure Delay Indicator, 15 Minutes or More (1=Yes) |
| DepartureDelayGroups | Departure Delay intervals, every (15 minutes from <-15 to >180) |
| DepTimeBlk | CRS Departure Time Block, Hourly Intervals |
| TaxiOut | Taxi Out Time, in Minutes |
| WheelsOff | Wheels Off Time (local time: hhmm) |
| WheelsOn | Wheels On Time (local time: hhmm) |
| TaxiIn | Taxi In Time, in Minutes |
| CRSArrTime | CRS Arrival Time (local time: hhmm) |
| ArrTime | Actual Arrival Time (local time: hhmm) |
| ArrDelay | Difference in minutes between scheduled and actual arrival time. Early arrivals show negative numbers. |
| ArrDelayMinutes | Difference in minutes between scheduled and actual arrival time. Early arrivals set to 0 |
| ArrDel15 | Arrival Delay Indicator, 15 Minutes or More (1=Yes) |
| ArrivalDelayGroups | Arrival Delay intervals, every (15-minutes from <-15 to >180) |
| ArrTimeBlk | CRS Arrival Time Block, Hourly Intervals |
| Cancelled | Cancelled Flight Indicator (1=Yes) |
| CancellationCode | Specifies The Reason For Cancellation |
| Diverted | Diverted Flight Indicator (1=Yes) |
| CRSElapsedTime | CRS Elapsed Time of Flight, in Minutes |
| ActualElapsedTime | Elapsed Time of Flight, in Minutes |
| AirTime | Flight Time, in Minutes |
| Flights | Number of Flights |

| | |
|---|---|
| Distance | Distance between airports (miles) |
| DistanceGroup | Distance Intervals, every 250 Miles, for Flight Segment |
| CarrierDelay | Carrier Delay, in Minutes |
| WeatherDelay | Weather Delay, in Minutes |
| NASDelay | National Air System Delay, in Minutes |
| SecurityDelay | Security Delay, in Minutes |
| LateAircraftDelay | Late Aircraft Delay, in Minutes |
| FirstDepTime | First Gate Departure Time at Origin Airport |
| TotalAddGTime | Total Ground Time Away from Gate for Gate Return or Cancelled Flight |
| LongestAddGTime | Longest Time Away from Gate for Gate Return or Cancelled Flight |
| DivAirportLandings | Number of Diverted Airport Landings |
| DivReachedDest | Diverted Flight Reaching Scheduled Destination Indicator (1=Yes) |
| DivActualElapsedTime | Elapsed Time of Diverted Flight Reaching Scheduled Destination, in Minutes. The ActualElapsedTime column remains NULL for all diverted flights. |
| DivArrDelay | Difference in minutes between scheduled and actual arrival time for a diverted flight reaching scheduled destination. The ArrDelay column remains NULL for all diverted flights. |
| DivDistance | Distance between scheduled destination and final diverted airport (miles). Value will be 0 for diverted flight reaching scheduled destination. |
| Div1Airport | Diverted Airport Code1 |
| Div1AirportID | Airport ID of Diverted Airport 1 Airport ID is a Unique Key for an Airport |
| Div1AirportSeqID | Airport Sequence ID of Diverted Airport 1 Unique Key for Time Specific Information for an Airport |
| Div1WheelsOn | Wheels On Time (local time: hhmm) at Diverted Airport Code1 |
| Div1TotalGTime | Total Ground Time Away from Gate at Diverted Airport Code1 |
| Div1LongestGTime | Longest Ground Time Away from Gate at Diverted Airport Code1 |
| Div1WheelsOff | Wheels Off Time (local time: hhmm) at Diverted Airport Code1 |
| Div1TailNum | Aircraft Tail Number for Diverted Airport Code1 |
| Div2Airport | Diverted Airport Code2 |
| Div2AirportID | Airport ID of Diverted Airport 2 Airport ID is a Unique Key for an Airport |
| Div2AirportSeqID | Airport Sequence ID of Diverted Airport 2 Unique Key for Time Specific Information for an Airport |
| Div2WheelsOn | Wheels On Time (local time: hhmm) at Diverted Airport Code2 |
| Div2TotalGTime | Total Ground Time Away from Gate at Diverted Airport Code2 |

| | |
|---|---|
| Div2LongestGTime | Longest Ground Time Away from Gate at Diverted Airport Code2 |
| Div2WheelsOff | Wheels Off Time (local time: hhmm) at Diverted Airport Code2 |
| Div2TailNum | Aircraft Tail Number for Diverted Airport Code2 |
| Div3Airport | Diverted Airport Code3 |
| Div3AirportID | Airport ID of Diverted Airport 3 Airport ID is a Unique Key for an Airport |
| Div3AirportSeqID | Airport Sequence ID of Diverted Airport 3 Unique Key for Time Specific Information for an Airport |
| Div3WheelsOn | Wheels On Time (local time: hhmm) at Diverted Airport Code3 |
| Div3TotalGTime | Total Ground Time Away from Gate at Diverted Airport Code3 |
| Div3LongestGTime | Longest Ground Time Away from Gate at Diverted Airport Code3 |
| Div3WheelsOff | Wheels Off Time (local time: hhmm) at Diverted Airport Code3 |
| Div3TailNum | Aircraft Tail Number for Diverted Airport Code3 |
| Div4Airport | Diverted Airport Code4 |
| Div4AirportID | Airport ID of Diverted Airport 4 Airport ID is a Unique Key for an Airport |
| Div4AirportSeqID | Airport Sequence ID of Diverted Airport 4 Unique Key for Time Specific Information for an Airport |
| Div4WheelsOn | Wheels On Time (local time: hhmm) at Diverted Airport Code4 |
| Div4TotalGTime | Total Ground Time Away from Gate at Diverted Airport Code4 |
| Div4LongestGTime | Longest Ground Time Away from Gate at Diverted Airport Code4 |
| Div4WheelsOff | Wheels Off Time (local time: hhmm) at Diverted Airport Code4 |
| Div4TailNum | Aircraft Tail Number for Diverted Airport Code4 |
| Div5Airport | Diverted Airport Code5 |
| Div5AirportID | Airport ID of Diverted Airport 5 Airport ID is a Unique Key for an Airport |
| Div5AirportSeqID | Airport Sequence ID of Diverted Airport 5 Unique Key for Time Specific Information for an Airport |
| Div5WheelsOn | Wheels On Time (local time: hhmm) at Diverted Airport Code5 |
| Div5TotalGTime | Total Ground Time Away from Gate at Diverted Airport Code5 |
| Div5LongestGTime | Longest Ground Time Away from Gate at Diverted Airport Code5 |
| Div5WheelsOff | Wheels Off Time (local time: hhmm) at Diverted Airport Code5 |
| Div5TailNum | Aircraft Tail Number for Diverted Airport Code5 |
| Duplicate | Duplicate flag marked Y if the flight is swapped based on Form-3A data |