# WHY HAVE DELAYS INCREASED SO MUCH AT EWR IN 2022

Anilkumar Malyala

Kalyan Karre

Murali Krishna Avula

Nithin Kumar Allam

Siri Smitha Joginapally

**University of Maryland, Baltimore County (UMBC)**

**DATA 602: Introduction to Data Analysis and Machine Learning**

**Dr. Tony Diana**

**May 10, 2023**

## ABSTRACT

EWR (Newark Liberty International Airport) has been experiencing significant delays in 2022, which has led to increased public dissatisfaction and criticism of the airport's performance. Addressing the issue of delays at EWR will require a multi-faceted analysis of various delay factors and analyzing public opinion. Therefore, we have analyzed data related to EWR airport to find out the possible reason for the increased delays and performed sentiment analysis to find out public sentiment towards EWR airport.


## INTRODUCTION

Newark Liberty International Airport (EWR) is a major airport located in New Jersey that serves the New York City metropolitan area. In recent years, EWR has been experiencing a significant increase in flight delays, causing frustration and inconvenience for passengers and airlines alike. In fact, according to the Federal Aviation Administration (FAA), EWR has been one of the worst-performing airports in the United States regarding flight delays. In 2022, the situation has become even more severe, with the number of flight delays at EWR increasing by over 50% compared to the previous year (*Federal Aviation Administration*, 2023). Increased delays have resulted in missed connections, canceled flights, and lost revenue for airlines. The situation has also affected passenger experience and satisfaction, leading to negative reviews and public sentiment about the airport.

The reasons for the increase in delays at EWR are complex and multifaceted. They include a range of factors, such as air traffic congestion, runway closures, weather-related issues, and operational challenges. For example, during the summer of 2022, the airport experienced a record number of thunderstorms, which caused significant disruptions to flight schedules (*NextGen Weather*, n.d.)

To address the issue, airport operators and airlines have been working to identify the primary causes of delays and implement strategies to improve operational performance and customer satisfaction, which included measures such as reducing the number of flights during peak hours, optimizing runway usage, and improving communication between airlines and passengers during disruptions.

In this paper, we aim to investigate the causes of delays at EWR and evaluate public sentiment toward the airport. We will analyze a large dataset of flight information and social media posts to identify trends and patterns in flight delays and public sentiment. We aim to provide insights into the causes of delays and make recommendations for improving

operational performance and customer satisfaction. The research may also shed light on the importance of public sentiment in shaping an airport's reputation and the impact of social media on public perception. This analysis can be replicated and used to analyze delays at other airports.

## LITERATURE REVIEW

The issue of flight delays has been extensively studied in the literature. As a result, researchers have identified several factors contributing to flight delays.

A study by Ho et al. (2019) examined the impact of capacity constraints on airport performance and found that delays caused by capacity constraints are more likely to occur during peak hours. The study also highlighted the importance of investing in new technology and infrastructure to increase airport capacity and reduce delays.

Another study by Evans and Schäfer (2016) examines the impact of airport capacity constraints on future growth in the US air transportation system. The study concludes that such constraints will significantly affect the ability of the system to accommodate future demand for air travel. To address this issue, the paper recommends investing in measures to increase airport capacity, such as building new runways and expanding existing airport facilities (Evans & Schäfer, 2011).

In addition, the Federal Aviation Administration (FAA) has recognized the issue of airport capacity constraints and has implemented several initiatives to address the problem. These initiatives include the Next Generation Air Transportation System (NextGen), which aims to modernize air traffic control systems and increase airport capacity (*Federal Aviation Administration*, 2023).

A study by Borsky and Unterberger (2019) studied the impact of sudden and slow onset weather events on flight delays. The study found that sudden and slow-onset weather events significantly impacted flight delays, with thunderstorms and snowstorms having the most significant impact. Airlines responded differently to sudden and slow onset weather events, with airlines being more likely to cancel flights in response to sudden onset weather events. The study recommended that airlines should develop better contingency plans for managing the impact of sudden and slow-onset weather events on flight schedules (Borsky & Unterberger, 2019b).

Another study by Bombelli and Sallan (2019) studied the impact of extreme weather events on the US domestic air network using a delay and cancellation propagation network approach. The study found that extreme weather events such as hurricanes, snowstorms, and

thunderstorms significantly impacted the US domestic air network, causing widespread delays and cancellations that affected millions of passengers. The study recommended that airports and airlines invest in better weather forecasting and monitoring tools and develop contingency plans to manage the impact of extreme weather events on flight schedules (Bombelli & Sallan, 2023b).

The literature suggests that capacity constraints and weather are significant issues for airports, leading to delays and increased costs for airlines and passengers. Therefore, addressing these issues is essential in improving operational efficiency and providing a better travel experience for passengers.

## PROBLEM STATEMENTS

Based on the literature review, we will address the following questions.

- Is weather significantly responsible for increased flight delays at EWR in 2022?
- Is the capacity of EWR airport a primary reason responsible for increased flight delays at EWR in the year 2022?

## DATASET DESCRIPTION

Table 1 describes the data used in the analysis.

Table 1. Dataset Variables

| Variable | Type | Description |
|---|---|---|
| Scheduled Departures | Integer | Number of departures listed in a published schedule. Most cargo carriers do not have a published schedule. |
| Scheduled Arrivals | Integer | Number of arrivals listed in a published schedule. Most cargo carriers do not have a published schedule. |
| Average Gate Departure Delay | Float | The sum of minutes of Gate Departure Delay of 1 minute or more, divided by all departures. Gate Departure Delay is the difference between the Actual Gate Out time and Scheduled or Flight Plan Gate Out time, in minutes. |
| Average Taxi Out Delay | Float | The sum of minutes of Taxi Out Delay of 1 minute or more, divided by all departures. Taxi Out Delay equals Actual Taxi Out Time minus Unimpeded Taxi Out time |

| | | |
|---|---|---|
| Average Airport Departure Delay. | Float | The Actual Wheels Off minus the Scheduled Gate Out plus the Unimpeded Taxi Out time, in minutes. Negative values contribute to the average if the report has “Include Early Flights” checkbox selected on the filter tab in the Airport Analysis section. |
| Average Airborne Delay. | Float | The difference between Actual Airborne Time and the flight plan Estimated Time Enroute (Filed ETE), in minutes. |
| Average Taxi in Delay. | Float | The sum of minutes of Taxi In Delay of 1 minute or more, divided by all arrivals. Taxi in Delay equals actual Taxi In Time minus Unimpeded Taxi In Time. |
| Average Block Delay. | Float | The difference between Actual Gate-To-Gate time and Scheduled Gate-To-Gate, in minutes. |
| Average Gate Arrival Delay. | Float | The sum of minutes of Gate Arrival Delay of 1 minute or more, divided by all arrivals. Gate Arrival Delay is the difference between the Actual Gate In Time and the Scheduled or Flight Plan Gate In Time. |
| Capacity Airport Arrival Rate (Capacity AAR) | Integer | The number of arrivals an airport can support per unit of time when no traffic management initiatives are in effect. |
| Efficiency Airport Arrival Rate (Efficiency AAR) | Integer | The number of arrivals an airport can support per unit of time when traffic management initiatives are in effect. |
| Airport Departure Rate (ADR) | Integer | The Airport Departure Rate, or the number of departures an airport can support, per unit of time. |
| VMC | Integer | Visual Meteorological Conditions |
| IMC | Integer | Instrument Meteorological Conditions |
| Total for Efficiency Computation | Integer | Sum of number of departures and number of arrivals |
| % Capacity Utilized | Float | The sum of Arrivals and Departures for Efficiency Computation divided by the sum of the ADR and the Capacity AAR. |

| | | |
|---|---|---|
| Departure Efficiency. | Float | The ratio of Departures for Efficiency Computation to the lower of the ADR for a given runway configuration expressed as a percentage. |
| Arrival Efficiency | Float | The ratio of Arrivals for Efficiency Computation to the lesser of the Efficiency AAR expressed as a percentage. |
| System Airport Efficiency Rate (SAER) | Float | The SAER is the weighted average (by demand) of the Arrival Efficiency Rate and Departure Efficiency Rate. |

## INSIGHTS FROM THE DATASET

The data consists of various metrics to assess airlines' performance regarding their flights' timeliness. These metrics include the number of scheduled departures and arrivals, on-time departures and arrivals, gate departure and arrival delays, taxi times, and airport and airborne delays. For example, the percentage of on-time gate departures and arrivals indicates the proportion of flights that leave and arrive within 15 minutes of their scheduled times, respectively. This metric provides insight into an airline's ability to adhere to its published schedule.

The average gate departure delay and average taxi out delay provide information on how long flights spend waiting at the gate and taxiing on the runway, respectively. High values for these metrics may indicate issues with airport congestion, weather, or operational inefficiencies. The average airport departure and airborne delays mean the time flights wait on the ground and in the air before taking off. These metrics can be affected by various factors, including air traffic congestion, weather, and the performance of the airline's ground crew.

The average taxi in delay provides information on when flights wait to park at the gate after landing. In contrast, the average block delay indicates when flights remain for a gate to become available. These metrics can be influenced by factors such as airport layout and the efficiency of the airline's ground crew. Finally, the average gate arrival delay indicates when flights wait at the gate after landing. High values for this metric may indicate issues with airport congestion or operational inefficiencies. Overall, these metrics provide valuable insights into an airline's performance and can be used to identify areas for improvement.

## ALGORITHM

LDA (Latent Dirichlet Allocation) is a machine learning algorithm commonly used in natural language processing to identify topics within a corpus of text data. It works by recognizing patterns in the co-occurrence of words in documents and clustering those patterns into topics.

In analyzing airport delay data, LDA can be useful for identifying the underlying topics or causes of delays. For example, suppose we have a corpus of text data consisting of reports on airport delays. In that case, LDA can help us identify the most common topics or causes of delays by clustering related words and phrases into distinct topics.

By using LDA, we can gain insights into the primary causes of delays at an airport and develop targeted solutions to address them. For example, suppose LDA reveals that weather-related delays are the most common cause. In that case, the airport may invest in better weather forecasting tools or infrastructure to mitigate the impact of adverse weather conditions on flight operations.

Overall, LDA helps analyze large volumes of text data to identify underlying patterns and topics that may take time to be apparent. By applying LDA to airport delay data, we can gain insights into the root causes of delays and develop targeted solutions to address them, improving operational efficiency and the passenger experience.

## OUTCOMES

The delay at EWR in the year 2022 can be attributed to various factors, such as the weather at the airport and capacity. Therefore, we have analyzed the data containing weather, capacity, and delay at EWR airport from 2018 to 2022 to determine the reason for the increased delays in 2022.

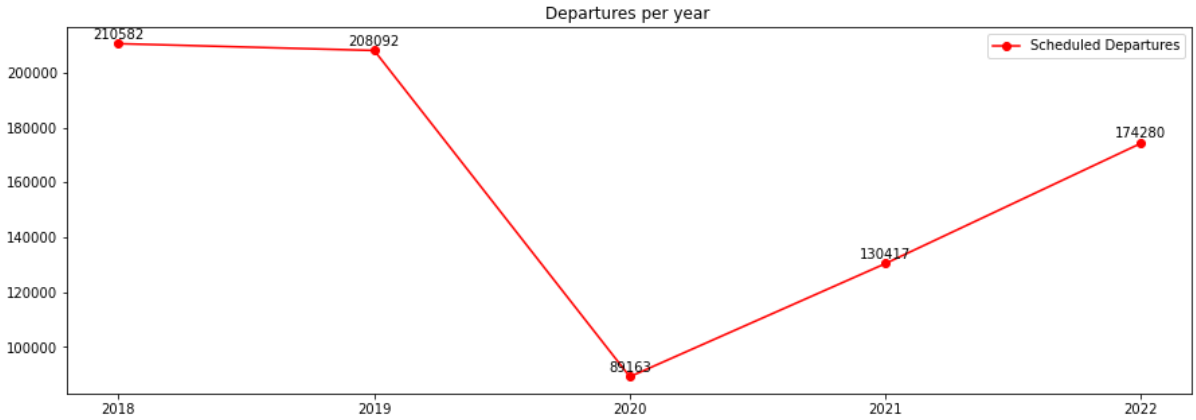


Figure 1. Departures per year

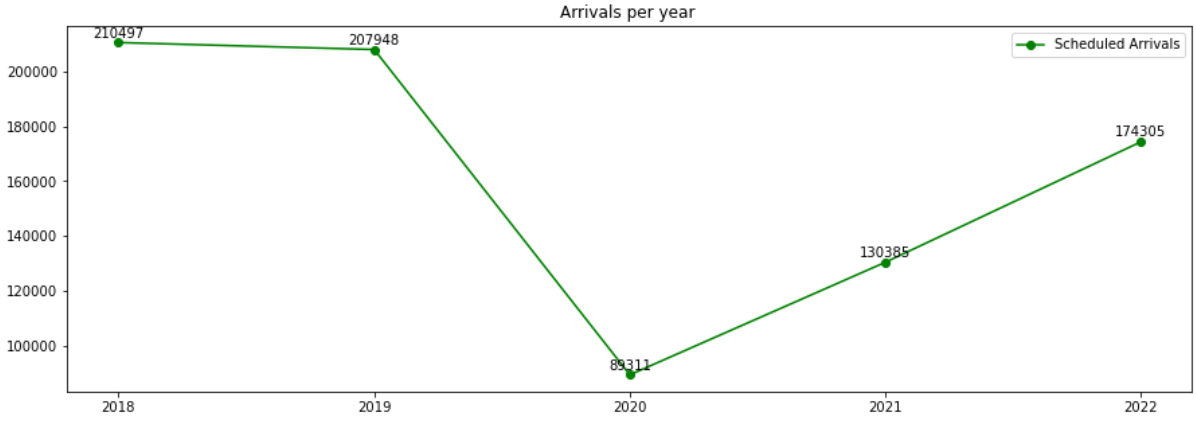


Figure 2. Arrivals per year

Arrivals and departures have been almost constant in the year 2018 and 2019. Later in the year, we can see that arrivals and departures have decreased to a great extent. This decrease can be attributed to the Covid-19 outbreak. Later in 2021 and 2022, we see that travel has started booming and returning to normal. However, the arrivals and departures are still less in 2022 compared to 2018 and 2019. The delays at EWR pre-Covid were high, and traffic was at an all-time high. Although it declined due to travel restrictions during the Covid period and post-Covid due to ease in travel restrictions, traffic is on a continuous rise which will raise problems related to airport capacity, arrivals, and departures, increasing in turn delays. Another factor for an increase in delays in 2022 may be attributed to the reduction in the airport staff post-Covid.

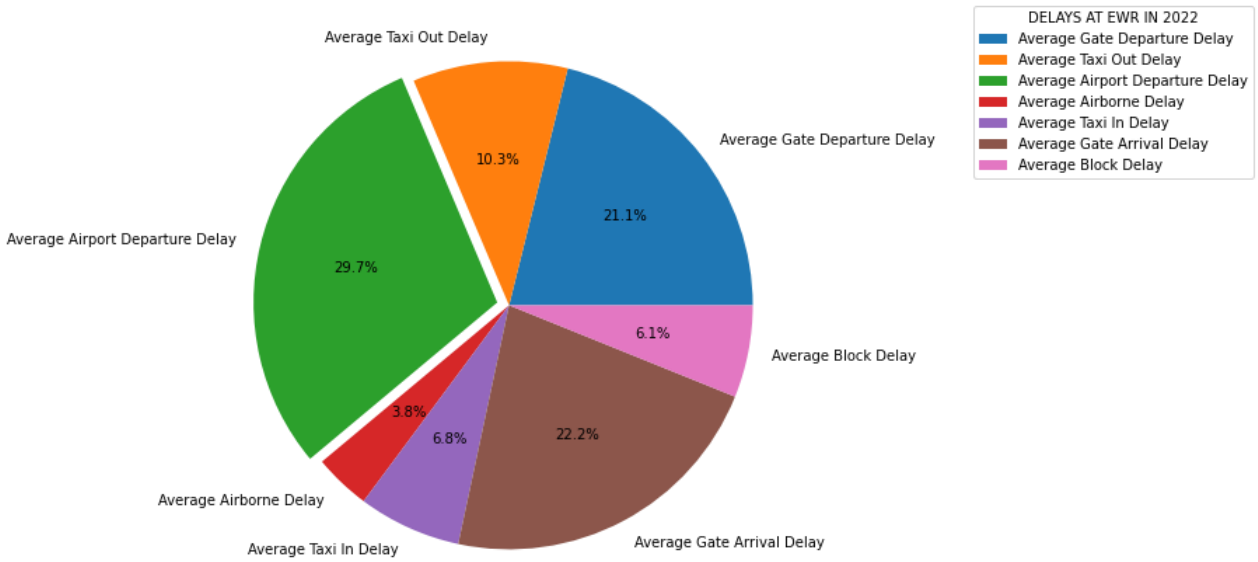


Figure 3. Delays at EWR in the year 2022

Different types of delays at EWR airport in 2022 are plotted as a pie chart in Figure 3. Again, average airport departure delay constitutes a significant part of delays with 29.7%.

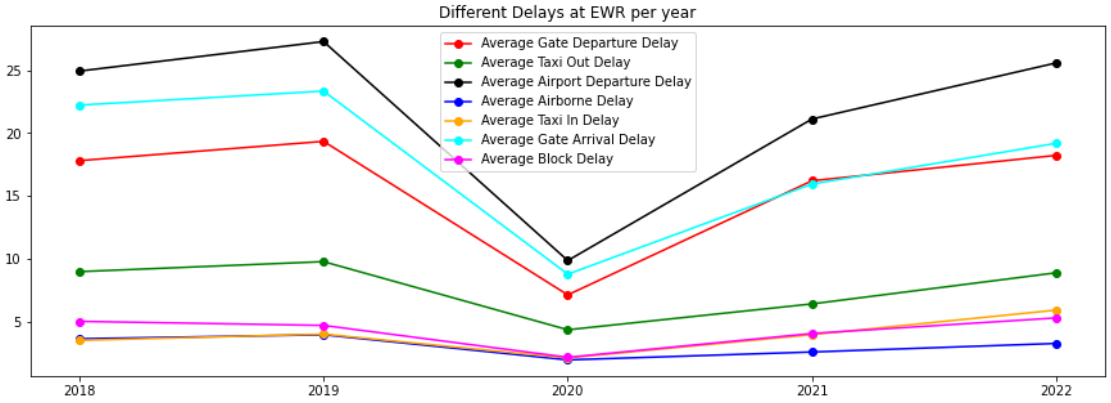


Figure 4. Delays at EWR per year

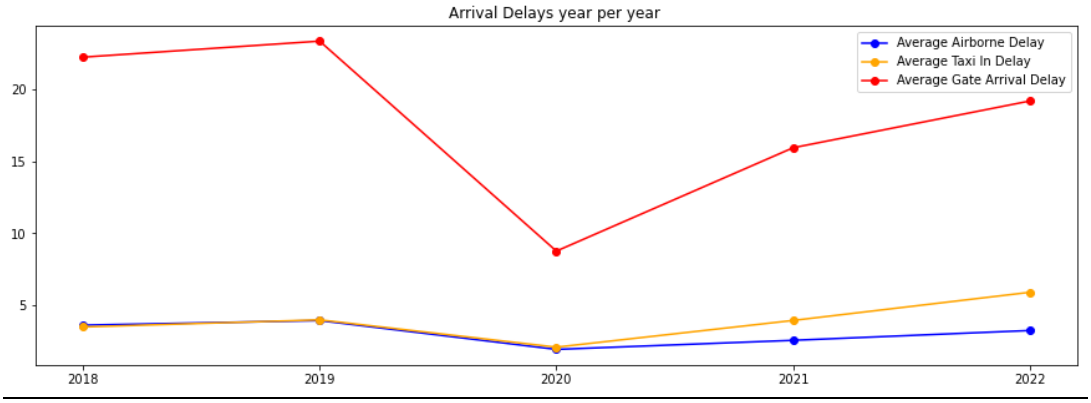


Figure 5. Arrival Delays at EWR per year

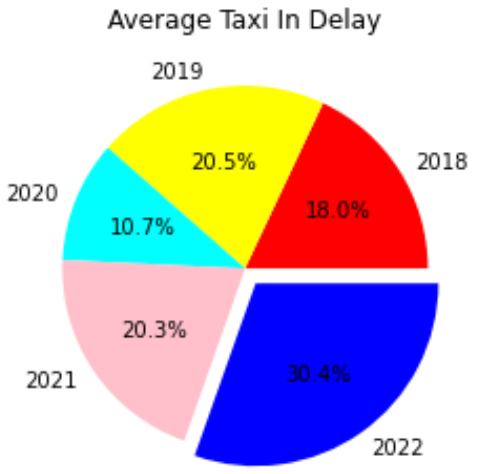


Figure 6. Average Taxi-In Delays at EWR per year

Figure 4 shows various delays that occurred at EWR airport from the year 2018 to 2022. All the delays except the taxi-in delay have happened in the same proportions each year. Even though the number of arrivals and departures in 2022 is less than in 2018, the taxi-in delay has increased from 18% in 2018 to 30.4% in 2022, as seen in Figure 6. Taxi-In delays might be one of the reasons for increased overall delays at EWR airport in 2022.

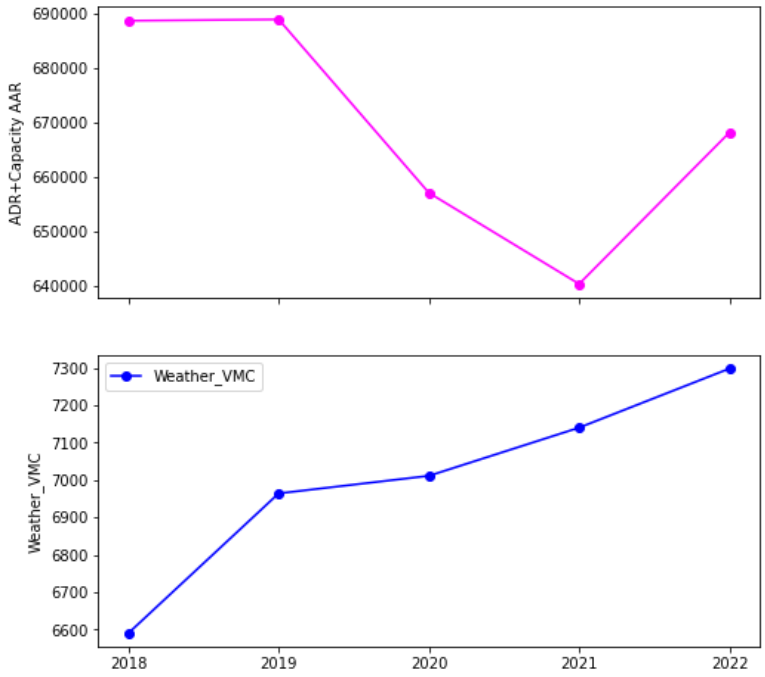


Figure 7. ADR + Capacity AAR and VMC Weather per year

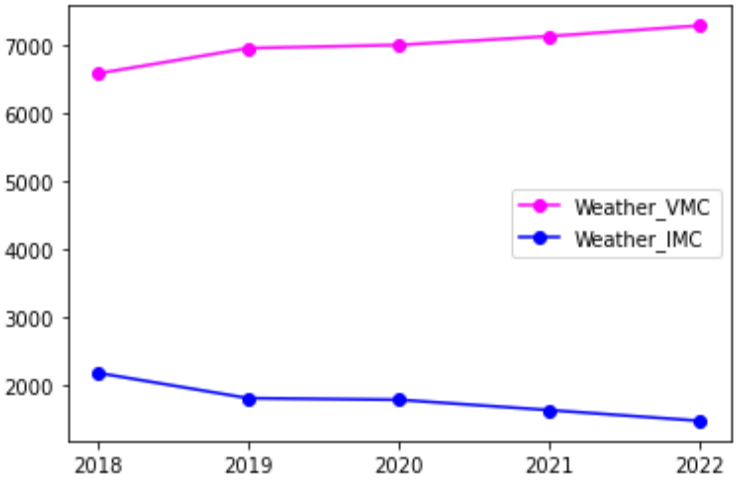


Figure 8. ADR and Efficiency AAR per year

The arrivals and departures handling capacity of EWR decreased over the years, as seen in Figure 7. In contrast, the number of hours with Visual Meteorological Conditions(VMC) has increased from 2018 to 2022. So, in our analysis, we found that weather is not the primary factor behind the increase in delays in the year 2022.

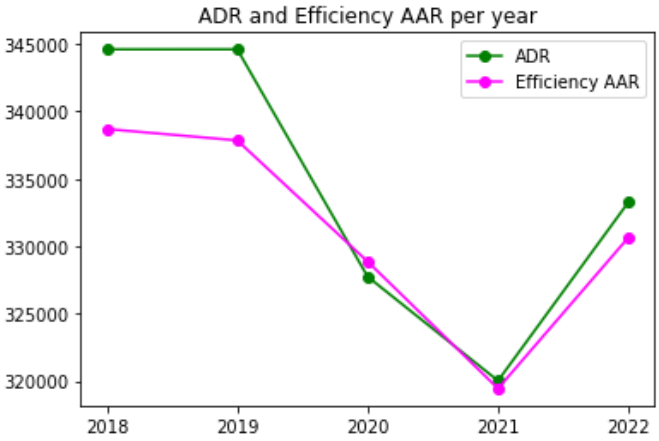


Figure 9. ADR and Efficiency AAR per year

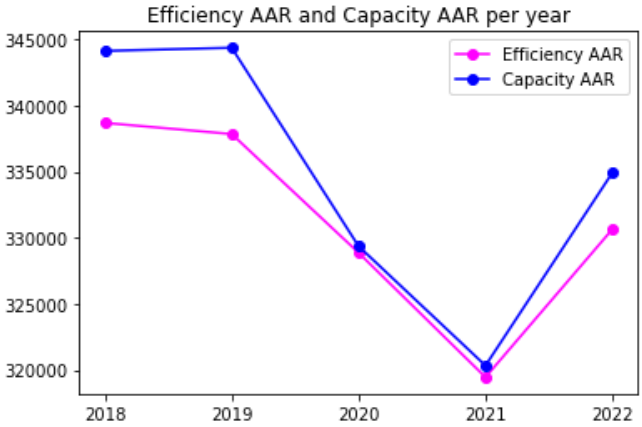


Figure 10. Efficiency AAR and Capacity AAR per year

Figure 9 shows that Airport Departure Rates and Efficiency Airport Arrival Rates are less in 2022 than in 2018. Therefore, it can be inferred that the arrivals and departures handling capacity of EWR airport have been reduced to a great extent. This might be one of the reasons for increased delays in 2022. Also, EWR airport has been working at Efficiency Airport Arrival Rates rather than Capacity Airport Arrival Rates over all the years. This might be due to traffic management initiatives in place. This shows that EWR airport is working at less capacity, which might be one of the reasons for the increase in delays at EWR in 2022.

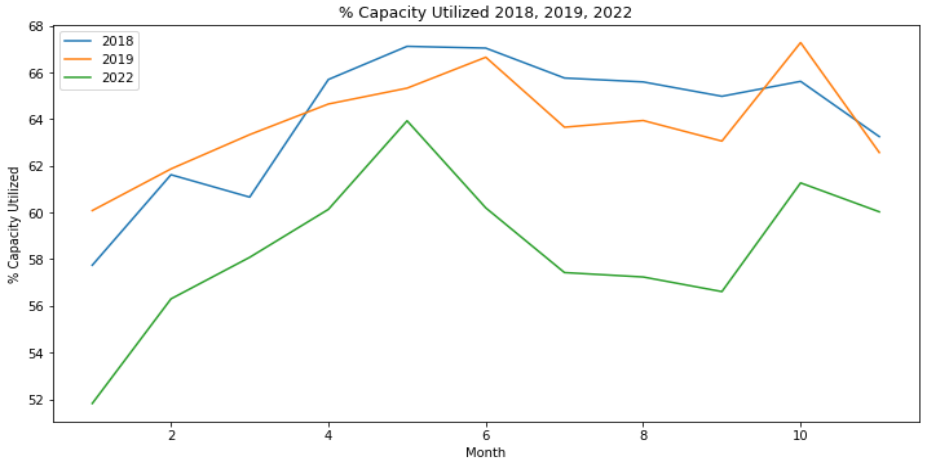


Figure 11. % Capacity of EWR utilized per year

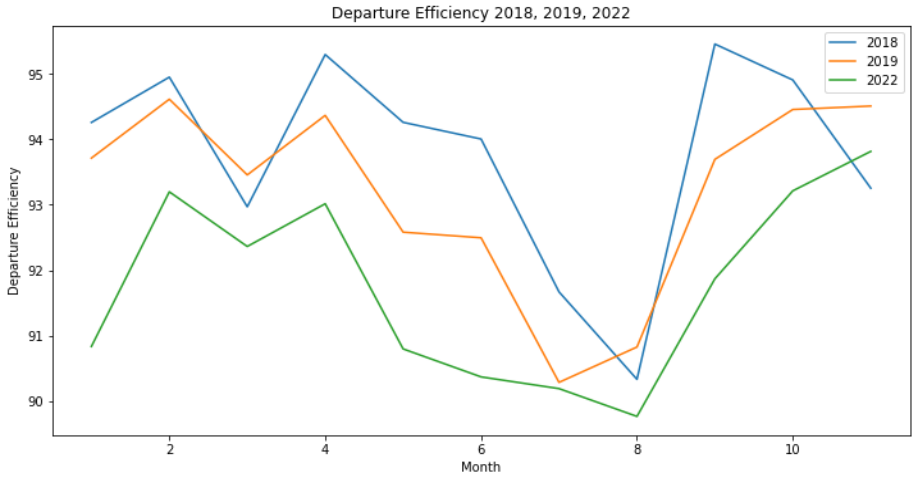


Figure 12. Departure Efficiency of EWR utilized per year

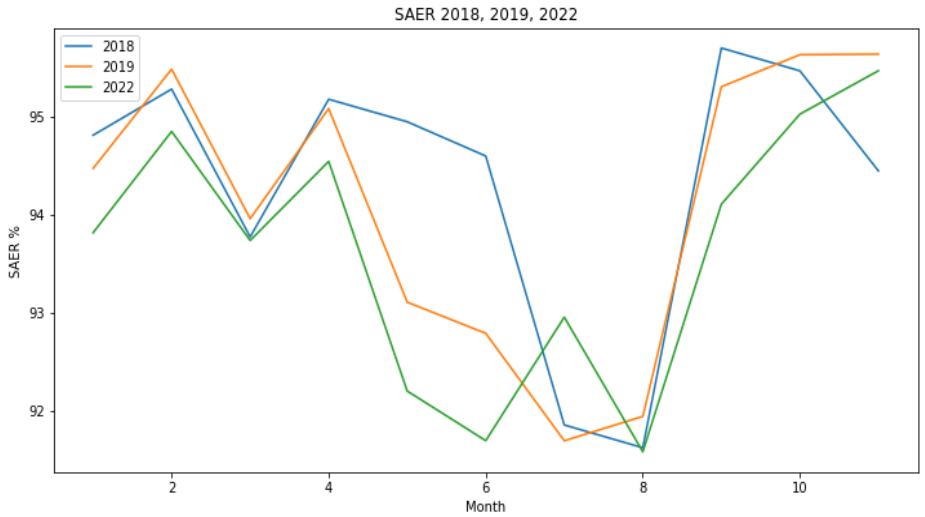


Figure 11. SAER of EWR per year

An airport's percentage capacity is defined as the capacity it has used out of its actual capacity. System Airport Efficiency Rate (SAER) measures the extent to which an airport handled the number of aircraft it indicated it could accommodate and how well the demand was met. From the above line graphs, the percentage utilized capacity of EWR airport and System Airport Efficiency Rate (SAER) is reduced in the year 2022 compared to the years 2018 and 2019. Hence, capacity is the primary reason behind the increase in delays at EWR in 2022.

## SENTIMENT ANALYSIS

We have also performed sentiment analysis on the tweets from Twitter to understand public opinion about EWR airport. We have collected over 2100 tweets from Twitter with search terms including 'EWR delay', 'EWR baggage claim', and 'EWR weather'. We found that around 59.7% of tweets are negative, indicating that people face delays issues at EWR airport.

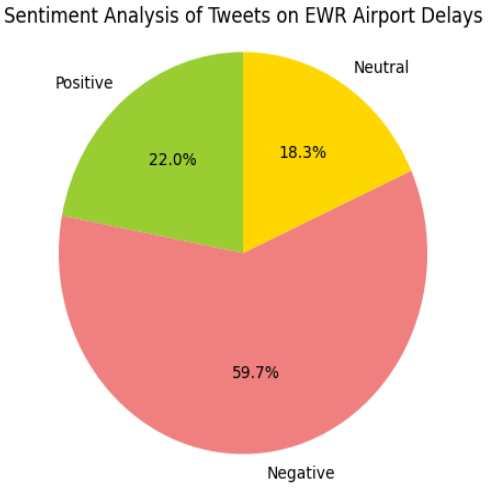


Figure 12. Sentiment Analysis of Tweets on EWR airport delays


Figure 13. Word Cloud of tweets on EWR airport delays

The word cloud of the tweets also suggests that people mainly discuss the delays related to taxi-in, taxi-out, and gate delays. These delays occur mainly due to capacity issues which supports our analysis.

Latent Dirichlet Allocation (LDA) is a statistical model discovering topics hidden within an extensive collection of text documents. An unsupervised learning algorithm represents each document as a mixture of topics, where each topic is a probability distribution over words. GridSearchCV provides the best combination of hyperparameters by specifying a range of values for each hyperparameter and evaluating the model's performance on the validation set. Furthermore, it will train and evaluate the model for each combination of hyperparameters and return the best hyperparameters that give the highest performance score. Overall, GridSearchCV can help to automate the hyperparameter tuning process and find the optimal configuration of the model for Twitter sentiment analysis.

We deployed the Latent Dirichlet Allocation (LDA) model to discover the topics in our tweets. To tune the hyperparameters of LDA, we used the GridSearchCV technique and found the set of hyperparameters for LDA with high performance. We then trained the LDA model with the hyperparameters suggested by GridSearchCV and modeled tweets into five topics. For example, from Figure 15, we can see that people are discussing more ‘Traffic Management Initiatives’, ‘Gate Delay’, and ‘Taxi Delay’. These delays occur due to capacity constraints, and the topics modeled support our above analysis.

```
Best Hyperparameters:  {'learning_decay': 0.5, 'max_iter': 100, 'n_components': 5}
```

Figure 14. Hyperparameters of LDA suggested by GridSearchCV

Topics Modeled By LDA

Topic: 0
**delay** program **traffic management** prior lifted land gt carrier business

Topic: 1
flight **delay** minute causing arriving **traffic mgmt** averaging prgrm hour

Topic: 2
**delay gate** minute hold **taxi** hour tm increasing **initiatives**swapwx general

Topic: 3
**delay ewr** newark amp airline international got **service** liberty currently

Topic: 4
**delay** ewr flight hour experiencing time **arrival traffic airborne** airport

## CONCLUSION

In conclusion, the analysis has revealed that many of the delays at EWR are related to "Traffic management initiatives", "Gate delays", "Taxi delays", and "Airborne delays". This analysis has been done, including the pre and post-Covid period providing a good picture of how Covid-19 has affected air traffic. We also analyzed that the rise in delays at EWR airport will go back to the pre-Covid level. These delays can occur due to capacity constraints, which can arise when there is insufficient infrastructure, staff, or resources to handle the volume of flights and passengers. As a result, airports need to monitor and address these delays to improve their operational performance and reduce passenger frustration. Addressing capacity constraints may involve investing in infrastructure upgrades, increasing staffing levels, and implementing new technologies and procedures. By taking proactive measures to address these issues, airports such as EWR can improve their operational efficiency and provide a better travel experience for passengers.

## ACKNOWLEDGEMENTS

We would like to thank our professor Dr. Tony Diana for providing access to aviation performance data used in this analysis.

# REFERENCES

Amira Sweilem, NJ Advance Media for NJ.com. (2022, September 20). Newark airport ranks worst for flight delays through first half of 2022. *Nj*. https://www.nj.com/data/2022/09/newark-airport-ranks-worst-for-flight-delays-through-first-half-of-2022.html

*ASPMHelp*. (n.d.). https://aspm.faa.gov/aspmhelp/index/Main_Page.html

Bombelli, A., & Sallan, J. M. (2023). Analysis of the effect of extreme weather on the US domestic air network. A delay and cancellation propagation network approach. *Journal of Transport Geography*, *107*, 103541. https://doi.org/10.1016/j.jtrangeo.2023.103541

Borsky, S., & Unterberger, C. (2019). Bad weather and flight delays: The impact of sudden and slow onset weather events. *Economics of Transportation*, *18*, 10–26. https://doi.org/10.1016/j.ecotra.2019.02.002

*Bureau of Transportation Statistics*. (2023, April 26). https://www.bts.gov/

Evans, A., & Schäfer, A. (2011). The impact of airport capacity constraints on future growth in the US air transportation system. *Journal of Air Transport Management*, *17*(5), 288–295. https://doi.org/10.1016/j.jairtraman.2011.03.004

Federal Aviation Administration. (n.d.). Capacity. Retrieved from https://www.faa.gov/air_traffic/capacity/

Fitch Ratings. (2021). Newark Liberty International Airport. Retrieved from https://www.fitchratings.com/research/us-public-finance/fitch-rates-port-authority-of-ny-nj-1b-series-2021a-bonds-23-stb-23-27s-08-10-2021

Ho, H. W., Li, Y., Oum, T. H., & Yan, J. (2019). Airport capacity constraint, delay, and pricing: Insights from slot allocation reform at Seoul Incheon International Airport.

Transportation Research Part A: Policy and Practice, 120, 35-49. https://doi.org/10.1016/j.tra.2018.11.019

Kammoun, M. B. A., & Rezg, N. (2018d). An efficient hybrid approach for resolving the aircraft routing and rescheduling problem. *Journal of Air Transport Management*, *71*, 73–87. https://doi.org/10.1016/j.jairtraman.2018.06.005

National Oceanic and Atmospheric Administration. (2022). Climate Data Online. Retrieved from https://www.ncdc.noaa.gov/cdo-web/datatools/normals

NextGen. (n.d.). Next Generation Air Transportation System. Retrieved from https://www.faa.gov/nextgen/

*NOAA NCEI U.S. Climate Normals Quick Access*. (n.d.). https://www.ncdc.noaa.gov/cdo-web/datatools/normals

Port Authority of New York and New Jersey. (2022). Newark Liberty International Airport. Retrieved from https://www.panynj.gov/airports/newark-liberty.html