CITADEL SUMMER INVITATIONAL DATATHON

WEATHER IMPACT ON AIR TRAFFIC

A CLOSER LOOK AT THE CORRELATION BETWEEN WEATHER CONDITIONS ON FLIGHT DELAYS

TEAM 11

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Air travel is a key pillar of modern global connectivity, facilitating not only business but also cultural exchange and tourism. Yet, flight delays remain a major inconvenience for travelers worldwide. These disruptions can cause a ripple effect, disrupting plans, increasing costs, and inducing stress. Our team, comprising of avid travelers and data enthusiasts, decided to delve into one of the crucial factors impacting flight delays: weather conditions. Understanding this factor is not just a matter of academic interest but can lead to more effective communication from airlines, better passenger management, and improved scheduling protocols.

To unravel this complex issue, we focused on two main research questions:

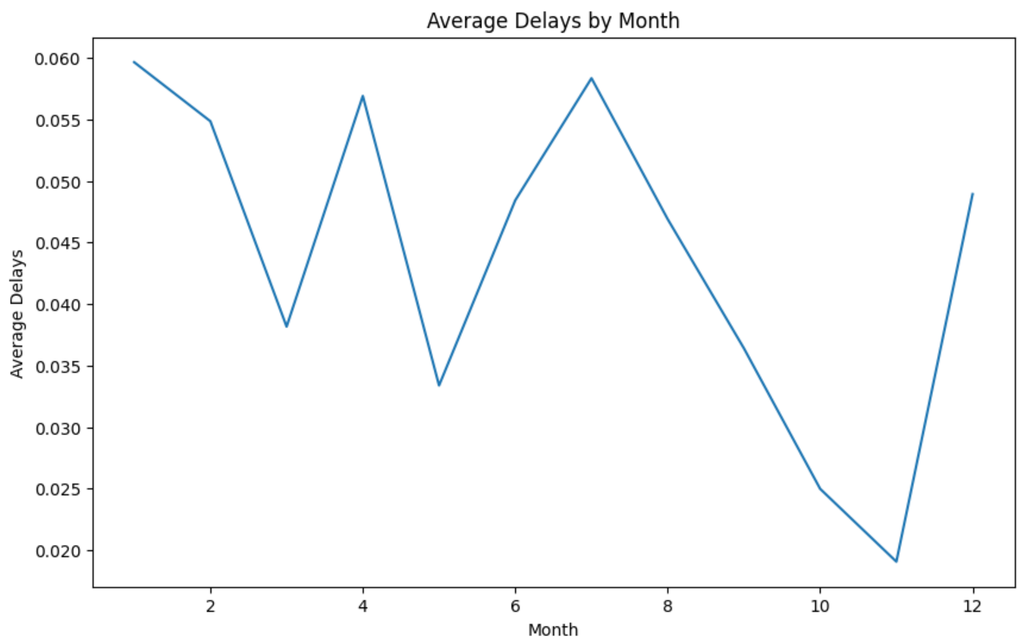
01 How do the severity and seasonality of weather conditions correlate with the magnitude of flight delays?

02 Is there a particular season during which flights are disproportionately impacted by delays?

The choice of these questions was motivated by our experiences as travelers and our realization that flight delays are often too easily attributed to airline inefficiencies. In contrast, we hypothesized that weather, a factor beyond human control, might be a significant contributor to these delays.

03

Through a meticulous exploratory data analysis and modeling process, we discovered that the severity of weather conditions indeed significantly impacts the likelihood and duration of flight delays.



High wind speed, poor visibility, and extreme temperatures were among the key weather conditions contributing to longer delays.

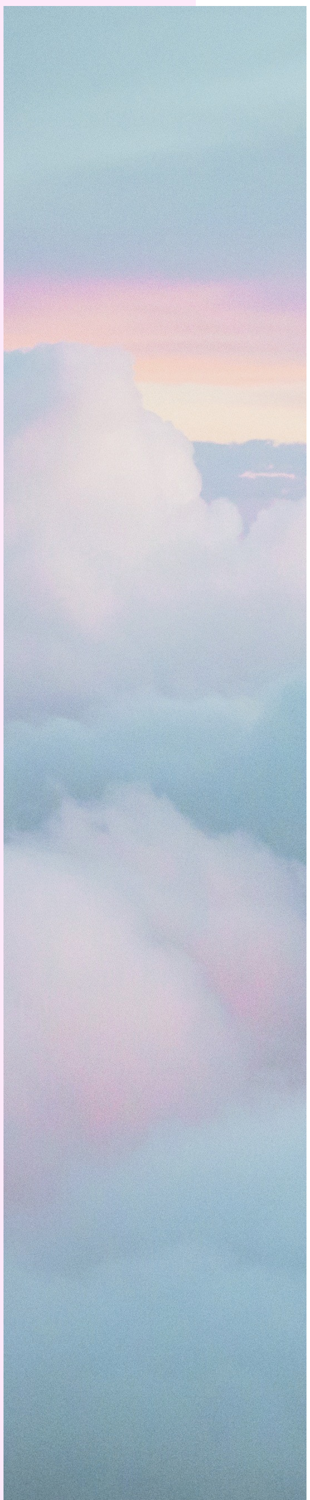
However, our findings also revealed the complexity of this issue. There was no single weather severity threshold identified that would predict a delay. Instead, it was the combination of various severe weather conditions that tended to correlate with flight disruptions.

Regarding seasonality, we found that periods associated with harsher weather conditions—such as winters in colder regions and hurricane season in tropical areas—tended to see a higher rate of flight delays. Nevertheless, our data did not allow us to declare one specific season as the most problematic universally.

To make our findings more tangible, we employed visualizations like heatmaps to showcase the correlations between weather factors and flight delays. This demonstrated the multidimensional nature of our issue at hand, showing the interconnectedness of different weather conditions in their impact on flight schedules.

In summary, our research underscores the significant but complex relationship between weather severity and flight delays. This complexity is a reminder of the challenges airlines face in maintaining punctuality in the face of nature's unpredictability. Consequently, our findings call for continued research in this domain and for travelers to be mindful of these factors when dealing with flight delays. After all, there is much more to the story than what meets the eye.

BACKGROUND



I had a 30 hour delay on my flight it was absolutely mortifying.

- An angry airline customer

Flight delays are now a common occurrence in the aviation industry, causing significant inconvenience to both passengers and airline companies themselves. Although the factors contributing to these delays are multifaceted, weather conditions are the most common reason behind the greatest fear of every traveler (Whitmore, 2022). The interplay between weather and flight delays has always been a point of interest, and the severity and seasonality of weather conditions have been observed to have a profound impact on these delays.

Weather conditions directly affect flight schedules and operational efficiency. Severe weather events such as snowstorms or low visibility can cause substantial delays or even cancellations of flights. The seasonality of weather is another crucial factor. Weather conditions vary by each season, and certain times of the year, like winter or hurricane season, may bring about more severe weather, potentially causing an increased frequency of delays.

CAUSE

According to the Federal Aviation Administration, weather conditions account for approximately 40% of all flight delays, the most significant cause of aircraft delays. A range of weather scenarios, including but not limited to storms and high winds, pose serious challenges to flight operations, often leading to inevitable delays in flight schedules.

CONSEQUENCE

Flight delays induce considerable extra costs for airlines, airport terminals, and passengers alike. Airports Council International estimates that the global cost of airport delays amounted to $75.5 billion in 2017. Airports too suffer financially due to lower revenues from missed arrival slots and increased operational costs.

Meanwhile, passengers face inconveniences in the form of missed connections, increased travel expenses, and lost time. PwC estimates around $60 billion for 2017. This significant financial impact necessitates improved air traffic control, operational strategies, and stakeholder collaboration. The adoption of advanced technologies and preventive maintenance programs could also contribute to reducing delays and increasing aviation system efficiency.

RESEARCH QUESTIONS

We test two main questions to determine the relationship between severe weather events and flight delays:

01 How does the severity and seasonality of weather conditions correlate with the magnitude of flight delays?

02 Is there a specific weather severity threshold or a particular season during which flights are disproportionately impacted by delays?

MOTIVATION

Flight delays pose significant operational, economic, and reputational challenges to the aviation industry, alongside causing substantial inconvenience to passengers. Weather, as a leading cause of these delays, thus necessitates closer examination.

Identifying a specific weather severity threshold or a particular season during which flights are disproportionately impacted by delays can provide a tangible benchmark for airline decision-making. Ultimately, the insight could assist the industry in mitigating the negative impacts of severe weather and seasonal changes, leading to improved operational efficiency, cost savings, and an optimal customer experience.

DATA WRANGLING

The first step was the process of transforming and mapping data from its raw form into a more compatible version for further data analysis.

IMPORTING DATA

The pandas library is utilized for this purpose. Two datasets are imported: 'flight_traffic.csv' and 'weather.csv', containing flight traffic information and weather data respectively.

```
flight_traffic = pd.read_csv("flight_traffic.csv")
weather = pd.read_csv("weather.csv")
```

FORMATTING DATA

For both datasets, a 'datetime' column was created to facilitate the merging of the two datasets. The 'datetime' column was constructed using the pandas to_datetime function which allows conversion of string or multiple columns to a datetime object. The 'flight_traffic' dataset's 'datetime' column was created using the 'year', 'month', and 'day' columns, while for the 'weather' dataset it was already available.

```
df["datetime"] = pd.to_datetime(df[["year", "month", "day"]])

weather["datetime"] = pd.to_datetime(weather["datetime"])
```

MERGING DATA

The flight traffic data was merged with the weather data for both the origin and destination airports. This was done using the merge_asof function, a special merge order that allows for merging on nearest key rather than exact matches. In this case, the data was merged based on the nearest available weather data within a 24-hour range. This is achieved using the 'tolerance' argument set to 24 hours.

```
types = [("origin", "origin_airport"), ("destination", "destination_airport")]
props = weather.columns[5:]
for u, v in types:
    df = df.sort_values(["datetime"])
    df.rename(columns={v: "airport_id"}, inplace=True)
    df = pd.merge_asof(df, weather, on='datetime', by='airport_id', direction='nearest', tolerance=pd.Timedelta(hours=24))
    df.rename(columns={"airport_id": v}, inplace=True)
    for column in weather.columns:
        if column not in ['airport_id', 'datetime']:
            df.rename(columns={column: f'{u}_' + column}, inplace=True)
```

DATA CLEANING

The last step in the data wrangling process involved cleaning the data. Any rows with missing values in the selected features were dropped. This was done to ensure that only quality data is used in the subsequent modeling process. The dropna function was used to drop missing values in the specified features.

```
features = []
for x in weather.columns[6:-2]:
    features.append(f"origin_{x}")
    features.append(f"destination_{x}")
features.append("distance")
features.append("scheduled_departure")
df = df.dropna(subset=features)
```

Each weather attribute was assigned respectively to both the origin airport as well as the destination report to ensure that the entire flight route was considered appropriately.

METHODOLOGY

The second step was to conduct exploratory data analysis in order to gain a deeper insight.

The exploratory data process began with a comprehensive overview of the flight traffic and weather datasets. A thorough examination of each feature's data type was conducted to observe their descriptive statistics. Missing values, outliers, and the distribution of values were also inspected.

We noticed a significant class imbalance in our target variable, i.e., the 'delayed' column, with an overwhelming majority of flights experiencing no weather-related delays. Such a class imbalance could lead a machine learning model to predict the majority class overwhelmingly, thereby providing misleading accuracy metrics. To address this, we decided to implement oversampling, a method which duplicates minority class instances until we attain a balanced dataset.

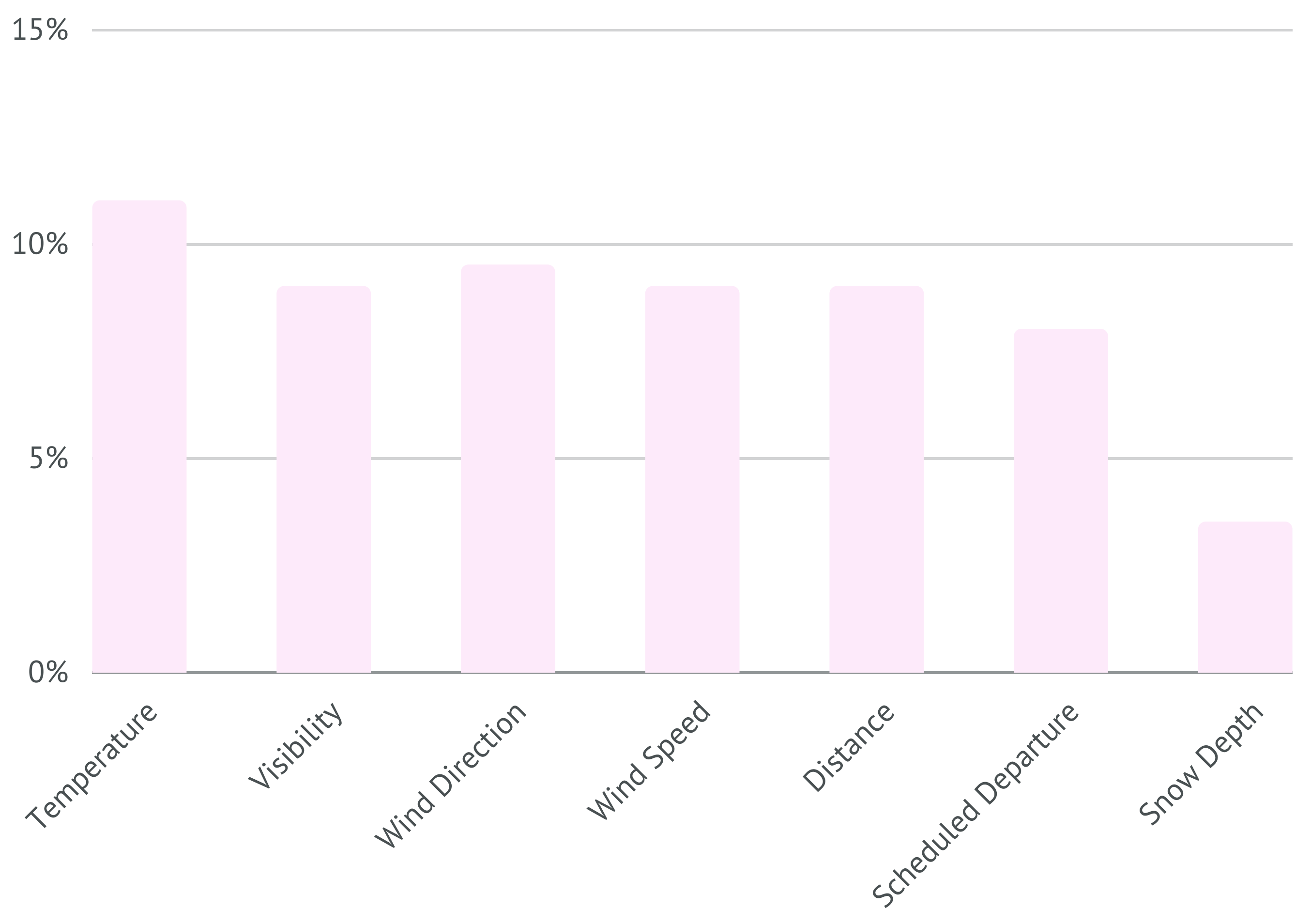
While assessing our weather dataset, we observed that some attributes, like cloud cover and snow depth, contained numerous missing values. To maintain data integrity, we decided to transform all null snow depth values as zero because many airports do not encounter snow. Cloud cover was removed entirely from the data, since visibility is a more important factor.

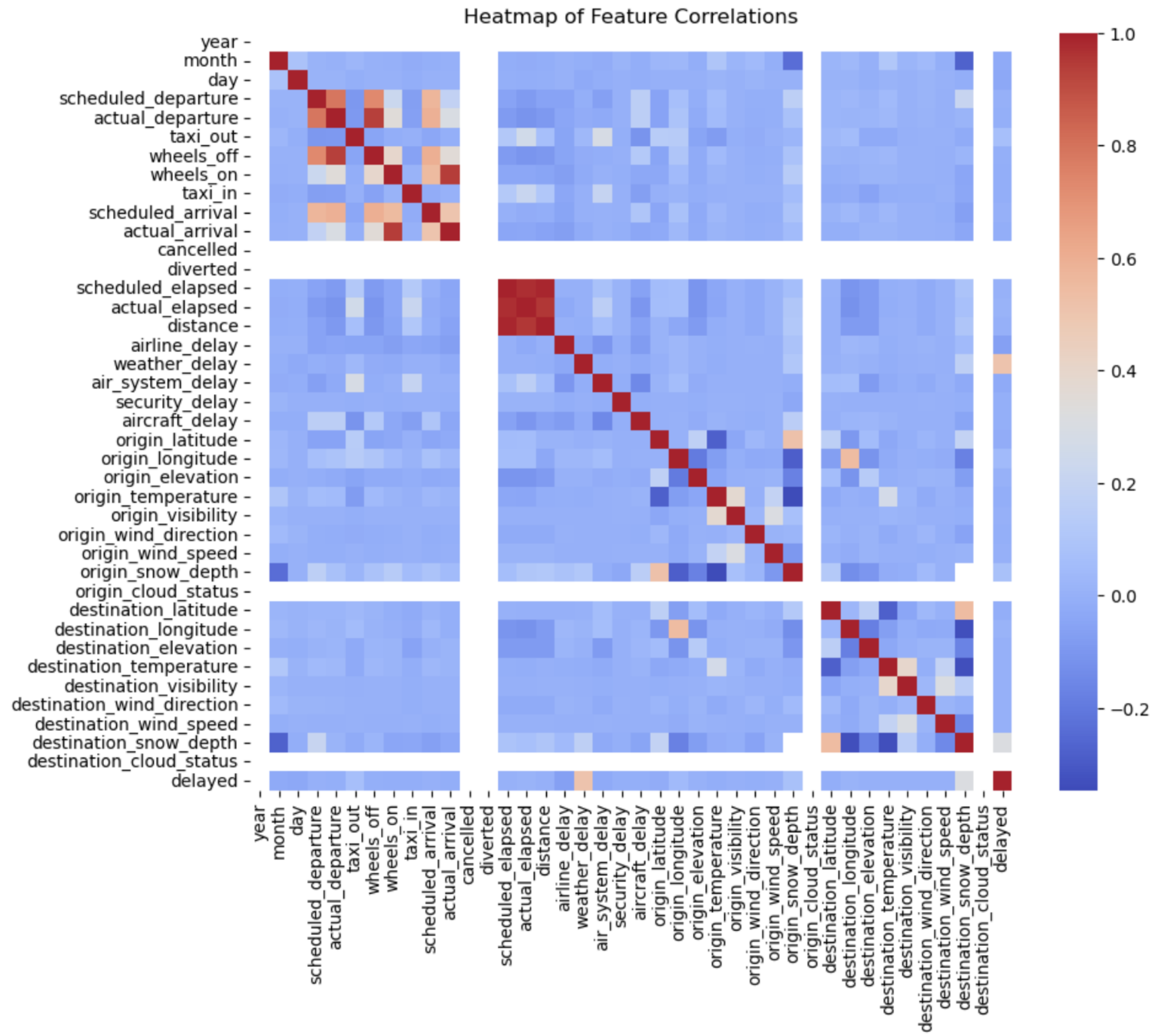
10

Our research question concerning the relationship between weather severity, seasonality, and flight delays led us to formulate a hypothesis that certain weather conditions impact flight delays.

Additionally, we embarked on an ad-hoc analysis, creating individual models for each airport. The rationale behind this decision was rooted in the geographical specificity of weather conditions and their consequent impact on flight operations, which could significantly differ between airports. For instance, snow in New York would completely be normal during December but there would not be any amount of snow in Miami at any time of the year.

In the below bar graph, feature importance is demonstrated. The origin and destination airports have been combined to show the significance of each feature on flights that have been delayed. As expected, temperature is the leading factor, as freezing temperatures would indicate snow and snowstorms. In regions where winters are often an issue, an abnormally low temperature could cause unexpected delays in their airports.

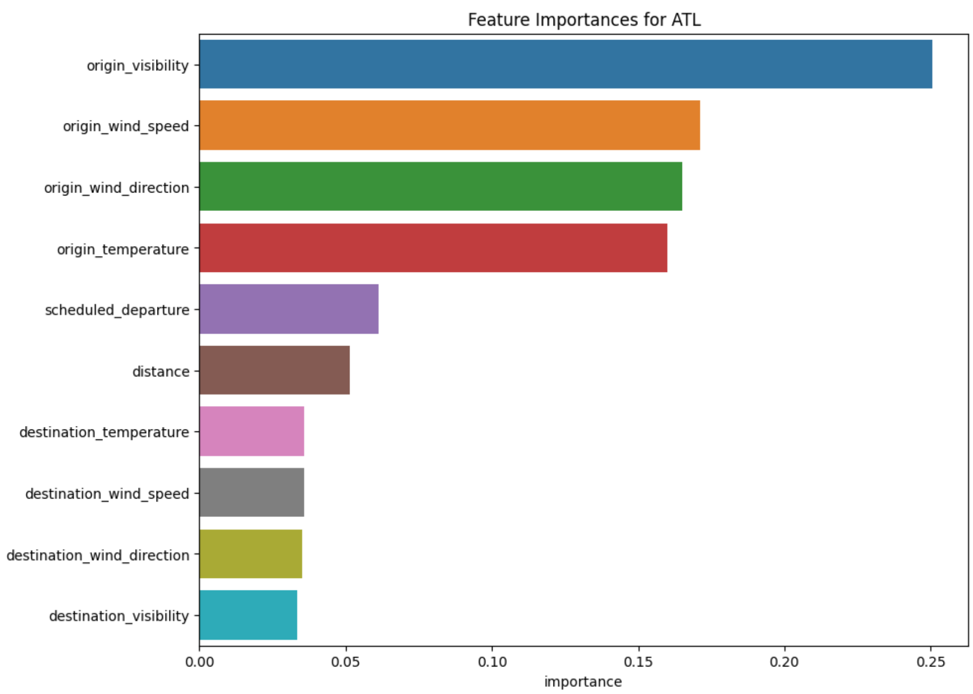




HEATMAP

Creating a heatmap helps visualize the correlation between different features in the dataset. This understanding can help guide further steps in the data analysis and modeling process, such as feature selection. For example, if two features are highly correlated, it might not be beneficial to include both in a predictive model as they could be providing redundant information. We can see that the information gained from the top left corner is not very useful, hence we decided to drop these features in our model design.

12 RESULTS



The above bar graph shows our model findings for the nation's busiest airport in Atlanta. Reading feature importance allows to consider the reasons behind delays, specific to this region. Atlanta, in this case, has hot and humid summers, with added precipitation, which increases the chance of fog and consequently has low visibility rates. This is believed to cause extensive flight delays particularly in July and August, when weather conditions are most severe.

ANALYTICS

The selection of XGBoost classifier as our machine learning model was driven by its inherent strengths. XGBoost is robust to outliers and efficiently handles a mix of categorical and numerical features. Additionally, it performs an implicit feature selection by attributing an importance score to each feature, guiding us in understanding the features' relative significance.

Our data exhibited a severe class imbalance, with the 'no delay' instances significantly outnumbering the 'delay' instances. To rectify this, we employed the RandomOverSampler method from the imbalanced-learn library, which uses random sampling to duplicate instances from the minority class, thereby balancing the class distribution in our training data.

13 MODELING RIGOUR

For feature selection, we initially included all weather attributes available, excluding those with a high number of missing values. Post-model training, we used XGBoost's built-in feature importance mechanism to understand the contribution of each feature towards our model's decision-making process.

We rigorously analyzed the performance of our models using a variety of metrics, including accuracy, precision, recall, and F1-score. These metrics were calculated for both our general model, which was trained on the complete dataset, and the airport-specific models, which were trained exclusively on data from individual airports.

EVALUATION METRICS

Accuracy: The ratio of correct predictions to the total number of predictions. For a binary classification problem, it can be defined as:

Accuracy = (True Positives + True Negatives) / (True Positives + False Positives + True Negatives + False Negatives)

Precision: The ratio of correct positive predictions to the total predicted positives. It answers "What proportion of positive identifications was correct?".

Precision = True Positives / (True Positives + False Positives)

Recall: The ratio of correct positive predictions to the total actual positives. It answers "What proportion of actual positives was identified correctly?".

Recall = True Positives / (True Positives + False Negatives)

F1-Score: The harmonic mean of precision and recall. It tries to find the balance between precision and recall.

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

Airport	Accuracy	Precision	Recall	F1 Score
ATL	0.840	0.857	0.840	0.848
DEN	0.930	0.935	0.930	0.933
LAX	0.960	0.957	0.960	0.959
JFK	0.924	0.915	0.924	0.920
LAS	0.960	0.957	0.960	0.959

The above and below charts summarizes our model evaluations for the top five most busiest airports in the United States for the year 2017, although the same has been done for the entire list of airports in the code itself. Accuracy and precision have been optimized to produce the best results possible.

Airport	Accuracy	Precision	Recall	F1 Score
ATL	0.882	0.922	0.882	0.900
DEN	0.900	0.927	0.900	0.913
LAX	0.852	0.924	0.852	0.884
JFK	0.936	0.944	0.936	0.940
LAS	0.924	0.938	0.924	0.931

CONCLUSION

Addressing our first research question on how the severity and seasonality of weather conditions correlate with the magnitude of flight delays, we discovered that there is indeed a strong correlation between certain weather conditions and flight delays.

From our XGBoost model, we observed that the severity of conditions such as wind speed, visibility, and temperature had high feature importance scores, indicating that they play significant roles in contributing to flight delays. For instance, high wind speeds or reduced visibility can create unsafe conditions for takeoff or landing, hence causing delays.

Regarding seasonality, our model did not directly incorporate specific seasonal information to avoid over-generalizations for region-specific airports. From a logical standpoint, Chicago and Los Angeles weather behave completely differently, and it would incorrect to assume that their respective delays share the same weather correlations, such as snow depth. However, some seasonal trends might be inferred indirectly via temperature fluctuations, and other weather patterns associated with specific seasons. For example, a drop in temperature might imply winter months, which could see a surge in weather-related delays due to snow or icy conditions.

The second research question sought to find a particular season during which flights are disproportionately impacted by delays. Our analysis did not define an exact numerical weather severity threshold beyond which delays become significantly more likely. This is because weather is multifactorial, and flight delays can depend on the combination and interaction of several weather conditions, not a single factor exceeding a certain limit.

As for seasonality, our model could not explicitly discern a 'worst' season for flight delays as our features did not directly contain such information. However, by interpreting weather patterns such as colder temperatures, it is possible to infer and validate that seasons traditionally associated with such weather, like winter, could see a higher rate of flight delays.

In summary, our research suggests that both the severity of weather conditions and the time of the year can significantly impact the likelihood of flight delays. While it is difficult to label a single season as the most troublesome, since the criteria for severe weather is so diverse across the country, we highlight the importance of considering multifaceted weather patterns when predicting and managing flight delays. Future research could focus on refining these observations, perhaps by directly incorporating seasons into the model, or exploring specific weather thresholds in more detail.

LIMITATIONS

Although oversampling is an effective way to ensure that the model isn't biased towards the majority class, it comes with its own limitations. It may lead to overfitting since it replicates the minority class events. Additionally, although oversampling ensures a better model performance metric on the training data, it does not guarantee that the model will perform similarly well on unseen data.

FUTURE RESEARCH

In terms of future research, where there is less of a time and data shortage, we could explore handling class imbalance, like Adaptive Synthetic Sampling which generate synthetic examples in a less arbitrary manner.

Another possible extension to the current work would be to consider airport-specific weather impact factors. Weather data inherently has a temporal component to it and a time-series model like ARIMA can capture patterns missed by the current models.

Finally, more granular weather data could help to improve the performance and accuracy of the predictive model.

REFERENCES

Anupkumar , Ashmith. “Investigating the Costs and Economic Impact of Flight Delays in The ...” CSUSB ScholarWorks, scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=2885&context=etd. Accessed 30 July 2023.

“Connectivity and Growth 2017 Edition - PWC.” PWC, www.pwc.com/gx/en/industries/assets/connectivity-and-growth-2017-edition.pdf. Accessed 30 July 2023.

Whitmore, Geoff. “What’s Causing Flight Delays and Cancellations?” *Forbes*, 23 June 2022, www.forbes.com/sites/geoffwhitmore/2022/06/22/whats-causing-flight-delays-and-cancellations/?sh=42ae4e604873.

“List of the Busiest Airports in the United States.” *Wikipedia*, 25 July 2023, en.wikipedia.org/wiki/List_of_the_busiest_airports_in_the_United_States.

“Weatherspark.Com.” *Atlanta Climate, Weather By Month, Average Temperature (Georgia, United States) - Weather Spark*, weatherspark.com/y/15598/Average-Weather-in-Atlanta-Georgia-United-States-Year-Round. Accessed 30 July 2023.