Flight Forecast: Predicting Summer Travel Disruptions

- Background
- Data Visualization
- Modeling Approach

Group 4

Jiapeng Wang

Yifan Chen

Zhixing Liu

Background

Background:

- Summer is the busiest air travel season in the U.S.
- Passengers often face frequent delays and cancellations
- Understanding disruption patterns can help travelers make smarter decisions

Objective:

- Predict the probability of flight cancellation
- Predict departure and arrival delay durations
- Provide actionable travel recommendations based on predictions

Data Scope:

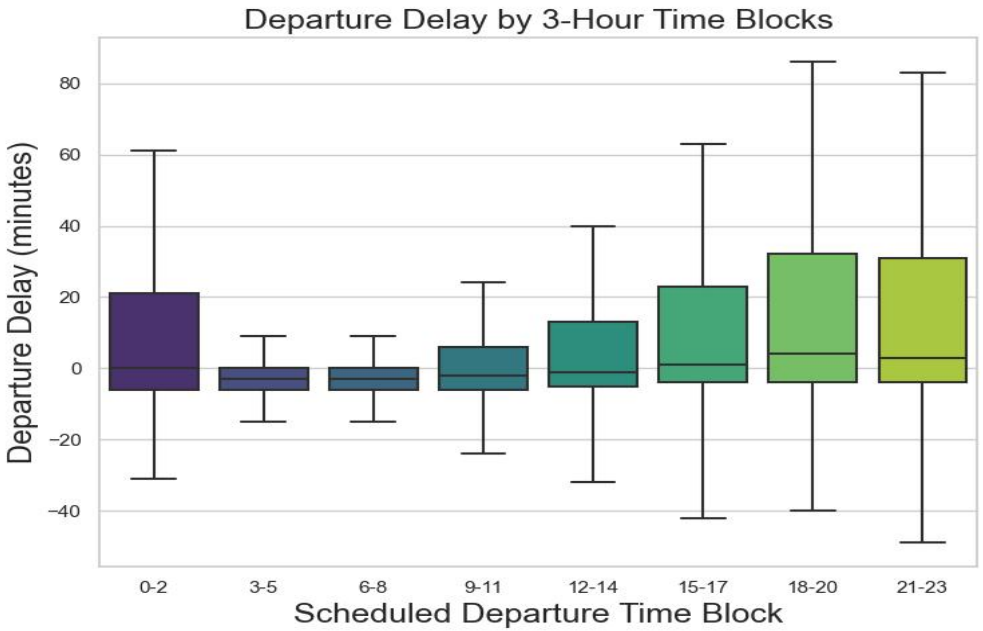
- Focused on flights from May to August 2024, Captures the peak summer travel season in the U.S.

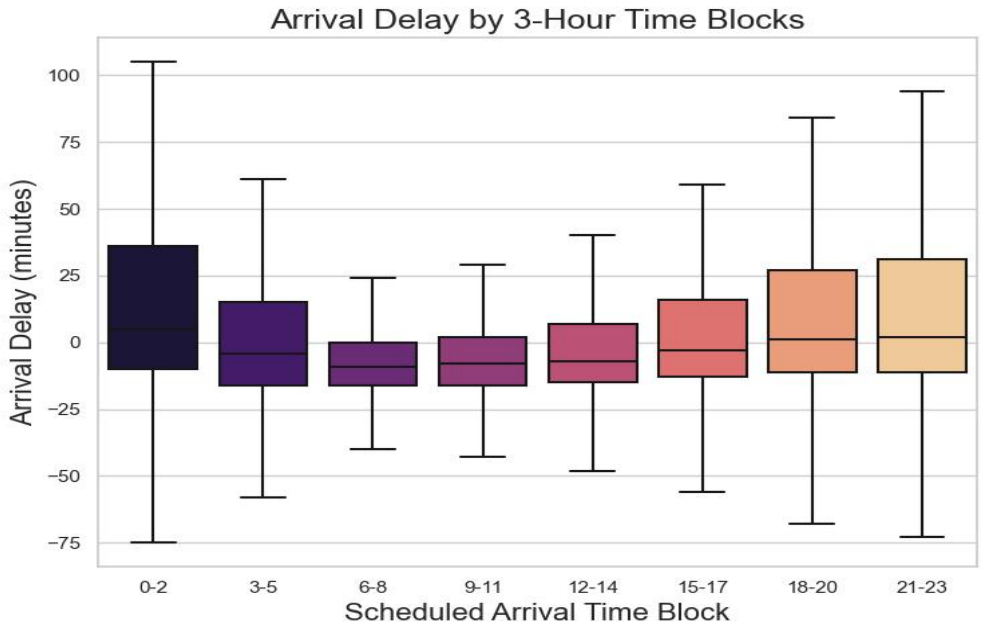
Data Sources & Enrichment:

- Core dataset from the U.S. Department of Transportation
- Includes flight schedules, actual times, and airline info
- Weather data was added using the Meteostat package, based on airport latitude and longitude

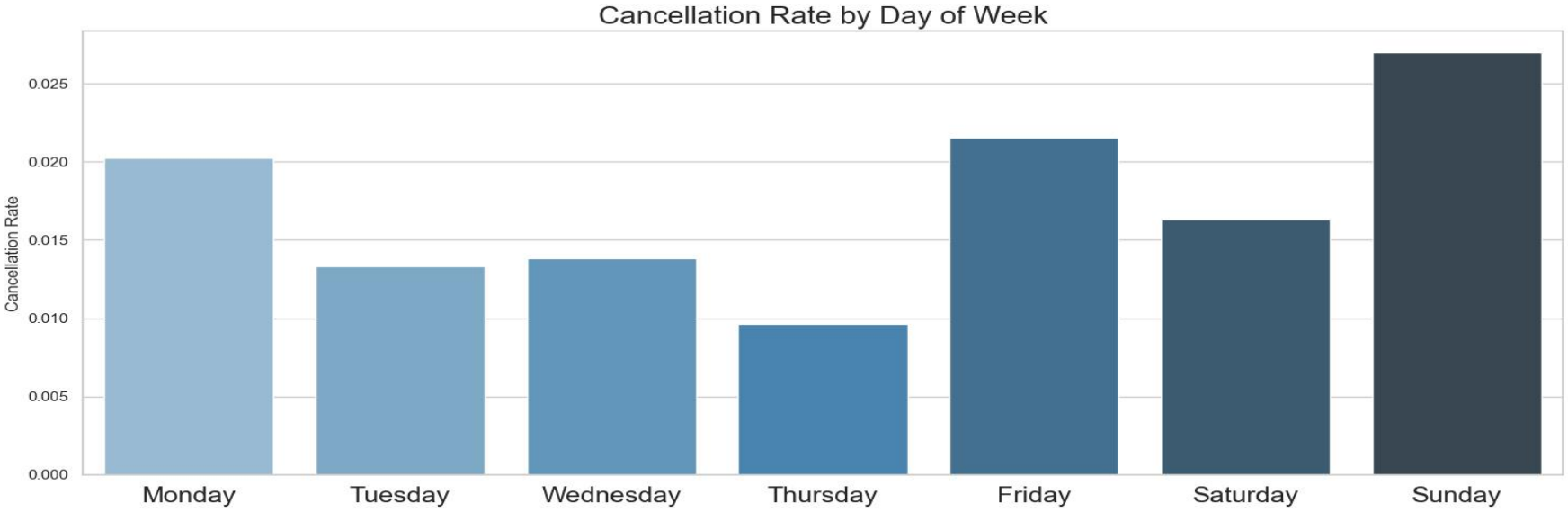
Data Visualization: Time Patterns

Delays by 3-Hour Time Blocks



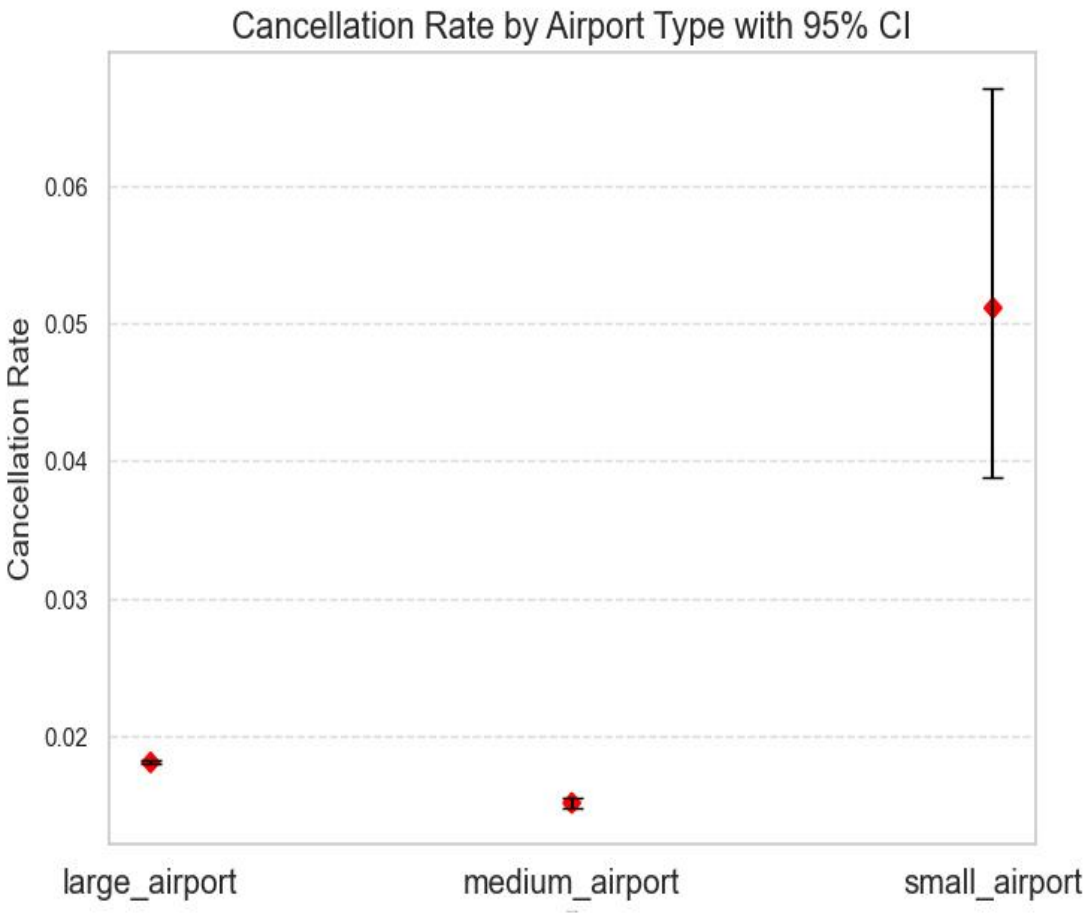


Cancellation Patterns by Weekday

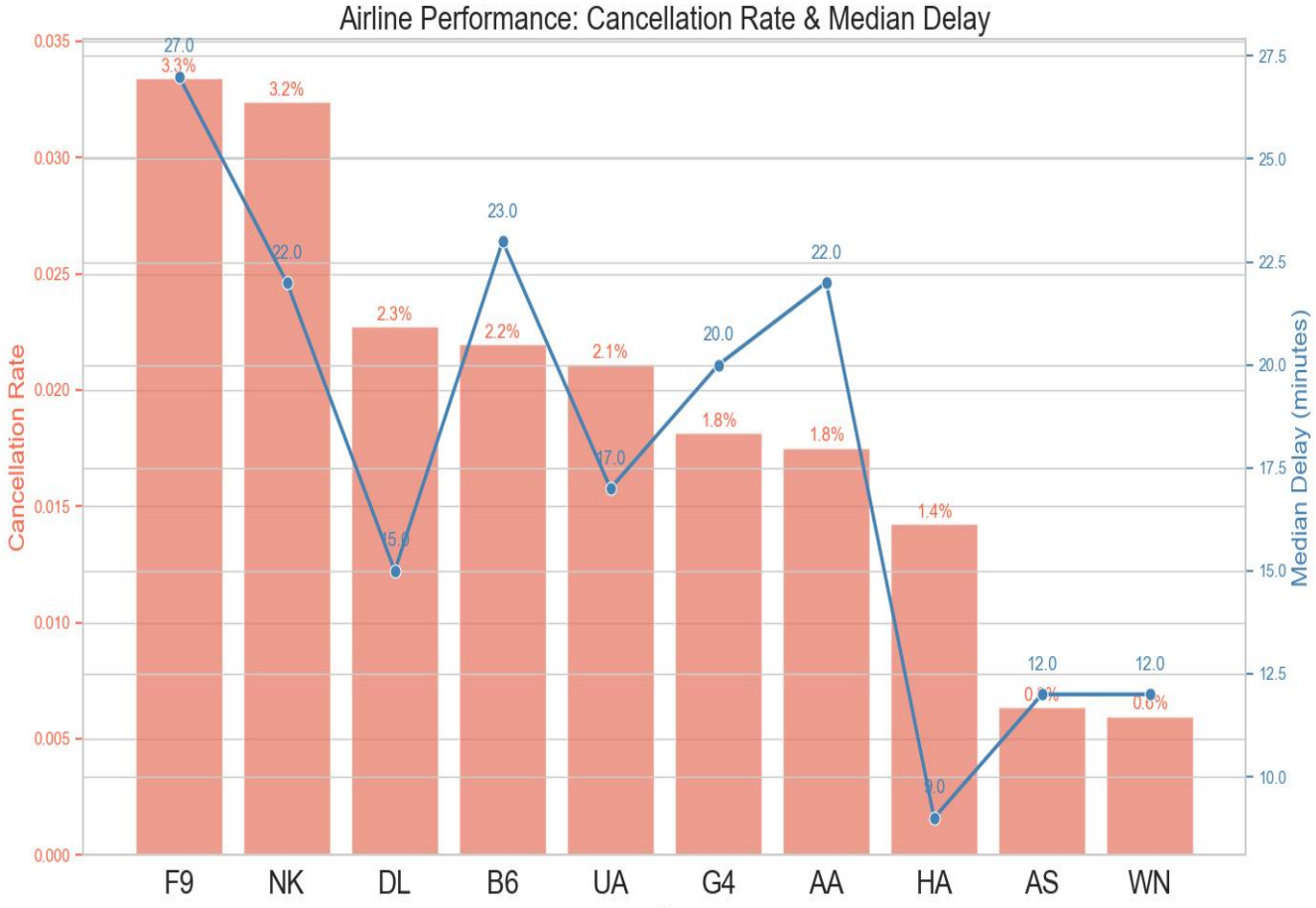


Data Visualization: Airline & Airport Performance

Cancellation by Airport Type

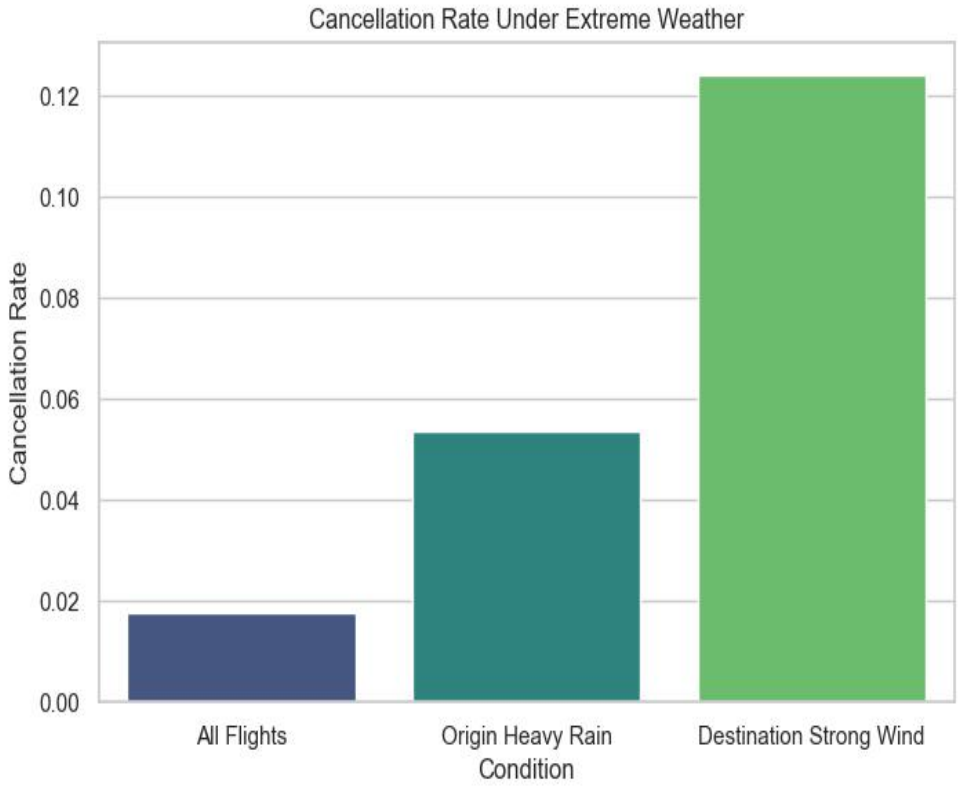


Airline Cancellation & Delay

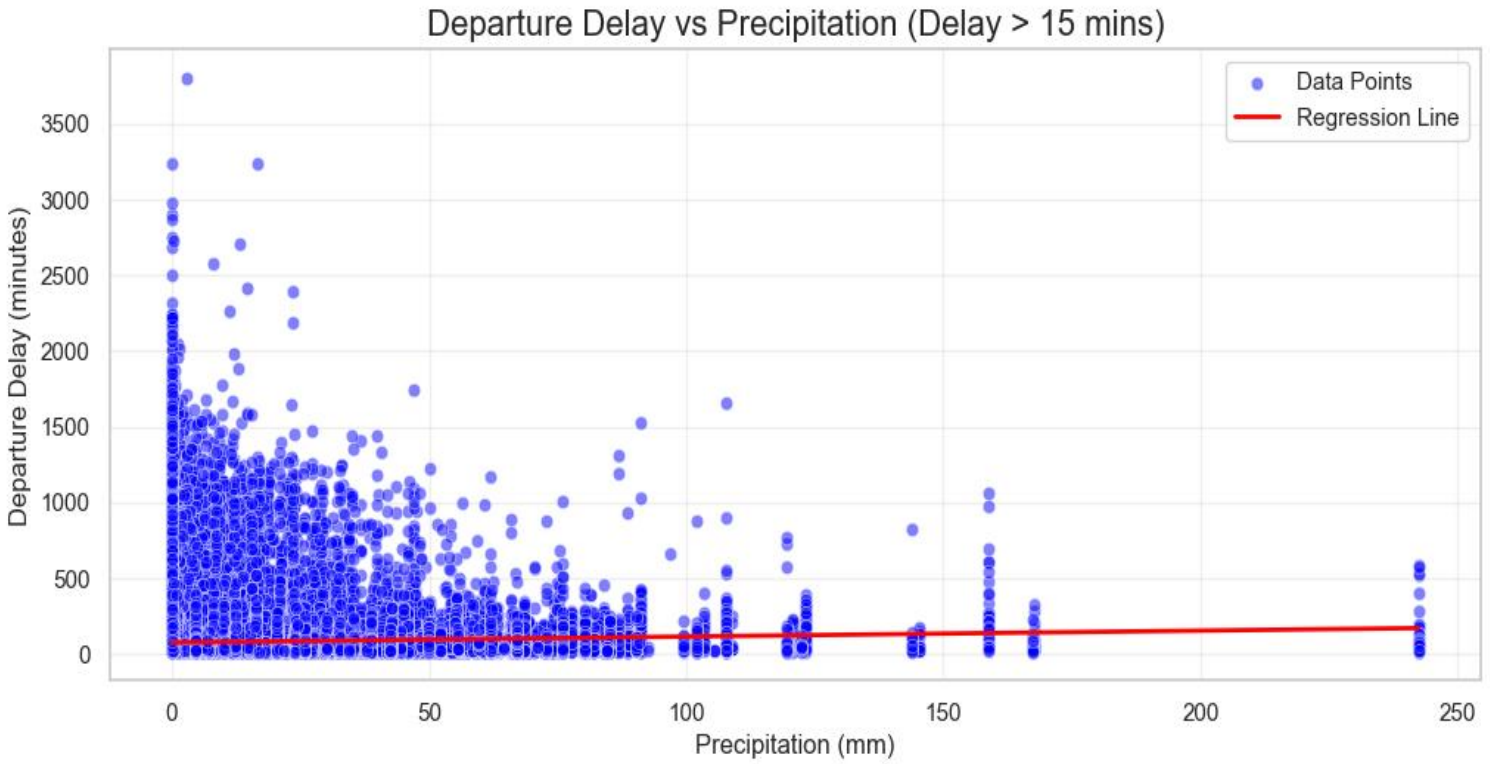


Data Visualization: Weather Impact

Cancellation Rate Under Extreme Weather



Departure Delay vs. Precipitation



Modeling approach - XGBoost

- Apply one-hot encoding and label encoding
- Handle data imbalance for classification model
- Remove outliers for regression models using the IQR method

Model	AUC_MSE
Cancellation	0.87
Dep_delay	136.07
Arr_delay	219.92

Cancellation	Dep_delay	Arr_delay
origin_snow_mm	MKT_AIRLINE	MKT_AIRLINE
dest_snow_mm	SCH_DEP_TIME	SCH_DEP_TIME
small_airport	DEP_HOUR	DEP_HOUR
WEEK_Sun	ORIGIN_IATA	ORIGIN_IATA
medium_airport	medium_airport	precipitation_mm

Modeling approach - XGBoost

- Models without weather use base information on flights as inputs
- Models with weather add weather features such as snow, precipitation, air temperature, and wind speed to the inputs.
- The APP will detect if the entered departure time is within one week from the current day. If yes, the model with weather is called.

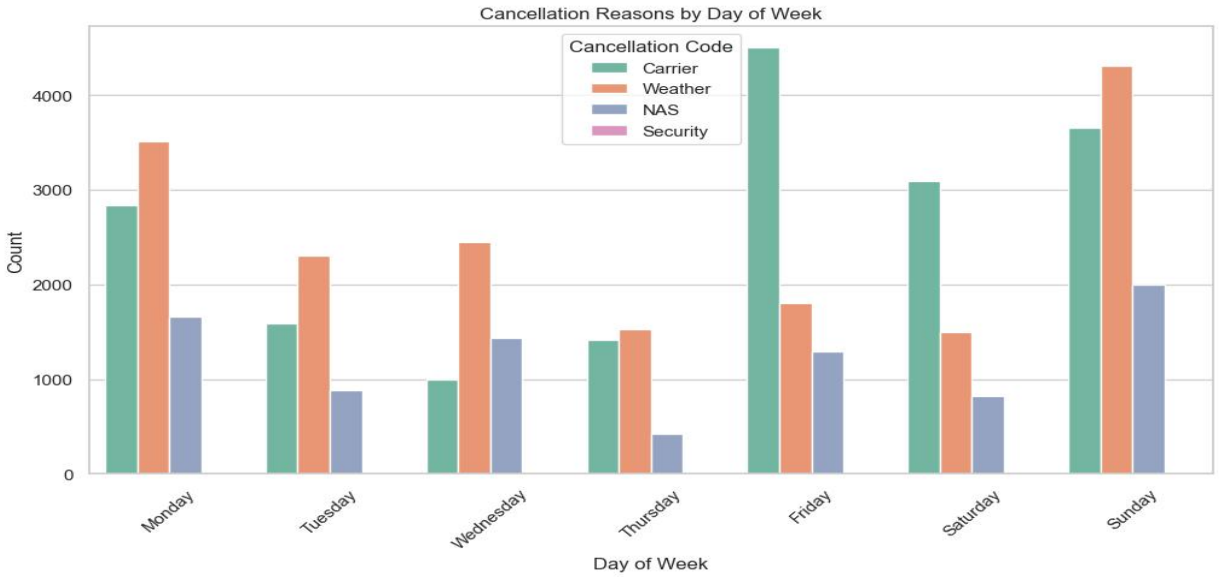
Version_1	Version_2
model_arr_no_weather	model_arr_weather
model_dep_no_weather	model_dep_weather
model_cancel_no_weather	model_cancel_weather

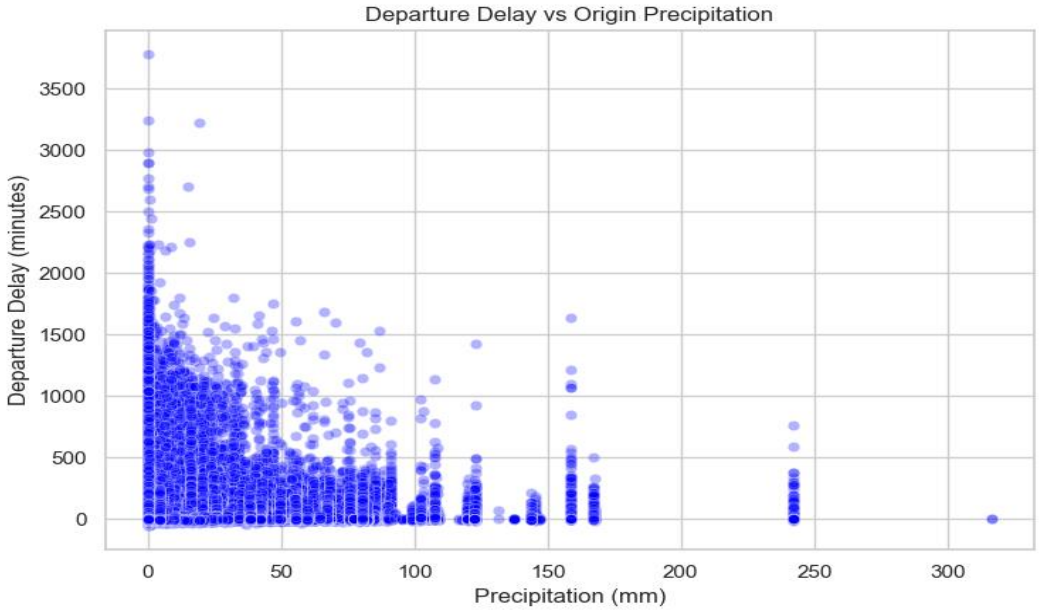
- In the regression model, we use the RMSE of the model to approximate the volatility of the forecast.

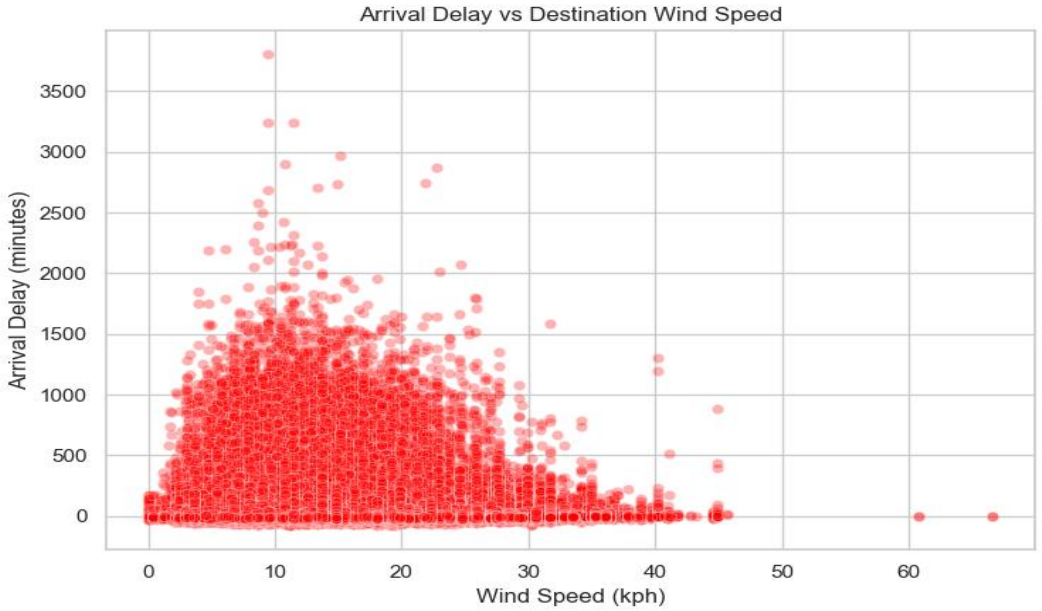
Thank you

Q&A

Backup


Cancellation Reasons by Day of Week
Cancellation Code
Carrier
Weather
NAS
Security
Count
0
1000
2000
3000
4000
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of Week


Departure Delay vs Origin Precipitation
Departure Delay (minutes)
0
500
1000
1500
2000
2500
3000
3500
0
50
100
150
200
250
300
Precipitation (mm)


Arrival Delay vs Destination Wind Speed
Arrival Delay (minutes)
0
500
1000
1500
2000
2500
3000
3500
0
10
20
30
40
50
60
Wind Speed (kph)

Note: Only use red icons on white or light gray backgrounds