San Diego Departure Taxi Time Analysis

Agenda


Introduction
Data Description
Data analysis & Prediction
Recommendation
ETL Architecture

Introduction

San Diego International Airport is a single runway airport servicing ~300 flights annually.

Taxi time in runway impacts delays, fuel usage and other costs

Objective

- Analyze the taxi out time for San Diego Airport
- Analyze potential factors impacting taxi out time
- Predict the future taxi out time
- Recommend potential exercises to decrease taxi out time
- Proposal for the E2E ETL/ML pipeline for automated predictions and BI dashboard

Data Description

Datasets used

1. Flights Data: The data provided contains every departure from San Diego International Airport between 1/1/2017 and 12/31/2018.

2. Airline Names: Names obtained using IATA codes[2]

3. Events and Holiday Days: Major events dates in San Diego (like Comic con) and Public Holidays to gauge impact on Taxi time

2. https://www.iata.org/en/publications/directories/code-search/

Data Analysis and Insights

Features Considered

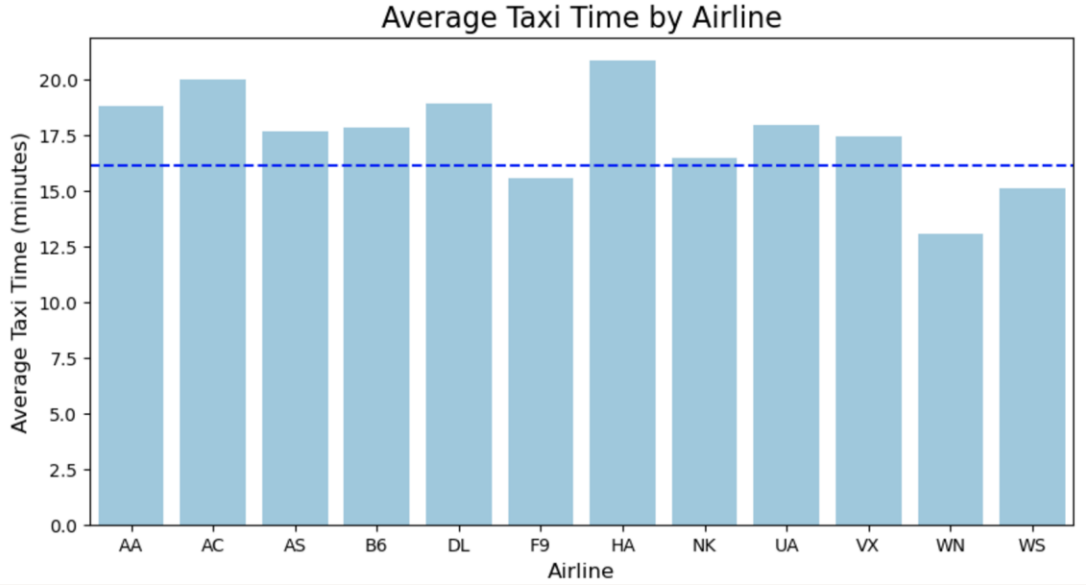
1. Date and Time components- Time of day, Day of week, etc.

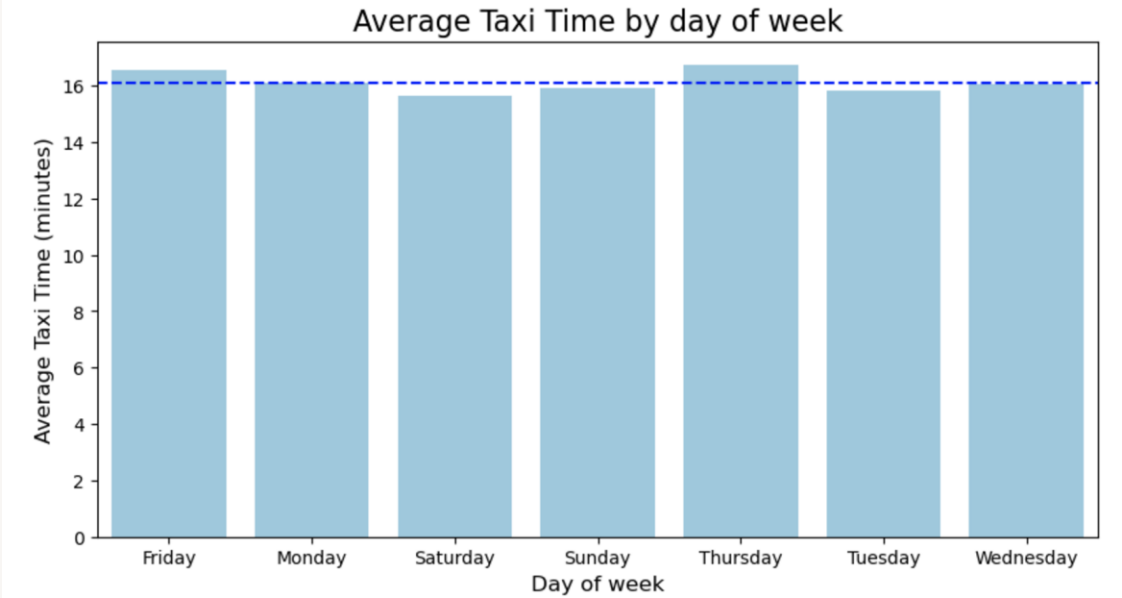
2. Airport congestion metrics- Number of other departures at same time as each flight [1]

3. Events and Public Holidays

1. https://www.mit.edu/~hamsa/pubs/ClewlowSimaiakisBalakrishnanGNC2010.pdf

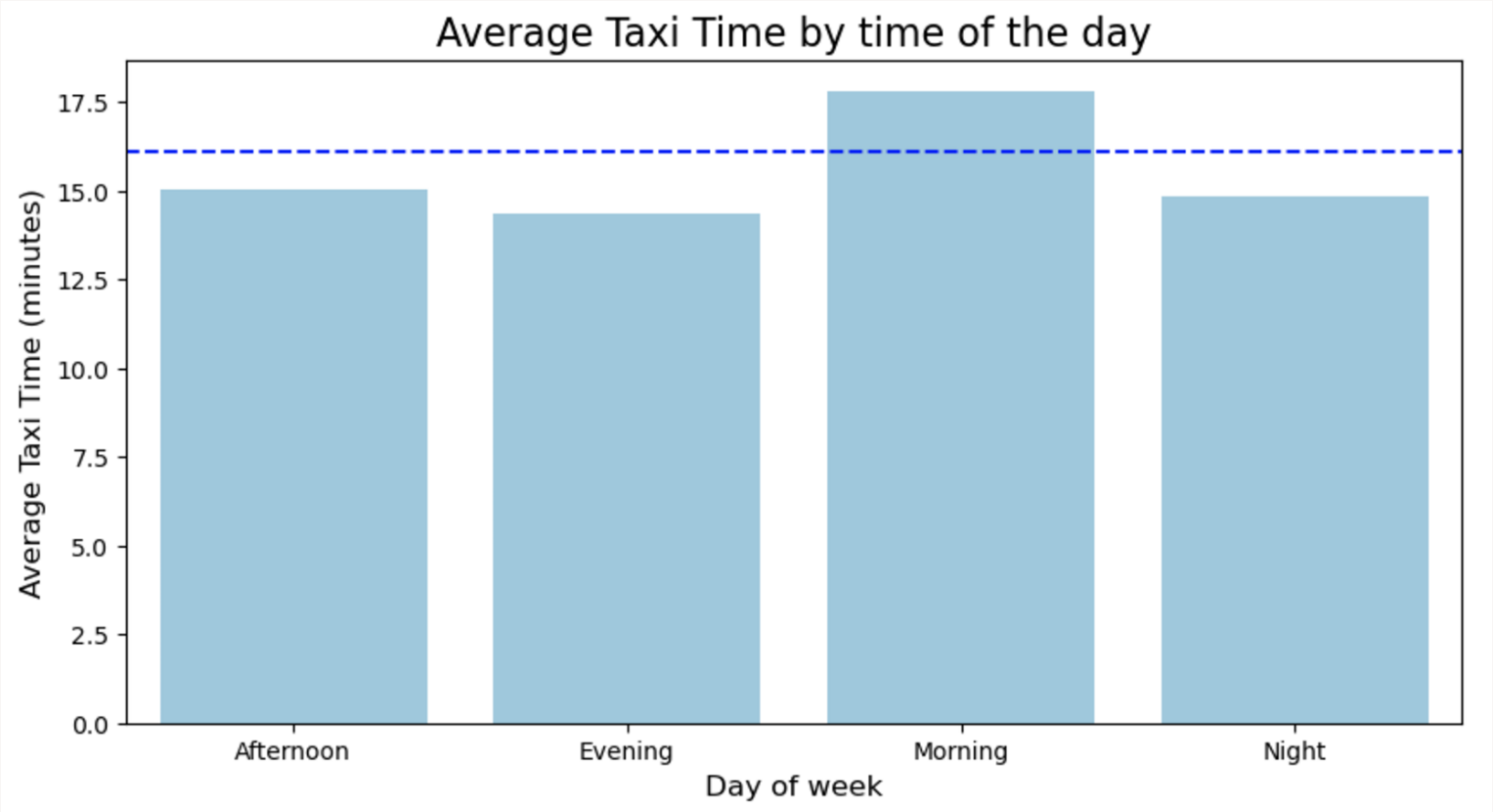
Exploratory Data Analysis





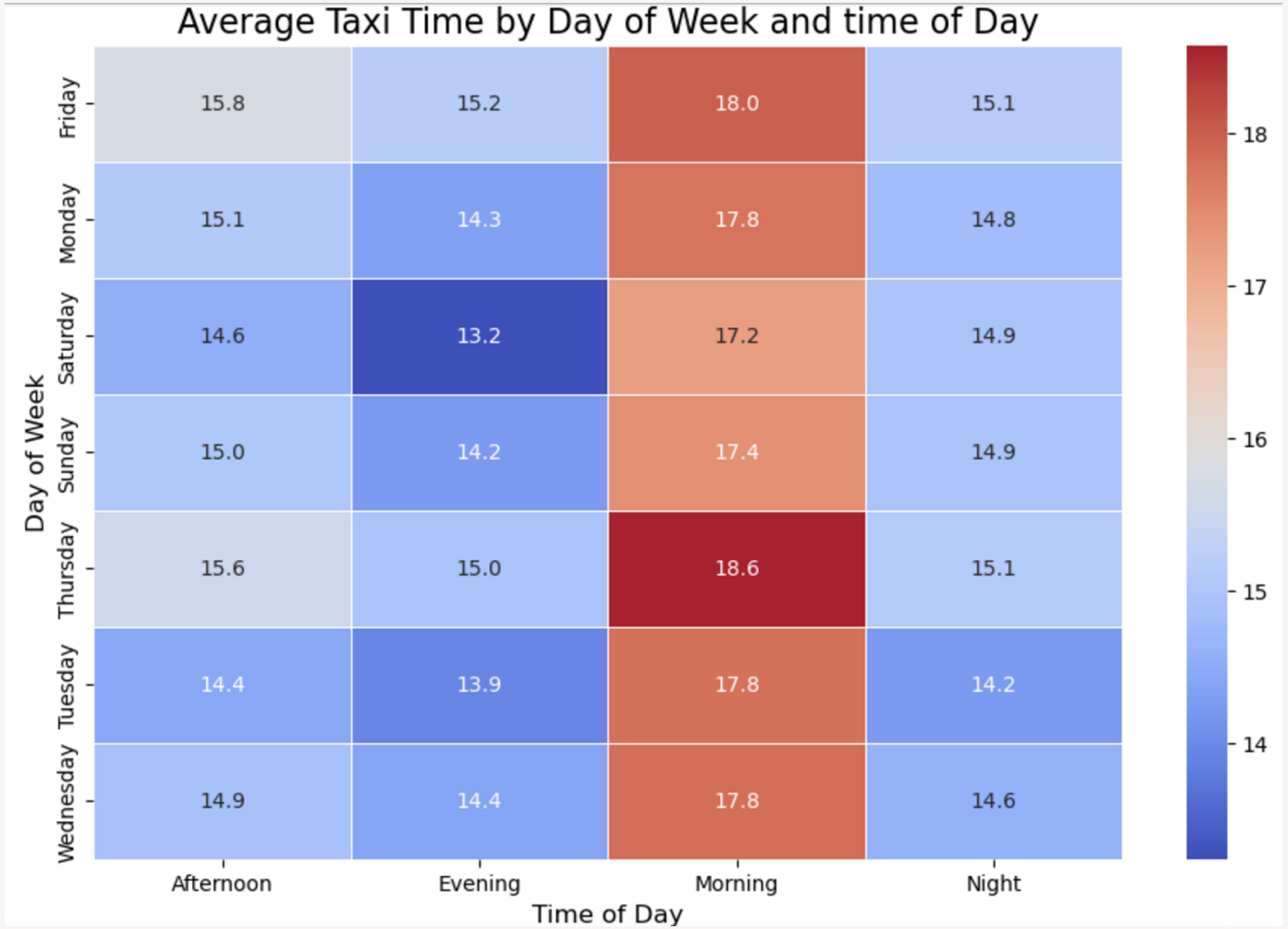
No Correlation seen with Airlines or Day of the week

Exploratory Data Analysis

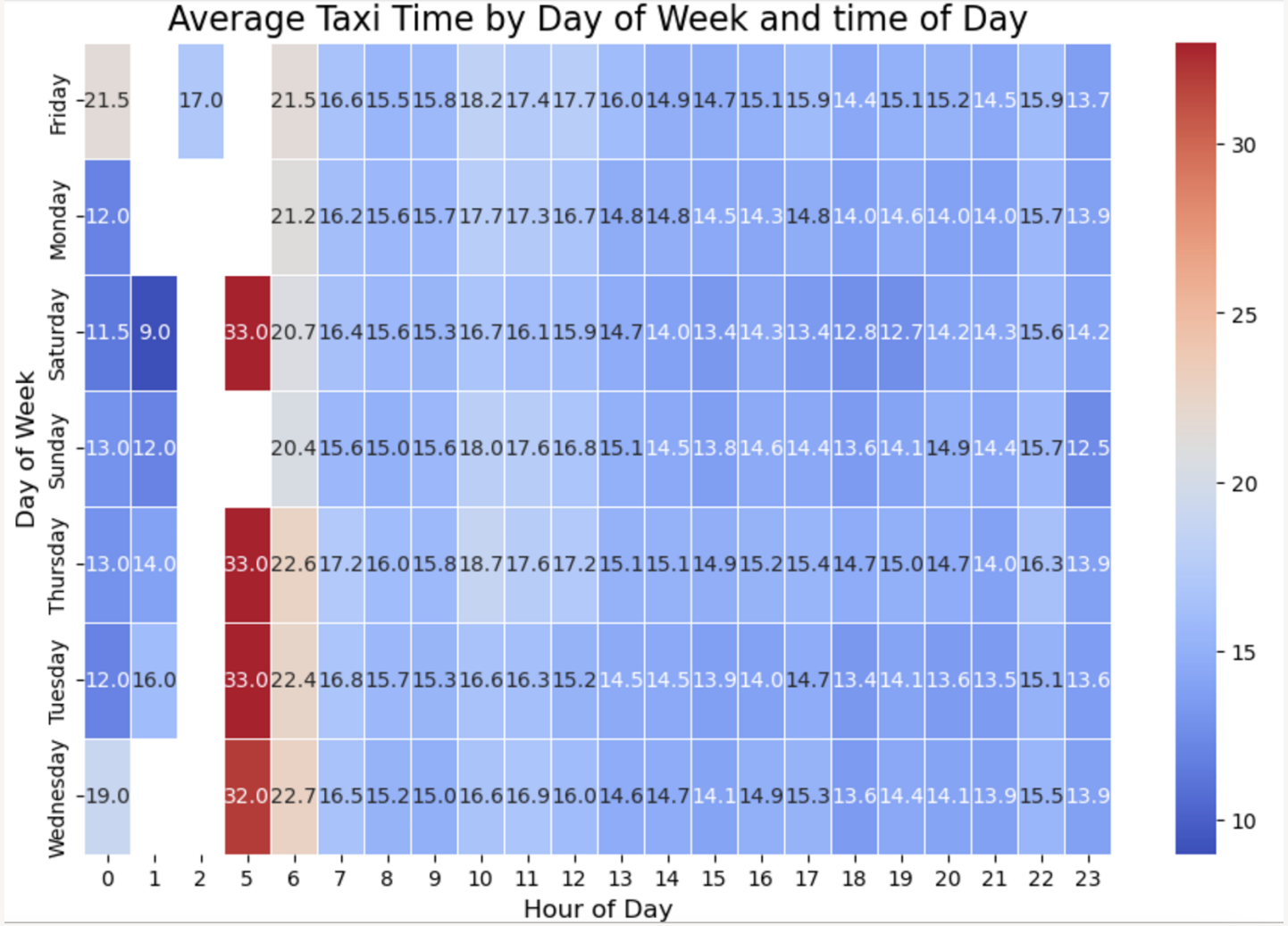


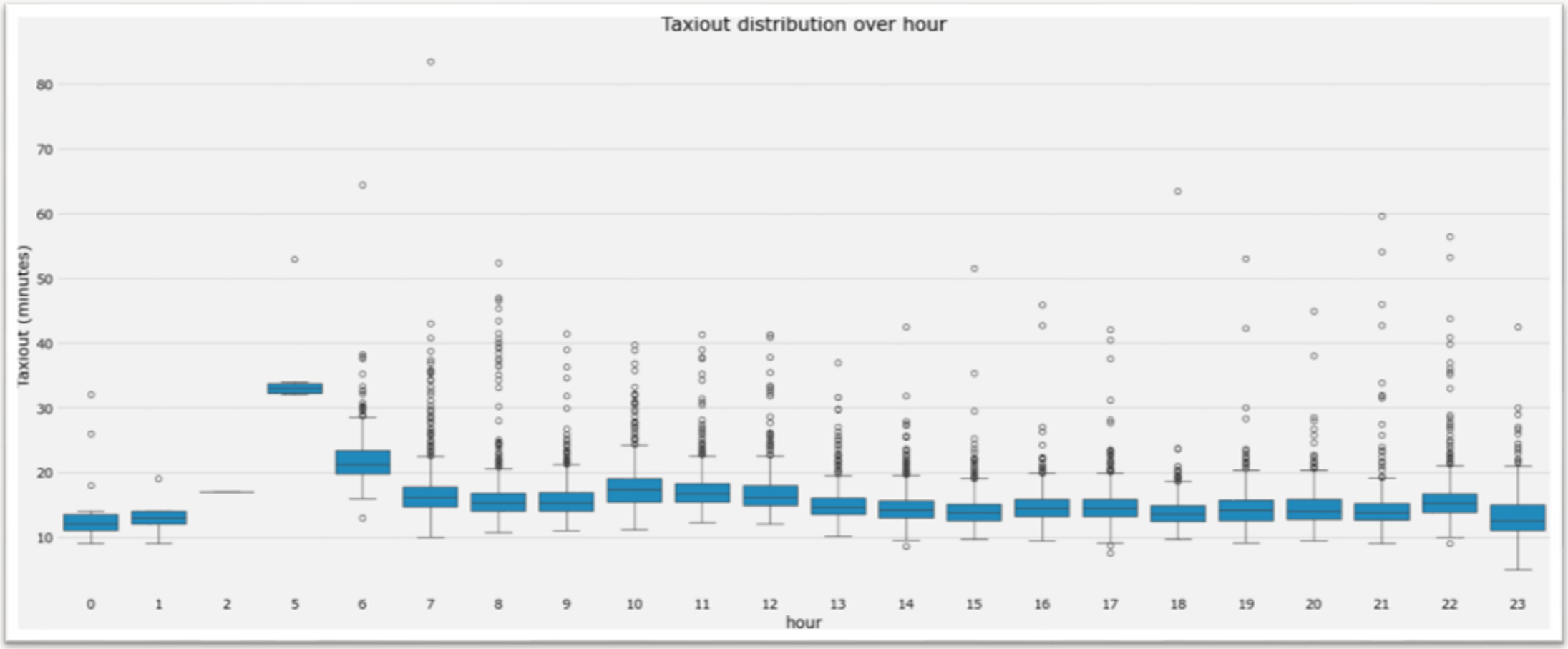
Morning: 4 AM to 12 PM
Afternoon: 12 PM to 5 PM
Evening: 5 PM to 9 PM
Night: After 9 PM

Exploratory Data Analysis

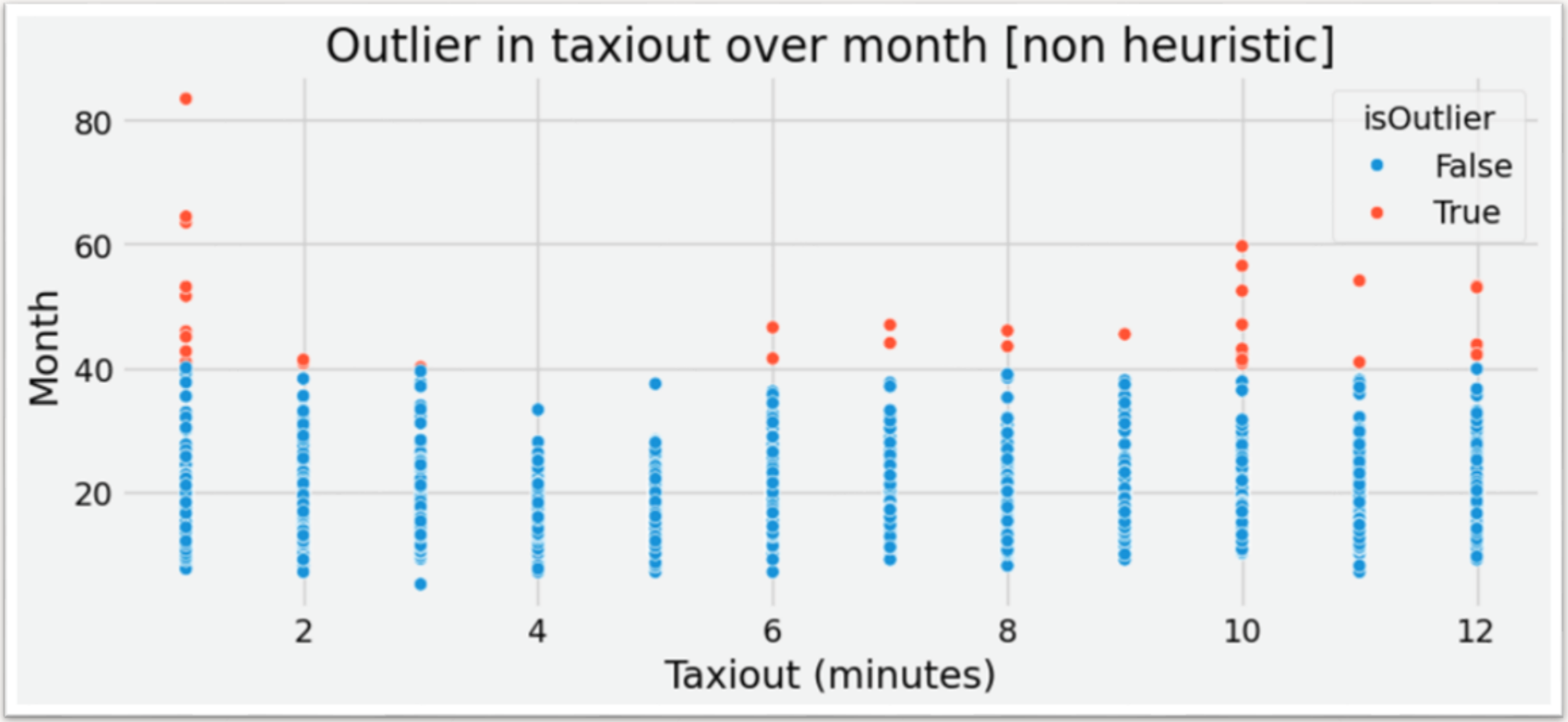


Exploratory Data Analysis

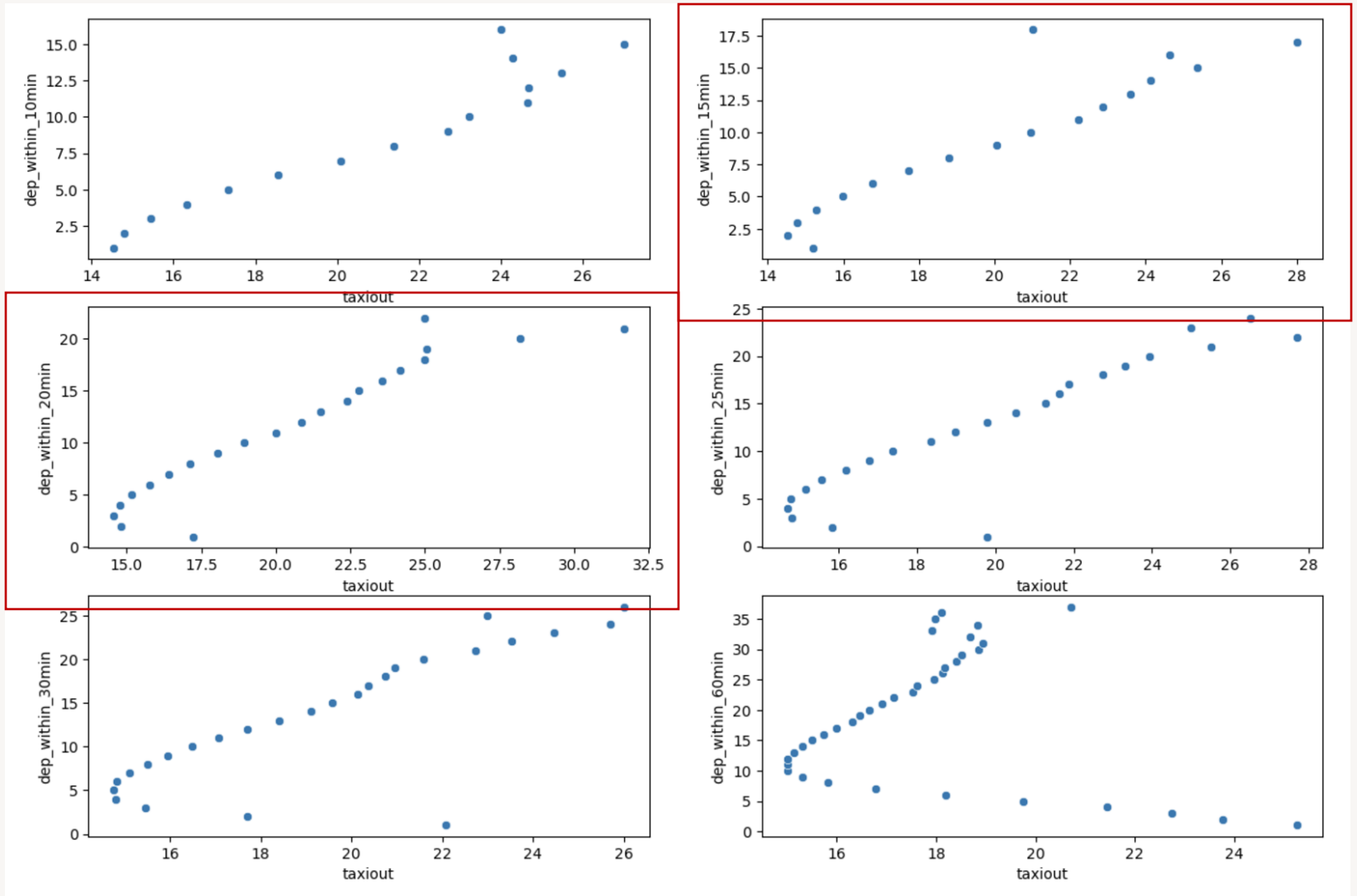




- Almost no traffic between 2nd, 3rd and 4th hour
- Consistently high taxi time between 5th and 6th hour



- Less outliers (high taxiout) from February to September
- High outliers in January



- No. of departures within given time window for each flight
- High correlation between Taxi out time and no. of departures from before 15-20 min of each flight

EDA Conclusion

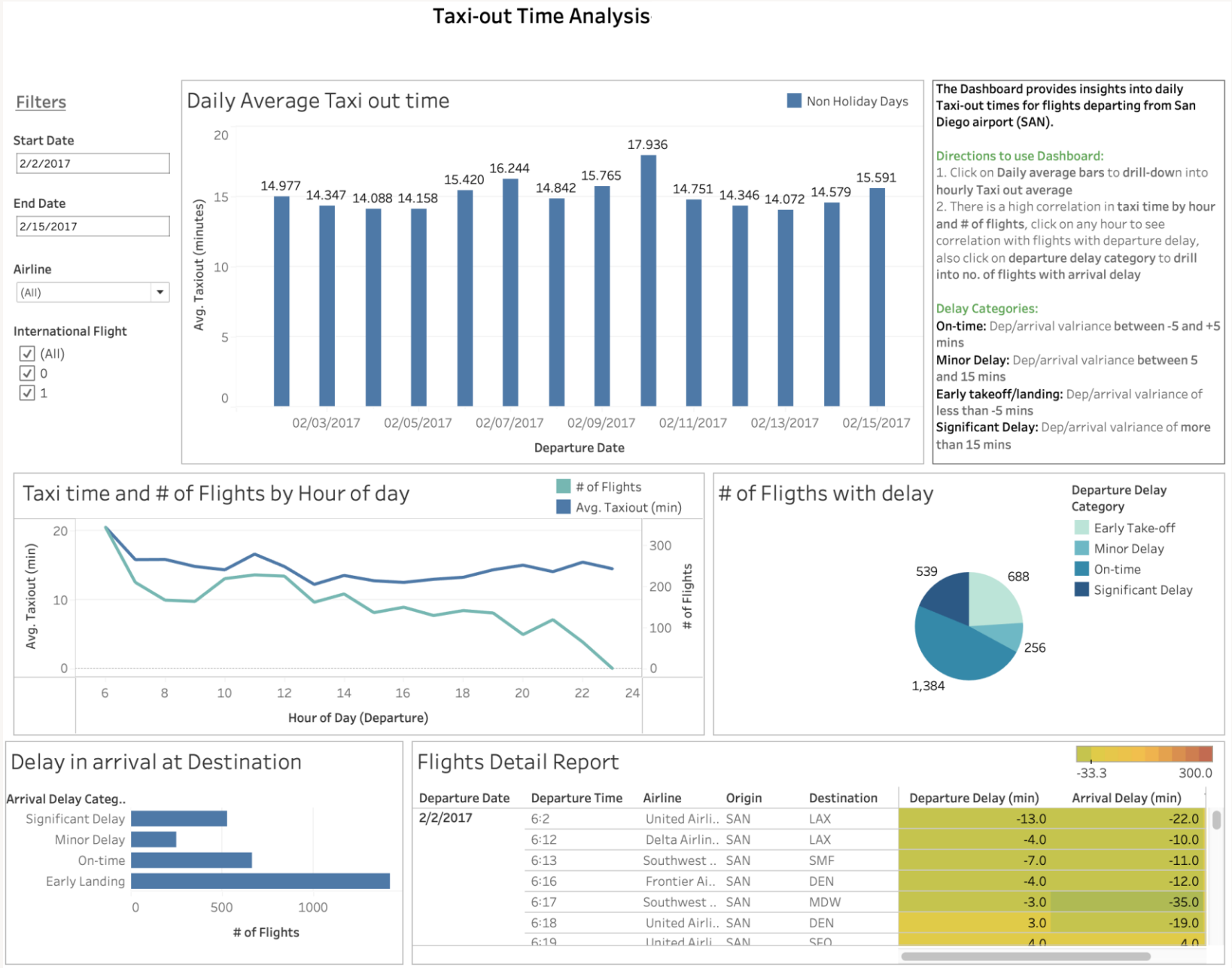
The features with most impact of Taxi out Time:

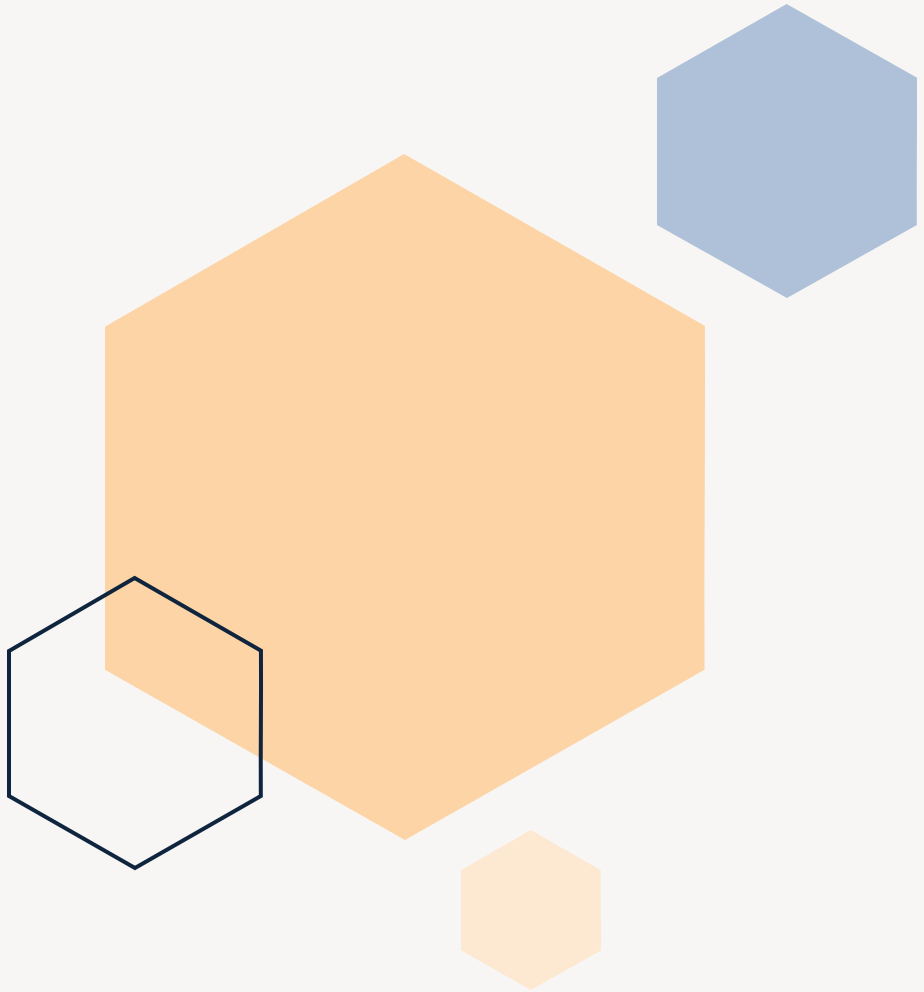
1. Time of the Day – 5th and 6th hour

2. Airport congestion from 15-20 mins before flight departure

3. Seasonality impact- January shows high Taxi out time

Tableau Dashboard for Insights





Taxi Time Prediction

Data Preparation and Model Features

Goal: Predict average taxi out time per hour into couple of months in the future

Features:

1. Date time metrics [Hour of day, Day of week, Week of year, Month]
2. Lag features
    1. Average taxi out time for the previous year (364 days)
    2. Fuzzy average taxi out time from previous year [Average taxi time over +- 7 days]
    3. Fuzzy max taxi out time from previous year [Max taxi time over +- 7 days]
3. Congestion metrics [Number of flight departures within X minutes of the flight

    This metric can't be used for date range outside of the given dataset as we don't have schedules for that time [2019].

Data Preparation and Model Features

Goal: Predict average taxi out time per hour into couple of months in the future

Features:

Model1

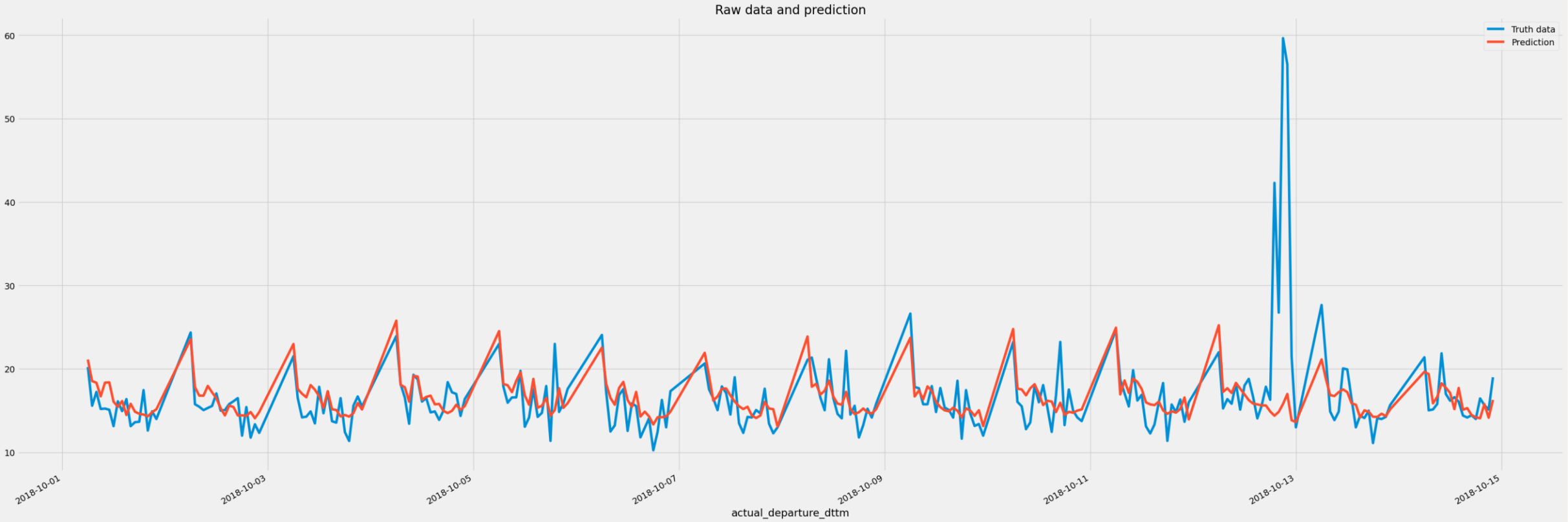
1. Date time metrics [Hour of day, Day of week, Week of year, Month]

Model 2- forecast

2. Lag features
    1. Average taxi out time for the previous year (364 days)
    2. Fuzzy average taxi out time from previous year [Average taxi time over +- 7 days]
    3. Fuzzy max taxi out time from previous year [Max taxi time over +- 7 days]
3. Congestion metrics [Number of flight departures within X minutes of the flight

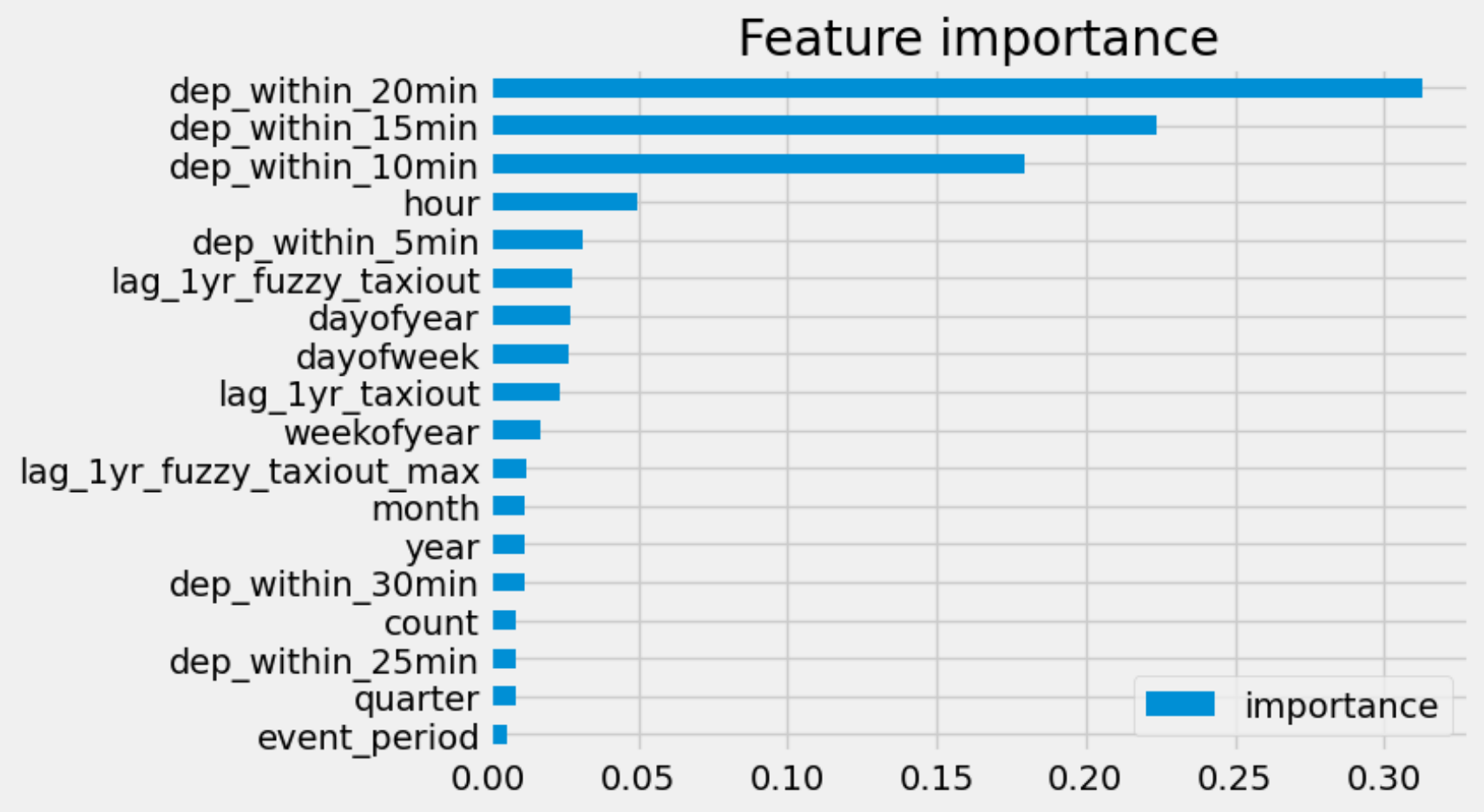
    This metric can't be used for date range outside of the given dataset as we don't have schedules for that time [2019].

Congestion metrics-based Model



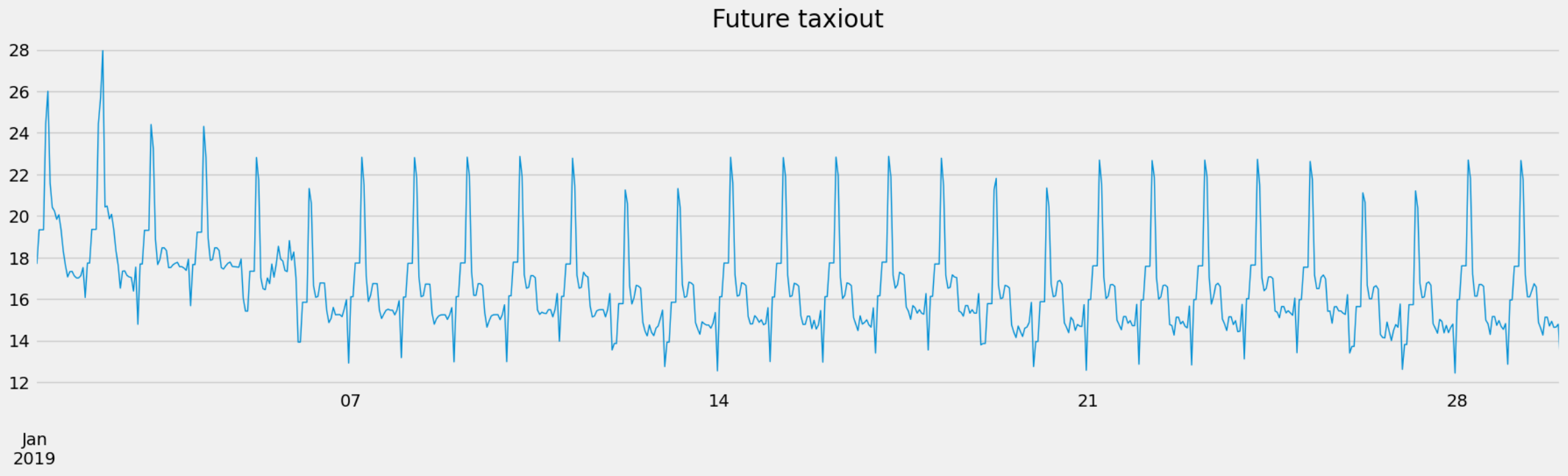
Model used: XGBoostRegressor
Root mean squared error: 3.704

Congestion metrics-based Model: Feature Importance



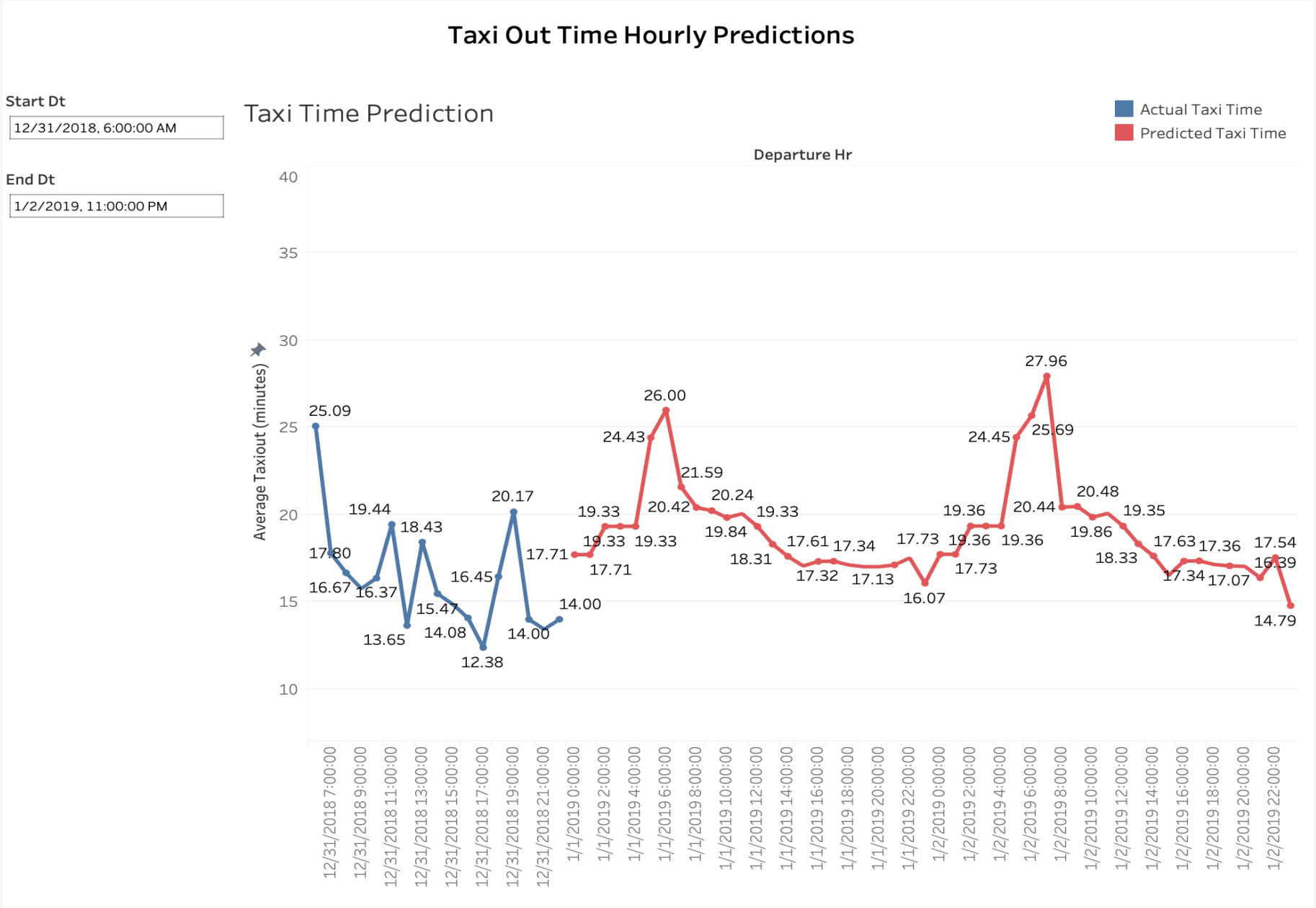
- Congestion metrics are considered very important

Model with no congestion metrics: 2019 Jan Forecast

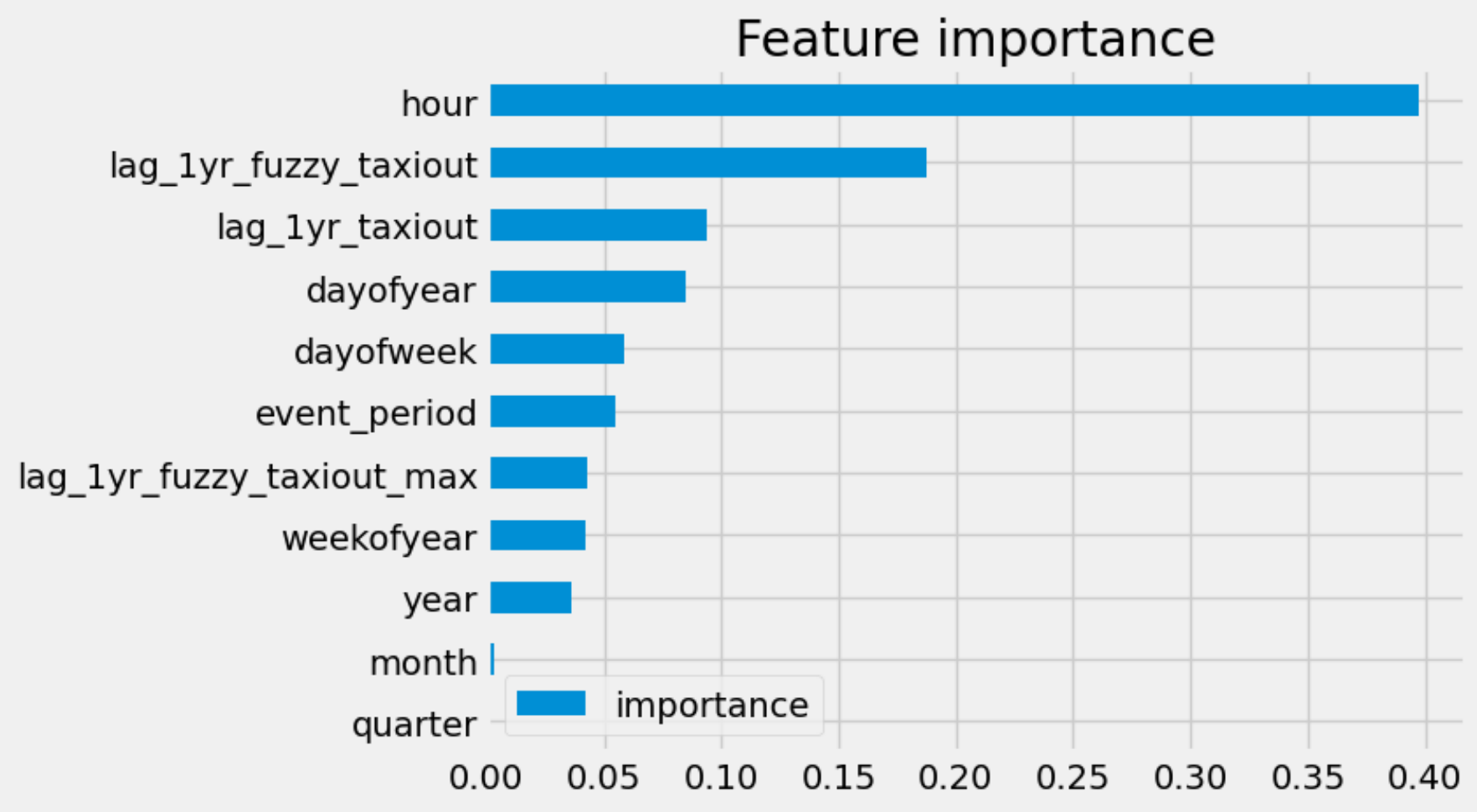


Model used: XGBoostRegressor
RMSE: 3.7306

Showing Forecast in Tableau Dashboard



Model with no congestion metrics: Feature Importance



- Without congestion metrics, hour of day and lag features become more important

Recommendations

Increasing Airport/Ground Staff

- Increasing staff in morning times could enhance coordination and efficiency in managing aircraft movements.
- This would help in reduces bottlenecks during peak time and ultimately shortens taxi times.
- Assumption: There is limited airport staff at early morning shifts at 5 AM

Managing Departure times

- Scheduling departures more than 20 minutes apart from each other if possible
- For multiple Morning flights schedules within 20-minute window, departing later could potentially reduce taxi time and fuel consumption

Recommendations for Additional Data Analysis

1. Weather Data

- Harsh weather conditions like heavy rain can have impact on taxi time
- Include historical and forecasted weather Data

2. Airport Details

- Number of Runways
- Gates and location (distance from runway)

3. Analyze results of different ML Models

- Facebook Prophet, AWS Forecast

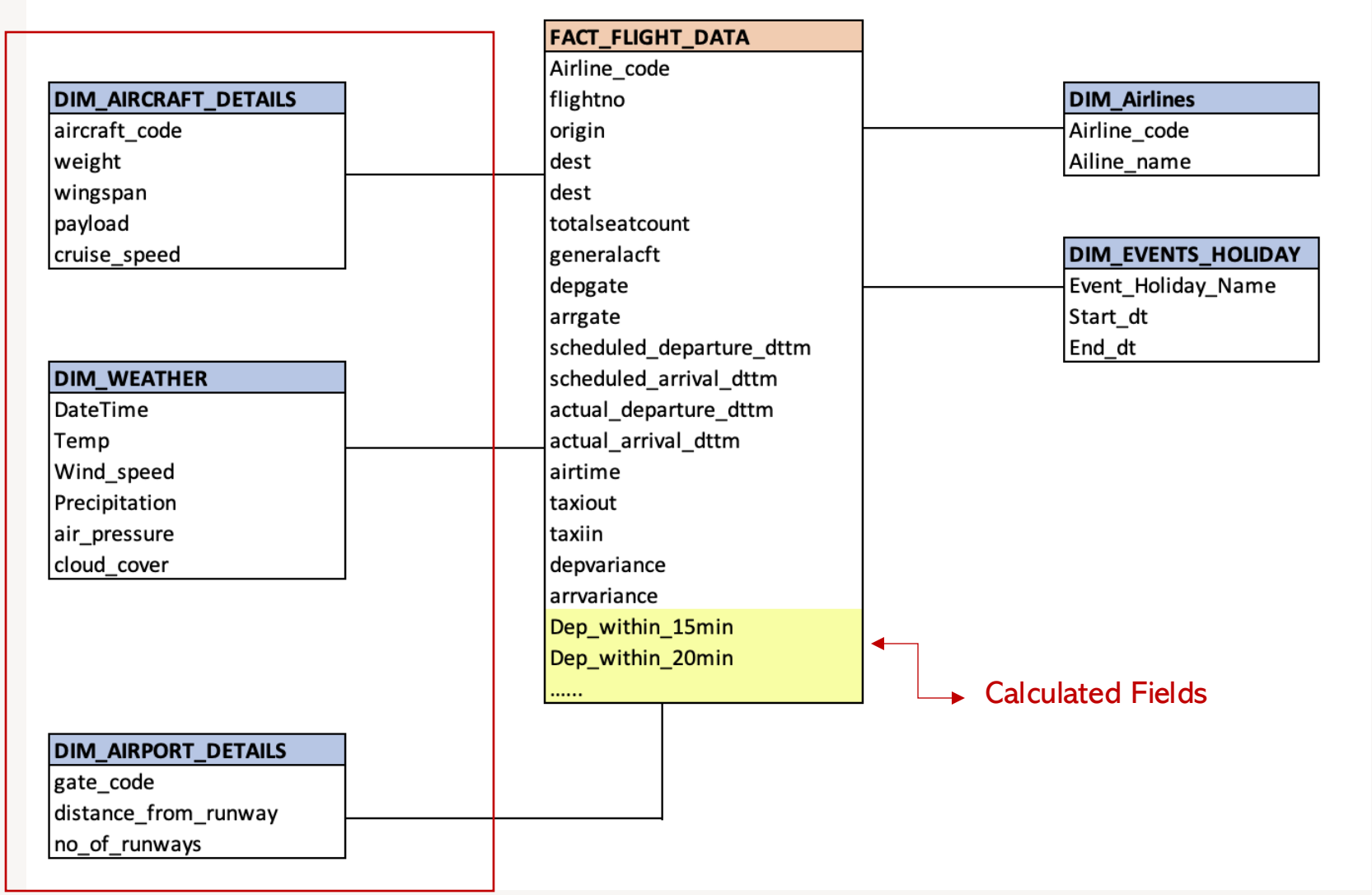
4. Arrivals at Airport

- In addition to departures, include all arrivals at the airport to check congestion impact

5. Events/Holiday Data

- Popular sports and events happing in city and public holiday information

Data Model


Suggested Data Dimensions
DIM_AIRCRAFT_DETAILS
aircraft_code
weight
wingspan
payload
cruise_speed
DIM_WEATHER
DateTime
Temp
Wind_speed
Precipitation
air_pressure
cloud_cover
DIM_AIRPORT_DETAILS
gate_code
distance_from_runway
no_of_runways
FACT_FLIGHT_DATA
Airline_code
flightno
origin
dest
dest
totalseatcount
generalacft
depgate
arrgate
scheduled_departure_dttm
scheduled_arrival_dttm
actual_departure_dttm
actual_arrival_dttm
airtime
taxiout
taxiin
depvariance
arrvariance
Dep_within_15min
Dep_within_20min
.....
DIM_Airlines
Airline_code
Ailine_name
DIM_EVENTS_HOLIDAY
Event_Holiday_Name
Start_dt
End_dt
Calculated Fields

ETL Architecture