

# Traffic Flow Forecasting

Joshua, Clayton, Bryan



# Presentation Plan

Topics to be presented

01

Motivation

02

Set the  
Stage

03

Core  
Analysis

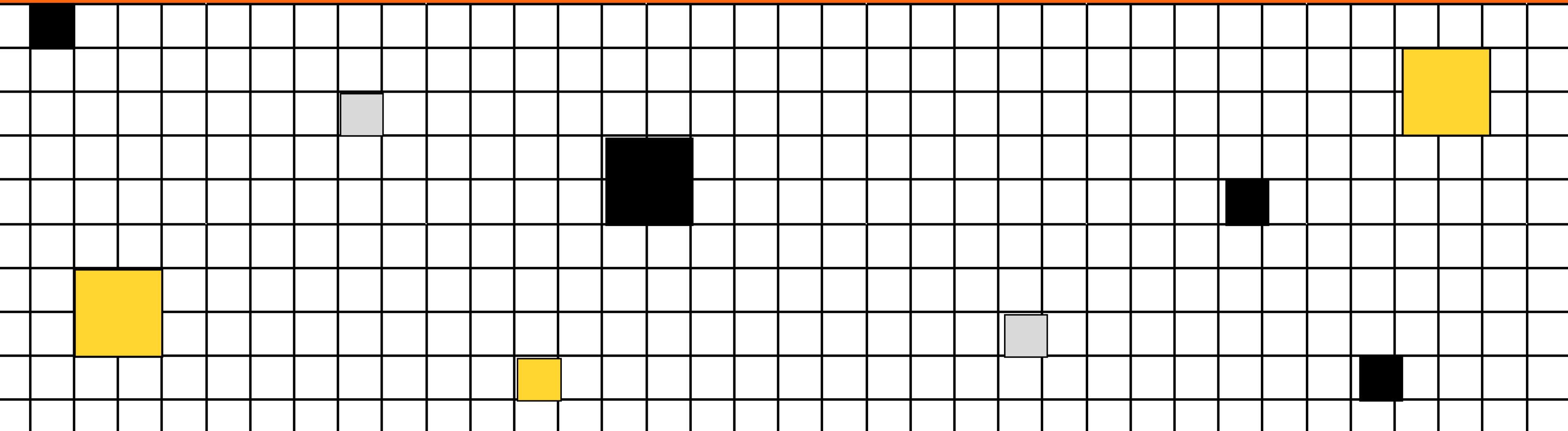
04

Conclusion



# Motivation

Why this Project?



# PEMS-08 Dataset



Location: San Bernardino

Time: July to August 2016

170 Locations

Traffic Data Captured every 5 minutes

# Probem Definition



**Ranking the locations in a city based on traffic occupancy.**

Formally, we formulated the following problem statement: For  $N$  prediction of the `occupy` variable and  $K$  locations, make a predictive model to minimize  $\sum_{t=0}^{N-1} \rho_t$ , where  $\rho_t$  is defined as the [Spearman correlation](#) of the prediction values and the actual values.

# Why Traffic?



Traffic  
Congestions



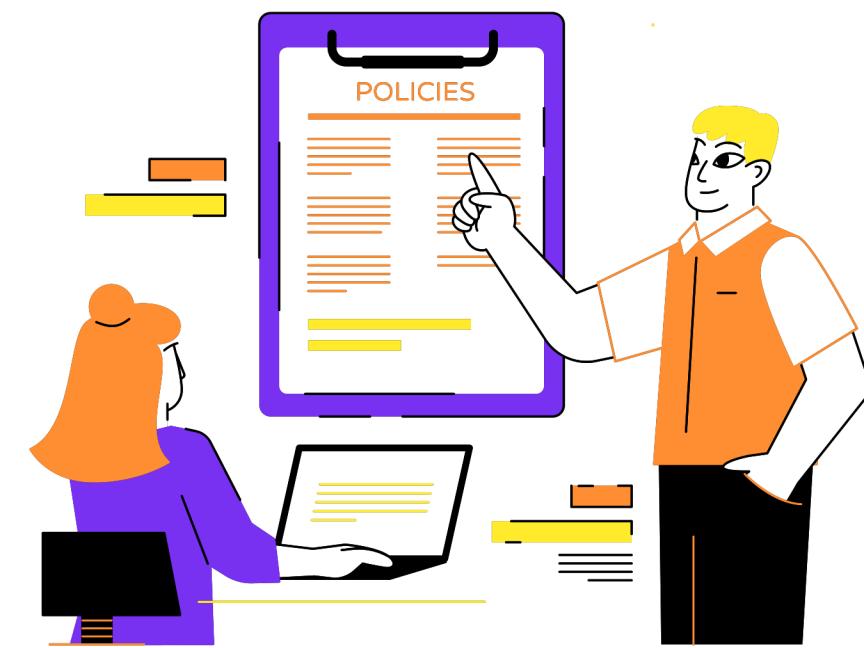
Road  
Accidents



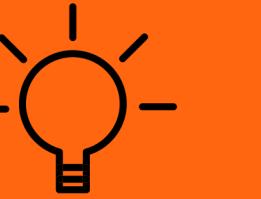
# Tools for Police & Government



Planning  
Patrol



Implement Key  
Policies



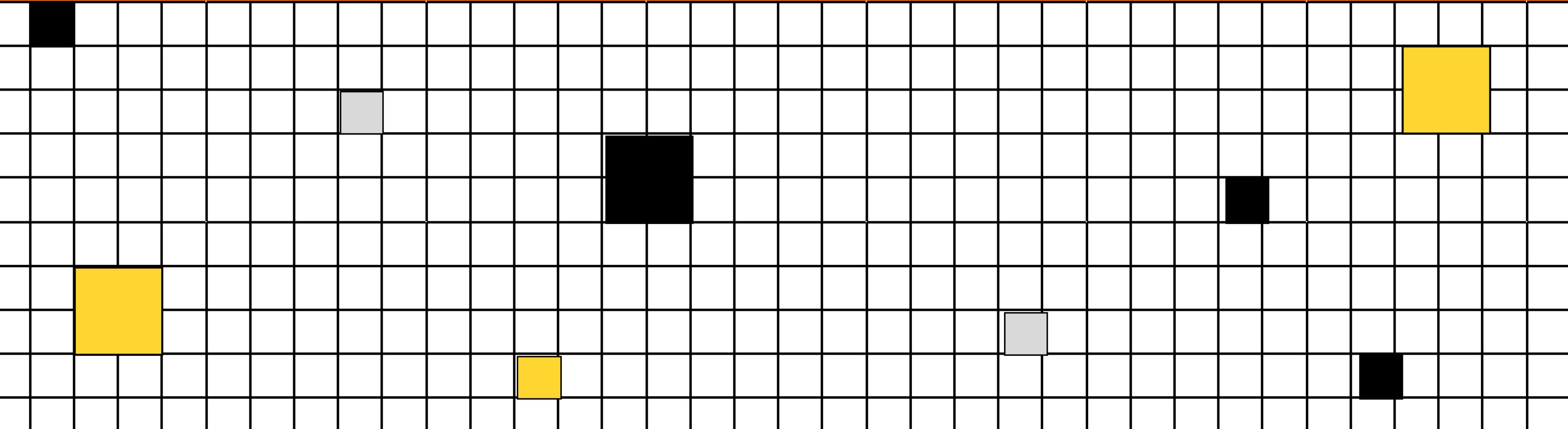
# Outcome

Improve road safety and reduce traffic congestions

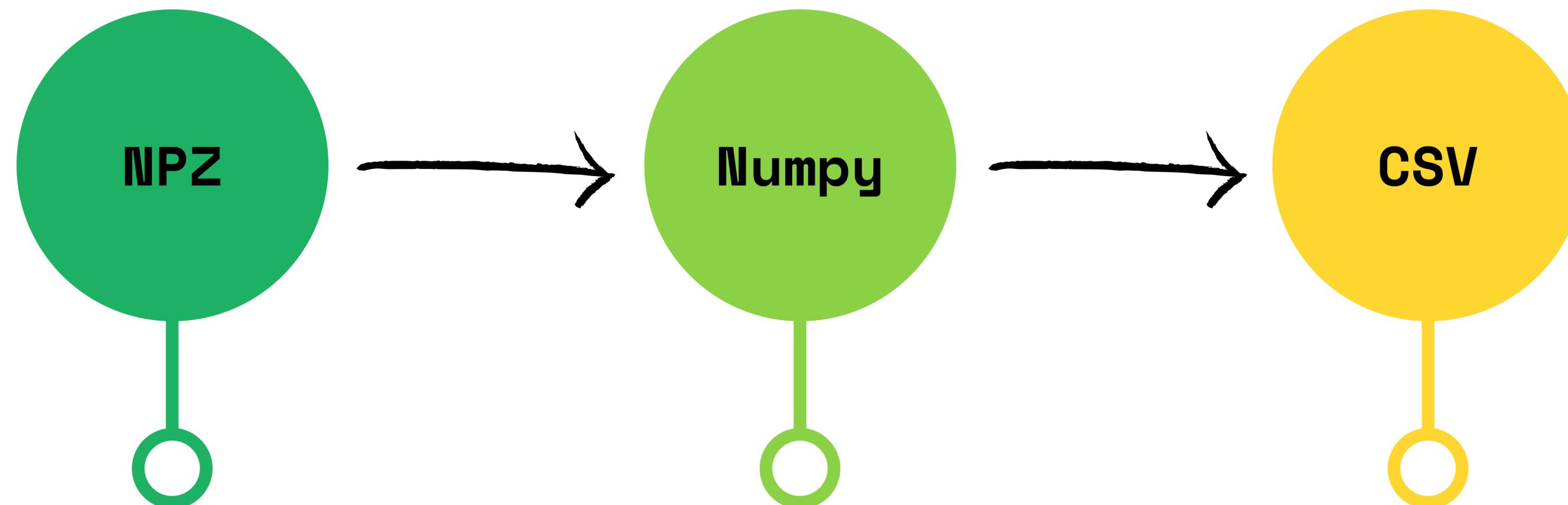


# Set the Stage

Data Exploration



# Data Preparation



# Dimension

(17856, 170, 3)

(Timesteps, Location, Features)

# Features

---

Occupy

Speed

Flow

# Data Types & Quantity

---

3 Million Rows

# Data Types & Quantity

---

Integer or Float

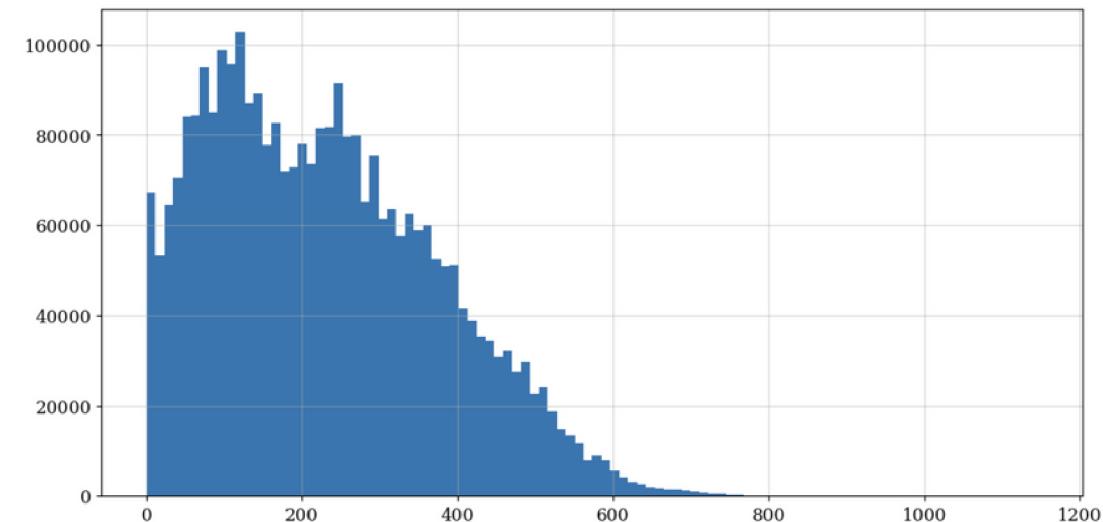
# Data Types & Quantity

---

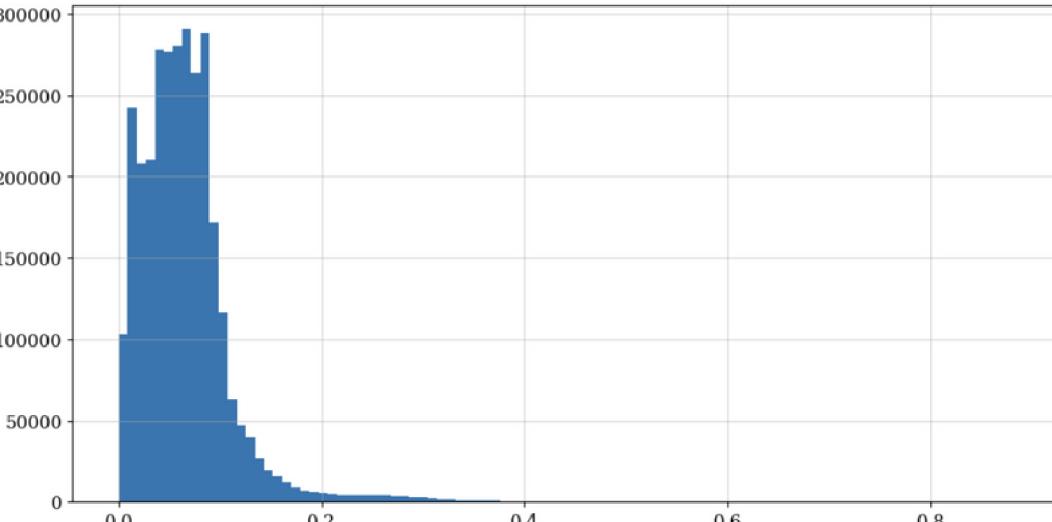
Numerical &  
Continuous Features

# Distribution of Features

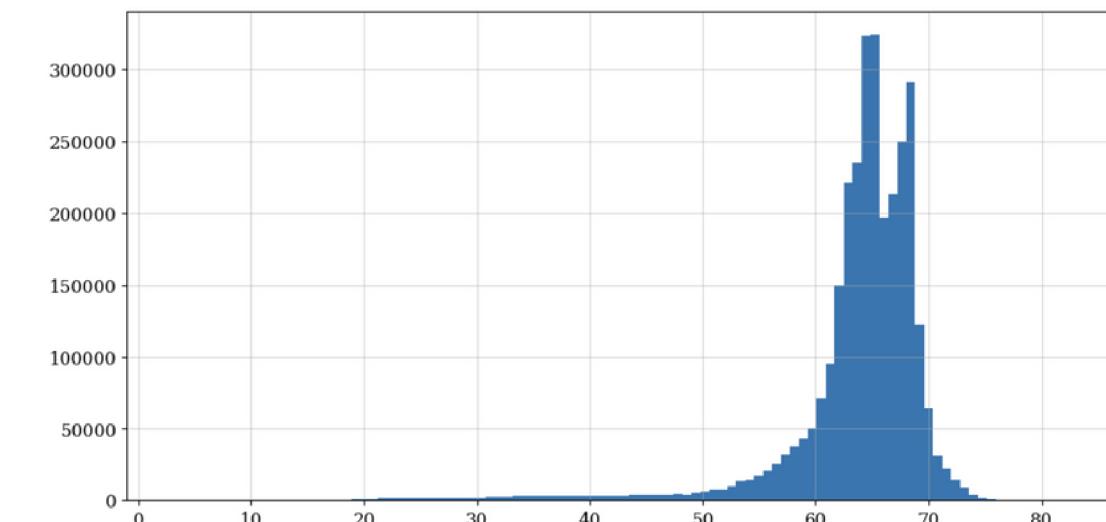
Slightly Skewed Normal Distribution



Flow



Occupy



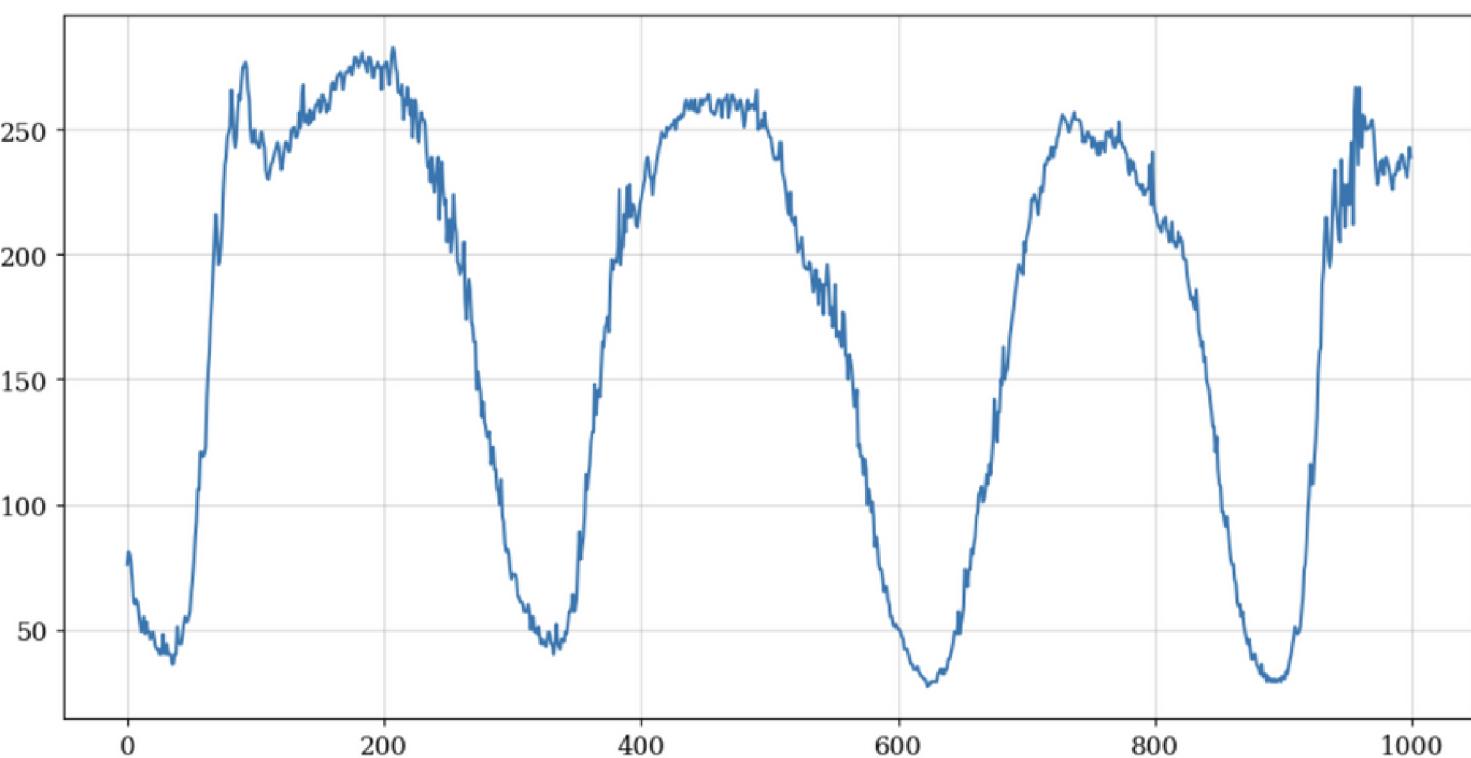
Speed

# Distribution of Features

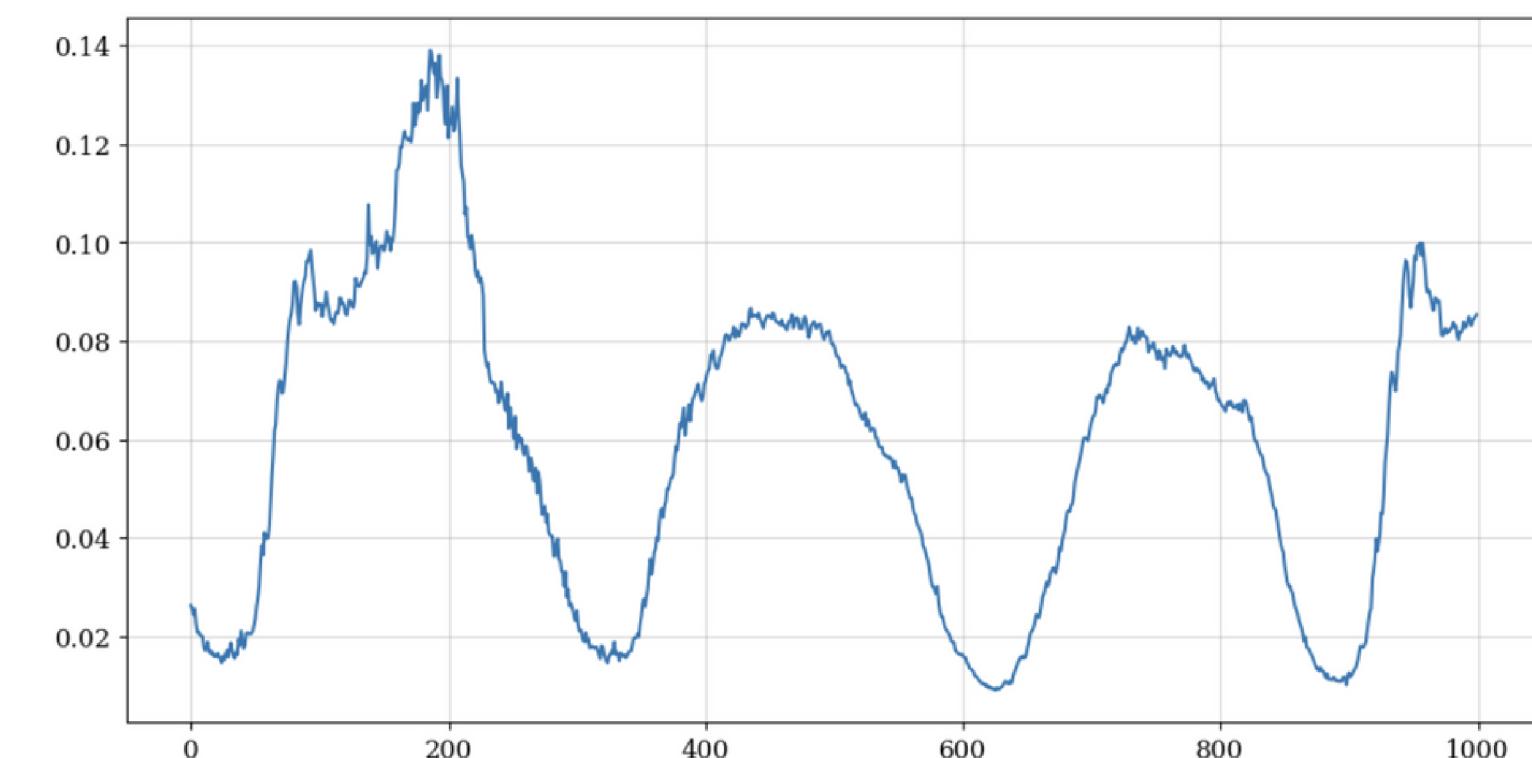


5 Minutes to 60 Minutes Interval

# Explore Seasonality



Flow

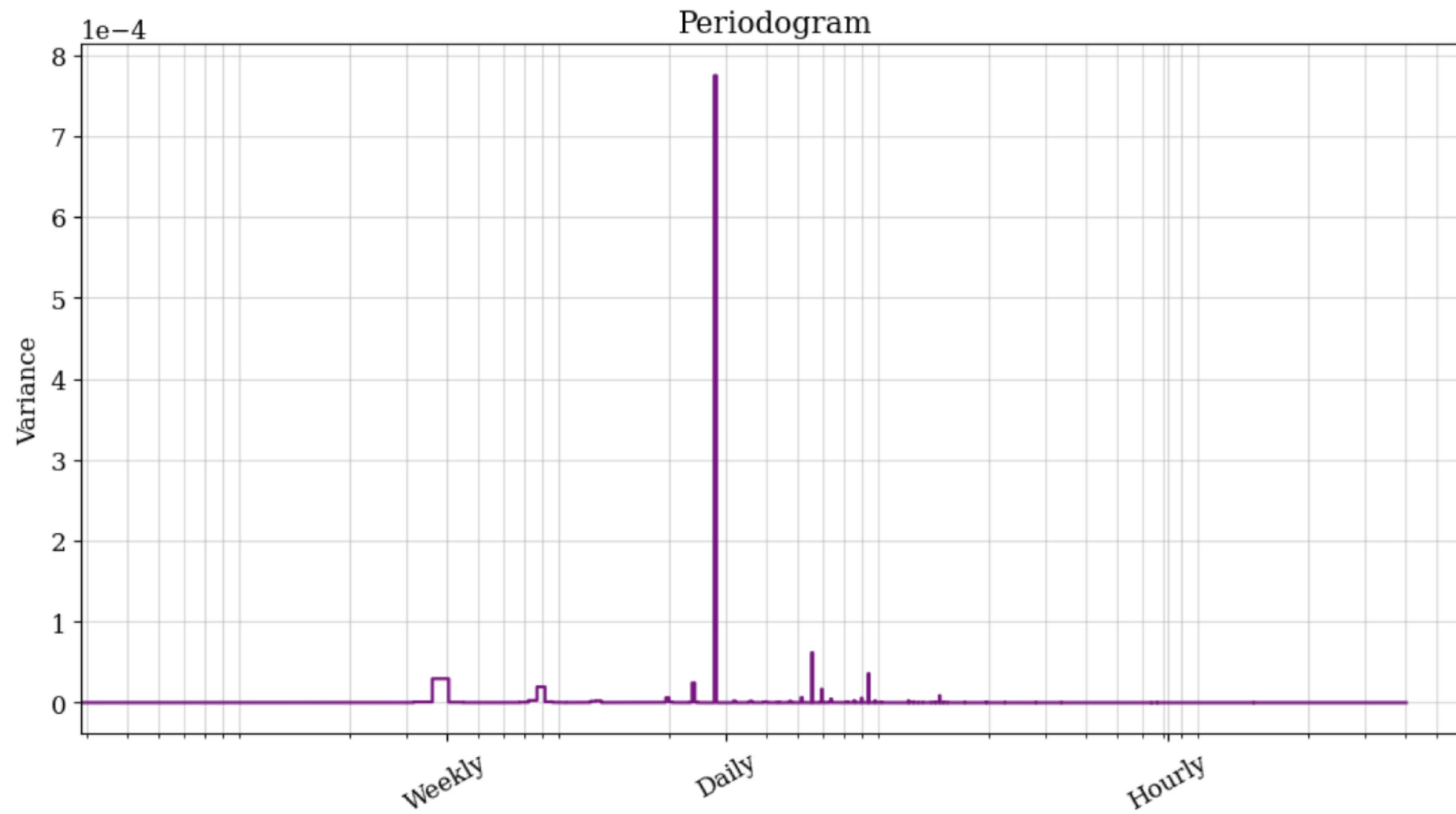


Occupy

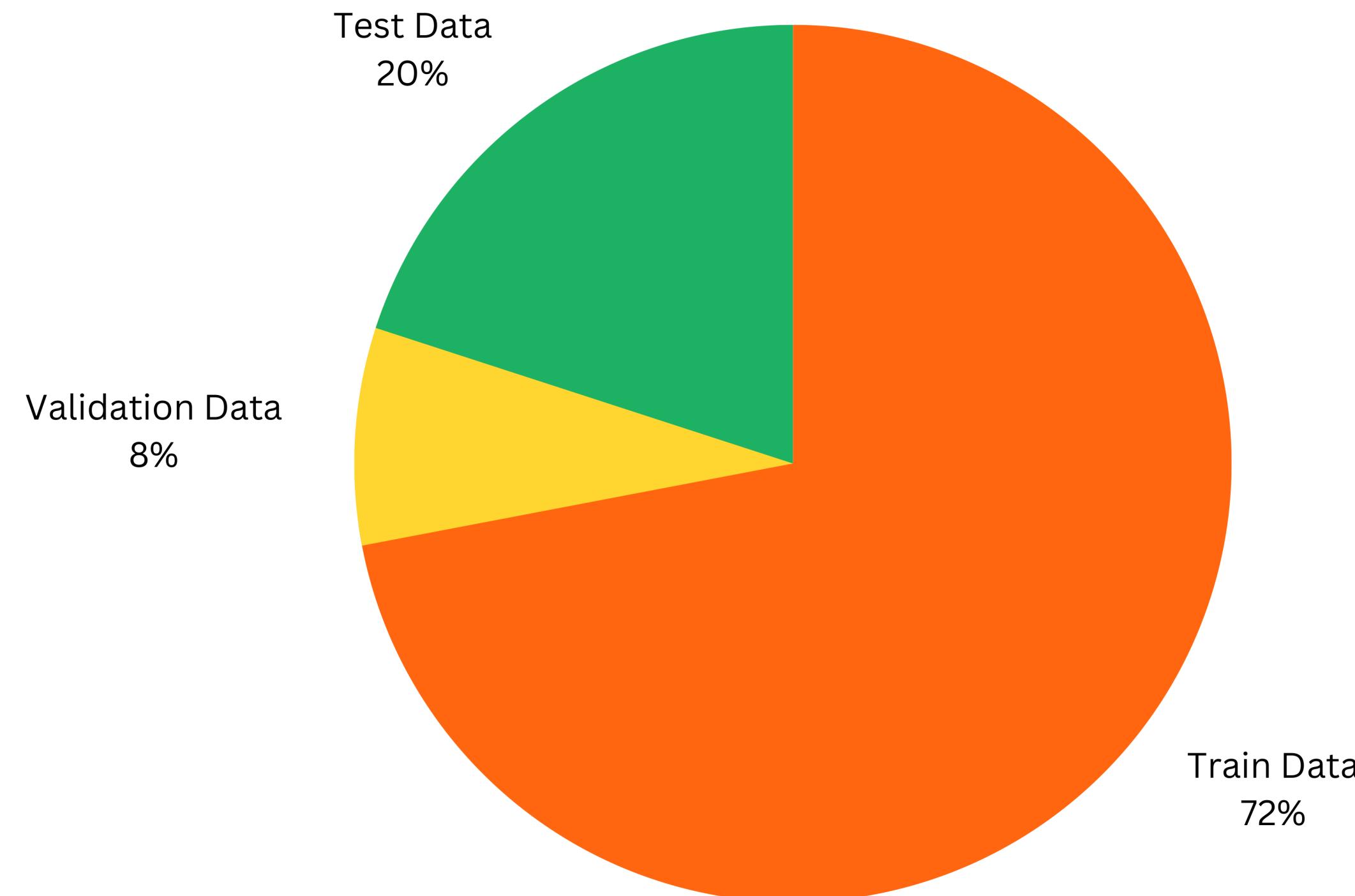
# Explore Correlation

	<b>flow</b>	<b>occupy</b>	<b>speed</b>	<b>flow_future</b>	<b>occupy_future</b>	<b>speed_future</b>
<b>flow</b>	1.000000	0.674039	-0.296332	0.535235	0.450192	-0.235030
<b>occupy</b>	0.674039	1.000000	-0.752040	0.445282	0.477379	-0.303858
<b>speed</b>	-0.296332	-0.752040	1.000000	-0.228266	-0.275180	0.233537
<b>flow_future</b>	0.535235	0.445282	-0.228266	1.000000	0.674040	-0.296331
<b>occupy_future</b>	0.450192	0.477379	-0.275180	0.674040	1.000000	-0.752040
<b>speed_future</b>	-0.235030	-0.303858	0.233537	-0.296331	-0.752040	1.000000

# Occupy Variable's Seasonality



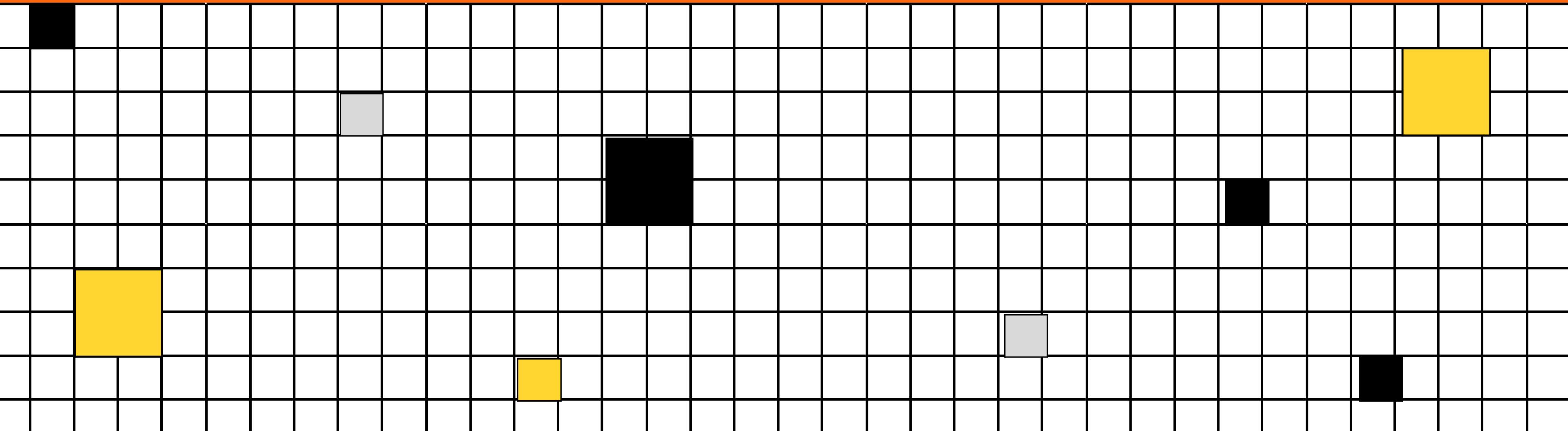
# Split the Data





# Core Analysis I

Training the Model



# Model Used

---

Time-Series  
Problem

# Model Used

---

Long-Short Term  
Memory

# Model Used

---

Mean - Squared  
Error

# Model Used

---

Root Mean Square  
Error

# Model Used

---

Rectified Linear  
Unit

# Model Used

---

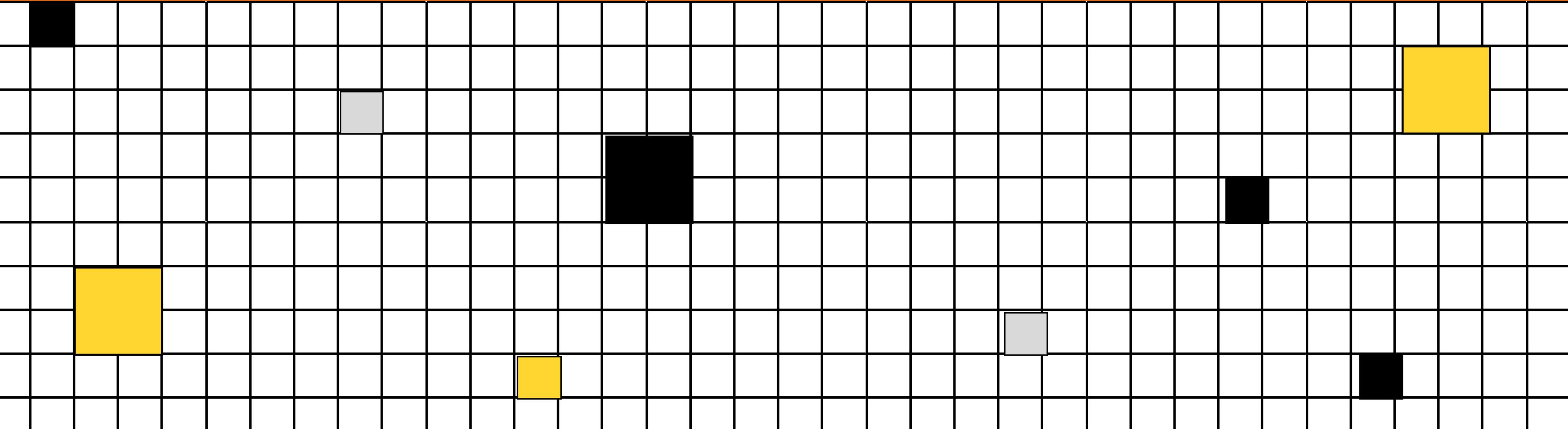


Hyperparameter  
Tuning

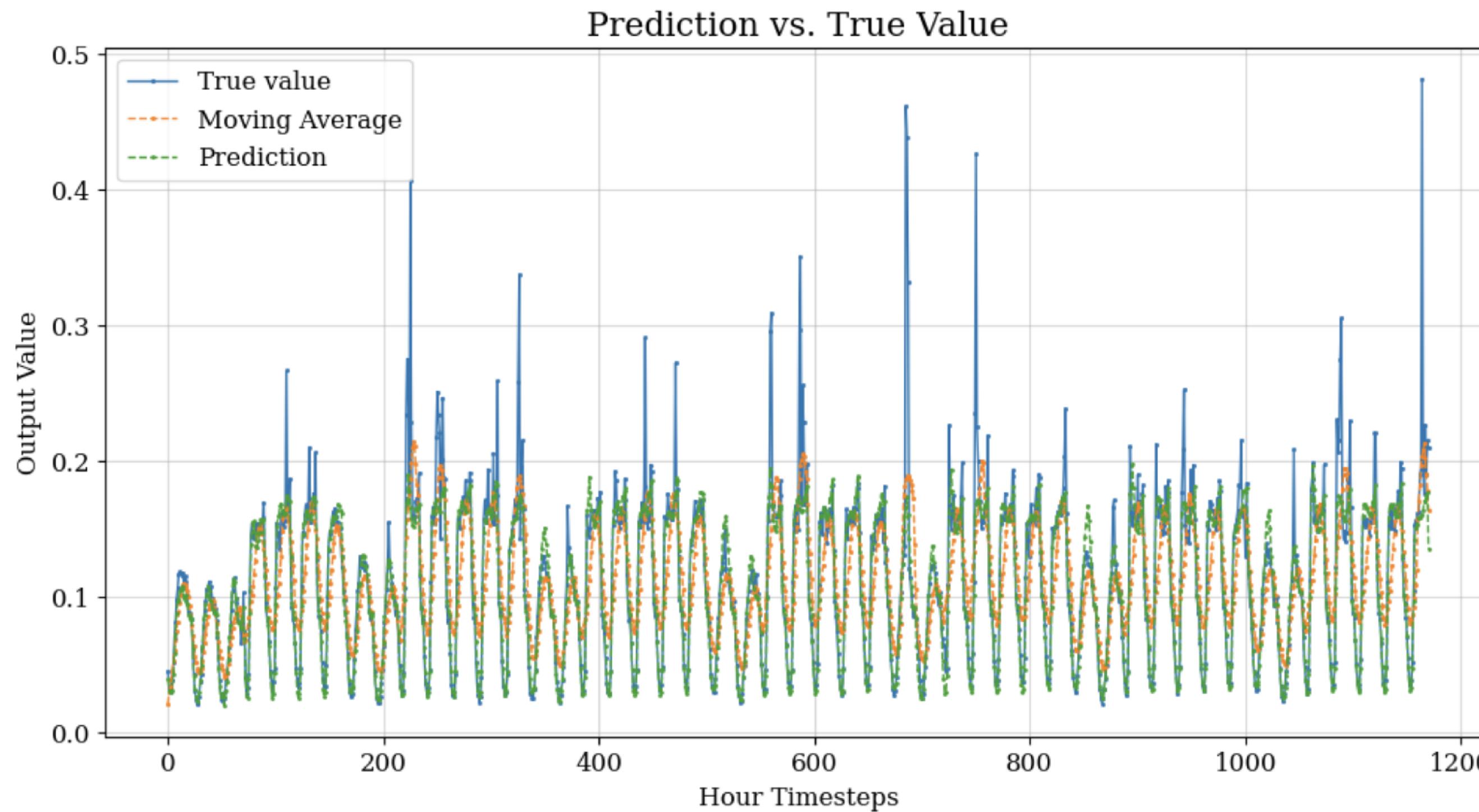


# Core Analysis II

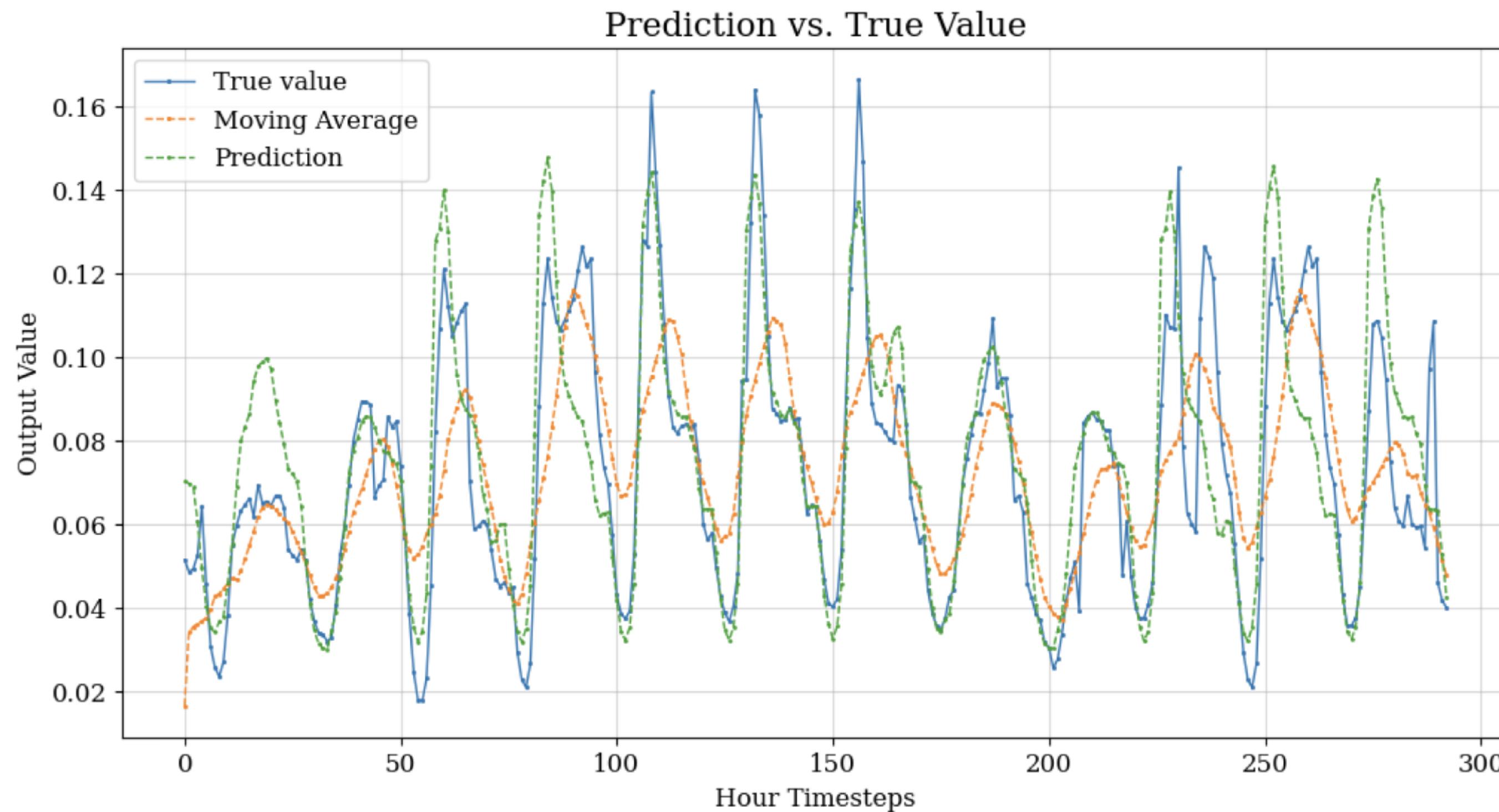
Prediction Evaluation



# Train Dataset Evaluation



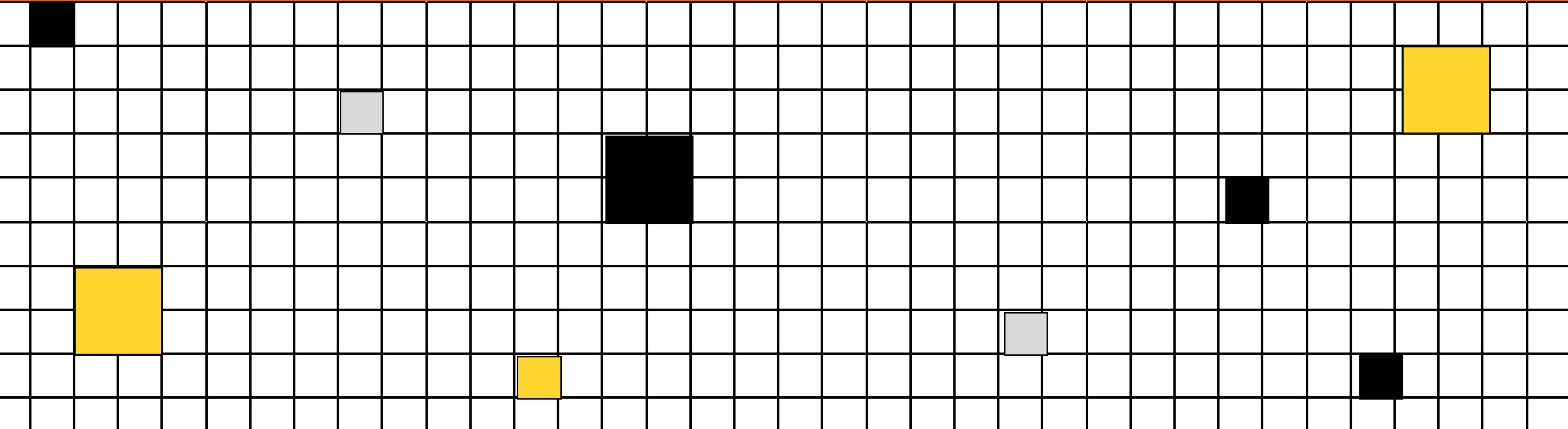
# Test Dataset Evaluation





# Conclusion

Concluding Statement



# Checklist



01

Predictive Model using LSTM



02

Included hour and lag features



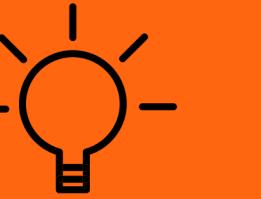
03

Scale & Split the data



04

Able to predict better than baseline



# Excellent Result

Improve road safety and reduce  
traffic congestions

# Thank you for your time!



Email or message for any  
questions or clarifications