

# Modernizing U.S. Waterway Logistics: A Strategic Barge Forecasting & Optimization Project

AI for Supply Chain & Logistics Management

Jul 18, 2025

Richard Jiang: [jiang736@mit.edu](mailto:jiang736@mit.edu)  
HanShen Pan: [hansh423@mit.edu](mailto:hansh423@mit.edu)  
ZhuoYing Lei: [zylei@mit.edu](mailto:zylei@mit.edu)

# Contents

**1. Introduction**

**2. Data Exploration**

**3. Model Analysis**

**4. Future  
Projections**

# Introduction - Mississippi River

This map highlights the Mississippi River and its surrounding states in the central and southern United States.

The Mississippi River begins at Lake Itasca in Minnesota.

The purpose of our project is to forecast the demand of the number of containers.



**The Mississippi River Basin: Surrounding States**



# Introduction - Louisiana

Louisiana is located at the southernmost end of the lower Mississippi River, serving as the final gateway where the river flows into the Gulf of Mexico. All agricultural and industrial products from the Midwest converge here for export.

The Port of New Orleans and the Port of South Louisiana rank as the 6th and 1st largest ports in the U.S. by cargo volume.



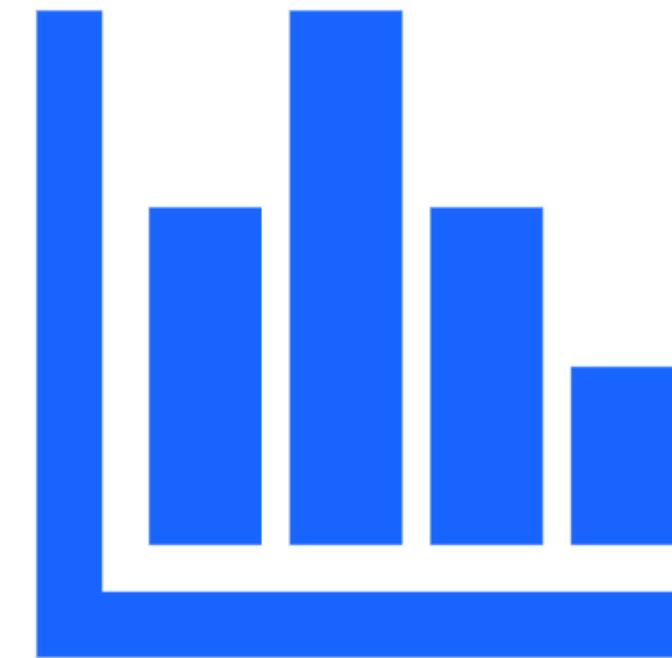
MIT Center for  
Transportation &  
Logistics

## 1. Data Preprocessing

- Data Aggregation, Imputation, Normalisation
- Data Pipelines
- Feature Engineering

## 2. Data Visualization

- Geographic Visualization
- Comparative Analysis
- Correlation Analysis



# Data Exploration - Feature Description

**TABLE 1 Feature Description**

Feature Category	Description
Time Series Features	Trend, Seasonality, Lags
Sea Distance (km)	Distance from the last foreign port to the first US port calculated by searoute API
Shipment Weight (kg)	Product declared weight
Weather	Maximum Temperature, Minimum Temperature, Palmer Drought Severity Index, Palmer_Z index
Product Type	Product type determines which container to use

PCPN: Precipitation

PDSI: Drought and Flood Index

ZNDX: Drought Index

TMPC: Average Temperature

PHDI: Hydrological Drought Index



# Data Preprocessing – Product Type Encoding

The screenshot shows the Hugging Face Model Hub page for the 'all-MiniLM-L6-v2' model. At the top, there's a navigation bar with links for 'Sentence Similarity', 'sentence-transformers', 'PyTorch', 'Tensor', 'text-embeddings-inference', 'arxiv:5 papers', and 'License: apache-2.0'. Below the navigation bar, there's a 'Model card' section with a link to 'Files and versions', a 'xet' button, and a 'Community' section with 123 members. The main content area displays the model's product descriptions.

	productDescription
0	PALLET 40' HIGH CUBE CONTAINER DONATED RELIEF...
1	PALLETS DONATED RELIEF CARGO: MEDICAL SUPPLIES
2	BIN 3901400000 (HS) EPM OFF GRADE HTS: 390140

all-MiniLM-L6-v2

Return the hscode (encoding)

The first 99 Classes (01-99)

	hscode	description
0	01	Animals; live
1	0101	Horses, asses, mules and hinnies; live
2	010121	Horses; live, pure-bred breeding animals



# Data Preprocessing – Sea Distance Calculation With SeaRoute

## Sample Data

Start Port	End Port
Antwerp	Wilmington, North Carolina
Pusan	Wilmington, North Carolina
Shanghai	Wilmington, North Carolina



01

Get Longitude + Latitude for Start and Target Ports

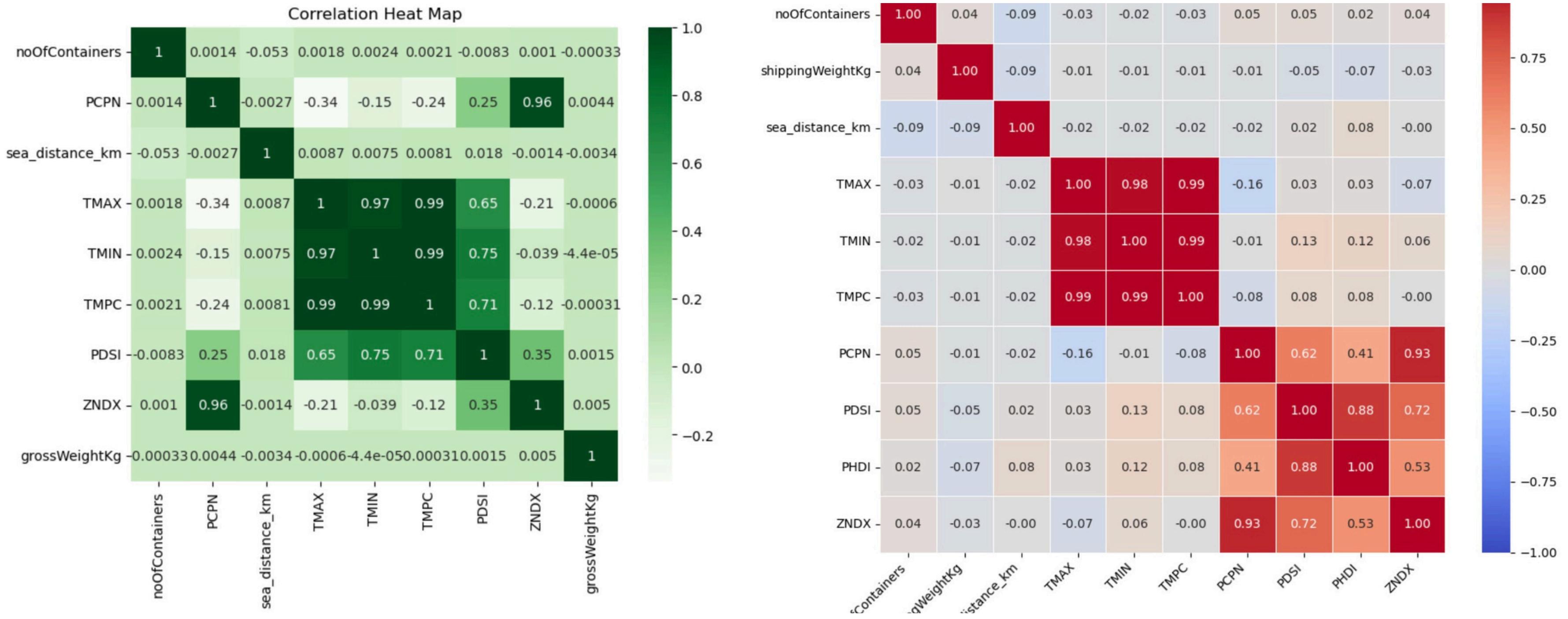
02

Calculate the actual sea travel distance with this information

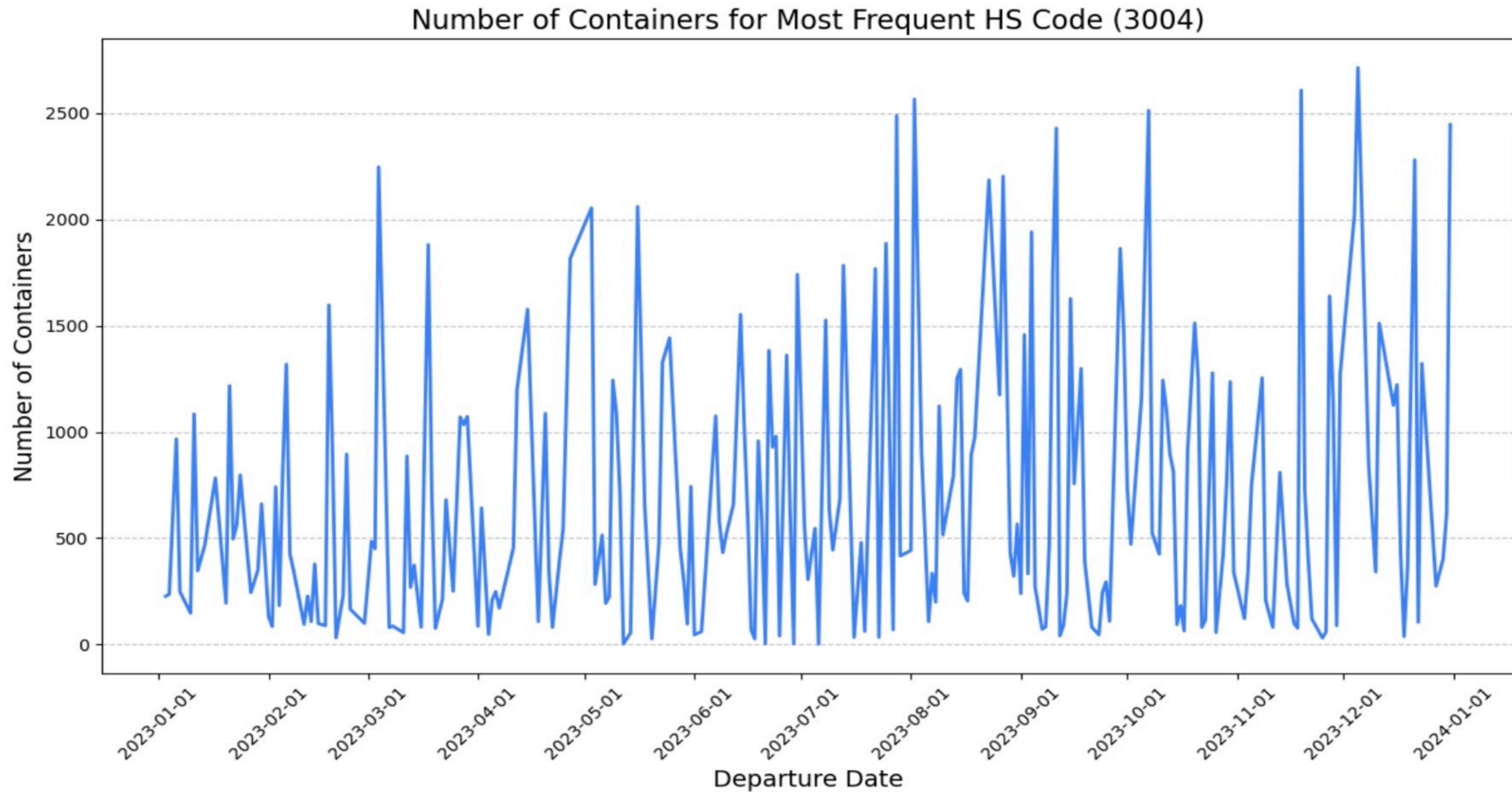
03

Drop the two port columns in the end

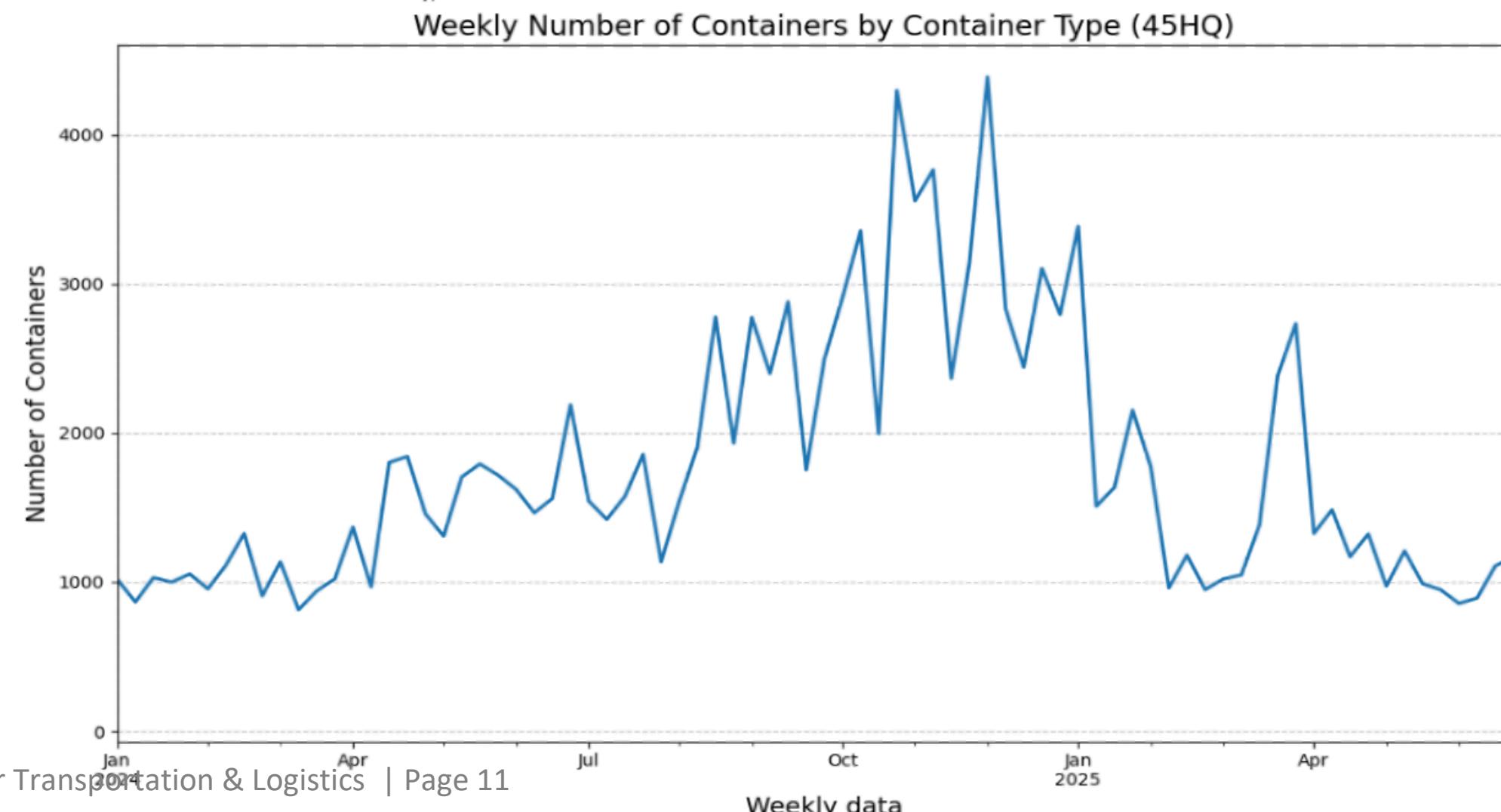
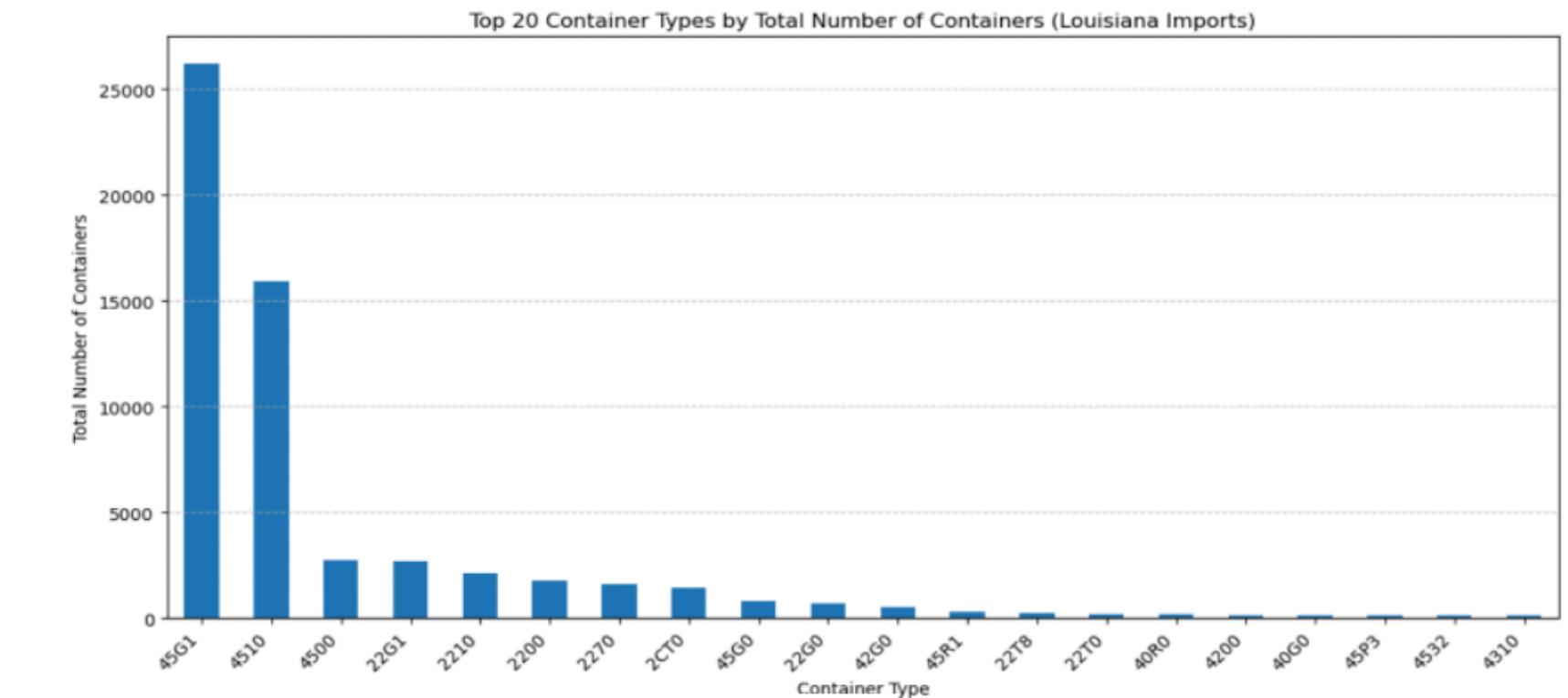
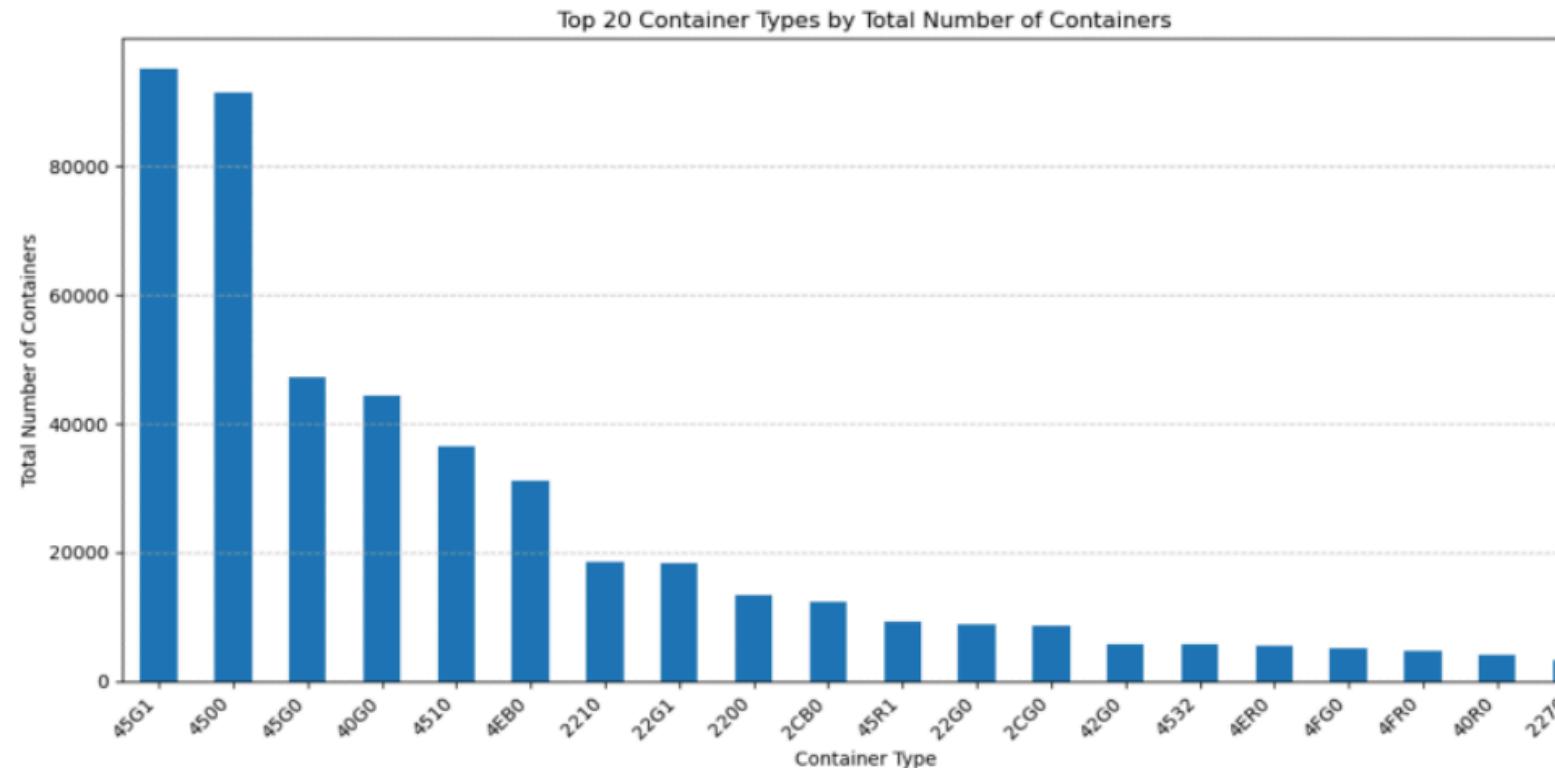
# Data Visualisation – Correlation Heat Map



# Data Visualisation – Medicaments and # of Containers



# Data Visualisation - Container Types



# Model Analysis – Models We Used

## ML Models

### Hybrid Model (2):

- Linear Regression
- XGBoost / RandomForest

## DL Model

### LSTM

- Great with temporal data understanding

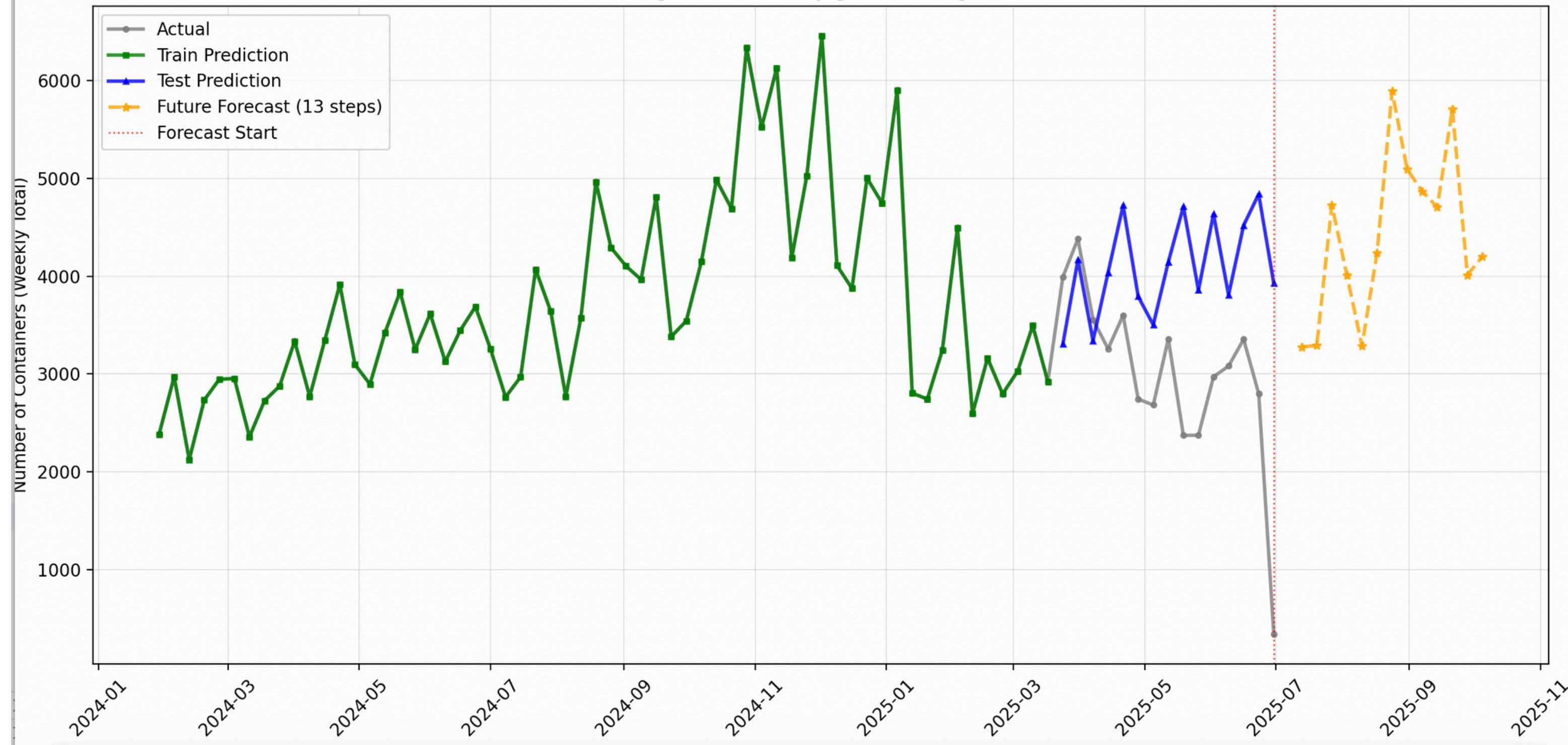
# Model Analysis – Result for Linear Regression + XGBoost

## Linear Regression + XGBoost

Container Import Predictions vs Actual with 13-Step Forecast  
40-foot high cube 9'6" dry general usage container

TABLE 1 Metrics of Linear+XGBoost Model

Split	MAE	MSE	MAPE
Training	65.89	4341.49	62.70
Testing	188.07	35370.32	173.99



# Model Analysis – Result for Linear Regression + XGBoost

## Linear Regression + RandomForest

Container Import Predictions vs Actual with 13-Step Forecast  
40-foot high cube 9'6" dry general usage container

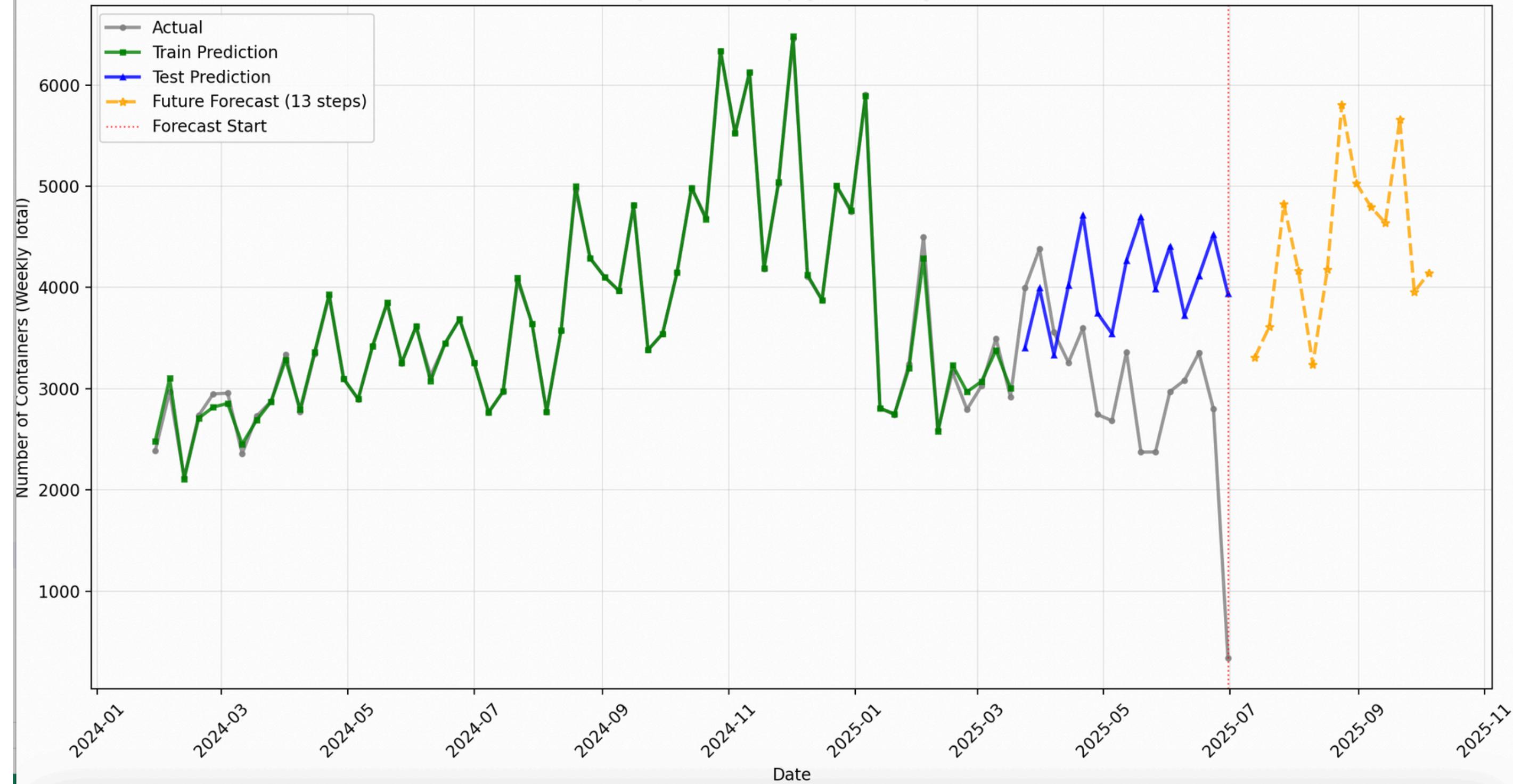
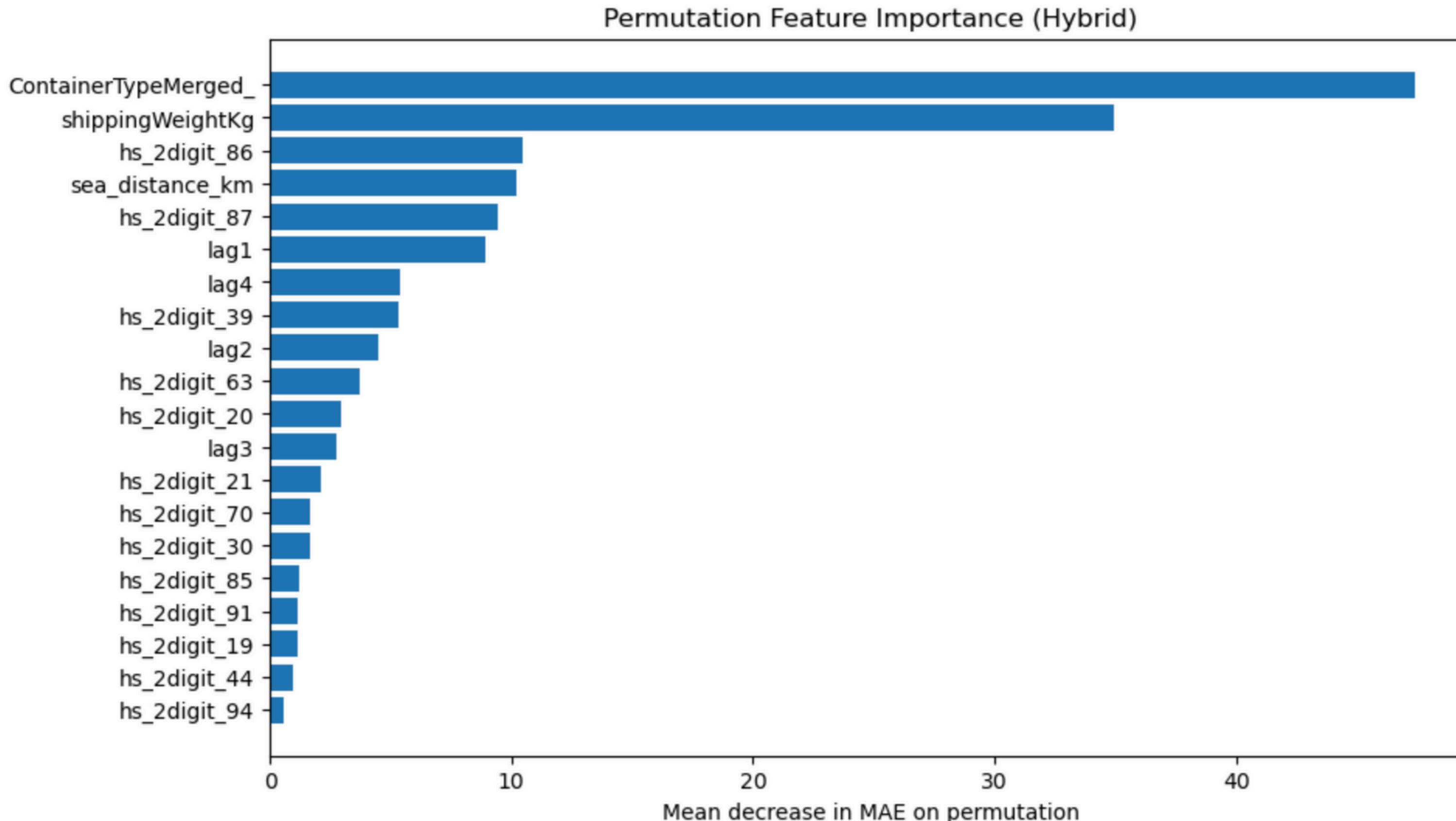


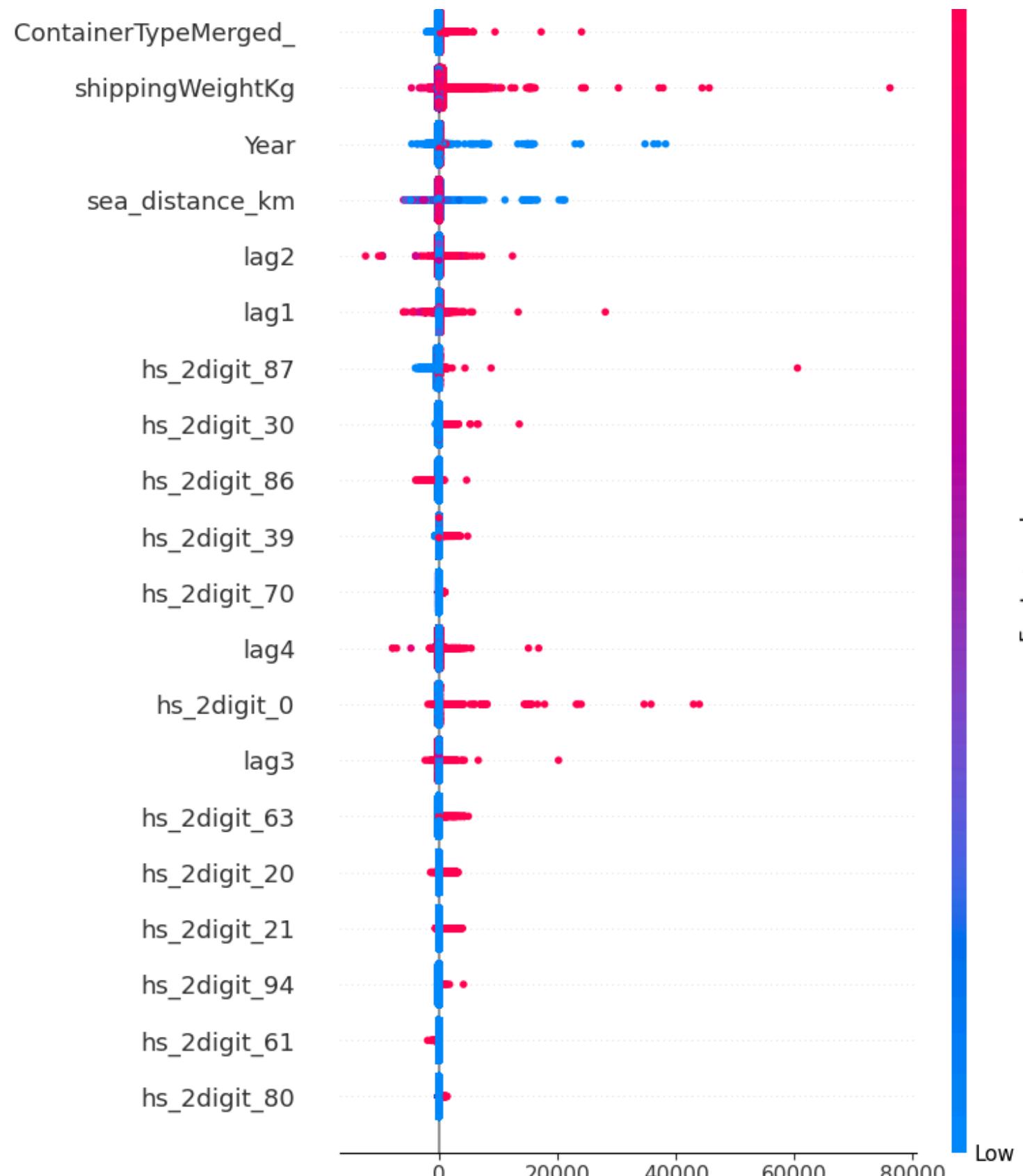
TABLE 1 Metrics of Linear+XGBoost Model

Split	MAE	MSE	MAPE
Training	65.89	4341.49	62.70
Testing	188.07	35370.32	173.99

# Model Analysis – Permutation (Hybrid Models)



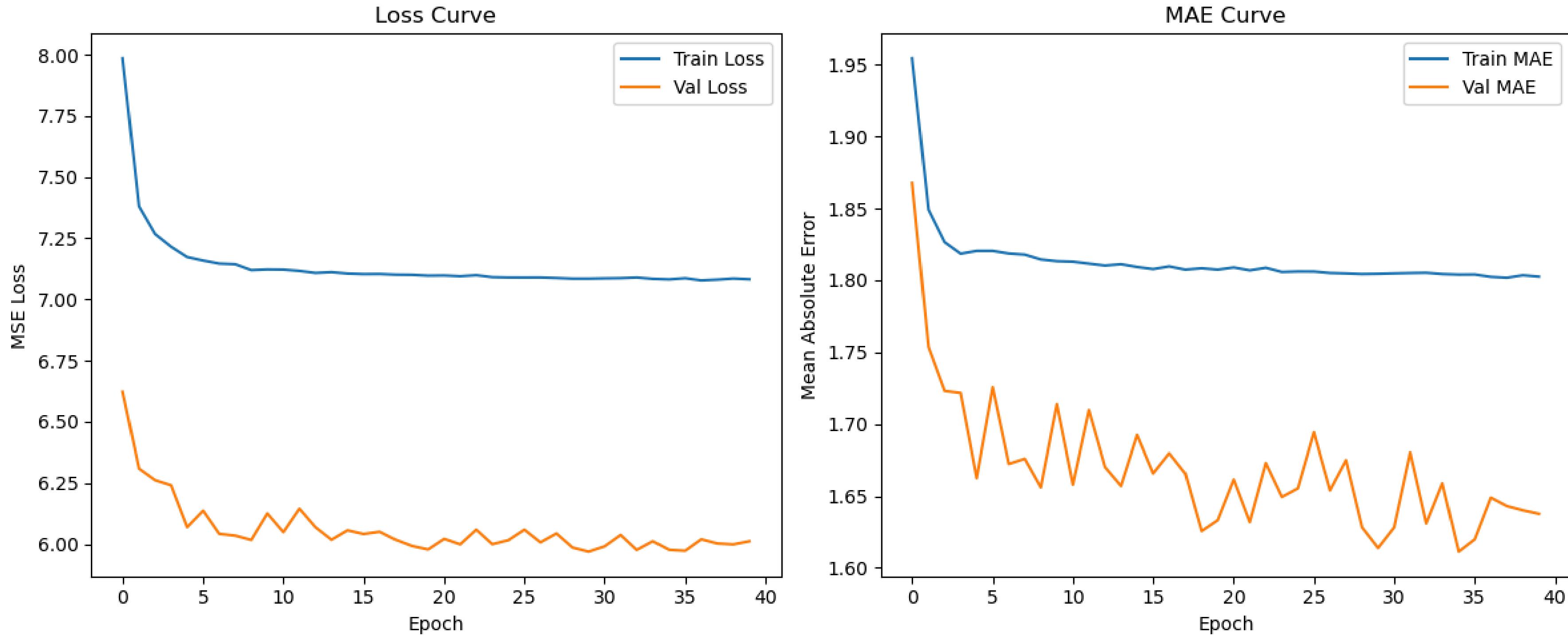
# Model Analysis - SHAP Value (Hybrid Models)



Top Determining Features:

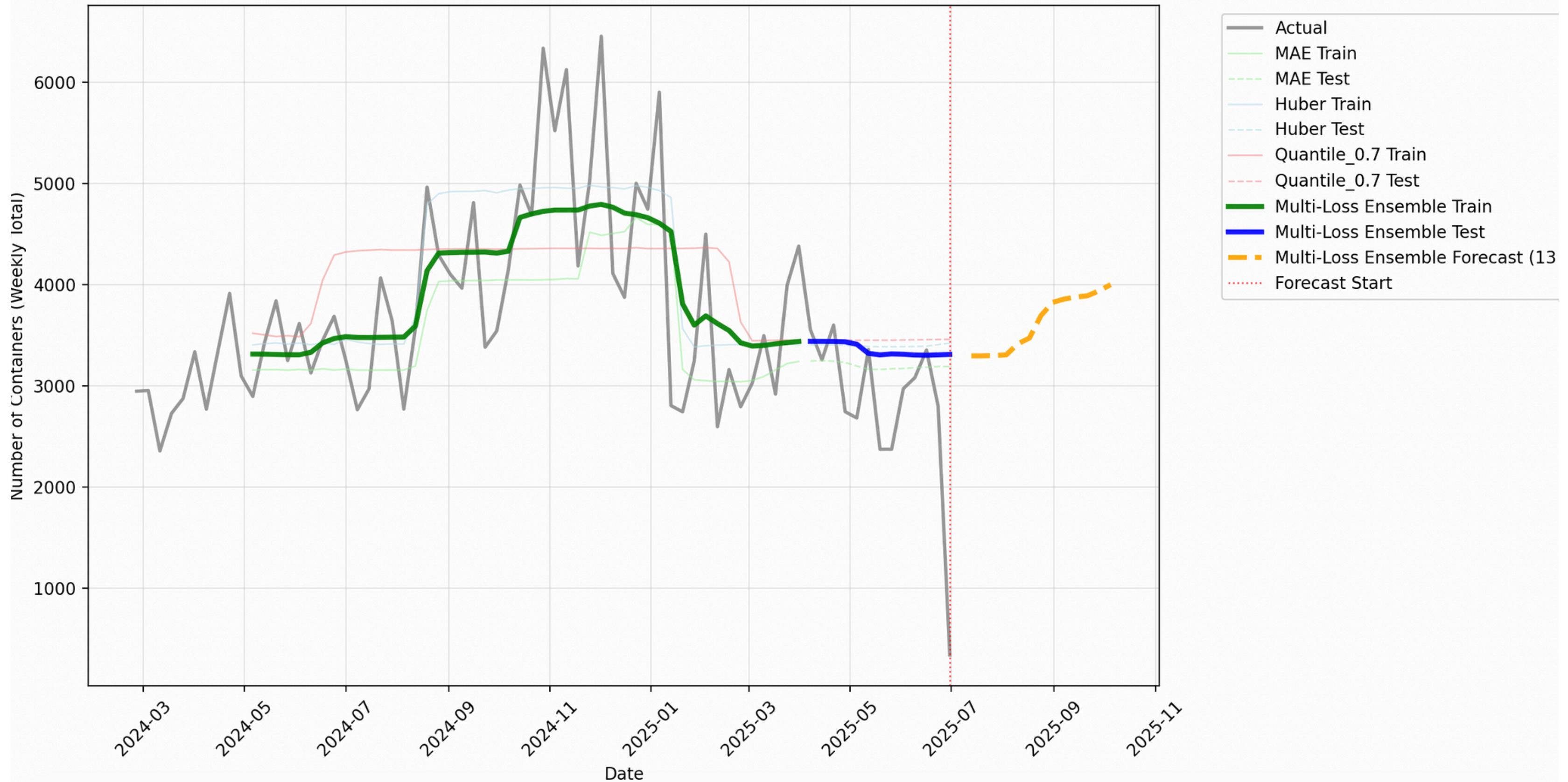
1. Container Type
2. Shipping Weight
3. Time-Step and Time-Series
4. Sea Distance Km
5. Product Types

# Model Analysis – LSTM Loss Curves



# Model Analysis – LSTM Forecast

Advanced Multi-Loss LSTM Container Import Predictions vs Actual  
40-foot high cube 9'6" dry general usage container



# Future Projections – Spatial-Temporal Analysis

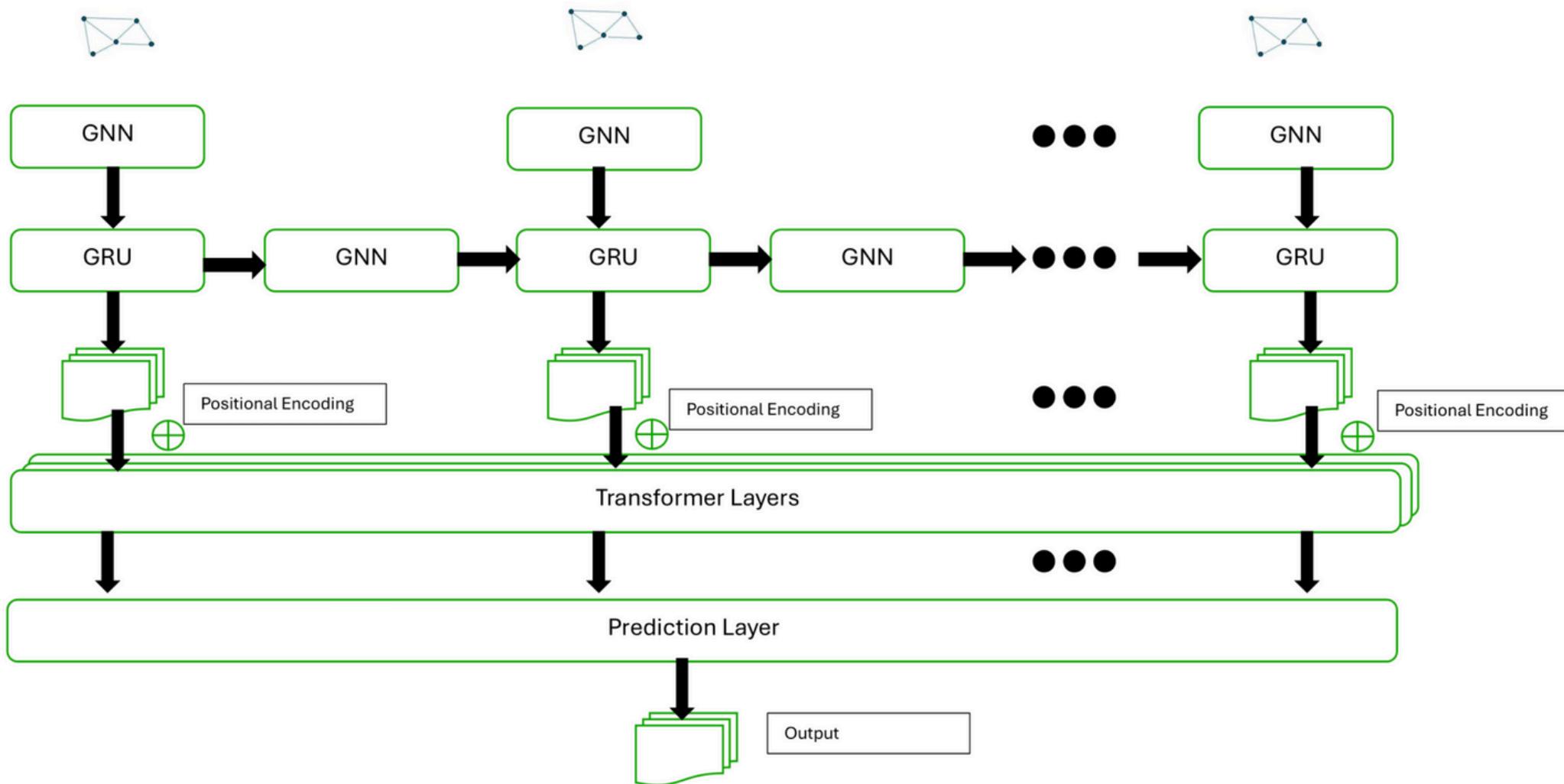


Fig. 1. ST-GNN Model representation, redrawn from [4]

## Spatial-Temporal Analysis

### 1. Spatial (GNN)

- Construct a graph, with nodes being the features for each city
- The edge being the relative distance between each city

### 2. Temporal (1-3B LLM)

- Autoregressive attribute
- Process temporal (each graph updates) part of the data using previous data



# Thank You!

---

Richard Jiang: [jiang736@mit.edu](mailto:jiang736@mit.edu)  
HanShen Pan: [hansh423@mit.edu](mailto:hansh423@mit.edu)  
ZhuoYing Lei: [zylei@mit.edu](mailto:zylei@mit.edu)