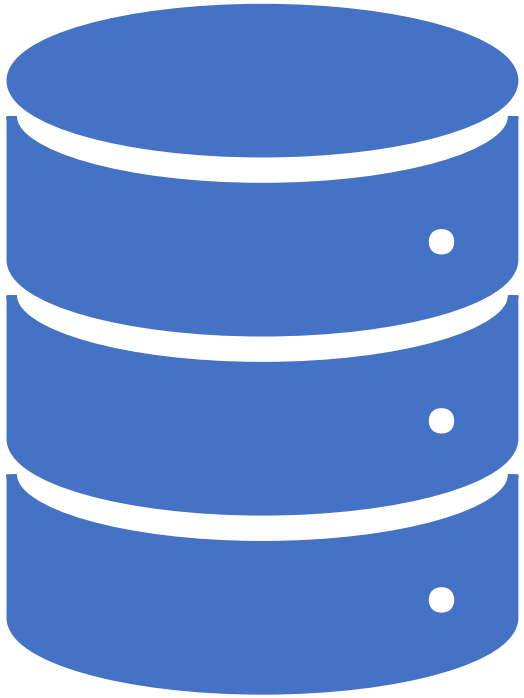


"Autumn 2020 Order Datasets Exploration and Predictive Modeling for Estimated Delivery Time"

By Rabiul Islam

"Enhancing Delivery Service Efficiency"

Introduction

- **In this presentation:**
  - Datasets exploration and Analysis
  - Time series Analysis
  - Linear Regression and Random Forest Model
  - Key Findings, Comparison and Future steps.
- **Goals:**
  - Improve delivery time estimation.
  - Enhance customer satisfaction and operational efficiency.

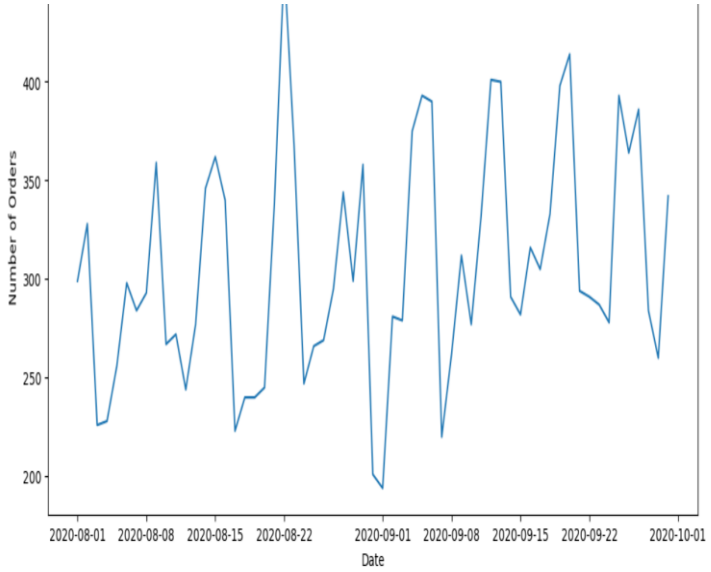
Summary statistics of the DataFrame

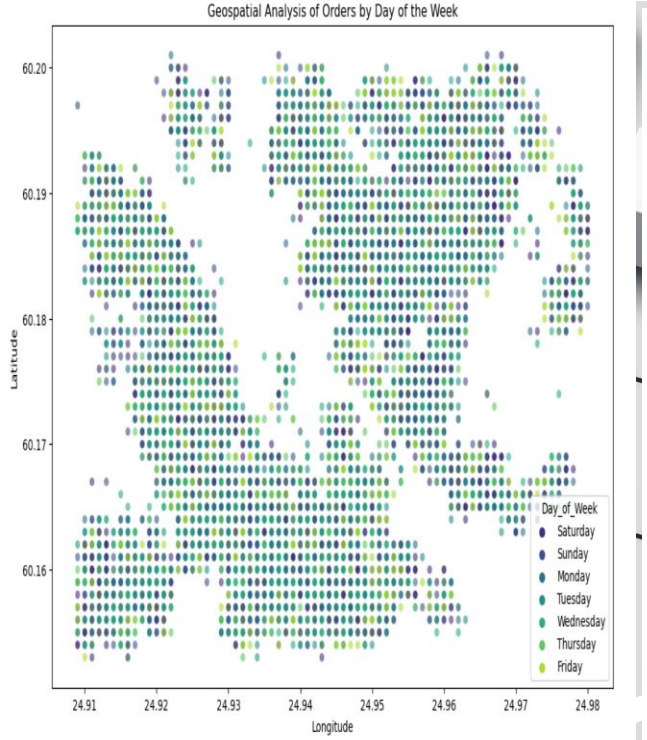
25%			-7.000000	1.000000	
50%			-2.000000	2.000000	
75%			5.000000	4.000000	
max			34.000000	11.000000	

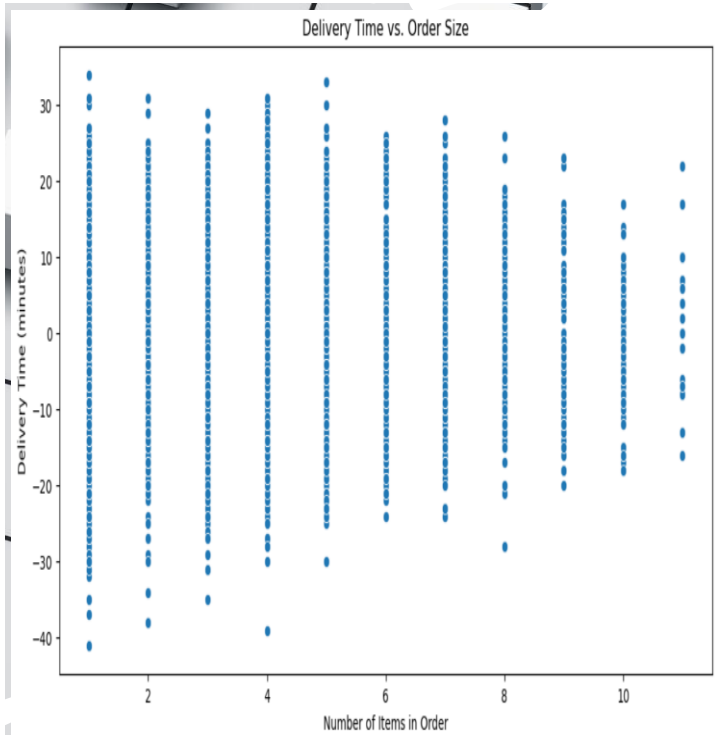
	USER_LAT	USER_LONG	VENUE_LAT	VENUE_LONG	\
count	18706.000000	18706.000000	18706.000000	18706.000000	
mean	60.175234	24.941244	60.175643	24.941214	
std	0.012674	0.016540	0.011509	0.014482	
min	60.153000	24.909000	60.149000	24.878000	
25%	60.163000	24.926000	60.167000	24.930000	
50%	60.175000	24.943000	60.170000	24.941000	
75%	60.186000	24.954000	60.186000	24.950000	
max	60.201000	24.980000	60.219000	25.042000	

	ESTIMATED_DELIVERY_MINUTES	ACTUAL_DELIVERY_MINUTES	CLOUD_COVERAGE	\
count	18706.000000	18706.000000	18429.000000	
mean	33.809313	32.608254	11.996853	
std	7.340283	10.018879	23.812605	
min	10.000000	6.000000	0.000000	
25%	28.000000	25.000000	0.000000	
50%	33.000000	32.000000	0.000000	
75%	38.000000	40.000000	25.000000	
max	82.000000	58.000000	100.000000	

	TEMPERATURE	WIND_SPEED	PRECIPITATION
count	18429.000000	18429.000000	18706.000000
mean	16.973536	3.790991	0.332756
std	3.411900	1.456017	1.129234
min	6.100000	0.077419	0.000000
25%	14.400000	2.696190	0.000000
50%	16.700000	3.631970	0.000000
75%	18.900000	4.692530	0.000000
max	26.700000	9.857300	6.315790

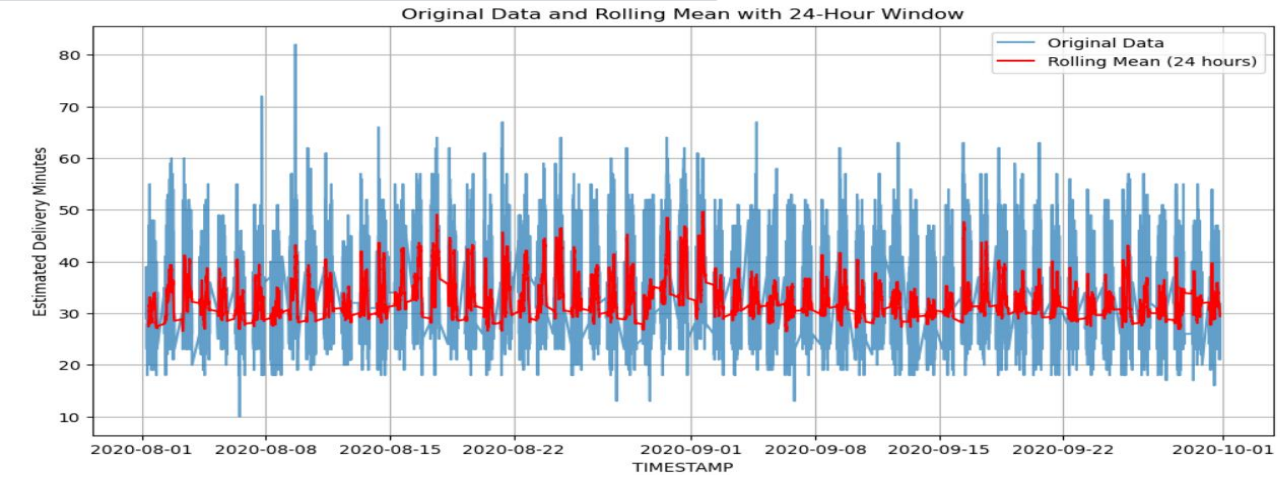


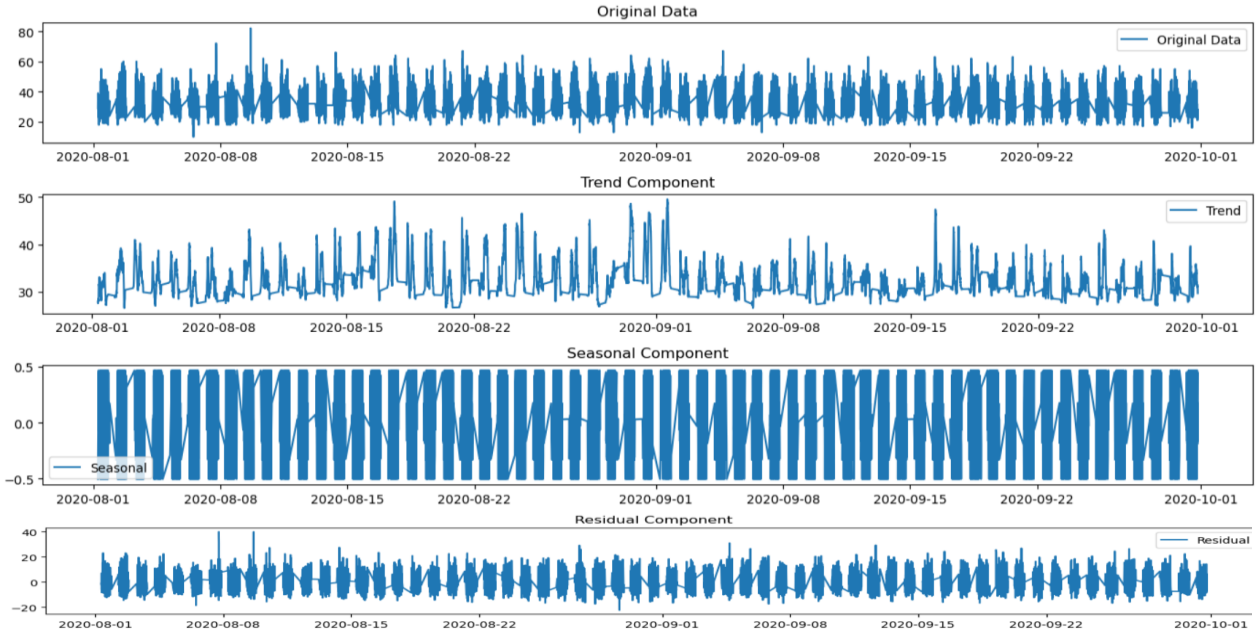


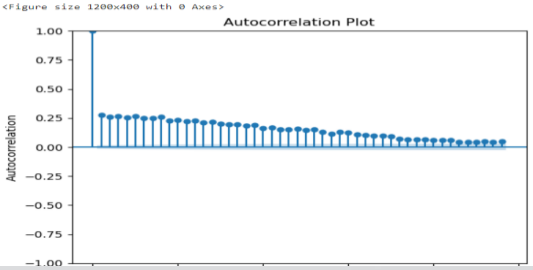


Datasets exploration and Analysis

- Basic information and Summary statistics of the Data frame.
- Identify Nan Values and visualization.
- Identify Missing, Duplicate values and visualization.
- Order Frequency over Time
- Distribution of delivery time
- Delivery Time vs Order Size
- Geospatial Analysis of orders by Day of the week



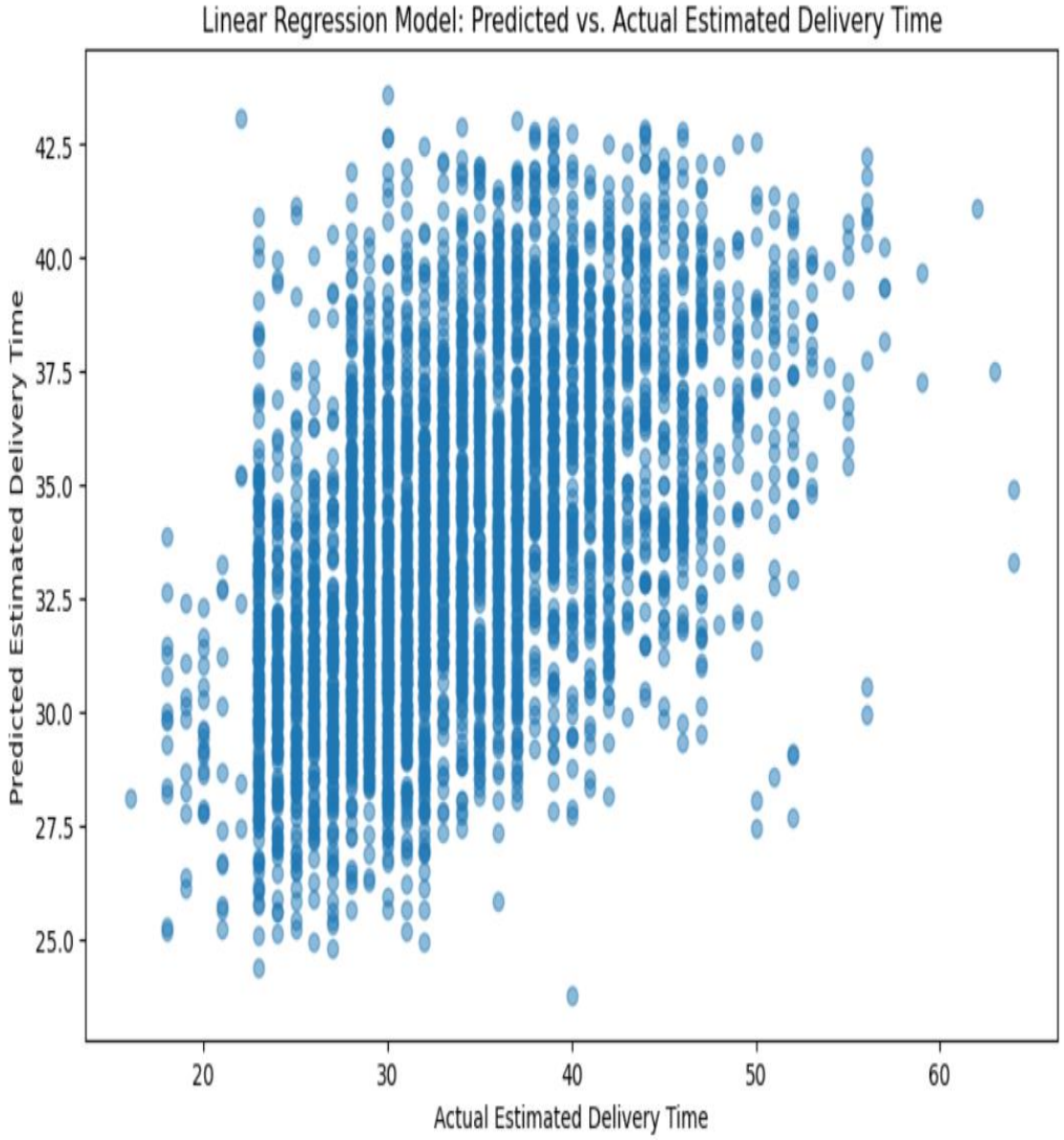




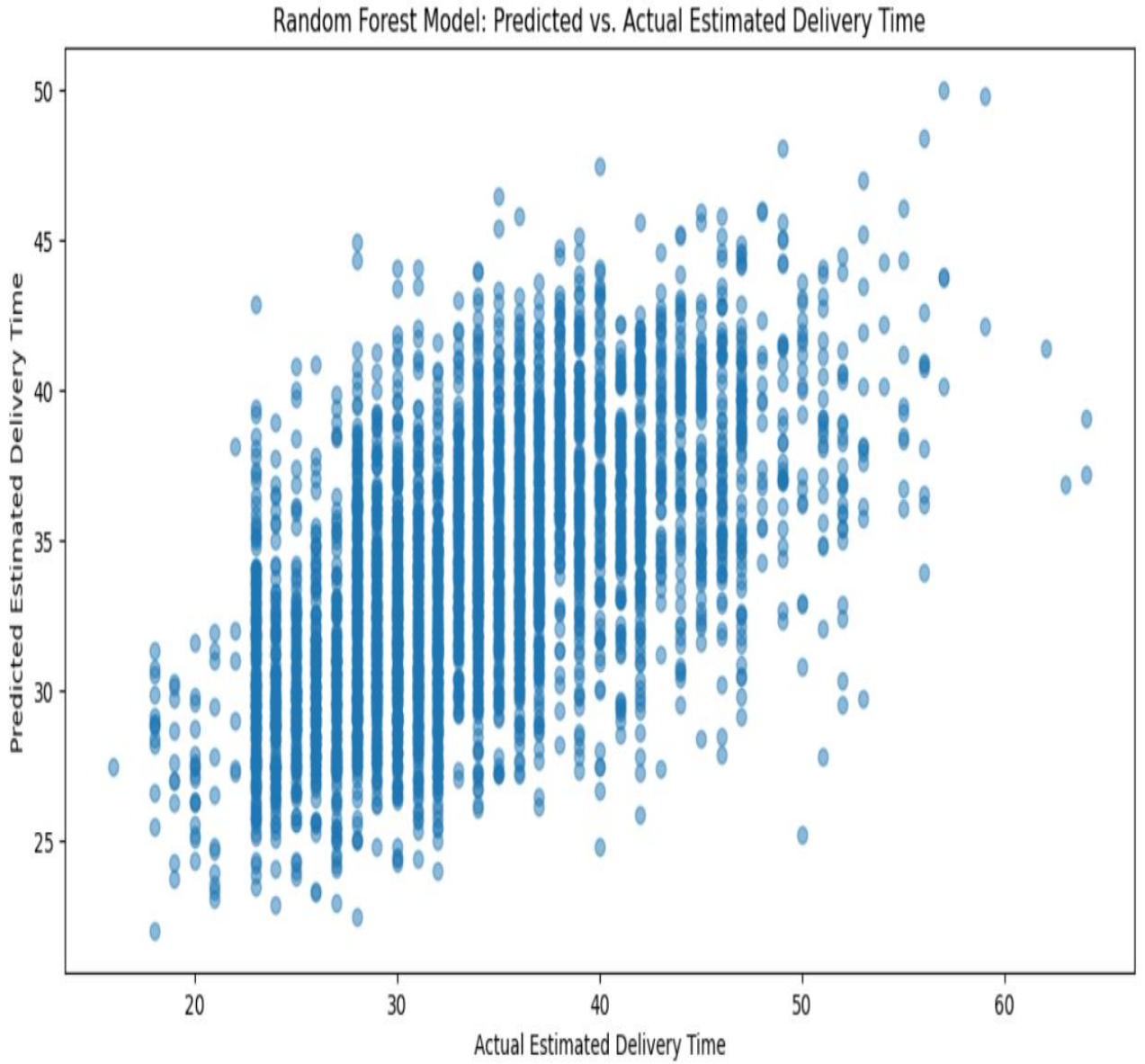
Time series Analysis

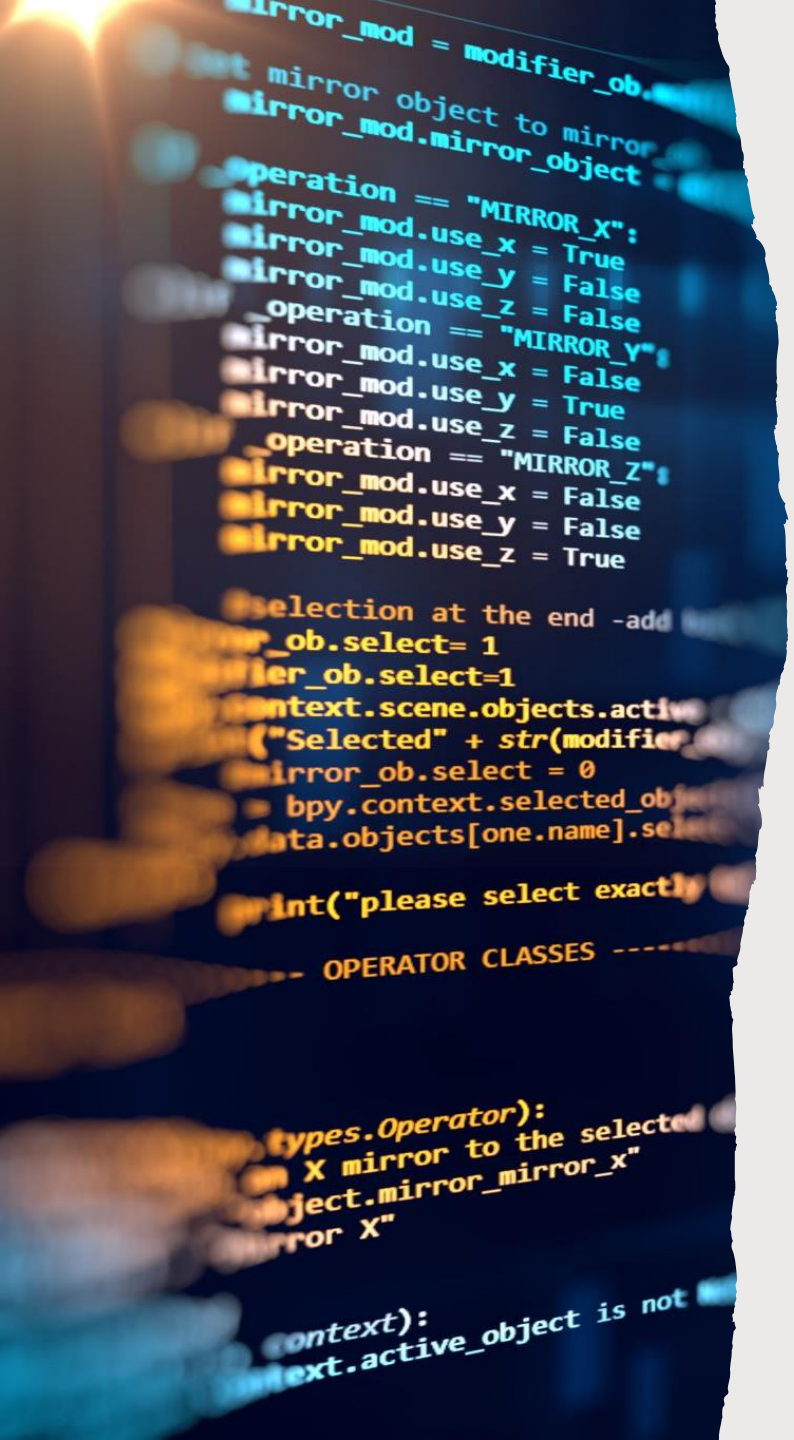
- Original Estimated Delivery Time data and Rolling Mean
- Seasonality

Mean Absolute Error: 5.019069619106365
Mean Squared Error: 40.8672142406891



Random Forest - Mean Absolute Error: 4.632170374389582
Random Forest - Mean Squared Error: 35.54697352143244





Key Findings, Comparison and potential steps

- Understanding feature importance can guide operational decisions and highlight areas for improvement.
- Random Forest outperformed Linear Regression in accurately predicting estimated delivery times.
- Assess the inclusion of external factors (e.g., traffic, special events) to further improve model accuracy.

Thank you for time!