ENERGY CONSUMPTION FORECASTING

(Energy Consumption Dataset)
- Kaung Myat San

PROBLEM STATEMENT

Predicting energy consumption 60 months in the future as accurately as possible

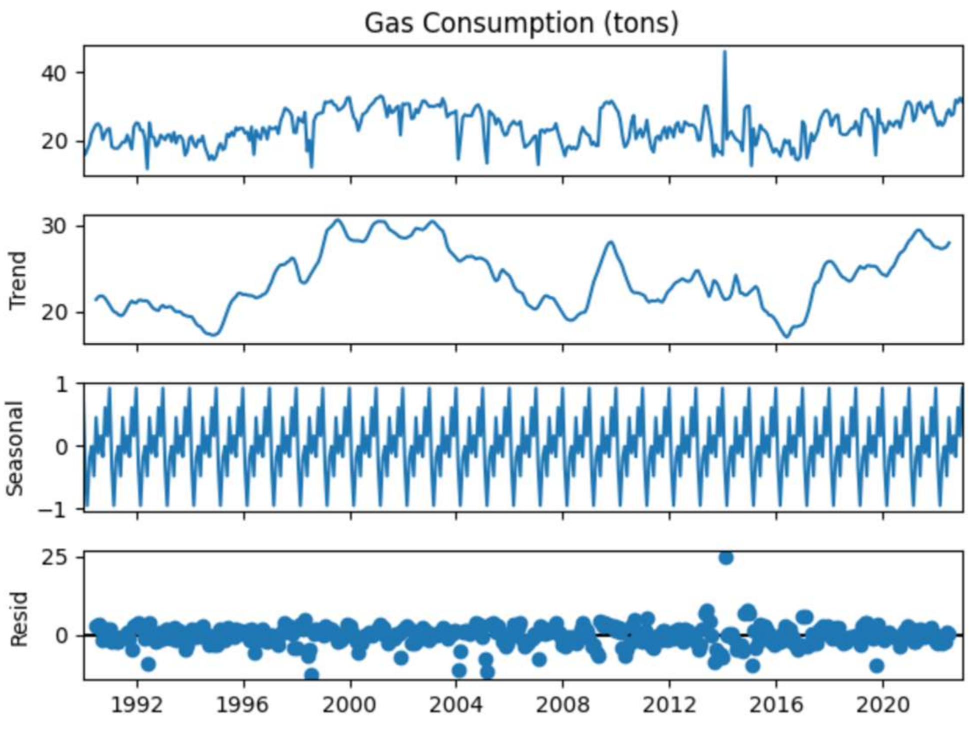
STATIONARY ANALYSIS

	P-VALUE	ADF STATS	CRITICAL 1%	CRITICAL 5%	CRITICAL 10%
GAS	0.0108	-3.40	-3.45	-2.869	-2.57
ELECTRICITY	0.186	-2.26	-3.45	-2.869	-2.57
WATER	8.98e-05	-4.68	-3.45	-2.869	-2.57

Stationary Results					
GAS	YES	ELECTRICITY	NO	WATER	YES
P-Value (0.01 < 0.05) ADF STATS (-3.4 < -2.869- 5%)		P-Value (0.186 > 0.05) ADF STATS (-2.26 > -2.57- 10%)		P-Value (8.98e-05 < 0.05) ADF STATS (-4.68 < -3.45 -1%)	

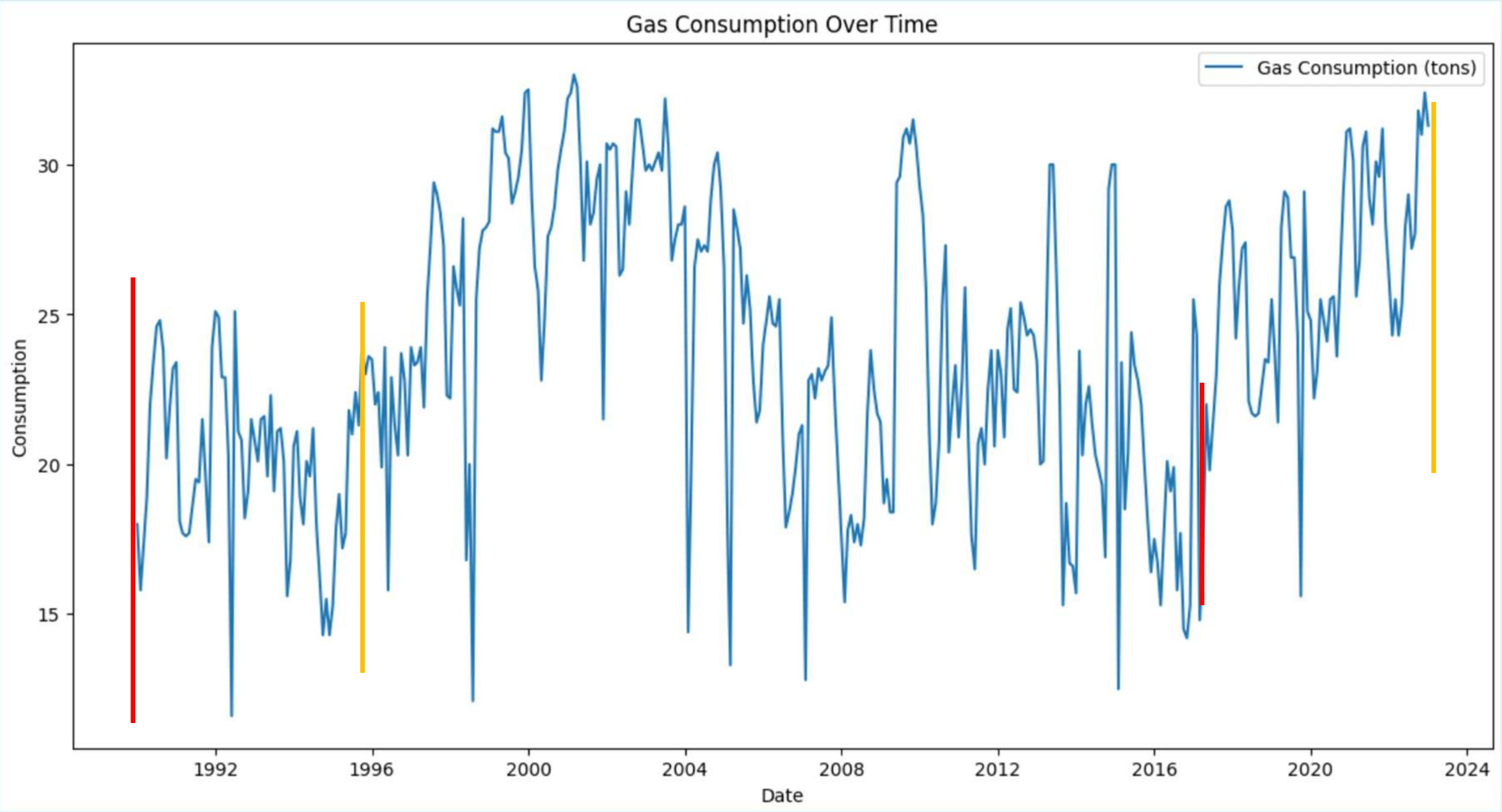
GAS CONSUMPTION ANALYSIS

GAS SEASONAL DECOMPOSITION

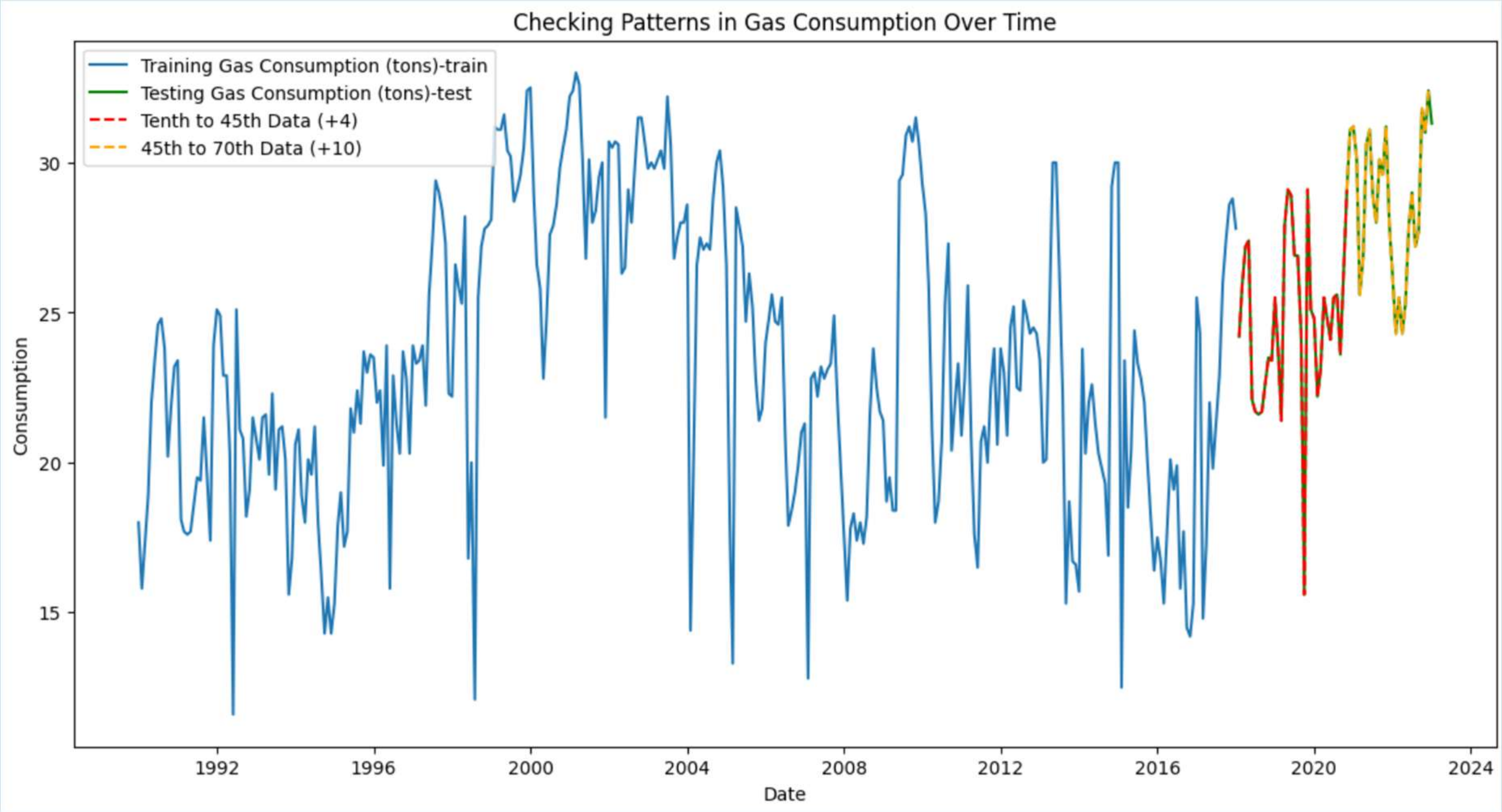


- Overall Gas Consumption has increase since 1990
- Trend-
  - 1990 to 1996: stable
  - 1996 to 2004: growth
  - 2004 to 2016: decline with some spike
  - 2016 to 2023: growth
- Consistent Seasonal Pattern (every year)
- Outlier in 2014 (Replaced with Mean)

GAS CONSUMPTION TIME SERIES



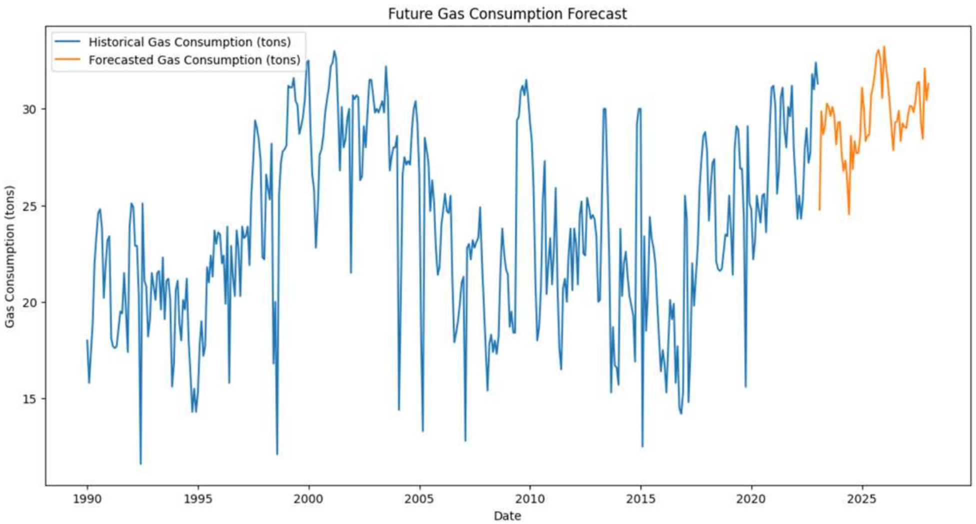
CONFIRMING ANALYSIS



GAS CONSUMPTION FORECAST

<u>Holt-Winter's Seasonal Smoothing model</u>

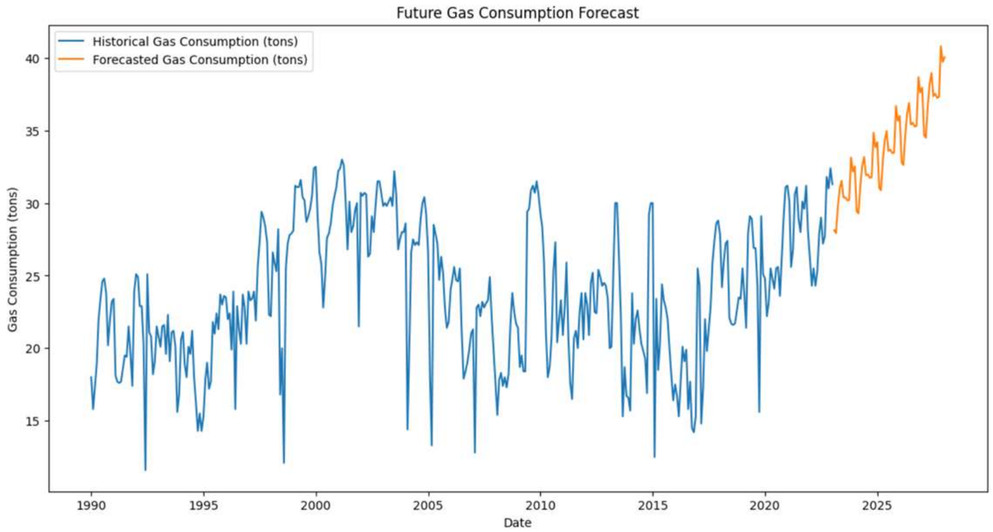
Trend = Add , Seasonal = Add, Period = 96



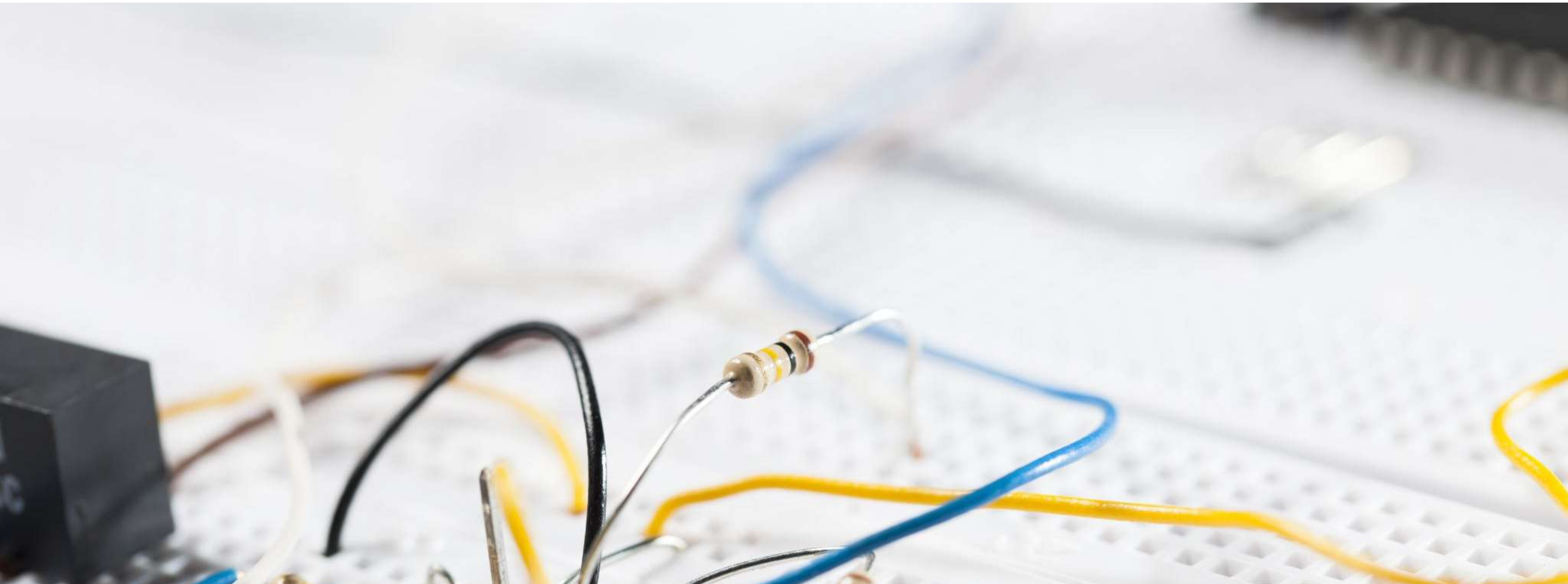
MSE: 30.6 MAE: 4.72 MAPE: 19.18%

<u>SARIMAX</u>

(1, 2, 2) x (0, 2, 2, 12)

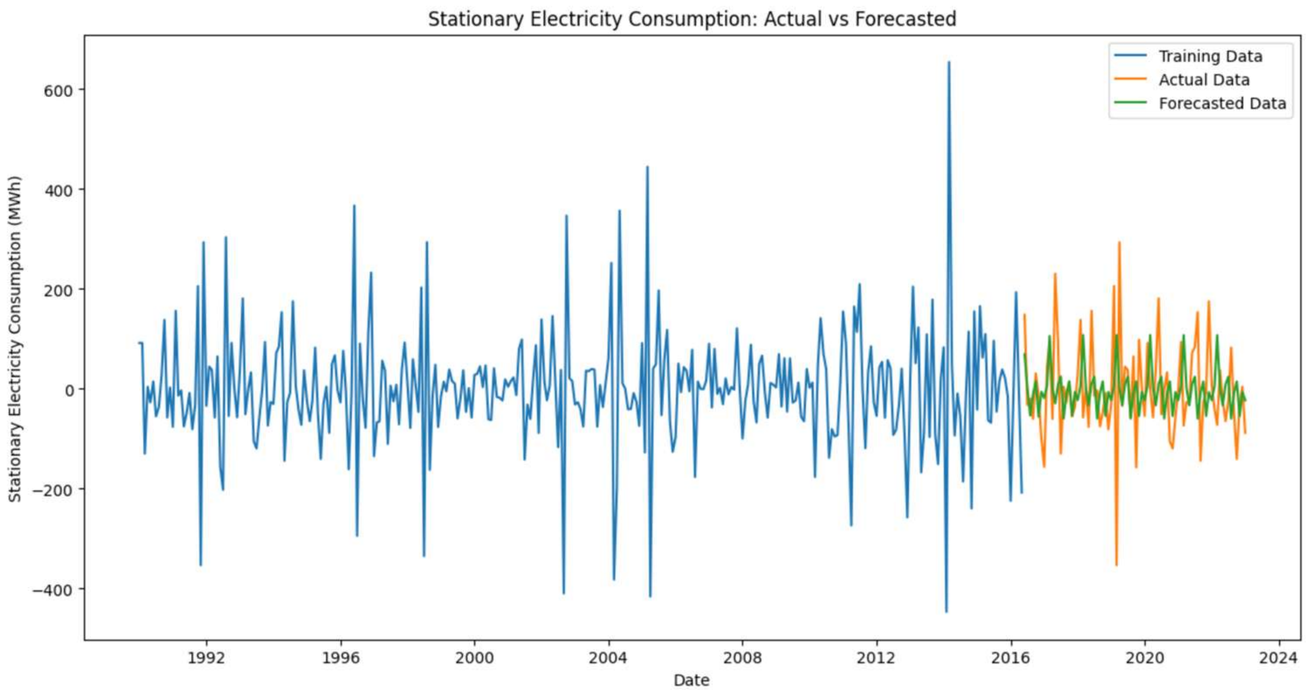


MSE: 12.56, MAE: 2.87, MAPE: 11.479%



ELECTRICITY CONSUMPTION ANALYSIS

STATIONARY ELECTRICTY CONSUMPTION USEFULNESS

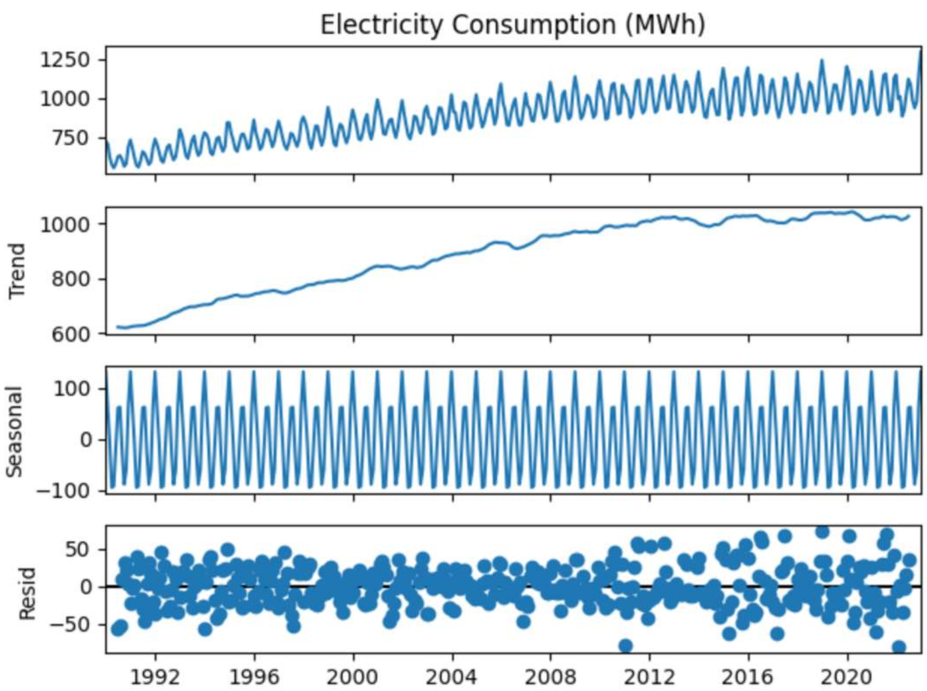


MSE: 12160, MAE: 79.7, MAPE: 395%

- MAPE of 395% indicates the model's prediction is 4times off from the actual value.
- Poor fit on the graph
- Lowest Gotten MAPE is 99%

- Stationary Electricity Consumption is not useful for forecasting

ELECTRICITY SEASONAL DECOMPOSITION

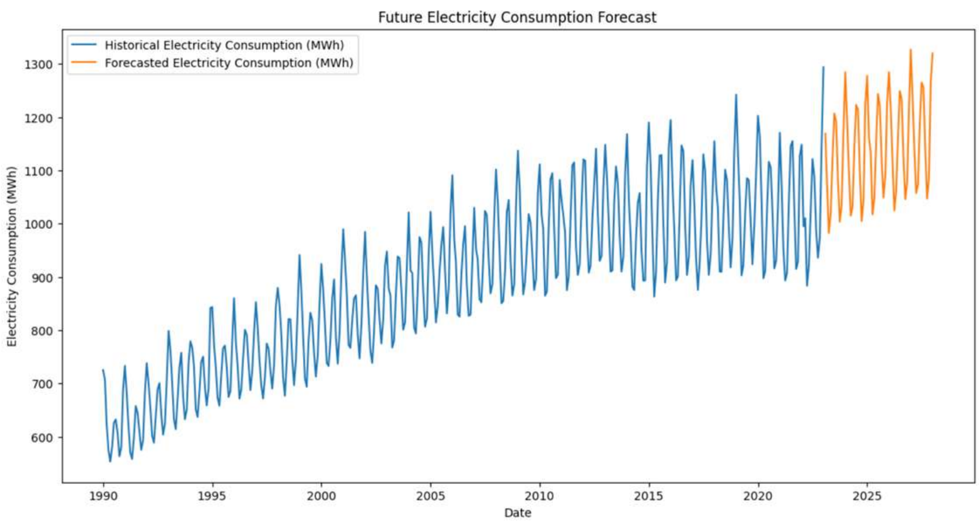


- Overall Electricity Consumption has increased by about 400 MWh
- Trend- Upward Non-linear Trend
- Seasonal-
  - fluctuation between 100 and -100 MWh
  - Consistent Seasonal Pattern
- Increase Variation in Residual as year increases

ELECTRICITY CONSUMPTION FORECAST

Holt-Winter's Seasonal Smoothing model

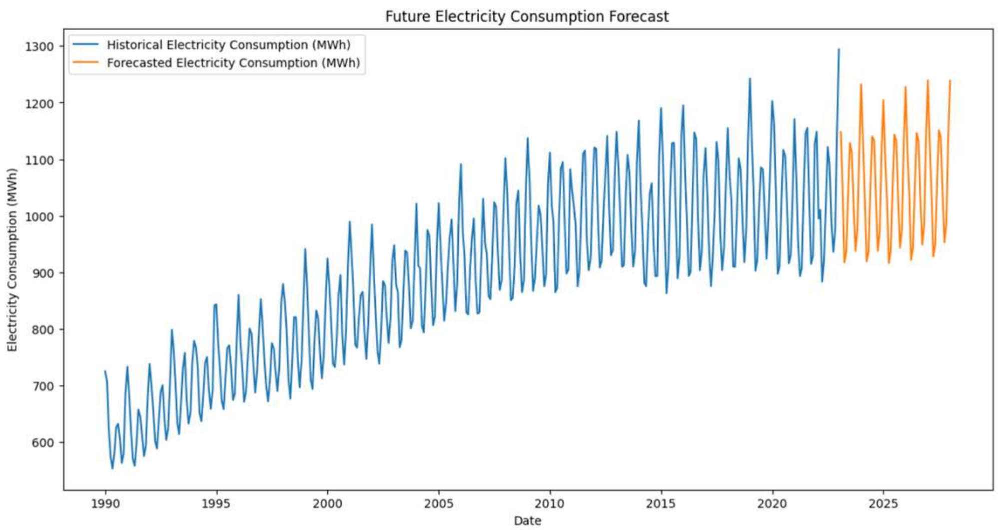
Trend = Add , Seasonal = Add, Period = 36



MSE = 1478, MAE =28.9, MAPE = 2.76%

SARIMAX

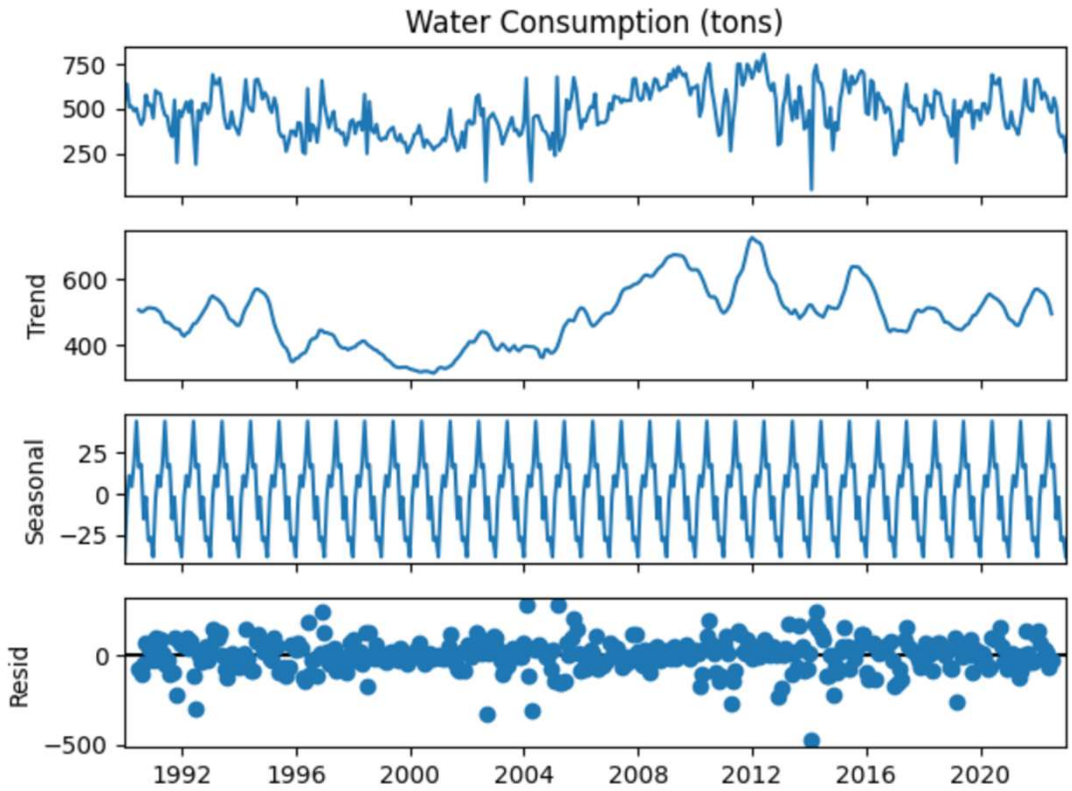
(2, 0, 2) x (2, 1, 1, 12)



MSE = 1290, MAE =26, MAPE = 2.44%

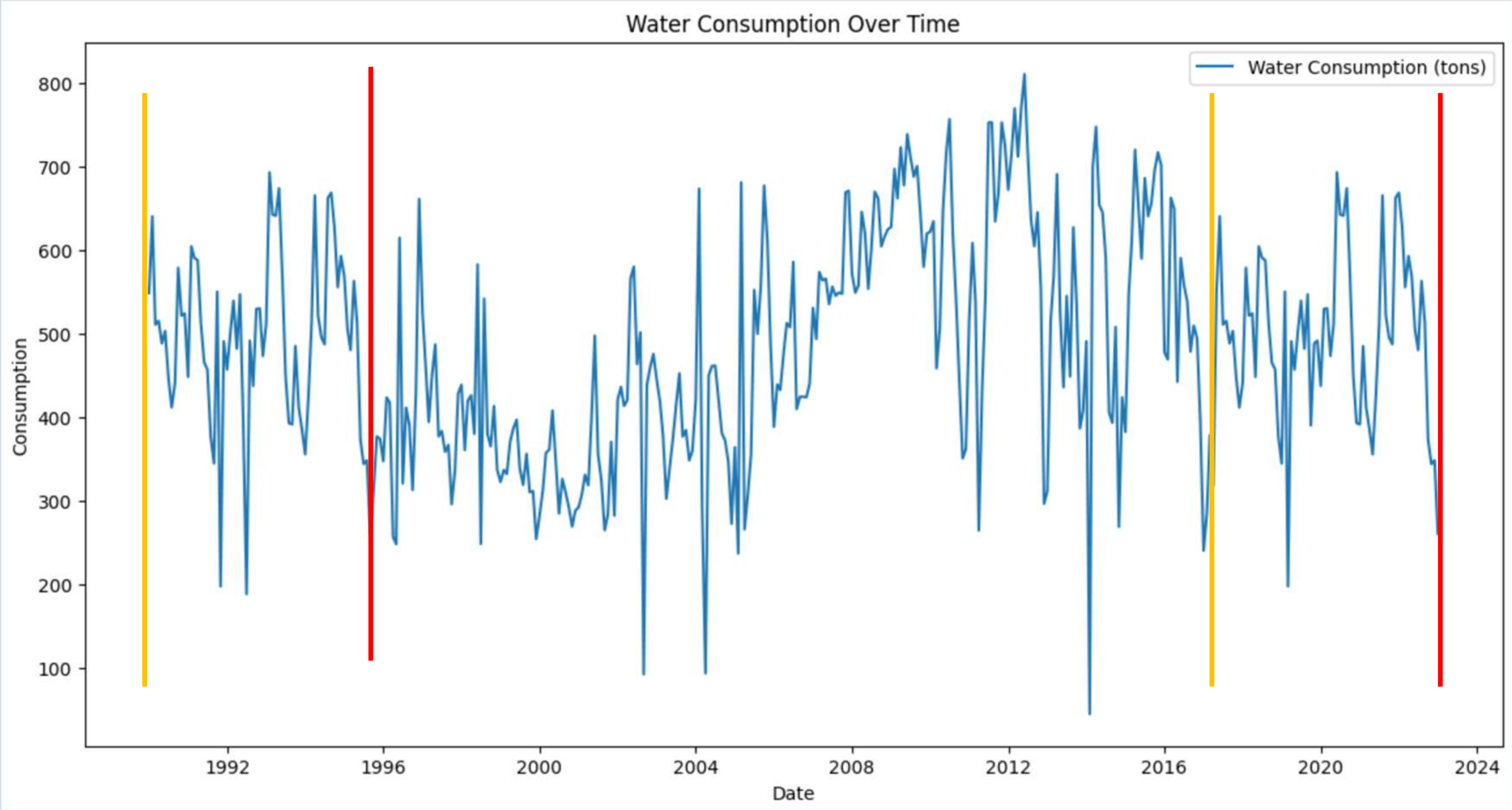
WATER CONSUMPTION ANALYSIS

WATER SEASONAL DECOMPOSITION

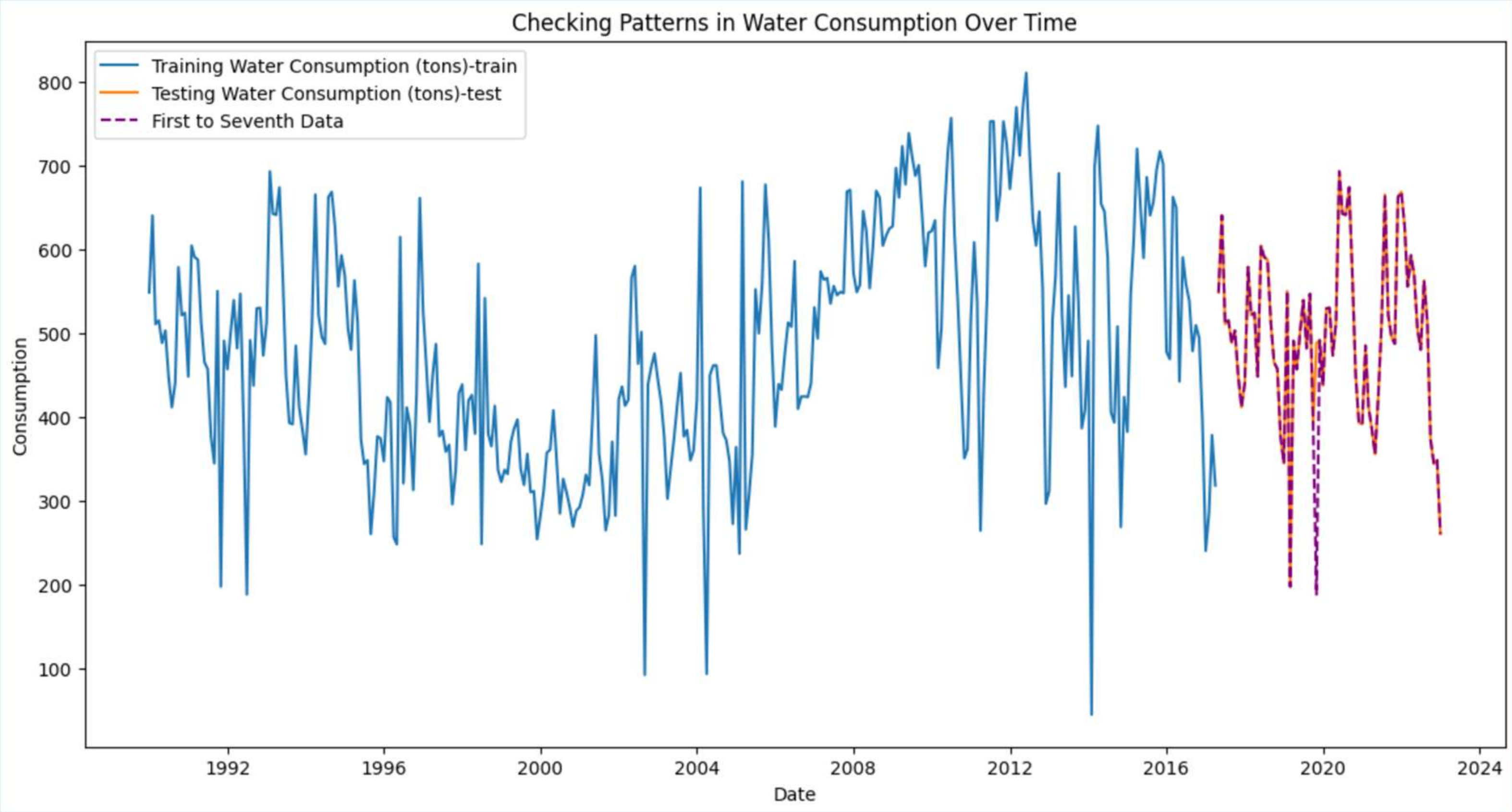


- High Fluctuation over time
- Sharp Drop at some periods (2013, 2004)
- Trend:
  - 1990 to 1995- stable
  - 1995 to 2000 – decline
  - 2000 to 2010 – gradual increase
  - 2010 to 2023 – several peaks and drops
- Consistent Seasonal Pattern
- Constant Variation in Residual as year increases (slight outliers)

WATER CONSUMPTION TIME SERIES



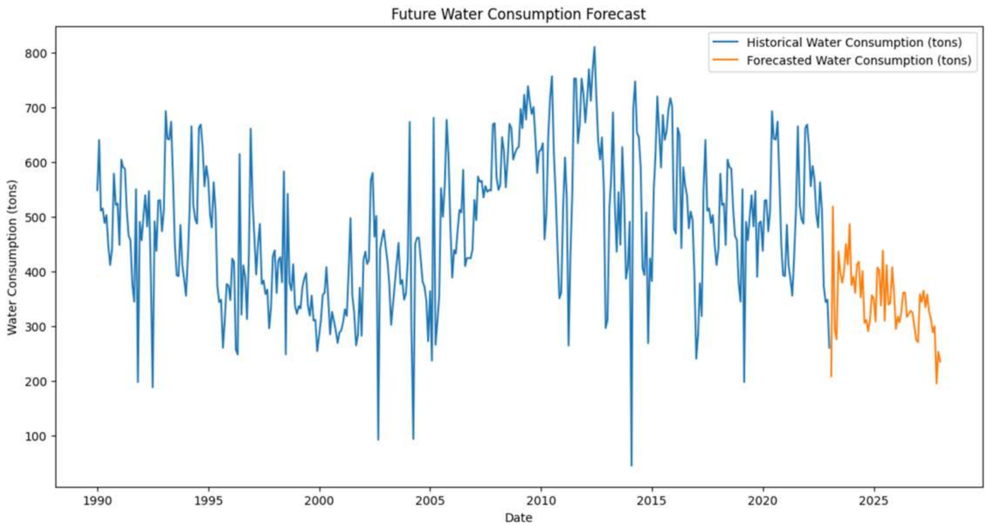
CONFIRMING ANALYSIS



WATER CONSUMPTION FORECAST

Holt-Winter's Seasonal Smoothing model

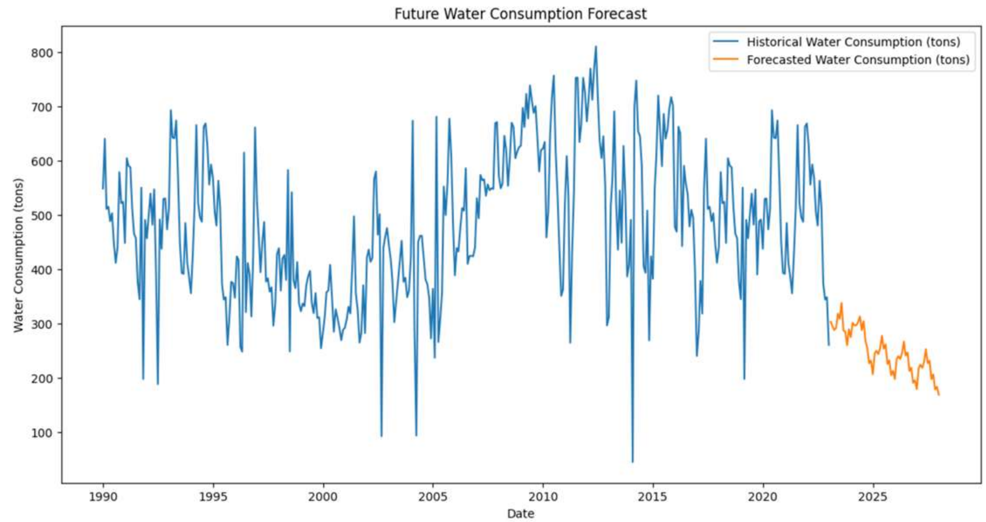
Trend = Add , Seasonal = Add, Period = 108



MSE = 10676, MAE =81.9, MAPE = 18.5%

SARIMAX

(2, 0, 2) x (2, 1, 1, 12)



MSE = 10871, MAE =81.6, MAPE = 17.5%

FINAL CONCLUSION

- Gas Consumption will increase in the near future
- Electricity Consumption will increase in the near future
- Water Consumption will slightly decrease in the near future

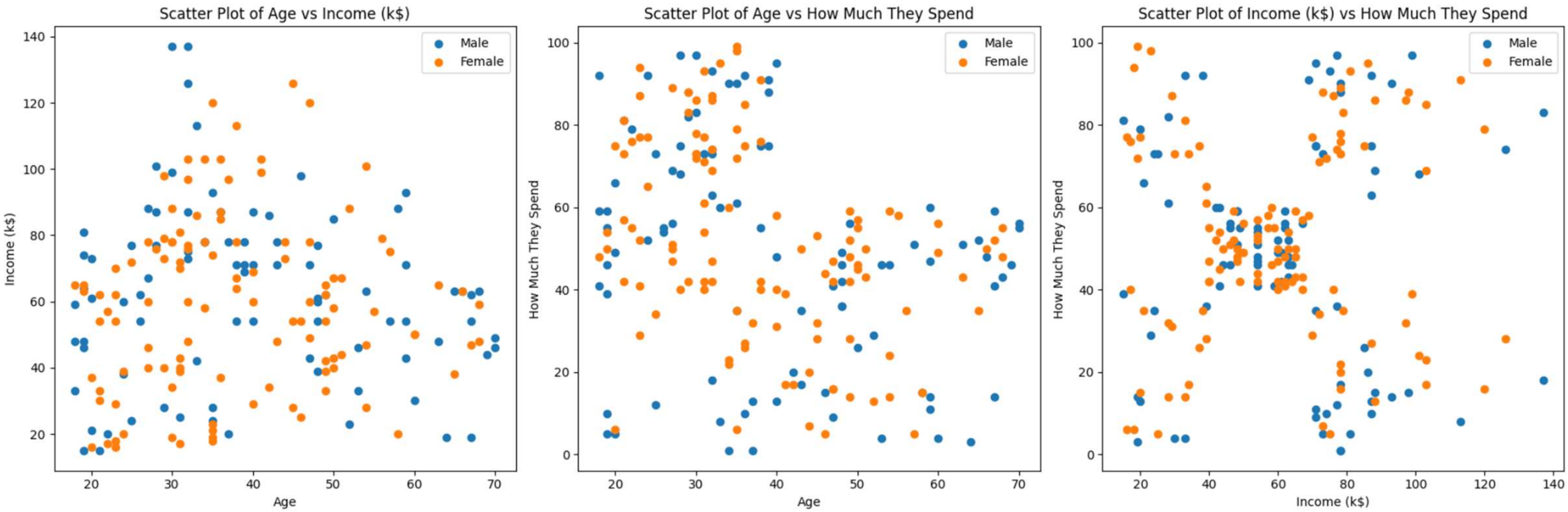
CUSTOMER SEGMENTATION

(Customer Dataset)
- Kaung Myat San

PROBLEM STATEMENT

Creating clusters that will allow us to effectively group customer with different needs

DATA ANALYSIS



FEATURES USABILITY

In all the graphs, both genders overlap across all features, indicating that gender is not a strong factor in forming distinct clusters

FEATURE ENGINEERING

I created a new feature called "Spending to Income Ratio," which represents the proportion of an individual's spending relative to their income.

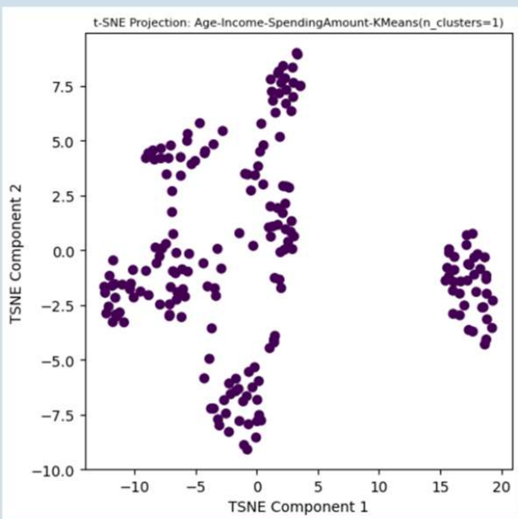
DOES SPENDING TO INCOME RATIO CONTRIBUTES TO CLUSTERING?

Variance of Spending to Income Ratio (with outliers)			1.0	
Variance of Spending to Income Ratio (without outliers)			0.153	
Data with Spending to Income Ratio				
Explained Variance Ratios (with outliers)	0.499	0.286	0.175	0.04
Explained Variance Ratios (without outliers):	0.472	0.275	0.247	0.00567
Data without Spending to Income Ratio				
Explained Variance Ratios (with outliers)	0.443	0.333	0.223	
Explained Variance Ratios (without outliers):	0.493	0.269	0.238	

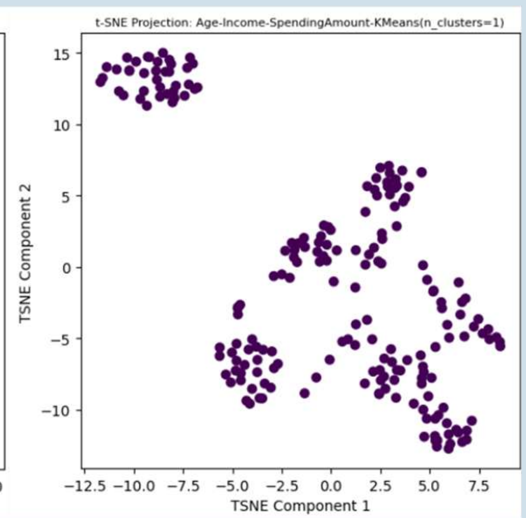
The variance for Spending to Income Ratio is low without outliers suggesting it does not separate cluster well.
From looking at the PCA with and without **"Spending To Income Ratio"**, it shows that **"Spending To Income Ratio"** contributes at most 0.57% which is very low.
Therefore, it should be better to use without it.

FEATURES SELECTION

Standard Scaler



Min-Max Scaler



Chosen Features
Age
Income
Spending Amount
Chosen Standardizing
Standard Scaler

K- Means	Silhouette Score
Z-Score (5 clusters)	0.421
Z-Score (6 clusters)	0.437
Min Max (5 clusters)	0.315
Min Max (6 clusters)	0.430

Based on the graph, both the Standard Scaled and Min-Max Scaled data appear to form 5 or 6 clusters. However, the clusters in the Standard Scaled data are more distinct and clearer.
As a result, I have decided to choose the Standard Scaled data.

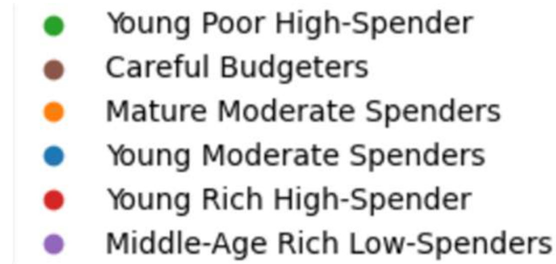
MODEL SELECTION

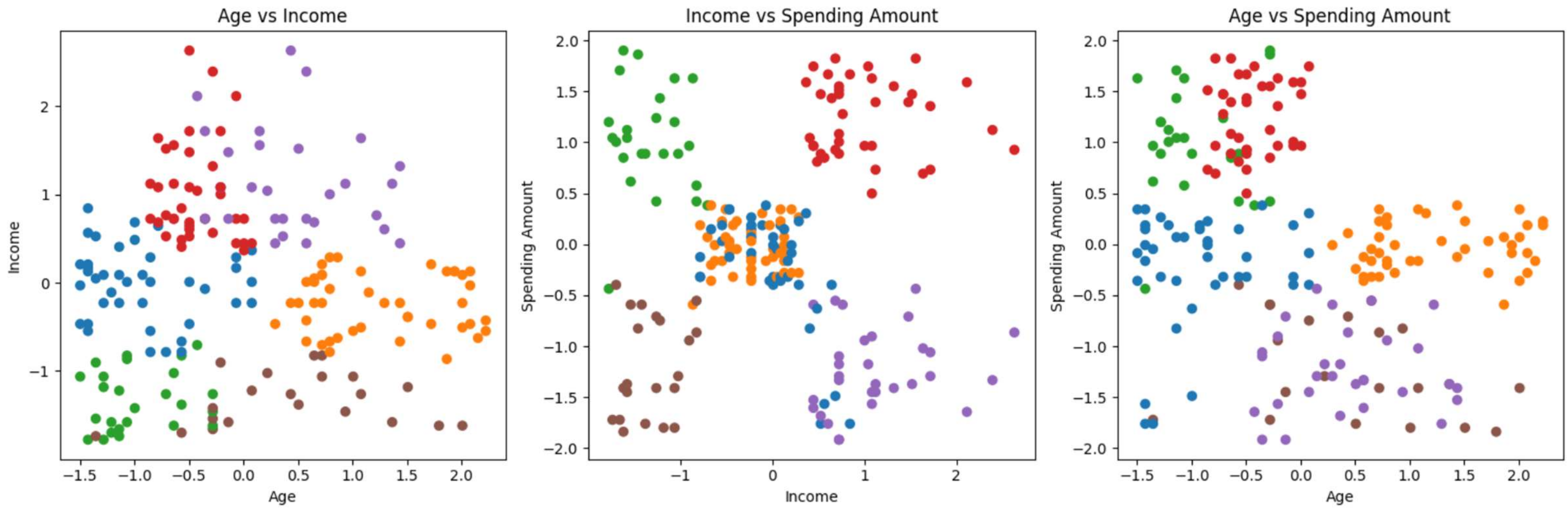
MODELS PERFORMANCE

Final Model: KMeans (n_clusters=6)

Models	Silhouette Score
DBSCAN (ep = 0.59, min_samples = 7)	0.276
GMM (covariance_type='tied', n_components=6)	0.434
AgglomerativeClustering(linkage='average', metric='manhattan', n_clusters=6)	0.426
OPTICS(max_eps=0.5, metric='cosine', min_samples=15)	0.378
SpectralClustering(affinity='nearest_neighbors', gamma=0.71, n_clusters=6)	0.430
KMeans(n_clusters=6)	0.437

FINAL SEGMENTATION





Segment	Profile	Age Quartiles	Income Quartiles	Spending Quartiles	Analysis
Green	Young Poor High-Spender	0 to 37.5 (0 to 3/8)	0 to 20	62.5 to 100 (5/8 to 1)	group living beyond their means or possibly supported by family
Red	Young Rich High-Spender	12.5 to 37.5(1/8 to 3/8)	50 to 100	62.5 to 100 (5/8 to 1)	group of young professionals with high disposable income
Purple	Middle-Age Rich Low-Spenders	25 to 75(2/8 to 6/8)	50 to 100	0 to 37.5 (0 to 3/8)	financially conservative individuals with good saving habits
Brown	Careful Budgeters	0 to 100	0 to 20	0 to 37.5 (0 to 3/8)	Shows financially constrained behavior across all age groups
Blue	Young Moderate Spenders	0 to 37.5 (0 to 3/8)	20 to 60	0 to 62.5 (0 to 5/8)	Represents relatively balanced financial behavior
Orange	Mature Moderate Spenders	37.5 to 100 (3/8 to 1)	20 to 50	37.5 to 62.5 (3/8 to 5/8)	Shows stable, balanced financial behavior among older individuals

CLUSTERS SUMMARY

CATERING RECOMMENDATION

Segment	Profile	Characteristics	Recommendation
Green	Young Poor High-Spender	Young, Low Income, High Spenders	Trendy, budget-friendly Items (Fast Fashion, Mobile Games)
Red	Young Rich High-Spender	Young, High Income, High Spenders	Luxury, High-end Items (designer brands, High-end Electronics)
Purple	Middle-Age Rich Low-Spenders	Middle-Age, High Income, Low Spenders	Investment & Financial Services (retirement planning)
Brown	Careful Budgeters	All Ages, Low Income, Low Spenders	Budget Items (Second-hand Items, cheap grocery)
Blue	Young Moderate Spenders	Young, Moderate Income, Moderate Spenders	Mid-Range Items (Items with good-values)
Orange	Mature Moderate Spenders	Older, Moderate Income, Moderate Spenders	Health & wellness products (supplements, fitness products)

	Age			Income (k$)			How Much They Spend			Count
	mean	min	max	mean	min	max	mean	min	max	count
Profile										
Careful Budgeters	45.523810	20	67	26.285714	16	39	19.380952	3	40	21
Mature Moderate Spenders	56.340909	43	70	53.704545	38	67	49.386364	35	60	44
Middle-Age Rich Low-Spenders	44.800000	33	59	88.200000	71	126	18.500000	1	39	30
Young Moderate Spenders	26.125000	18	40	59.425000	40	81	44.450000	5	60	40
Young Poor High-Spender	25.560000	18	35	26.480000	15	42	76.240000	39	99	25
Young Rich High-Spender	32.763158	27	40	85.210526	69	126	82.105263	63	97	38

CLUSTERS SUMMARY

FINAL CONCLUSION

- **Primary Focus: High-Spenders**
  - Spend **twice as much** as other groups.
  - Their count is **63 (30% of sample population).**
  - **Best products to target them:** Luxury, high-end gadgets, and premium items.
- **Secondary Focus: Moderate Spenders**
  - Spend **half as much** as high-spenders.
  - **Larger customer base** than high-spenders.
  - **Best products to target them:** Mid-range gadgets and items.
- **Overall Strategy:**
  - Prioritize selling **luxury and high-end** products to **maximize revenue**.
  - Supplement revenue by offering **mid-range products** to cater to a broader audience