# Report of Univariant Forecasting Models

When captured different customer groups, I found some of customer groups had different behavior when Covid-Pandemic occurred. So, I decided to make forecasting models for each customer group separately. Also, I made a model for each product separate in each group and found the best model in a select usage of each RMSE/mean groups. Besides having a consistency approach.

The models were really interesting which are lately supported by my findings in explanation and analysis.

## I. Individual approach for modling

List of each product of customer group

### Commercial Daily Usage

#### 1) stationary at 0% significance level

**Results of Dickey-Fuller Test:**

test statistic -3.251356

p-value 0.017198

Critical Value 1% -3.436224

Critical Value 5% -2.864127

Critical Value 10% -2.567997

**2) There is no special trend**

**3) Daily usage doesn't follow seasonality**

**4) Daily usage is not a white noise, because:**

Mean not zero and not constant over time, Daily usage values are not independent, they are correlated with many values in the latter months (autocorrelation at many lags are not zero)

**5) Daily Usage - Difference order 1: stationary at all significance levels**

**ARIMA Test Result:**

Test Statistic -5.162324

p-value 0.000009

Critical Value 1% -3.436150

Critical Value 5% -2.864095

Critical Value 10% -2.567979

Critical Value 1% 0.739000

**6) Performance of Univariant Time Series Forecasting Models**

**The Best model to forecast Commercial usage is Simple Exponential Smoothing with smoothing factor of 0.71**

- Model: Simple Exponential Smoothing
- α: 0.71
- AIC: 1.0
- BIC: 5.0
- AICc: 1.0
- HQ: 0.20

| | Model | Params | RMSE | MAE | MAPE | HQ |
|---|---|---|---|---|---|---|
| 0 | Simple Exponential Smoothing | alpha: 0.71 | 0.287 | 0.15 | 9.04 | 0.20 |
| 1 | Exponential Smoothing | α: 0.73, β: nan, γ: nan, φ: nan | 0.31 | 0.15 | 9.04 | 0.28 |
| 2 | Holt-Winters' Additive | α: 0.54, β: 0.0000, γ: 0.0000, φ: 0.99 | 0.31 | 0.165 | 10.34 | 0.76 |
| 3 | Holt-Winters' Multiplicative | α: 0.54, β: 0.0000, γ: 0.0000, φ: 0.99 | 0.31 | 0.165 | 10.34 | 0.76 |
| 4 | Holt's Linear Trend | α: 0.54, β: 0.0000, φ: 0.99 | 0.31 | 0.165 | 10.34 | 0.76 |
| 5 | Damped Trend | α: 0.54, β: 0.0000, φ: 0.99 | 0.31 | 0.165 | 10.34 | 0.76 |
| 6 | Moving Avg | q: 5, min_obs: 5 | 0.35 | 0.19 | 12.01 | 0.75 |
| 7 | Moving Avg Centered | q: 5, min_obs: 5, centered: True | 0.35 | 0.19 | 12.01 | 0.75 |
| 8 | AutoRegressive Moving Avg | lag: 2, ma_lag: 2, min_obs: 5 | 0.35 | 0.19 | 12.01 | 0.75 |
| 9 | AutoRegressive Moving Avg | lag: 2, ma_lag: 2, min_obs: 5 | 0.35 | 0.19 | 12.01 | 0.75 |
| 10 | ARIMA | lag=(2, 0, 2), ma_lag=0, 144=0 | 0.369 | 0.205 | 14.73 | 0.25 |
| 11 | ARIMA | lag=(2, 0, 2), ma_lag=0, 144=0 | 0.369 | 0.205 | 14.73 | 0.25 |
| 12 | SARIMAX | trend = [c], order = (1, 0, 1), s=7, 144=0 | 0.369 | 0.205 | 14.73 | 0.25 |
| 13 | SARIMAX | trend = [c], order = (1, 0, 1), s=7, 144=0 | 0.369 | 0.205 | 14.73 | 0.25 |
| 14 | SARIMAX | trend = [c], order = (1, 0, 1), s=7, 144=0 | 0.369 | 0.205 | 14.73 | 0.25 |

### Industrial Daily Usage

#### 1) stationary at 0% significance level

**Results of Dickey-Fuller Test:**

test statistic -3.811141

p-value 0.002802

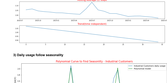
Critical Value 1% -3.436150

Critical Value 5% -2.864095

Critical Value 10% -2.567979

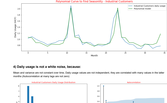
**2) There is no special trend**

**3) Daily usage follow seasonality**




**4) Daily usage is not a white noise, because:**

Mean not zero and not constant over time, Daily usage values are not independent, they are correlated with many values in the latter months (autocorrelation at many lags are not zero)




**5) Industrial Daily Usage - Difference order 1: stationary at all significance levels**

**ARIMA Test Result:**

Test Statistic -5.588953

p-value 0.000013

Critical Value 1% -3.436150

Critical Value 5% -2.864095

Critical Value 10% -2.567979

Critical Value 1% 0.739000

**6) Performance of Univariant Time Series Forecasting Models**

**The Best model to forecast Industrial usage is ARIMA (1,**

- Model: AutoRegressive Moving Average
- ARIMA: (1, 0, 0)
- AIC: 1.0
- BIC: 5.0
- AICc: 1.0
- HQ: 0.20

| | Model | Params | RMSE | MAE | MAPE | HQ |
|---|---|---|---|---|---|---|
| 0 | Simple Exponential Smoothing | α: 0.72 | 0.287 | 0.15 | 9.04 | 0.20 |
| 1 | Exponential Smoothing | α: 0.73, β: nan, γ: nan, φ: nan | 0.29 | 0.149 | 10.10 | 0.29 |
| 2 | Holt-Winters' Additive | α: 0.54, β: 0.0000, γ: 0.0000, φ: 0.99 | 0.29 | 0.149 | 10.10 | 0.29 |
| 3 | Holt-Winters' Multiplicative | α: 0.54, β: 0.0000, γ: 0.0000, φ: 0.99 | 0.29 | 0.149 | 10.10 | 0.29 |
| 4 | Holt's Linear Trend | α: 0.54, β: 0.0000, φ: 0.99 | 0.29 | 0.149 | 10.10 | 0.29 |
| 5 | Damped Trend | α: 0.54, β: 0.0000, φ: 0.99 | 0.29 | 0.149 | 10.10 | 0.29 |
| 6 | Moving Avg | q: 5, min_obs: 5 | 0.33 | 0.18 | 12.01 | 0.29 |
| 7 | Moving Avg Centered | q: 5, min_obs: 5, centered: True | 0.33 | 0.18 | 12.01 | 0.29 |
| 8 | AutoRegressive Moving Avg | lag: 2, ma_lag: 2, min_obs: 5 | 0.33 | 0.18 | 12.01 | 0.29 |
| 9 | AutoRegressive Moving Avg | lag: 2, ma_lag: 2, min_obs: 5 | 0.33 | 0.18 | 12.01 | 0.29 |
| 10 | ARIMA | lag=(2, 0, 1), ma_lag=0, 144=0 | 0.347 | 0.197 | 14.73 | 0.29 |
| 11 | ARIMA | lag=(2, 0, 1), ma_lag=0, 144=0 | 0.347 | 0.197 | 14.73 | 0.29 |
| 12 | SARIMAX | trend = [c], order = (1, 0, 1), s=7, 144=0 | 0.347 | 0.197 | 14.73 | 0.29 |
| 13 | SARIMAX | trend = [c], order = (1, 0, 1), s=7, 144=0 | 0.347 | 0.197 | 14.73 | 0.29 |
| 14 | SARIMAX | trend = [c], order = (1, 0, 1), s=7, 144=0 | 0.347 | 0.197 | 14.73 | 0.29 |

### Irrigation Group

#### 1) Stationary at 0% significance level

**Results of Dickey-Fuller Test:**

test statistic -3.538205

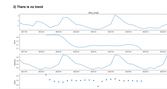
p-value 0.007000

Critical Value 1% -3.436150

Critical Value 5% -2.864095

Critical Value 10% -2.567979

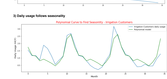
**2) There is no trend**




**3) Daily usage follow seasonality**

**4) Daily usage is not a white noise, because:**

Mean not zero and not constant over time, Daily usage values are not independent, they are correlated with many values in the latter months (autocorrelation at many lags are not zero)




**5) Irrigation Daily Usage - Difference order 1: stationary at all significance levels**

**ARIMA Test Result:**

Test Statistic -5.555217

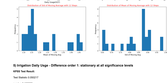
p-value 0.000001

Critical Value 1% -3.436150

Critical Value 5% -2.864095

Critical Value 10% -2.567979

Critical Value 1% 0.739000



**6) Performance of Univariant Time Series Forecasting Models**

**The Best model to forecast Irrigation usage is SARIMAX**

- Model: SARIMAX
- order: (1, 0, 1)
- seasonal_order: (1, 0, 1, 7)
- AIC: 1.0, BIC: 5.0, AICc: 1.0, HQ: -38.60
- Model: SARIMAX
- AIC: 1.0
- BIC: 5.0
- AICc: 1.0
- HQ: -38.60

| | Model | Params | RMSE | MAE | MAPE | HQ |
|---|---|---|---|---|---|---|
| 0 | Simple Exponential Smoothing | alpha: 0.71 | 0.279 | 0.139 | 10.15 | 0.36 |
| 1 | Exponential Smoothing | α: 0.73, β: nan, γ: nan, φ: nan | 0.29 | 0.149 | 10.10 | 0.29 |
| 2 | Holt-Winters' Additive | α: 0.54, β: 0.0000, γ: 0.0000, φ: 0.99 | 0.29 | 0.149 | 10.10 | 0.29 |
| 3 | Holt-Winters' Multiplicative | α: 0.54, β: 0.0000, γ: 0.0000, φ: 0.99 | 0.29 | 0.149 | 10.10 | 0.29 |
| 4 | Holt's Linear Trend | α: 0.54, β: 0.0000, φ: 0.99 | 0.29 | 0.149 | 10.10 | 0.29 |
| 5 | Damped Trend | α: 0.54, β: 0.0000, φ: 0.99 | 0.29 | 0.149 | 10.10 | 0.29 |
| 6 | Moving Avg | q: 5, min_obs: 5 | 0.33 | 0.18 | 12.01 | 0.29 |
| 7 | Moving Avg Centered | q: 5, min_obs: 5, centered: True | 0.33 | 0.18 | 12.01 | 0.29 |
| 8 | AutoRegressive Moving Avg | lag: 2, ma_lag: 2, min_obs: 5 | 0.33 | 0.18 | 12.01 | 0.29 |
| 9 | AutoRegressive Moving Avg | lag: 2, ma_lag: 2, min_obs: 5 | 0.33 | 0.18 | 12.01 | 0.29 |
| 10 | ARIMA | lag=(2, 0, 1), ma_lag=0, 144=0 | 0.347 | 0.197 | 14.73 | 0.29 |
| 11 | ARIMA | lag=(2, 0, 1), ma_lag=0, 144=0 | 0.347 | 0.197 | 14.73 | 0.29 |
| 12 | SARIMAX | trend = [c], order = (1, 0, 1), s=7, 144=0 | 0.347 | 0.197 | 14.73 | 0.29 |
| 13 | SARIMAX | trend = [c], order = (1, 0, 1), s=7, 144=0 | 0.347 | 0.197 | 14.73 | 0.29 |
| 14 | SARIMAX | trend = [c], order = (1, 0, 1), s=7, 144=0 | 0.347 | 0.197 | 14.73 | 0.29 |

### Duplex Group

#### 1) Daily usage stationary at 10% significance level

**Results of Dickey-Fuller Test:**

test statistic -2.439563

p-value 0.130739

Critical Value 1% -3.436150

Critical Value 5% -2.864095

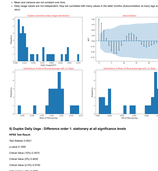
Critical Value 10% -2.567979

**2) There is no special trend**

**3) Daily usage doesn't follow seasonality**

### c) Daily usage is not a white noise, because:

- Mean and variance are not constant over time.
- Daily usage shows skewed distribution, they are correlated with many values in the latter months (Autocorrelation at many lags are not zero).




### d) Daily Data Usage - Difference order 1: stationary at all significance levels

**HYIU Test Result:**

Test Statistic: 0.001

p-value 0.100

Critical Value (10%) 0.1870

Critical Value (5%) 0.2610

Critical Value (2.5%) 0.3120

Critical Value (1%) 0.3480

### e) Performance of Univariant Time Series Forecasting Models:

**The Best model is Holt-Winters Additive Seasonal Smoothing:**

- Model: Simple Exponential Smoothing
- AIC: 1.0
- MAE: 0.00
- MAPE: 0.00
- RMSE: 0.1
- R2: 0.00

| | Model | Parameters | RMSE | MAE | MAPE | R2 |
|---|---|---|---|---|---|---|
| 0 | Simple Exponential Smoothing | α = 0.2 and α = 0.2 | 0.010 | 0.000 | 0.000 | 1.000 |
| 1 | Exponential Smoothing | α = 0.2, β = 0.2, γ = 0.2, and m = 30 | 0.030 | 0.000 | 0.000 | 0.000 |
| 2 | Holt-Winters Additive Seasonal Smoothing | α = 0.2, β = 0.2, γ = 0.2, and m = 30 | 0.030 | 0.000 | 0.000 | 0.000 |
| 3 | Auto Regression | α = 0.2, lags = 30, and seasonal Trends = False, AIC = 0.0 | 0.010 | 0.000 | 0.000 | 0.000 |
| 4 | Moving Avg | α = 0.2, lags = 30 | 0.010 | 0.000 | 0.000 | 0.000 |
| 5 | Moving Avg | α = 0.2, lags = 30 | 0.010 | 0.000 | 0.000 | 0.000 |
| 6 | Arim(Regression Moving Avg | (p,q) = (0, 0), and (0,0) | 0.010 | 0.000 | 0.000 | 0.000 |
| 7 | Arim(Regression Moving Avg | (p,q) = (0, 0), and (0,0) | 0.010 | 0.000 | 0.000 | 0.000 |
| 8 | Arim(Regression Moving Avg | (p,q) = (0, 0), and (0,0) | 0.010 | 0.000 | 0.000 | 0.000 |
| 9 | ARIMA | (p,d,q) = (1, 0, 1), and (0,0,0,30) | 0.010 | 0.000 | 0.000 | 0.000 |
| 10 | SARIMAX | Order = (p,d,q) = (1, 0, 1), (P,D,Q,m) = (0, 0, 1, 30), and (0,0,0,0) | 0.010 | 0.000 | 0.000 | 0.000 |

## Single Family

### f) Daily usage is not stationary at any significance level

**HYIU Test Result:**

Results of Kviatq Pyfer Test:

test statistic: 2.0100

p-value: 0.9900

Critical Value 1%: 0.2390000

Critical Value 5%: 2.3610000

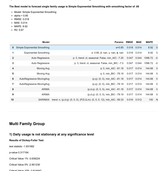
Critical Value 10%: 3.0210000

### g) There is no trend

### h) Daily usage doesn't follow seasonality

### i) Daily usage is not a white noise, because:

- Mean and variance are not constant over time.
- Daily usage shows skewed distribution, they are correlated with many values in the latter months (Autocorrelation at many lags are not zero).




### j) Single Family Daily Usage - Difference order 1: stationary at all significance levels

**HYIU Test Result:**

Test Statistic: 0.0000

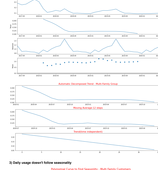
p-value 0.1000

Critical Value (10%) 0.1870000

Critical Value (5%) 0.2610000

Critical Value (2.5%) 0.3120000

Critical Value (1%) 0.3480000




### k) Performance of Univariant Time Series Forecasting Models:

**The Best model is Holt-Winters Additive Seasonal Smoothing with smoothing factor of 30:**

- Model: Simple Exponential Smoothing
- AIC: 1.0
- MAE: 0.00
- MAPE: 0.00
- RMSE: 0.0
- R2: 0.00

| | Model | Parameters | RMSE | MAE | MAPE | R2 |
|---|---|---|---|---|---|---|
| 0 | Simple Exponential Smoothing | α = 0.2 and α = 0.2 | 0.010 | 0.000 | 0.000 | 1.000 |
| 1 | Exponential Smoothing | α = 0.2, β = 0.2, γ = 0.2, and m = 30 | 0.010 | 0.000 | 0.000 | 1.000 |
| 2 | Holt-Winters Additive Seasonal Smoothing | α = 0.2, β = 0.2, γ = 0.2, and m = 30 | 0.010 | 0.000 | 0.000 | 1.000 |
| 3 | Auto Regression | α = 0.2, lags = 30, and seasonal Trends = False, AIC = 0.0 | 0.010 | 0.000 | 0.000 | 1.000 |
| 4 | Moving Avg | α = 0.2, lags = 30 | 0.010 | 0.000 | 0.000 | 1.000 |
| 5 | Moving Avg | α = 0.2, lags = 30 | 0.010 | 0.000 | 0.000 | 1.000 |
| 6 | Arim(Regression Moving Avg | (p,q) = (0, 0), and (0,0) | 0.010 | 0.000 | 0.000 | 1.000 |
| 7 | Arim(Regression Moving Avg | (p,q) = (0, 0), and (0,0) | 0.010 | 0.000 | 0.000 | 1.000 |
| 8 | Arim(Regression Moving Avg | (p,q) = (0, 0), and (0,0) | 0.010 | 0.000 | 0.000 | 1.000 |
| 9 | ARIMA | (p,d,q) = (1, 0, 1), and (0,0,0,30) | 0.010 | 0.000 | 0.000 | 1.000 |
| 10 | SARIMAX | Order = (p,d,q) = (1, 0, 1), (P,D,Q,m) = (0, 0, 1, 30), and (0,0,0,0) | 0.010 | 0.000 | 0.000 | 1.000 |

## Multi Family Group

### l) Daily usage is not stationary at any significance level

**HYIU Test Result:**

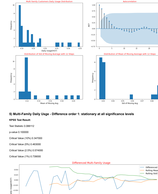
test statistic: 1.0400000

p-value 0.2710000

Critical Value (1%) 0.3480000

Critical Value (5%) 0.2610000

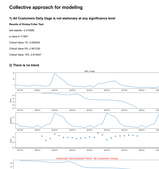
Critical Value (10%) 0.1870000




### m) Daily usage doesn't follow seasonality

### n) Daily usage is not a white noise, because:

- Mean and variance are not constant over time.
- Daily usage shows skewed distribution, they are correlated with many values in the latter months (Autocorrelation at many lags are not zero).




### o) Multi-Family Daily Usage - Difference order 1: stationary at all significance levels

**HYIU Test Result:**

Test Statistic: 0.0000 1.0

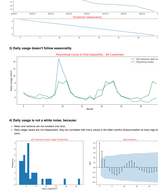
p-value 0.1000 0.1000

Critical Value (10%) 0.1870 0.1870

Critical Value (5%) 0.2610 0.2610

Critical Value (2.5%) 0.3120 0.3120

Critical Value (1%) 0.3480 0.3480




### p) Performance of Univariant Time Series Forecasting Models:

**The Best model is Holt-Winters Additive Seasonal Smoothing with smoothing factor of 1:**

- Model: Simple Exponential Smoothing
- AIC: 1.0
- MAE: 0.00
- MAPE: 0.00
- RMSE: 0.0
- R2: 0.00

| | Model | Parameters | RMSE | MAE | MAPE | R2 |
|---|---|---|---|---|---|---|
| 0 | Simple Exponential Smoothing | α = 0.2 and α = 0.2 | 0.010 | 0.000 | 0.000 | 1.000 |
| 1 | Exponential Smoothing | α = 0.2, β = 0.2, γ = 0.2, and m = 30 | 0.010 | 0.000 | 0.000 | 1.000 |
| 2 | Holt-Winters Additive Seasonal Smoothing | α = 0.2, β = 0.2, γ = 0.2, and m = 30 | 0.010 | 0.000 | 0.000 | 1.000 |
| 3 | Auto Regression | α = 0.2, lags = 30, and seasonal Trends = False, AIC = 0.0 | 0.010 | 0.000 | 0.000 | 1.000 |
| 4 | Moving Avg | α = 0.2, lags = 30 | 0.010 | 0.000 | 0.000 | 1.000 |
| 5 | Moving Avg | α = 0.2, lags = 30 | 0.010 | 0.000 | 0.000 | 1.000 |
| 6 | Arim(Regression Moving Avg | (p,q) = (0, 0), and (0,0) | 0.010 | 0.000 | 0.000 | 1.000 |
| 7 | Arim(Regression Moving Avg | (p,q) = (0, 0), and (0,0) | 0.010 | 0.000 | 0.000 | 1.000 |
| 8 | Arim(Regression Moving Avg | (p,q) = (0, 0), and (0,0) | 0.010 | 0.000 | 0.000 | 1.000 |
| 9 | ARIMA | (p,d,q) = (1, 0, 1), and (0,0,0,30) | 0.010 | 0.000 | 0.000 | 1.000 |
| 10 | SARIMAX | Order = (p,d,q) = (1, 0, 1), (P,D,Q,m) = (0, 0, 1, 30), and (0,0,0,0) | 0.010 | 0.000 | 0.000 | 1.000 |

## Collective approach for modeling

### q) All Customers Daily Usage is not stationary at any significance level

**Results of Kviatq Pyfer Test:**

test statistic: 2.0100

p-value 0.1000

Critical Value 1%: 0.2390

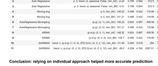
Critical Value 5%: 2.3610

Critical Value 10%: 3.0210

### r) Daily usage doesn't follow seasonality

### s) Daily usage is not a white noise, because:

- Mean and variance are not constant over time.
- Daily usage shows skewed distribution, they are correlated with many values in the latter months (Autocorrelation at many lags are not zero).




### t) All Customer groups Daily Usage - Difference order 1: stationary at all significance levels

**HYIU Test Result:**

Test Statistic: 0.0010 0.00

p-value 0.1000 0.1000

Critical Value (10%) 0.1870 0.1870

Critical Value (5%) 0.2610 0.2610

Critical Value (2.5%) 0.3120 0.3120

Critical Value (1%) 0.3480 0.3480




### u) Performance of Univariant Time Series Forecasting Models

**The Best model is Holt-Winters Additive Seasonal Smoothing with smoothing factor of 1:**

- Model: Simple Exponential Smoothing
- AIC: 1.0
- MAE: 0.00
- MAPE: 0.00
- RMSE: 0.0
- R2: 0.00

| | Model | Parameters | RMSE | MAE | MAPE | R2 |
|---|---|---|---|---|---|---|
| 0 | Simple Exponential Smoothing | α = 0.2 and α = 0.2 | 0.010 | 0.000 | 0.000 | 1.000 |
| 1 | Exponential Smoothing | α = 0.2, β = 0.2, γ = 0.2, and m = 30 | 0.010 | 0.000 | 0.000 | 1.000 |
| 2 | Holt-Winters Additive Seasonal Smoothing | α = 0.2, β = 0.2, γ = 0.2, and m = 30 | 0.010 | 0.000 | 0.000 | 1.000 |
| 3 | Auto Regression | α = 0.2, lags = 30, and seasonal Trends = False, AIC = 0.0 | 0.010 | 0.000 | 0.000 | 1.000 |
| 4 | Moving Avg | α = 0.2, lags = 30 | 0.010 | 0.000 | 0.000 | 1.000 |
| 5 | Moving Avg | α = 0.2, lags = 30 | 0.010 | 0.000 | 0.000 | 1.000 |
| 6 | Arim(Regression Moving Avg | (p,q) = (0, 0), and (0,0) | 0.010 | 0.000 | 0.000 | 1.000 |
| 7 | Arim(Regression Moving Avg | (p,q) = (0, 0), and (0,0) | 0.010 | 0.000 | 0.000 | 1.000 |
| 8 | Arim(Regression Moving Avg | (p,q) = (0, 0), and (0,0) | 0.010 | 0.000 | 0.000 | 1.000 |
| 9 | ARIMA | (p,d,q) = (1, 0, 1), and (0,0,0,30) | 0.010 | 0.000 | 0.000 | 1.000 |
| 10 | SARIMAX | Order = (p,d,q) = (1, 0, 1), (P,D,Q,m) = (0, 0, 1, 30), and (0,0,0,0) | 0.010 | 0.000 | 0.000 | 1.000 |

## Conclusion: relying on individual approach helped more accurate prediction