Report

Report

When explained different customer groups, I found some of customer groups had different behavior when Cross-Functional Teams Buying Group 2 (Mary) and Industrial Daily Usage (Commercial Daily Usage) were used. Thus, I decided to use the hierarchical approach to segment and find the best model in predict usage of each RFM/Demand group. Besides having a robust time series model.

The residual were really interesting which are totally supported by my findings in explanation and analysis.

Individual approach for modling

Let's look around 10 customer groups

Commercial Daily Usage

1) stationary at 0% significance level

Results of Dickey-Fuller Test:

test statistic: -3.20188

p-value 0.01798

Lags Used 0.000000

Observation Used 34.000000

Critical Value 1% -3.65024

Critical Value 5% -2.95106

Critical Value 10% -2.61607

2) There is no special trend

Automatic Decomposed Trend - Commercial Usage

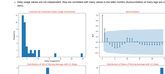
Trend(Non Independent)

3) Daily usage doesn't follow seasonality

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

Mean and variance are 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

DF0 Test Result:

Test Statistic -5.15204

p-value 0.00009

Lags Used 0.000000

Critical Value (1%) -3.65024

Critical Value (5%) -2.95106

Critical Value (10%) -2.61607

Critical Value (1%) 2.75800

6) Performance of Univariant Time Series Forecasting Models:

The Best model is Innovat Commercial usage is Simple Exponential Smoothing with smoothing factor of 0.71

- Model: Simple Exponential Smoothing
- SMOOTHING_LEVEL: 0.710000
- NAME: 0.107
- AIC: 1.000
- BIC: 5.000
- SSE: 0.000

	Model	Process	RMSE	MAE	MAPE	R2
0	Simple Exponential Smoothing	α=0.71	0.107	0.074	0.04	0.00
1	Exponential Smoothing	α 0.72, & beta= 0.01	0.107	0.074	0.04	0.00
2	Holt Regression	α, β fixed & seasonal Freq=(M_QS)	0.108	0.080	157.42	0.18
3	Holt Regression	α, β fixed & seasonal Freq=(M_QS)	0.108	0.080	157.41	0.18
4	Holt Winter	α, β, γ fixed, and Freq=(M_QS)	0.108	0.080	157.41	0.18
5	Holt Winter	α, β, γ fixed, and Freq=(M_QS)	0.108	0.080	157.41	0.18
6	ExpRegression Damping	(α, β, γ, ϕ), and (M, M)	0.108	0.080	167.75	25.25
7	ExpRegression Damping	(α, β, γ, ϕ), and (M, M)	0.108	0.080	168.25	25.25
8	ARIMA	(p,d,q) = (0,0,0), and (M1, M2)	0.108	0.080	167.75	25.25
9	ARIMA	(p,d,q) = (0,0,0), and (M1, M2)	0.108	0.080	168.25	25.25
10	SARIMAX	trend = (p,d,q) = (1,0,1), SP(D,Q,S) = (1,1,1), (M1, M2)	0.108	0.080	168.25	25.25
11	SARIMAX	trend = (p,d,q) = (1,0,1), SP(D,Q,S) = (1,1,1), (M1, M2)	0.108	0.080	640.01	-0.3

Industrial Daily Usage

1) stationary at 0% significance level

Results of Dickey-Fuller Test:

test statistic: -3.20188

p-value 0.01798

Lags Used 0.000000

Observation Used 34.000000

Critical Value 1% -3.65024

Critical Value 5% -2.95106

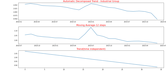
Critical Value 10% -2.61607

2) There is no special trend

Automatic Decomposed Trend - Industrial Usage

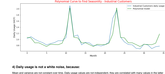
Trend(Non Independent)

3) Daily usage follow seasonality



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

Mean and variance are 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

DF0 Test Result:

Test Statistic -5.00004

p-value 0.00002

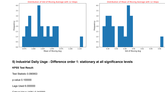
Lags Used 0.000000

Critical Value (1%) -3.65024

Critical Value (5%) -2.95106

Critical Value (10%) -2.61607

Critical Value (1%) 2.75800



6) Performance of Univariant Time Series Forecasting Models:

The Best model is Innovat Industrial usage is ARIMA(1,1)

- Model: ExpRegression Damping Package
- AIC: 1.000
- BIC: 5.000
- NAME: 0.108
- SSE: 0.000
- MAE: 0.080
- MAPE: 168.25
- R2: 25.25

	Model	Process	RMSE	MAE	MAPE	R2
0	Simple Exponential Smoothing	α=0.62	0.224	0.205	0.03	-0.28
1	Exponential Smoothing	α 0.72, & beta= 0.01	0.224	0.205	0.03	-0.28
2	Holt Regression	α, β fixed & seasonal Freq=(M_QS)	0.105	0.080	134.95	0.17
3	Holt Regression	α, β fixed & seasonal Freq=(M_QS)	0.105	0.080	134.95	0.17
4	Wrong Arg	α, β, γ, and (M, M)	0.108	0.080	168.25	25.25
5	Wrong Arg	α, β, γ, and (M, M)	0.108	0.080	168.25	25.25
6	ExpRegression Damping	(α, β, γ, ϕ), and (M, M)	0.108	0.080	168.25	25.25
7	ExpRegression Damping	(α, β, γ, ϕ), and (M, M)	0.108	0.080	168.25	25.25
8	ARIMA	(p,d,q) = (1,1,0), and (M1, M2)	0.108	0.080	168.25	25.25
9	ARIMA	(p,d,q) = (1,1,0), and (M1, M2)	0.108	0.080	168.25	25.25
10	SARIMAX	trend = (p,d,q) = (1,0,1), SP(D,Q,S) = (1,1,1), (M1, M2)	0.108	0.080	168.25	25.25
11	SARIMAX	trend = (p,d,q) = (1,0,1), SP(D,Q,S) = (1,1,1), (M1, M2)	0.108	0.080	640.01	-0.3

Irrigation Group

1) Stationary at 1% significance level

Results of Dickey-Fuller Test:

test statistic -3.53101

p-value 0.01080

Lags Used 1.000000

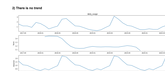
Observation Used 33.000000

Critical Value 1% -3.66137

Critical Value 5% -2.96127

Critical Value 10% -2.61907

2) There is no trend



Automatic Decomposed Trend - Irrigation Group

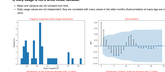
Trend(Non Independent)

3) Daily usage follows seasonality

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

Mean and variance are 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

DF0 Test Result:

Test Statistic -5.05017

p-value 0.00001

Lags Used 0.000000

Critical Value (1%) -3.65024

Critical Value (5%) -2.95106

Critical Value (10%) -2.61607

Critical Value (1%) 2.75800

6) Performance of Univariant Time Series Forecasting Models:

The Best model is Innovat Irrigation usage is SARIMAX

- Model: SARIMAX
- AIC: 1.000
- BIC: 5.000
- NAME: 0.107
- P,D,Q,S: (1, 0, 1), 12), MU: (M1, M2)
- SSE: 0.000
- MAE: 0.078
- MAPE: 167.75
- R2: 25.25

	Model	Process	RMSE	MAE	MAPE	R2
0	Simple Exponential Smoothing	α=0.71	0.107	0.074	0.04	0.00
1	Exponential Smoothing	α 0.72, & beta= 0.01	0.107	0.074	0.04	0.00
2	Holt Regression	α, β fixed & seasonal Freq=(M_QS)	0.107	0.077	170.08	-0.21
3	Holt Regression	α, β fixed & seasonal Freq=(M_QS)	0.108	0.077	170.08	-0.21
4	Wrong Arg	α, β, γ, and (M1, M2)	0.108	0.074	167.25	0.07
5	Wrong Arg	α, β, γ, and (M1, M2)	0.108	0.074	167.25	0.07
6	ExpRegression Damping	(α, β, γ, ϕ), and (M1, M2)	0.108	0.074	167.25	0.07
7	ExpRegression Damping	(α, β, γ, ϕ), and (M1, M2)	0.108	0.074	167.25	0.07
8	ARIMA	(p,d,q) = (1,1,1), and (M1, M2)	0.108	0.074	167.25	0.07
9	ARIMA	(p,d,q) = (1,1,1), and (M1, M2)	0.108	0.074	167.25	0.07
10	SARIMAX	trend = (p,d,q) = (1,0,1), SP(D,Q,S) = (1,1,1), (M1, M2)	0.108	0.074	167.25	0.07
11	SARIMAX	trend = (p,d,q) = (1,0,1), SP(D,Q,S) = (1,1,1), (M1, M2)	0.108	0.074	640.01	-0.3

Duplex Group

1) Daily usage stationary at 10% significance level

Results of Dickey-Fuller Test:

test statistic -2.33818

p-value 0.16157

Lags Used 1.000000

Observation Used 33.000000

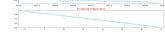
Critical Value 1% -3.66137

Critical Value 5% -2.96127

Critical Value 10% -2.61907

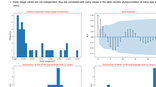
2) There is no special trend

Automatic Decomposed Trend - Duplex Group



Trend(Non Independent)

3) Daily usage doesn't follow seasonality



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

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

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

HYPO Test Results:

Test statistic: 0.0001

p-value of 0.1000

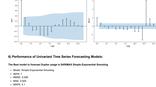
Lags Used: 23000

Critical Value: 10%: 0.3475

Critical Value: 5%: 0.4630

Critical Value: 2.5%: 0.7140

Critical Value: 1%: 0.7390


Differenced Digital Usage
Differenced Digital Usage
Rolling Mean
Rolling Mean

B) All Customer groups Daily Usge - Difference order 1: stationary at all significance levels

ADFtest Test Result:

Test Statistic -5.6413-05

p-value 0.10000

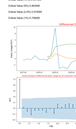
Lags Used 1.00000E

Critical Value 1%(1) -3.42169E

Critical Value 5%(1) -2.86300E

Critical Value 10%(1) -2.57000E

Critical Value 1%(1) 1.94000E


Differenced Daily Usage by all Customer Groups
Differenced
Residential
Analog Sales
Industrial
Jan 01
Jan 01
Jan 01