title: “Milestone 3”
author: “Abdulrahman , Ahmed Fahad, Ahmad Al Marri, Khalid , Malik Umar, Marwan”
date: “September 28, 2022”
output:
pdf_document: null
toc: yes
toc_depth: 2
latex_engine: xelatex
word_document: default
header-includes:

The Manifesto

Modeling - Marwan & Malik discussed with Ahmad/Ahmad and Abdulrahman, & Khaled (Everyone as we contribute 9 hours zoom meeting).

Analytics -Marwan & Malik discussed with Ahmad/Ahmad and Abdulrahman & Khaled (Everyone as we contribute 6 hours zoom meeting).

Findings -Ahmad,Ahmad & Khaled, Abdulrahman discussed with Marwan & Malik (Everyone as we contribute 5 hours zoom meeting).

Research -Ahmad & Ahmad & Khaled, Abdulrahman discussed with Marwan & Malik Everyone as we contribute 10 hours zoom meeting).

Milestone 2 - Everyone contributed equally.

Highlight of the problem & the frame

The data size is at 1.07 GB-1.23 GB (before & after) with +50 million in length & eight features in width, the .csv files of 2020-2022 has features such as the rate category, consumption unit, billing portions, the community number, the period, contract number, business numbers, the month based calendar and finally the business partner additional data set as extra to observe the events alongside the consumption rate includes additional Covid-19 data set of the world however the only three features to be considered is the country, date and the new cases. Furthermore, an LME frame with trend of aluminum prices (two features: Date, Dollars) (reference to DEWA first largest consumer EGA, after DEWA itself) (the other two sets for Millstone 3 is we can use it after the full perfection).

we have planned to apply a machine learning algorithm to predict the future demand of energy & to analyze the difference between the normal time and the event based time to extract the knowledge from the set, and to solve the mystery behind the abnormal and correlated trends whether they are favorable or not, the problem is that the current prediction model of DEWA isn't actually working due to wrong entries since the model depends on the filter data rather than the raw one.

Methodology

According to my previous possible questions and outcomes of Milestone 1, it is planned to carry out a different type of forecast models, and we've inserted a seasonal naïve method since it doesn't take much time to construct & run the algorithm and it can forecast based on the previous outcomes without interfering much with the factors and it follows the formula of T + h, Additionally it's a good approach for time series seasonal based data frame, it is somewhat similar to what is used in python known is the fancy impute from lubridate & fpp2 The second was the ETS forecast which is the exponential smoothing from the name and the look of the outcome it is similar the excel exponential forecast macros pack (By Macros), it depends a lot on the timeline, and it is a good one for consumer-based trends and continuous data in addition to unchanged cycles between the different locations of the data. The third being the ARIMA, which will be the selected candidate which we will conclude later why, it is known as the Auto regressive integrated moving average of which is a time series model that is used in statistics and econometrics to assess events that occur over time. It forecasts the observation based on historical data of earlier time spots recorded for the same observation. The model is employed to comprehend historical data and forecast subsequent data in a series, and it is useful to forecast the future demands of energy. Our selected Algorithm were all supervised based ML , due to the existence of time series based data and the presence of labelled input and outputs in addition they are linear based as the correlation matrix "heatmap style" (after the cut of 2020) illustrates that the correlation is present & clear between the attributes of the data with the addition of the data-frame which are the and the set thereby it usually shows a normal line without jumps/drops since it's not an gas turbine comparison but rather the whole thing in the comparison. Note that LSTM would be used in the Milestone 3 Now going back on why exactly we've chosen the ARIMA well as you can see in the outputs the p values are less and the standard deviation residual is also less therefore in the graph below based on the basis of the F-statistics we will implement the ARIMA, as you can see the disadvantages of the seasonal is that it took the min peak to the highest amounts based on one year of 2020 which makes the model bias, and same it goes for the ETS it is closer to the actual outputs as indicated by the graph however we will still go with ARIMA due to the better %'s of the output and the higher favorable outputs , in the next milestone a high chance to conduct the LSTM as a fourth model and possibility of utilizing a hybrid models, more to added the Datasets of other frame will be inserted for further investigation,

Initialization

This section we are going to include all R library which included in this project.

```
## Loading required package: tidyverse

## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## Loading required package: data.table
##
##
## Attaching package: 'data.table'
##
##
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
##
##
## The following object is masked from 'package:purrr':
##
##     transpose
##
##
## Loading required package: xtable
##
## Loading required package: lubridate
##
##
## Attaching package: 'lubridate'
##
##
## The following objects are masked from 'package:data.table':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
##
##
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
##
##
## Loading required package: viridis
##
## Loading required package: viridisLite
##
```

```
## Loading required package: extrafont
##
## Registering fonts with R
##
## Loading required package: plyr
##
## ------------------------------------------------------------------------------
##
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
##
## ------------------------------------------------------------------------------
##
##
## Attaching package: 'plyr'
##
##
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
##
##
## The following object is masked from 'package:purrr':
##
##     compact
##
##
## Loading required package: magrittr
##
##
## Attaching package: 'magrittr'
##
##
## The following object is masked from 'package:purrr':
##
##     set_names
##
##
## The following object is masked from 'package:tidyr':
##
##     extract
##
##
## Loading required package: ggcorrplot
##
## Loading required package: corrplot
##
## corrplot 0.92 loaded
##
## Loading required package: fpp2
##
## Registered S3 method overwritten by 'quantmod':
```

```
##   method            from
##   as.zoo.data.frame zoo
##
## -- Attaching packages ---------------------------------------------- fpp2 2.4 --
##
## v forecast  8.17.0     v expsmooth 2.3
## v fma       2.4
##
## -- Conflicts ------------------------------------------------- fpp2_conflicts --
## x plyr::compact()         masks purrr::compact()
## x magrittr::extract()     masks tidyr::extract()
## x magrittr::set_names()   masks purrr::set_names()
## x data.table::transpose() masks purrr::transpose()
```

Data Set

Loading DEWA Dataset

Variable Names Dataset

```
## [1] "billing_portion"    "community"          "rate_category"
## [4] "consumption_period" "calendar_month"     "contract_account"
## [7] "business_partner"   "consumption_unit"
```

Summary Of Dataset

```
##  billing_portion      community     rate_category      consumption_period
##  Length:9972        Min.   :  0.0   Length:9972        Length:9972
##  Class :character   1st Qu.:313.0   Class :character   Class :character
##  Mode  :character   Median :381.0   Mode  :character   Mode  :character
##                     Mean   :420.9
##                     3rd Qu.:614.0
##                     Max.   :991.0
##                     NA's   :1
##  calendar_month   contract_account      business_partner   consumption_unit
##  Min.   :201703   Min.   :-1294962332   Min.   :    1062   Min.   :     0.0
##  1st Qu.:202003   1st Qu.: 2015847090   1st Qu.:10461535   1st Qu.:   152.6
##  Median :202107   Median : 2030913740   Median :11057883   Median :   410.0
##  Mean   :202096   Mean   : 1821980189   Mean   :11374045   Mean   :  2921.6
##  3rd Qu.:202112   3rd Qu.: 2036165974   3rd Qu.:11440151   3rd Qu.:  1201.5
##  Max.   :202205   Max.   : 3006011932   Max.   :17099140   Max.   :988000.0
##
```

```
## Classes 'data.table' and 'data.frame':   9972 obs. of  8 variables:
##  $ billing_portion   : chr  "B30" "B32" "B20" "B26" ...
##  $ community         : num  511 394 319 362 346 352 127 335 318 318 ...
##  $ rate_category     : chr  "RESIEXPE" "RESIEXPE" "RESIEXPE" "RESINATE" ...
##  $ consumption_period: chr  "17.03.2022" "25.02.2020" "05.08.2021" "18.03.2020" ...
##  $ calendar_month    : int  202203 202001 202108 202003 202205 202003 202201 202104 202112 202002 ..
##  $ contract_account  :integer64 2041277092 2002144214 2033430979 2001578237 2041929634 2031531875 202
##  $ business_partner  : int  10615466 10418039 10647258 10070945 11796473 10166549 10036650 17062092 1
##  $ consumption_unit  : num  100 2821 340 2446 124 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

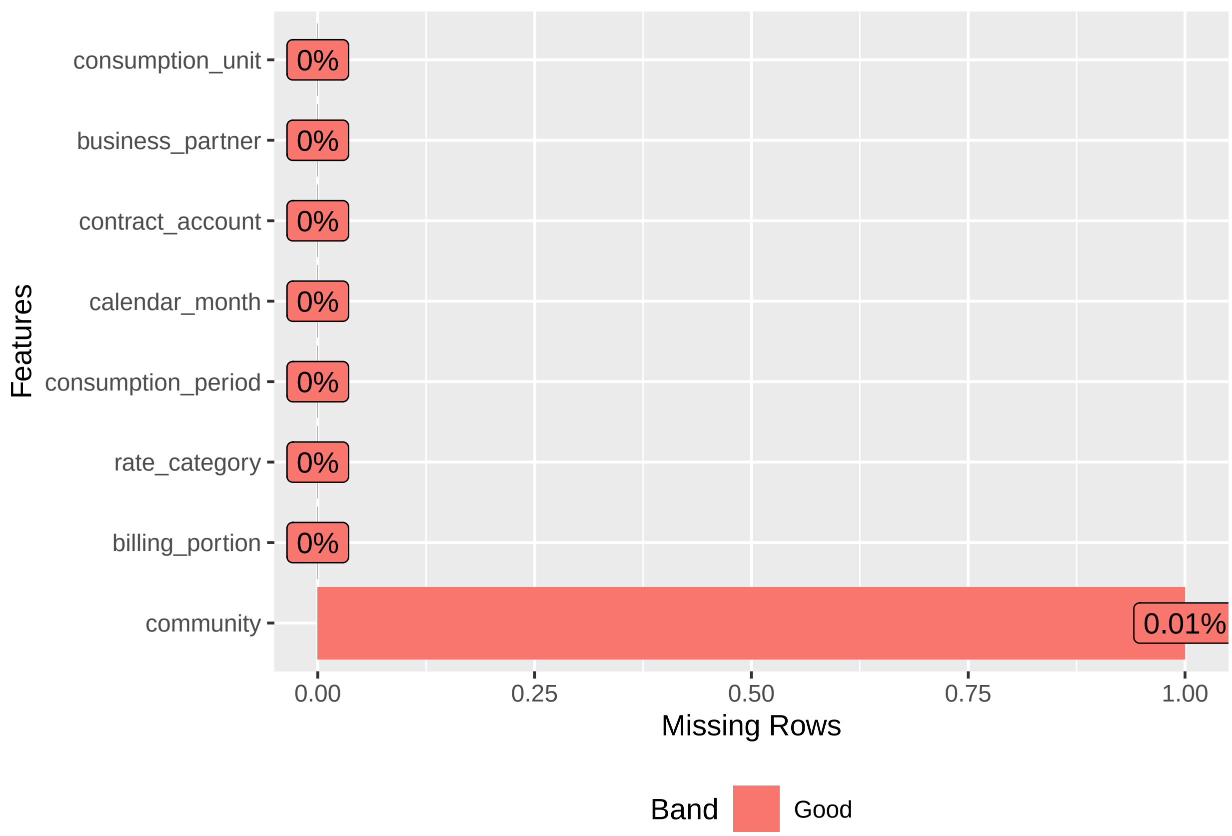
```
##    billing_portion community rate_category consumption_period calendar_month
## 1:             B30       511      RESIEXPE         17.03.2022         202203
## 2:             B32       394      RESIEXPE         25.02.2020         202001
## 3:             B20       319      RESIEXPE         05.08.2021         202108
## 4:             B26       362      RESINATE         18.03.2020         202003
## 5:             B28       346      RESIEXPE         13.05.2022         202205
## 6:             B24       352      COMMELEC         31.03.2020         202003
##    contract_account business_partner consumption_unit
## 1:       2041277092         10615466          100.000
## 2:       2002144214         10418039         2820.929
## 3:       2033430979         10647258          340.000
## 4:       2001578237         10070945         2446.210
```

```
## 5:       2041929634         11796473         124.000
## 6:       2031531875         10166549       32481.598
```

```
##    billing_portion community rate_category consumption_period calendar_month
## 1:             D31       648      RESIEXPE         17.10.2021         202110
## 2:             D33       911      RESINATE         23.04.2021         202104
## 3:             D29       421      RESIEXPE         14.02.2021         202102
## 4:             D23       241      COMMELEC         08.08.2021         202108
## 5:             D27       215      COMMELEC         12.06.2021         202106
## 6:             D19       129      RESIEXPE         10.03.2020         202003
##    contract_account business_partner consumption_unit
## 1:       2036580530         11554227          312.000
## 2:       2001730675         10079719         1880.000
## 3:       2015990046         10039230          151.000
## 4:       2037243529         10852787         4664.000
## 5:       2010349750         10373470          305.000
## 6:       2014158916         10137123          310.486
```

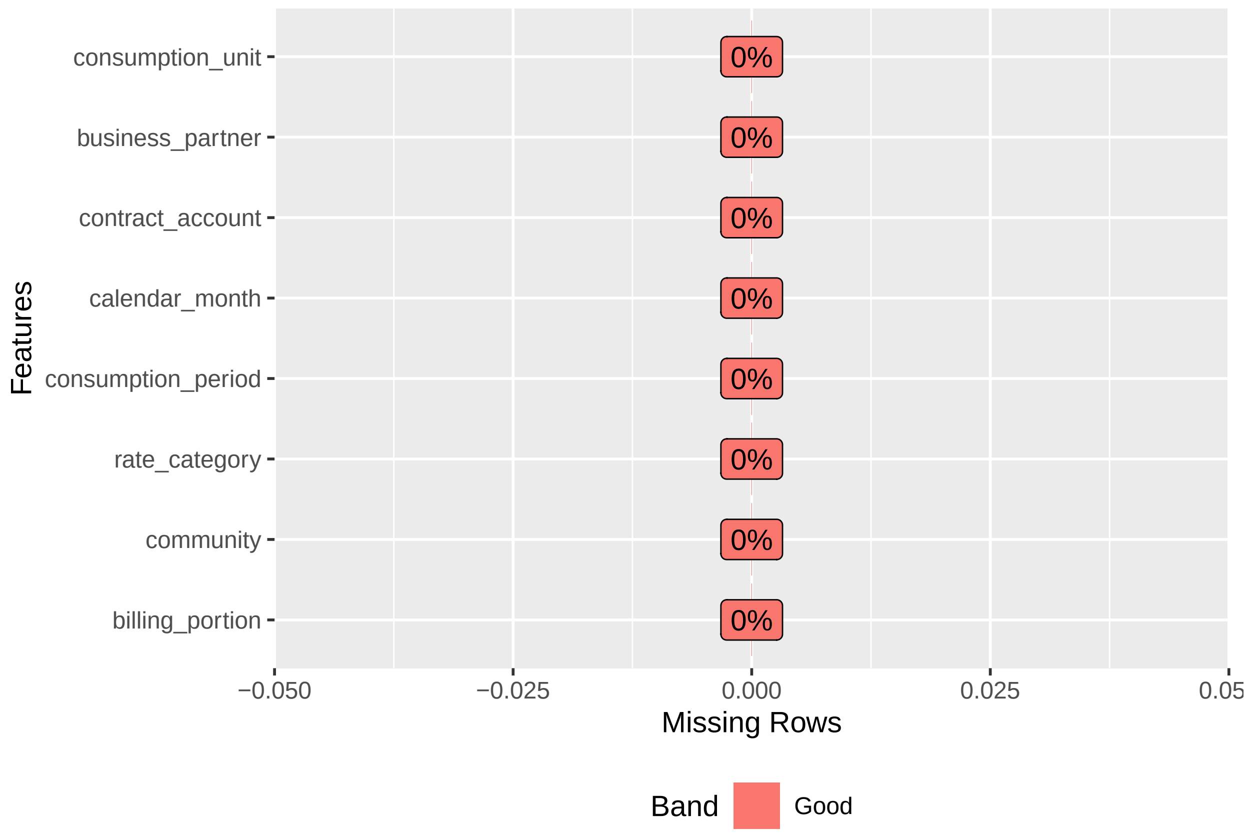
```
## [1] 9972    8
```

Data Explorer



Removed NA

Data Explorer After Removing NA



ETL

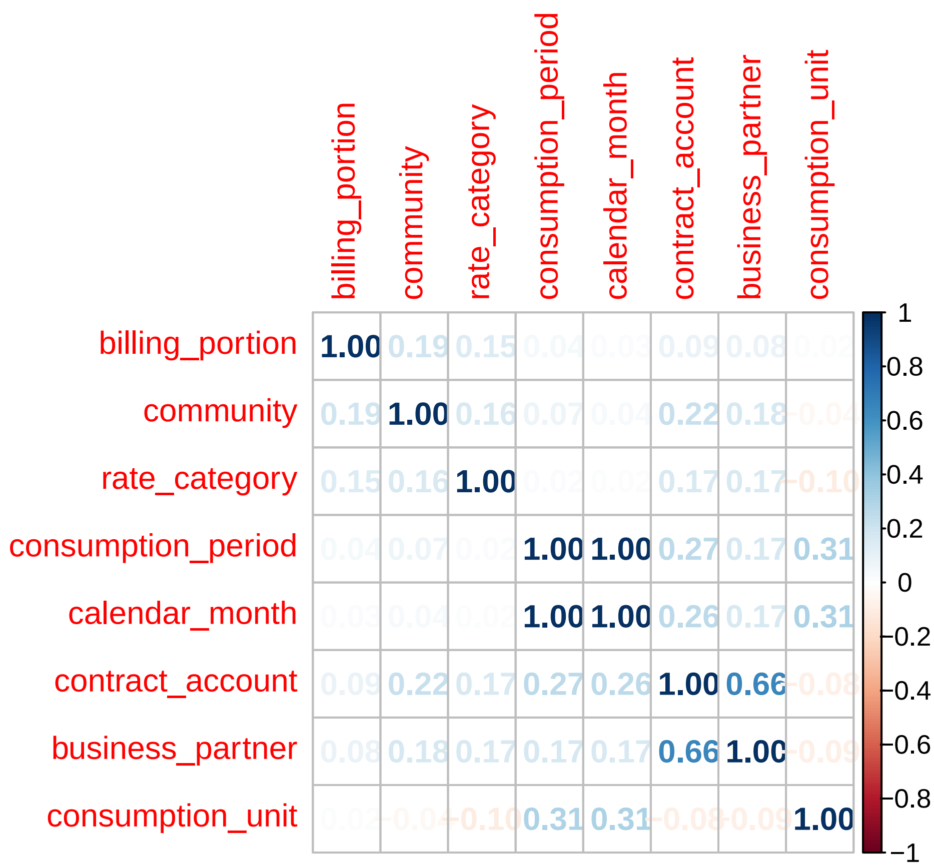
Transformation

You can also embed plots, for example:

Descriptive Analytics

Note COMMELEC: is the commercial. RESIEXPE: is a combined complex with residential and commercial examples being the RESINATE: is tower based residential building, basically a skyscraper that have studio or residential rooms. FREENRESIE: is multi-residential area like Emaar community or DAMAC Akoya as Empower, and DEWA are the provider, which a townhouse style. RESINATSE: is the semi resident basically a villa style house community not Emaar or DAMAC which is two or four houses attached together GOVTELEC: is the government. INDTELEC: is the industrial.

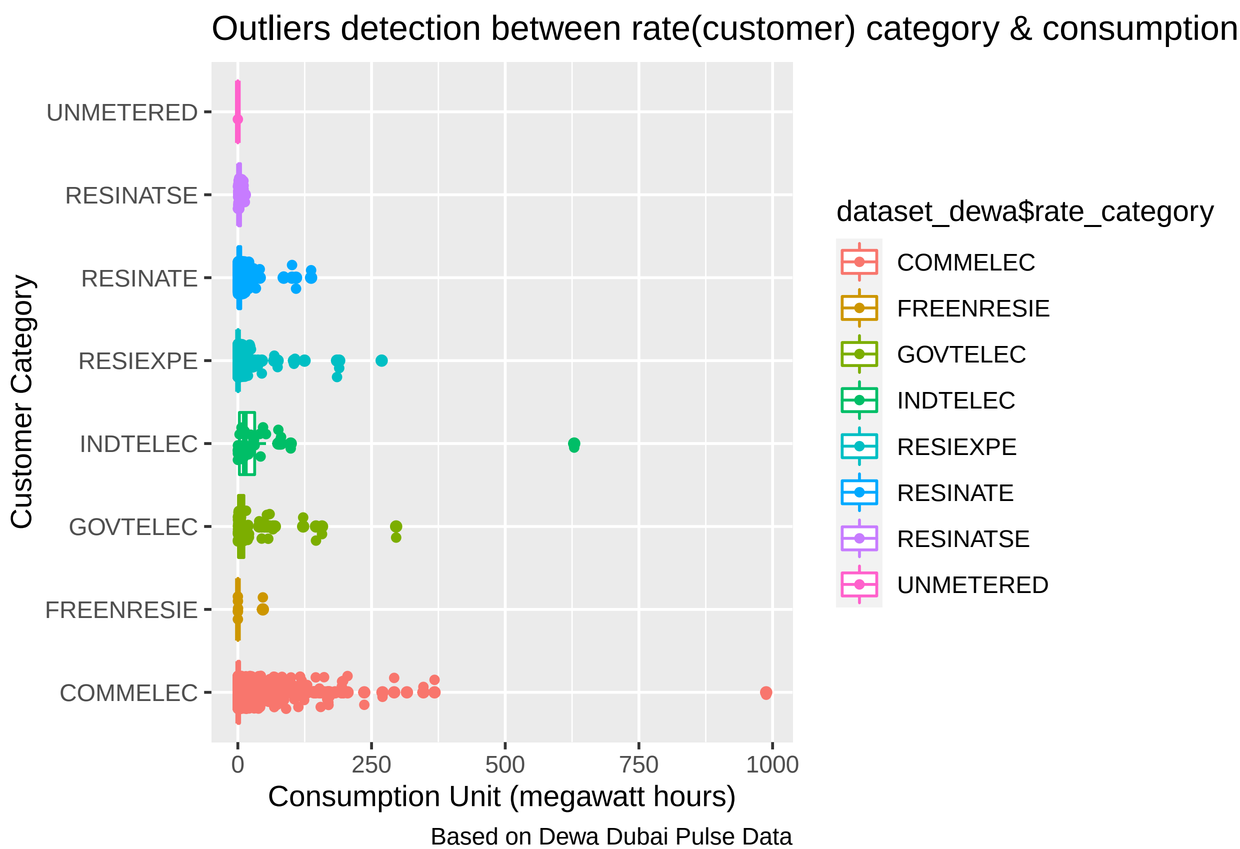
Correlation Analysis



Findings

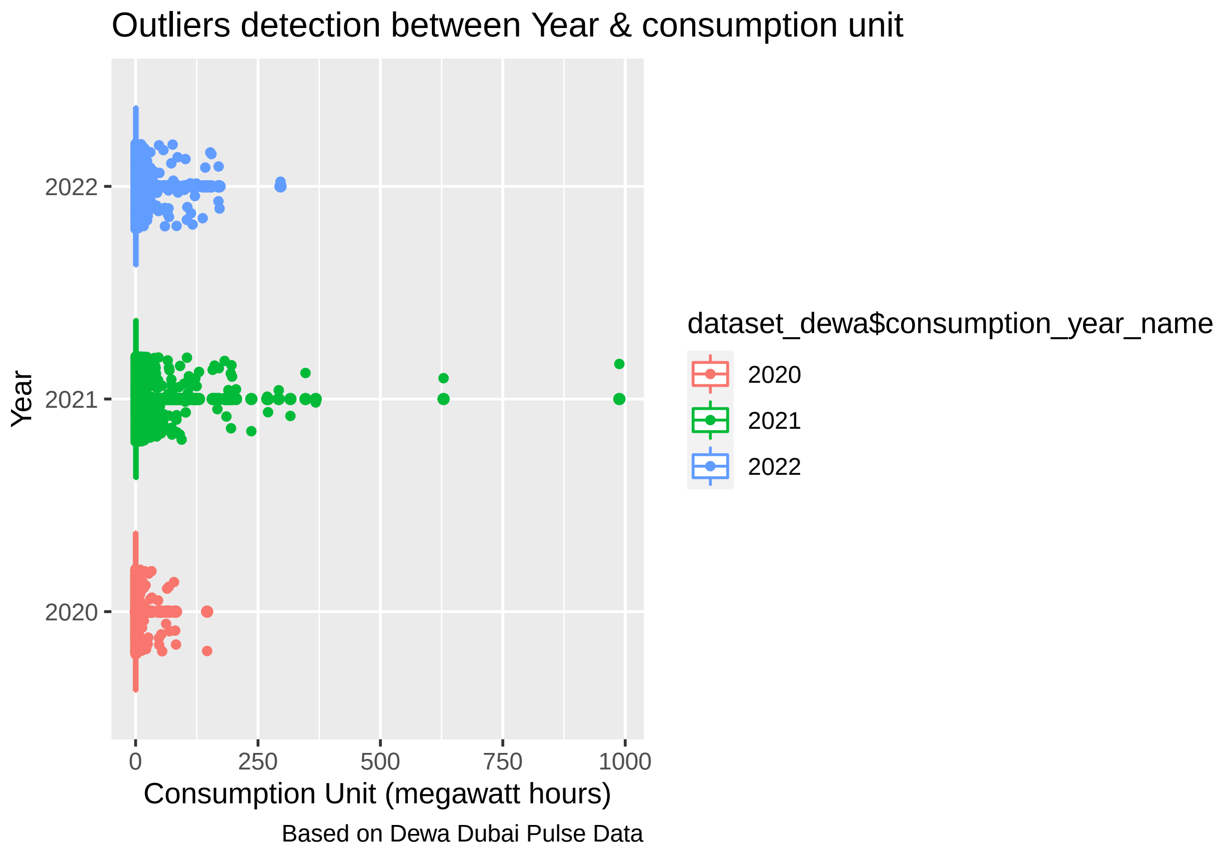
As it is being observed that there is a stong correlation between the Consumption unit and the calendar of the month, thereby it was also found the consumption unit and the calennder of the month in addition the business partner and the contract number (account), thereby it was also found that the community and the contract number have some kind of relation.

Outliers Detection



Findings • As the visual plot indicates that there not many outliers within all kinds of residential category whom are considered the first TWO categories: RESINATSE at 1800MW fully on DEWA and FREENRESIE at 800-900 MW due to different supplier dependence such as Empower, TAQA however the commercial 16750MW based consumers does have more outliers are it is known that usually the commercial consumers such as stores, plaza's, etc are known to operate true power from the daytime till midnight and usually the power generation are on lower voltage adder during the night time as the demand is less, the turbines are known to operation at rpms, thereby the power cost per night, similarly the one government facility 2700MW which is known as Dubai police & the UAE Armed force, as they are operational at 24/7, also what's interesting is the RESIEXPE which is the residential combined complex with stores 6800 MW, etc, an example of that would be festival city since it does have similar consumption to the commercial complex since it works in a similar way, same it goes with the Tower based residential complex 4000 MW since they act in similar way since most towers do have studio, workstations in addition to residential apartments with stores, however the industrial surprisingly emerged with 1400 MW meaning it's little less then expected a new knowledge gained as it seems for the 3 years period due to 2020 effect the production were in the lower side thereby it is was also known that these giant tend to have some support internally through in-house generation • It is seen that 2021 has most of the outlier due to high change in the events which would be discussed even more in milestone 3, already I know that cause based on the python segregation the Industrial had a boom on that period near the end of 2021.

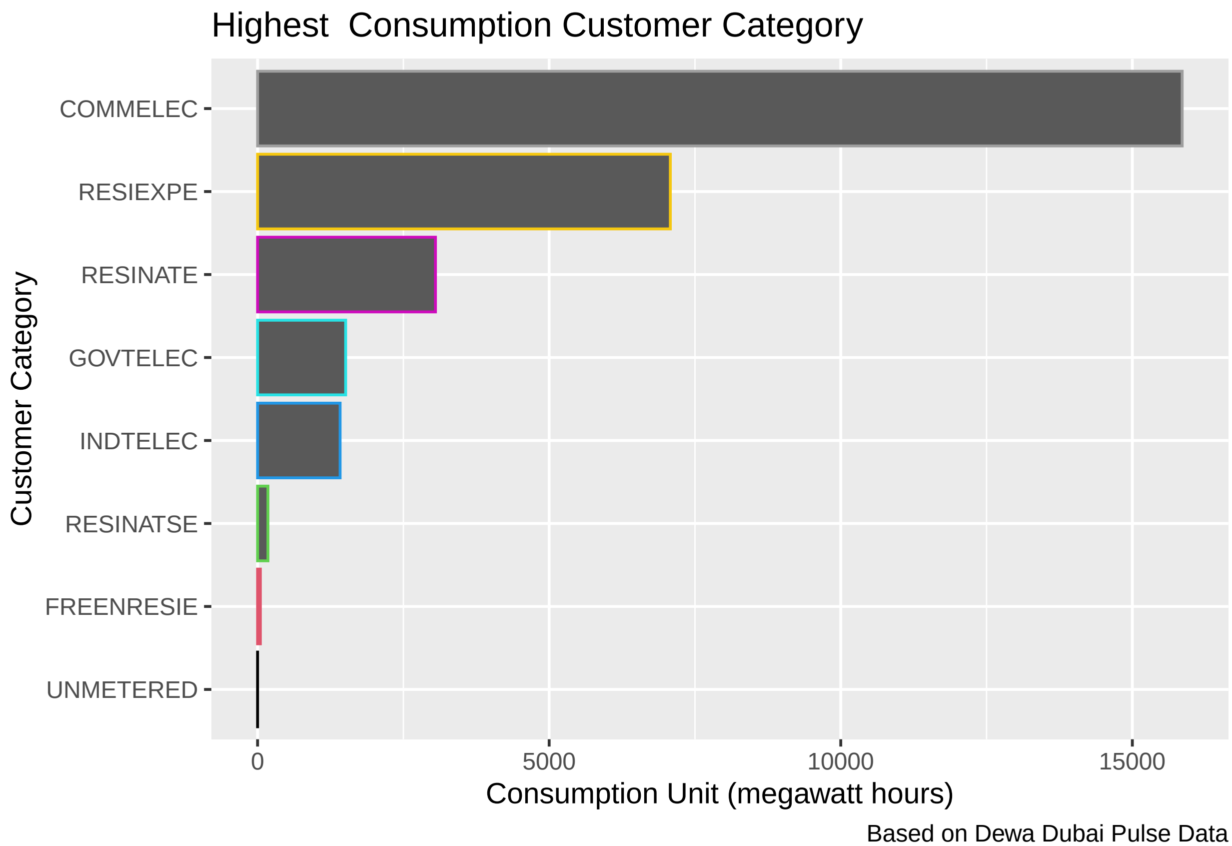
Boxplot Consumption



Findings

• The consumption graph illustrates that during mid-end of 2020 the electrical power consumption was much less then that of 2021, 2022 • It is known that DEWA were installing new metering unit for the transformers which were from Tokia, Areva and Alstom smart meter as according to Mr Al Tayer said that "The new grid covers the generation, transmission, and distribution systems with investments of up to AED 7 billion that will be completed in the short, medium, and long-term until 2035". • 2021 consumption rate was the highest among the rest, near end of 2021 we can observe that tremendous jump in the consumption levels were due to high increase within the industrial sector specially one of the largest power consumers of Dubai such as EGA, DP World. Additionally, the global aluminum price was on continuous raise, in addition to a good reduction in Covid-19 cases and the return of the demand in the commercial sector specially the community-based market & other workplaces which are considered part of the commercial.

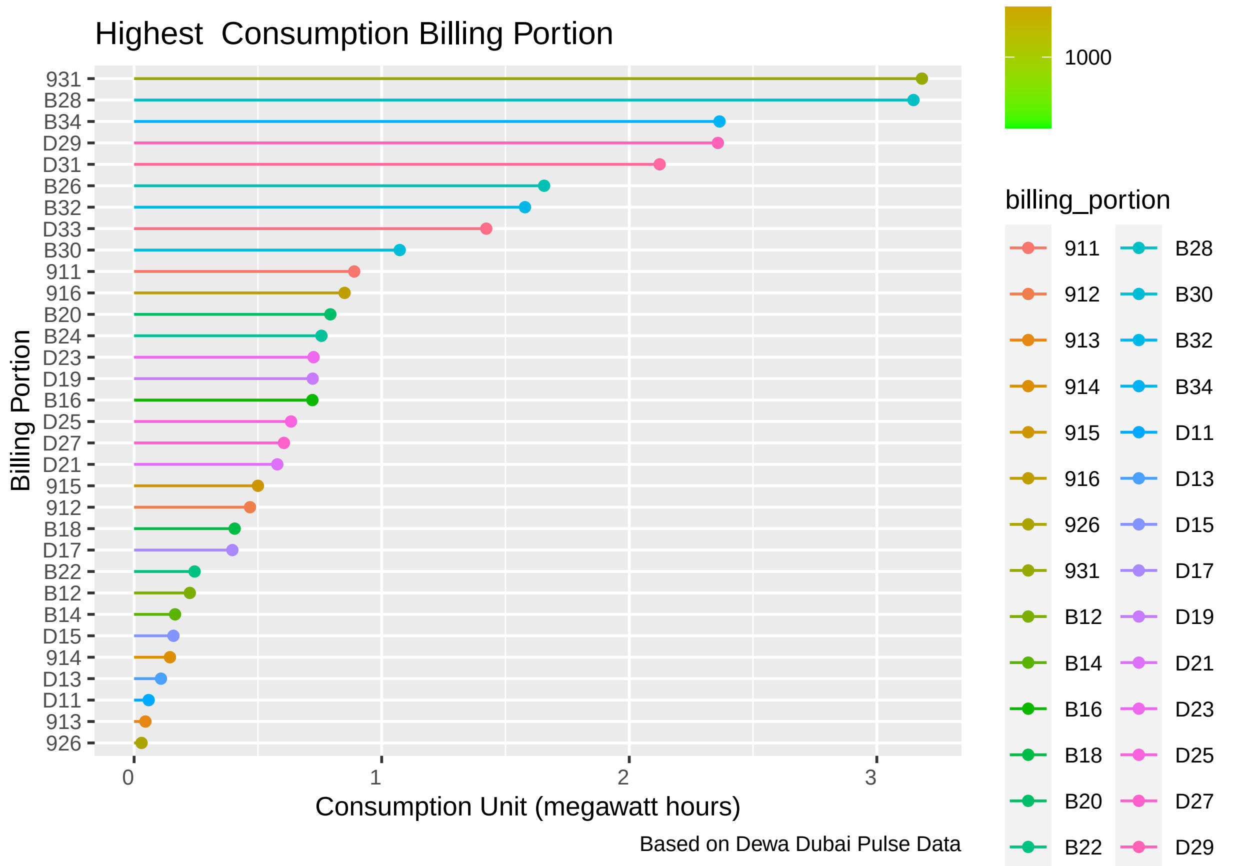
Highest consumption Customer Category



Findings

- As the visual plot indicates that there not many outliers within all kinds of residential category whom are considered the first TWO categories: RESINATSE at 1800MW fully on DEWA and FREENRESIE at 800-900 MW due to different supplier dependence such as Empower, TAQA however the commercial 16750MW based consumers does have more outliers are it is known that usually the commercial consumers such as stores, plaza's, etc are known to operate true power from the daytime till midnight and usually the power generation are on lower voltage adder during the night time as the demand is less, the turbines are known to operation at rpms, thereby the power cost per night, similarly the one government facility 2700MW which is known as Dubai police & the UAE Armed force, as they are operational at 24/7, also what's interesting is the RESIEXPE which is the residential combined complex with stores 6800 MW, etc, an example of that would be festival city since it does have similar consumption to the commercial complex since it works in a similar way, same it goes with the Tower based residential complex 4000 MW since they act in similar way since most towers do have studio, workstations in addition to residential apartments with stores, however the industrial surprisingly emerged with 1400 MW meaning it's little less then expected a new knowledge gained as it seems for the 3 years period due to 2020 effect the production were in the lower side thereby it is was also known that these giant tend to have some support internally through in-house generation • It is seen that 2021 has most of the outlier due to high change in the events which would be discussed even more in milestone 3, already I know that cause based on the r segregation the Industrial had a boom on that period near the end of 2021.

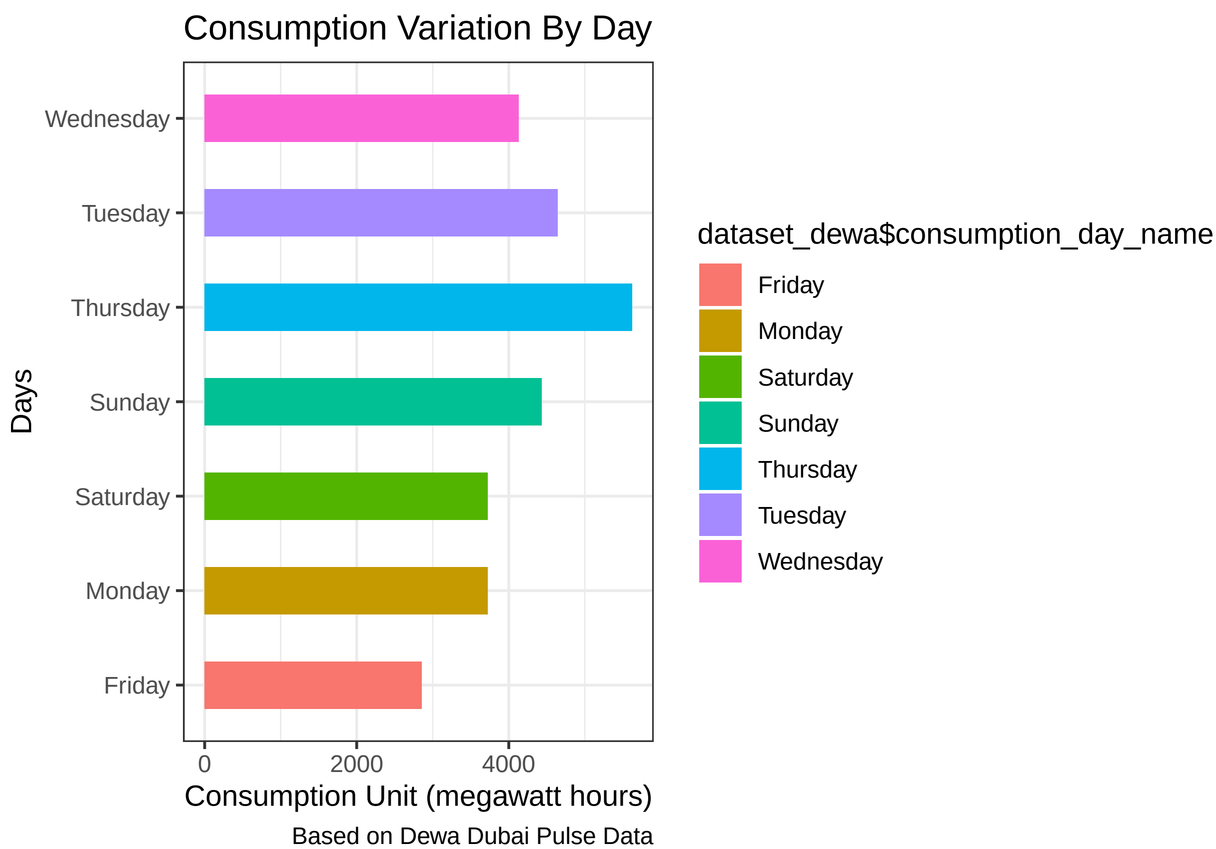
Highest Consumption Billing Portion



Findings • 931 at 4400MW, D31 standing at 2300MW and finally B34 at 2150MW as the billing remains anonymous, this is just for further information, note: since I already know that 931 is DXB airport in Al-Twar 3 region, which seems reasonable since DXB is a commercial and operating at 24/7.

Time Series Analysis

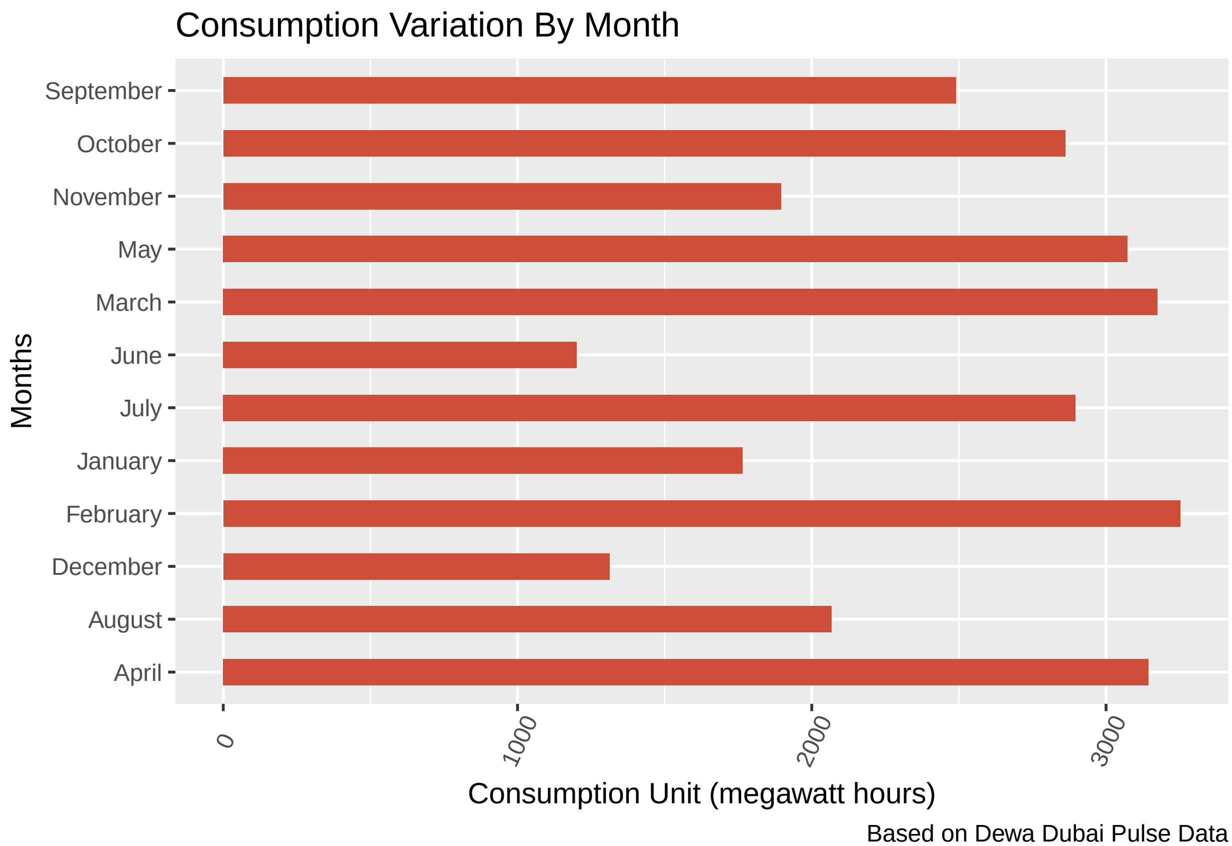
By Day



Findings

- It shows that the near start day of the week Tuesday 6200MW which has the highest consumption well that's for all the three years combined however let's take into the account the change of the weekends that has occurred in 2022 which changed the holidays from Friday-Sat to Sat-Sun, so it's considered as mid working of the week is the busiest and most electrical power consumed time, the second highest goes to Wednesday 5350MW again since this is a three years data the change would occur there by again it is the mid work day of the week, meanwhile Sun 5200MW which was considered as the first day of the week holds the third highest consumption period as it is.

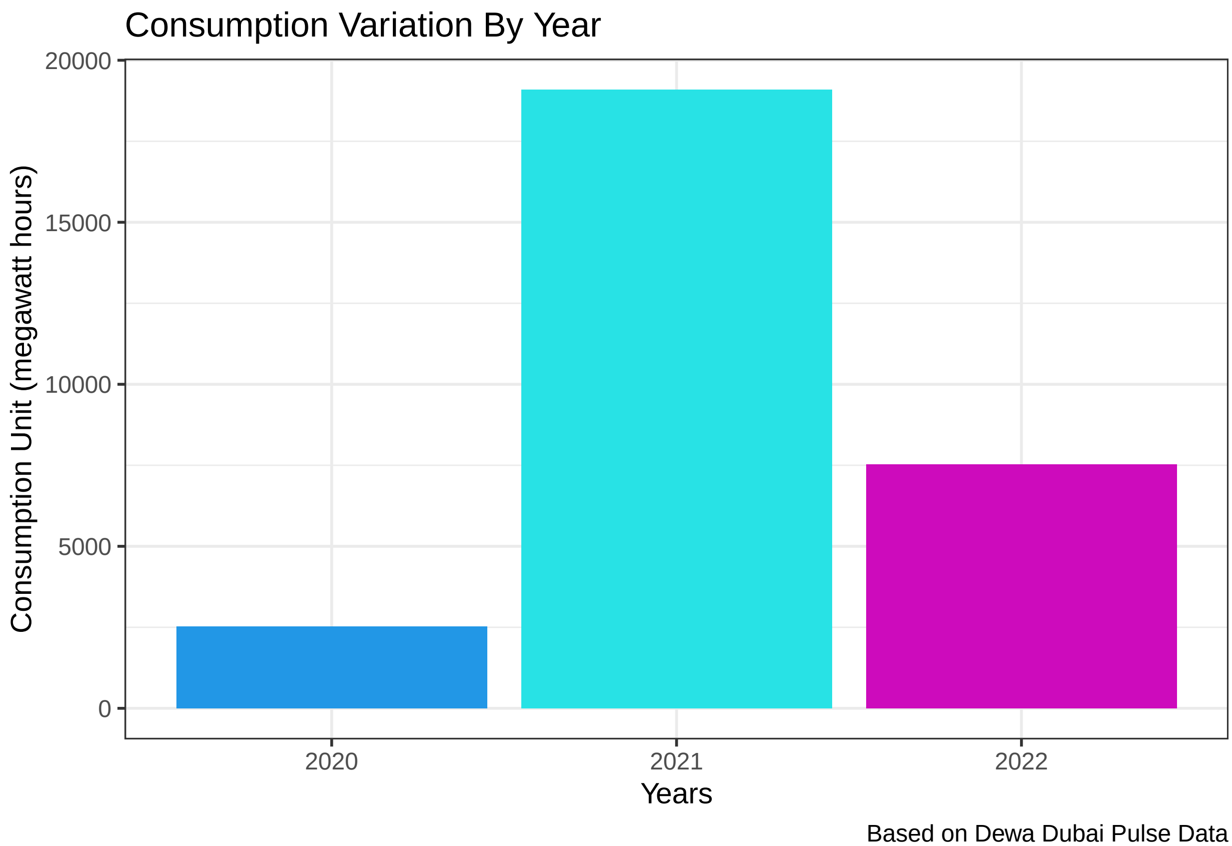
By Month



Findings

- It is clear that May 4200MW, Feb 3300MW and Oct with Sep (3199-3250MW) are the largest time of the consumption period for 3 year data, since what the graph does is that it add the months all together of all the three years, May was the largest since it seems that usually on this time in addition to the beginning of the summer is considered to start on 3-5 (Mar-May) months of year, moreover similar to march data since again the beginning of the summer takes place during that period in addition to but expect that usually people tend to stay within the country rather then travel abroad, since it is known that the majority of the travels takes place during the mid-summer time 5-8 (Mar-Jul) months of year.

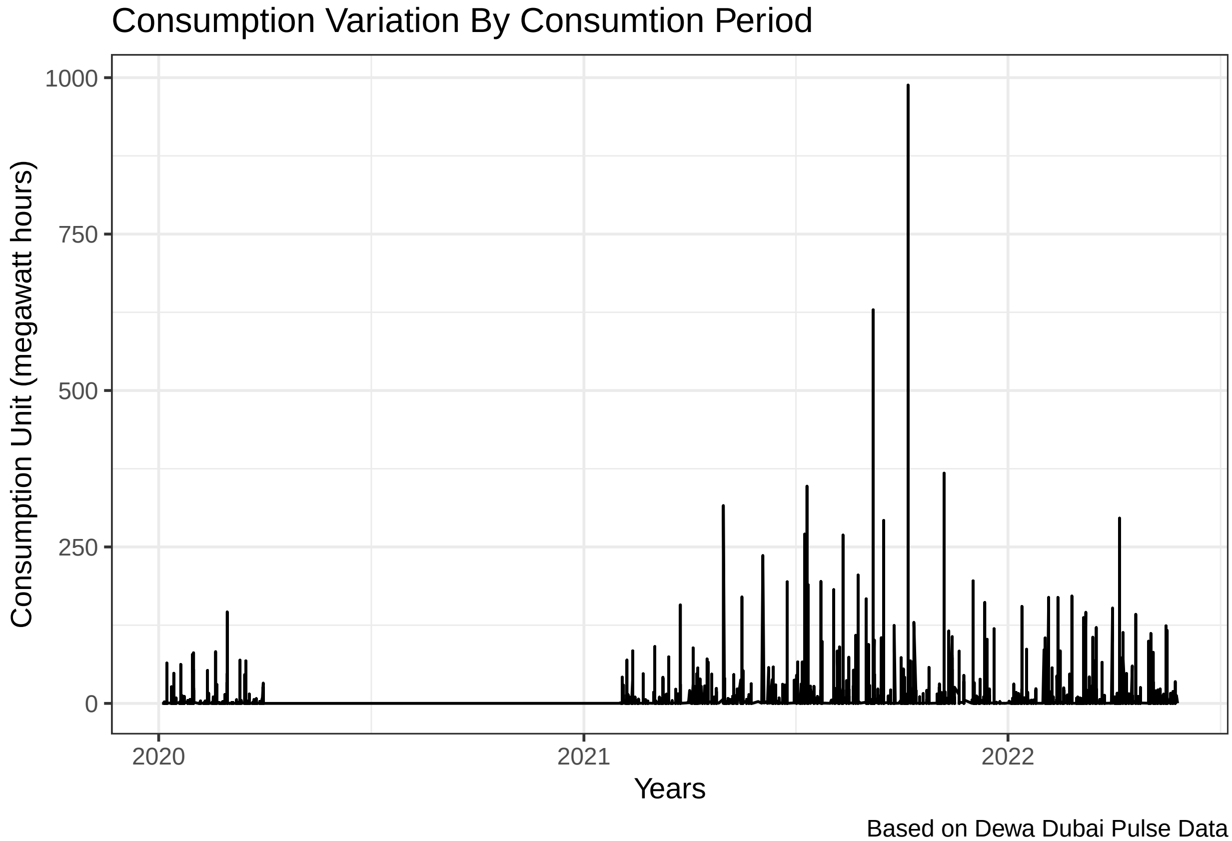
By Year



Findings

- 2021 tops all and second is 2022 meanwhile 2020 remains as the last for sensor change reasons as mentioned in the Consumption Variation by Consumption Period.

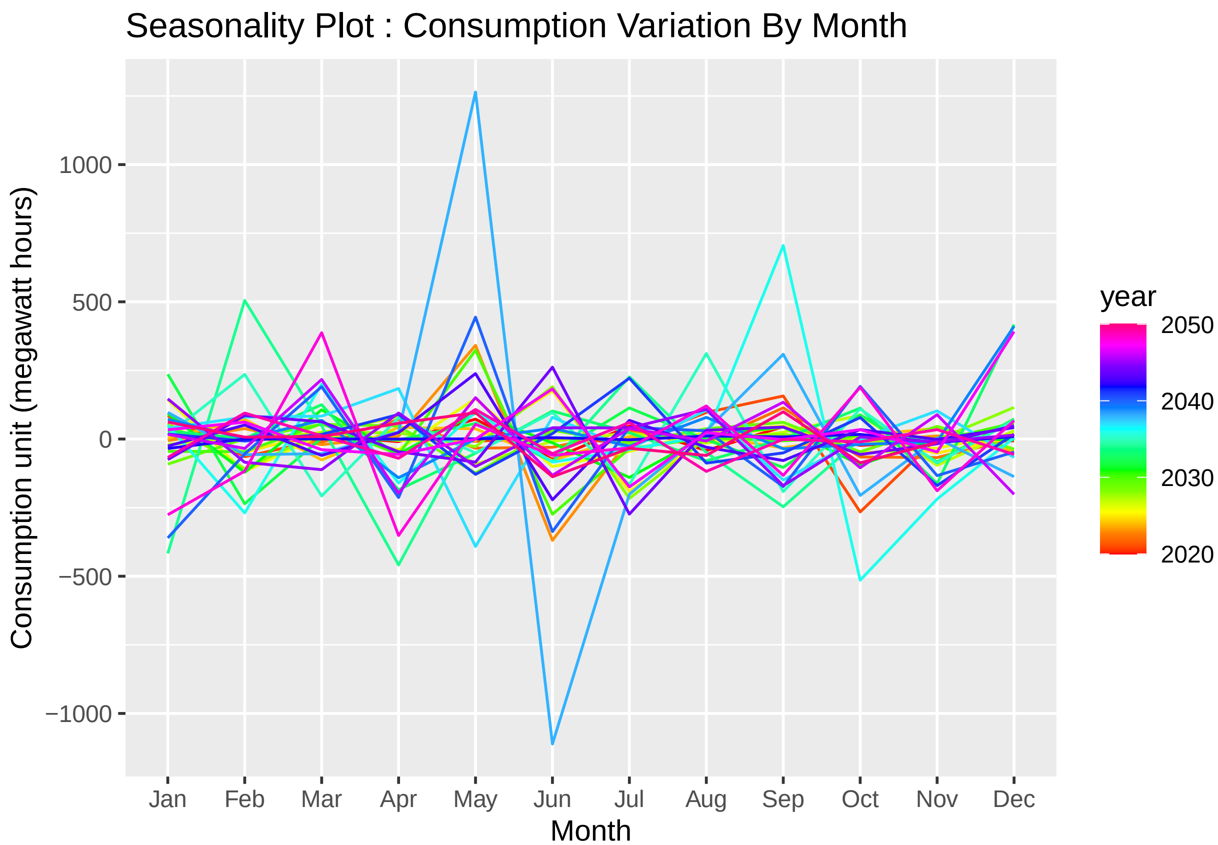
By Consumption Period

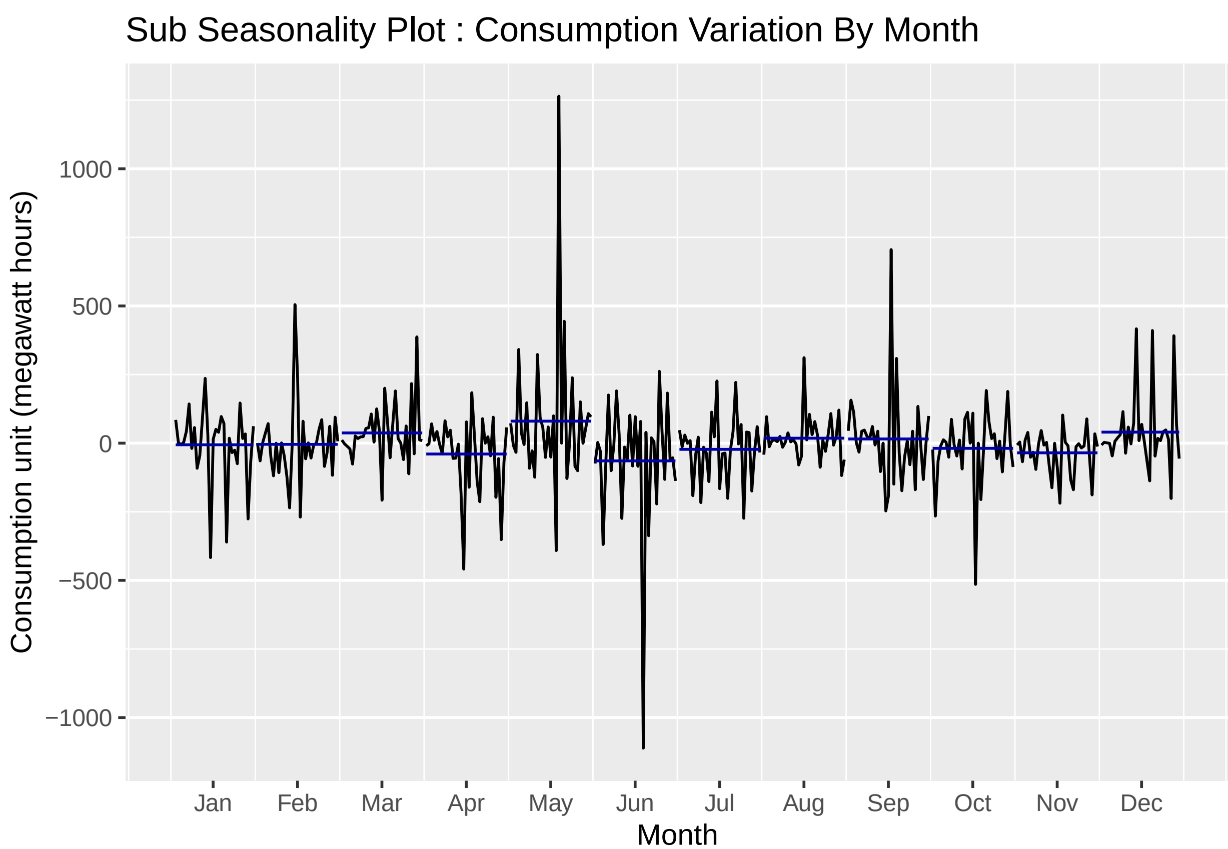


Findings

- The consumption graph illustrates that during mid-end of 2020 the electrical power consumption was much less then that of 2021, 2022 • It is known that DEWA were installing new metering unit for the transformers which were from Tokia, Areva and Alstom smart meter as according to Mr Al Tayer said that "The new grid covers the generation, transmission, and distribution systems with investments of up to AED 7 billion that will be completed in the short, medium, and long-term until 2035". • 2021 consumption rate was the highest among the rest, near end of 2021 we can observe that tremendous jump in the consumption levels were due to high increase within the industrial sector specially one of the largest power consumers of Dubai such as EGA, DP World. Additionally, the global aluminum price was on continuous raise, in addition to a good reduction in Covid-19 cases and the return of the demand in the commercial sector specially the community-based market & other workplaces which are considered part of the commercial.

Seasonal Plot



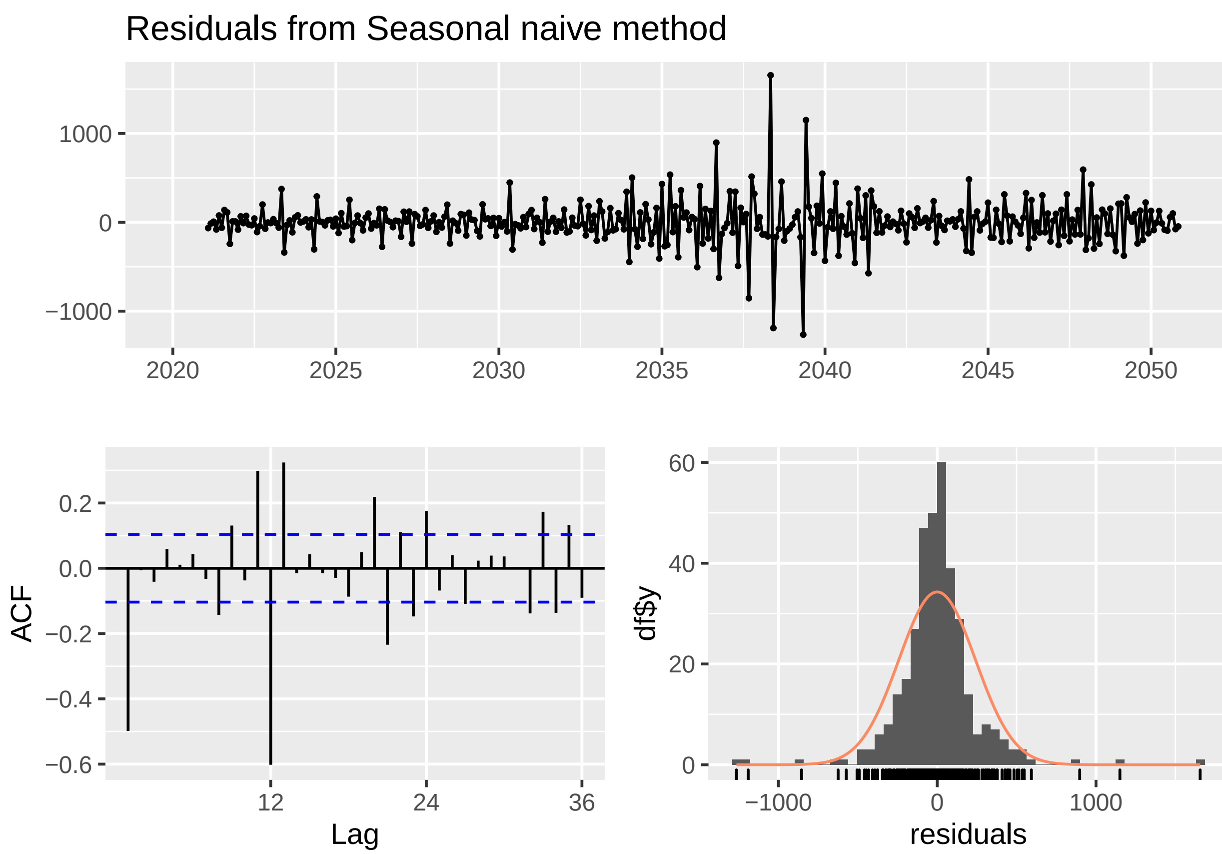


Predictive Analytics

Time Series Prediction

Bench Mark

```
##
## Forecast method: Seasonal naive method
##
## Model Information:
## Call: snaive(y = DF)
##
## Residual sd: 241.3946
##
## Error measures:
##                      ME     RMSE      MAE  MPE MAPE MASE       ACF1
## Training set -0.4366572 241.3946 153.9816 -Inf  Inf    1 -0.4983757
##
## Forecasts:
##          Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
## Dec 2050        -56.281 -365.6406 253.0786 -529.4057 416.8437
## Jan 2051         61.898 -247.4616 371.2576 -411.2267 535.0227
## Feb 2051          6.908 -302.4516 316.2676 -466.2167 480.0327
## Mar 2051          9.032 -300.3276 318.3916 -464.0927 482.1567
## Apr 2051         57.537 -251.8226 366.8966 -415.5877 530.6617
## May 2051         95.835 -213.5246 405.1946 -377.2897 568.9597
## Jun 2051       -137.644 -447.0036 171.7156 -610.7687 335.4807
## Jul 2051        -33.725 -343.0846 275.6346 -506.8497 439.3997
## Aug 2051        -59.932 -369.2916 249.4276 -533.0567 413.1927
## Sep 2051         99.076 -210.2836 408.4356 -374.0487 572.2007
## Oct 2051        -86.833 -396.1926 222.5266 -559.9577 386.2917
## Nov 2051        -12.226 -321.5856 297.1336 -485.3507 460.8987
## Dec 2051        -56.281 -493.7815 381.2195 -725.3803 612.8183
## Jan 2052         61.898 -375.6025 499.3985 -607.2013 730.9973
## Feb 2052          6.908 -430.5925 444.4085 -662.1913 676.0073
## Mar 2052          9.032 -428.4685 446.5325 -660.0673 678.1313
## Apr 2052         57.537 -379.9635 495.0375 -611.5623 726.6363
## May 2052         95.835 -341.6655 533.3355 -573.2643 764.9343
## Jun 2052       -137.644 -575.1445 299.8565 -806.7433 531.4553
## Jul 2052        -33.725 -471.2255 403.7755 -702.8243 635.3743
## Aug 2052        -59.932 -497.4325 377.5685 -729.0313 609.1673
## Sep 2052         99.076 -338.4245 536.5765 -570.0233 768.1753
## Oct 2052        -86.833 -524.3335 350.6675 -755.9323 582.2663
## Nov 2052        -12.226 -449.7265 425.2745 -681.3253 656.8733
```

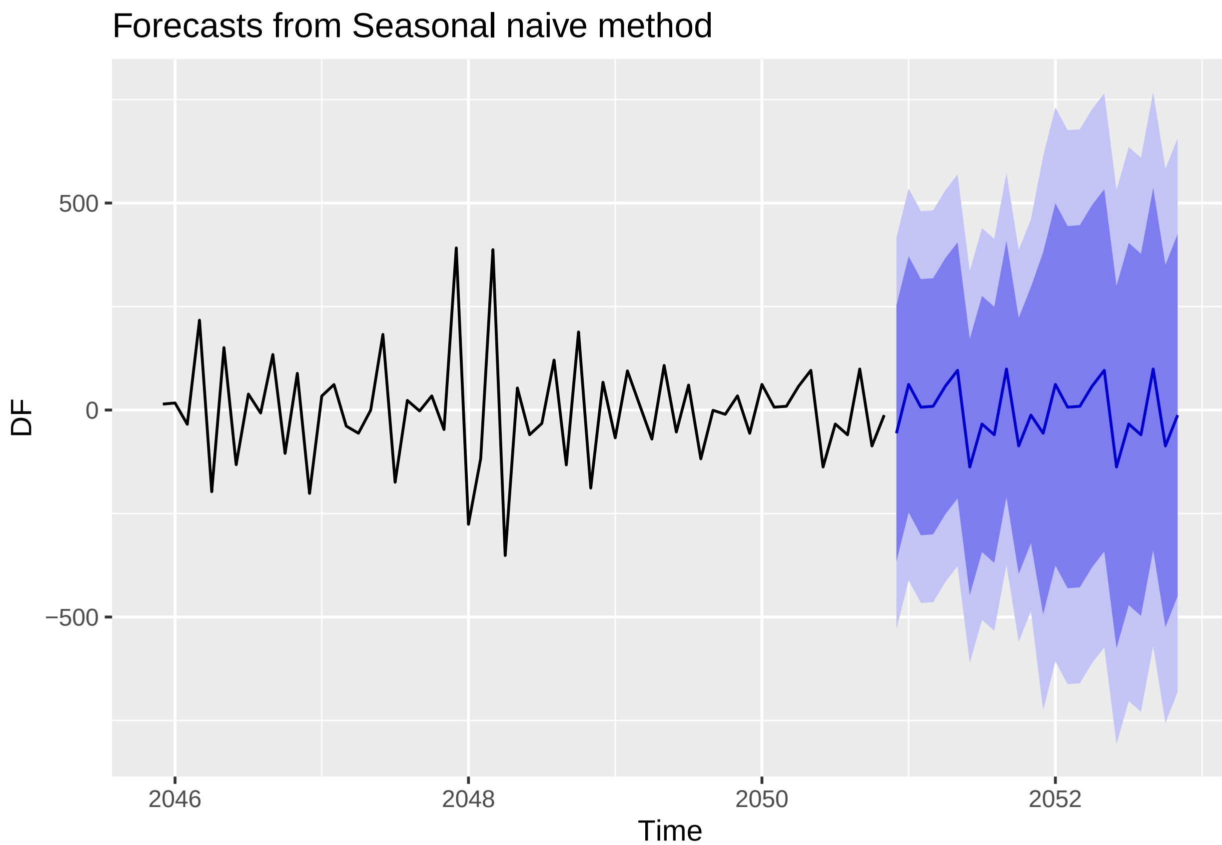



```
##
##  Ljung-Box test
##
## data:  Residuals from Seasonal naive method
## Q* = 383.45, df = 24, p-value < 2.2e-16
##
## Model df: 0.   Total lags used: 24
```

ETS

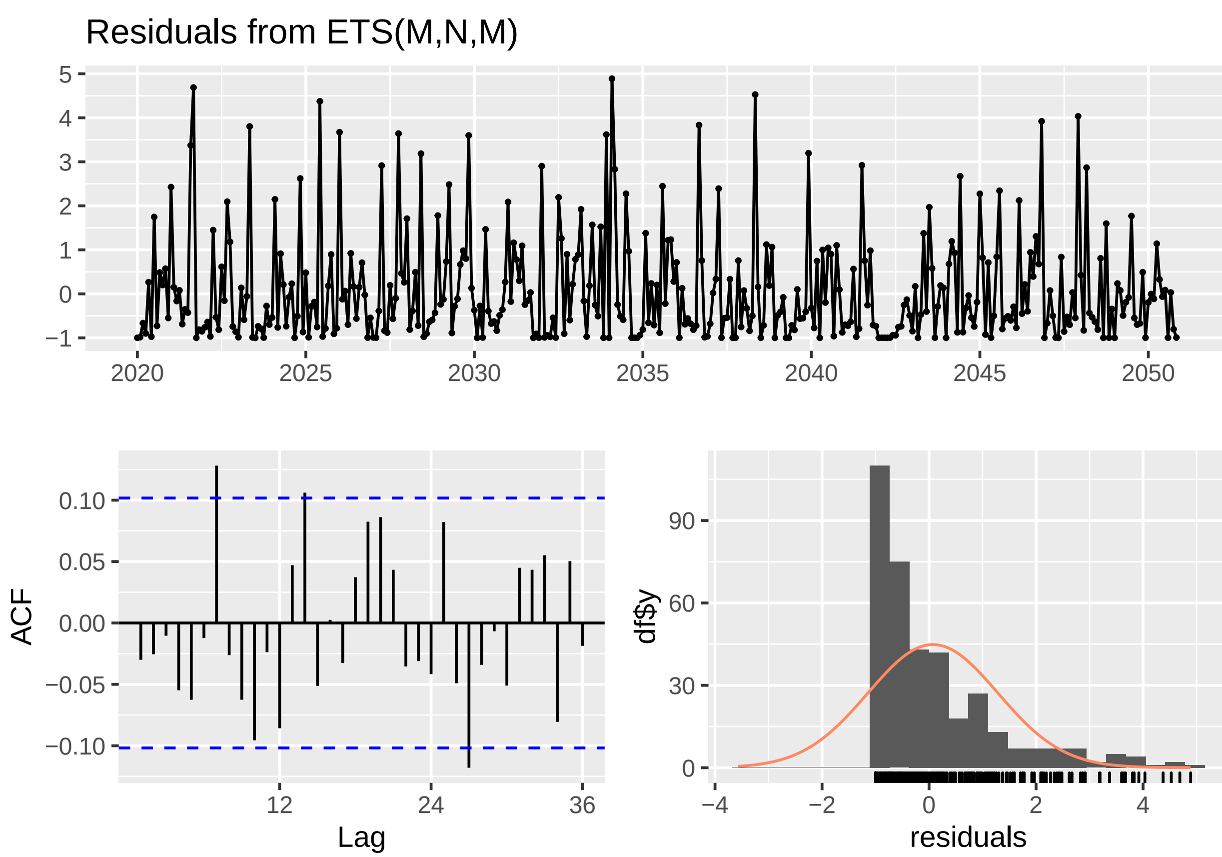
```
## ETS(M,N,M)
##
## Call:
##  ets(y = Y)
##
##   Smoothing parameters:
##     alpha = 0.0662
##     gamma = 1e-04
##
##   Initial states:
##     l = 36.2676
##     s = 1.077 0.4592 0.7891 1.2015 0.7835 0.6491
##            1.1707 2.0803 0.8133 1.1223 0.8837 0.9702
##
##   sigma:  1.239
```

```
##
##      AIC     AICc      BIC
## 5474.423 5475.775 5533.166
##
## Training set error measures:
##                      ME     RMSE      MAE      MPE     MAPE     MASE
## Training set -0.4165797 112.4293 69.52928 -9206.96 9241.209 0.690823
##                     ACF1
## Training set 0.007852389
```



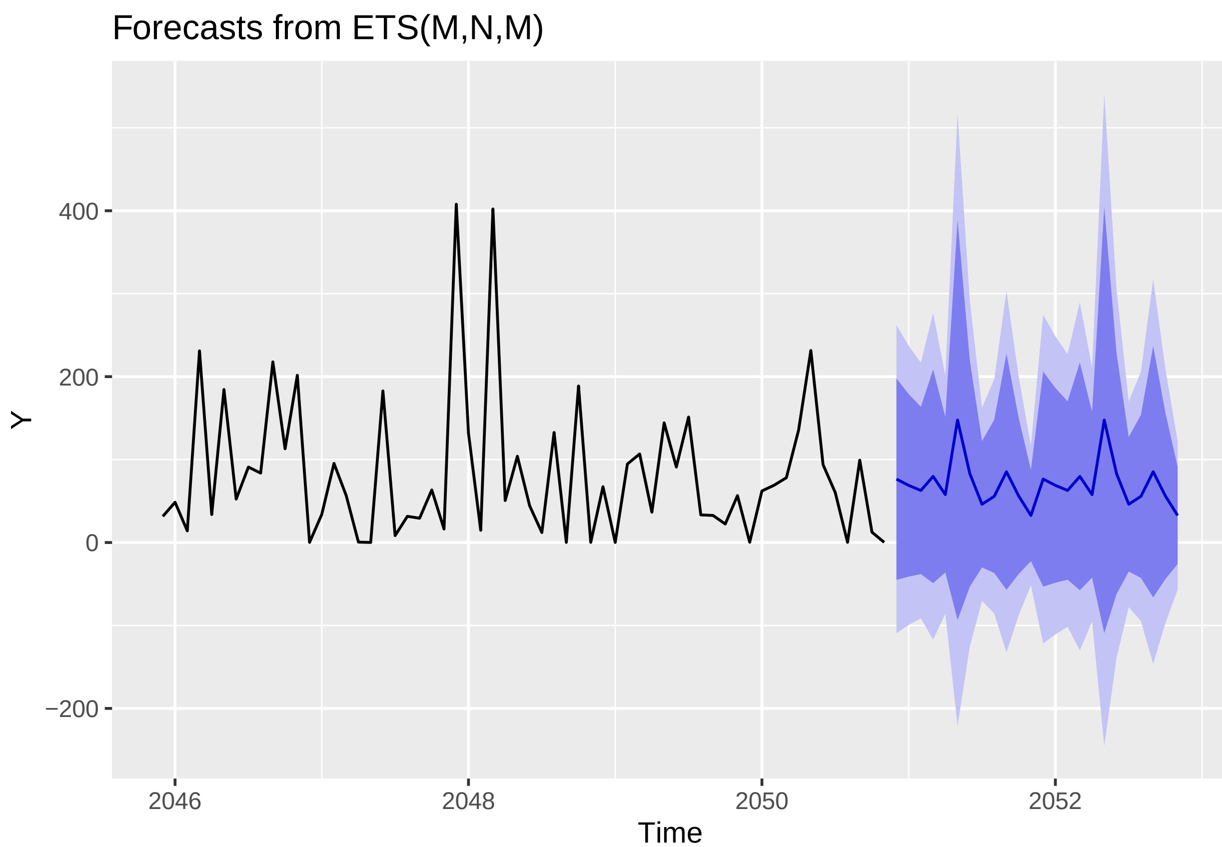

```
##
## Forecast method: Seasonal naive method
##
## Model Information:
## Call: snaive(y = DF)
##
## Residual sd: 241.3946
##
## Error measures:
##                      ME     RMSE      MAE  MPE MAPE MASE       ACF1
## Training set -0.4366572 241.3946 153.9816 -Inf  Inf    1 -0.4983757
##
## Forecasts:
##          Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
## Dec 2050        -56.281 -365.6406 253.0786 -529.4057 416.8437
```

```
## Jan 2051         61.898 -247.4616 371.2576 -411.2267 535.0227
## Feb 2051          6.908 -302.4516 316.2676 -466.2167 480.0327
## Mar 2051          9.032 -300.3276 318.3916 -464.0927 482.1567
## Apr 2051         57.537 -251.8226 366.8966 -415.5877 530.6617
## May 2051         95.835 -213.5246 405.1946 -377.2897 568.9597
## Jun 2051       -137.644 -447.0036 171.7156 -610.7687 335.4807
## Jul 2051        -33.725 -343.0846 275.6346 -506.8497 439.3997
## Aug 2051        -59.932 -369.2916 249.4276 -533.0567 413.1927
## Sep 2051         99.076 -210.2836 408.4356 -374.0487 572.2007
## Oct 2051        -86.833 -396.1926 222.5266 -559.9577 386.2917
## Nov 2051        -12.226 -321.5856 297.1336 -485.3507 460.8987
## Dec 2051        -56.281 -493.7815 381.2195 -725.3803 612.8183
## Jan 2052         61.898 -375.6025 499.3985 -607.2013 730.9973
## Feb 2052          6.908 -430.5925 444.4085 -662.1913 676.0073
## Mar 2052          9.032 -428.4685 446.5325 -660.0673 678.1313
## Apr 2052         57.537 -379.9635 495.0375 -611.5623 726.6363
## May 2052         95.835 -341.6655 533.3355 -573.2643 764.9343
## Jun 2052       -137.644 -575.1445 299.8565 -806.7433 531.4553
## Jul 2052        -33.725 -471.2255 403.7755 -702.8243 635.3743
## Aug 2052        -59.932 -497.4325 377.5685 -729.0313 609.1673
## Sep 2052         99.076 -338.4245 536.5765 -570.0233 768.1753
## Oct 2052        -86.833 -524.3335 350.6675 -755.9323 582.2663
## Nov 2052        -12.226 -449.7265 425.2745 -681.3253 656.8733
```




```
##
##  Ljung-Box test
```

```
##
## data:  Residuals from ETS(M,N,M)
## Q* = 32.958, df = 10, p-value = 0.0002769
##
## Model df: 14.   Total lags used: 24
```




```
##
## Forecast method: ETS(M,N,M)
##
## Model Information:
## ETS(M,N,M)
##
## Call:
##  ets(y = Y)
##
##   Smoothing parameters:
##     alpha = 0.0662
##     gamma = 1e-04
##
##   Initial states:
##     l = 36.2676
##     s = 1.077 0.4592 0.7891 1.2015 0.7835 0.6491
##            1.1707 2.0803 0.8133 1.1223 0.8837 0.9702
##
##   sigma:  1.239
```

```
##
##      AIC     AICc      BIC
## 5474.423 5475.775 5533.166
##
## Error measures:
##                      ME     RMSE      MAE      MPE     MAPE     MASE
## Training set -0.4165797 112.4293 69.52928 -9206.96 9241.209 0.690823
##                     ACF1
## Training set 0.007852389
##
## Forecasts:
##          Point Forecast      Lo 80     Hi 80      Lo 95    Hi 95
## Dec 2050       76.48722  -44.96425 197.93870 -109.25678 262.2312
## Jan 2051       68.93871  -41.13371 179.01113  -99.40253 237.2800
## Feb 2051       62.76260  -38.00187 163.52706  -91.34336 216.8685
## Mar 2051       79.74267  -48.98687 208.47221 -117.13217 276.6175
## Apr 2051       57.78636  -36.00950 151.58222  -85.66202 201.2347
## May 2051      147.74886  -93.37680 388.87451 -221.02101 516.5187
## Jun 2051       83.15413  -53.28984 219.59809 -125.51891 291.8272
## Jul 2051       46.13118  -29.97269 122.23506  -70.25964 162.5220
## Aug 2051       55.68188  -36.67265 148.03641  -85.56218 196.9259
## Sep 2051       85.35561  -56.97529 227.68651 -132.32072 303.0319
## Oct 2051       56.07974  -37.93301 150.09249  -87.70036 199.8598
## Nov 2051       32.62197  -22.35687  87.60082  -51.46091 116.7049
## Dec 2051       76.48800  -53.10489 206.08089 -121.70722 274.6832
## Jan 2052       68.93941  -48.48064 186.35946 -110.63906 248.5179
## Feb 2052       62.76323  -44.69957 170.22604 -101.58694 227.1134
## Mar 2052       79.74348  -57.50819 216.99515 -130.16484 289.6518
## Apr 2052       57.78695  -42.19311 157.76701  -95.11936 210.6933
## May 2052      147.75036 -109.20936 404.71007 -245.23562 540.7363
## Jun 2052       83.15497  -62.21322 228.52317 -139.16650 305.4764
## Jul 2052       46.13165  -34.93025 127.19356  -77.84183 170.1051
## Aug 2052       55.68245  -42.66537 154.03026  -94.72755 206.0925
## Sep 2052       85.35648  -66.17527 236.88823 -146.39133 317.1043
## Oct 2052       56.08031  -43.98662 156.14724  -96.95885 209.1195
## Nov 2052       32.62231  -25.88366  91.12827  -56.85485 122.0995
```

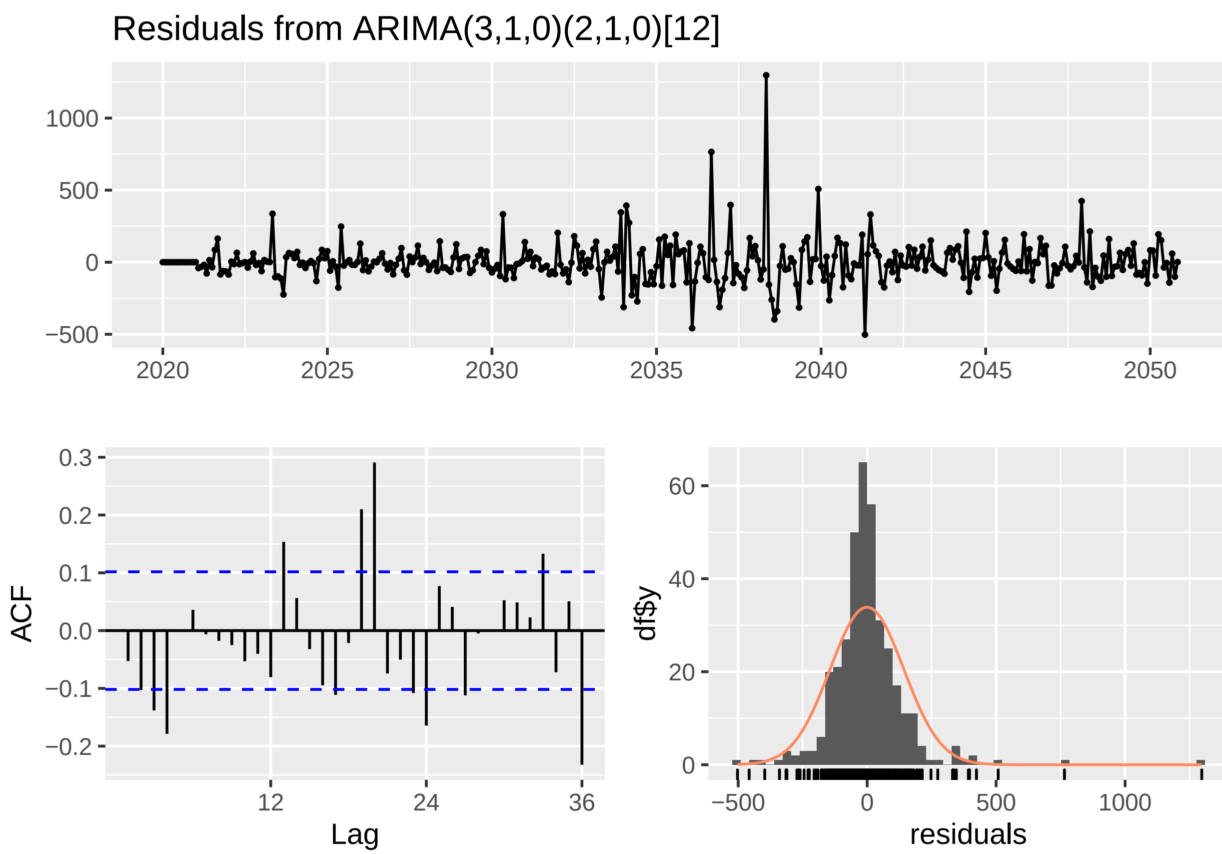
ARIMA Model

```
##
##  ARIMA(0,1,0)(0,1,0)[12]                    : 4946.257
##  ARIMA(0,1,0)(0,1,1)[12]                    : Inf
##  ARIMA(0,1,0)(0,1,2)[12]                    : Inf
##  ARIMA(0,1,0)(1,1,0)[12]                    : 4790.477
##  ARIMA(0,1,0)(1,1,1)[12]                    : Inf
##  ARIMA(0,1,0)(1,1,2)[12]                    : Inf
##  ARIMA(0,1,0)(2,1,0)[12]                    : 4758.55
##  ARIMA(0,1,0)(2,1,1)[12]                    : Inf
##  ARIMA(0,1,0)(2,1,2)[12]                    : Inf
##  ARIMA(0,1,1)(0,1,0)[12]                    : Inf
##  ARIMA(0,1,1)(0,1,1)[12]                    : Inf
##  ARIMA(0,1,1)(0,1,2)[12]                    : Inf
##  ARIMA(0,1,1)(1,1,0)[12]                    : Inf
```

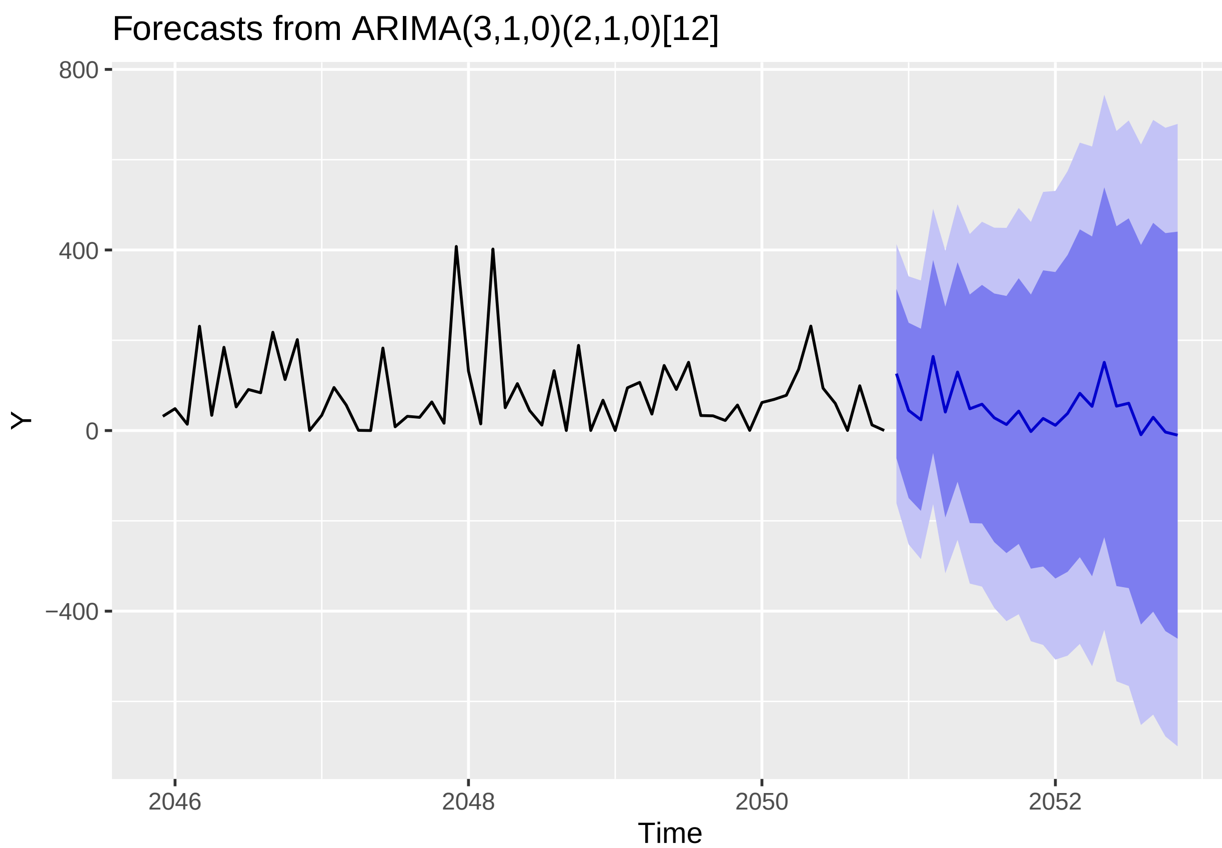
```
##  ARIMA(0,1,1)(1,1,1)[12]                    : Inf
##  ARIMA(0,1,1)(1,1,2)[12]                    : Inf
##  ARIMA(0,1,1)(2,1,0)[12]                    : Inf
##  ARIMA(0,1,1)(2,1,1)[12]                    : Inf
##  ARIMA(0,1,1)(2,1,2)[12]                    : Inf
##  ARIMA(0,1,2)(0,1,0)[12]                    : Inf
##  ARIMA(0,1,2)(0,1,1)[12]                    : Inf
##  ARIMA(0,1,2)(0,1,2)[12]                    : Inf
##  ARIMA(0,1,2)(1,1,0)[12]                    : Inf
##  ARIMA(0,1,2)(1,1,1)[12]                    : Inf
##  ARIMA(0,1,2)(1,1,2)[12]                    : Inf
##  ARIMA(0,1,2)(2,1,0)[12]                    : Inf
##  ARIMA(0,1,2)(2,1,1)[12]                    : Inf
##  ARIMA(0,1,3)(0,1,0)[12]                    : Inf
##  ARIMA(0,1,3)(0,1,1)[12]                    : Inf
##  ARIMA(0,1,3)(0,1,2)[12]                    : Inf
##  ARIMA(0,1,3)(1,1,0)[12]                    : Inf
##  ARIMA(0,1,3)(1,1,1)[12]                    : Inf
##  ARIMA(0,1,3)(2,1,0)[12]                    : Inf
##  ARIMA(0,1,4)(0,1,0)[12]                    : Inf
##  ARIMA(0,1,4)(0,1,1)[12]                    : Inf
##  ARIMA(0,1,4)(1,1,0)[12]                    : Inf
##  ARIMA(0,1,5)(0,1,0)[12]                    : Inf
##  ARIMA(1,1,0)(0,1,0)[12]                    : 4846.313
##  ARIMA(1,1,0)(0,1,1)[12]                    : Inf
##  ARIMA(1,1,0)(0,1,2)[12]                    : Inf
##  ARIMA(1,1,0)(1,1,0)[12]                    : 4699.748
##  ARIMA(1,1,0)(1,1,1)[12]                    : Inf
##  ARIMA(1,1,0)(1,1,2)[12]                    : Inf
##  ARIMA(1,1,0)(2,1,0)[12]                    : 4667.047
##  ARIMA(1,1,0)(2,1,1)[12]                    : Inf
##  ARIMA(1,1,0)(2,1,2)[12]                    : Inf
##  ARIMA(1,1,1)(0,1,0)[12]                    : Inf
##  ARIMA(1,1,1)(0,1,1)[12]                    : Inf
##  ARIMA(1,1,1)(0,1,2)[12]                    : Inf
##  ARIMA(1,1,1)(1,1,0)[12]                    : Inf
##  ARIMA(1,1,1)(1,1,1)[12]                    : Inf
##  ARIMA(1,1,1)(1,1,2)[12]                    : Inf
##  ARIMA(1,1,1)(2,1,0)[12]                    : Inf
##  ARIMA(1,1,1)(2,1,1)[12]                    : Inf
##  ARIMA(1,1,2)(0,1,0)[12]                    : Inf
##  ARIMA(1,1,2)(0,1,1)[12]                    : Inf
##  ARIMA(1,1,2)(0,1,2)[12]                    : Inf
##  ARIMA(1,1,2)(1,1,0)[12]                    : Inf
##  ARIMA(1,1,2)(1,1,1)[12]                    : Inf
##  ARIMA(1,1,2)(2,1,0)[12]                    : Inf
##  ARIMA(1,1,3)(0,1,0)[12]                    : Inf
##  ARIMA(1,1,3)(0,1,1)[12]                    : Inf
##  ARIMA(1,1,3)(1,1,0)[12]                    : Inf
##  ARIMA(1,1,4)(0,1,0)[12]                    : Inf
##  ARIMA(2,1,0)(0,1,0)[12]                    : 4804.906
##  ARIMA(2,1,0)(0,1,1)[12]                    : Inf
##  ARIMA(2,1,0)(0,1,2)[12]                    : Inf
##  ARIMA(2,1,0)(1,1,0)[12]                    : 4647.71
```

```
##  ARIMA(2,1,0)(1,1,1)[12]                    : Inf
##  ARIMA(2,1,0)(1,1,2)[12]                    : Inf
##  ARIMA(2,1,0)(2,1,0)[12]                    : 4624.672
##  ARIMA(2,1,0)(2,1,1)[12]                    : Inf
##  ARIMA(2,1,1)(0,1,0)[12]                    : Inf
##  ARIMA(2,1,1)(0,1,1)[12]                    : Inf
##  ARIMA(2,1,1)(0,1,2)[12]                    : Inf
##  ARIMA(2,1,1)(1,1,0)[12]                    : Inf
##  ARIMA(2,1,1)(1,1,1)[12]                    : Inf
##  ARIMA(2,1,1)(2,1,0)[12]                    : Inf
##  ARIMA(2,1,2)(0,1,0)[12]                    : Inf
##  ARIMA(2,1,2)(0,1,1)[12]                    : Inf
##  ARIMA(2,1,2)(1,1,0)[12]                    : Inf
##  ARIMA(2,1,3)(0,1,0)[12]                    : Inf
##  ARIMA(3,1,0)(0,1,0)[12]                    : 4768.111
##  ARIMA(3,1,0)(0,1,1)[12]                    : Inf
##  ARIMA(3,1,0)(0,1,2)[12]                    : Inf
##  ARIMA(3,1,0)(1,1,0)[12]                    : 4632.245
##  ARIMA(3,1,0)(1,1,1)[12]                    : Inf
##  ARIMA(3,1,0)(2,1,0)[12]                    : 4600.928
##  ARIMA(3,1,1)(0,1,0)[12]                    : Inf
##  ARIMA(3,1,1)(0,1,1)[12]                    : Inf
##  ARIMA(3,1,1)(1,1,0)[12]                    : Inf
##  ARIMA(3,1,2)(0,1,0)[12]                    : Inf
##  ARIMA(4,1,0)(0,1,0)[12]                    : 4751.133
##  ARIMA(4,1,0)(0,1,1)[12]                    : Inf
##  ARIMA(4,1,0)(1,1,0)[12]                    : 4624.507
##  ARIMA(4,1,1)(0,1,0)[12]                    : Inf
##  ARIMA(5,1,0)(0,1,0)[12]                    : 4745.375
##
##
##
##  Best model: ARIMA(3,1,0)(2,1,0)[12]
```

```
## Series: Y
## ARIMA(3,1,0)(2,1,0)[12]
##
## Coefficients:
##           ar1      ar2      ar3     sar1     sar2
##       -0.7344  -0.5107  -0.2729  -0.7373  -0.3050
## s.e.   0.0509   0.0579   0.0527   0.0502   0.0511
##
## sigma^2 = 21425:  log likelihood = -2294.34
## AIC=4600.69   AICc=4600.93   BIC=4623.97
##
## Training set error measures:
##                      ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -0.8036084 142.7776 88.56394 -12073.29 12779.19 0.8799459
##                     ACF1
## Training set -0.05265921
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,1,0)(2,1,0)[12]
## Q* = 117.84, df = 19, p-value = 3.331e-16
##
## Model df: 5.   Total lags used: 24
```




```
##
## Forecast method: ARIMA(3,1,0)(2,1,0)[12]
##
## Model Information:
## Series: Y
## ARIMA(3,1,0)(2,1,0)[12]
##
## Coefficients:
##           ar1      ar2      ar3     sar1     sar2
##       -0.7344  -0.5107  -0.2729  -0.7373  -0.3050
## s.e.   0.0509   0.0579   0.0527   0.0502   0.0511
##
## sigma^2 = 21425:  log likelihood = -2294.34
## AIC=4600.69   AICc=4600.93   BIC=4623.97
##
## Error measures:
##                      ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -0.8036084 142.7776 88.56394 -12073.29 12779.19 0.8799459
##                     ACF1
## Training set -0.05265921
##
## Forecasts:
##          Point Forecast       Lo 80    Hi 80     Lo 95    Hi 95
## Dec 2050     126.077754  -61.50626 313.6618 -160.8072 412.9627
## Jan 2051      44.736161 -149.34960 238.8219 -252.0924 341.5647
## Feb 2051      23.875740 -177.90745 225.6589 -284.7250 332.4765
```

```
## Mar 2051     164.244065  -49.47382 377.9619 -162.6092 491.0974
## Apr 2051      40.991175 -192.51871 274.5011 -316.1314 398.1137
## May 2051     129.729503 -113.33012 372.7891 -241.9981 501.4571
## Jun 2051      48.258212 -204.89643 301.4128 -338.9084 435.4248
## Jul 2051      58.459401 -205.70826 322.6271 -345.5502 462.4690
## Aug 2051      28.315778 -247.01268 303.6442 -392.7628 449.3943
## Sep 2051      13.427978 -271.34604 298.2020 -422.0963 448.9522
## Oct 2051      43.164136 -251.05275 337.3810 -406.8018 493.1300
## Nov 2051      -2.221523 -305.83757 301.3945 -466.5622 462.1191
## Dec 2051      26.901996 -301.13365 354.9376 -474.7852 528.5892
## Jan 2052      11.721420 -327.70283 351.1457 -507.3832 530.8260
## Feb 2052      38.038902 -312.84594 388.9237 -498.5932 574.6710
## Mar 2052      82.586277 -280.44471 445.6173 -472.6217 637.7943
## Apr 2052      53.633298 -322.72139 429.9880 -521.9515 629.2181
## May 2052     151.209121 -236.25971 538.6780 -441.3733 743.7916
## Jun 2052      54.071672 -344.42338 452.5667 -555.3739 663.5173
## Jul 2052      60.573831 -348.93873 470.0864 -565.7216 686.8693
## Aug 2052      -9.243964 -429.60442 411.1165 -652.1298 633.6419
## Sep 2052      29.475444 -401.17722 460.1281 -629.1510 688.1018
## Oct 2052      -3.397058 -444.17296 437.3788 -677.5056 670.7115
## Nov 2052     -10.158503 -460.88651 440.5695 -699.4875 679.1705
```

Findings

In this predictive analysis we selected three algorithm to predict the DEWA energy consumption. snaive, ets and ARIMA. To finding goodness of fit or accuracy for all these time series algorithm we have a concept of auto correlation plot, lags and residual plots.

Autocorrelation plot gives an idea how data trends are best fit time series model, where "lags" within the plot describes the stationary and un stationary data points. as per above observation we can conclude that "ARIMA" model have 99% best fit the time series data all the data points are in between the regression lines.

Further more residual plot also way where we can how the model is best fitted in model, but in our case there is each model have almost similar residual plot, from above plot we can find that frequency & density of residual are almost zero. means error between regression line and training data is fitted well.

Other factors AIC (Akaike's Information Criteria) and BIC means (Bayesian Information Criteria), lowest AIC and BIC values of ARIMA model interpret that selection of ARIMA is best of our time series data.

snaive model only consider previous value to predict , seasonality and multivariate is in appropriate with snaive so one of the major reason of rejecting snaive is limitations.

ets method random variation can not handle very well, seasonality and cyclic behavior of data is major but ets can not handle trends very well thats impact on the accuracy of non stationary data.

At the end we have ARIMA model that is AR (Auto regression) I (Integrator) MA (Moving Average), which deals all the trends and seasonality in the data, also
mutlivariate can handle very accurately in ARIMA models.

In our study, we tried multiple variable ARIMA to check the variation in AIC & BIC value. Following are the test for the ARIMA model.

1. Test 1 Porovide the whole data set.

AIC=7502.52 AICc=5409.89 BIC=5623.89

2. Test 2

Removed billing portion & contract account

AIC=6109.39 AICc=5409.63 BIC=5832.77

3. Test 3

Community, Rate Category and business partner.

AIC=5109.39 AICc=5109.63 BIC=5132.77

AIC=5109.39 AICc=5109.63 BIC=5132.77 result is best of out of other two.

Solutions Of Questions

The relationship between the consumption unit and period is a low-medium correlation however when adding all the three years data, as the correlation is standing at 0.29 %, of course it doesn't seems like a tremendous relation however when taken the data of 2020 out from the combined set, then definitely the correlation will increase to 0.33 %, overall there is a correlation but it really depends on the event more rather then the period itself as the period of the year doesn't behave the same on yearly bases, the only natural cause of the change is the weather change (summer & winter), which is an indication that the events matter more and would increase the correlation if suppose let's say an reoccurring man made event is happening routinely. the rate category is affecting the consumption negatively if taking all the three year together, in addition due to higher variance within the category for such as community-wise of 347 in comparison to a another community but with the same category rate, in this example both were commercial, the variance between the two is very high even when taking only the high consumers of the commercial industry, even if the years were removed, the consumption unit and the variant of the consumer seems to be unrelated, thus it is recommended for DEWA to introduce a new procedure to classify the category rate rather then generalizing, it is better the breakup the category in sub-parts, additionally there is a clear difference when breaking them into parts. the predictor models that were produced seem to be relatively close in-terms of ETS and ARIMA meanwhile the seasonal naïve has a huge difference even when perfecting and fining the tune, ETS and ARIMA becomes even closer together then they were. the largest consumers within the area were the 500 series of the community, 500 means Jebel Ali as a whole, the largest among them were 531 industrial, 516 commercial, 594 commercial, 500 industrial and lastly the 599 industrial since Jebel Ali is a known destination for warehouses of goods, plants and factories, however in terms of pure power consumer would be the government facility of 915 which is Empower distribution center as they are the ones whom take and transformer, connect different bus bars to the suppliers.

Literature Review

It has been demonstrated that certain conventional statistical models have strong energy intake forecasting capabilities and have been widely employed in quasi-time period prediction issues. Besides that, the most recent methodologies, such as the "Autoregressive Integrated Moving Average Model" or ARIMA, the "Autoregressive Moving Average Model" or ARMA, as well as linear regression, are the main topics of this literature review. The "ARMA Model" combines the advantages of the autoregressive and moving average models, making it one of the most frequently implemented suitable models in static random series research. However, the bulk of time series is, in fact, quasi. Dansana et al., (2020) stated that the "ARIMA" model gets around this restriction by using a discretization step. Incorporating both mobile & autoregressive contexts, the suggested technique predicts static and trends in time-series information. In this sense, the researchers examined the "ARIMA" & "ARMA" electric usage designs for domestic usage. Moreover, the information obtained for the different models was done so between 2010 November & 2006 December. The researcher's data revealed that the "ARIMA" model is superior for quarterly & monthly forecasting. The ARMA model is probably suitable for regular and weekly forecasting. The development of energy usage may be predicted using the statistical technique of linear regression analysis, which is focused on causal relationships. The researchers carried out further evaluations in the same study. In a different study, the researchers used hourly and daily data, single and multiple linear regression models and other methods to forecast home energy demand. The accuracy of the predictions is significantly impacted by the temporal granularity of the observable information, as the authors have demonstrated. An increasing number of machine learning approaches are being used to solve problems with those techniques. Studies have suggested using "Artificial Neural Networks" or ANNs to forecast and optimise power consumption. For instance, a stance tool integrates a multi-layer perceptron as well as a two-layer ANN to swiftly predict how much energy will be used in business and learning institutions. They used the information from the energy analysis of one-hundred and fifty-one public infrastructures in four areas of "South Italy" to determine which of the two ANNs had the best architectural design. The first ANN provided the building's real energy efficiency in its recommended design, while the subsequent ANN assessed the best remodelling options. After receiving a lot of data-rich instruction, they both did well. As per the findings of Yang & Wang (2020), in spite of having a strong learning capacity of the ANN models, they frequently result in issues with under as well as overfitting.
The newly developed deep learning technique for forecasting has shown encouraging results in a number of fields. Deep networks have been actively used in recent years to tackle the challenge of predicting energy usage. Abdar et al., (2021) mentioned that a system implements deep learning for predicting the short-term consumption of energy.

Conclusion & Recomendations

The Commercial sector is the highest consumer of electrical power within the state of Dubai, specifically the number 221 which is actually known as Jumeriah city, since most of the shops, commercial market are all focused towards this area and second was community number 354 which is Dubai Jebel Ali community since some of the major commercial warehouses are present within Jebel Ali thus it's the second highest consumer of electrical power, it is seen by the prediction model, it indicates that the commercial is still would be the highest consumer of electrical power and also in the future, previously assumed that the industrial would be the highest, however due to the outcome we can recommend that furthermore 25 kV transformer system would be required in the future and additionally buck-boost converters as some of the bus-bar design within the malls are DC power (underground bus bars) rather than AC to have a unified amperage at first meanwhile then it turns into AC power once it enter the local transformers, more to added some major 531 community power consumer which is Jebel Ali again but the port side as the port side, has some industrial complex's such as factories, plants and so on, thereby it is recommended to build a new transmission line under the ground rather then overhead, which is actually being built as I'm writing this, which will exchange power among the factors rather line a line connection, basically a spider based connection rather then line based, as it would be more efficient, more to added a little increase within the trend of the ARIMA for 2037 indicates a little raise on the power consumption of the industrial sector since that's pretty accurate and it match the reality as new factories are being built, one of them is EGA scrap metal melting plant which will be located in that area, additionally the raise of the in the prediction model of ARIMA indicates that new market & malls to be opened in the future, which seems unlikely unfortunately, however if these entertainment zones got trend it can take this effect as it would attack costumers to these location, more to added is that these location are the first tourist zones meaning the power trend should be consistent, however the prediction has some consistency but with some few changes or error at some points like 2028, additionally it is recommended for DEWA to use this prediction for the power changes during the seasonal times of the year, as it is shown during the middle days of the week, the consumption tends to be higher then usual

References

M. Al Kalban, “Dubai Electricity & Water Authority (DEWA): Dewa completes 100 of the short-term goals of its AED 7 billion smart grid strategy, 28-Feb-2022. [Online]. Available: https://www.dewa.gov.ae/en/about-us/media-publications/latest-news/2022/02/dewa-completes-100-of-the-short-term-goals-of-its-aed-7-billion-smart-grid-strategy. [Accessed: 14-Sep-2022].

T. L. M. Exchange, “LME aluminium: London Metal Exchange,” Lme, 2022. [Online]. Available: https://www.lme.com/en/Metals/Non-ferrous/LME-Aluminium. [Accessed: 14-Sep-2022].

Dawsoon, “Global price of aluminum,” FRED, 24-Aug-2022. [Online]. Available: https://fred.stlouisfed.org/series/PALUMUSDQ. [Accessed: 14-Sep-2022]. H. Ritchie, E. Mathieu, L. Rodés-Guirao, C. Appel, C. Giattino, E. Ortiz-Ospina, J. Hasell, B. Macdonald, D. Beltekian, and M. Roser, “United Arab Emirates: Coronavirus pandemic country profile,” Our World in Data, 05-Mar-2020. [Online]. Available: https://ourworldindata.org/coronavirus/country/united-arab-emirates. [Accessed: 14-Sep-2022].

W. I. O. N. W. Team, “UAE residents work on a Friday for the first time; Saturday-Sunday to be weekends,” WION, 08-Jan-2022. [Online]. Available: https://www.wionews.com/world/uae-residents-work-on-a-friday-for-the-first-time-saturday-sunday-to-be-weekends-443126#:~:text=The%20UAE%20government%20had%20announced,set%
[Accessed: 14-Sep-2022]. “Arima¶,” ARIMA - sktime documentation. [Online]. Available: https://www.sktime.org/en/v0.8.0/api_reference/auto_generated/sktime.forecasting.arima.ARIMA.html. [Accessed: 14-Sep-2022].

“Dubai Electricity & Water Authority: Bill Explanation,” . [Online]. Available: https://www.dewa.gov.ae/en/consumer/billing/understand-bill. [Accessed: 14-Sep-2022].

M. Pallante, “Time Series Forecasting models,” Medium, 20-Feb-2020. [Online]. Available: https://medium.com/analytics-vidhya/time-series-forecasting-models-726f7968a2c1. [Accessed: 14-Sep-2022].

J. Albanna, “Dubai Electricity & Water Authority (DEWA): Shams Dubai,” . [Online]. Available: https://www.dewa.gov.ae/en/consumer/useful-tools/shams-dubai. [Accessed: 14-Sep-2022].