

## gowtham.final

gowtham chakri/811240979

2022-12-08

### *#Loading all the libraries*

```
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.1.3

## -- Attaching packages ----- tidyverse
## 1.3.1 --

## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v dplyr  1.0.8
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1

## -- Conflicts -----
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(factoextra)

## Warning: package 'factoextra' was built under R version 4.1.3

## Welcome! Want to learn more? See two factoextra-related books at
## https://goo.gl/ve3WBa

library(class)
library(dplyr)
library(caret)

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
## lift

#Loading the datasets
data_1=read.csv("C:/Users/desineni/Downloads/fuel_receipts_costs_eia923.csv")
summary(data_1)
```

```

##      rowid      plant_id_eia plant_id_eia_label report_date
## Min.   :      1   Min.   :      3   Length:608565   Length:608565
## 1st Qu.:152142   1st Qu.: 2712   Class :character Class :character
## Median :304283   Median : 6155   Mode  :character Mode  :character
## Mean   :304283   Mean   :18290
## 3rd Qu.:456424   3rd Qu.:50707
## Max.   :608565   Max.   :64020
##
## contract_type_code contract_type_code_label contract_expiration_date
## Length:608565      Length:608565      Length:608565
## Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character
##
##
##
## energy_source_code energy_source_code_label fuel_type_code_pudl
## Length:608565      Length:608565      Length:608565
## Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character
##
##
##
## fuel_group_code      mine_id_pudl      mine_id_pudl_label supplier_name
## Length:608565      Min.   :      0   Min.   :      0   Length:608565
## Class :character   1st Qu.: 42     1st Qu.: 42     Class :character
## Mode  :character   Median : 972    Median : 972    Mode  :character
##                      Mean   :1577    Mean   :1577
##                      3rd Qu.:3121    3rd Qu.:3121
##                      Max.   :4562    Max.   :4562
##                      NA's   :391947  NA's   :391947
## fuel_received_units fuel_mmbtu_per_unit sulfur_content_pct
ash_content_pct
## Min.   :      1   Min.   :      0.000   Min.   : 0.0000   Min.   : 0.000
## 1st Qu.: 3700     1st Qu.: 1.025     1st Qu.: 0.0000   1st Qu.: 0.000
## Median : 21565    Median : 1.061     Median : 0.0000   Median : 0.000
## Mean   : 242967    Mean   : 8.839     Mean   : 0.5145   Mean   : 3.606
## 3rd Qu.: 106164    3rd Qu.: 17.809    3rd Qu.: 0.4900   3rd Qu.: 5.800
## Max.   :48159765   Max.   :1049.000    Max.   :11.0100   Max.   :72.200
##
## mercury_content_ppm fuel_cost_per_mmbtu primary_transportation_mode_code
## Min.   :0.00      Min.   : -71.9     Length:608565
## 1st Qu.:0.00      1st Qu.: 2.3      Class :character
## Median :0.00      Median : 3.3      Mode  :character
## Mean   :0.01      Mean   : 14.2
## 3rd Qu.:0.00      3rd Qu.: 4.8
## Max.   :1.82      Max.   :562572.2
## NA's   :289482    NA's   :200240
## primary_transportation_mode_code_label secondary_transportation_mode_code

```

```
## Length:608565          Length:608565
## Class :character       Class :character
## Mode :character        Mode :character
##
##
##
## secondary_transportation_mode_code_label natural_gas_transport_code
## Length:608565          Length:608565
## Class :character       Class :character
## Mode :character        Mode :character
##
##
##
## natural_gas_delivery_contract_type_code moisture_content_pct
## Length:608565          Min. : 0.0
## Class :character       1st Qu.: 6.6
## Mode :character        Median : 11.9
##                        Mean : 15.6
##                        3rd Qu.: 26.8
##                        Max. : 247.0
##                        NA's : 516589
## chlorine_content_ppm data_maturity data_maturity_label
## Min. : 0.0          Length:608565 Length:608565
## 1st Qu.: 0.0        Class :character Class :character
## Median : 0.0        Mode :character Mode :character
## Mean : 59.2
## 3rd Qu.: 0.0
## Max. : 3747.0
## NA's : 516589
```

*#selecting the rows for developing model*

```
data<-data_1[ , c(15,16,18)]
summary(data)
```

```
## fuel_received_units fuel_mmbtu_per_unit ash_content_pct
## Min. : 1          Min. : 0.000      Min. : 0.000
## 1st Qu.: 3700      1st Qu.: 1.025      1st Qu.: 0.000
## Median : 21565     Median : 1.061      Median : 0.000
## Mean : 242967      Mean : 8.839       Mean : 3.606
## 3rd Qu.: 106164    3rd Qu.: 17.809    3rd Qu.: 5.800
## Max. : 48159765    Max. : 1049.000    Max. : 72.200
```

```
map(data,~sum(is.na(.)))
```

```
## $fuel_received_units
## [1] 0
##
## $fuel_mmbtu_per_unit
```

```

## [1] 0
##
## $ash_content_pct
## [1] 0

nrow(data)

## [1] 608565

#dealing with missing values
data$ash_content_pct [is.na(data$ash_content_pct )]<-
median(data$ash_content_pct , na.rm = T)

#creating data partition

set.seed(0088)
trail<-createDataPartition(data$ash_content_pct,p=0.02,list=FALSE)
trail_dataset<-data[trail,]
ncol(trail_dataset)

## [1] 3

nrow(trail_dataset)

## [1] 12172

#data partitioning for training and testing

set.seed(0088)
division<-createDataPartition(trail_dataset$ash_content_pct,p=0.75,list =
FALSE)
trainset<-trail_dataset[division,]
testset<-trail_dataset[-division,]
nrow(trainset)

## [1] 9130

nrow(testset)

## [1] 3042

normalized_values<-preProcess(trainset ,method = c('center','scale'))

trainset_norm<-predict(normalized_values,trainset)

testset_norm<-predict(normalized_values,testset)

summary(trainset)

## fuel_received_units fuel_mmbtu_per_unit ash_content_pct
## Min. : 1 Min. : 0.076 Min. : 0.000
## 1st Qu.: 3938 1st Qu.: 1.025 1st Qu.: 0.000

```

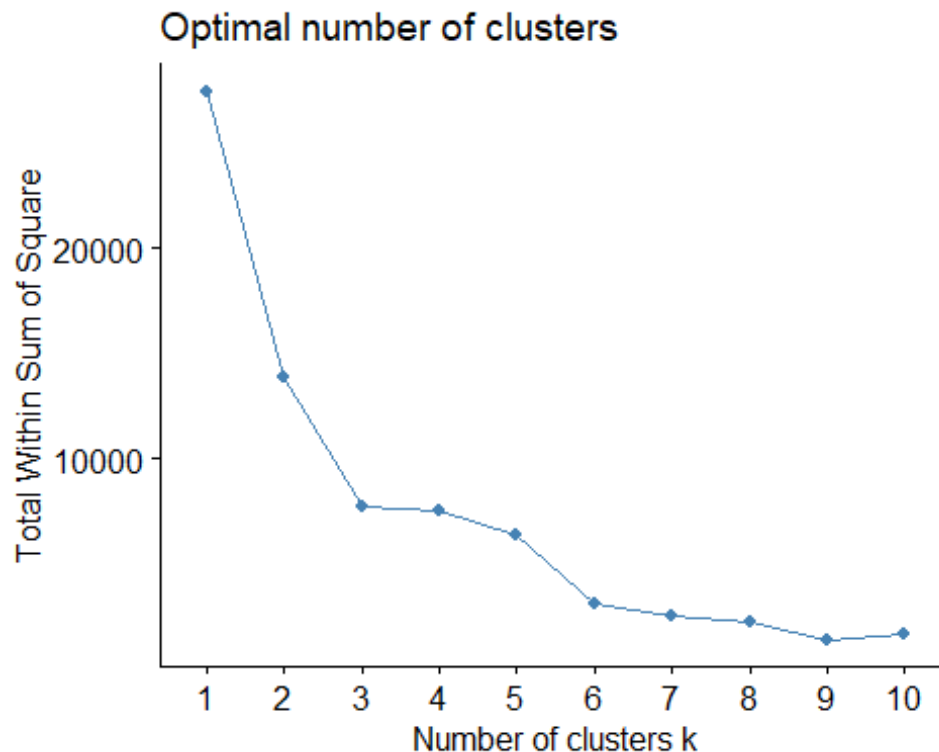
```
## Median : 22738      Median : 1.060      Median : 0.000
## Mean   : 240991     Mean   : 8.788      Mean   : 3.626
## 3rd Qu.: 105552     3rd Qu.:17.770     3rd Qu.: 5.800
## Max.   :12672784    Max.   :30.000     Max.   :63.000
```

```
summary(trainset_norm)
```

```
## fuel_received_units fuel_mmbtu_per_unit ash_content_pct
## Min.   :-0.3386      Min.   :-0.8895      Min.   :-0.5406
## 1st Qu.: -0.3330     1st Qu.: -0.7926     1st Qu.: -0.5406
## Median :-0.3066      Median :-0.7890     Median :-0.5406
## Mean    : 0.0000      Mean    : 0.0000     Mean    : 0.0000
## 3rd Qu.: -0.1903     3rd Qu.: 0.9172      3rd Qu.: 0.3241
## Max.    :17.4657      Max.    : 2.1660     Max.    : 8.8526
```

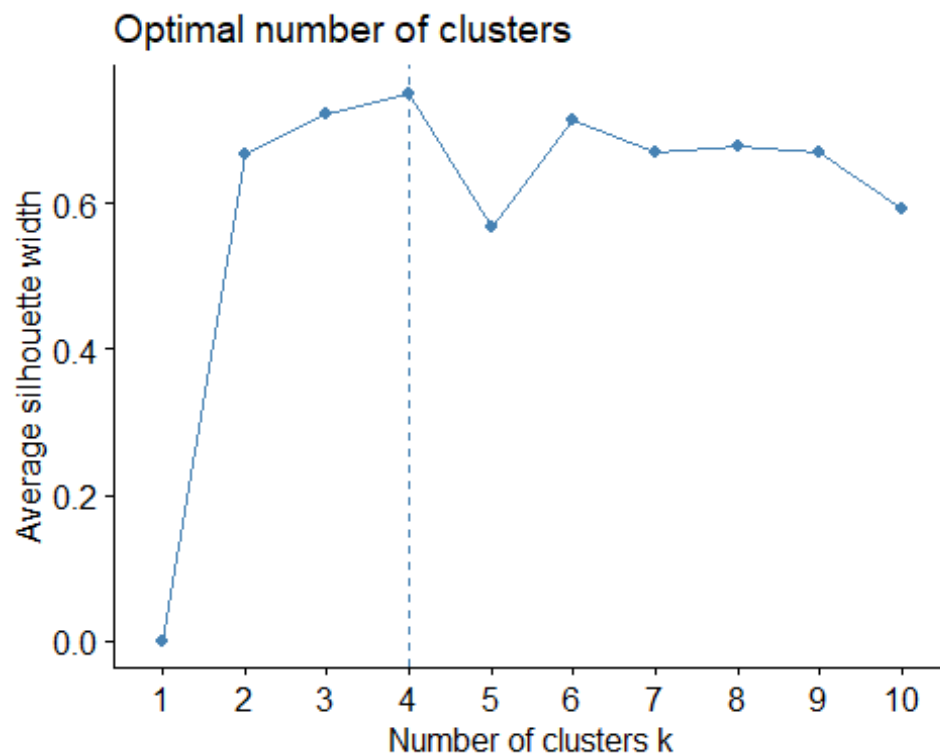
```
wss_k<-fviz_nbclust(trainset_norm,kmeans,method="wss")
```

```
wss_k
```



```
silhouette_k<-fviz_nbclust(trainset_norm,kmeans,method="silhouette")
```

```
silhouette_k
```



```
summary(trainset_norm)
```

```
## fuel_received_units fuel_mmbtu_per_unit ash_content_pct
## Min. :-0.3386 Min. :-0.8895 Min. :-0.5406
## 1st Qu.:-0.3330 1st Qu.:-0.7926 1st Qu.:-0.5406
## Median :-0.3066 Median :-0.7890 Median :-0.5406
## Mean : 0.0000 Mean : 0.0000 Mean : 0.0000
## 3rd Qu.:-0.1903 3rd Qu.: 0.9172 3rd Qu.: 0.3241
## Max. :17.4657 Max. : 2.1660 Max. : 8.8526
```

```
kmeans_4<-kmeans(trainset_norm,centers = 4, nstart = 25)
```

```
plot_kmeans_4<-fviz_cluster(kmeans_4,data = trainset_norm)
```

```
plot_kmeans_4
```



```
final_data<-trainset
final_data$cluster<-kmeans_4$cluster
head(final_data)
```

	fuel_received_units	fuel_mmbtu_per_unit	ash_content_pct	cluster
## 49	86684	22.600	10.0	4
## 93	2164	1.030	0.0	1
## 220	336	1.084	0.0	1
## 270	7	1.000	0.0	1
## 318	1860	1.033	0.0	1
## 366	22280	25.390	8.7	4

```
final_data%>%group_by(cluster)%>%
  summarize(avg_units=mean(fuel_received_units),
            avg_ash=mean(ash_content_pct),
            avg_mmbtu=mean(fuel_mmbtu_per_unit))
```

```
## # A tibble: 4 x 4
##   cluster avg_units    avg_ash avg_mmbtu
##   <int>    <dbl>    <dbl>    <dbl>
## 1     1  139643.  0.0000187    1.72
## 2     2   36131.  37.1         14.9
## 3     3 2873502.  0           1.01
## 4     4   47811.  8.37         21.6
```