ANOMALY DETECTION

IYLINE CHUMO

10/09/2021

Defining the Question

Performing analysis on Carrefour Kenya dataset to draw insightsthat could help the marketing department on the most relevant marketing strategies that will result in the highest no. of sales (total price including tax).

Loading the libraries

```
# Load libraries
suppressWarnings(
        suppressMessages(if
                         (!require(tidyverse, quietly=TRUE))
                 install.packages("tidyverse")))
library(tidyverse)
suppressWarnings(
        suppressMessages(if
                         (!require(anomalize, quietly=TRUE))
                 install.packages("anomalize")))
library(anomalize)
suppressWarnings(
        suppressMessages(if
                         (!require(tibbletime, quietly=TRUE))
                 install.packages("tibbletime")))
library(tibbletime)
suppressWarnings(
        suppressMessages(if
                         (!require(dplyr, quietly=TRUE))
                 install.packages("dplyr")))
library(dplyr)

library(tibbletime)
library(dplyr)
```

Loading the Dataset

```
sales <- read.csv("http://bit.ly/CarreFourSalesDataset")
sales$Date <- as.Date(sales$Date, format ="%m/%d/%Y")
sales$Date <- sort(sales$Date, decreasing = FALSE)
sales <- as_tbl_time(sales, index = Date)
sales <- sales %>%
    as_period("daily")
```

previewing the dataset

Checking the dimensions of our dataset

```
dim(sales)
```

```
## [1] 89  2
```

our dataset has 89rows and 2 variables

previewing the top of the dataset

```
head(sales)
```

```
## # A time tibble: 6 x 2
## # Index: Date
##   Date       Sales
##   <date>     <dbl>
## 1 2019-01-01  549.
## 2 2019-01-02  246.
## 3 2019-01-03  452.
## 4 2019-01-04  464.
## 5 2019-01-05  418.
## 6 2019-01-06  536.
```

Previewing the bottom of the dataset

```
tail(sales)
```

```
## # A time tibble: 6 x 2
## # Index: Date
##   Date       Sales
##   <date>     <dbl>
## 1 2019-03-25 361.
## 2 2019-03-26 188.
## 3 2019-03-27  43.9
## 4 2019-03-28 271.
## 5 2019-03-29 244.
## 6 2019-03-30 633.
```

Exploratory Data Analysis

Univariate Analysis

Finding the mean of sales column

```
colMeans(sales[sapply(sales,is.numeric)])
```

```
##    Sales
## 346.1541
```

Finding the median

```
Sales_median <- median(sales$Sales)
Sales_median

## [1] 253.26
```

Finding the mode

```
getmode <- function(v) {
   uniqv <- unique(v)
   uniqv[which.max(tabulate(match(v, uniqv)))]}
getmode(sales$Sales)

## [1] 548.9715
```

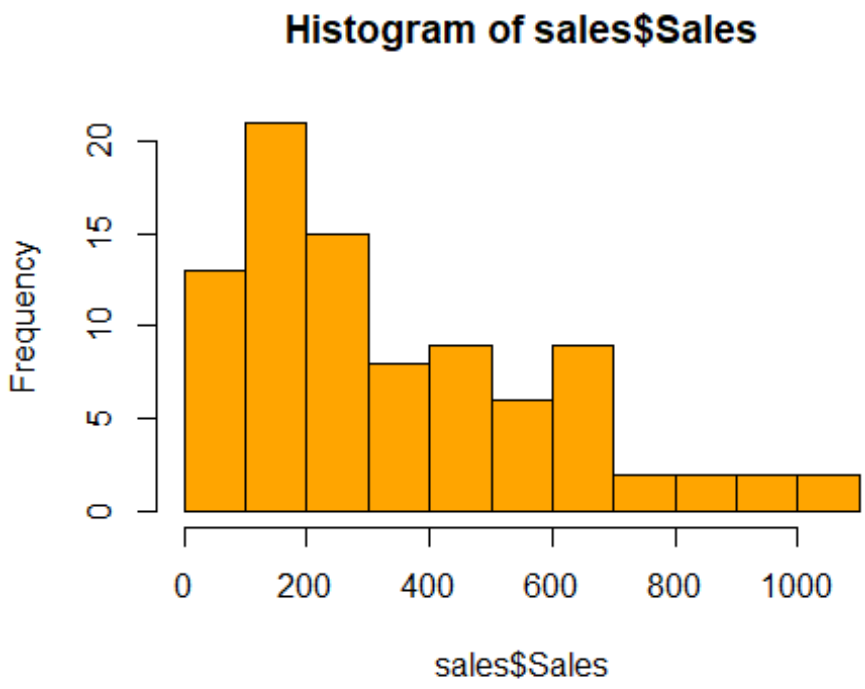
Finding the quantiles

```
quantile(sales$Sales)

##        0%       25%       50%       75%      100% 
##   18.6375  150.0975  253.2600  517.9650 1034.4600
```

Plotting the gistograms

```
hist(sales$Sales, col  = "orange")
```



Detecting Anomalies

```
sales %>%
    time_decompose(Sales) %>%
    anomalize(remainder) %>%
```

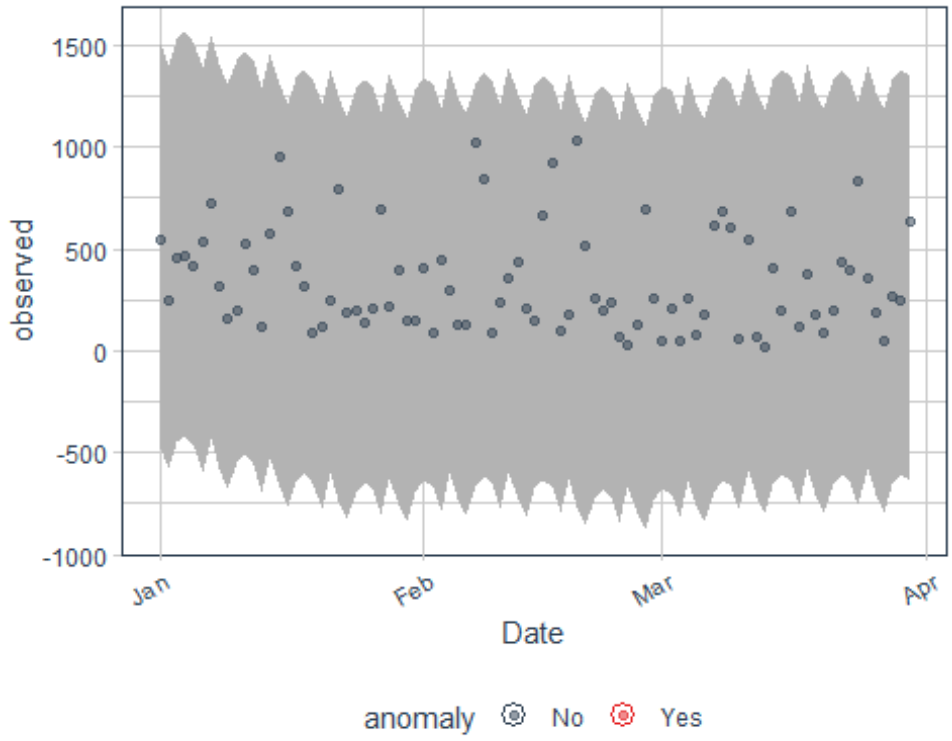
```
    time_recompose() %>%
    plot_anomalies(time_recomposed = TRUE, ncol = 3, alpha_dots = 0.5)

## frequency = 7 days

## trend = 30 days

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

## Warning: `type_convert()` only converts columns of type 'character'.
## - `df` has no columns of type 'character'
```



There are no anomalies in our dataset.