Carrefour-Marketing-Project Applying Anomaly Detection

Margaret Gathoni

2022-07-29

Introduction

Carrefour has 13 outlets mostly located in the suburbs of Kenya's capital city, Nairobi. Their mission is to provide our customers with quality services, products and food accessible to all across all distribution channels.

Problem Statement

The project aim to inform the marketing department on the most relevant marketing strategies that will result in the highest no. of sales (total price including tax).

Metric of Success

a. Identification of anomalies In the given data set.
b. Provide insights from the analysis on whether there was fraud or not.

Anomaly Detection

Loading libraries

```
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse
1.3.2 ──
## ✓ ggplot2 3.3.6     ✓ purrr   0.3.4
## ✓ tibble  3.1.8     ✓ dplyr   1.0.9
## ✓ tidyr   1.2.0     ✓ stringr 1.4.0
## ✓ readr   2.1.2     ✓ forcats 0.5.1
## ── Conflicts ──────────────────────────────────────────
tidyverse_conflicts() ──
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()

library(anomalize)

## ══ Use anomalize to improve your Forecasts by 50%!
═════════════════════════════
## Business Science offers a 1-hour course - Lab #18: Time Series Anomaly
Detection!
## </> Learn more at: https://university.business-science.io/p/learning-labs-
pro </>
```

```
library(tibbletime)

##
## Attaching package: 'tibbletime'
##
## The following object is masked from 'package:stats':
##
##     filter

library(anomalize)
library(timetk)
```

Loading the data set

```
data <- read.csv('http://bit.ly/CarreFourSalesDataset')
head(data)

##        Date    Sales
## 1  1/5/2019 548.9715
## 2  3/8/2019  80.2200
## 3  3/3/2019 340.5255
## 4 1/27/2019 489.0480
## 5  2/8/2019 634.3785
## 6 3/25/2019 627.6165
```

Change Date column to Date format

```
data$Date <- as.Date(data$Date, format ="%m/%d/%Y")
data$Date <- sort(data$Date, decreasing = FALSE)
```

Will Convert data to a tibble

```
data <- as_tbl_time(data, index = Date)
head(data)

## # A time tibble: 6 × 2
## # Index: Date
##   Date       Sales
##   <date>     <dbl>
## 1 2019-01-01 549.
## 2 2019-01-01  80.2
## 3 2019-01-01 341.
## 4 2019-01-01 489.
## 5 2019-01-01 634.
## 6 2019-01-01 628.

data <- data %>%
    as_period("daily")
```

Previewing the number of objects in the data set

```
dim(data)
```

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

We have 89 rows and 2 columns

Visualize the Anomalies

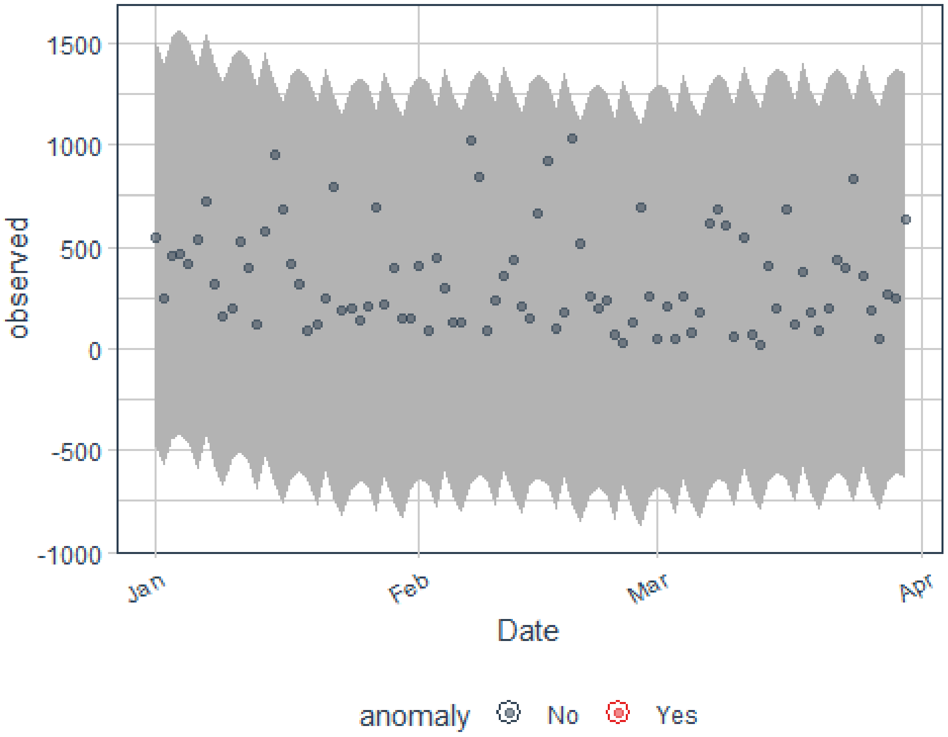
Plotting Anomalies using different alphas a. alpha 0.5

```
data %>%
    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
```

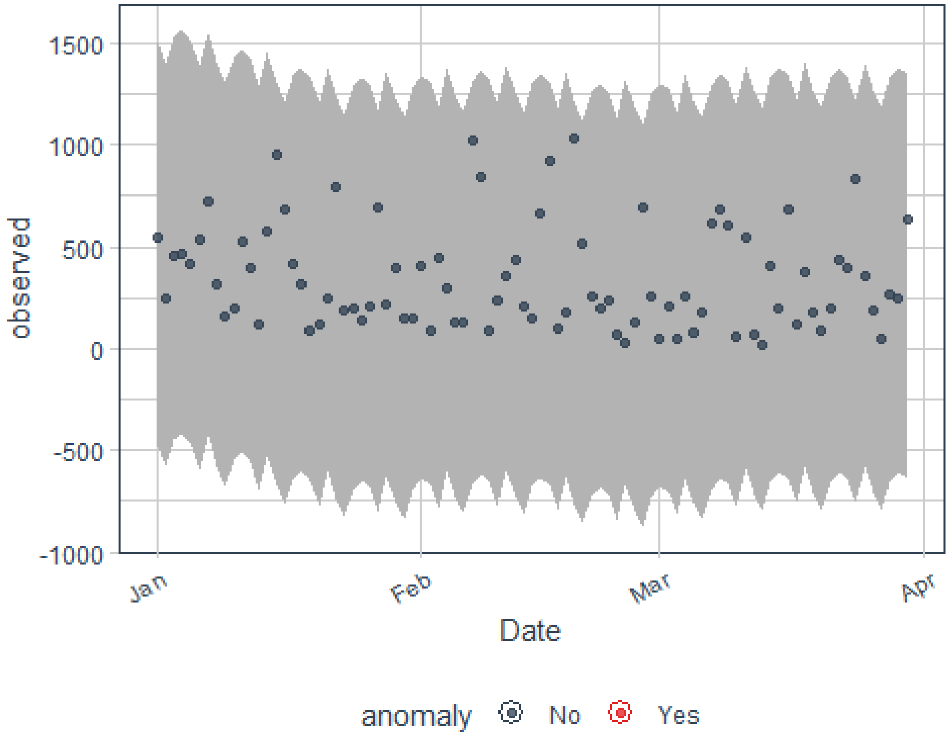


We can see there was no anomalies (red dot) detected

```
data %>%
    time_decompose(Sales) %>%
    anomalize(remainder) %>%
    time_recompose() %>%
    plot_anomalies(time_recomposed = TRUE, ncol = 3, alpha_dots = 0.75)
```

```
## frequency = 7 days

## trend = 30 days
```

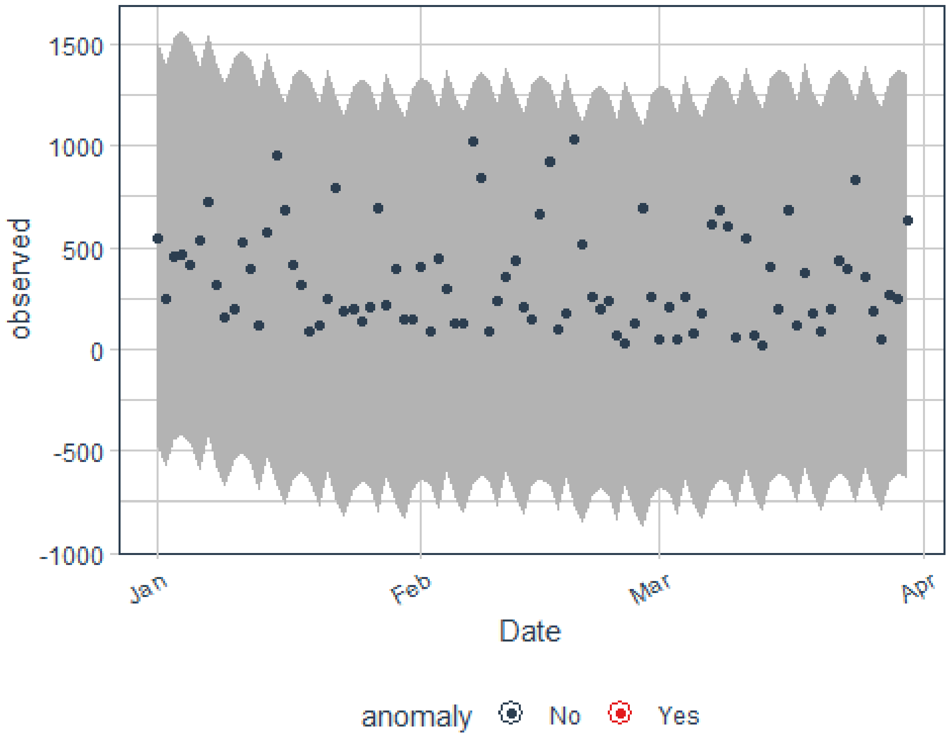


We can see there was no anomalies (red dot) detected

```
data %>%
    time_decompose(Sales) %>%
    anomalize(remainder) %>%
    time_recompose() %>%
    plot_anomalies(time_recomposed = TRUE, ncol = 3, alpha_dots = 1)

## frequency = 7 days

## trend = 30 days
```

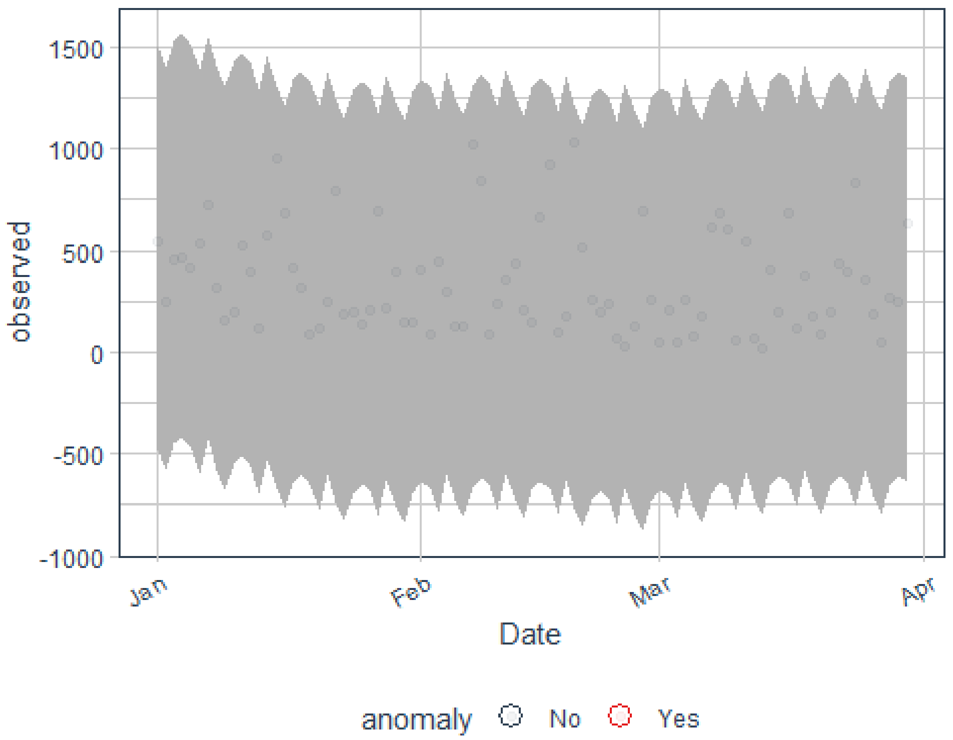


We can see there was no anomalies (red dot) detected

```r
data %>%
    time_decompose(Sales) %>%
    anomalize(remainder) %>%
    time_recompose() %>%
    plot_anomalies(time_recomposed = TRUE, ncol = 3, alpha_dots = 0.05)
```

```
## frequency = 7 days

## trend = 30 days
```



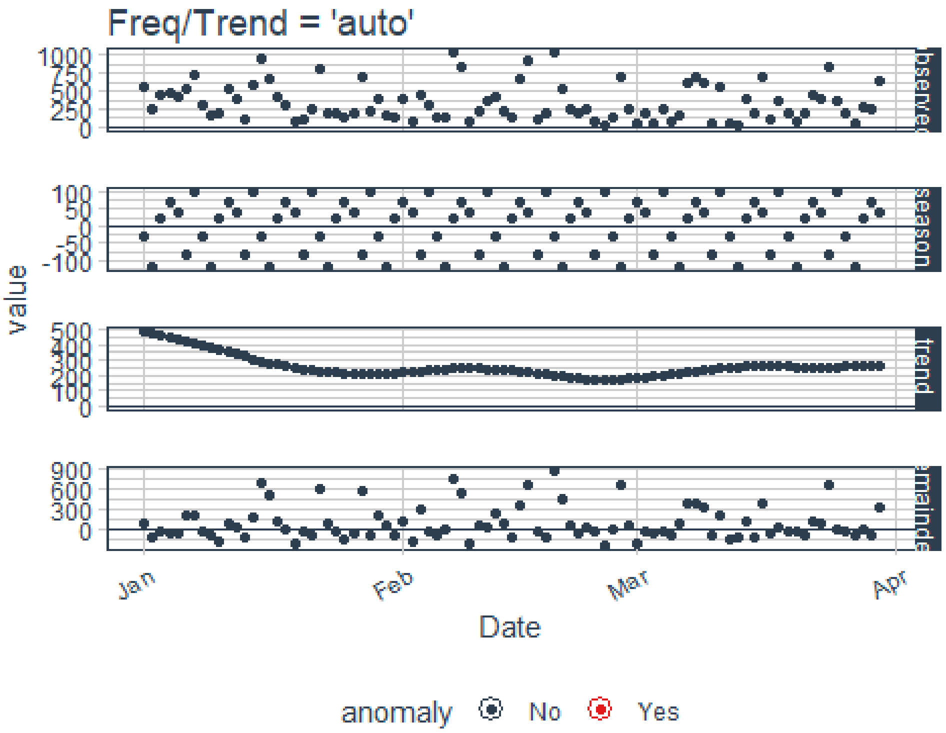
We can see there was no anomalies (red dot) detected

Adjusting Trend and Seasonality

```
data %>%
    time_decompose(Sales) %>%
    anomalize(remainder) %>%
    time_recompose() %>%
    plot_anomaly_decomposition() +
    ggtitle("Freq/Trend = 'auto'")

## frequency = 7 days

## trend = 30 days
```



We still have no anomalies detected.

Let’s view logical frequency and trend spans based on the scale of the data.

```
get_time_scale_template()

## # A tibble: 8 × 3
##   time_scale frequency trend
##   <chr>      <chr>     <chr>
## 1 second     1 hour    12 hours
## 2 minute     1 day     14 days
## 3 hour       1 day     1 month
## 4 day        1 week    3 months
## 5 week       1 quarter 1 year
## 6 month      1 year    5 years
## 7 quarter    1 year    10 years
## 8 year       5 years   30 years
```

The tibble shows that the scale, the frequency and the trend used for our time series by the system.

Conclusion

We can confidently advise Carrefour marketing department that there was no anomaly detected.