

Anomaly Detection Analysis

Naomi Chebet

9/19/2020

1. Business Understanding

1.1 Define the question: the objective of this project is to find out if the given time series sales data from Carrefour has any anomalies.

1.2 Metric for Success

Our project will be successful if we are able to plot an anomaly graph to demonstrate the presence/absence of anomaly in the given data.

1.3 Experimental Design

The approach for the project includes:

1. Loading libraries, and data
2. Preprocessing
3. Implementing the solution

2. Loading Libraries and Data

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.2      v purrr   0.3.4
## v tibble  3.0.3      v dplyr   1.0.2
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

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

library(tibbletime)

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

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(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
# Loading the data
```

```
sales.date <- read.csv("~/Moringa School/R Programming/R datasets/Supermarket_Sales_Forecasting - Sales")
glimpse(sales.date)
```

```
## Rows: 1,000
## Columns: 2
## $ Date <chr> "1/5/2019", "3/8/2019", "3/3/2019", "1/27/2019", "2/8/2019", ...
## $ Sales <dbl> 548.9715, 80.2200, 340.5255, 489.0480, 634.3785, 627.6165, 43...
```

The date is listed as a character, let's convert it to datetime.

3. Preprocessing

```
#converting the date to dtm
```

```
sales.date <- sales.date %>%
  mutate(Date = Date %>% str_c(" 00:00:00") %>% mdy_hms())
glimpse(sales.date)
```

```
## Rows: 1,000
## Columns: 2
## $ Date <dtm> 2019-01-05, 2019-03-08, 2019-03-03, 2019-01-27, 2019-02-08, ...
## $ Sales <dbl> 548.9715, 80.2200, 340.5255, 489.0480, 634.3785, 627.6165, 43...
```

```
#checking for class of our dataset
```

```
class(sales.date)
```

```
## [1] "data.frame"
```

Our dataset object is a dataframe, we need to convert it to a `tbl_time` object before running the `anomalize` algorithm.

```
sales.date <- as_tbl_time(sales.date, index = Date)
sales.date
```

```
## # A time tibble: 1,000 x 2
## # Index: Date
##   Date          Sales
##   <dtm>         <dbl>
## 1 2019-01-05 00:00:00 549.
## 2 2019-03-08 00:00:00  80.2
## 3 2019-03-03 00:00:00 341.
## 4 2019-01-27 00:00:00 489.
## 5 2019-02-08 00:00:00 634.
```

```
## 6 2019-03-25 00:00:00 628.
## 7 2019-02-25 00:00:00 434.
## 8 2019-02-24 00:00:00 772.
## 9 2019-01-10 00:00:00 76.1
## 10 2019-02-20 00:00:00 173.
## # ... with 990 more rows
```

4. Implementing the solution

```
#applying the anomalization function
```

```
sales.date.anomalized <- sales.date %>%
  time_decompose(Sales, merge = TRUE) %>%
  anomalize(remainder) %>%
  time_recompose()
```

```
## Note: Index not ordered. tibblertime assumes index is in ascending order. Results may not be as desired
```

```
## frequency = 12 seconds
```

```
## Note: Index not ordered. tibblertime assumes index is in ascending order. Results may not be as desired
```

```
## trend = 12 seconds
```

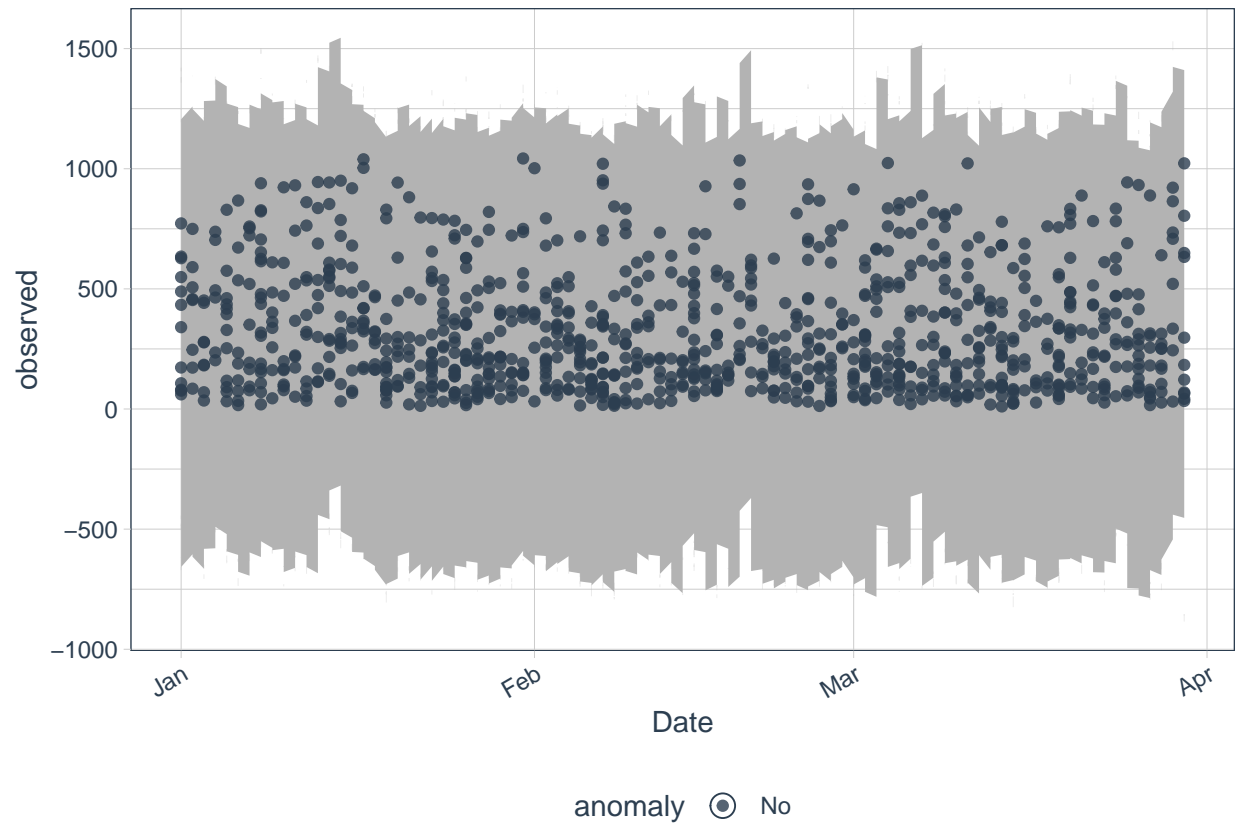
```
## Registered S3 method overwritten by 'quantmod':
```

```
##   method      from
```

```
##   as.zoo.data.frame zoo
```

```
#plotting after anomalization to see if we have any anomalies
```

```
sales.date.anomalized %>%
  plot_anomalies(time_recomposed = TRUE, ncol = 1, alpha_dots = 0.8)
```



Based on the results obtained, there are no anomalies that were detected.

5. Follow up questions

Did we have the right data? yes we did

Do we need more data? Absolutely. The data given was for 3 months only.