Research Question

You are a Data analyst at Carrefour Kenya and are currently undertaking a project that will inform the marketing department on the most relevant marketing strategies that will result in the highest no. of sales (total price including tax)

Anomaly Detection

```
# Installing anomalize package
# ---
#
install.packages("anomalize")
```

```
Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)

also installing the dependencies ‘timetk’, ‘sweep’
```

```
# Load tidyverse and anomalize
# ---
#
library(tidyverse)
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 </>
```

```
# Collect our time series data
# http://bit.ly/CarreFourSalesDataset
#Reading the dataset
#Previewing the first six rows of the dataset.
library("data.table")
df = fread('http://bit.ly/CarreFourSalesDataset')
head(df)
```

```
Attaching package: ‘data.table’


The following objects are masked from ‘package:dplyr’:

    between, first, last


The following object is masked from ‘package:purrr’:

    transpose
```

A data.table: 6 × 2

Date	Sales
<chr>	<dbl>
1/5/2019	548.9715
3/8/2019	80.2200
3/3/2019	340.5255

```
#Checking the dimensions of the dataset
dim(df)
```

```
1000 · 2
```

The dataset has 1000 rows and 2 columns.

```
#Checking the structure of the dataset
str(df)
```

```
Classes ‘data.table’ and 'data.frame':  1000 obs. of  2 variables:
 $ Date : chr  "1/5/2019" "3/8/2019" "3/3/2019" "1/27/2019" ...
 $ Sales: num  549 80.2 340.5 489 634.4 ...
 - attr(*, ".internal.selfref")=<externalptr>
```

```
#Changing the date column to its appropriate data type.
df$Date <- as.Date(df$Date, "%m/%d/%Y")
str(df)
```

```
Classes ‘data.table’ and 'data.frame':  1000 obs. of  2 variables:
 $ Date : Date, format: "2019-01-05" "2019-03-08" ...
 $ Sales: num  549 80.2 340.5 489 634.4 ...
 - attr(*, ".internal.selfref")=<externalptr>
```

```
# totalling the sales based on their common shared dates
sales_tot <- aggregate(df$Sales, by = list(Date = df$Date), FUN = sum)
```

```
head(sales_tot)
```

A data.frame: 6 × 2

	Date	x
	<date>	<dbl>
1	2019-01-01	4745.181
2	2019-01-02	1945.503
3	2019-01-03	2078.128
4	2019-01-04	1623.688
5	2019-01-05	3536.684
6	2019-01-06	3614.205

```
#Changing the column name x to total sales.
names(sales_tot)[2] <- 'Total_sales'
head(sales_tot)
```

A data.frame: 6 × 2

	Date	Total_sales
	<date>	<dbl>
1	2019-01-01	4745.181
2	2019-01-02	1945.503
3	2019-01-03	2078.128
4	2019-01-04	1623.688
5	2019-01-05	3536.684
6	2019-01-06	3614.205

```
#Checking the dimensions of sales tot
dim(sales_tot)
```

89 · 2

```
#Checking for any duplicates
anyDuplicated(sales_tot)
```

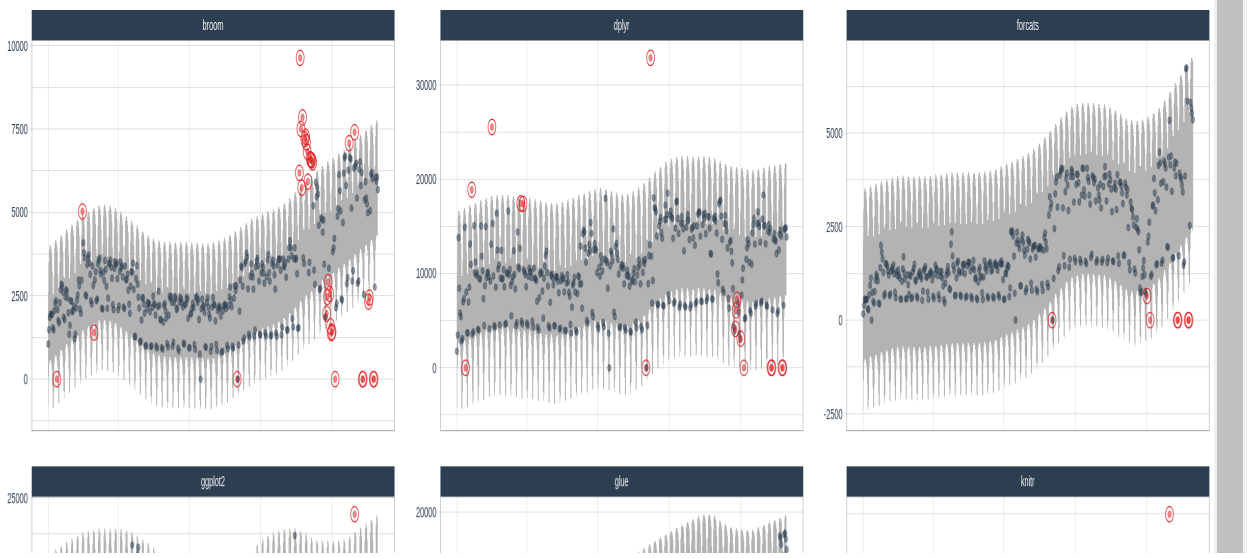
0

```
#Checking for any missing values
colSums(is.na(sales_tot))
```

```
Date:         0 Total_sales:         0
```

```
# time_decompose() - this function would help with time series decomposition.
# anomalize() -
# We perform anomaly detection on the decomposed data using
# the remainder column through the use of the anomalize() function
# We create the lower and upper bounds around the “observed” values
# through the use of the time_recompose() function, which recomposes
# the lower and upper bounds of the anomalies around the observed values.
# we now plot using plot_anomaly_decomposition() to visualize out data.
options(repr.plot.width = 20, repr.plot.height = 20)
tidyverse_cran_downloads %>%
    time_decompose(count) %>%
    anomalize(remainder) %>%
    time_recompose() %>%
    plot_anomalies(time_recomposed = TRUE, ncol = 3, alpha_dots = 0.5)
```

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



Conclusion and Recommendation

There are anomalies in the dataset hence indicating that there could be some cases of fraud.