

PORTFOLIO

BY: SIDDARTH KUMAR SAMPATH KUMAAR
MAIL: SSAMPATH@SYR.EDU
CELL NUMBER: 315-751-4932

The report consists of Insights derived from the dataset provided by the company

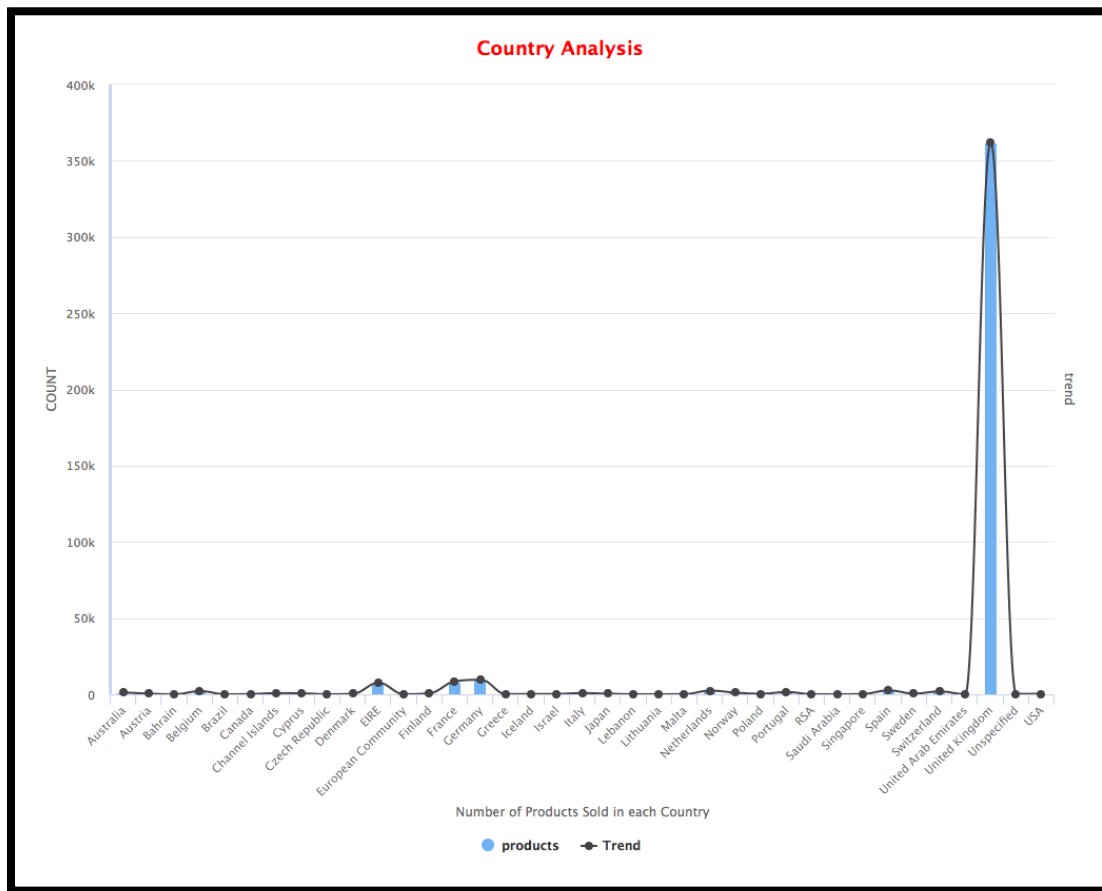
- **Dataset:** The dataset consists of **541,917** observations and 8 columns which represent the features of a sales report. The dataset consisted of 136,590 missing values or NA's.

```
> sum(is.na(mydata))  
[1] 136590
```

- In this Analysis, I have omitted all the missing or null values to have more qualitative and accurate analysis.
 - Presence of Na's in the Data could lead to inaccurate and improper analysis and report generation.
- **Tools:** For this analysis, R-Studio, Tableau, MS Excel have been used. Thanks to the freedom given by the company to use any tools, I was able to select most suitable tool for that particular analysis.
 - R- Studio and MS EXCEL have been the mostly used tools for Data cleaning, Data munging and Visualizations.
 - For more interactive visualizations, Tableau has been used in some instances.
- **Insights that were focused in this analysis:**
 - Which Country has the biggest market?
 - What are most frequently bought items in that particular country?
 - How much quantity is sold per Item in that country?
 - How much revenue is generated by each of these products in that country?
 - Which product has the sold the most in that country?
 - What is the sales pattern based on the time of the day in that country?
 - Which product generated the highest revenue in that country?
 - What is the buying pattern of this product over the period of time?
 - What is the Sales vs Revenue pattern for these products?
 - Who is the largest buyer in that country?
 - Which Customer generates highest revenue in the country?
 - How is the market in other countries?

COUNTRY WISE PERFORMANCE –

In order to focus our analysis on a particular market segment to get insights, let us analyze how is the market trend in different countries based on the sales report.



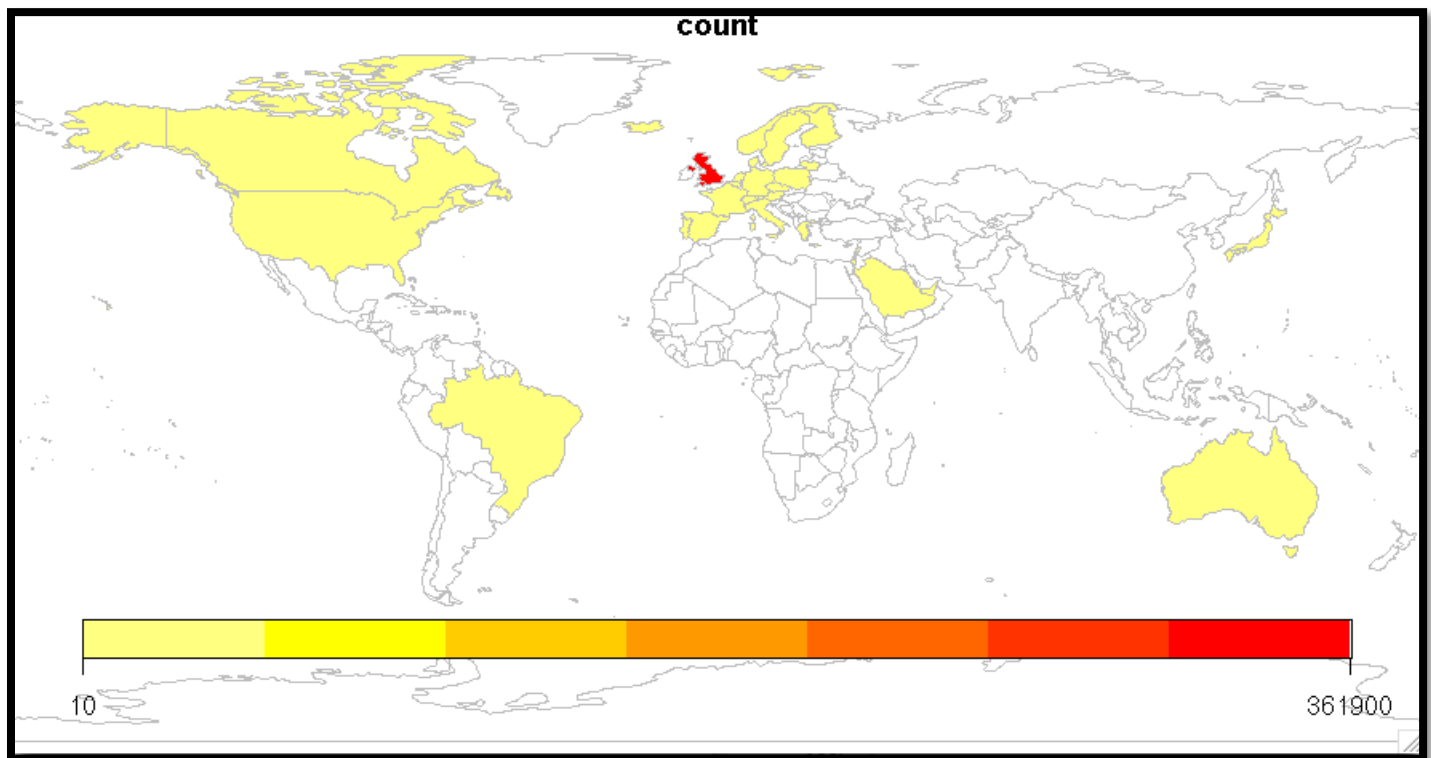
- From the above plot, we can clearly determine that the majority of the sales have been in the United Kingdom and the product has much bigger market in United Kingdom compared to other countries.

R-Code:

```
mydata$Country <- as.factor(mydata$Country)
country_count = data.frame(summary(mydata$Country))
colnames(country_count) <- "count"
View(country_count)

highchart() %>% hc_title(text = "<b>Country Analysis</b>",
  margin = 20, align = "center",
  style = list(color = "red", useHTML = TRUE)) %>%
  hc_yAxis_multiples(
    list(lineWidth = 3, title = list(text = "COUNT"), min=0),
    list(showLastLabel = FALSE, opposite = TRUE, title = list(text = "trend"))
  ) %>%
  hc_xAxis(title=list(text = "Number of Products Sold in each Country" ), categories = row.names(country_count)) %>%
  hc_add_series(name = "products", data = country_count$count ,type = "column") %>%
  hc_add_series(name = "Trend", data = country_count$count ,type = "spline") %>%
  hc_plotOptions(series = list(stacking = FALSE)) %>%
  hc_chart(type = "column")
```

To make it Interactive,



- From this Map, we can clearly observe the global trend of the sales. The scale below the map represents the sales per country with color intensities. With that reference, we can clearly observe that United Kingdom has the largest market in the world.
- We can also observe the other countries involved in the trade, which can help us tap into potential markets.

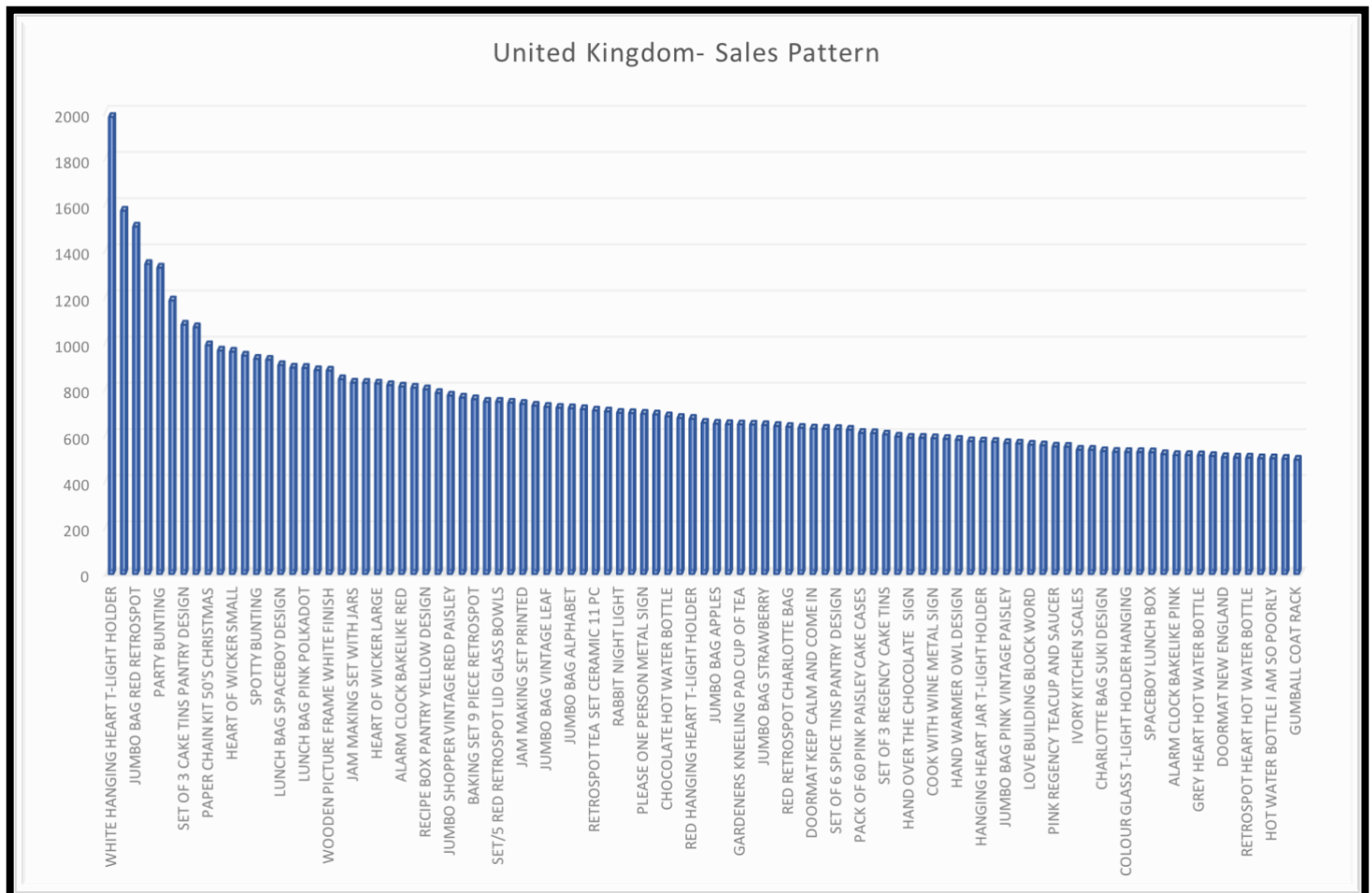
R Code:

```
library(rworldmap)
mapDevice('x11')
country_count$name <- row.names(country_count)
spdf <- joinCountryData2Map(country_count, joinCode="NAME", nameJoinColumn="name")
mapCountryData(spdf, nameColumnToPlot="count", catMethod="fixedWidth")
```

- Now that we have observed that United Kingdom has the largest market in the world, let us focus in to the that particular country to get more details insights.

Products Frequently Bought in United Kingdom:

- This particular analysis lets us know the frequently and popular products in the market. NOTE: This analysis does not determine the amount of quantity sold or the revenue generated.

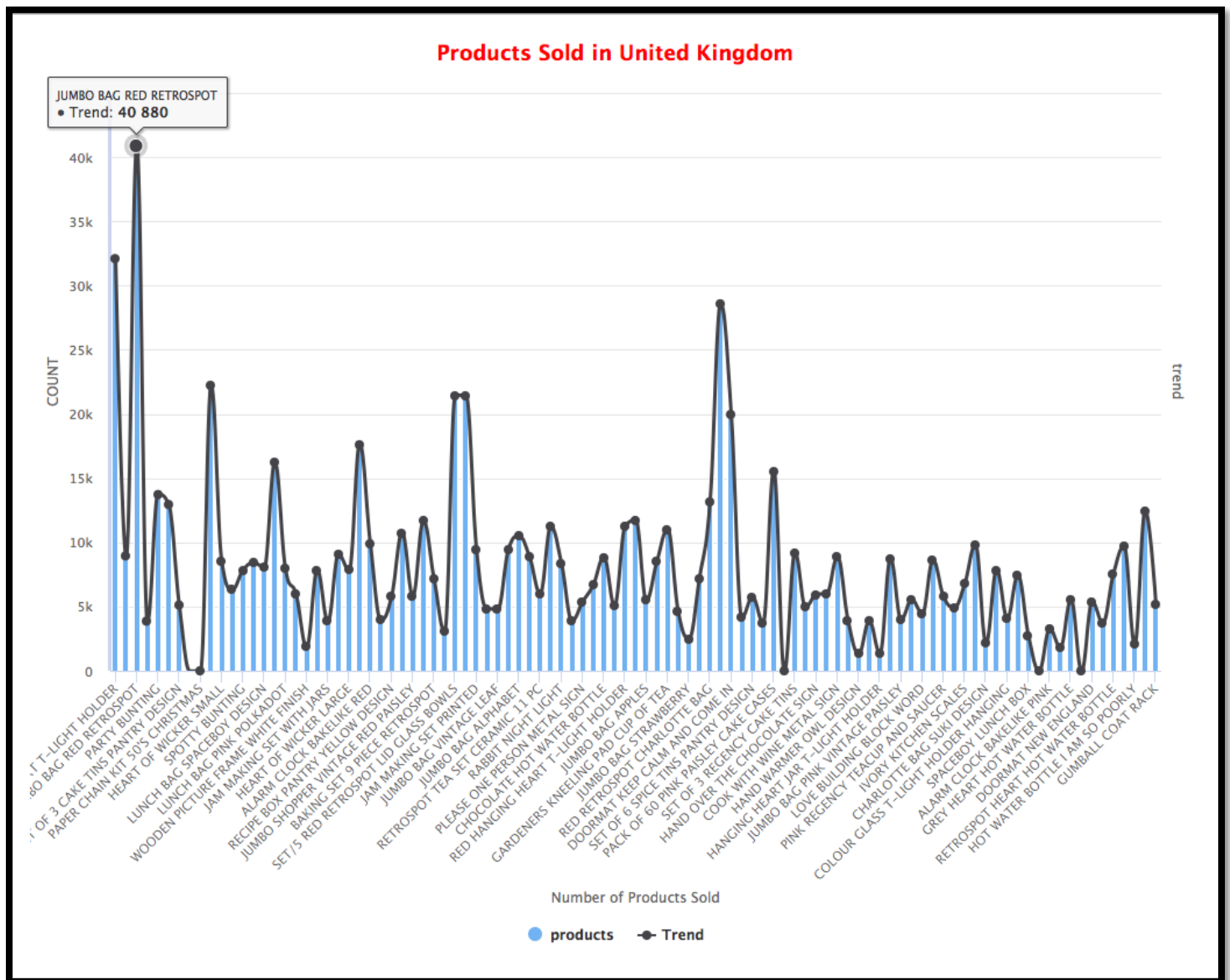


- From This analysis, we can infer that the most frequently bought product in the United Kingdom is WHITIE HANGING HEART T-LIGHT HOLDER. This particular item sold nearly 1980 pieces. The second most frequently bought items would be REGENCY CAKESTAND 3TIER which sold nearly 1574 pieces.

MS EXCEL: This graph was generated using PIVOT tables and charts in MS excel

- To Look into the Sales and Revenue aspect of the products, let us consider these frequently bought objects and determine the trend.
- We have Selected the Top 100 Frequently bought items in the market for this particular analysis.

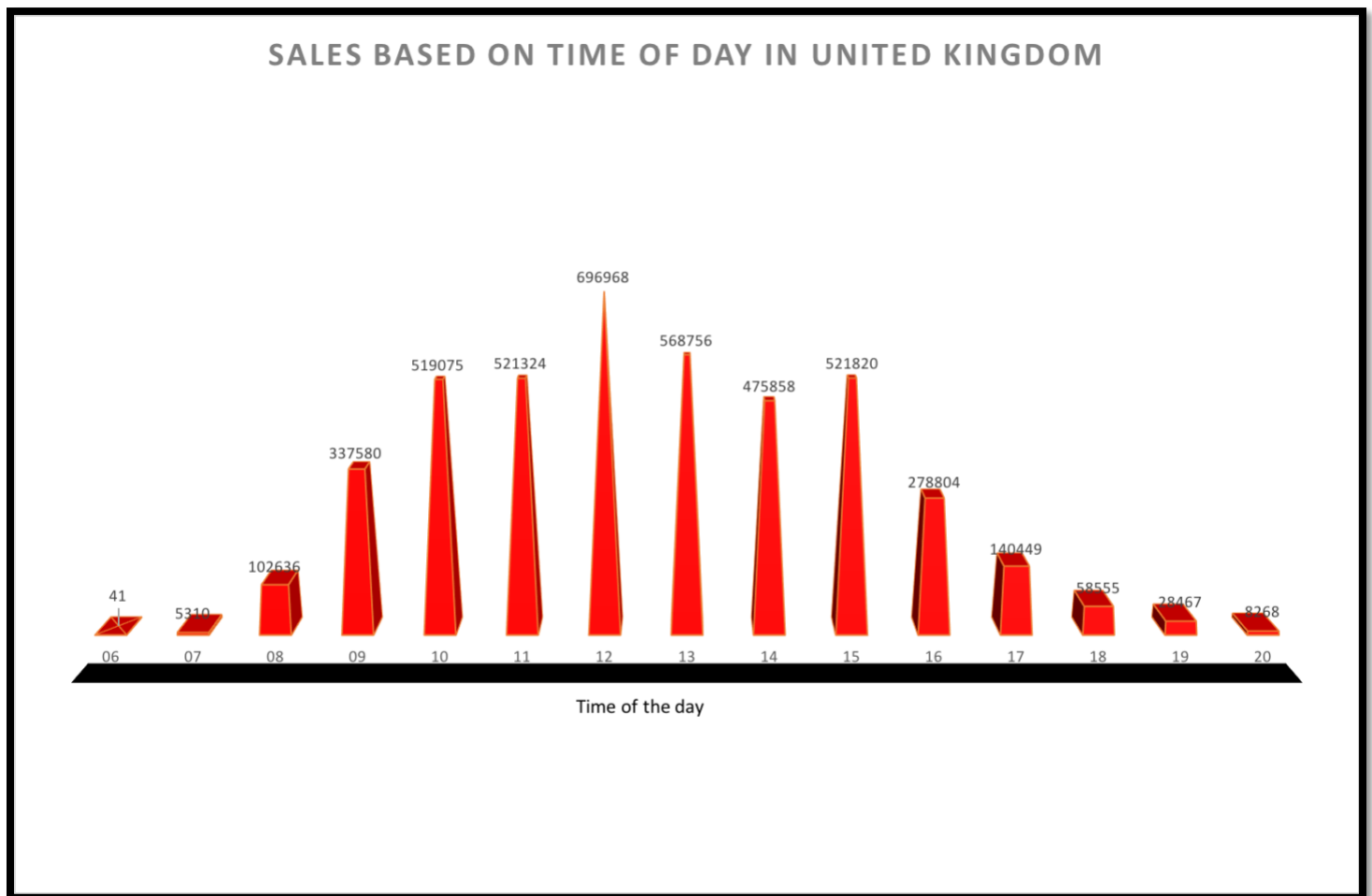
SALES TREND OF PRODUCT IN UNITED KINGDOM:



- From this plot, we can determine that the product which sold most number of quantities would be JUMBO BAG RED RETROSPOT. This particular item has sold nearly 40,880 pieces in United Kingdom. WHITE HANGING HEART T-LIGHT HOLDER closely follows with a count of 33,225 pieces.

SALES TREND BASED ON TIME OF THE DAY IN UNITED KINGDOM:

- This analysis focusses on the number of quantities sold based on the time of the day. This would help us focus on which particular time period we must focus to capitalize market or to have promotional offers that would generate more sales.



Scale: The numbers on x-axis represents time of the day (6 am – 8 pm)

- From this graph, we can see that the maximum number of sales happens during the time range from 10 am till 3 pm.
- We can also notice that the maximum number of sales happens during noon. We can also notice that from 3pm till rest of the day the sales decline drastically.
- This insight can be used by the company to have promotional offers, discounts during this time of the day to increase their sales and thereby their net revenue.

Tool: This analysis was developed using R and MS EXCEL. The invoice time was used to generate this graph.

R CODE:

```
#### White Hanging Heart T- Light Holder
whh <- uk[uk$Description=="WHITE HANGING HEART T-LIGHT HOLDER",]
total_sales <- sum(whh$Quantity)
total_revenue <- sum(whh$Quantity*whh$UnitPrice)
total_sales

### Regency###
rct <- uk[uk$Description=="REGENCY CAKESTAND 3 TIER",]
total_sales_rct <- sum(rct$Quantity)
total_revenue_rct <- sum(rct$Quantity*rct$UnitPrice)

####JUMBO BAG RED RETROSPOT###
jbr <- uk[uk$Description=="JUMBO BAG RED RETROSPOT",]
total_sales_jbr <- sum(jbr$Quantity)
total_revenue_jbr <- sum(jbr$Quantity*jbr$UnitPrice)

####ASSORTED COLOUR BIRD ORNAMENT###
acb <- uk[uk$Description=="ASSORTED COLOUR BIRD ORNAMENT",]
total_sales_acb <- sum(acb$Quantity)
total_revenue_acb <- sum(acb$Quantity*acb$UnitPrice)

####PARTY BUNTING###
pb <- uk[uk$Description=="PARTY BUNTING",]
total_sales_pb <- sum(pb$Quantity)
total_revenue_pb <- sum(pb$Quantity*pb$UnitPrice)
####LUNCH BAG RED RETROSPOT
lbr <- uk[uk$Description=="LUNCH BAG RED RETROSPOT",]
total_sales_lbr <- sum(lbr$Quantity)
total_revenue_lbr <- sum(lbr$Quantity*lbr$UnitPrice)

####SET OF 3 CAKE TINS PANTRY DESIGN###
sct <- uk[uk$Description=="SET OF 3 CAKE TINS PANTRY DESIGN",]
total_sales_sct <- sum(sct$Quantity)
total_revenue_sct <- sum(sct$Quantity*sct$UnitPrice)

####POSTAGE###
post <- uk[uk$Description=="POSTAGE",]
total_sales_post <- sum(post$Quantity)
total_revenue_post <- sum(post$Quantity*post$UnitPrice)
```

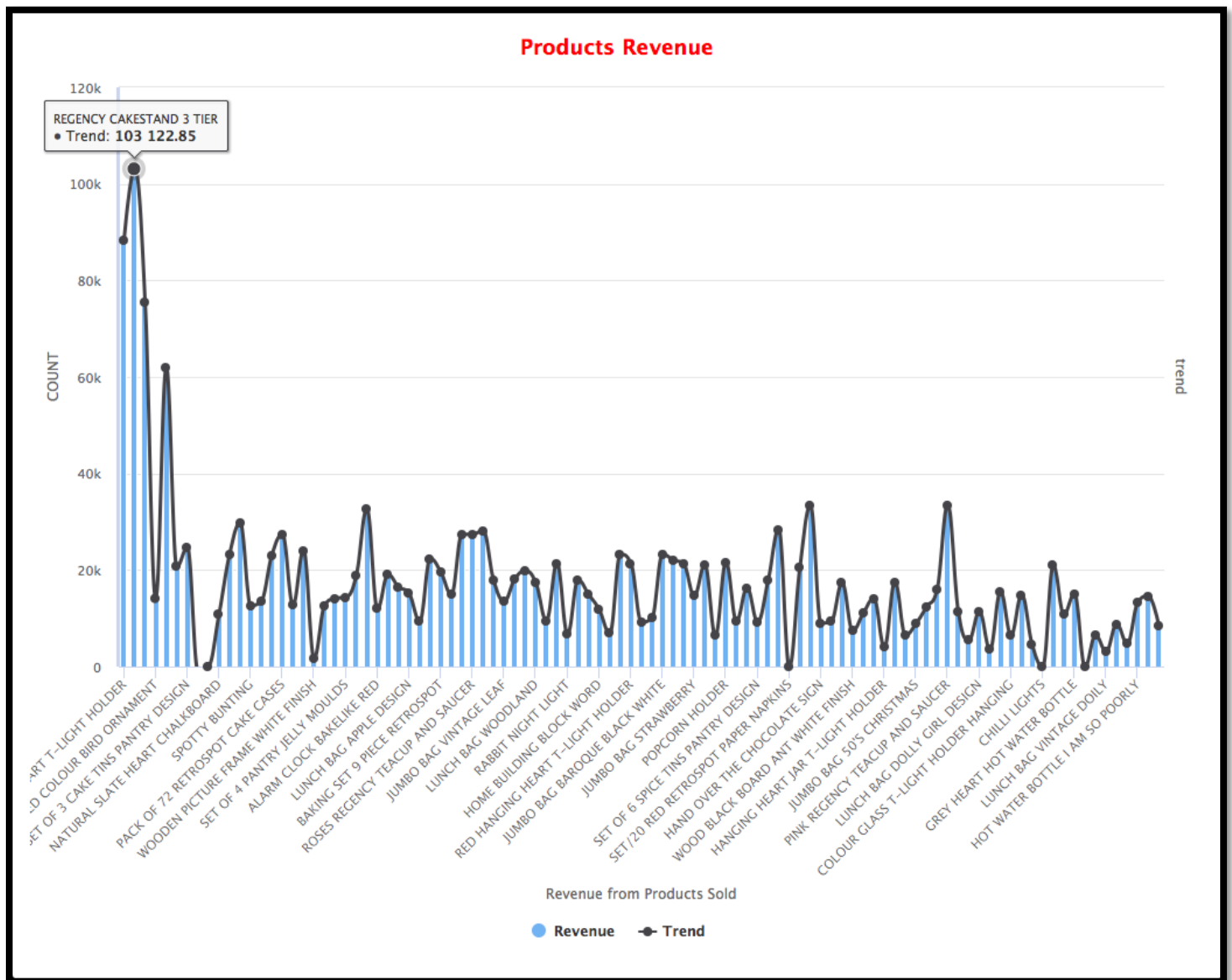
Similarly, the same steps were processed for 100 different products:

```
uk_pdt$sales= c(total_sales,total_sales_rct,total_sales_jbr,total_sales_acb,total_sales_pb,total_sales_lbr,
total_sales_sct,total_sales_post,total_sales_lbb,total_sales_ptc,total_sales_pck,total_sales_sb,
total_sales_lbs,total_sales_lbc,total_sales_nsh,total_sales_hws,total_sales_lbp,total_sales_jms,
total_sales_rcc,total_sales_lbs,total_sales_acb,total_sales_lba,total_sales_wpf,total_sales_jbpb,total_sales_spj,
total_sales_bsp,total_sales_rbp,total_sales_jmsp,total_sales_lbw,total_sales_rnl,total_sales_rrt,total_sales_rts,
total_sales_vgh,total_sales_wfa,total_sales_hwl,total_sales_abg,total_sales_rrl,total_sales_jsv,
total_sales_jsbs,total_sales_jbv,total_sales_gkp,total_sales_jbp,total_sales_rrcb,total_sales_srct,
total_sales_grt,total_sales_rrc,total_sales_rpn,total_sales_kc,total_sales_jbp,total_sales_vsc,total_sales_slb,
total_sales_pck,total_sales_opm,total_sales_chw,total_sales_hbbw,total_sales_rhh,total_sales_tfc,total_sales_ph,
total_sales_ppcc,total_sales_stp,total_sales_lbad,total_sales_sdh,total_sales_jbs,total_sales_jbb,total_sales_gtg,
total_sales_dke,total_sales_cpr,total_sales_gkp,total_sales_jbv,total_sales_dgl,total_sales_iks,total_sales_acb,
total_sales_rsb,total_sales_ctw,total_sales_lbd,total_sales_lbv,total_sales_prt,total_sales_jbpv,total_sales_wbb,
total_sales_gcr,total_sales_cbsd,total_sales_hhj,total_sales_pit,total_sales_jbc,total_sales_swh,total_sales_hwo,
total_sales_pts,total_sales_hocs,total_sales_drr,total_sales_lbbw,total_sales_pbr,total_sales_rhib,total_sales_gmcc,
total_sales_pnpd,total_sales_sctb,total_sales_pdcc,total_sales_wdcw,total_sales_astg,total_sales_lbv)
```

```
highchart() %>% hc_title(text = "<b>Products Sold in United Kingdom</b>",
margin = 20, align = "center",
style = list(color = "red", useHTML = TRUE)) %>%
hc_yAxis_multiples(
list(lineWidth = 3,title = list(text = "COUNT"),min=0),
list(showLastLabel = FALSE, opposite = TRUE,title = list(text = "trend"))
) %>%
hc_xAxis(title=list(text = "Number of Products Sold" ),categories = uk_pdt$X) %>%
hc_add_series(name = "products", data = uk_pdt$sales ,type = "column") %>%
hc_add_series(name = "Trend", data = uk_pdt$sales ,type = "spline") %>%
hc_plotOptions(series = list(stacking = FALSE)) %>%
hc_chart(type = "column")
```


- Let us further analyze the Revenue trend in United Kingdom, to understand which product is generating maximum revenue based on the number of products sold.
- We have Selected the Top 100 Frequently bought items in the market for this particular analysis.

REVENUE GENERATED BY PRODUCTS IN UNITED KINGDOM:



- From this above graph, we can determine that the highest grossing product would be REGENCY CAKE STAND 3 TIERS. This particular product generated around 103,122.85(currency) and is the 2nd most frequently bought item in the United Kingdom. WHITE HANGING HEART T-LIGHT HOLDER closely follows this product grossing around 90,000 (currency). This product was the most popular product bought in United Kingdom

R Code:

```
#### White Hanging Heart T- Light Holder
whh <- uk[uk$Description=="WHITE HANGING HEART T-LIGHT HOLDER",]
total_sales_whh <- sum(whh$Quantity)
total_revenue_whh <- sum(whh$Quantity*whh$UnitPrice)
total_sales

## Regency##
rct <- uk[uk$Description=="REGENCY CAKESTAND 3 TIER",]
total_sales_rct <- sum(rct$Quantity)
total_revenue_rct <- sum(rct$Quantity*rct$UnitPrice)

####JUMBO BAG RED RETROSPOT####
jbr <- uk[uk$Description=="JUMBO BAG RED RETROSPOT",]
total_sales_jbr <- sum(jbr$Quantity)
total_revenue_jbr <- sum(jbr$Quantity*jbr$UnitPrice)

##ASSORTED COLOUR BIRD ORNAMENT##
acb <- uk[uk$Description=="ASSORTED COLOUR BIRD ORNAMENT",]
total_sales_acb <- sum(acb$Quantity)
total_revenue_acb <- sum(acb$Quantity*acb$UnitPrice)

##PARTY BUNTING##
pb <- uk[uk$Description=="PARTY BUNTING",]
total_sales_pb <- sum(pb$Quantity)
total_revenue_pb <- sum(pb$Quantity*pb$UnitPrice)
####LUNCH BAG RED RETROSPOT
lbr <- uk[uk$Description=="LUNCH BAG RED RETROSPOT",]
total_sales_lbr <- sum(lbr$Quantity)
total_revenue_lbr <- sum(lbr$Quantity*lbr$UnitPrice)

####SET OF 3 CAKE TINS PANTRY DESIGN####
sct <- uk[uk$Description=="SET OF 3 CAKE TINS PANTRY DESIGN",]
total_sales_sct <- sum(sct$Quantity)
total_revenue_sct <- sum(sct$Quantity*sct$UnitPrice)

####POSTAGE##
post <- uk[uk$Description=="POSTAGE",]
total_sales_post <- sum(post$Quantity)
total_revenue_post <- sum(post$Quantity*post$UnitPrice)
```

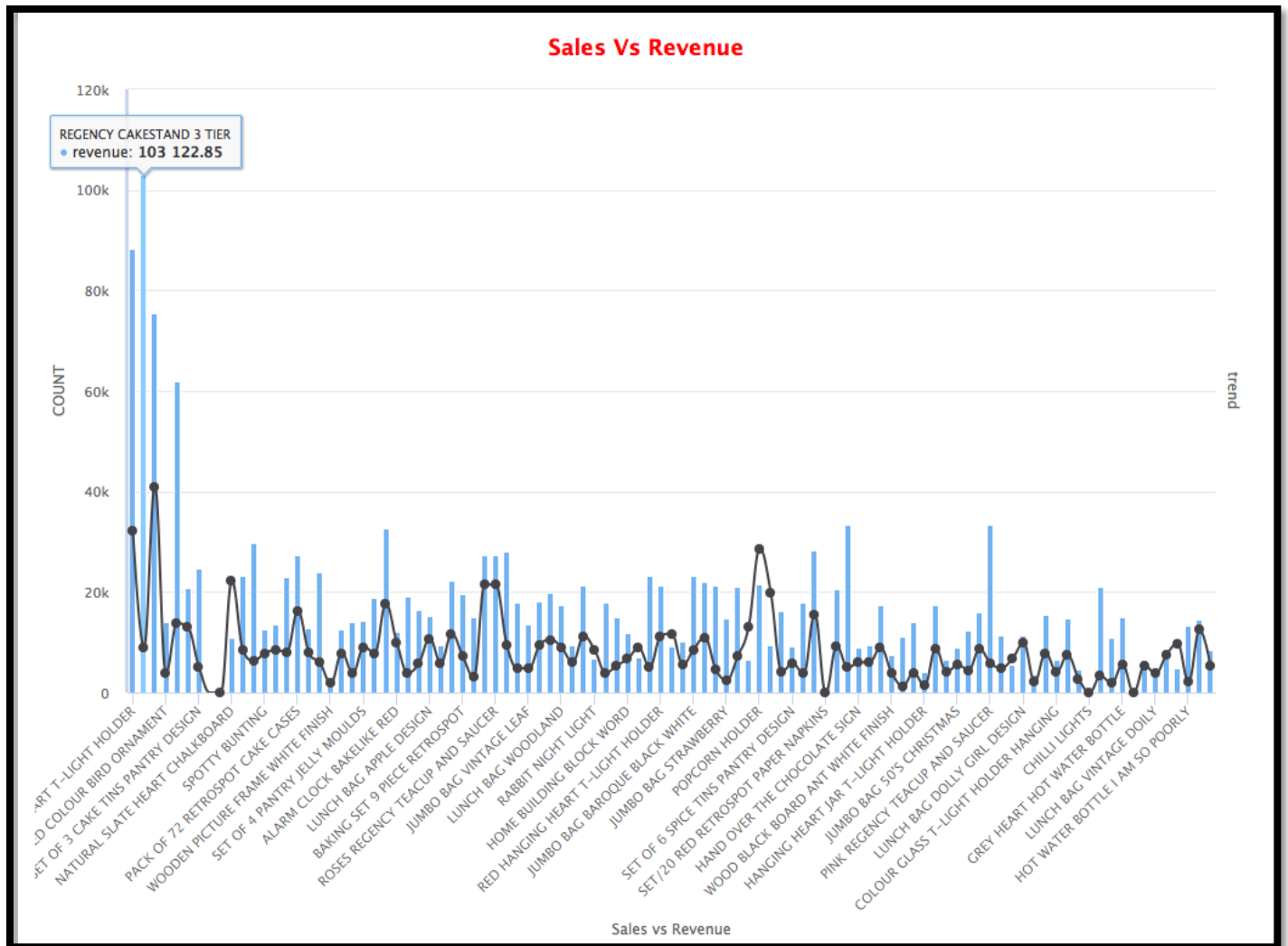
- Similarly, we calculate revenue for top 100 frequently bought items in United Kingdom:

```
k_pdt$revenue <- c(total_revenue,total_revenue_rct,total_revenue_jbr,total_revenue_acb,total_revenue_pb,total_revenue_lbr,
total_revenue_sct,total_revenue_post,total_revenue_lbb,total_revenue_ptc,total_revenue_pck,total_revenue_sb,
total_revenue_lbs,total_revenue_lbc,total_revenue_nsh,total_revenue_hws,total_revenue_lbp,total_revenue_jms,
total_revenue_rcc,total_revenue_lbs,total_revenue_acb,total_revenue_lba,total_revenue_wpf,total_revenue_jbpb,total_revenue_spj,
total_revenue_bsp,total_revenue_rbp,total_revenue_jmsp,total_revenue_lbw,total_revenue_rnl,total_revenue_rrt,total_revenue_rts,
total_revenue_vgh,total_revenue_wfa,total_revenue_hwl,total_revenue_abg,total_revenue_rrl,total_revenue_jsv,
total_revenue_jsbs,total_revenue_jbv,total_revenue_gkp,total_revenue_jbp,total_revenue_rrcb,total_revenue_srct,
total_revenue_grt,total_revenue_rrc,total_revenue_rpn,total_revenue_kc,total_revenue_jbp,total_revenue_vsc,total_revenue_slb,
total_revenue_pck,total_revenue_opm,total_revenue_chw,total_revenue_hbbw,total_revenue_rhh,total_revenue_tfc,total_revenue_ph,
total_revenue_ppcc,total_revenue_stp,total_revenue_lbad,total_revenue_sdh,total_revenue_jbs,total_revenue_jbb,total_revenue_gtg,
total_revenue_dkc,total_revenue_cpr,total_revenue_gkp,total_revenue_jbv,total_revenue_dgl,total_revenue_iks,total_revenue_acb,
total_revenue_rsb,total_revenue_ctw,total_revenue_lbd,total_revenue_lbv,total_revenue_prt,total_revenue_jbpv,total_revenue_wbb,
total_revenue_gcr,total_revenue_cbsd,total_revenue_hhj,total_revenue_pit,total_revenue_jbc,total_revenue_swh,total_revenue_hwo,
total_revenue_pts,total_revenue_hocs,total_revenue_drr,total_revenue_lbbw,total_revenue_pbr,total_revenue_rhib,total_revenue_gmcc,
total_revenue_pnpd,total_revenue_sctb,total_revenue_pdcc,total_revenue_wdcw,total_revenue_astg,total_revenue_lbv)
```

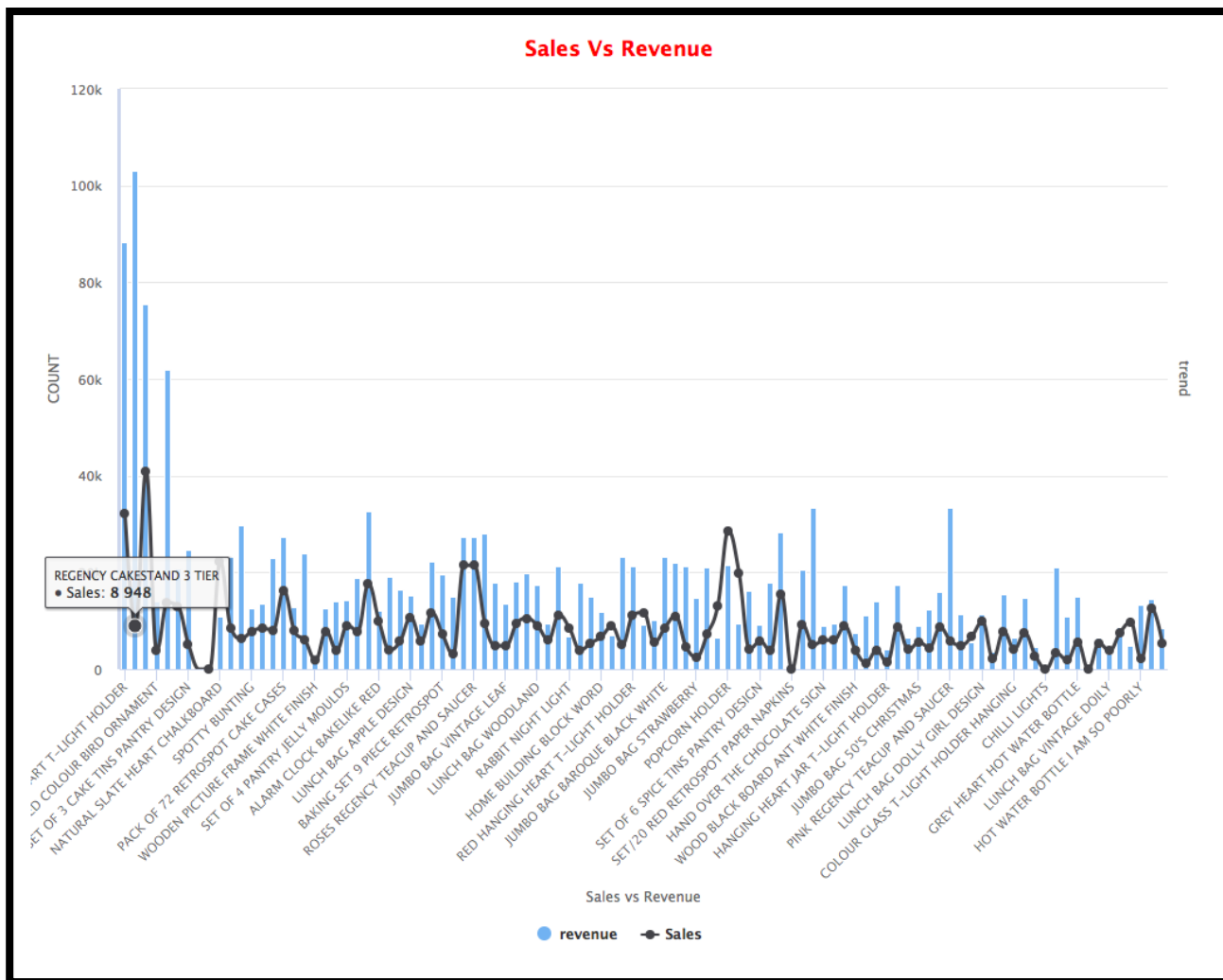
```
# total revenue by 100 products in UK
highchart() %>% hc_title(text = "<b>Products Revenue </b>",
margin = 20, align = "center",
style = list(color = "red", useHTML = TRUE)) %>%
hc_yAxis_multiples(
list(lineWidth = 3,title = list(text = "COUNT"),min=0),
list(showLastLabel = FALSE, opposite = TRUE,title = list(text = "trend"))
) %>%
hc_xAxis(title=list(text = "Revenue from Products Sold" ),categories = uk_pdt$X) %>%
hc_add_series(name = "Revenue", data = uk_pdt$revenue ,type = "column") %>%
hc_add_series(name = "Trend", data = uk_pdt$revenue ,type = "spline") %>%
hc_plotOptions(series = list(stacking = FALSE)) %>%
hc_chart(type = "column")
```

- Now that we have both sales and Revenue generated by different products, let us compare both of them to see how much sales have resulted in how much revenue.

SALES VS REVENUE FOR PRODUCTS IN UNITED KINGDOM:



- In this particular we can notice that the product generating the highest income would be the REGENCY CAKESTAND 3-TIER followed by WHITE HANGING HEART T-LIGHT HOLDER.
- Let us, see how many products have been sold for each of this item.



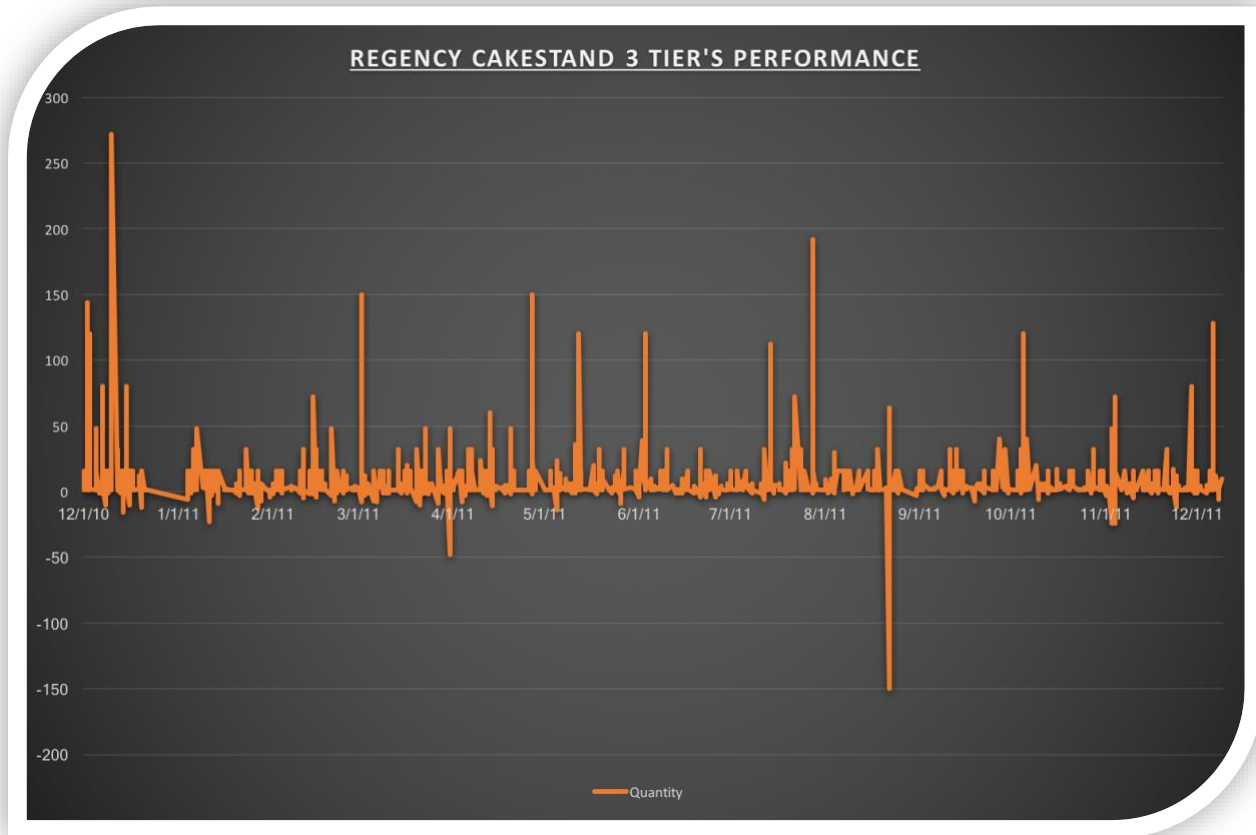
- From this graph, we can notice that REGENCY CAKESTAND 3 TIER has very less quantities sold. This product has been generating the highest revenue for the company in United Kingdom with very few sales and it is the second most frequently bought object in United Kingdom. The company can focus on products like this to increase the number of quantities sold and thereby the increase the revenue.

R CODE:

```
highchart() %>% hc_title(text = "<b>Sales Vs Revenue </b>",
  margin = 20, align = "center",
  style = list(color = "red", useHTML = TRUE)) %>%
  hc_yAxis_multiples(
    list(lineWidth = 3, title = list(text = "COUNT"), min=0),
    list(showLastLabel = FALSE, opposite = TRUE, title = list(text = "trend"))
  ) %>%
  hc_xAxis(title=list(text="Sales vs Revenue"), categories = uk_pdt$X) %>%
  hc_add_series(name = "revenue", data = uk_pdt$revenue, type = "column") %>%
  hc_add_series(name = "Sales", data = uk_pdt$sales, type = "spline") %>%
  hc_plotOptions(series = list(stacking = FALSE)) %>%
  hc_chart(type = "column")
```

- Now that we know REGENCY CAKESTAND 3TIER is the highest grossing product, let us look at how the product has been sold over the period of time. This analysis will let us understand if there is any seasonality or trend so that it could be used for promotion and boosting the sales number for the product

TIME SERIES ANALYSIS FOR HIGHEST REVENUE GENERATING PRODUCT:



- This particular Time series Analysis lets us determine the quantities sold over a period of time in United Kingdom. With this analysis, we can see which part of the year or month has maximum sales and which part has minimum. It will let us focus those regions to make promotional activities. From the above graph, we can observe that during the month of December 2010, there has been a spike in the quantities sold. The reason could be because of holiday season and Christmas and new year coming up.

R code:

```
btprdt <- uk[uk$Description=="REGENCY CAKESTAND 3 TIER",]
#View(btprdt)

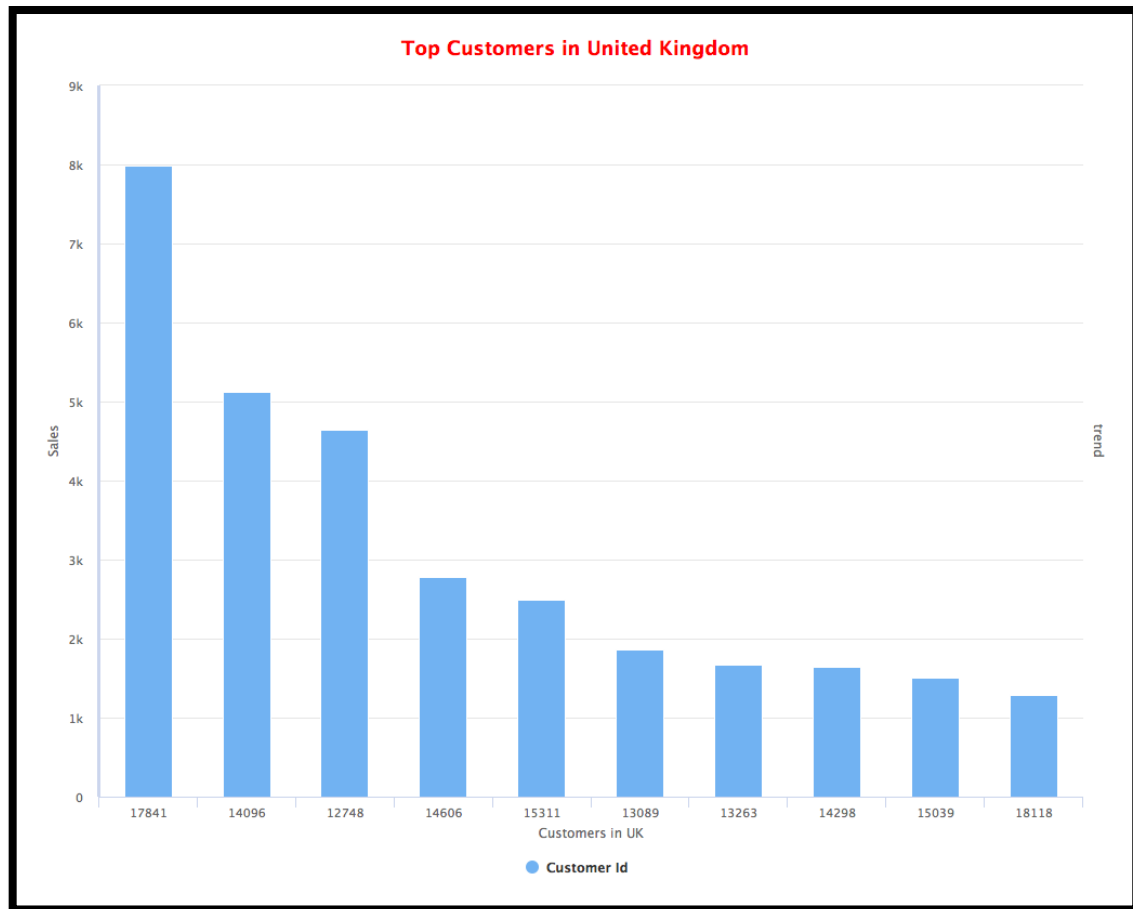
sur <- data.frame(btprdt$InvoiceDate,btprdt$Quantity)
View(sur)

sur$btprdt.InvoiceDate = as.Date(sur$btprdt.InvoiceDate, "%m/%d/%Y %H:%M")

## time series Analysis of Product over a period of time
ts <- ggplot(sur,aes(x=sur$btprdt.InvoiceDate,y=sur$btprdt.Quantity),col= sur$btprdt.Quantity ) +geom_line(aes(fill=sur$btprdt.Quantity))
ts<- ts+theme (axis.text.x = element_text(angle = 90,hjust = 1))
ts <- ts+labs(x="Product Name",y="Sold")
ts
```

- Now, let us focus on the Customers in United Kingdom. We use the Customer Id has a unique code for recognizing the customers.

LARGEST BUYER IN THE COUNTRY



- In this analysis, we focus on the top 10 buyers in the United Kingdom. We can notice that Customer with Customer ID 17841 has bought maximum number of quantities. This analysis could help us understand potential customers and very good customers to offer great deals and discounts to increase the sales and attract new customers to buy more.

R Code:

```
cust_tt <- cust[1:10,]

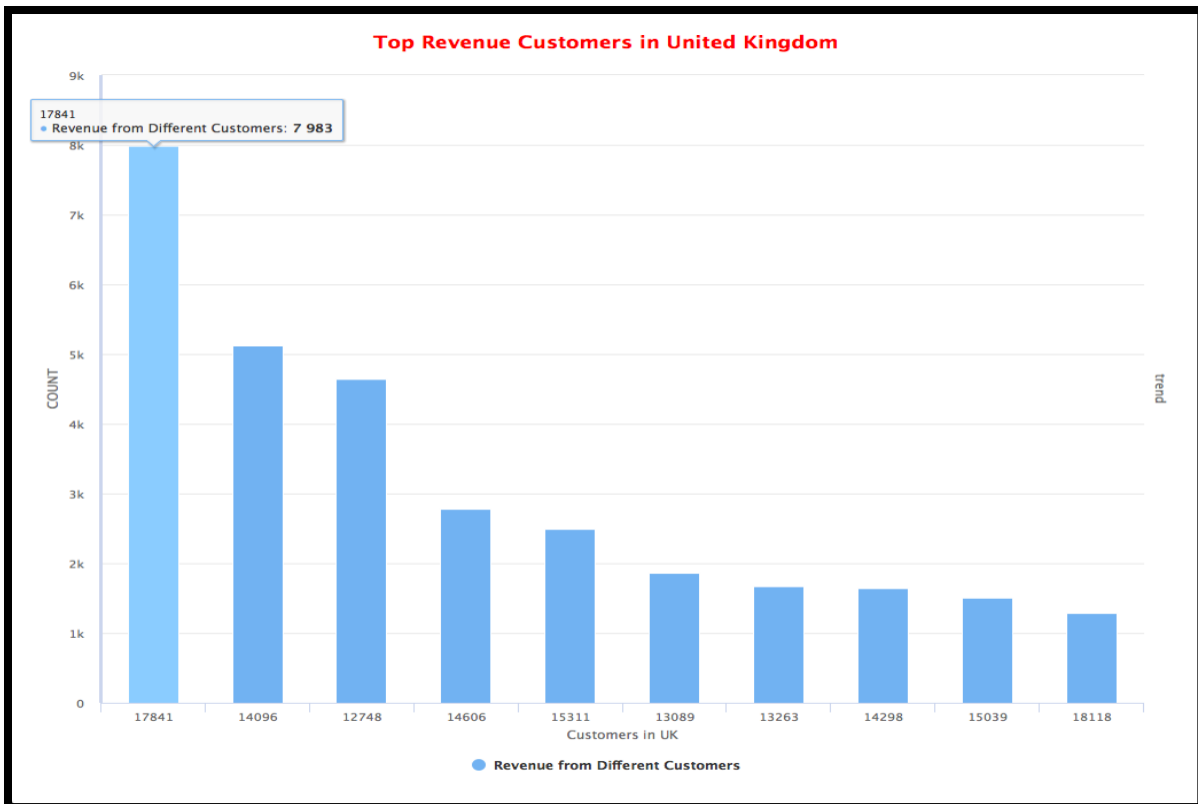
cc <- ggplot(cust_tt,aes(x=cust_tt$CustomerID,y=cust_tt$`Count of Quantity`)) +geom_col(aes(fill=cust_tt$`Count of Quantity`))
cc<- cc+theme (axis.text.x = element_text(angle = 90,hjust = 1))
cc <- cc+labs(x="Customer Id ",y="Quantity Bought by the Customer")
cc

highchart() %>% hc_title(text = "<b>Top Customers in United Kingdom</b>",
                        margin = 20, align = "center",
                        style = list(color = "red", useHTML = TRUE)) %>%

  hc_yAxis_multiples(
    list(lineWidth = 3,title = list(text = "Sales"),min=0),
    list(showLastLabel = FALSE, opposite = TRUE,title = list(text = "trend"))
  ) %>%
  hc_xAxis(title=list(text = "Customers in UK" ),categories = cust_tt$CustomerID) %>%
  hc_add_series(name = "Customer Id", data = cust_tt$`Count of Quantity` ,type = "column") %>%
  hc_plotOptions(series = list(stacking = FALSE)) %>%
  hc_chart(type = "column")
```

LARGEST REVENUE GENERATOR FOR THE COMPANY IN UNITED KINGDOM:

- In this analysis, we will focus on which customer produces the largest income for the company in United Kingdom.



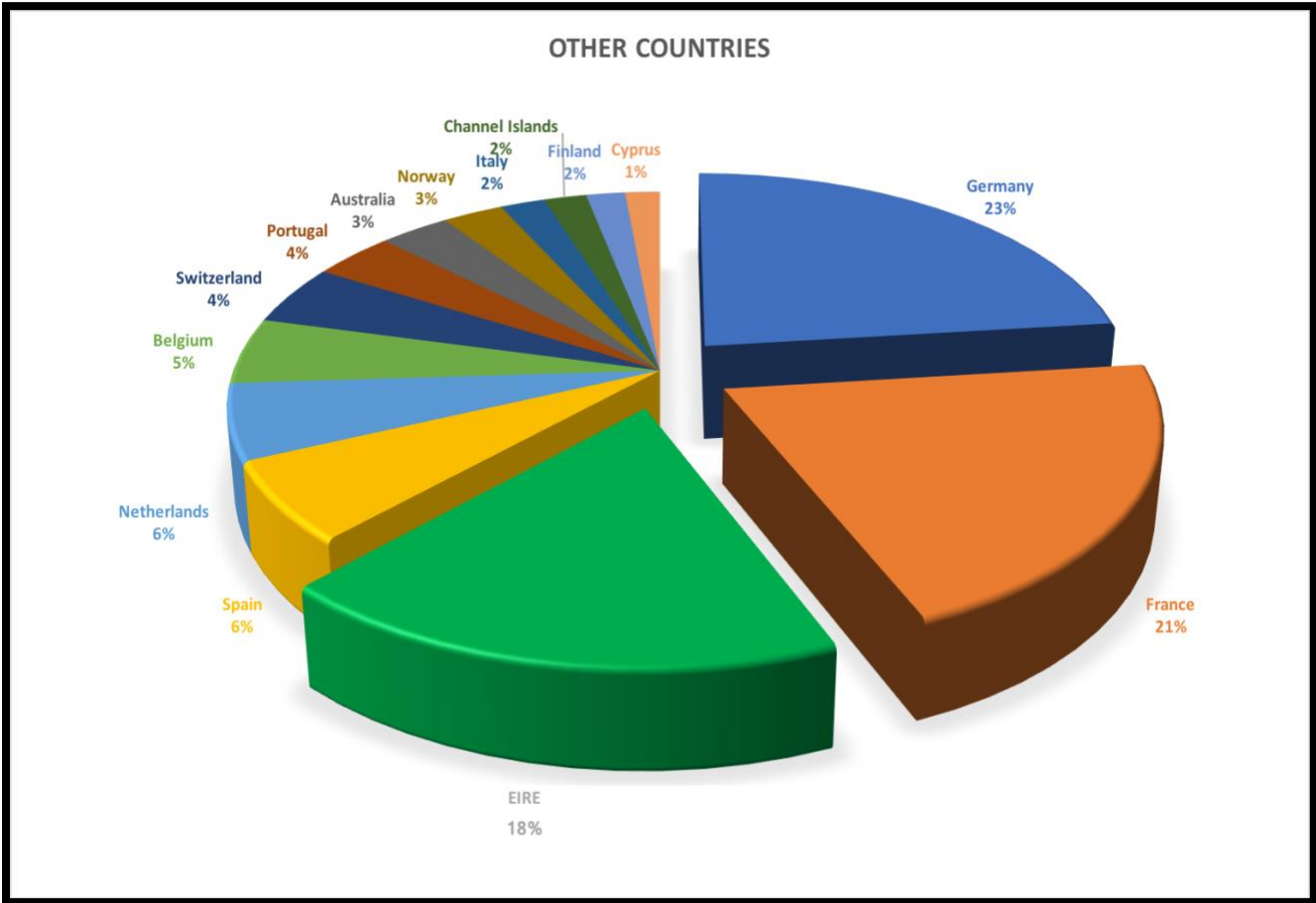
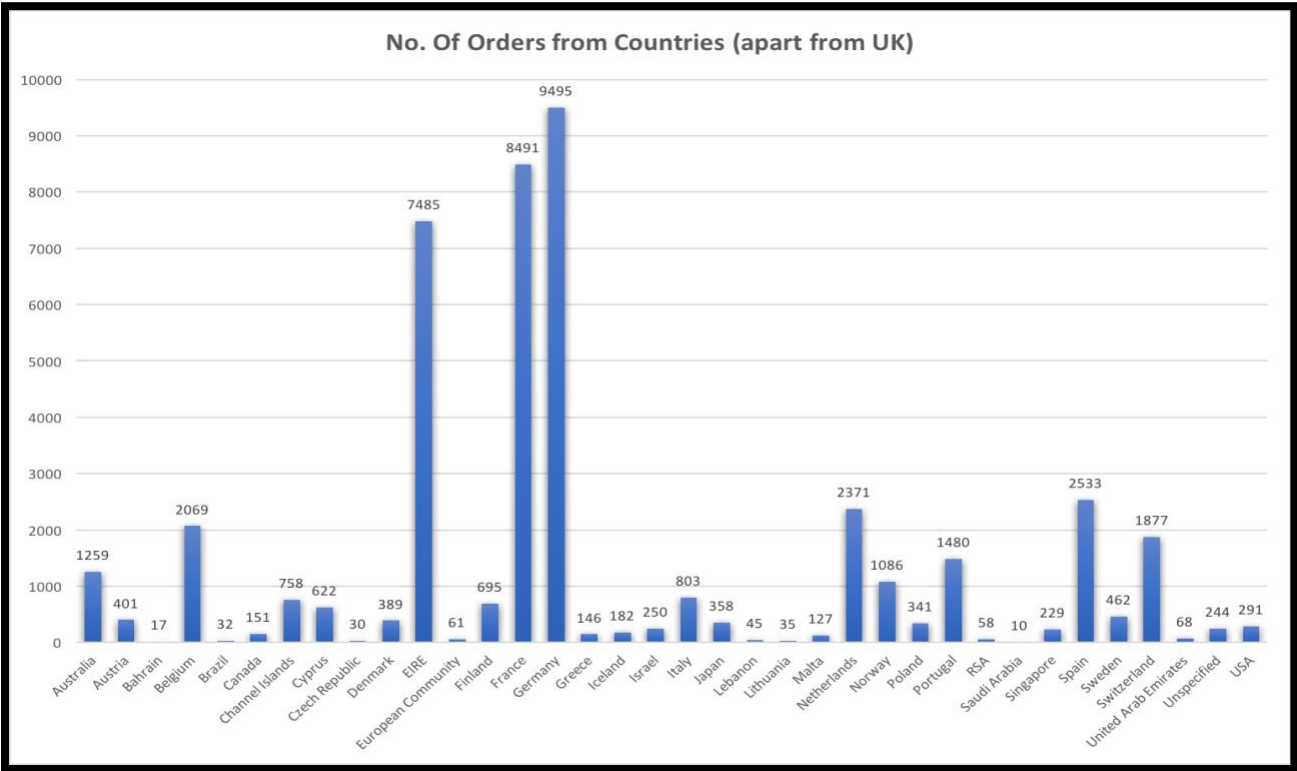
- From this graph, we can determine that Customer with Customer ID 17841 is the highest revenue generator for the company. Based on the previous analysis, Customer with Customer ID 17841 is the largest buyer and highest revenue generating customer for the company in United Kingdom.

R Code:

```
mydata$Revenue = mydata$Quantity*mydata$UnitPrice
revenue_cust = dataset
revenue_cust <- revenue_cust[order(-revenue_cust`Count of Revenue`),]

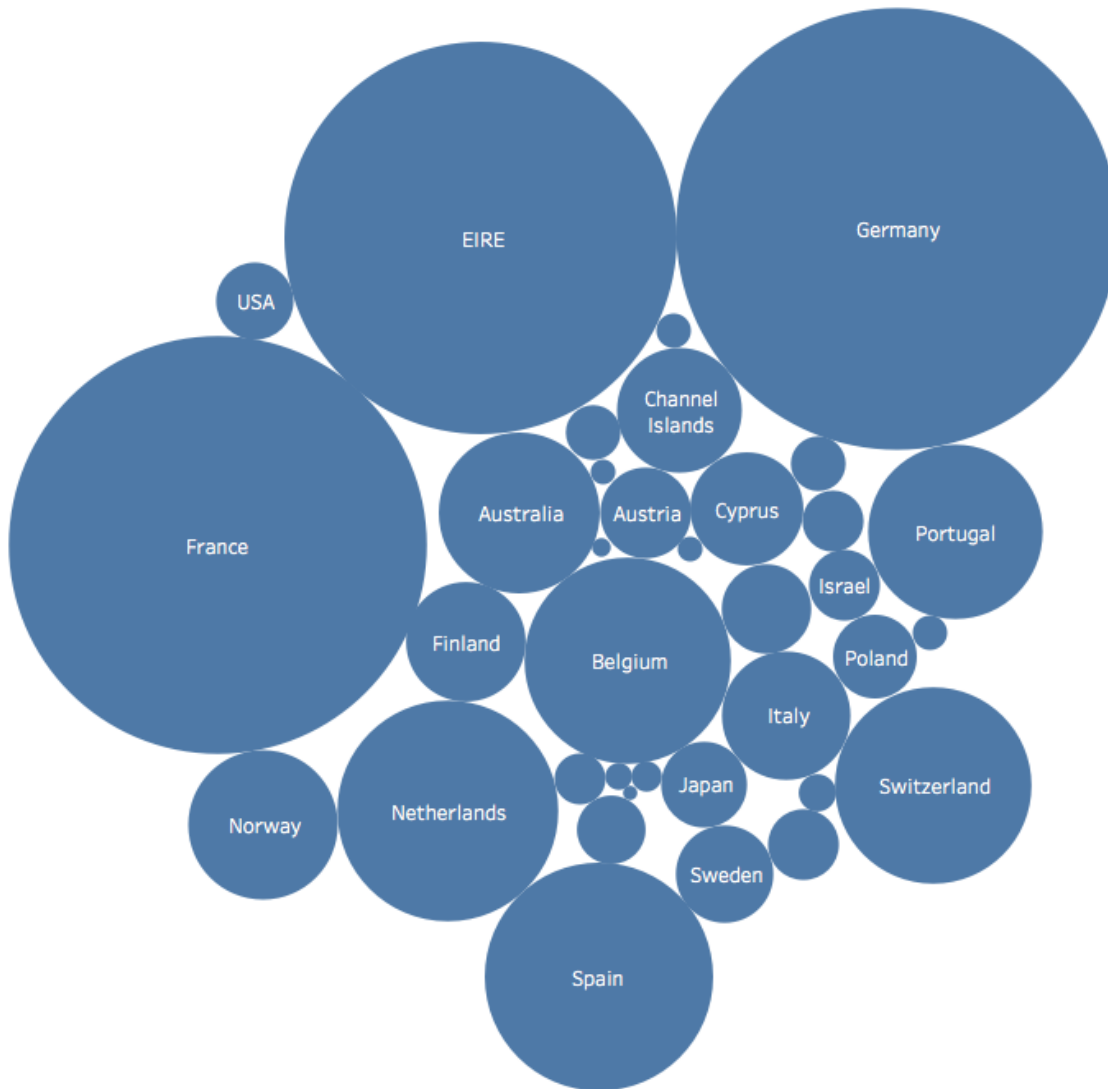
revenue <- revenue_cust[1:11,]
revenue <- revenue[-1,]
highchart() %>% hc_title(text = "<b>Top Revenue Customers in United Kingdom</b>",
                        margin = 20, align = "center",
                        style = list(color = "red", useHTML = TRUE)) %>%
  hc_yAxis_multiples(
    list(lineWidth = 3,title = list(text = "COUNT"),min=0),
    list(showLastLabel = FALSE, opposite = TRUE,title = list(text = "trend"))
  ) %>%
  hc_xAxis(title=list(text = "Customers in UK" ),categories = revenue$X1) %>%
  hc_add_series(name = "Revenue from Different Customers", data = revenue`Count of Revenue`,type = "column") %>%
  hc_plotOptions(series = list(stacking = FALSE)) %>%
  hc_chart(type = "column")
```


GLOBAL MARKET ANALYSIS: Let us focus on the market in other countries



- From the above graph, we can determine that apart United Kingdom the company has substantial market Germany, France, EIRE. From this analysis, the company can focus on which countries to expand their market

Sheet 2



- This is an interactive bubble graph to view the market trend in other countries apart from United Kingdom.

Tools: The above graphs were generated in Tableau for interactive visualizations.