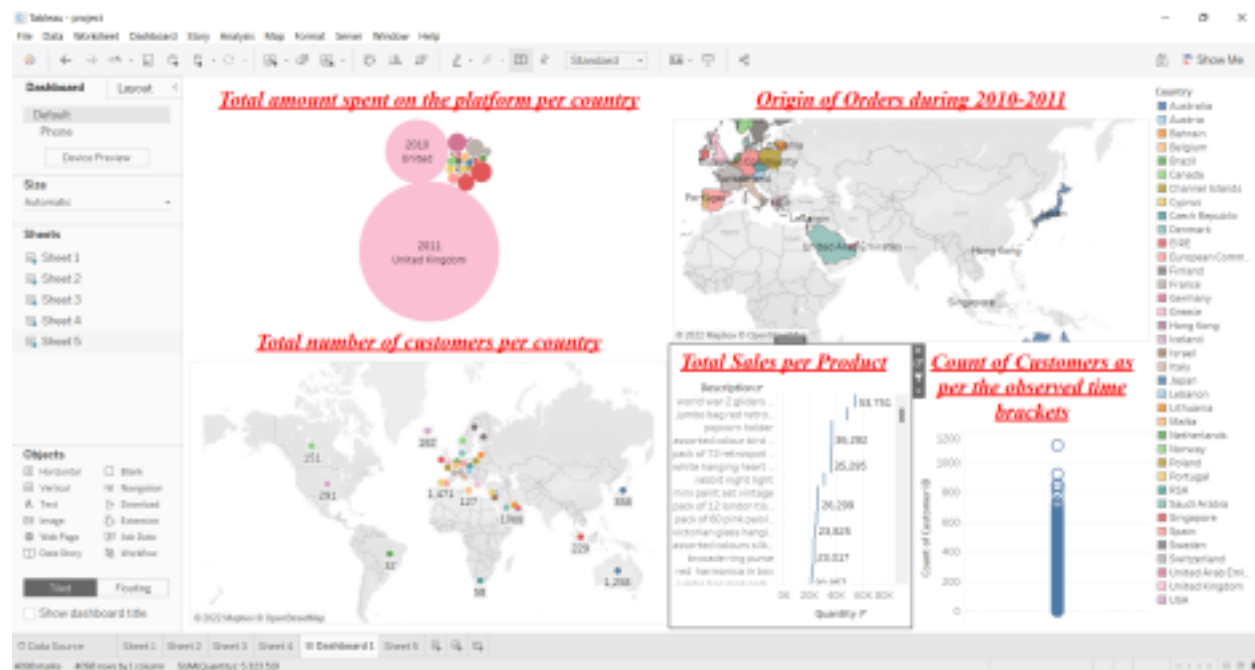


RAHUL SINGH THAKUR

**TITLE:** E-COMMERCE SALES DATA SCRAPED FROM AMAZON API DURING THE YEAR 2010-2011

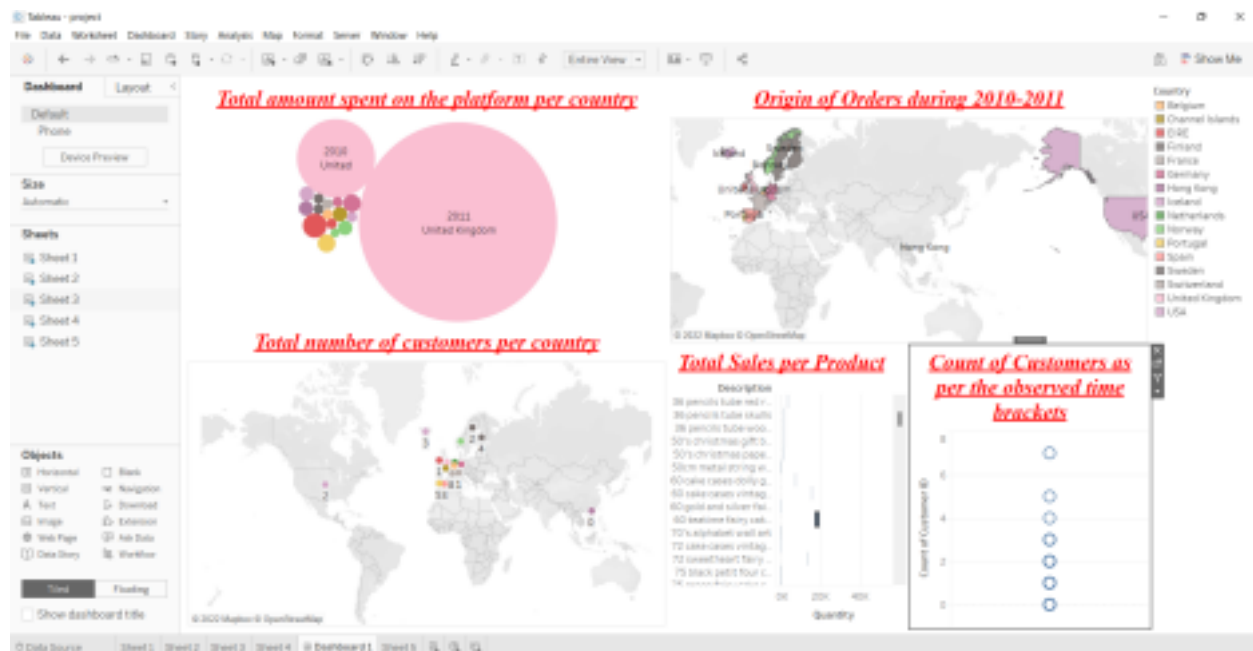
## **Pictures of Dashboard**



*Since this dashboard is interactive, we can have deeper insights if we look into the “Total Sales per Product” graph.*

*Upon clicking any illustration mapped to a certain product-sale group, the entire dashboard transforms to display information regarding that very product only.*

*This page shows the data for the product “60 teatime fairy cake cases”*



## R code used to clean the dataset.

```
##RAHUL SINGH THAKUR
##102017178
##3CSE-8
##DATA SCIENCE PROJECT

##Data Pre-Processing

library(ggplot2)
library(data.table)

setwd("C:\\users\\rahu1\\OneDrive\\Desktop\\3rd Year\\Data sci dashboard")

##data = read.csv("data.csv", encoding="UTF-8")

sales_data<-fread("C:\\users\\rahu1\\OneDrive\\Desktop\\3rd Year\\Data sci dashboard\\data.csv")
view(sales_data)

R 4.2.1 - C:/Users/rahu1/OneDrive/Desktop/3rd Year/Data sci dashboard/
> library(ggplot2)
> library(data.table)
> setwd("C:\\Users\\rahu1\\OneDrive\\Desktop\\3rd Year\\Data sci dashboard")
> sales_data<-fread("C:\\users\\rahu1\\OneDrive\\Desktop\\3rd Year\\Data sci dashboard\\data.csv")
> view(sales_data)
>
```

## Initial state of dataset

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	536385	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850	United Kingdom
2	536385	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850	United Kingdom
3	536385	844066	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850	United Kingdom
4	536385	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850	United Kingdom
5	536385	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850	United Kingdom
6	536385	22752	SET 7 BABUSHKA NESTING BOXES	2	12/1/2010 8:26	7.65	17850	United Kingdom
7	536385	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12/1/2010 8:26	4.25	17850	United Kingdom
8	536386	22633	HAND WARMER UNION JACK	6	12/1/2010 8:28	1.85	17850	United Kingdom
9	536386	22632	HAND WARMER RED POLKA DOT	6	12/1/2010 8:28	1.85	17850	United Kingdom
10	536387	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/1/2010 8:34	1.69	13047	United Kingdom
11	536387	22146	PORPY'S PLAYHOUSE BEDROOM	6	12/1/2010 8:34	2.10	13047	United Kingdom
12	536387	22148	PORPY'S PLAYHOUSE KITCHEN	6	12/1/2010 8:34	2.10	13047	United Kingdom
13	536387	22149	RESCRAFT PRINCESS CHARLOTTE DOLL	8	12/1/2010 8:34	5.75	13047	United Kingdom
14	536387	22310	NOIR KNITTED MUG COZY	6	12/1/2010 8:34	1.85	13047	United Kingdom
15	536387	84989	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	12/1/2010 8:34	4.25	13047	United Kingdom
16	536387	23623	BOX OF VINTAGE AGSAW BLOCKS	3	12/1/2010 8:34	4.95	13047	United Kingdom
17	536387	23622	BOX OF VINTAGE ALPHABET BLOCKS	2	12/1/2010 8:34	9.95	13047	United Kingdom
18	536387	21184	HOME BUILDING BLOCK WORD	3	12/1/2010 8:34	5.95	13047	United Kingdom
19	536387	21186	JOVE BUILDING BLOCK WORD	3	12/1/2010 8:34	5.95	13047	United Kingdom
20	536387	21177	RECIPE BOX WITH METAL HEART	4	12/1/2010 8:34	7.89	13047	United Kingdom
21	536387	48187	DOORMAT NEW ENGLAND	4	12/1/2010 8:34	7.95	13047	United Kingdom
22	536388	22980	JAM MAKING SET WITH JARS	6	12/1/2010 8:34	4.25	13047	United Kingdom
23	536388	22913	RED COAT RACK PARIS FASHION	3	12/1/2010 8:34	4.95	13047	United Kingdom

```
head(sales_data,10)
```

```
tail(sales_data,10)
```

```
##We don't have a use for the invoice number and stock code as they don't provide any
##relevant information and do not aid our analysis.
```

```
> head(sales_data,10)
  InvoiceNo StockCode Description Quantity InvoiceDate UnitPrice CustomerID Country
1:  536385  85123A  WHITE HANGING HEART T-LIGHT HOLDER      6 12/1/2010 8:26      2.55  17850 united kingdom
2:  536385  71053    WHITE METAL LANTERN                    6 12/1/2010 8:26      3.39  17850 united kingdom
3:  536385  844066  CREAM CUPID HEARTS COAT HANGER           8 12/1/2010 8:26      2.75  17850 united kingdom
4:  536385  84029G  KNITTED UNION FLAG HOT WATER BOTTLE      6 12/1/2010 8:26      3.39  17850 united kingdom
5:  536385  84029E  RED WOOLLY HOTTIE WHITE HEART.              6 12/1/2010 8:26      3.39  17850 united kingdom
6:  536385  22752    SET 7 BABUSHKA NESTING BOXES              2 12/1/2010 8:26      7.65  17850 united kingdom
7:  536385  21730  GLASS STAR FROSTED T-LIGHT HOLDER           6 12/1/2010 8:26      4.25  17850 united kingdom
8:  536386  22633    HAND WARMER UNION JACK                6 12/1/2010 8:28      1.85  17850 united kingdom
9:  536386  22632    HAND WARMER RED POLKA DOT                6 12/1/2010 8:28      1.85  17850 united kingdom
10: 536387  84879  ASSORTED COLOUR BIRD ORNAMENT                    32 12/1/2010 8:34      1.69  13047 united kingdom
> tail(sales_data,10)
  InvoiceNo StockCode Description Quantity InvoiceDate UnitPrice CustomerID Country
1:  581587  22726  ALARM CLOCK BAKELIKE GREEN                4 12/9/2011 12:50      1.75  12680 France
2:  581587  22730  ALARM CLOCK BAKELIKE IVORY                 4 12/9/2011 12:50      1.75  12680 France
3:  581587  22367  CHILDRENS APRON SPACEBOY DESIGN             8 12/9/2011 12:50      1.95  12680 France
4:  581587  22629  SPACEBOY LUNCH BOX                          12 12/9/2011 12:50      1.95  12680 France
5:  581587  22326  CHILDRENS CUTLERY SPACEBOY                  4 12/9/2011 12:50      4.15  12680 France
6:  581587  22613  PACK OF 20 SPACEBOY NAPKINS                 12 12/9/2011 12:50      0.85  12680 France
7:  581587  22899  CHILDREN'S APRON DOLLY GIRL                 6 12/9/2011 12:50      2.10  12680 France
8:  581587  21254  CHILDRENS CUTLERY DOLLY GIRL                 4 12/9/2011 12:50      4.15  12680 France
9:  581587  23255  CHILDRENS CUTLERY CIRCUS PARADE              4 12/9/2011 12:50      4.15  12680 France
10: 581587  22138  BAKING SET 9 PIECE RETROSPOT                 3 12/9/2011 12:50      4.95  12680 France
> |
```

```
5
6 cols_remove<-c("InvoiceNo","StockCode")
7 sales_data[, (cols_remove) := NULL]
8 head(sales_data,10)
9
```

```
> cols_remove<-c("InvoiceNo","StockCode")
> sales_data[, (cols_remove) := NULL]
> head(sales_data,10)
  Description quantity InvoiceDate unitPrice customerID Country
1: WHITE HANGING HEART T-LIGHT HOLDER      6 12/1/2010 8:26      2.55  17850 United Kingdom
2: WHITE METAL LANTERN                    6 12/1/2010 8:26      3.39  17850 United Kingdom
3: CREAM CUPID HEARTS COAT HANGER           8 12/1/2010 8:26      2.75  17850 United Kingdom
4: KNITTED UNION FLAG HOT WATER BOTTLE      6 12/1/2010 8:26      3.39  17850 United Kingdom
5: RED WOOLLY HOTTIE WHITE HEART.              6 12/1/2010 8:26      3.39  17850 United Kingdom
6: SET 7 BABUSHKA NESTING BOXES              2 12/1/2010 8:26      7.65  17850 United Kingdom
7: GLASS STAR FROSTED T-LIGHT HOLDER           6 12/1/2010 8:26      4.25  17850 United Kingdom
8: HAND WARMER UNION JACK                6 12/1/2010 8:28      1.85  17850 United Kingdom
9: HAND WARMER RED POLKA DOT                6 12/1/2010 8:28      1.85  17850 United Kingdom
10: ASSORTED COLOUR BIRD ORNAMENT                    32 12/1/2010 8:34      1.69  13047 United Kingdom
> |
```

```

sales_data$description<- tolower(sales_data$description)

#try to split the invoicedate column to date and time separately

sales_data$Time <- format(as.POSIXct(sales_data$InvoiceDate,format="%Y:%m:%d %H:%M:%S"),"%H:%M:%S")

sales_data$Date <- format(as.POSIXct(sales_data$InvoiceDate,,format="%Y:%m:%d %H:%M:%S"),"%Y:%m:%d")

```

## *This now transforms our dataset*

	V1	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1:	1	white hanging heart t-light holder	6	12-01-2010 08:26	2.55	17850	united kingdom
2:	2	white metal lantern	6	12-01-2010 08:26	3.39	17850	united kingdom
3:	3	cream cupid hearts coat hanger	8	12-01-2010 08:26	2.75	17850	united kingdom
4:	4	knitted union flag hot water bottle	6	12-01-2010 08:26	3.39	17850	united kingdom
5:	5	red woolly hottie white heart.	6	12-01-2010 08:26	3.39	17850	united kingdom
6:	6	set 7 babushka nesting boxes	2	12-01-2010 08:26	7.65	17850	united kingdom
7:	7	glass star frosted t-light holder	6	12-01-2010 08:26	4.25	17850	united kingdom
8:	8	hand warmer union jack	6	12-01-2010 08:28	1.85	17850	united kingdom
9:	9	hand warmer red polka dot	6	12-01-2010 08:28	1.85	17850	united kingdom
10:	10	assorted colour bird ornament	32	12-01-2010 08:34	1.69	13047	united kingdom

## *Looking at the newly formed date and time column*

	Date	Time
12 January 2010	08:26:00	AM
12 January 2010	08:26:00	AM
12 January 2010	08:26:00	AM
12 January 2010	08:26:00	AM
12 January 2010	08:26:00	AM
12 January 2010	08:26:00	AM
12 January 2010	08:26:00	AM
12 January 2010	08:28:00	AM
12 January 2010	08:28:00	AM
12 January 2010	08:34:00	AM

## *Brief view of the number of NA values in our dataset*

```

> is.na(sales_data)

```

	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
[1,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[2,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[3,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[4,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[5,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[6,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[7,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[8,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[9,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[10,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[11,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[12,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[13,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[14,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[15,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[16,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[17,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[18,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[19,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

```
##REMOVE DUPLICATES ACROSS THE ENTIRE DATASET
library(dplyr)
sales_data <- sales_data %>% distinct()

write.csv(sales_data,"cleandata.csv")

> ##REMOVE DUPLICATES ACROSS THE ENTIRE DATASET
> library(dplyr)
> sales_data <- sales_data %>% distinct()
> write.csv(sales_data,"cleandata.csv")
|
```

### **RELATIONSHIP ESTABLISHED IN TABLEAU**



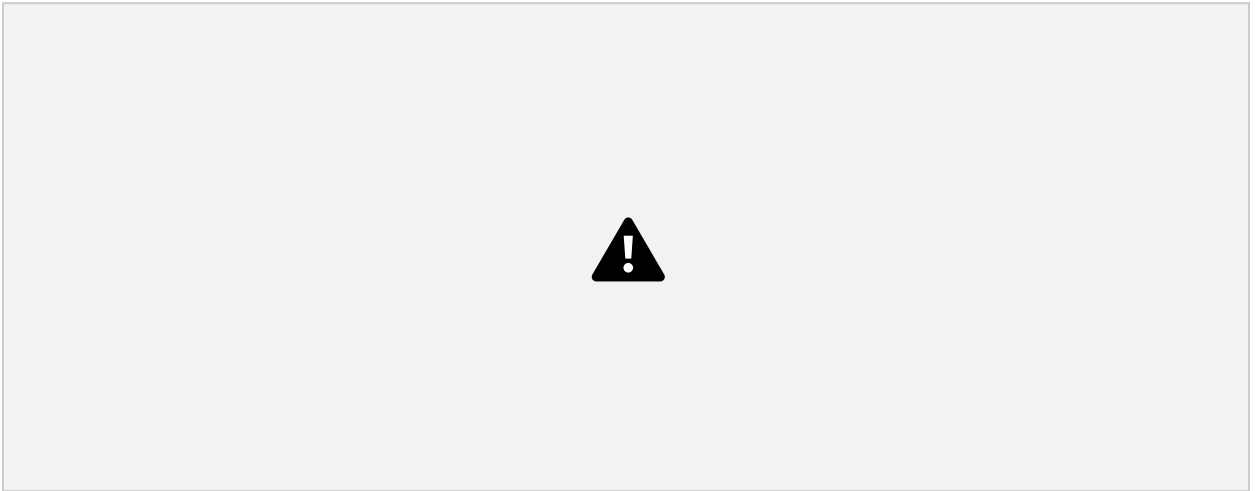
*This “ratings.csv” file contains data obtained from a survey where each customer was asked to give a rating based on the product price per unit. I did not add it to my dashboard as it wasn’t contributing towards deriving meaningful insights.*

**RATINGS.CSV HAS BEEN ADDED TO MY DATA SOURCES**



**Some Pictures of my final dataset “Cleandata.csv”**





*Some pictures of the “ratings.csv”*

*file*

