

# SALES TRANSACTION DATA ANALYSIS

---

- Sameer Kumar Mohapatra

## Data Overview:

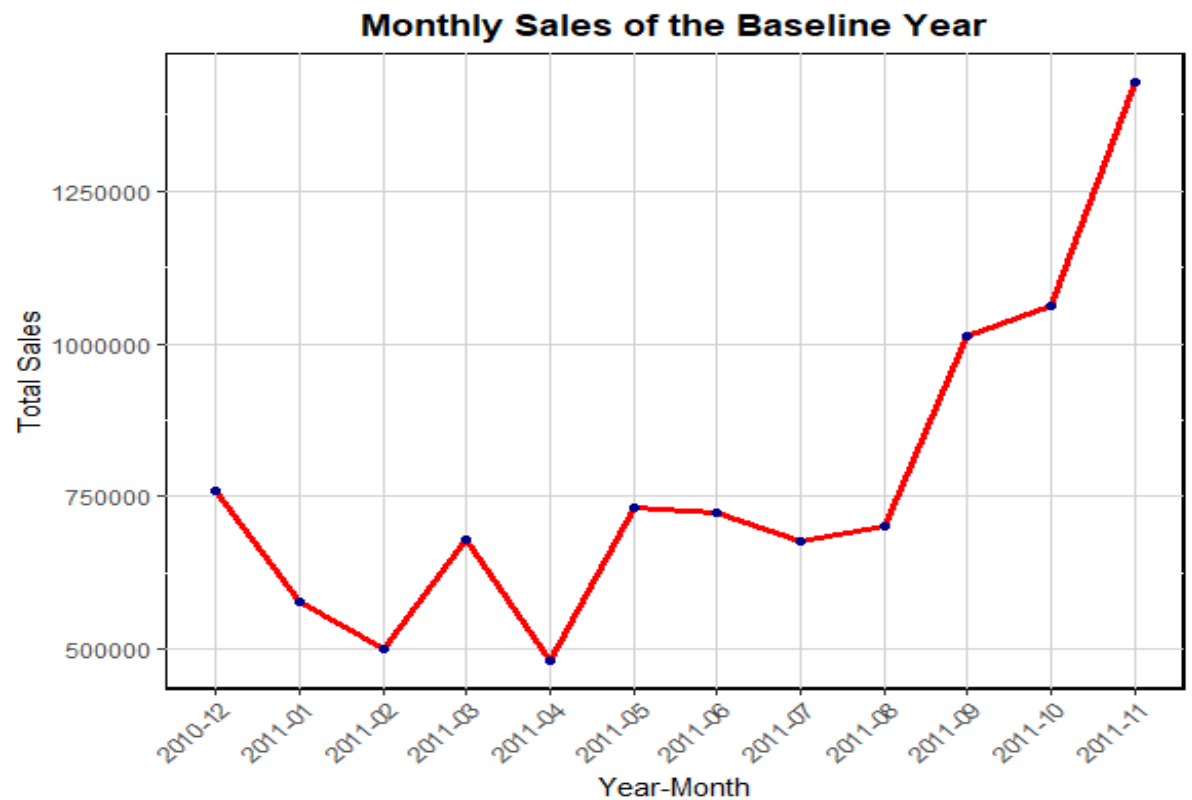
- The sales transaction data contains the transaction details of various product from '1 December 2010' till '9 December 2011'
- The transactions are identified with unique Transaction ID which has attributes that defines product ID, product description, transaction timestamp, quantities of the product purchased, unit price, customer ID of buyer and country of transaction
- The raw data contains 541909 rows and 8 columns
- A separate data containing date has been merged with sales transaction data to get attributes of timestamp

## Questions Covered in the Analysis:

- Company's sales performance over time
- Trends and insights related to sales
- Company's performance in terms of customer acquisition and customer loyalty
- Customer behaviour and related insights
- Initiatives to improve the sales

# SALES PERFORMANCE

# Monthly Sales:

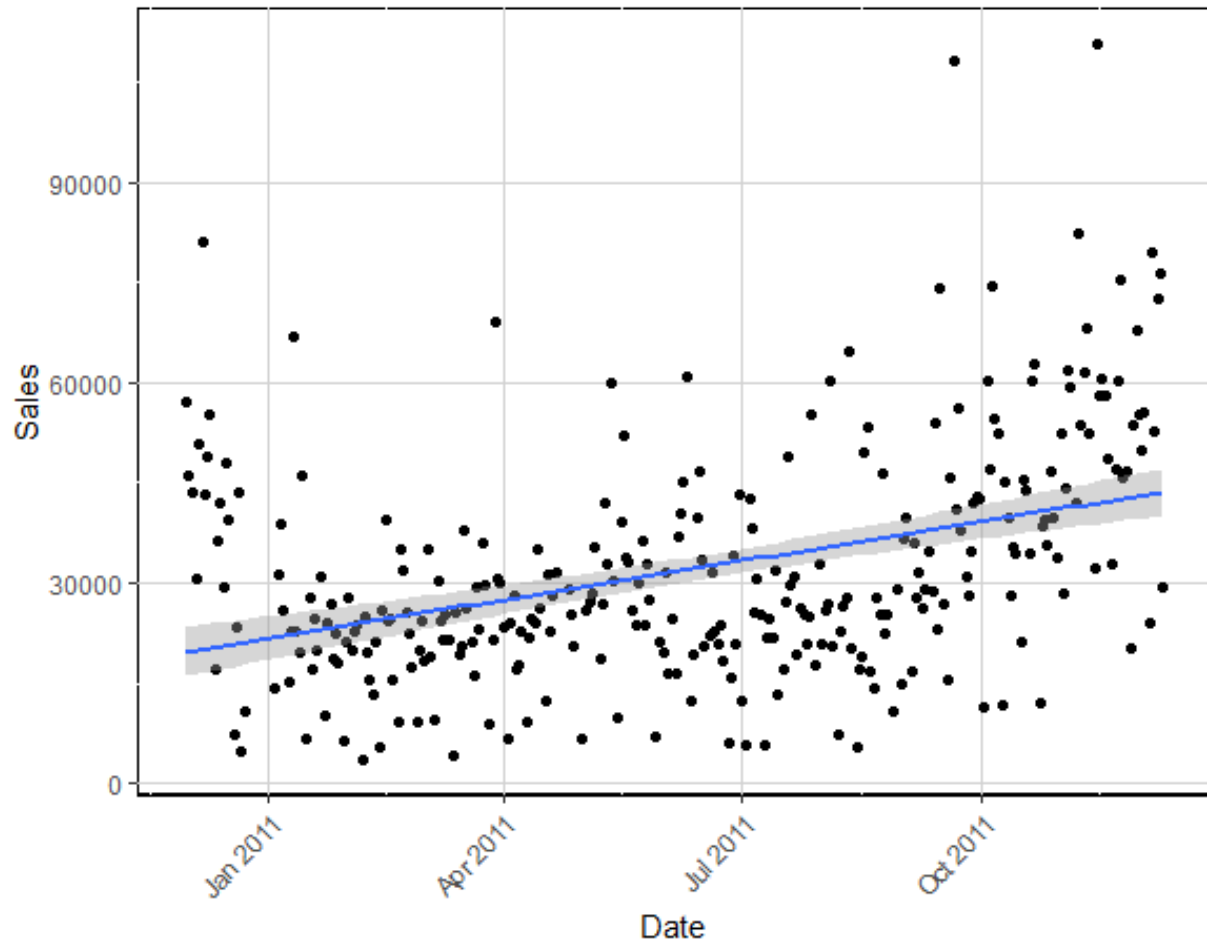


Year-Month	Sales
2010-12	758090.5
2011-01	578855.19
2011-02	499464.35
2011-03	679329.21
2011-04	482061.93
2011-05	731004.77
2011-06	723845.14
2011-07	676812.08
2011-08	701289.93
2011-09	1011339.74
2011-10	1061701.22
2011-11	1427124.54
2011-12	440399.56

- Sales improved significantly in Sep, Oct and Nov 2011.
- From Dec '2010 to Aug '2010, sales have been fluctuating without any improvement from initial sales recorded in Dec '2010
- Seasonality could be a factor in high sales in last three months
- In the above graph, sales of Dec '2011 are not recorded because sales of data upto 9 Dec '2011 were only present

# Daily Sales:

Linear Model



Residuals:

Min	1Q	Median	3Q	Max
-30574	-10092	-2444	7574	69768

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-9.362e+05	1.292e+05	-7.246	3.57e-12 ***
date	7.404e-04	9.879e-05	7.495	7.37e-13 ***

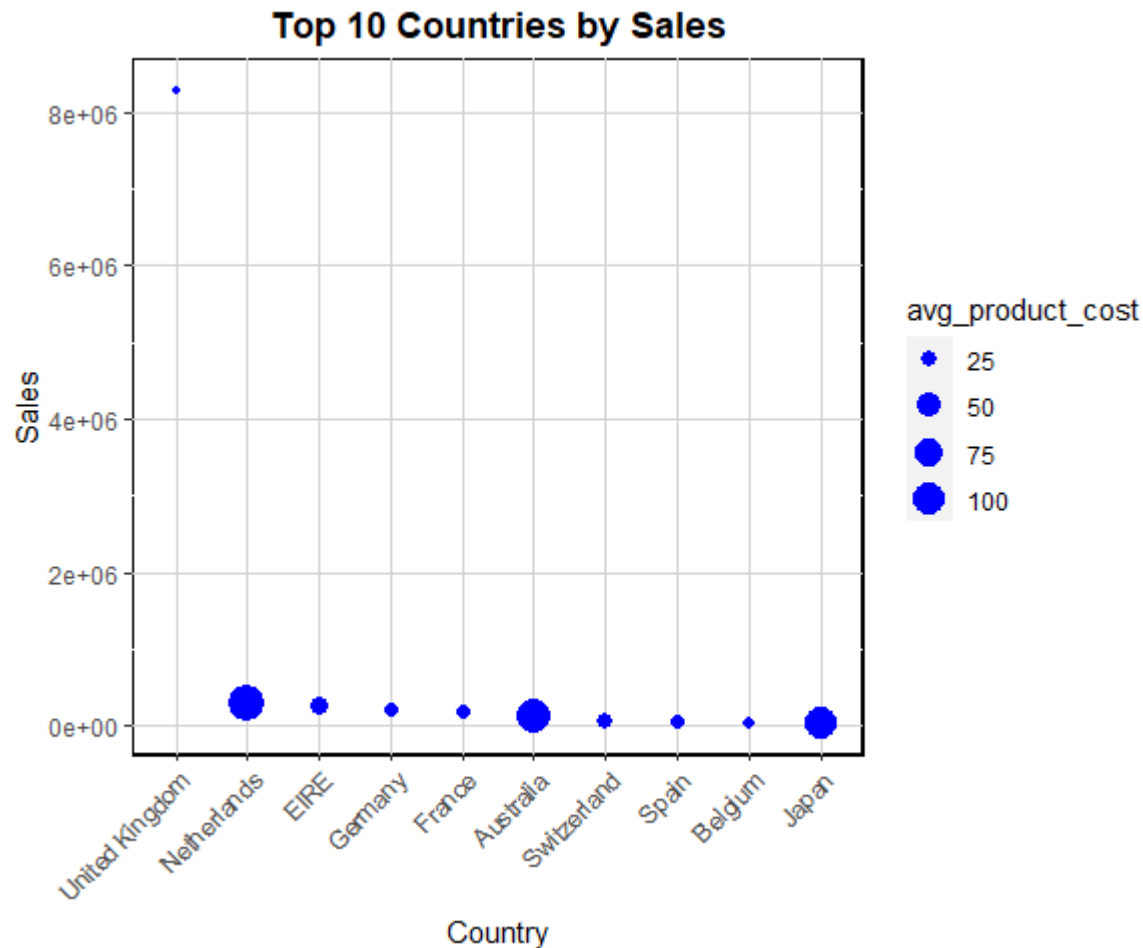
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15970 on 303 degrees of freedom  
Multiple R-squared: 0.1564, Adjusted R-squared: 0.1536  
F-statistic: 56.17 on 1 and 303 DF, p-value: 7.369e-13

- It can be concluded that the sales on daily level is increasing but that **growth is not very high**
- Hence, overall sales performance has slightly improved mainly because of sales in last three months

# SALES INSIGHTS

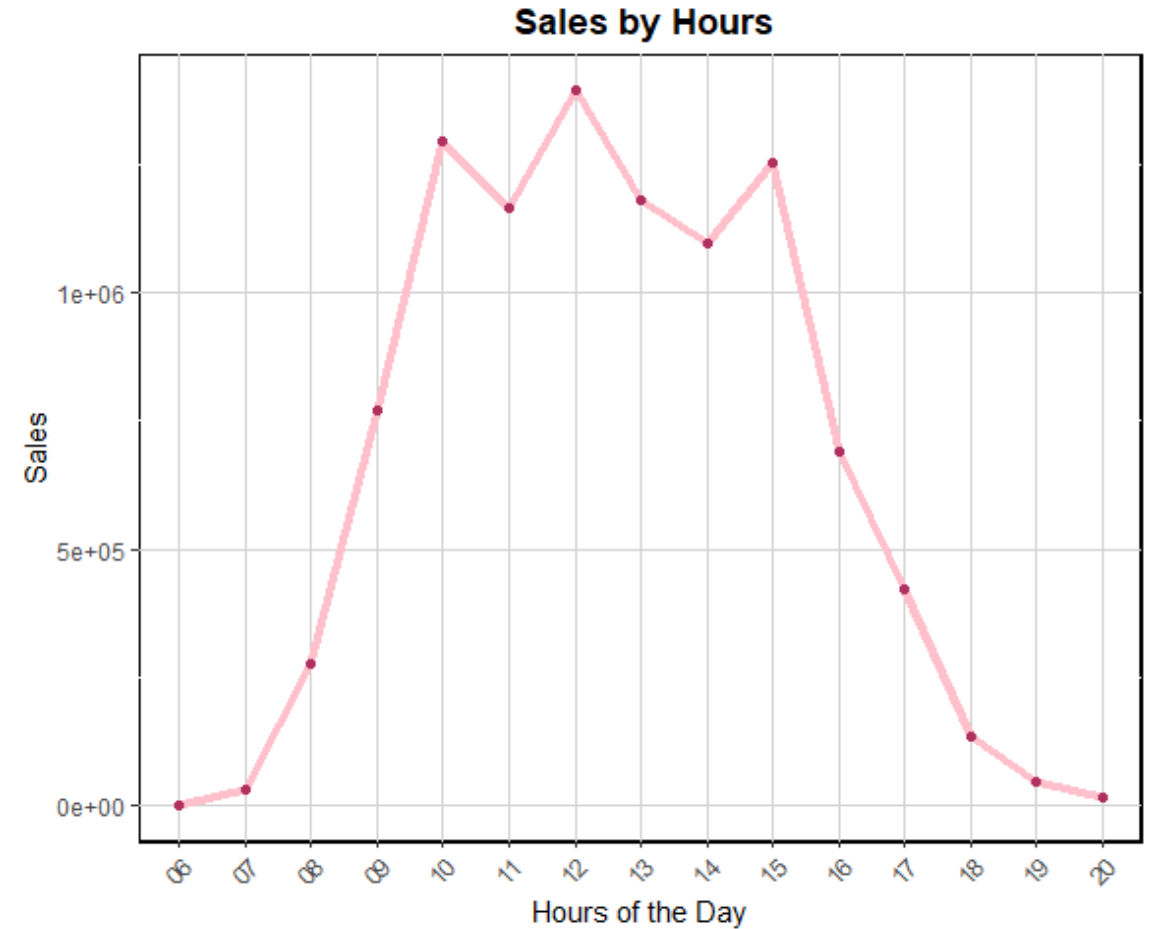
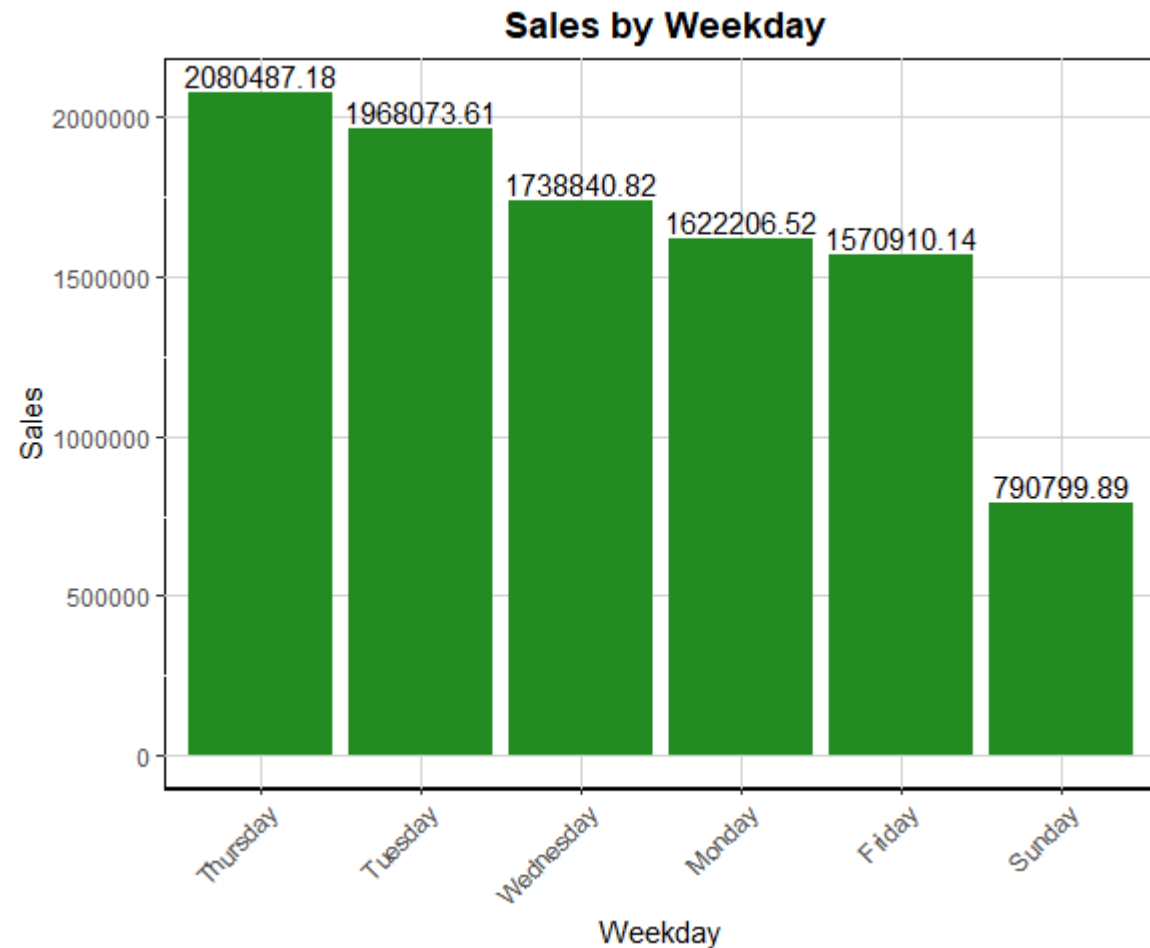
## Sales by Country:



Country	Sales	Transaction Count	Average Cost of Product
United Kingdom	8281390.36	486200	17.03
Netherlands	283479.54	2326	121.87
EIRE	259380.02	8059	32.19
Germany	200619.66	9080	22.09
France	182076.6	8218	22.16
Australia	136922.5	1253	109.28
Switzerland	52483.05	1960	26.78
Spain	51746.65	2462	21.02
Belgium	36662.96	1971	18.60
Japan	35419.79	355	99.77

- UK has the **highest sales** but the average product cost is very less
- **Netherlands, Australia and Japan** have low sales compared to UK but have **high average cost of product** bought

# Sales by Weekday and Hours:

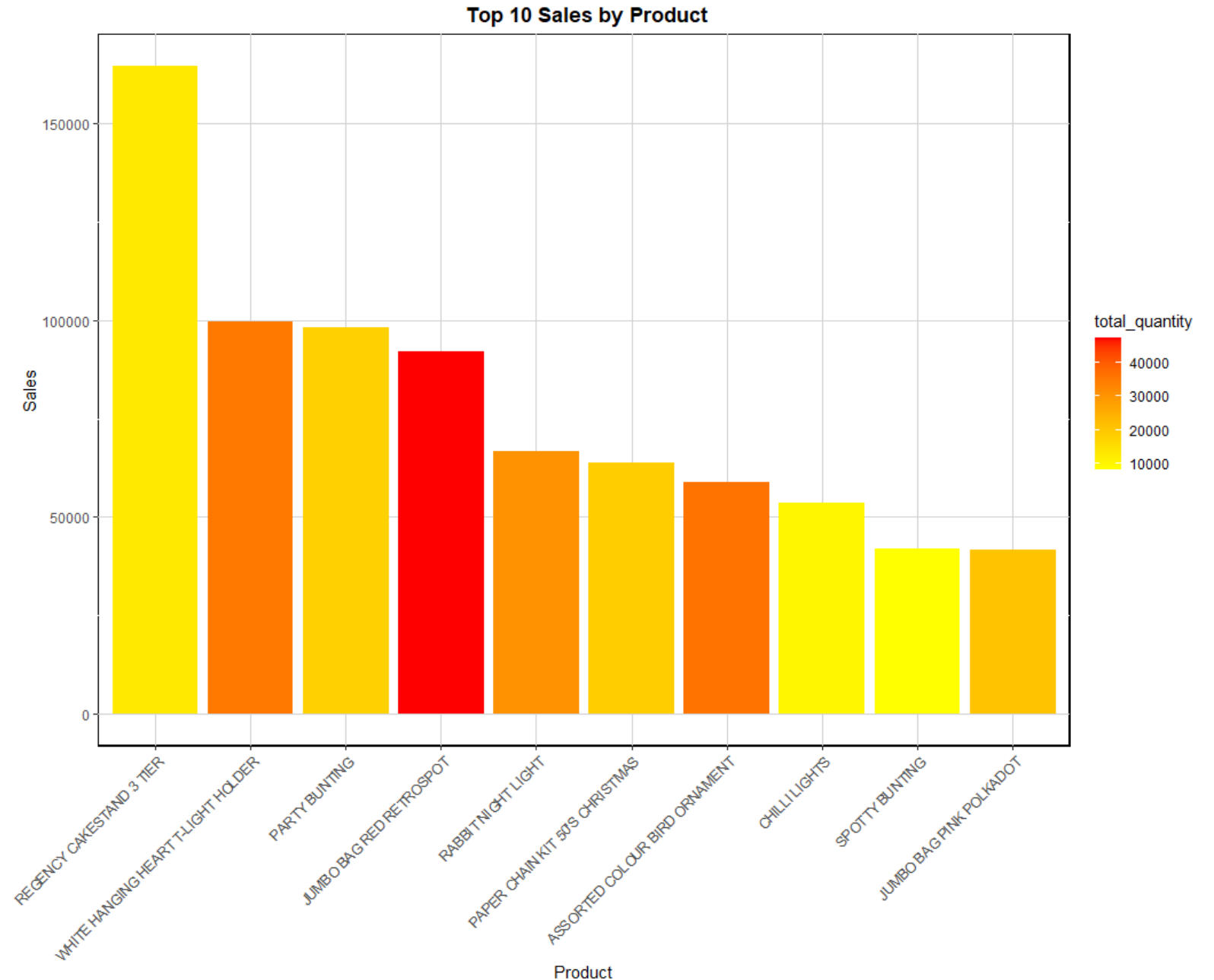


- Sales on **Sunday** are very low compared to other weekdays
- Sales are **very high** from **10:00 AM to 3:00 PM**



## Sales by Product:

- “Regency Cakestand 3 Tier” has generated highest sales
- “Jumbo Bag Red Retrospot” has very high number of quantity sold
- “Regency Cakestand 3 Tier” has less quantity sold than many others in this bar chart that means it is a product that generates sales because of being a product with good cost as well as having decent quantity number.

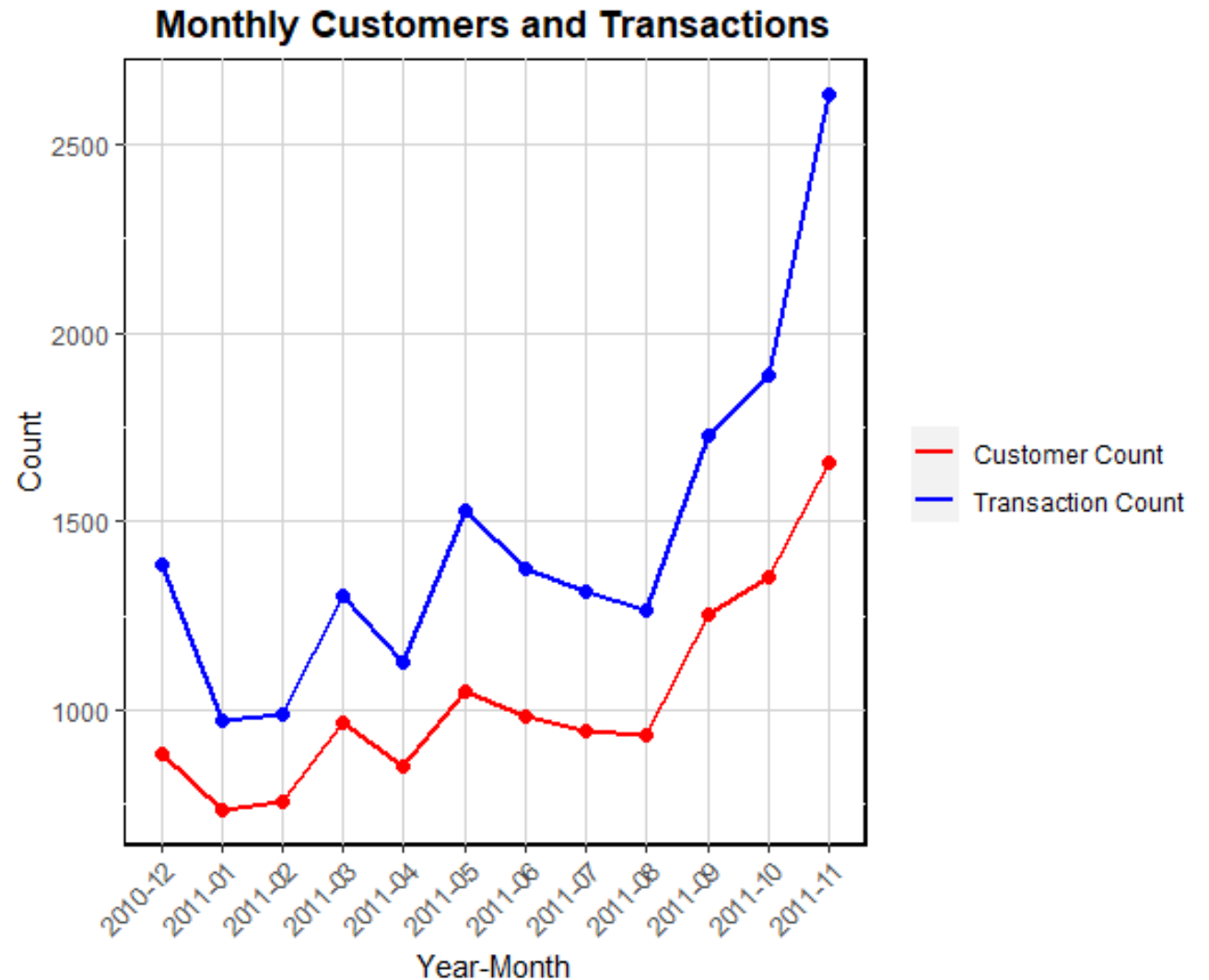


# CUSTOMER GROWTH AND RETENTION

# Monthly Customers and Transactions:

- Monthly unique customers count is fluctuating till Aug '2011 but from Sep'2011 to Nov '2011 sharp growth in customer count can be seen
- Just like the customer count, number of unique transaction has sharp growth from Sep'2011 to Nov '2011
- In each month number of transactions are almost double of number of customers

- Average Transaction Number (Yearly) = 4.22
- Median Transaction Number = 2
- Average Quantity Purchased = 269
- Average Purchase Amount) = 457



# Cohort Analysis: Customer Retention

COHORT	Month.01	Month.02	Month.03	Month.04	Month.05	Month.06	Month.07	Month.08	Month.09	Month.10	Month.11	Month.12
2010-12	883	323	286	337	317	352	318	307	312	347	330	443
2011-01	414	89	111	95	132	119	103	100	123	136	152	
2011-02	381	71	71	108	103	93	97	104	93	115		
2011-03	451	66	112	88	101	76	120	104	126			
2011-04	300	64	61	63	59	69	64	78				
2011-05	283	55	49	49	59	65	75					
2011-06	242	42	38	64	56	80						
2011-07	187	33	38	42	51							
2011-08	169	34	40	41								
2011-09	296	70	90									
2011-10	356	84										
2011-11	322											

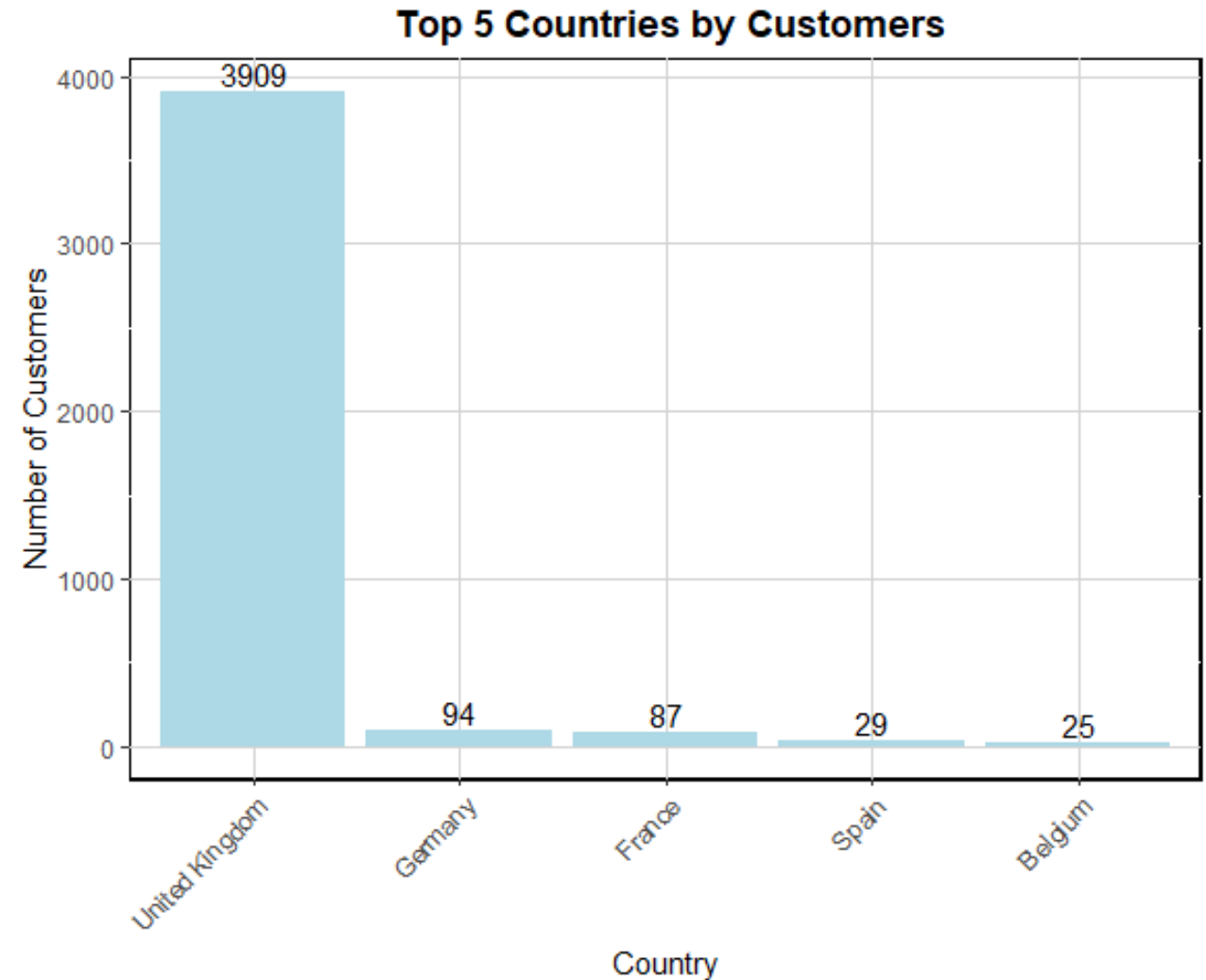
COHORT	Month.01	Month.02	Month.03	Month.04	Month.05	Month.06	Month.07	Month.08	Month.09	Month.10	Month.11	Month.12
2010-12	100.00%	36.58%	32.39%	38.17%	35.90%	39.86%	36.01%	34.77%	35.33%	39.30%	37.37%	50.17%
2011-01	100.00%	21.50%	26.81%	22.95%	31.88%	28.74%	24.88%	24.15%	29.71%	32.85%	36.71%	
2011-02	100.00%	18.64%	18.64%	28.35%	27.03%	24.41%	25.46%	27.30%	24.41%	30.18%		
2011-03	100.00%	14.63%	24.83%	19.51%	22.39%	16.85%	26.61%	23.06%	27.94%			
2011-04	100.00%	21.33%	20.33%	21.00%	19.67%	23.00%	21.33%	26.00%				
2011-05	100.00%	19.43%	17.31%	17.31%	20.85%	22.97%	26.50%					
2011-06	100.00%	17.36%	15.70%	26.45%	23.14%	33.06%						
2011-07	100.00%	17.65%	20.32%	22.46%	27.27%							
2011-08	100.00%	20.12%	23.67%	24.26%								
2011-09	100.00%	23.65%	30.41%									
2011-10	100.00%	23.60%										
2011-11	100.00%											

- Customer retention for the customers in Dec '2010 is 50% by the 12<sup>th</sup> month which is the highest ever in comparison to any other months
- However, customer retention for months other than Dec '2010 are not so high

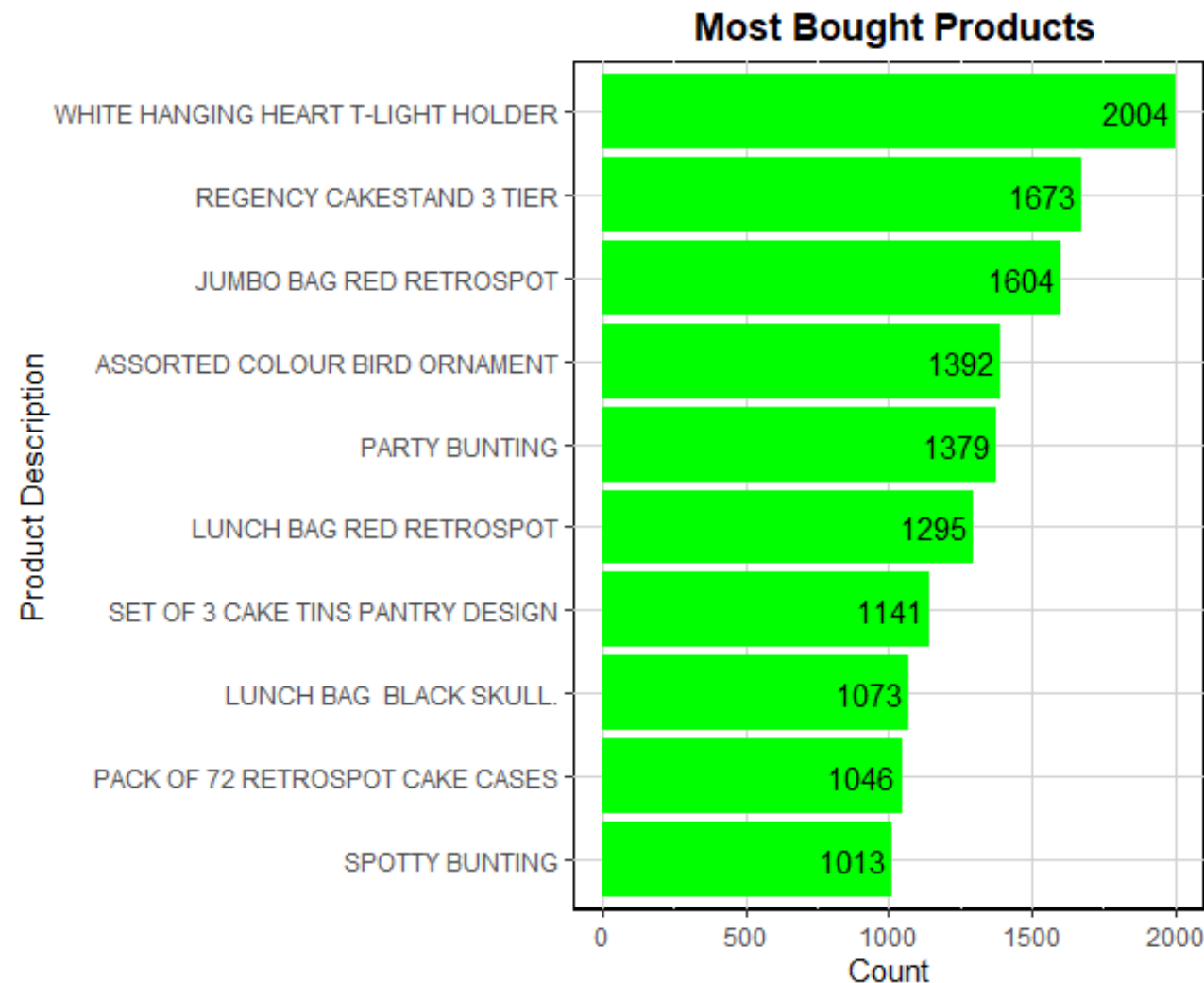
# CUSTOMER INSIGHTS

## Top Customers by Country:

- **United Kingdom** has very high number of customer base at **3909** unique customers
- **Germany** and **France** are next highest after UK with **94** and **87** customer base respectively



# Most Bought Products:



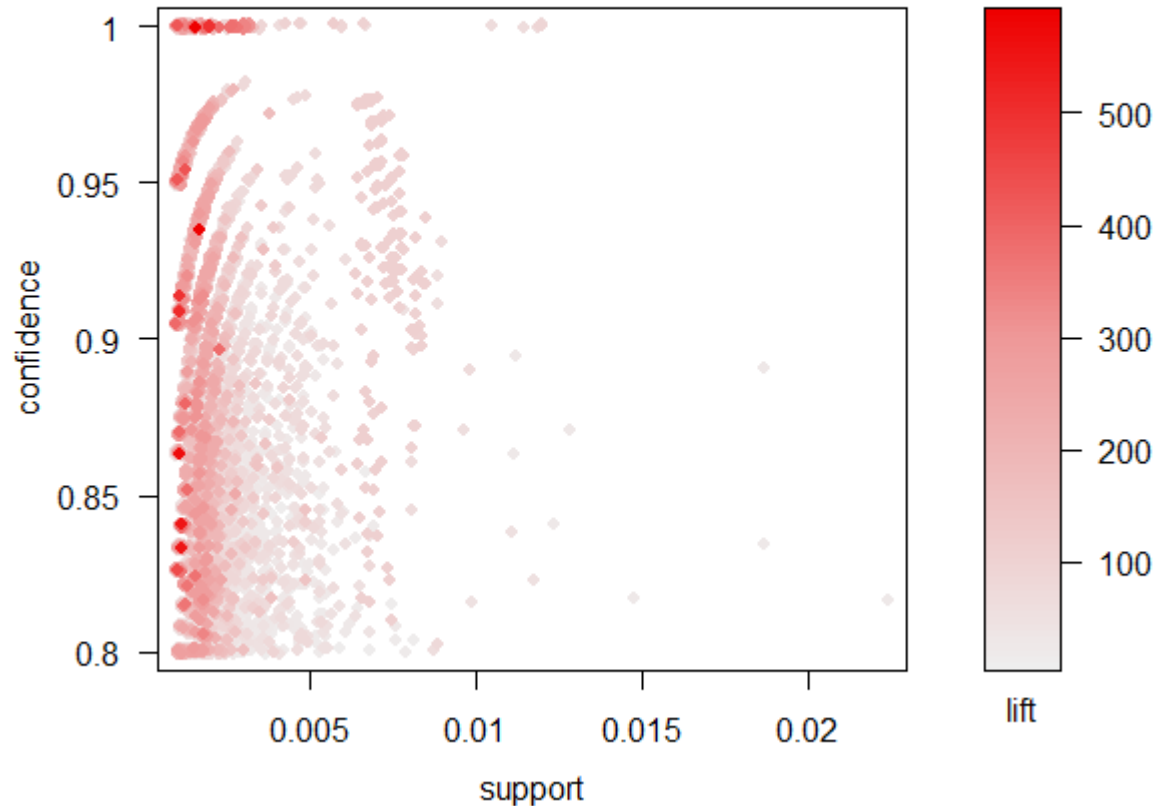
- “White Hanging Heart T-Light Holder” is the most bought product by the customers with 2004 unique transactions

# MARKET BASKET ANALYSIS

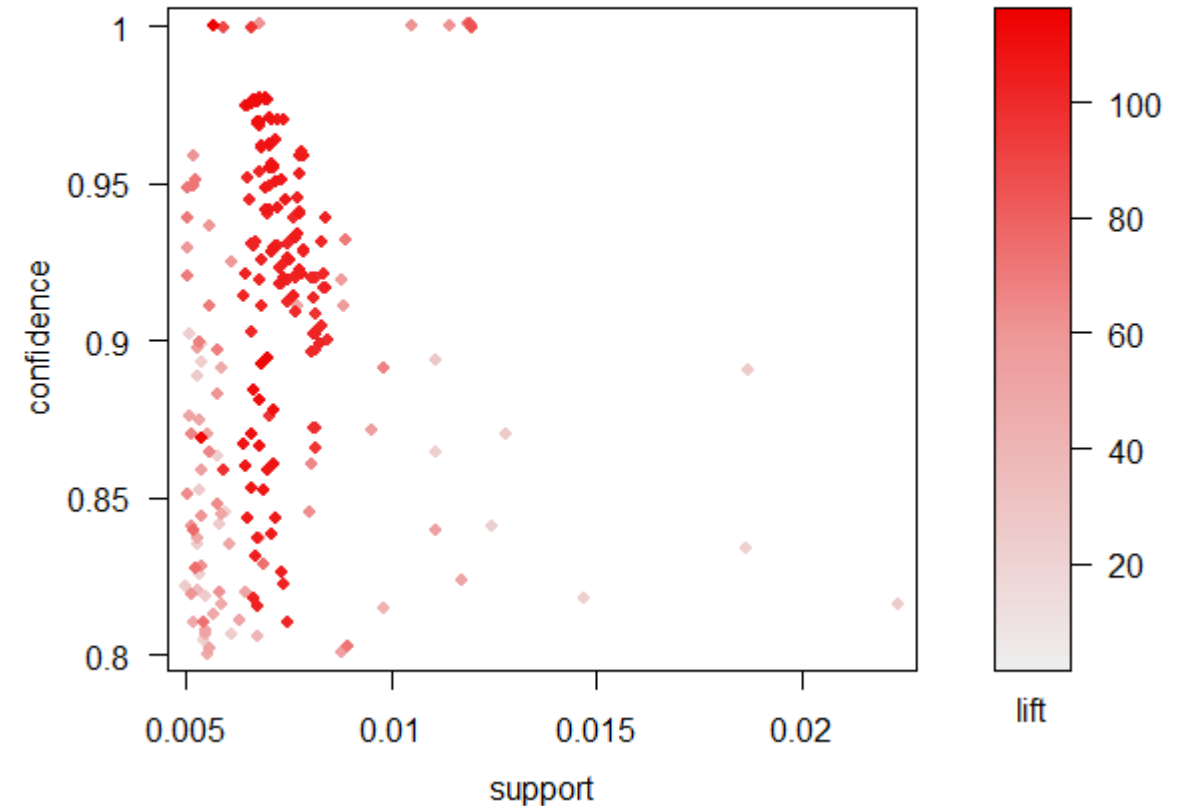


# Association Rules Scatter Plot:

Scatter plot for 12168 rules



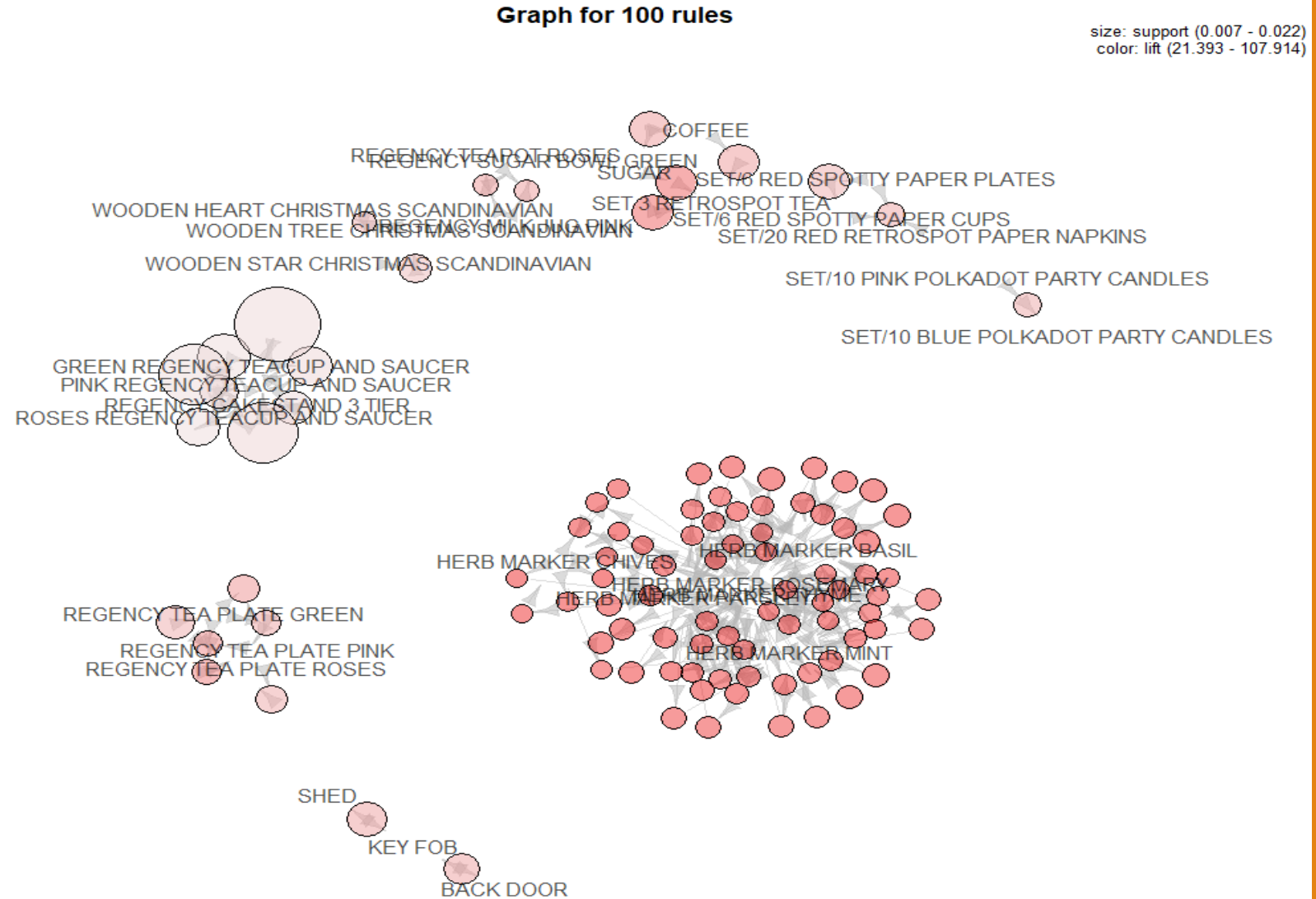
Scatter plot for 231 rules



- The scatter plot on left tells the **distribution of association rules** for products with minimum **Support=0.001** and **Confidence=0.8** and **maximum length of association to 4**
- The scatter plot on right is filtered version of first graph with **Support > 0.005** and **Lift > 20**

# Graph of top 100 rules:

- Clusters of the products seen in the graph tells us the closely associated rules for various set products
- These clusters represent that the **products in a cluster** are closely associated and hence, are **likely to be purchased together**



## Conclusion:

- The sales performance can be improved by creating a recommendation system where a customer will be recommended a product based on previous purchased products.
- The recommendation rules can be obtained by Market Basket Analysis as mentioned in previously
- Based on the insights generated, the out of stock situation can be avoided by looking at the most bought products chart
- Marketing/Promotional campaigns can be done during 10:00 AM to 3:00 PM which is peak timing
- Offers on products can be introduced and weekdays should be preferred for that as on Sunday, very less sales are generated.
- Customer base can be increased by targeting specific countries outside the UK such as Germany and France