

Actionable Insights through Analytics for a Consumer Durables Retailer

Final Report- Group 3 (August 2016 Batch, Pune)

Submitted by:

Rahul Bhardwaj

Amar Vivek

Harshvardhan Mirajkar

Badri Narayan Patnaik

Anirudh Bhatjiwale

Sandeep Pangul

Research Supervisor:

Mr. Snehamoy Mukherjee

July 9, 2017

Post Graduate Program in Business Analytics & business intelligence

Great Lakes Institute of Management



**Acknowledgment**

This project would not have been possible without the kind support and help of many individuals and organizations. We would like to extend our sincere gratitude to all of them.

We are highly indebted to Great Lakes Institute of Management for their guidance and constant supervision as well as for providing necessary information regarding the project and also for their support in completing the project.

We would like to express our gratitude towards our Mentor, Mr. Snehamoy Mukherjee and other members of GLIM for their kind co-operation and encouragement which helped us in completion of this project.

**Certificate of Originality**

This is to certify that the research work presented in this project entitled **Actionable Insights through Analytics for a Consumer Durables Retailer** for the Post Graduation Program at Great Lakes Institute of Management embodies the results of original and scholarly work carried out by the undersigned. This Capstone does not contain words or idea taken from published sources or written work that have been accepted as basis for the award of a degree from any higher education institution, except where proper referencing and acknowledgment were made.

**Names Sign**

Rahul Bhardwaj …………………………………

Amar Vivek …………………………………

Harshvardhan Mirajkar …………………………………

Badri Narayan Patnaik …………………………………

Anirudh Bhatjiwale …………………………………

Sandeep Pangul …………………………………

**Project Completion Certificate**

Date: 09/07/2017

This is to certify that the below listed students of Post Graduate Program in Business Analytics, Great Lakes Institute of Management, have undergone the Capstone Project work from December 01, 2016 to July 08, 2017 titled **Actionable Insights through Analytics for a Consumer Durables Retailer.**

**Student Names:**

1. Rahul Bhardwaj
2. Amar Vivek
3. Harshvardhan Mirajkar
4. Badri Narayan Patnaik
5. Anirudh Bhatjiwale
6. Sandeep Pangul

****

**Research Supervisor**

Mr. Snehamoy Mukherjee

**Table of Contents**

[List of tables and figures 6](#_Toc487396399)

[1. Introduction 7](#_Toc487396400)

[2. Literature Review 8](#_Toc487396401)

[3. Scope and Objectives 10](#_Toc487396402)

[4. Data Sources 10](#_Toc487396403)

[5. Tools 10](#_Toc487396404)

[6. Abbreviations 10](#_Toc487396405)

[7. Executive Summary 11](#_Toc487396406)

[8. Challenges 12](#_Toc487396407)

[9. Exploratory Analysis 12](#_Toc487396408)

[*Weekly Sales* 12](#_Toc487396409)

[*Branch Data* 13](#_Toc487396410)

[*Brand Analysis* 14](#_Toc487396411)

[*Product Analysis* 15](#_Toc487396412)

[10. Market Basket Analysis 16](#_Toc487396413)

[*Description* 16](#_Toc487396414)

[*Analysis* 16](#_Toc487396415)

[11. RFM Analysis for customer segmentation 17](#_Toc487396416)

[12. Twitter analysis for Samsung 18](#_Toc487396417)

[*Data Extraction* 18](#_Toc487396418)

[*Data Analysis* 18](#_Toc487396419)

[*WordCloud* 18](#_Toc487396420)

[*Word Correlation* 19](#_Toc487396421)

[*Sentiment Analysis* 20](#_Toc487396422)

[*Streamgraph* 21](#_Toc487396423)

[13. Dashboards 23](#_Toc487396424)

[*Dashboard 1: Brand wise sales performance* 23](#_Toc487396425)

[*Dashboard 2: Sales across store branches and brand-wise sales* 24](#_Toc487396426)

[*Dashboard 3: Sales and Footfall trends* 25](#_Toc487396427)

[*Dashboard 4: Market Basket Analysis* 26](#_Toc487396428)

[14. Recommendations and Conclusions 27](#_Toc487396429)

# List of tables and figures

Table: Consumer Durables Category – Page 7

Figure: Size of Consumer Durables Market – Page 7

Figure: Indian Retail Industry – Page 8

Table: Data Sources – Page 9

Table: Abbreviations – Page 9

Figure: Weekly Sales – Page 11

Figure: Weekly Footfall – Page 12

Figure: Top 5 and bottom 5 branches by sales – Page 12

Figure: Pareto chart: Top branches bringing 80% of sales – Page 13

Figure: Pareto chart: Top brands bringing 80% of sales – Page 13

Figure: Pareto chart: Top products bringing 80% of sales – Page 14

Figure: Pareto chart: Branches: Rs. 5 Cr plus sales – Page 14

Figures: RFM analysis – Page 17

Figures depicting Twitter analysis – Page 18 to 22

Figure: Dashboard 1: Brand wise sales performance – Page 23

Figure: Dashboard 2: Sales across store branches and brand-wise sales – Page 24

Figure: Dashboard 3: Sales and Footfall trends – Page 25

Figure: Dashboard 4: Market Basket Analysis – Page 26

## Introduction

Today’s Consumers are markedly different from those a few years ago. They carry out research on pricing, delivery time, product reviews and online alternatives and make an informed decision to buy from a retailer who provides maximum value. They are highly informed, enabled by new technologies and tools right from their mobile and influenced to a great extent by Social Media before making buying decisions. Their expectations are extremely high. It is thus challenging for Retailers to win their wallet share.

Retailers have to listen to the voice of the consumer, differentiate their offerings to the multi-generational customer base with drastically different needs, shopping habits and interactions with the brand. Developing a loyal customer base for their brand is crucial for retailers in order to differentiate their product from the competition and build an ongoing relationship with customers in an aggressive business ecosystem. Retailers want their customers purchasing their products and recommending them to friends and family. Some of the questions that the retailers want answers for are:-

What are the potential customers they thinking? What are their needs?

Our Capstone is aimed at exploration of data from a consumer durable retail outlet based in Chennai by applying Statistical Techniques, so as to recommend actionable insights for the Retailer in delivering a compelling value proposition to the customer.

## Literature Review

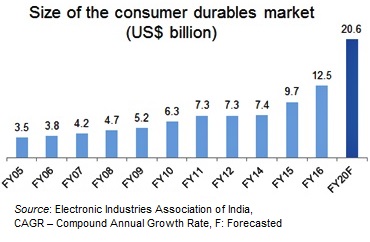
The Consumer Durables industry consists of durable goods and appliances for domestic use such as televisions, refrigerators, Air Conditioners and washing machines. Instruments such as cellphones and kitchen appliances like microwave ovens are also included in this category. The sector has been witnessing significant growth in recent years, helped by several drivers such as the emerging retail boom, real estate and housing demand, greater disposable income and an overall increase in the level of affluence of a significant section of the population.

The consumer durables industry can be broadly classified into two segments: Consumer Electronics and Consumer Appliances. Consumer Appliances can be further categorized into Brown Goods and White Goods.

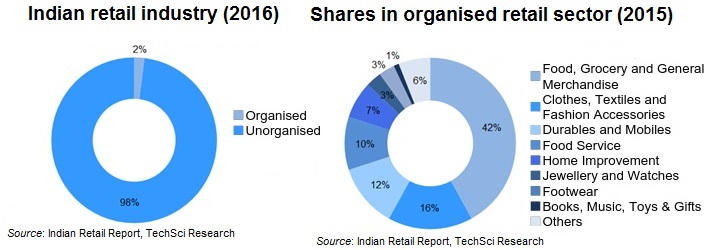
|  |  |  |
| --- | --- | --- |
| **Consumer Durables** | | |
| White Goods | Kitchen Appliances/Brown Goods | Consumer Electronics |
| * Refrigerators * Washing Machine * Air-conditioners * Speakers and Audio Equipment | * Mixers * Grinders * Microwave Owens * Iron * Electric Fans * Cooking Range * Chimneys | * Mobile Phones * Televisions * MP3 Players * DVD Players * VCD Players |

Indian consumer segment is broadly segregated into urban and rural markets, and is attracting marketers from across the world. The sector comprises of a huge middle class, relatively large affluent class and a small economically disadvantaged class, with spending anticipated to more than double by 2025.

The growing purchasing power and rising influence of the social media have enabled Indian consumers to splurge on good things. The Indian consumer sector has grown at an annual rate of 5.7 per cent between FY2005 to FY 2015. Annual growth in the Indian consumption market is estimated to be 6.7 per cent during FY2015-20 and 7.1 per cent during FY2021-25.



In 2015, revenue from consumer durables sector in India stood at US$ 9.7 billion, which further increased to US$ 12.5 billion in FY16. Consumer durable market is expected to grow at CAGR of 13 per cent from FY05 to FY20. Around two third of the total revenue is generated from urban population and rest is generated from rural population.



The consumer durables industry in India is set for sustained growth over the long term, fueled by favorable consumer demographics, overall growth in services and industrial sectors and infrastructure development in suburban and rural areas. Several Indian and MNC players are looking to strengthen their presence in India to leverage this opportunity.

Source: <https://www.ibef.org/download/Consumer_Durables_10708.pdf>

<https://www.ibef.org/industry/indian-consumer-market.aspx>

<https://www.ibef.org/download/Consumer-Durables-June-2017.pdf>

## Scope and Objectives

The scope of the project is to gain insight into the retail stores performance by studying various KPIs across different branches of the store. We have conducted exploratory analysis to investigate the data and used unsupervised learning techniques like Clustering, RFM analysis and Market-basket analysis to identify the customer segment and recommend the product basket. We have also prepared a dashboard of key performance indicators for the retailer that will help in taking informed decisions. We have also conducted sentiment analysis using Twitter to study the impact of opinions on social media, on the brand performance.

***Objectives of the Research:***

* To perform customer segmentation and product associations that can help the Retailer to increase the business at both customer and transaction level
* To build analytical models to forecast the sales
* To evaluate the brand performance and store performance across different stores
* To enable the retailer with appropriate dashboard based on KPIs to take decisions

## Data Sources

POS data from consumer durables (electronic items) from Chennai based Retailer for the year 2010-11 and 2011-12 was used for analysis, hypothesis testing and developing the predictive model. The 2012-13 data was used for validation. The dataset contains 1.2 Million records for 3 years.

The following are the variables available in the data.

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Attribute** | **Indicative Usage** |
| 1 | Bill No | Transaction Identification |
| 2 | Branch | Geographical segmentation |
| 3 | Class of Product | Product group clustering |
| 4 | Product | Exploratory analysis, market-basket |
| 5 | Brand | Exploratory analysis, wallet share, dimensions, profiling |
| 6 | QBC | Cluster Identifier |
| 7 | Item Description | Text mining for similarity, brands etc |
| 8 | Address | Customer segmentation, profiling, geospatial clustering |
| 9 | Pincode | Geospatial segmentation, Prediction |
| 10 | Qty | Detection of outliers, determine count of sales volume |
| 11 | Value | Detection of outliers, determine sales revenue |

**Table: Data Sources**

## 

## Tools

Following tools were used for the project:

* Excel
* RStudio
* Tableau
* KNIME
* Python

## Abbreviations

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Abbreviations** | **Term** |
| 1 | RFM score | Recency Frequency Monitory score |
| 2 | MBA | Market Basket Analysis |
| 3 | QBC | Quick Barcode |

## Executive Summary

The project involved analyzing three years’ (2010-2013) sales data for a chain of electronic retailer stores in Chennai. The retail data set was consolidated and imported into R and Tableau. Exploratory Data Analysis was done to understand branch, product, and category and brand wise sales. Monthly and Day-wise (throughout the week) data was also studied. We identified top branches/product/brands in terms of sales. Pareto analysis was carried out to see which product/branch is contributing to the majority of sales.

After the exploratory analysis above, we dived deeper into the data and carried out RFM analysis using KNIME tool. Twitter analysis for sentiments towards Samsung brand was carried out by writing a Python code. We also created a number of dashboards on key parameters using Tableau.

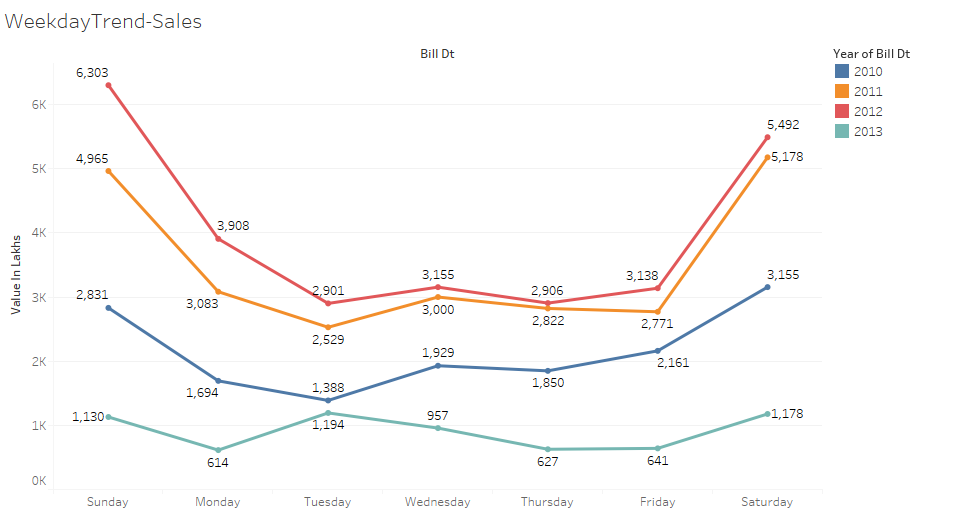
## Challenges

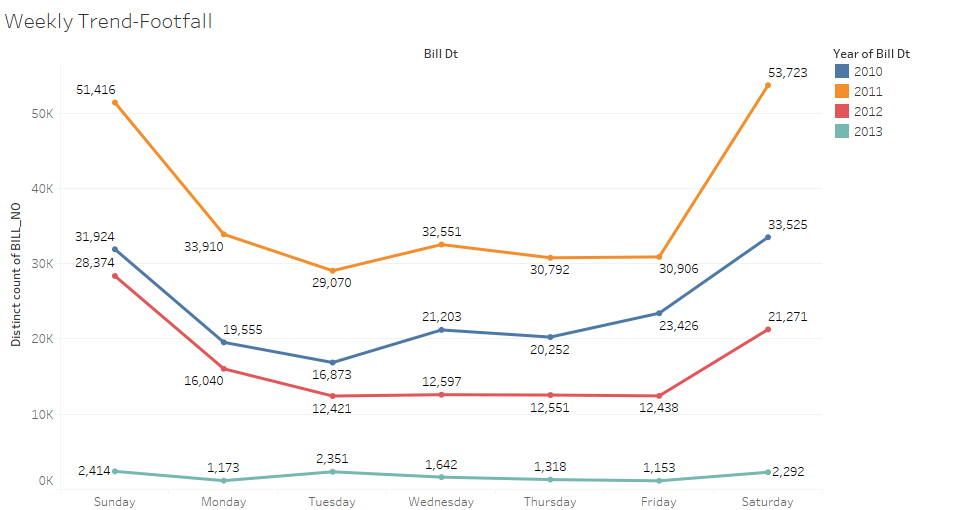
1. Unique identifier for customer is not available thus difficult to identify customers to perform RFM
2. Other supporting data like square footage of stores, neighborhood type etc. is not present. It could have helped us gain some additional insights

## Exploratory Analysis

### *Weekly Sales*

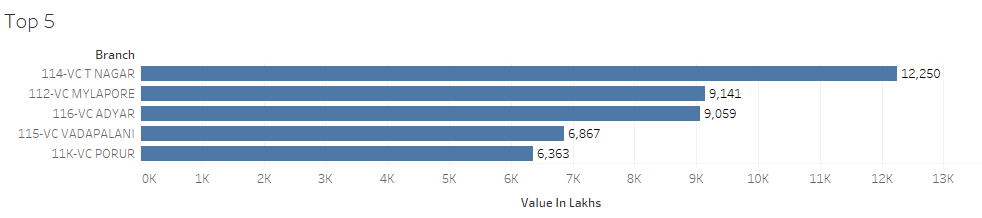
* Sum of Sales on Weekends is significantly higher than on weekdays as depicted in the graph below
* Sum of Sales has been increasing over the years. 2013 shows a dip because the data is available only for first quarter
* When comparing to the distinct count of bill generated, we see that there has been a dip from 2011 to 2012.

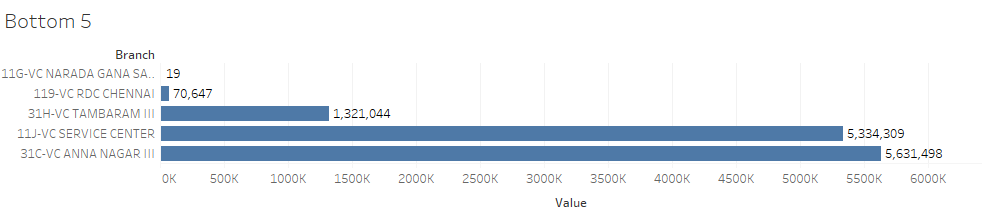


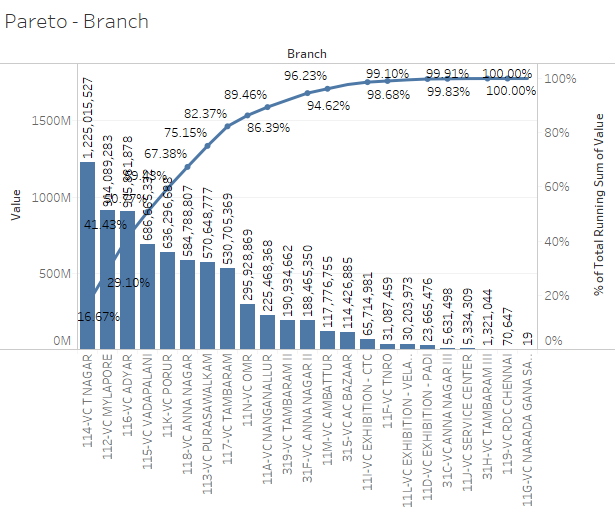


### *Branch Data*

* T Nagar is the branch that is doing the best in terms of sales followed by Mylapore
* On the other side Narada Gana Sabha is not doing much in sales followed by RDC Chennai branch
* Around 82% of sales are coming from 8 branches out of the 23 branches

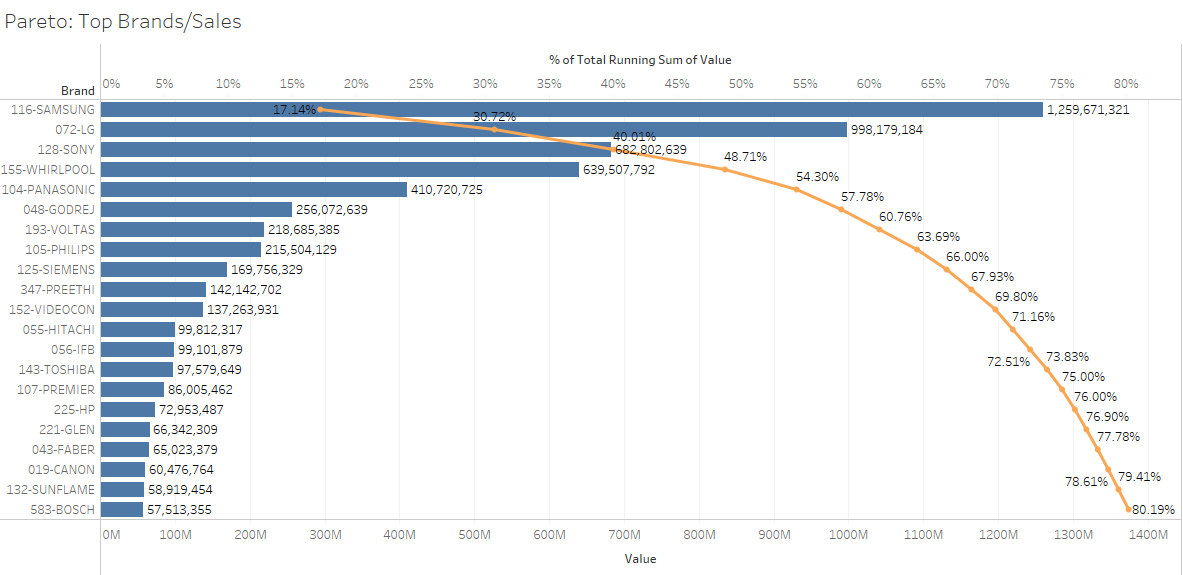






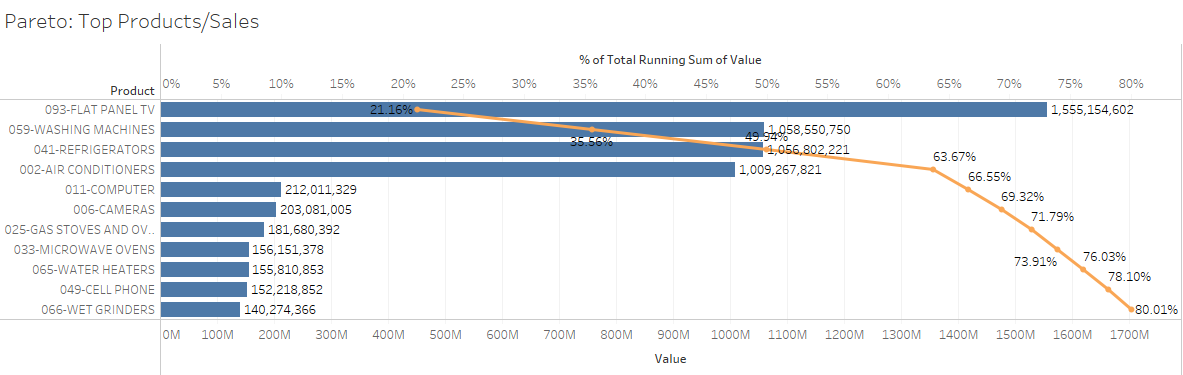
### *Brand Analysis*

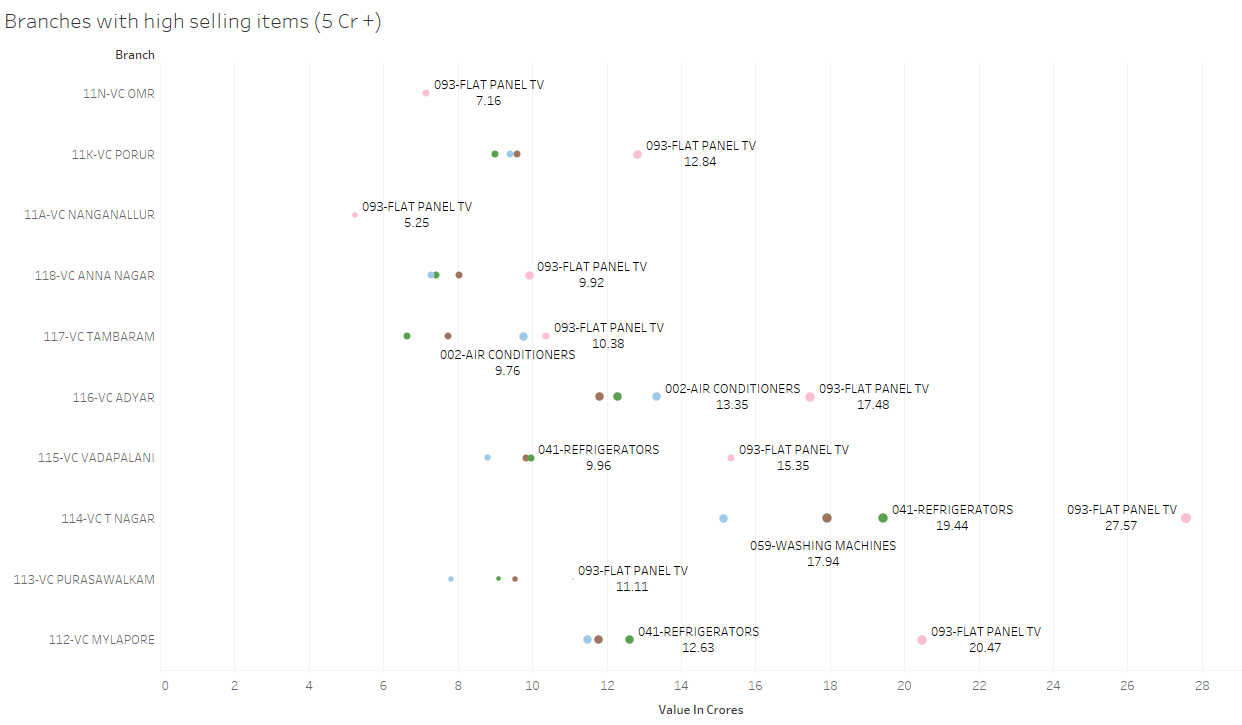
* Samsung is the brand which is selling the most
* Around 60% of sales are coming from 7 brands



### *Product Analysis*

* Flat Panel TV is the most selling product in terms of sales value followed by washing machine
* Around 80% of sales are coming from 11 products





## Market Basket Analysis

### *Description*

Market Basket Analysis is a modelling technique based upon the theory that if you buy a certain group of items, you are more (or less) likely to buy another group of items. For example, if we go out grocery shopping and buy bread then there is a very high probability that we would also buy butter or eggs. This type of analysis helps in designing shelf arrangement in stores and also helps in designing combo offers.

For our study, Market Basket Analysis was carried out using KNIME for all the 3 year data.

### *Analysis*

* After performing Market Basket Analysis we were able to find the associations between products and how they were selling together. It seems **Stabilizers** were being bought by the customers when they were buying **Air Conditioners**. In fact, Stabilizers were being bought along with **Washing Machines, Refrigerators and Flat Panel TVs** also.
* There were very close associations between **Gifts** and various other items like **Air Conditioners, Refrigerators and Washing Machine** etc. It looks like there was a lot of sale when gifts were on offer by the store.
* There was also a high association between **Refrigerators and Washing Machines** along with **Refrigerators and Air Conditioners.**
* When customers were coming to buy **Washing Machine** they were most likely to buy another product like **Irons, Gas Stoves, Water Heaters and other Kitchen Appliances.**
* There is an opportunity here to create offers for the customers so that they are tempted to buy more. The offers could be the following combination of products:

1. Stabilizer should be given as an accessory on almost all the bigger electric appliances like Refrigerator, Washing Machine, Air Conditioner, TVs.
2. Washing Machine and Refrigerator can also be turned into a lucrative offer for potential customers.
3. There is also an opportunity wherein offer could be created **for first time home buyer**. The offer could be created by combining products that a first time home buyer would be interested to furnish his house, such as Air Conditioner, TV, Washing Machine, Refrigerator and Kitchen Appliances.
4. Gifts seem to be attracting consumers, so it is a good idea to continue with the practice of giving free gifts to consumers on regularly sold products. This would give a boost to the sales of the store.

A spreadsheet showing association rules after carrying out the Market Basket Analysis is attached for reference.



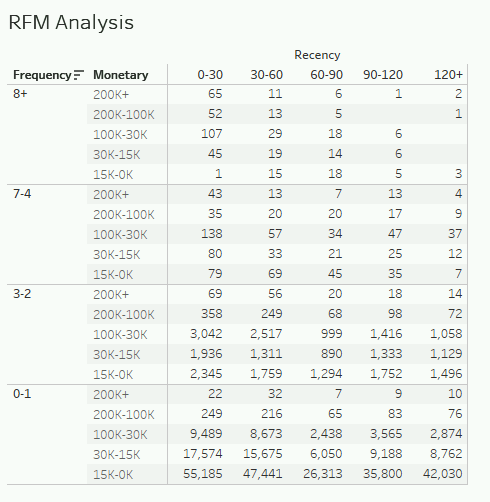
## RFM Analysis for customer segmentation

Under RFM analysis, each customer is scored based on three factors, namely Recency, Frequency, and Monetary value. RFM analysis can help companies identify customers that are most likely to respond to a new offer.

* Recency: Recency is the most important predictor of who is more likely to respond to an offer. Customers who have purchased recently from you are more likely to purchase again from you compared to those who did not purchase recently.
* Frequency: The second most important factor is how frequently these customers purchase from you. The higher the frequency, the higher is the chances of them responding to your offers.
* Monetary: The third factor is the amount of money these customers have spent on purchases. Customers who have spent higher are more likely to purchase based on the offer compared to those who have spent less.

The given datasets were not having the IDs of the customer. Thus, the unique IDs are created using customers’ unique address (with the help of string matching) and assigning to unique IDs.

318647 Unique Customers were identified.



Based on the RFM analysis, we found that customers buying frequently are either dealers or a distributors i.e. they are bulk buyers.

## Twitter analysis for Samsung

We chose Samsung as the brand for twitter analysis as it consistently came out as the best performing Brand for this store.

### *Data Extraction*

It was a challenge getting twitter data for a period of 2010-2013. Twitter inherently provides a search utility through which we can search for old tweets. This is a very manual and cumbersome process. We wrote a Python Code to create the same URL as Twitter does and then scrapped data out of it.

We modified the search URL as per our need to get to the results that we wanted and extracted the data using web scrapping. We had to iterate the URL creation for every 10 days for the period from 1st April 2010 to 31st March 2013 so that we get data spanning across the needed time period. 10 twitter feeds were collected for every 10 day period.

The Python code that we used, takes the brand names and the date range as input. Then it starts creating the web URL for the given search criteria like below:

<https://twitter.com/search?l=en&q=%23#Samsung%20since%3A2010-04-01%20until%3A2010-04-11&src=typd>

Here, we can see that the date range given is from 1st April 2010 to 11th April 2010 and the searched hashtag is of Samsung. Along with that, we have also given the language to be English. Once this URL is hit, the code picks up the top 10 tweets and writes them to a text file.

This loop runs until the end date is reached which in this case is 31st March 2013. Once the date range is exhausted, it moves onto the next brand name given. This way we were able to get Twitter data for the desired date range and required hashtag.

### *Data Analysis*

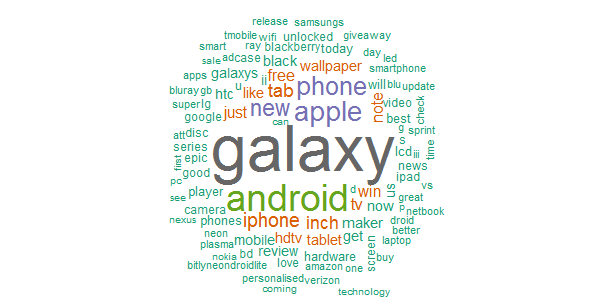
After extraction, the twitter data was cleaned for following elements:

1. Remove web links
2. Convert everything to lowercase
3. Remove everything except English like numbers and punctuation marks
4. Removed stop words and Samsung since Samsung is what we had searched for
5. Removed single letter words
6. Removed extra white spaces

Once the cleaning was done, we started with the analysis:

### *WordCloud*

The word cloud shows that ‘Galaxy’ was the most talked about thing about Samsung in that period. Other features like ‘Android’, ‘note’, ‘tab’ were also talked about. It is also apparent from the wordcloud that people were comparing Samsung phones with its main competitor Apple and its product iPhone as we see a lot of reference of it in the wordcloud.

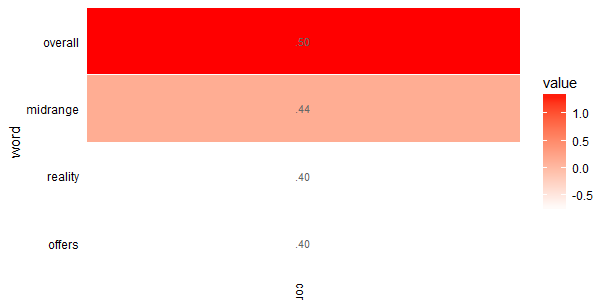


### *Word Correlation*

We ran a few word correlations just to see what two words were being used most commonly. We ran a word correlation on “good”, “great” and “iPhone”. The word correlation diagrams are given below.

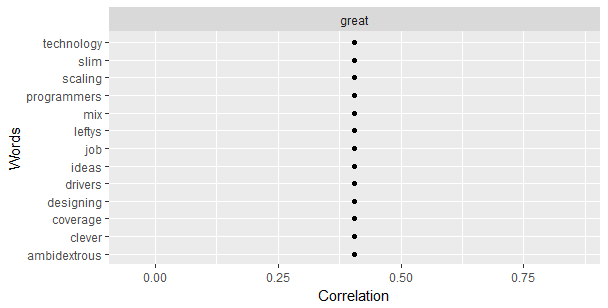
***“Good”***

From the correlation diagram below, it was evident that people considered Samsung products ‘good’, overall in the midrange products category. It also seems that they were talking about good offers when talking about Samsung products.



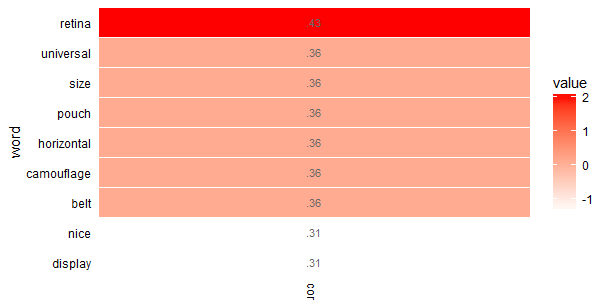
***“Great”***

When it comes to thigs referred to as ‘great’, people were saying that Samsung has great technology, Samsung team has done great job, great designing etc.



***“iPhone”***

For iPhone people were comparing their displays (talking about the retina display in particular, which was probably missing in Samsung).



### *Sentiment Analysis*

Out of the 1000+ tweets that were analyzed, the connotation used on them was mostly negative as depicted below.

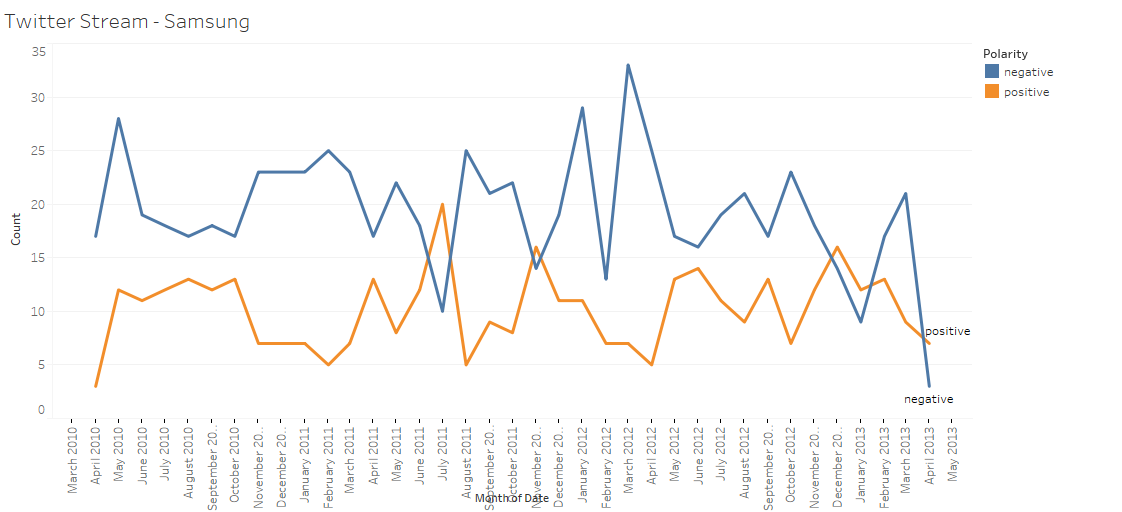
|  |  |
| --- | --- |
| **Negative** | **Positive** |
| 714 | 377 |

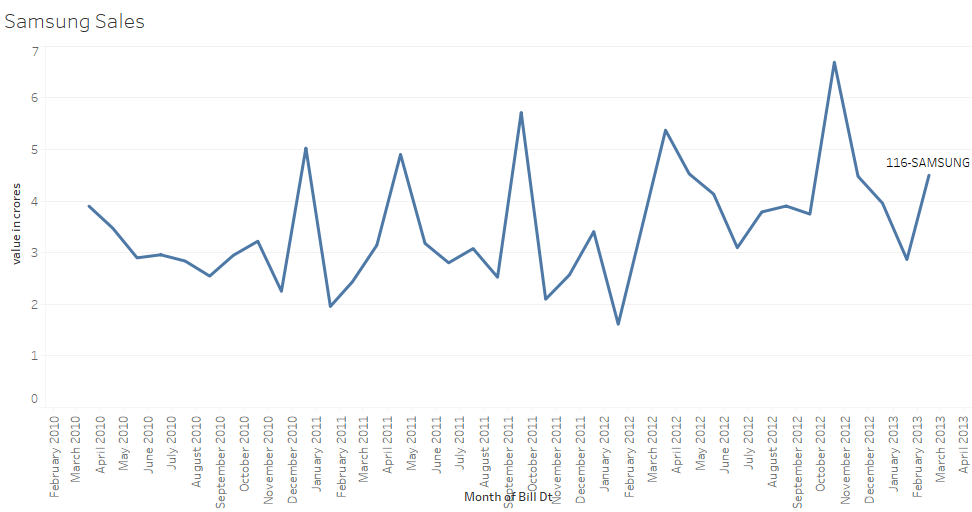
### *Streamgraph*

Below streamgraph shows the polarity of the tweets over the entire time period. We also compared it with the sales seen in the store in the graph below that. What we can see here is that towards the end of the time period, the positive sentiments were more than the negative sentiments and this is reflected in the sales as well. The sales were growing towards the end of the time period.

Around October 2010, when the sentiments were about neutral, the sales increased for December. Same can be seen when in July 2011, the positive sentiments increased, and then sales also increased in the month of September 2011.

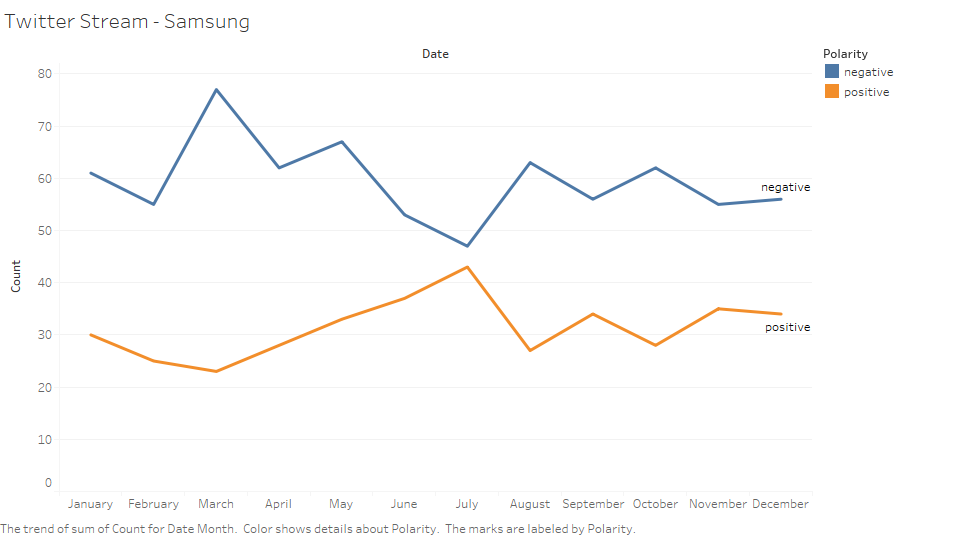
On the contrary, in the month of April 2012, when negative sentiments were the highest, the sales were also the third highest for this time period. This may also be because of low negative sentiments back in February 2012.

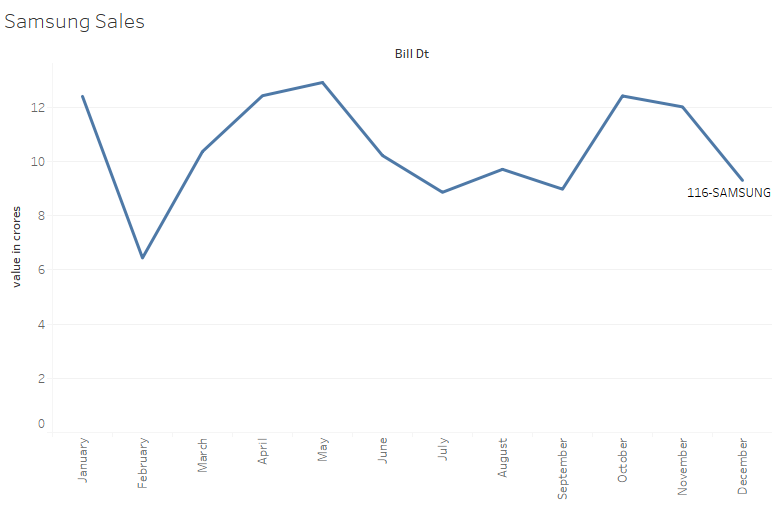




Next are streamgraphs consolidated over the months to see if there was an effect on the month of the year on what people were talking and its effects on the sales.

There does not seem to be any correlation for the month of the year on how twitter is reacting and the sales in the store.





## Dashboards

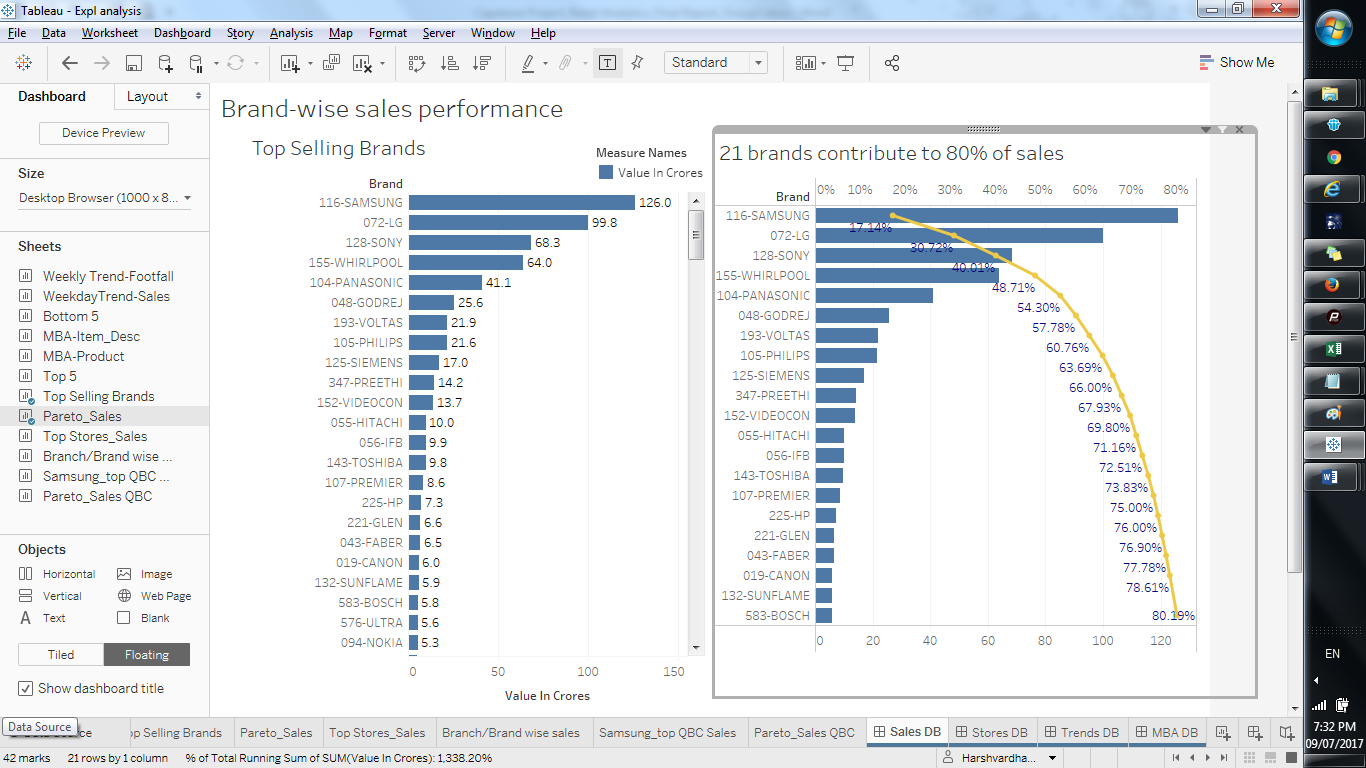
We used Tableau in creating various dashboards that could help the store management by providing visual insights into the store operations for the period of 2010-2013.

Due to the large data size, it was difficult to publish the dashboards using tableau servers; hence the dashboards are given as images below:

### *Dashboard 1: Brand wise sales performance*

This first view shows the sales in Rs Cr, showing the highest selling brands.

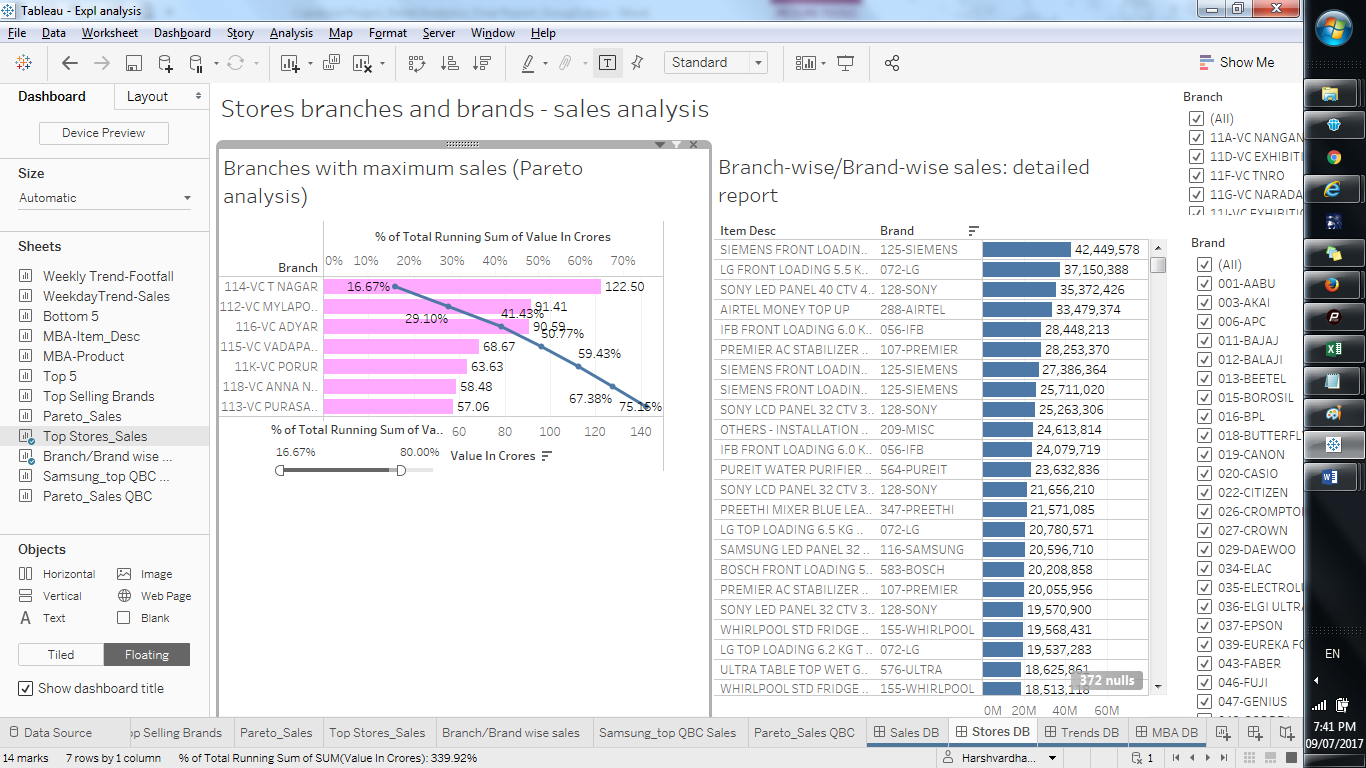
The second view (Pareto chart) shows that 21brands bring in about 80% of the revenue. Top 5 brands are Samsung, LG, Sony, Whirlpool and Panasonic



### *Dashboard 2: Sales across store branches and brand-wise sales*

The first view (Pareto chart) shows the top7 store branches which bring in about 80% of revenue

The second view shows which products bring in the most revenue; this view can be filtered for brands and store branches in Tableau

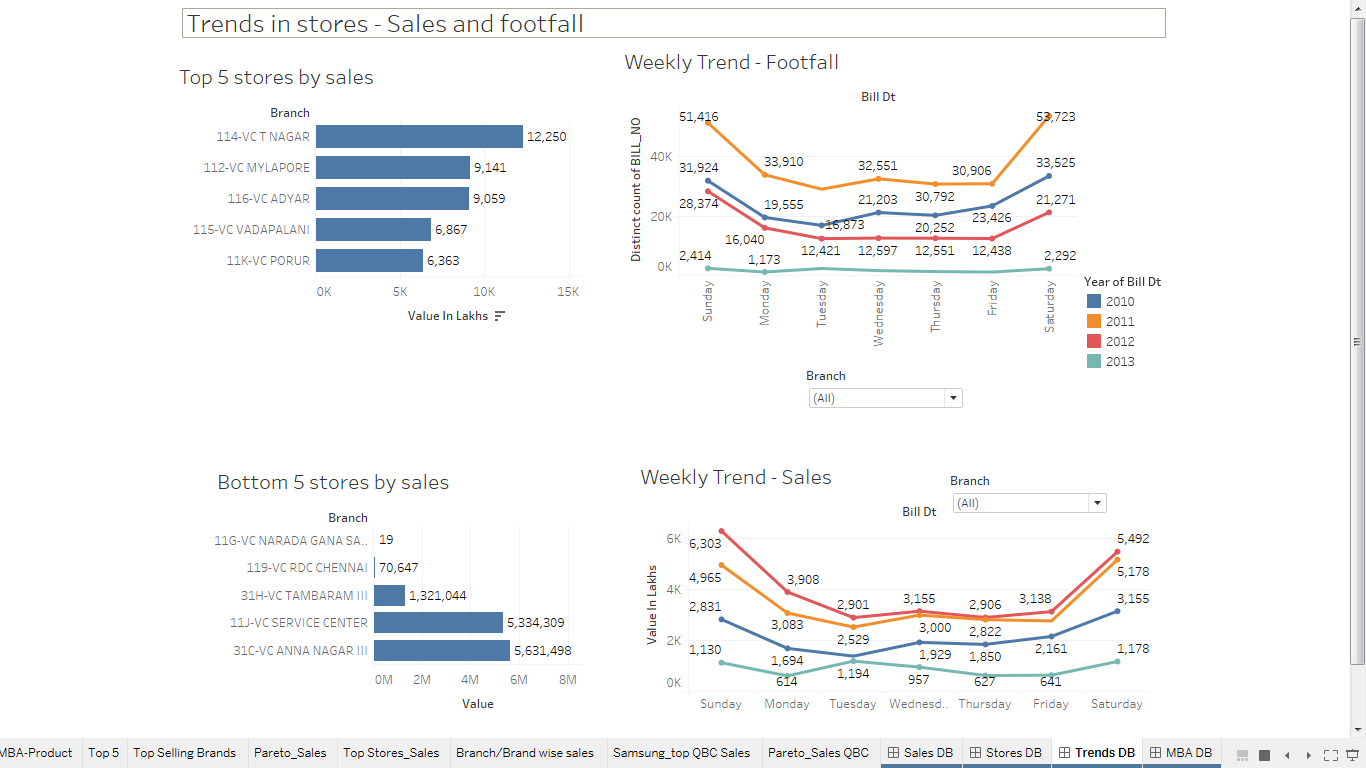


### *Dashboard 3: Sales and Footfall trends*

This dashboard gives visual analysis across 4 dimensions

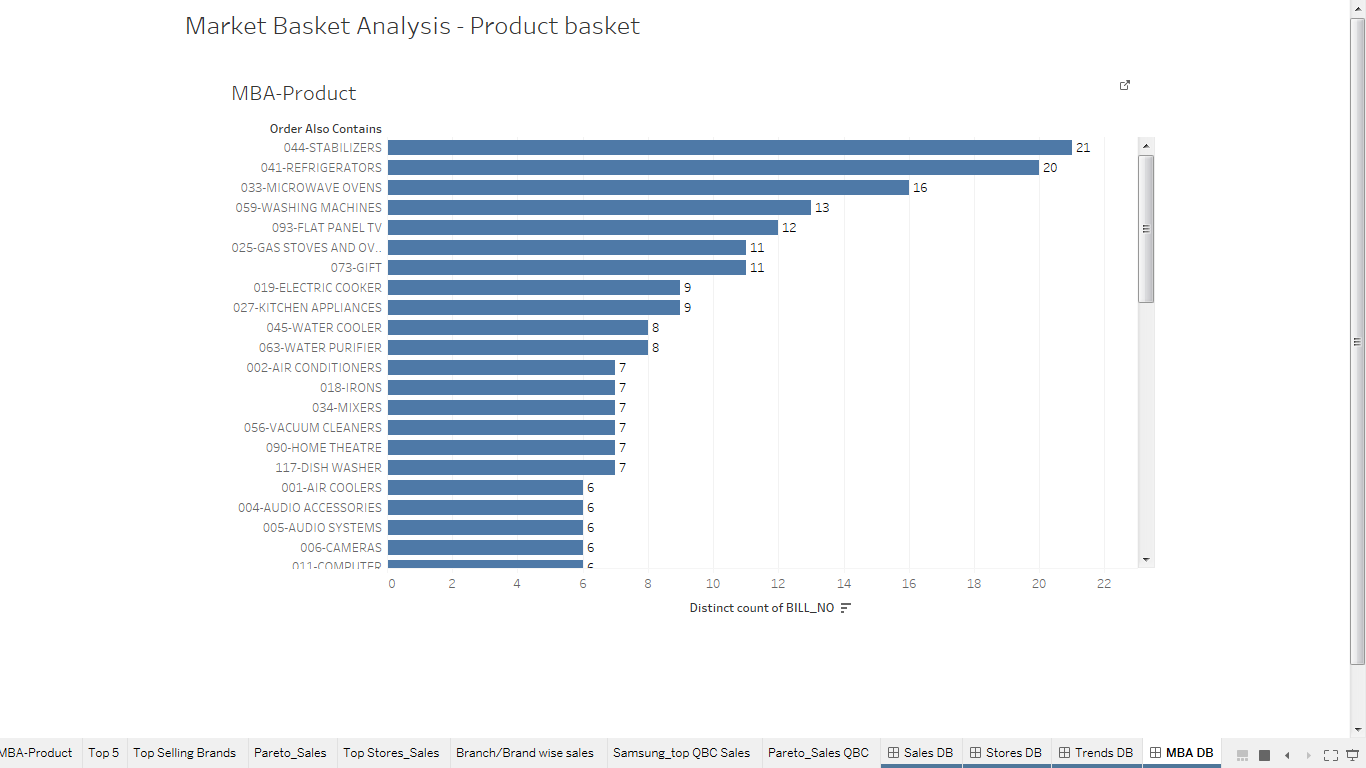
The first view simply shows the top 5 stores by sales; the view below it shows the bottom 5

The next two views show the weekly footfall and associated sales performance



### *Dashboard 4: Market Basket Analysis*

This dashboard shows how many distinct orders contain a particular product; Stabilizers, refrigerators, microwave ovens, washing machines and flat panel TVs are the top 5 products most frequently bought along with other products.



## Recommendations and Conclusions

* The association rules from MBA show that consumers are drawn towards offers which include free gifts. Thus, the stores can increase footfall if they bundle the free gifts with high value products
* Stores can also create a bundled offering for people setting up their homes for the first time or updating their appliances. This is evident from the fact that the top 5 items bought in the same order include Stabilizers, refrigerators, microwave ovens, washing machines and flat panel TVs
* RFM analysis shows that the most frequent buyers are bulk buyers – may be smaller retailers or industrial buyers Thus, the stores can target such buyers by offering customized buying options and discounts