**IFSC 5345: Information Visualization**

**Final Project Report**

Visualizing Rossmann Store Sales

Mahesh Babu Neelam

T00604964

**Objective**

To visualize the data that helps compare the stores’ sales throughout a period and impact of promotions and holidays on the sales of Rossmann Stores a drug store chain based out of Germany .

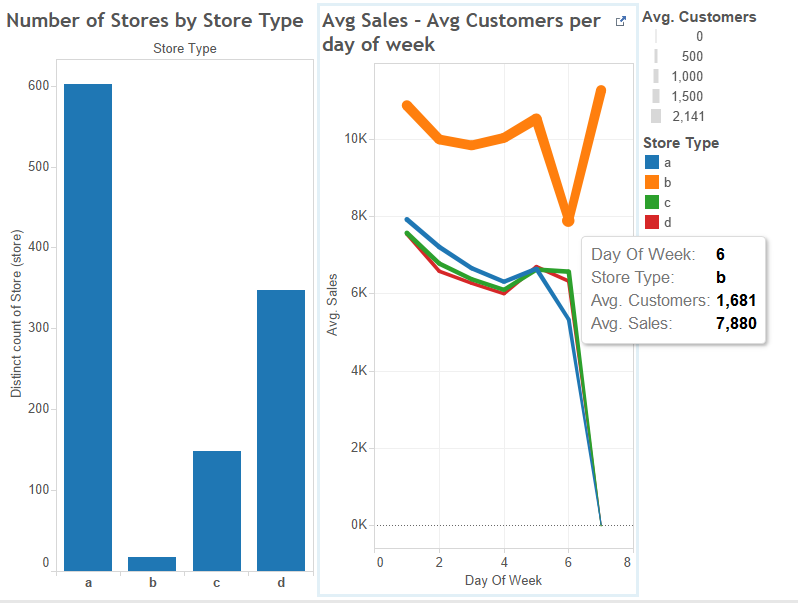
**Data**

The data has been collected from the website www.kaggle.com. There are three datasets available for the Rossmann stores which contains sales information in one dataset , stores information and the details about their competitors. The dataset collected is in the format of .csv which made easier to use. The attributes that are used for the creating visualizations are Store\_ID, promotion, sales, store type, state,country,competitor distance.,etc. The data available is for three years 2013-2015.

**Visualizations**

The visualizations which are created using the data would answer many questions and can be used by higher management level team to take appropriate actions to increase the sales. The main motive of creating the visualizations is to analyze the sales taking many factors into consideration. The visualizations include

1. **Average Sales by Store type**

****

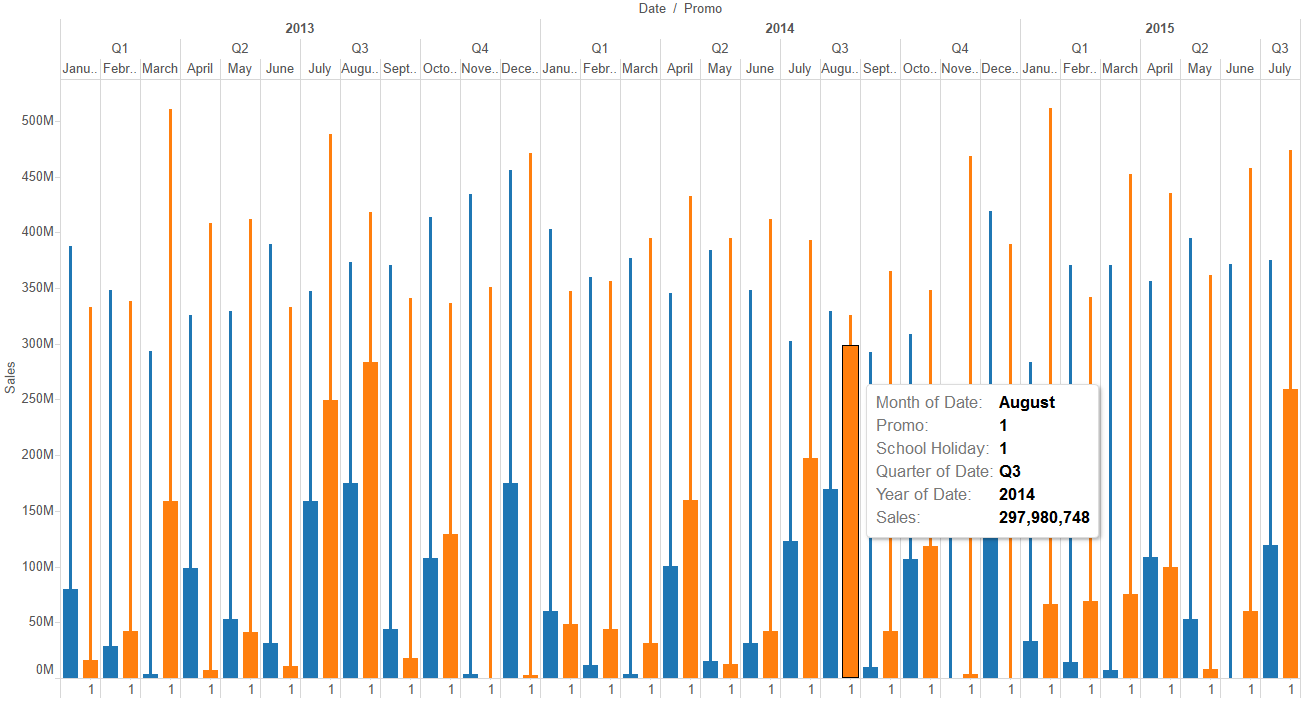
In the above cited visualization, the bars represent the number of stores with respect to store type. The line graph represent the average sales done by all types of stores ( here considered Store type a,b,c,d).

*Questions Answered using the visualization 1*

From this visualization we can infer that the sales of store type ‘b’ have the highest amount of sales and also we can observe that there are less number of stores of type ‘b’. So, we can increase the stores of type ‘b’ so that it increases the amount of sales overall.

And we can also see that on Sunday except the store ‘b’ all other stores are closed. If we can suggest that if they open some of the stores from other type there will be increase in sales.

1. **Impact of Holidays and Promotions on Sales**

****

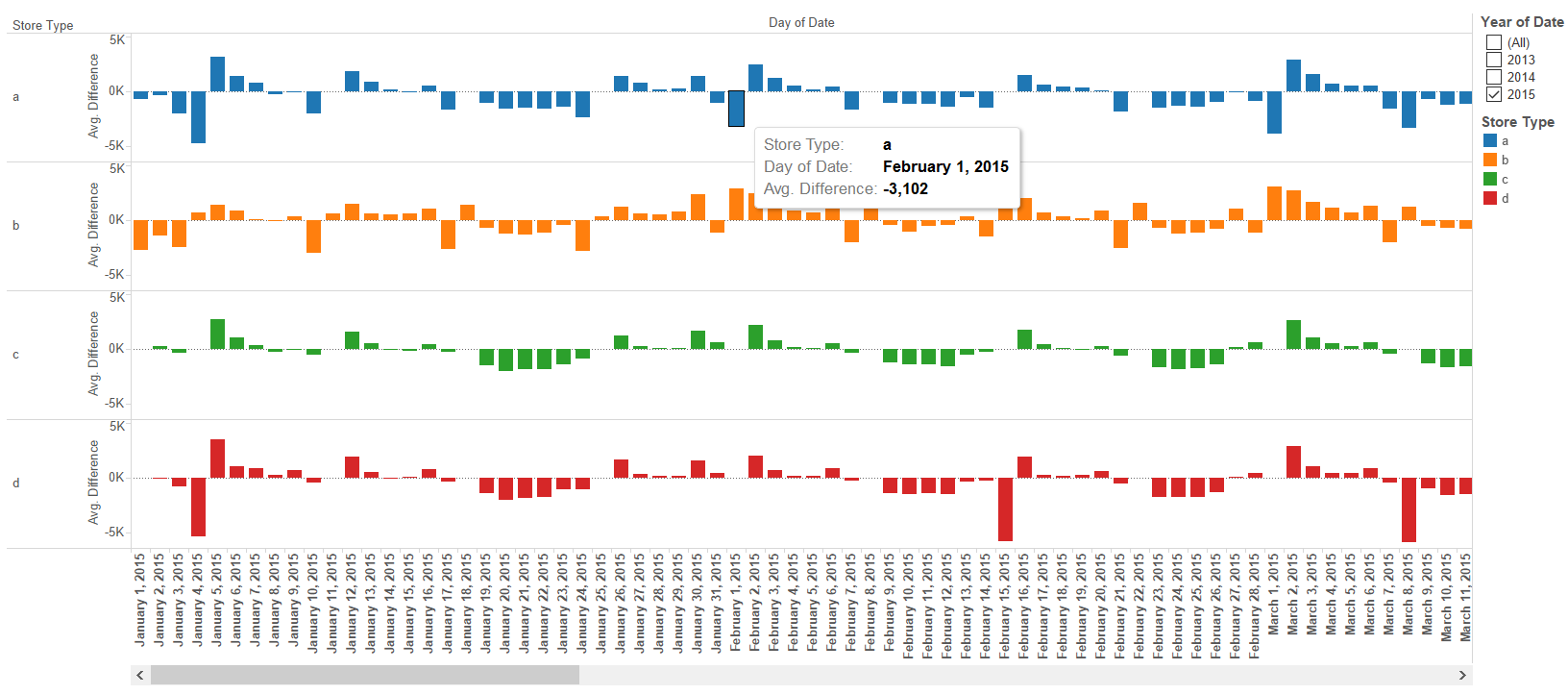
In the above visualization the impact of school holidays and the promotions on sales have been depicted . The orange bars represent the sales with both promo and holiday on the same day where as the orange lines represent the sales with only promotions in place. The blue bars represent the sales with sonly school holiday and the blue lines represent sales with neither holiday nor promotion is in place. This data is represented for every month through out the years 2013-2015.

*Questions Answered using the visualization 2*

From the above visualization we can infer that on an average the sales are high when there is a holiday and promotion both are in place. And we can also observe that for the three years time from 2013-2015 in the month of July and August we have more sales when both are in place.

From this we can suggest that if we have more promotions in the months of july and august there is a possibility of increase in sales to a considerable amount.

1. **Comparison of Daily Sales with Average Sales**

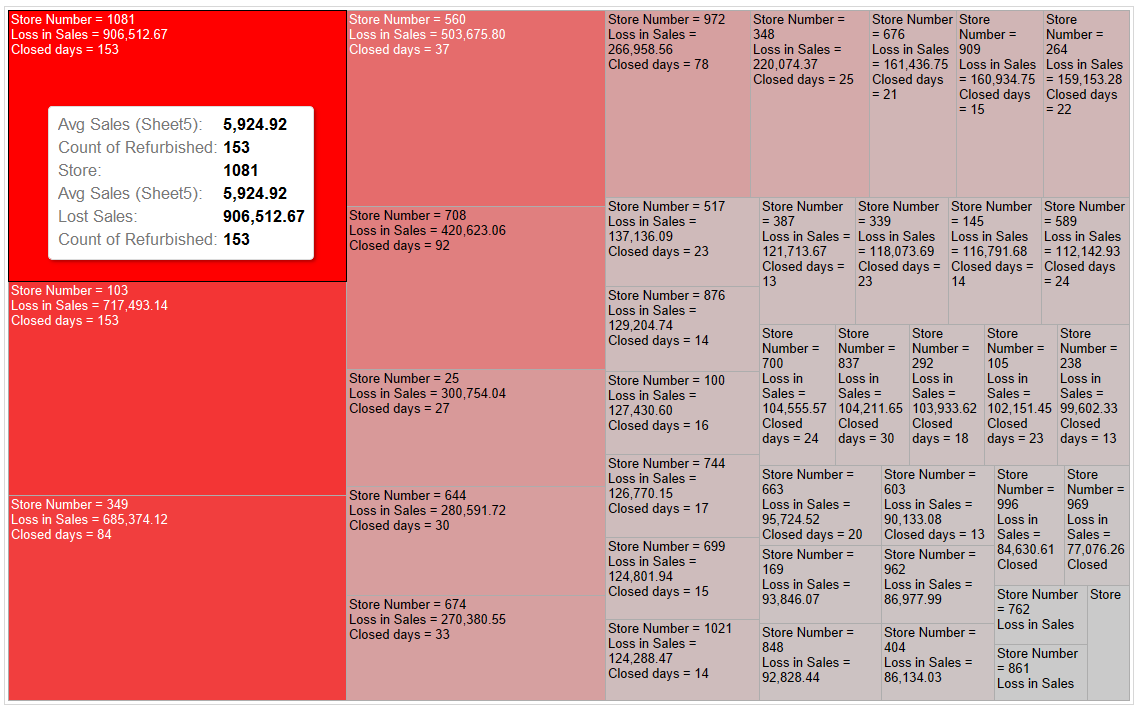
****

In the visualization shown above the average sales of each store type is calculates and has been compared with daily sales and shown the increase or decrease of sales each and every day for three years . This is represented by plotting the average sales of each store type and the increase in sales is represented by the bar raising above the average line and for decrease is the vice-versa.

*Questions Answered using the visualization 3*

From this visualization we can compare everyday sales with the average sales so that appropriate actions can be taken by the managers to decrease the loss of sales.

1. **Loss in Sales due to Refurbishment**

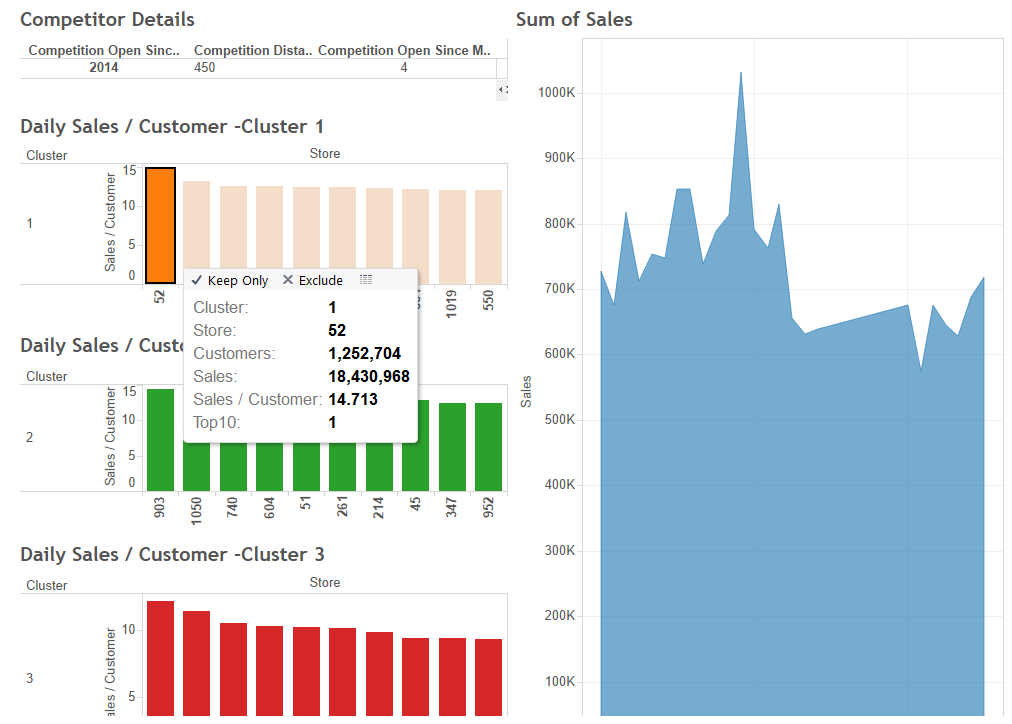
****

The tree-map visualization above gives us the loss in sales for the stores closed for refurbishment.

*Questions Answered using the visualization 4*

The visualization is used to know the loss in sales due to refurbishment and though there are less number of stores but if you consider the loss in sales that is a huge loss. From this we can say that the store should not be completely closed instead make alternative arrangements nearby if possible so that the loss incurred may be reduced.

1. **The Competitive Effect on sales**

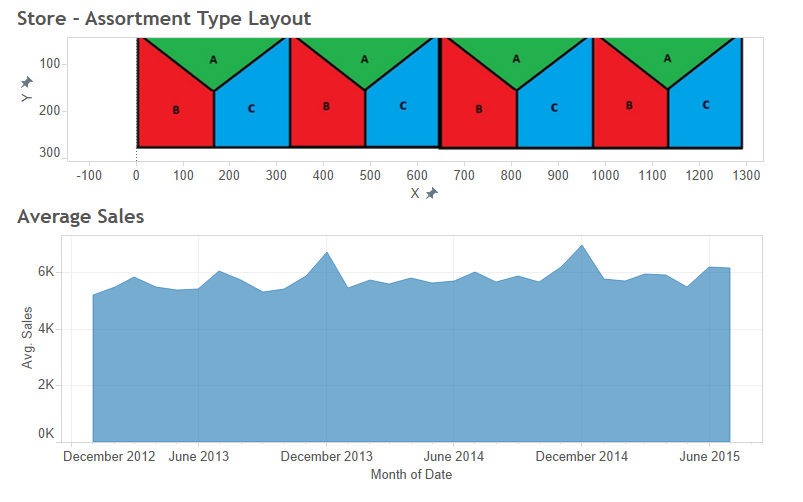


The effect of sales based on competitors near by the data has been collected the distance from which they are located and calculated according to the distance as nearer average distance and far away forming three clusters. The bars with different colors on the left represent the stores with total sales information in three clusters and on the right it represents the sum of sales of that particular store.

*Questions Answered using the visualization 5*

From this visualization we can determine the effect of competitor stores based on the distance they are separated. Using the management team can predict the reasons for the sales loss for the stores and improve them accordingly to minimize the impact on sales.

1. **Assortment Analysis**



From the visualization shown above we have the store types and including them we also have the assortment type which is categorized as a,b,c. The assortment is maintaining different types of products at different types of stores.

*Questions Answered using the visualization 6*

In the visualization above we can get the sales count according to the store type and assortment type so that we can maintain the assortment to maximize the sales in all types of stores. The Store assortment layout has been shown above and the average sales per assortment type with respect to store type is also shown.

**Tools Used and Improvements**

The tools used to implement these visualizations are Tableau 9.2 and Microsoft Excel for the data. The tool tableau is used in this project to implement the visualization as this tool doesn’t require IT skills like programming. Tableau follows good GUI and easy to understand. It is also used for the business people enabling them to analyze and build reports themselves. And as I choose the dataset is in the format of excel and the data is of sales it is easy to implement using Tableau.

In this project initially I have tried to do it with d3.js but felt that for the sales data Tableau is the best tool to visualize. And I don’t think of any other option that effective in visualizing the sales data that is the main reason Tableau is selected to implement this.

The plan is to apply few options in order to progress this project further. The options taken in to consideration are implementing either Java Script Library D3.js or InfoVis tools to develop the visualizations for the audience to use.

I would like to implement this next time by connecting to the data stored in database and create the templates and automate the process using high level programming language and visualizations are created on demand. And also I am planning to work out the implementation using matplotlib.

The purpose of the visualization is to equip the higher authority and management of ROSSMANN to have an easy access to ROSSMANN stores’ sales information and allow them

to strategically analyze if promotional offers in these stores effect the sales.