

Fall 2019

Business Metrics Dashboard for FCStone

Prasanth Vijaya Baskaran

Follow this and additional works at: <https://lib.dr.iastate.edu/creativecomponents>



Part of the [Business Analytics Commons](#), and the [International Business Commons](#)

Recommended Citation

Vijaya Baskaran, Prasanth, "Business Metrics Dashboard for FCStone" (2019). *Creative Components*. 437.
<https://lib.dr.iastate.edu/creativecomponents/437>

This Creative Component is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Creative Components by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

BUSINESS METRICS DASHBOARD FOR FCStone

By
Prasanth Vijaya Baskaran

*A Creative Component submitted in fulfillment of the requirements
for the degree of Master of Science in Information Systems at the
IOWA STATE UNIVERSITY OF SCIENCE AND TECHNOLOGY*

Submission Date: 18th November 2019

Creative Component Committee
Dr Anthony Townsend, Major Professor

Ivy College of Business

Iowa State University

Ames, Iowa

2019

Acknowledgment

Foremost, I would like to express my sincere gratitude to Dr. Anthony Townsend, my major professor for his continuous support and guidance throughout my graduate studies and related research, for his patience, motivation, enthusiasm, and immense knowledge. This project would not have been possible without his precious support and guidance.

Additionally, I would like to thank the graduate college faculty and staff and all the professors for their constant support and guidance throughout my graduate studies.

Also, I would like to extend my gratitude to my family, friends, and peers for their unfailing emotional support, timely encouragement, and endless patience.

Contents

1. Abstract.....	3
2. Introduction	3
3. Literature Review	6
4. Implementation	9
5. Conclusion.....	19
6. Future scope	20
7. References	22

1. Abstract

Dashboards are increasingly becoming high in demand in every industry since they are the easiest way to monitor business and get information to give better insight and thereby help make better decisions. The aim of this project is to leverage the use of the analytical tool Shiny to create a dashboard for FCStone (referred as “the organization” henceforth) that helps them visualize a variety of Business Metrics like customer count, brand awareness, customer experience, social media sentiment and Stock Market details.

Text Mining was done on data from Customer Feedback Survey and Twitter Data to perform sentiment analysis. This analysis along with customer count data, online customer activity data and stock market data were represented in the form of graphs or charts. These graphs or charts are generated with the help of packages in R and displayed together in a Shiny dashboard. This will assist the organization to display their key business metrics in a way that is easy to consume for their Business partners.

2. Introduction

There are many definitions for Business Intelligence but the underlying theme in each of these definitions is to provide useful information that can be helpful to decision making. Traditionally this information was provided in the form of spreadsheets consisting of rows and columns. But with the advent of big data, this is no longer a rational way of presenting information. The alternative is to use visualization tools that present this data in the form of graphs or charts. And the most generic form of visualization that has become ubiquitous to organizations is the Dashboard.

Dashboard is a tool for the end user that consists of multiple visual data that help them to monitor key business metrics and get insightful information to make better decisions. Dashboards are designed in such a way that they have a specific purpose and cater to a set of audience. The important points while designing a dashboard is to understand who the dashboard is designed for and how the information displayed will help the people using it. This dashboard designed for FCStone is for providing insights about brand awareness and customer sentiment about the organization. This is achieved with the help of the following graphs or charts.

Stock Details:

This chart provides information about the stock market history of the organization. It gives details about open and close price, volume traded each day, moving average graph along with Bollinger bands. This helps the end user to get all the details about the organization's performance in the stock market.

Twitter Sentiment:

This graph provides the details about what customers have expressed about the organization in the social media. This is a good indicator to understand the sentiment of the customers and the consensus about the organization.

Web Sentiment:

This graph is a representation of the data that is collected from online survey ratings from the users to understand their experience about the ease of access to their account, customer service provided and the design of the website. The information from this graph could help the organization realize the areas for potential improvement in their website and thereby enhance customer experience.

Customer Feedback:

The data used for generating this graph is a bundle of comments received from the customers about their experience with the organization. The graph displays the most common issues faced by the customer. This insight about the customer experience could help the organization to refine their processes to serve the customer better.

Customer Count:

This graph is a heatmap based on the number of customers in each state of the country, United States. This helps the organization to determine their presence in each state and understand their regions of stronghold. This information could help them to further drill down and understand what makes certain regions tick. They could use this insight and implement it in the regions where they still have not established themselves.

Each of this information provides a value on its own and could be utilized to enhance customer experience but when they are combined and viewed in the form of a dashboard, they add a whole new dimension into understanding customer experience and thereby help the organization make informed decisions backed up by solid visual evidence.

The report is organized in the following manner: Section 3 gives the literature review on dashboard design, Sentiment analysis of unstructured text and the data visualization techniques. It is followed by the Section 4 that discusses the methods used to design and implement the dashboard. Section 5 provides the conclusion and discusses the potential value that the dashboard provides to the organization. The concluding section discusses about the future scope of the Shiny Dashboard tool along with its pros and cons.

3. Literature Review

The literature review is done on all different topics involved in the design of the dashboard, that includes research works done on dashboard design, Data Visualization, Twitter Sentiment Analysis and Text Mining to understand the best way to create the dashboard by implementing best practices.

Dashboard Design and Data Visualization:

Majority of the research literature on dashboard iterate the fact that dashboards are driven by data and they should support the end user in making informed decisions that help Businesses. One of the most cited resource for dashboards is by Stephen Few which defines dashboard as “A visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance” (Few, 2006). The research paper also provides some valuable guidelines in designing a dashboard describing what is appropriate. For example, the paper advises against the use of 3D graphics and recommends the use of simple graphs. This is based on the study of how human brain processes visual information. The dashboard designer should ensure complex designs are avoided wherever possible. This helps the end user to determine patterns and relationships to make relevant decisions.

Dashboards have evolved quite significantly since the advent of big data. They started out as static reports given to management. They have now turned into a visual tool in which key business metrics can be monitored. Of late, Dashboards consists of technological functions that integrate with the Business workflow management. According to a survey conducted by Howson in 2012, Dashboards are on top of organization's priority list as shown below.

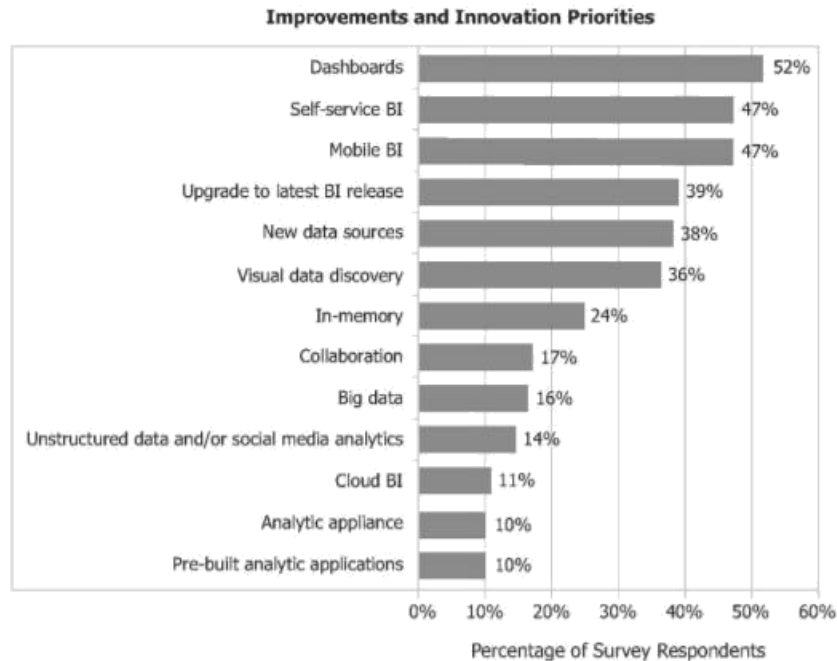


Figure 1 Survey showing priority lists of organizations

The research papers on Dashboards also emphasize the importance of not overwhelming the end user with an overload of information. An ideal dashboard should be able to provide quick insights, help make better decisions, focus on the user requirement, track the progression of Business and most important of all, should be understandable to the user with minimal training.

There is scope for future research in dashboard design in the following areas: Functional Fit, Cognitive Fit with User Tasks and Knowledge, Cognitive Fit with User Personality, New Dashboard Design Features, and Organizational Factors (Yigitbasioglu & Velcu, 2012).

Text Mining and Twitter Sentiment Analysis:

Text Mining or Text Analytics is the process in which unstructured text is converted into structured data to identify key relationships or patterns that can help improve Business. In this project, the textual data is a collection of comments from customer feedback and a series of tweets about the organization. The general process to convert unstructured text involves having the data organized in

a repository where it is easy to access, followed up by cleaning the data to convert it into a structured format followed by creating the term document matrix which is the most commonly used format for calculation (Feinerer et al, 2008). This is the method used in this project to perform text mining. The process is done on customer feedback comments and tweets about the organization that might help them understand their areas of strength and their reputation among people.

Every organization has started dedicating resources and investing money to establish their presence on the social media platforms. Social media analysis has become inevitable for organizations in this day and age. By analyzing the tweets organizations can understand their brand awareness among people. The sentiment analysis of tweets is a natural language processing task that has come a long way from the time it was analyzed first. The early sentiment analysis of the tweets was to recognize tweets ending with smile emoticons like ':)' or ': -)' as positive tweets and to classify tweets with sad emoticons like ': (' or ': - (' as negative tweets (Pak and Paroubek, 2010). This was a very basic model introduced when Twitter was a relatively new social media platform. From there research has come a long way in using Tweets to analyze people sentiment. Twitter data has become a representation of the people. If properly utilized it can be a source to understand the pulse of the people. It is to be noted that the data retrieved, viz. the tweets, is unstructured. It requires data pre-processing and comparison with a pre-defined corpus. Each word of the tweet is compared, and an overall score is assigned to determine whether the tweet is positive or negative (Agarwal et al, 2011). This technique is used in this project to determine the polarity of customer sentiment in the Twitter platform for the organization. This is a good indicator of the brand awareness among people. Sentiment Analysis can be done at various levels of granularity like word level, phrase level, sentence level or document level. In this project the analysis is done on a word level to ensure that there is a high level of accuracy in determining the customer sentiment.

4. Implementation

The dashboard is implemented as a Shiny application. Shiny is an open source package provided by RStudio providing a framework for interactive web applications. It has two components namely UI. R and Server.R. The UI. R file is created for designing the user interface and determining the front-end design for the end user. The Server.R file contains the logic to display the desired output.

The dashboard essentially consists of three components: a header that displays the title information, a sidebar that displays the subtitles and finally the body that shows the actual information for the user.

The dashboard designed in this project consists of the following information: Stock Details, Customer Feedback, Web Sentiment, Customer Count and Twitter Sentiment. The implementation of each of these components in the dashboard is described below.

4.1 Stock Details:

The Stock Details tab of the dashboard provides the information about the stock history of the organization. The getsymbols function in the “quantmod” package of R is used to get stock data from Yahoo finance website. The data is converted to a data frame and then saved in the form of .csv file. This data consists of the following details of the company’s stock: Date, Open price, Close Price, Volume traded, Maximum and Minimum stock value on each day. These details are all gathered and displayed in the form of a chart as shown below.



Figure 2 Stock Details chart

The type of chart used is a candlestick chart because of its intuitiveness. It shows the user the movement of stock on each day, the high and low along with the open and close price range. The volume traded is displayed as a bar chart right below the stock price details candle chart. In addition to these details, the moving average for 30 and 60 days is overlapped onto the stock chart as per the end user request. Bollinger bands, a useful tool in determining the trend of a stock is added as well. Bollinger bands has an upper band and lower band. The closer the prices move to the upper band, the more overbought the market, and the closer the prices move to the lower band, the more oversold the market. A stock is more likely to fall in value as it is overbought and increase in value when it is oversold. A range selector button is provided to select specific time interval and view the stock trends in that timeframe. The chart is also provided with options to choose stock details for a month or 3 months or a year as per the end user's request. A reset button is provided to set the chart back to default. The user can hover over the chart to check the values of the stock and provided with options of zooming.

4.2 Customer Count:

The customer count chart is a heatmap based on the number of customers in each state in the United States of America. A state in darker shades means that it has a higher number of customers and a very light shade indicates that the organization does not have a strong presence in that state. The data was provided by the FCStone organization. The heatmap in the dashboard is as shown below.

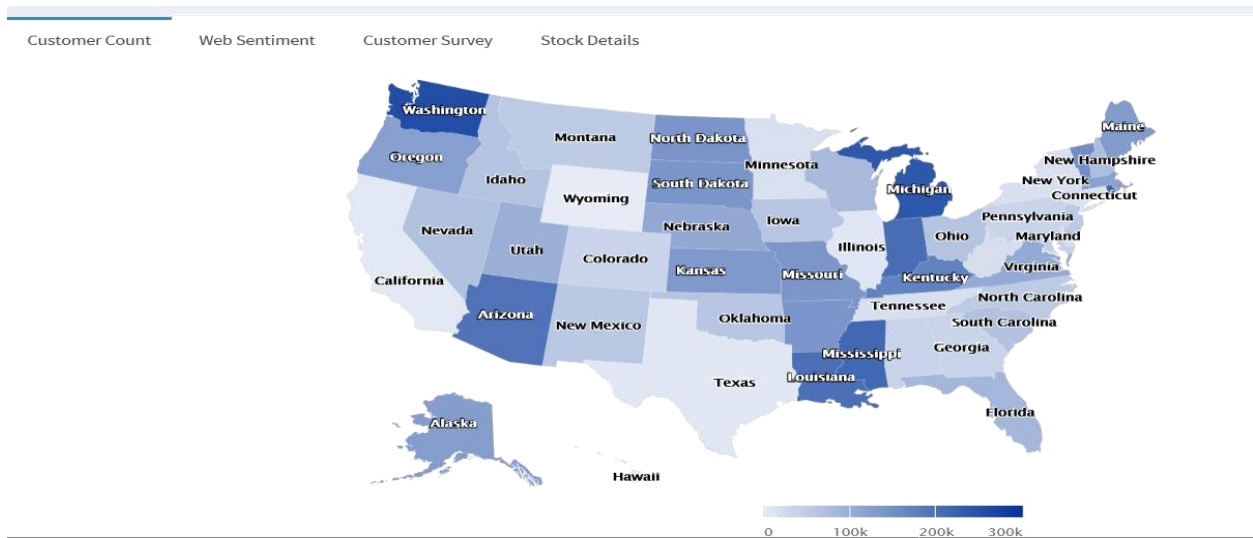


Figure 3 Customer Count Details

The coloring used is in line with the organization's design and if the user hovers over each state the count of customers is displayed along with the state name as a tooltip. A heat map is used in this case because they have the capability to display data in a simple visual format that is very intuitive to the end user. It is easy to figure out which region has a strong presence for the organization and the regions in which the organization is yet to establish its presence.

4.3 Customer Survey and Web Sentiment:

This part of the dashboard shows the chart that displays the details showing customer behavior online. The data is provided by the organization and the first part of visualization shows the details of network traffic through an organic search. The second part of the visualization is for data for paid search campaigns. The charts are shown below.

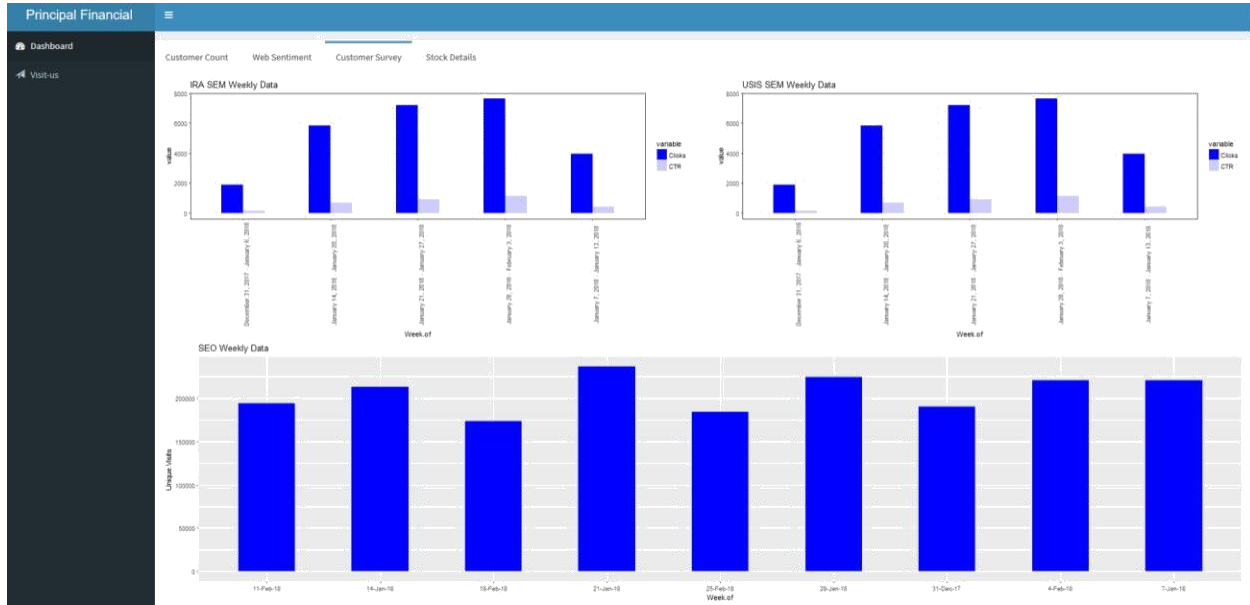


Figure 4 Network traffic and Unique Visits

The next chart in the dashboard shows the Web sentiment based on customer survey. It gives the details of the customer's response online to various options provided to them by the organization's website. The data was processed by the organization and given for this project in a structured format. Only the design of the chart was the activity here as data processing was already done.

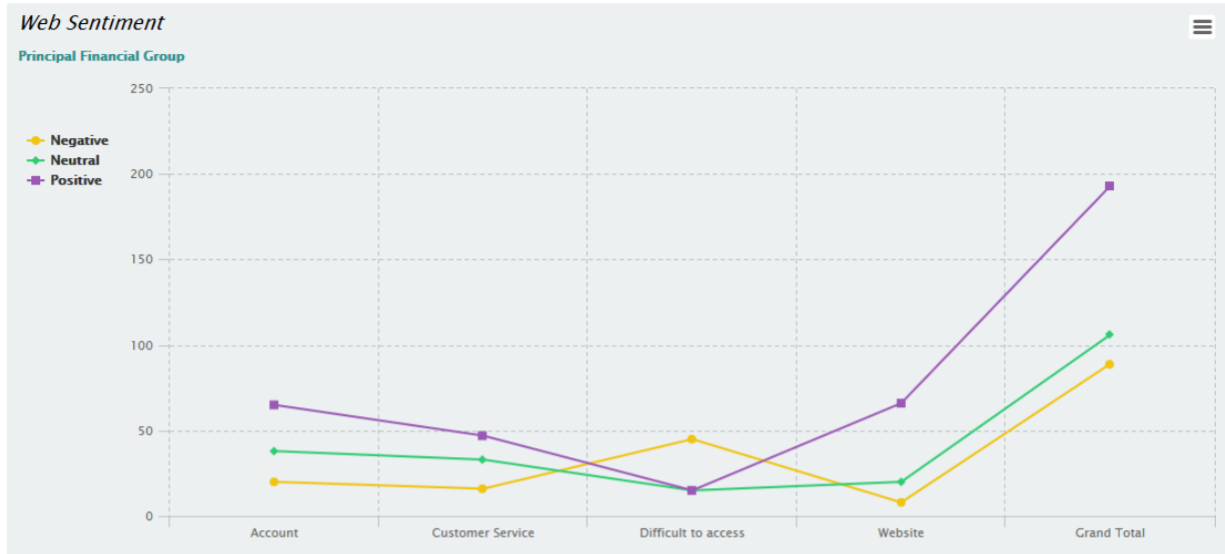


Figure 5 Web Sentiment Chart

There is nothing easier to understand than a simple line chart as it encapsulates the information very precisely. Contrasting colors are used to distinguish between the polarity of the feedback to ensure the end user gets a clear picture. The count of values is presented as a tooltip to the user when they hover over the graph. Prominent gridlines are shown to help the user to differentiate between categories and to get an idea of the values.

4.4 Twitter Sentiment:

From an organization perspective it is important to understand the people's feedback on any of their products or services. This can be done without much expense by using the data from Twitter. Hence sentiment analysis on the tweets is performed in this project to understand the pulse of the customers.

The process of tweet sentiment analysis is done as follows: 1. Get tweet data from Twitter application 2. Clean the data to perform analysis 3. Calculate sentiment of the tweets 4. Visualize the mean sentiment to get the overall sentiment analysis.

Twitter has enabled the user to analyze the tweets posted with the help of an API. This API allows the user to create a Twitter app provided there is a Twitter account. The app asks for some basic details like name and description. Once these details are entered Twitter provides the user with Consumer Key, Consumer Secret, Access Key Access tokens. These four parameters are implemented in the R script to extract tweets.

The package “twitterR” in R help us extract tweets from the Twitter application. In addition to this package we use the “tm” and “SnowballC” packages to help us with text mining and analysis which includes pre-processing the data and cleaning to convert it into a structured format. Once we have invoked the app with the help of our keys and tokens we can download the data from twitter. The initial data downloaded is unstructured and has a total of 16 variables. The focus for our project is the variable “text”. This column consists of the body of the tweet. It is to be noted that this field has to be cleaned by removing the unnecessary details for sentiment analysis like URLs, hashtags and other peoples’ twitter handles mentioned in the tweet. The “gsub” function in R does the job. It removes these terms from the body of the Text field.

Now the Text field only has the relevant text of the tweet on which the sentiment analysis is done.

To perform sentiment analysis, there is a package in R named “Syuzhet” that has an in-built dictionary consisting of more than 10000 words to compare to. Each word in a tweet is chosen and categorized as positive or negative and then the overall weight is calculated to get the mean sentiment of the tweet. This action is performed with the help of a built-in function in R named “get_sentiment”. Once this is finished the tweets have their own individual sentiment score. This sentiment score of each tweet is displayed as a chart in the dashboard to enable the end user to view the overall sentiment of the organization. In addition to this another visual showing the specific emotional values of tweets is also displayed. This can be achieved with the use of “get_nrc_sentiment” function in R.

The charts are as shown below.

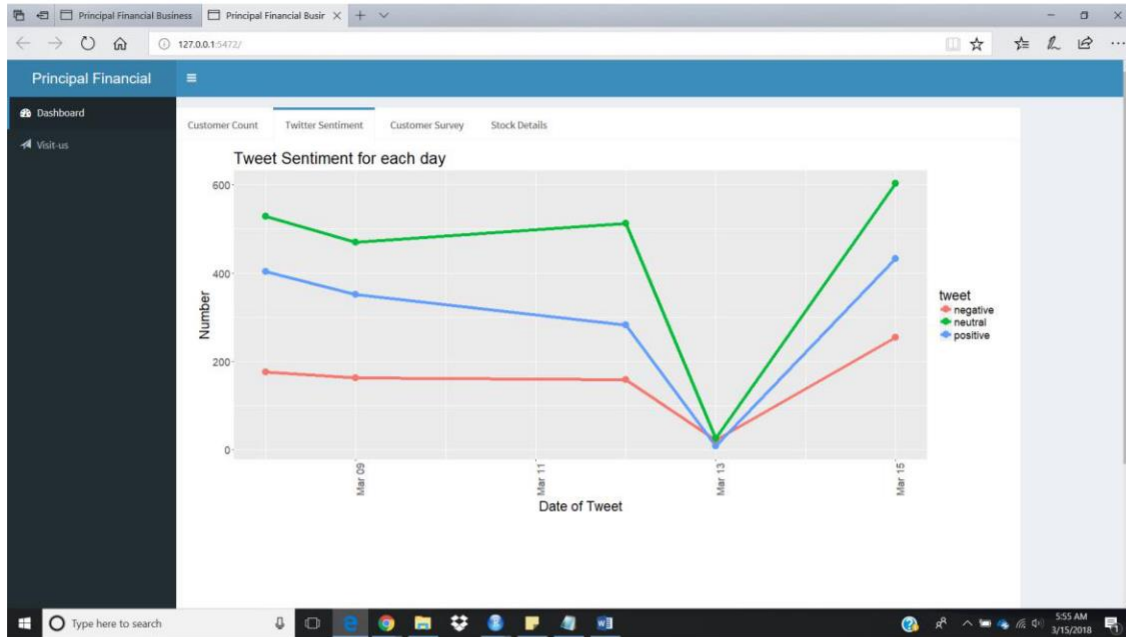


Figure 6 Tweet Sentiment

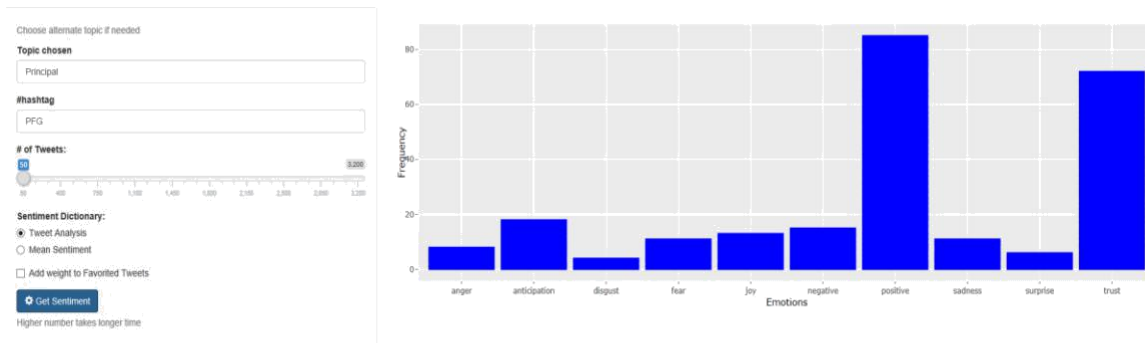


Figure 7 Tweet Split Sentiment

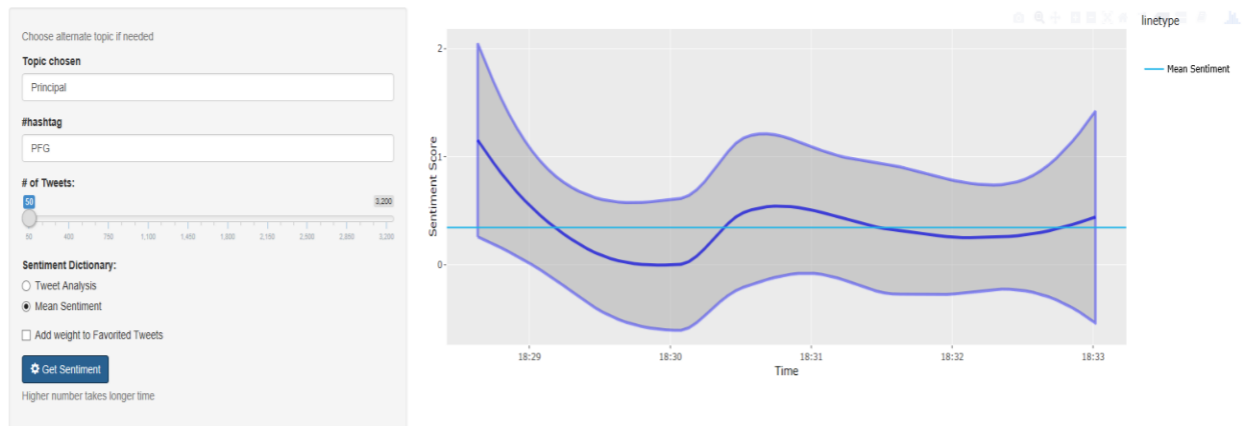


Figure 8 Mean Tweet Sentiment

The charts are designed to make it easy for the user to understand and follow the dashboard. The line charts and bar charts provide an uncomplicated way to understand the information being presented and make informed decisions. The colors are used in a way to enhance the user experience and are inline with the organization's website design. An option has been given to the end user to search for any specific term in Twitter and get the sentiment thereby not restricting the user capabilities to use the dashboard and making the dashboard more interactive.

4.5 Customer Feedback:

The data used for this part of the dashboard was provided by the organization and consisted of customer feedback comments about their experience on their website. The data was mined to identify where the customers found it most difficult to access. This can be achieved by plotting the most frequently used words in the comments. The process of text mining done is explained as below.

Loading data:

The data provided was loaded into R studio and then as it was done in Twitter Sentiment analysis, the packages “tm” is loaded. Each comment in the provided data is treated as a separate document and a corpus of these comments is created.

Pre-processing:

The data has to be cleaned to perform further analysis. The tm package has various functions that can be used to remove numbers, punctuation and white space. These functions are utilized to clean the documents. The next step in pre-processing is to convert all the text in the documents to lower case. The `content_transformer(tolower)` function can be used to achieve this. Once this step is completed, we need to ensure that the common words or stop words in the English Language do not occur in the final output. The advantage of using tm package is that it has a list of stop words and the documents can be compared to this list to remove these words using the `removeWords` function. The `removewords` function can also be customized to remove any words we deem unnecessary to be excluded from the documents. This completes the pre-processing task and we have data that is devoid of punctuations, numbers and stop words.

Stemming:

Stemming is the next step in Text Mining. The corpus contains a large number of words that will have a common root word. For example, run, ran and running all have the same root word run. The process of Stemming is done to ensure that all the related words in different documents are reduced to their roots. The SnowballC package is used for this purpose. It has an in-built function called `stemdocument` for the purpose of stemming. After stemming is complete we create a document term matrix. This matrix is a list of all occurrences of the words in a document. In this project there is a total of 2016 unique words and a total of 1624 comments. Since each comment is a document, the document term

The first chart is a histogram showing the most frequent terms in the customer feedback comments and the number of times they appear. The second chart is a Word cloud of the most frequent terms which is a fancy way of representation of the same details.

All these charts are integrated together and displayed as a dashboard and published as a web application for the end user. The packages used for plotting the charts are ggplot2, plotly, highcharts and wordcloud. The packages shiny, shinydashboard and shinythemes are used for the design of the dashboard.

5. Conclusion

Dashboards have become an important part of the organization and they serve as a one stop point for key business analysis. A dashboard is more than just a summary of the data and they can be used to drive the strategy of an organization, discover key insights and deliver operational excellence. It is important that appropriate charts are used in the design of a dashboard so that they reveal patterns and are easy for the user to follow. Dashboard designs need to be consistent and provide clarity in the information that it provides. It is also vital that an overdose of information is avoided. It is very easy to fall into the trap of trying to provide too much. An ideal dashboard focusses on the needs of the end user and provides aesthetic, easy to interpret visualization that could help drive Business solutions.

6. Future scope

Shiny is a very useful and upcoming tool for dashboard and visualizations. It may not have the powerful automated features of Tableau or Power BI but these features and many more can be implemented as a Shiny web application by coding the requirements in R which forms the base for the creation of Shiny apps. With new features and functionalities being added very frequently and a strong R community, it is not long before organizations turn their focus towards this tool. The pros and cons of Shiny are discussed with respect to the existing functionalities.

Advantages:

- The dashboard can be made interactive, user friendly and can be completely automated.
- To design Shiny dashboard as a web application the knowledge of HTML, CSS or JavaScript is not necessary.
- The Shiny application is very powerful and advance analytics can be performed. It can handle data irrespective of the complexity as it is coded in R and the language has very good functionalities and packages that can handle complex data.
- It is available as an open source tool and building your app in Shiny is completely free.

Disadvantages:

- Since it is still in the process of development the app may require frequent updates.
- There is no selective access in the free version. It is accessible to everyone once it is deployed.

Shiny Server and Pricing Details:

For any shiny app to be implemented, it needs Shiny Server to run. There are 3 different options for implementing the application.

1. Implementation of Shiny Server is free via ShinyApps. (Used in the project)

Completely free and there is no need to pay for your own server. ShinyApps should be online, which has so far not been a problem since it is hosted on the cloud. It has a reputation for being robust.

2. If there is an own server in the organization, then you can install the open source version of Shiny Server.

Secure because it is on your own server. However, someone has to install Shiny Server and update it whenever it's necessary to do so. The app is as reliable as the server on which it is hosted.

3. Pay for Shiny Server Pro and can install it in your own server.

This option allows you to have high level password security (via SSL) to any Shiny app you host. RStudio provides technical support for Shiny. The main disadvantage is that this option is very expensive. It costs \$9,995 per year per server. More details on <https://www.rstudio.com/pricing/>.

7. References

1. Howson, C. (2014). *Successful Business Intelligence: Unlock the Value of BI and Big Data* (Vol. Second edition). McGraw-Hill Education.
2. Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, 13, 41–59.
3. W Presthus, CA Canales (2015). *Business Intelligence Dashboard Design. A Case Study of a Large Logistics Company*. Norsk konferanse for organisasjoners bruk av IT.
4. P. Paroubek and A. Pak (2010). "Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In *Proceedings of the Seventh Conference on International Language Resources and Evaluation*, pp.1320–1326.
5. R. Bhayani, A. Go, L. Huang (2009). "Twitter Sentiment Classification Using Distant Supervision". Stanford University, Technical Paper.
6. Bifet A. and E. Frank (2010). "Sentiment Knowledge Discovery in Twitter Streaming Data", In *Proceedings of the 13th International Conference on Discovery Science, Berlin, Germany*: Springer, pp. 1–15.
7. Agarwal A., B. Xie, I. Vovsha, O. Rambow, R. Passonneau (2011). "Sentiment Analysis of Twitter Data", In *Proceedings of the ACL 2011 Workshop on Languages in Social Media*, pp. 30–38
8. Feinerer I., K. Hornik, D. Meyer (2008). "Text Mining Infrastructure in R", *Journal of Statistical Software*, Vol.25
9. Temple Lang D (2004). *Rstem: Interface to Snowball Implementation of Porter's Word Stemming Algorithm*.

10. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.
11. González-Ibáñez, R., Muresan, S., & Wacholder, N. (2011). Identifying Sarcasm in Twitter: A Closer Look. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. IBM.