**ANALYSIS OF SOCIAL MEDIA FOOTPRINT OF COMPANIES**

**Social Media Platform Chosen: Facebook**

**Team Members:**

**16BCE0717 – SHIVAM BHAGWANI**

**16BCE0594-MOHITH KRISHNA V**

***Report submitted for the final Project Review of***

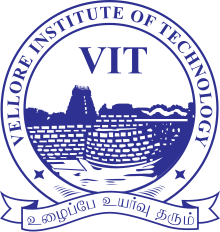
**Course Code: CSE3021**

**Course Title: Social and Information Networks**

***to***

**Professor: Dr. Anuradha J**

**Slot: A2 + TA2**



**DECLARATION**

We hereby declare that the project entitled “**ANALYSIS OF SOCIAL MEDIA FOOTPRINT OF COMPANIES**” being submitted in partial fulfilment for the degree of Bachelor of Technology in Computer Science Engineering at Vellore Institute of Technology is the authentic record of our own work done under guidance of our guide Prof. Anuradha J.

Shivam Bhagwani

B. Tech. Computer Science

Vellore Institute of Technology

Vellore, Tamil Nadu

Mohith Krishna V

B. Tech. Computer Science

Vellore Institute of Technology

Vellore, Tamil Nadu

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S No.** | **TITLE** | **PAGE NO.** |
|  | **ABSTRACT** | **3** |
|  | **Introduction** | **3** |
|  | **Literature Review** | **4** |
|  | **Tools Used** | **6** |
|  | **Methodology** | **6** |
|  | **Implementation and Outputs Obtained** | **9** |
|  | **Conclusion** | **14** |
|  | **References** | **15** |

1. **Abstract**

Facebook Pages provide a key aspect of social media marketing to businesses all around the world. In recent times, social networking sites have provided a medium through which people can interact and gain and share vast knowledge among themselves.

Most companies have Facebook pages and regularly upload news about their latest products from their line-up. This feed, due to Facebook’s sponsorship policy with companies, appear on our timeline.

On a Facebook Page we have the option to like, react and also follow to the posts put up by companies which means that whenever a new post comes up the user will be notified about the post as soon as the users logs in.

In most experimental approaches for social network analysis the study is based on social interactions between the users of the social network. The engagement between posts and users on social media is something that is rarely studied upon. In our work, we are showing the interactions between posts and users, rather than interactions among users. We are modelling the network using a parsing.

1. **Introduction**

On a public Facebook page users have the ability to comment, like or share posts. This is the interaction between the users with the company/individual.

The three primary ways people engage with a post is:

* Comment
* Like
* Share

Analysing these metrics will help in determining which posts resonate best with the audience.

Facebook Pages play a key role in a company’s success and hence the importance of the success of its Facebook is ascertained. We will be performing Social Network Analytics on Facebook Pages of two competitors and perform a comparison between them to identify the connectivity, centrality and other metrics in this network.

In the project parsing the Facebook pages is done and data about the posts is collected. The number of likes and reacts the post has got and the number of comments. This data is collected over a period of time so we can get a better understanding of people’s opinions on a company’s Facebook posts.

Social network analysis aspects such as parsing, clustering and mining are used to produce useful and meaningful data. Once a company gets to know what posts and what products people have liked over time, they use the information to replicate the success.

1. **Literature Review**

**Summary Table:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Authors and Year (References)** | **Title (Study)** | **Concept / Theoretical model/ Framework** | **Methodology used/ Implementation** | **Dataset details/ Analysis** | **Relevant Finding** | **Limitations/ Future Research/ Gaps identified** |
| Kevin Lewis , Jason Kaufmana, Marco Gonzalez , Andreas Wimmer , Nicholas Christakis  2006 | Tastes, ties, and time: A new social network dataset using Facebook.com | The framework used here is to apply network analysis on a given dataset | Obtain permission to get data from Facebook, then  Use Profile Data of all users to perform comprehensive analysis on the network | Facebook.com public data from students at Harvard University | computerized data collection “requires fewer research resources than do personal interviews or mailed questionnaires,” making replications and meta-evaluations much more easy | students differ tremendously in the extent to which they “act out their social lives” on Facebook |
| BongwonSuh, Lichan Hong, Peter Pirolli, and Ed H. Chi  2010 | Large Scale Analytics on Factors Impacting Retweet in Twitter Network | To identify the important factors that cause people to retweet tweets | Extract Data from Twitter,  Then perform a reduction technique where correlated features will be reduced into a smaller number these are called principal components, which accounts for variance of individual components. This is followed by selecting the right number of factors then interpreting them | 10k tweets from a downloaded dataset,  74M from Twitter API | URLs and hashtags have strong relationships with retweetability. Amongst contextual features, the number of followers and followees as well as the age of the account seems to affect retweetability | Future research includes generating a predictive model which can predict retweet ability based on past retweets |
| David Ediger Karl Jiang Jason Riedy David A. Bader  2014 | Massive Social Network Analysis: Mining Twitter for Social Good | The concept used here is parallel processing to do a large scale mining and social network analysis | The methodology used here is to set up a supercomputer architecture CrayXMT and then utilize graphCT to visualize graphs and perform analysis | Entire twitter feed of Sept 09 | Social Network Analysis of Big Data is done | Future research includes utilizing the full capability of supercomputer architecture |
| Lydia ManikondaYuheng Hu SubbaraoKambhampati  2014 | Analyzing User Activities, Demographics, Social Network Structure and User-Generated Content on Instagram |  | Obtain Unique id of profiles of main feed and then crawl through their users to obtain a large number of users and avoid sampling bias. Then perform network analysis | Live profiles of certain celebrities and their followers | Social Network Analysis along without Geo location analysis is done. We find that the reciprocity is not as high as in flickr and the clustering coefficient is higher than twitter | Dataset taken is just a small portion of the total users of Instagram |
| Backstrom, Lars, and Jon Kleinberg  2014 | Romantic partnerships and the dispersion of social ties: a network analysis of relationship status on facebook | The concept used here is dispersion which looks not only at the number of mutual friends but a network structure among mutual friends | Random sampled Facebook data is scraped where partner/spouse information is enlisted, then theoretical dispersion is performed and on using machine learning to study the social structure of the facebook friends the partner can be identified | Randomly sampled Facebook dataset where users have declared a relationship | dispersion is a structural means of capturing the notion that a friend spans many contexts in one’s social life | Does not test on people who haven’t declared a relationship which means that we can’t test what the dispersion method will do in case where no relationship is present. |

1. **Tools Used**

*Softwares used:*

* ParseHub for data parsing
* Python 3.7 for Data collection
* R for Data visualization
* Facebook developer API
* Google Chrome- To Access Facebook Data

*Libraries used:*

* ggplot2
* plotly

Windows XP/Vista/7/8/10.

Computer with minimum 2GB RAM and Intel**™** i3 processor

1. **Methodology**

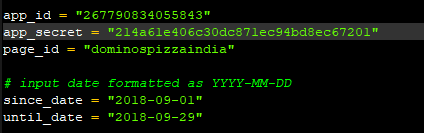
At first, A Developer Facebook account is created. On going through all the types of verifications by Facebook this account is made active. On getting an active account, a user can create an app within it. This app needs to get some permission from the Facebook App handling and verification department. The procedure is time taking.

That is why, a method which can function in the same manner is found, that is by parsing the respective pages and getting output in JSON format.

**Web Parsing:** It is a type of Web Scraping in which the data visible over the browser to a user can be saved on the user system for analysis and many other purposes. This data can be saved in many forms including SQL, JSON, CSV, Spreadsheet, etc.

Web scraping software automatically loads and extracts data from page based on user’s requirement.

After the data is obtained with the help of this software in JSON format, it is converted to CSV in this case and then one by one, loaded to R Variables and is processed.



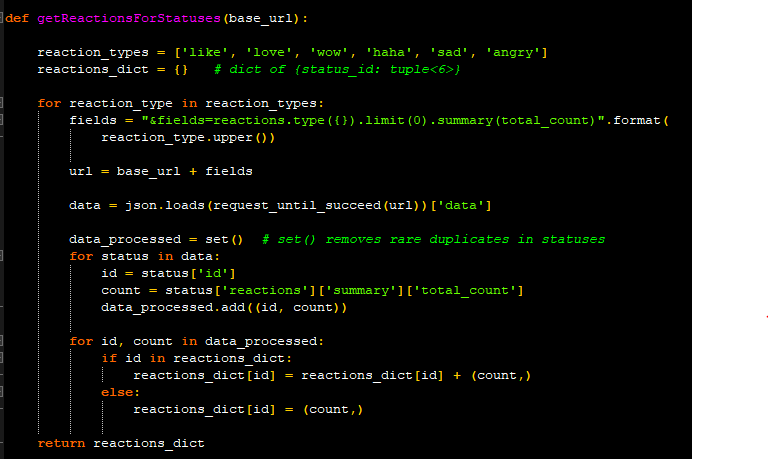


Figure : Python Parser

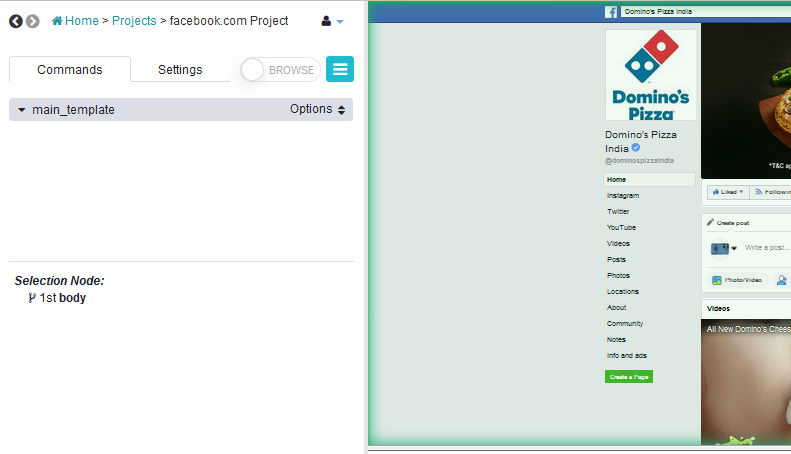


Figure : ParseHUB Software

* The data that this project uses is fetched in JSON format using the tool called ParseHUB. This JSON format is first converted to CSV Formate for further processing.
* In this project data from the online social media site Facebook about two companies which work in the same field:

1. Dominos

2. Pizza Hut

* From the pages of both of these companies on Facebook, Data about most recent posts like

1. Number of Reactions on posts.
2. Number of Comments on posts.
3. Number of Shares that a post gets.

is collected.

* In this project, the mentioned data is collected for the posts dated back upto the month of July.
* For each of the corresponding column in the CSV File, A Variable in the R Studio is created.
* For a better arrangement of data, Framing is done.
* The variables referring to data fields that have to be compared side by side are then used for obtaining best fit.
* This fitting of data is done using Linear Model which is a technique used for Linear Regression in two Variables.
  + Syntax for Linear Model Fitting is: lm(x~y). x being the independent variable and y being the dependent variable whose values are obtained on giving values to variable x at the later stage.
* After getting a relation and representing y in the form of x in each of the conditions, a graph is obtained. This graph has been analysed and at last, conclusion of the project has been made.

1. **Implementation and Outputs Obtained**

* **R Code and Graph Obtained for Shares**

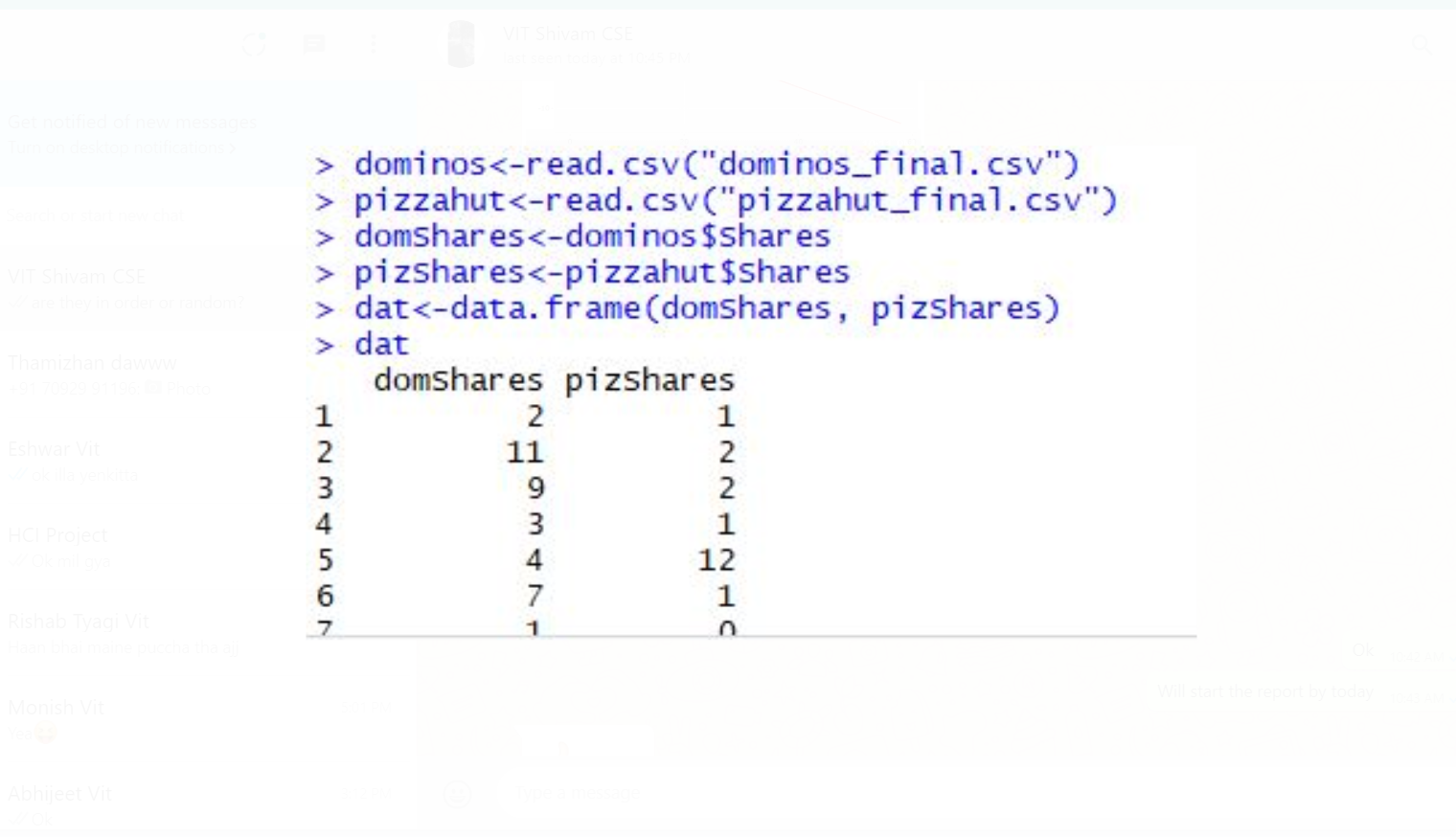
****

Figure : Data for number of shares is stored in variables

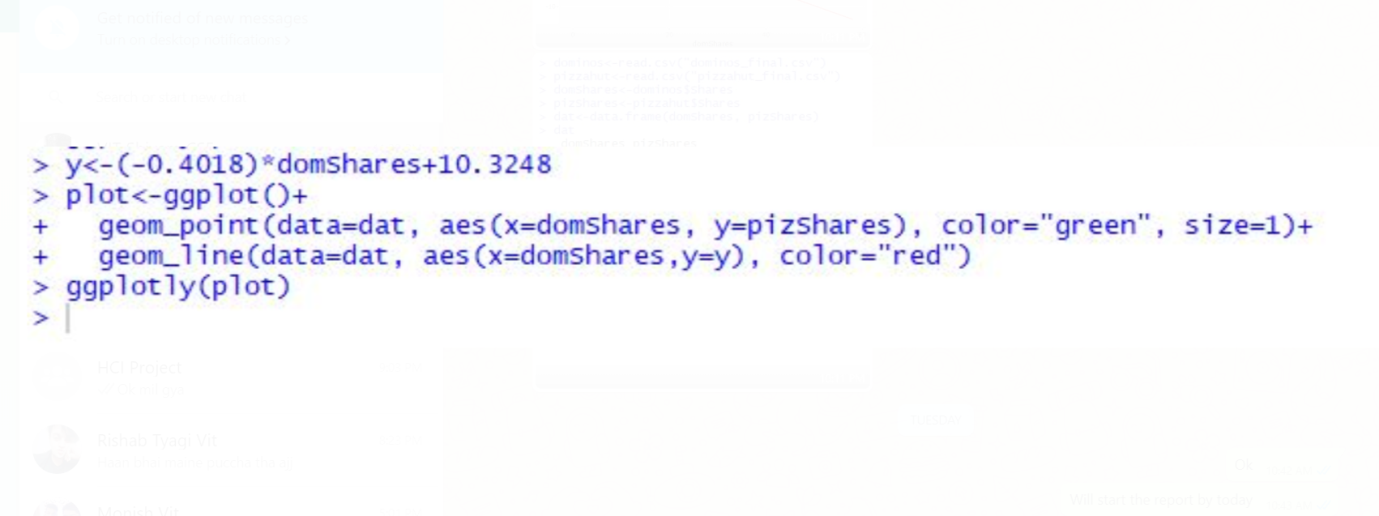


Figure : Code Snippet for obtaining Regression line based on number of shares

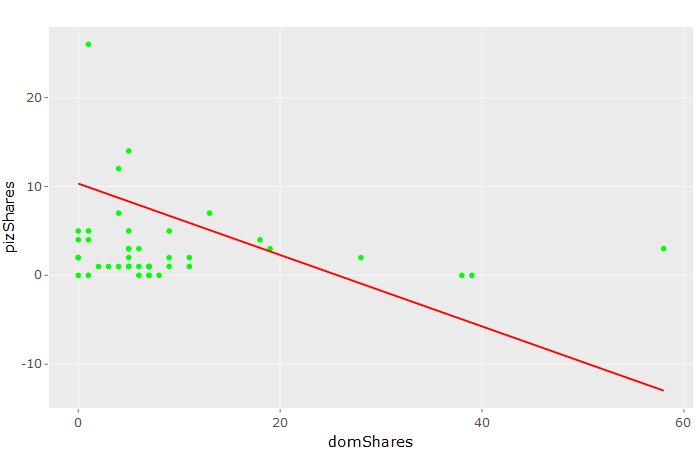
****

Figure : Regression Line Obtained from Shares

* **R Code and Graph for Reactions**

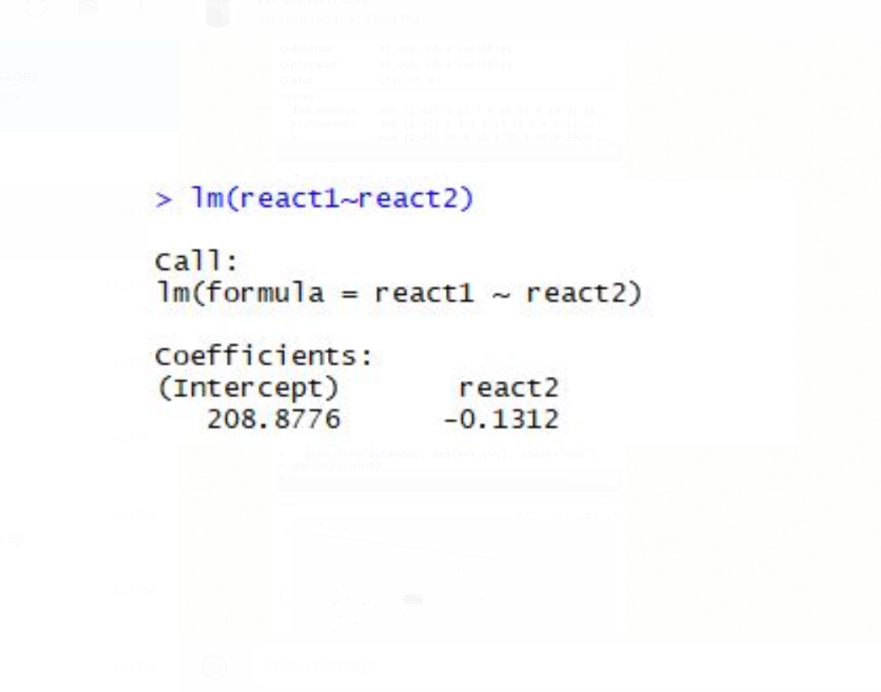
****

Figure : Finding Linear Relation between two variables

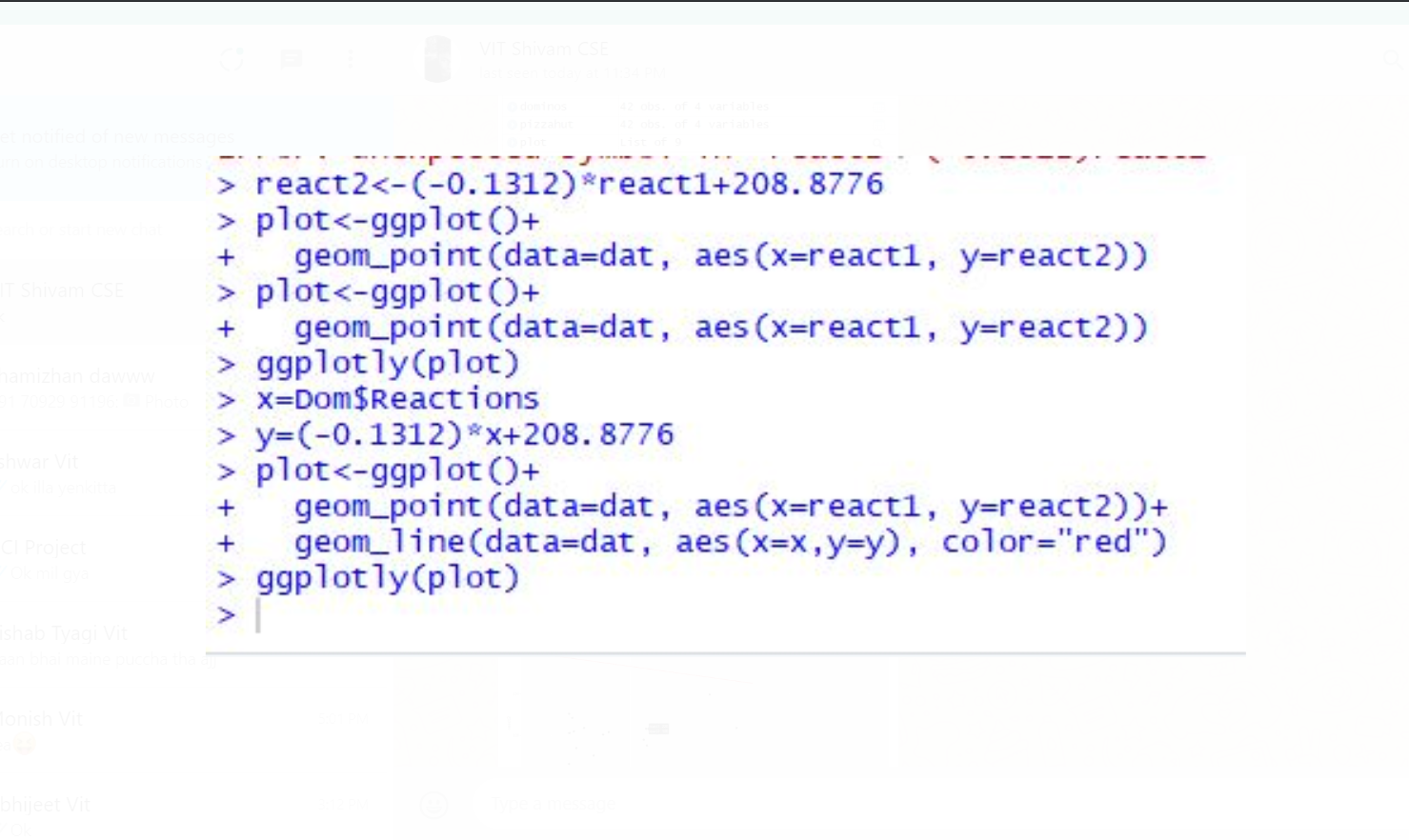
****

Figure : Code Snippet for obtaining Regression line based on number of Reactions

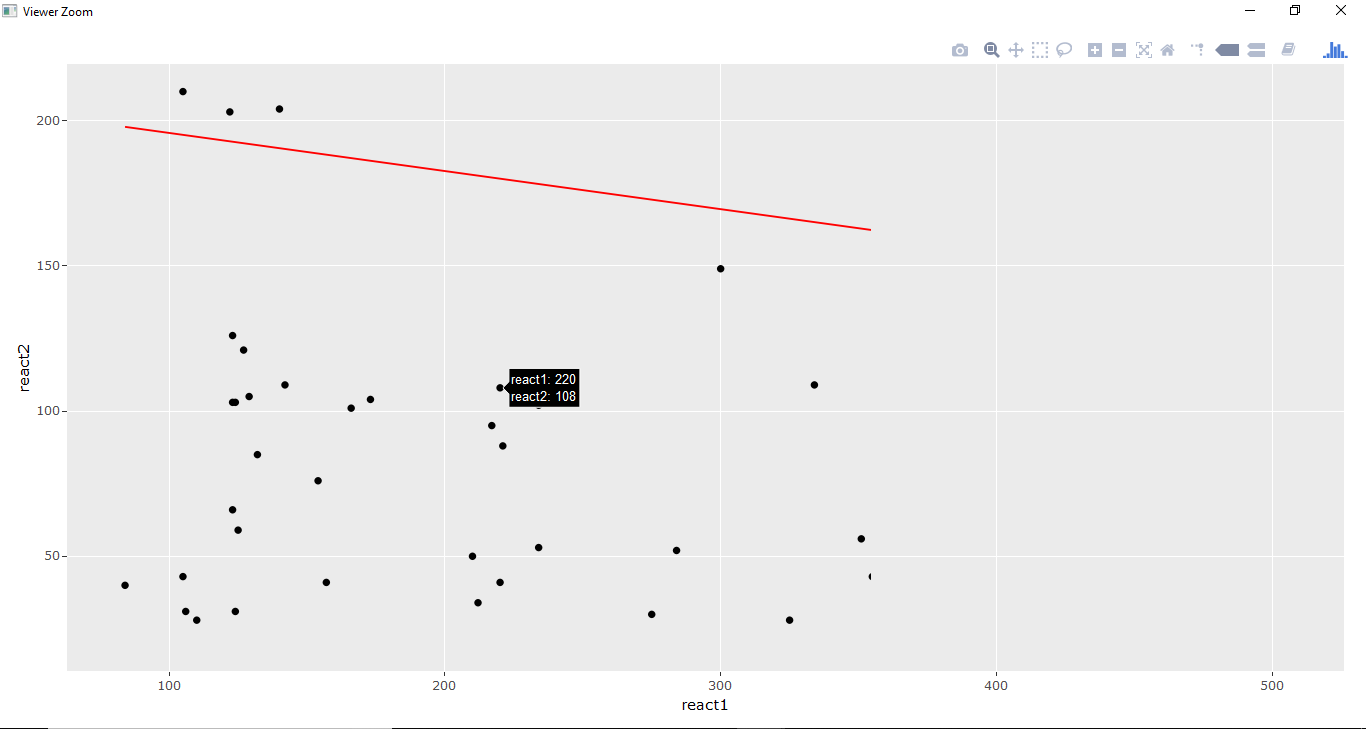


Figure : Regression Line Obtained from Reactions

* **R Code and Graph for Comments**

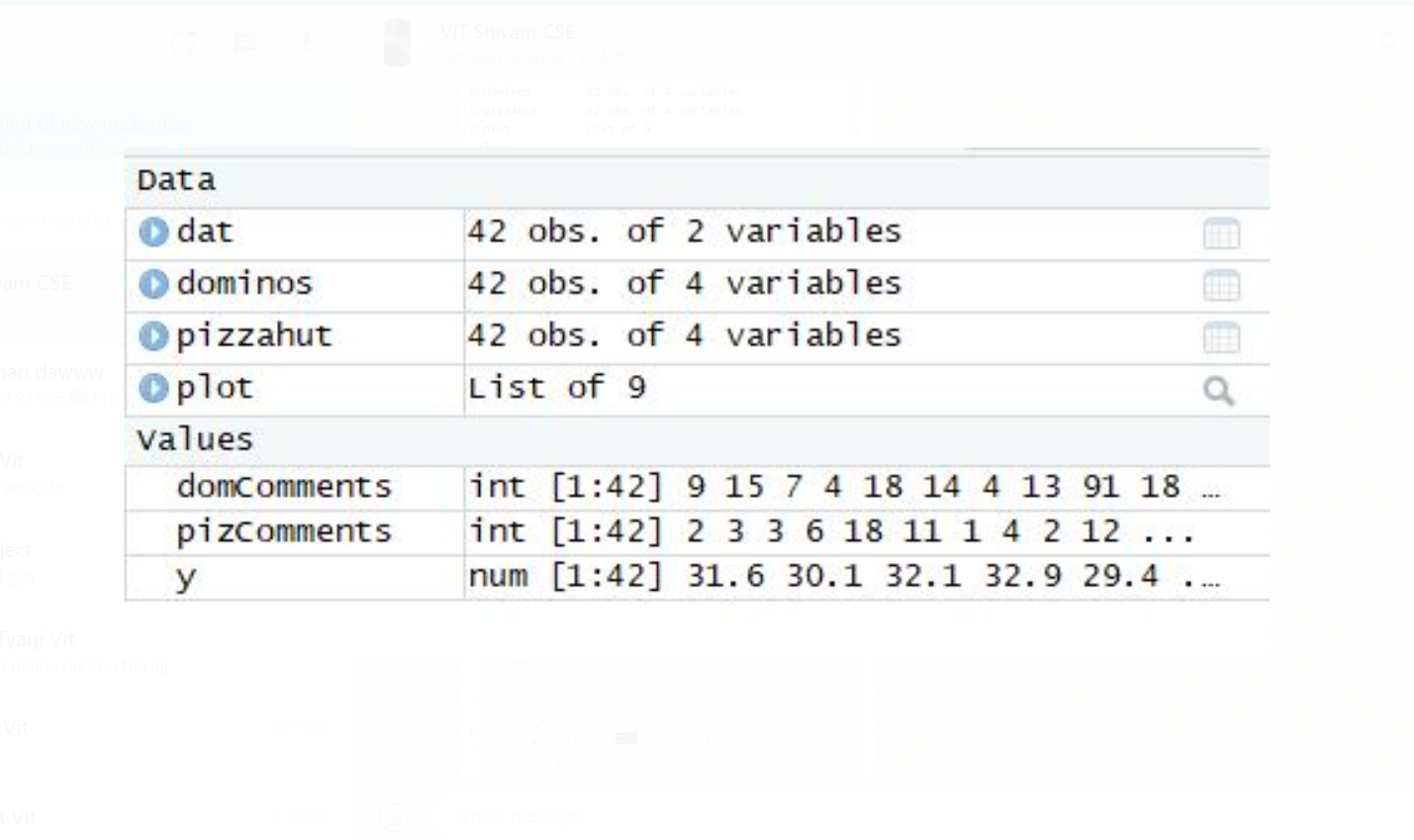


Figure : Data Representation in R Studio

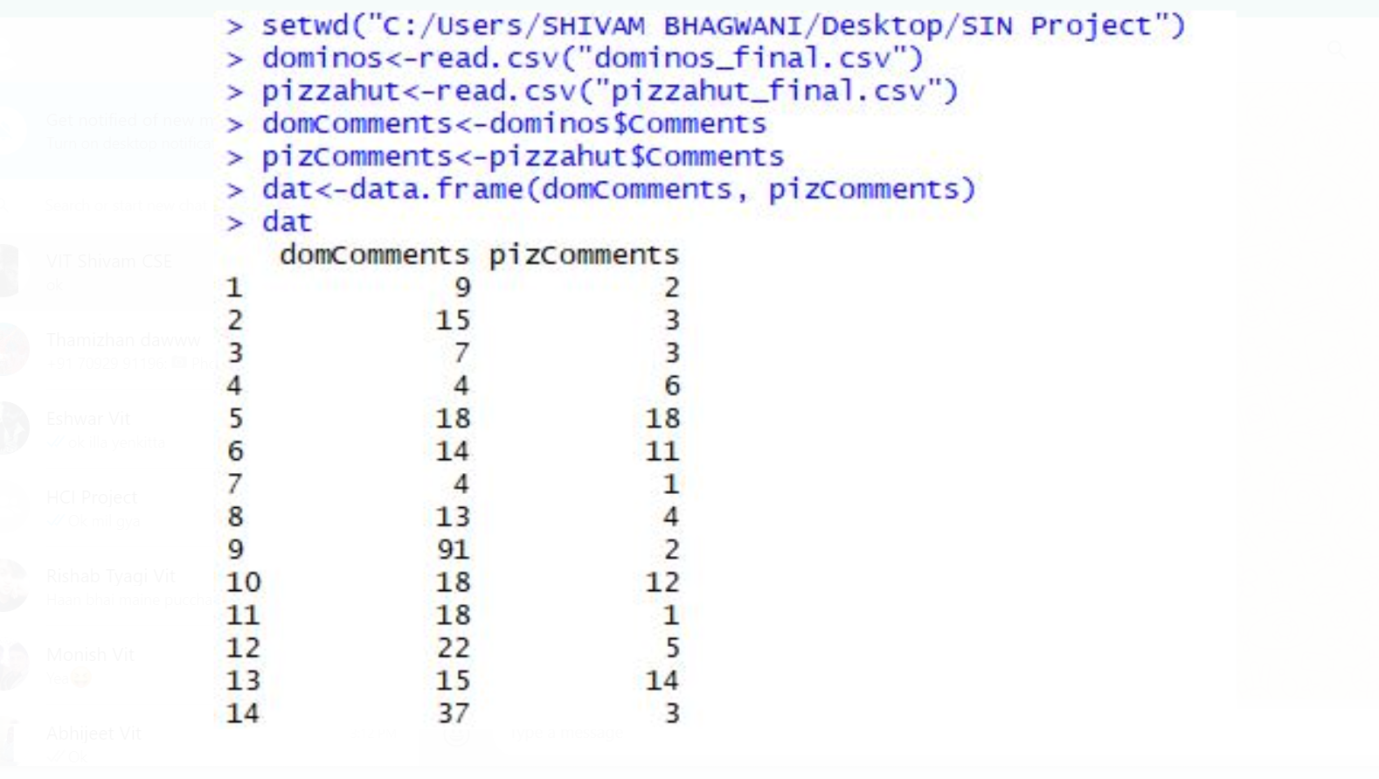


Figure :: Data for number of comments is stored in variables

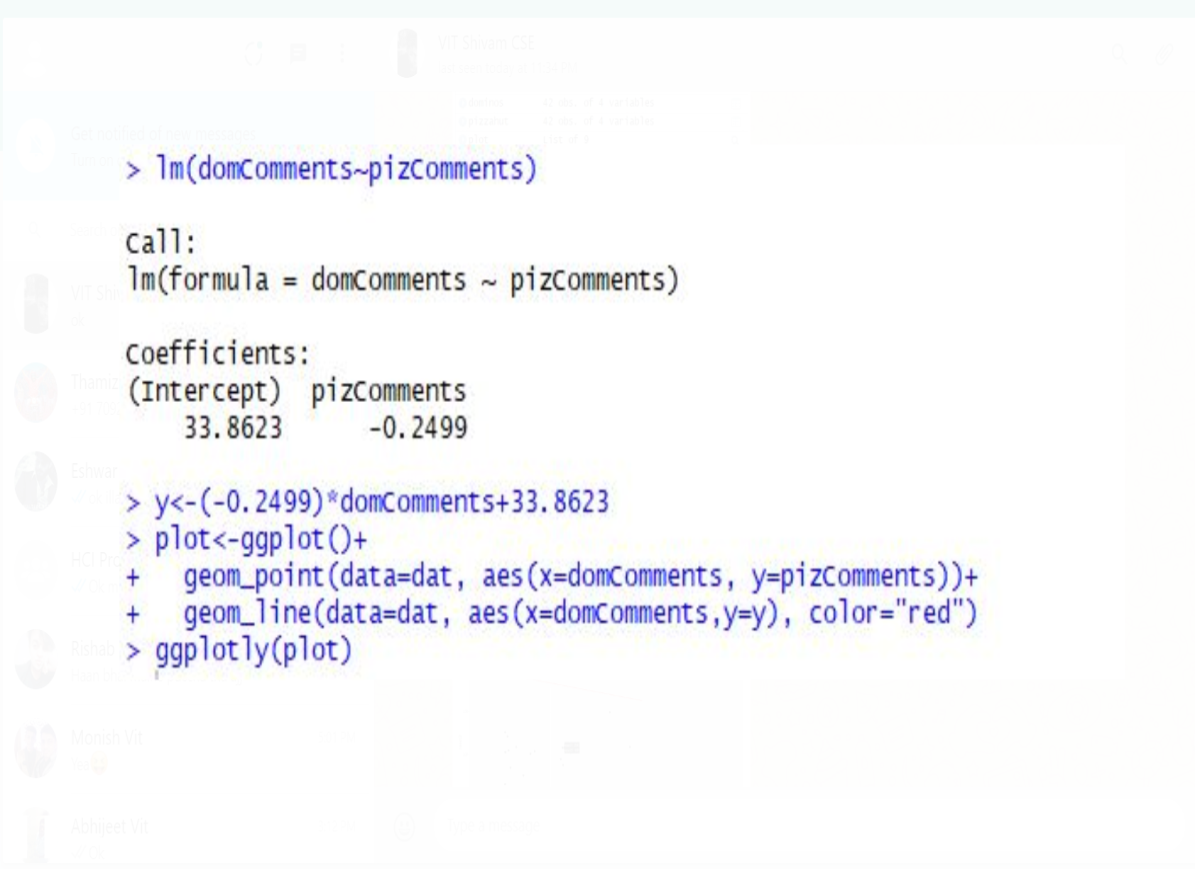
****

Figure : Finding Linear Relation between two variables

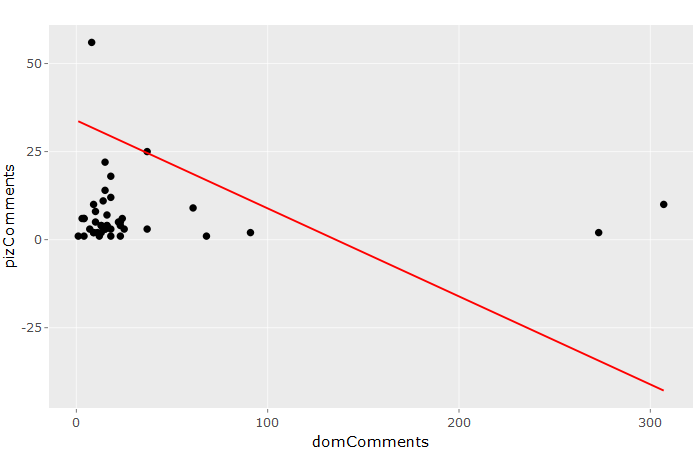


Figure : Regression Line Obtained from Comments

1. **Conclusion**

In the project, 3 comparisons were made using 3 different plots for each of them. These comparisons were as follows:

1. Reactions on Dominos posts VS Reactions on PizzaHut posts. (Figure 5)
2. Number of Shares for Dominos VS Number of Shares for PizzaHut. (Figure 8)
3. Number of Comments on Dominos Posts VS Number of Comments on PzzaHut Posts. (Figure 12)

It can be seen in the graphs that these have a NEGATIVE SLOPE.

As the most significant data would be the number of reactions in this case, the whole procedure can have been done with Reactions VS Reactions graphical method if more than 2 companies are selected for similar type of comparison.

In the course of collecting data, it was also observed that posts that included promotional offers bagged more attention than the posts that did not have any such thing.

From this analysis, finally it can be said that Dominos has better presence and reach in the online social media platform of Facebook than PizzaHut. This definitely creates a positive image in the minds of potential customers and thus PizzaHut should also work towards it.

**References**

Ediger, David, et al. "Massive social network analysis: Mining twitter for social good." *Parallel Processing (ICPP), 2010 39th International Conference on*. IEEE, 2010.

Manikonda, Lydia, Yuheng Hu, and SubbaraoKambhampati. "Analyzing user activities, demographics, social network structure and user-generated content on Instagram." *arXiv preprint arXiv:1410.8099* (2014).

Backstrom, Lars, and Jon Kleinberg. "Romantic partnerships and the dispersion of social ties: a network analysis of relationship status on facebook." *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 2014.

Suh, Bongwon, et al. "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network." *Social computing (socialcom), 2010 ieee second international conference on*. IEEE, 2010.

Lewis, Kevin, et al. "Tastes, ties, and time: A new social network dataset using Facebook. com." *Social networks* 30.4 (2008): 330-342.