PayPal Mobile Payment Landscape

Text Mining and Cluster Analysis using R

MSBA 324 – Final Project

TEAM MEMBERS: KANUPRIYA, MIRIAM O'CALLAGHAN, REENA SEHITYA, SWETA KUMARI

Submitted to: Professor Stephan Sorger

Agenda

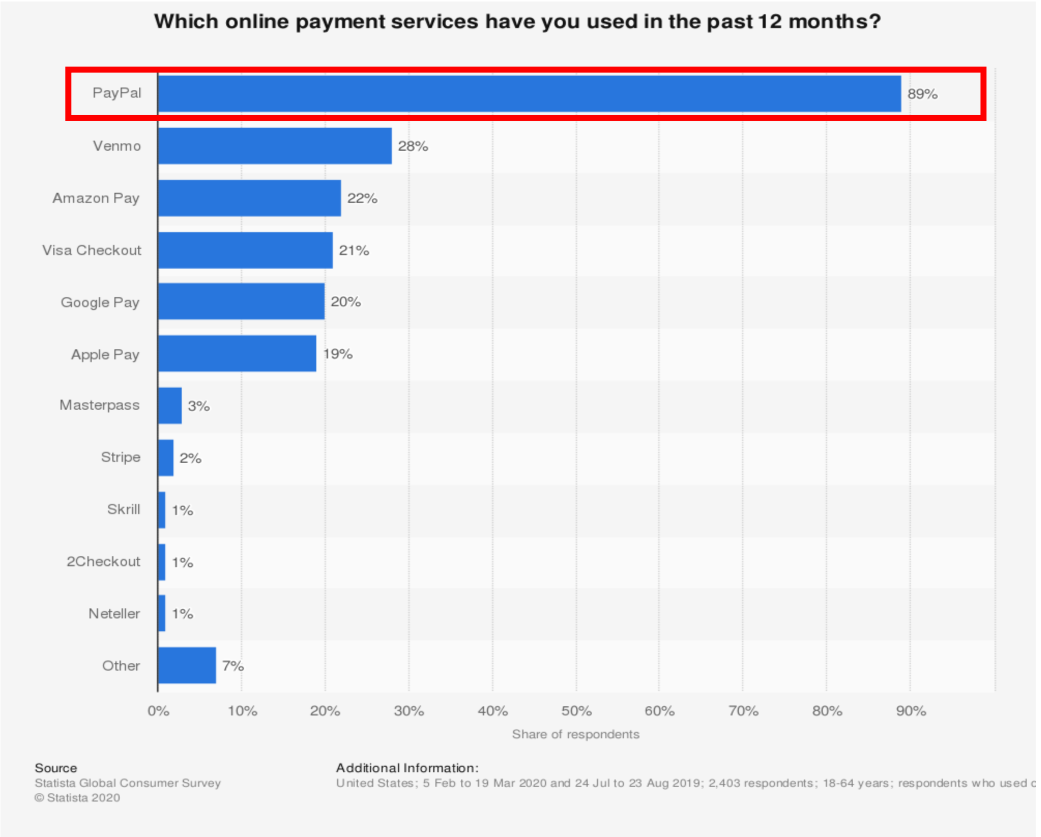
1. Situation (Reena, Kanupriya)
2. Problem Statement (Miriam)
3. Model Selection (all team members)
4. Solution Process (Miriam with all team members)
5. Research (Miriam with all team members)
6. Software (Reena and Sweta)
7. Model Results (Reena, Sweta, Miriam)
8. Visualization (Reena, Sweta, Kanupriya)
9. Results Interpretation (Miriam)
10. Situation Comparison (Sweta and Miriam)
11. Conclusion (Miriam with all team members)
12. Recommendations (Miriam with all team members)

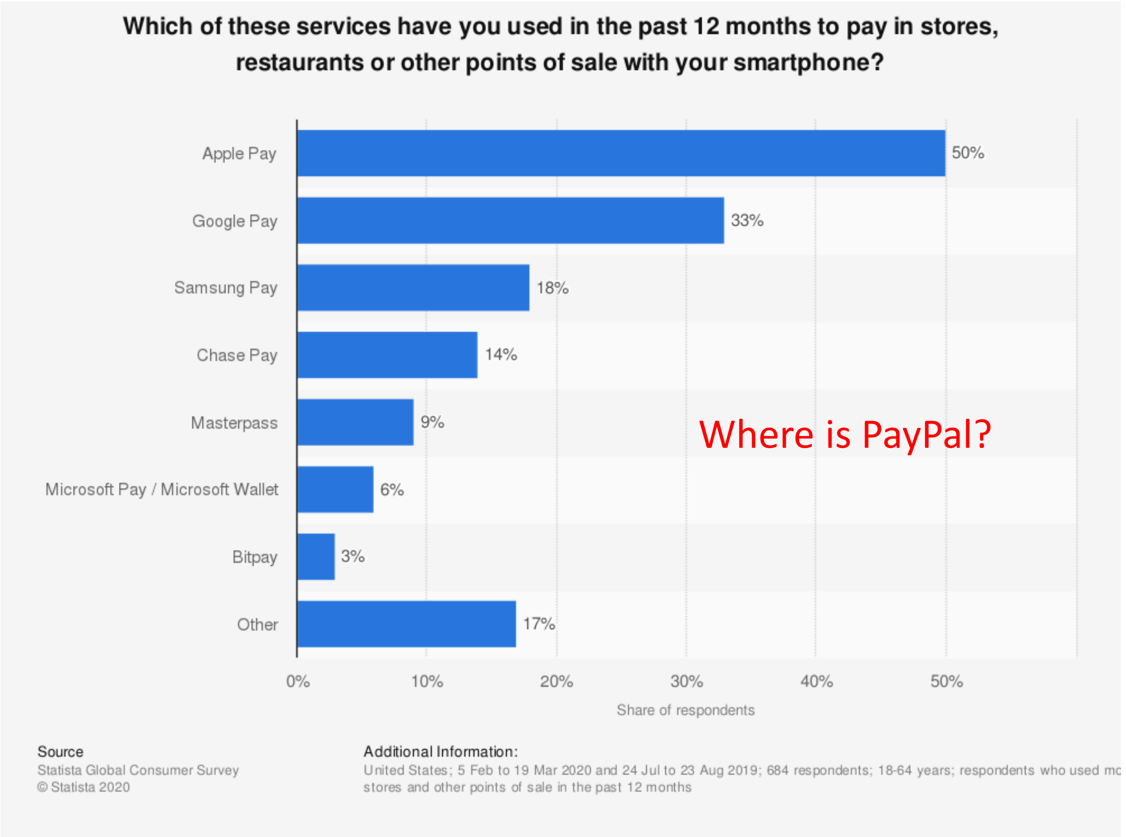
Note: the references for each slide are available in APA format both in the notes section of a respective slide and in the bibliography section in the end.

Situation

Brand: PayPal

Symptoms: Despite being a leader in the online payment sector, PayPal is nowhere to be seen in the mobile payment landscape.





Problem Statement

Problem to Solve: Strengthening mobile payment business by understanding people's **sentiments** about PayPal and knowing customers' **priorities** to create more relevant mobile payment service offerings and effective marketing strategies.

Approach: Social media sentiment analysis using text mining, market segmentation through cluster analysis.

Success Criteria:
<u>Long term:</u>
Increase PayPal's revenue by $2.7 billion (~15% increase) in 1 year.
<u>Short term:</u>
1. Identify social media sentiments for PayPal compared to the sentiments expressed for its competitor Google Pay.
2. Supported by market research, identify top three, high priority customer segments for PayPal's mobile payment service.

Model Selection

Models and Rationale – Why we chose these models?

Model 1. Sentiment Analysis using Text Mining – this model will be used to determine the **words or phrases** that represent **people's sentiments and perceptions** about PayPal and its competitor. Using text mining, we will be able to develop a **word cloud** that will represent the **most common positive and negative sentiments** people exhibit about the two companies. The model is primarily used for **brand comparison** purpose.

Model 2. Cluster Analysis (hierarchical agglomerative clustering approach) : this model will be used to determine **significant groups of customers** that PayPal should focus on while designing and marketing its mobile payment services. The model is primarily used for **customer segmentation** purpose. Why hierarchical agglomerative clustering? Please refer to the notes section.

Models are not yet adopted, no other models suitable: we found **no evidence by our online research that PayPal is already using these models**. Also, only these two models are suitable to find effective solution of our problem - understand people's sentiments about PayPal and creating customer segments by identifying their priorities to for mobile payments. We cannot apply other models that are used to solve other specific problems (example: regression that is used to make predictions).

Solution Process

To make important recommendations for PayPal as a team of consultants, we will go through this ten-step process.

1. Conduct background research
2. Collect sentiment data from Twitter for PayPal and its competitor - Google Pay
3. Perform sentiment analysis for PayPal and Google Pay
4. Collect market data with a survey
5. Execute cluster analysis
6. Identify segments and interpret the potential clusters
7. Compare clusters to market research
8. Situation analysis PayPal and Google Pay
9. Concluding the research
10. Make recommendations

Research

How the data was gathered; the sources that we used and why we believe that they are valid.

Background Research Data : most of the data for background research is collected from **statista.com** which is a well-known market and consumer data company. **Harvard Business School** library recommends Statista for research. Many **renowned research organizations and universities** are using Statista and that's why we believe it is a valid and credible source of data.

Sentiment Analysis and Text Mining Data: we scraped data directly from **Twitter** for text mining and sentiment analysis. To do that, we developed a developer account on Twitter and scraped the data we wanted. It is a first-hand, primary data directly coming from Twitter and therefore we believe it is highly credible.

Cluster Analysis Data: we collected this data through a **market survey**. We created our own questionnaire and collected responses using the Google survey tool. We also tested our questionnaire instrument on the sample of 15 respondents to evaluate its validity. It was able to accurately collect the data that we wanted for this study. The details of the questionnaire are discussed in the upcoming sections. It is a first-hand, primary data coming directly from the respondents and therefore we believe it is highly valid and credible data that can be used for this analysis.

Situation Comparison Data: part of this data is **scraped directly from Twitter**. It is a first-hand, primary data directly coming from Twitter and therefore we believe it is highly credible. We also used information available at **SimilarTech**, and **Statista** websites for comparing PayPal vs. Google Pay. SimilarTech, is a SaaS technology company and its services are used by Google, Facebook, and Amazon. We also used data from **Techcrunch**, an American online newspaper that publishes news on high tech topics. In both academic and industry research, these sources are considered as valid and credible.

Software

- R programming language (R Console in PC and Mac) is used to analyze and visualize PayPal and Google Pay data.

- Libraries used for text mining and sentiment analysis:
base64enc, twitteR, ROAuth, RCurl, tm, SnowballC, syuzhet, glue, ggplot2, ggpubr, tidyverse, tidytext, rtweet, ggmap, stringr, wordcloud, lubridate, data.table, reshape2

Software

Sentiment Analysis using Text Mining : Collecting Twitter data for PayPal

- To collect the tweets, the first step is to set up the authentication (API keys, token, etc.).

```
> #Authentication setup and token
> consumer_key <- '[redacted]'
> consumer_secret <- '[redacted]
> access_token <- '[redacted]'
> access_secret <- '[redacted]'
```

- The next step is to extract the data from Twitter and import it in data frame format into R.

```
> searchTerm <- "PayPal OR PayPalMobilePayment OR PayPalOnlineMobilePayment" #Set words to get the twitter data
> trendingTweetsPayPAL = searchTwitter(searchTerm,n=500, lang="en", since="2020-01-01", until="2020-11-24" )
> PayPal <- twListToDF(trendingTweetsPayPAL)
> attach(PayPal)
> head(PayPal$text,3)
[1] "RT @blockfolio: CEO of PayPal, \"We're going to allow cryptocurrencies
[2] "RT @blockfolio: CEO of PayPal, \"We're going to allow cryptocurrencies
[3] "RT @TheMangaScholar: The stream was so much fun! Will be doing it again
```

Software

Sentiment Analysis using Text Mining : Data Cleaning

- Number of Rows and Columns in data frame

```
> dim(PayPal)
[1] 500  16
```

- The next step is to clean the text data. The code below is used to remove http elements and punctuations, convert text to lowercase, add id for each tweet, and remove stop words.

```
> PayPal$stripped_text<- gsub("http.*","", PayPal$text) # remove http elements manually
> PayPal$stripped_text<- gsub("https.*","", PayPal$stripped_text) # remove https elements manually
> PayPal_clean <- PayPal %>% dplyr::select(stripped_text) %>% unnest_tokens(word,stripped_text)
> PayPal_keywords <- PayPal_clean %>% anti_join(stop_words) # remove stop words from your list of words
Joining, by = "word"
```

Software

Sentiment Analysis using Text Mining : Sentiments for PayPal

- To understand and visualize the sentiments for PayPal mobile payment, below code is used. It displays the sentiments in five categories: “Very Positive”, “Positive”, “Neutral”, “Very Negative”, “Negative”.

```
> encodeSentiment <- function(x) {
+   if(x <= -0.5){
+     "Very Negative"
+   }else if(x > -0.5 & x < 0){
+     "Negative"
+   }else if(x > 0 & x < 0.5){
+     "Positive"
+   }else if(x >= 0.5){
+     "Very Positive"
+   }else {
+     "Neutral"
+   }
+ }
> tweetSentiments <- get_sentiment (PayPal$text, method = "syuzhet") #syuzhet method for getting sentiment
> tweets <- cbind(PayPal, tweetSentiments) #combining columns
> tweets$sentiment <- sapply(tweets$tweetSentiments,encodeSentiment) #apply encodesentiment function
> ggplot(tweets, aes(sentiment)) + #plot the graph using ggplot
+ geom_bar(fill = "aquamarine4") + #use aquamarine4(any color) to fill the bar
+ theme(legend.position="none",  #No legend position is mentioned
+ axis.title.x = element_blank()) + #hide the x axis title
+ ylab("Number of tweets") + #put the label on y axis
+ ggtitle("Tweets by Sentiment") #title of the graph
```

Software

Sentiment Analysis using Text Mining : Frequently used words and Word Cloud for PayPal

- To understand how frequently the positive and negative terms are used for PayPal, we used this code and created a bar chart for positive and negative words.

```
> PayPal_bing <- PayPal_clean %>% inner_join ( get_sentiments ( "bing" )) %>% count ( word,
+ sentiment, sort = TRUE ) %>% ungroup ( )
Joining, by = "word"
> PayPal_bing %>% group_by ( sentiment ) %>% top_n ( 10 ) %>% ungroup ( ) %>%
+ mutate ( word = reorder ( word, n )) %>% ggplot ( aes ( word, n, fill = sentiment )) +
+ geom_col ( show.legend = FALSE ) + facet_wrap ( ~sentiment, scales = "free_y" ) +
+ labs ( title = "Keywords associated with PayPal's negative & positive sentiment", y = "Contribution to Sentiment",
+ x = NULL ) + coord_flip()
Selecting by n
```

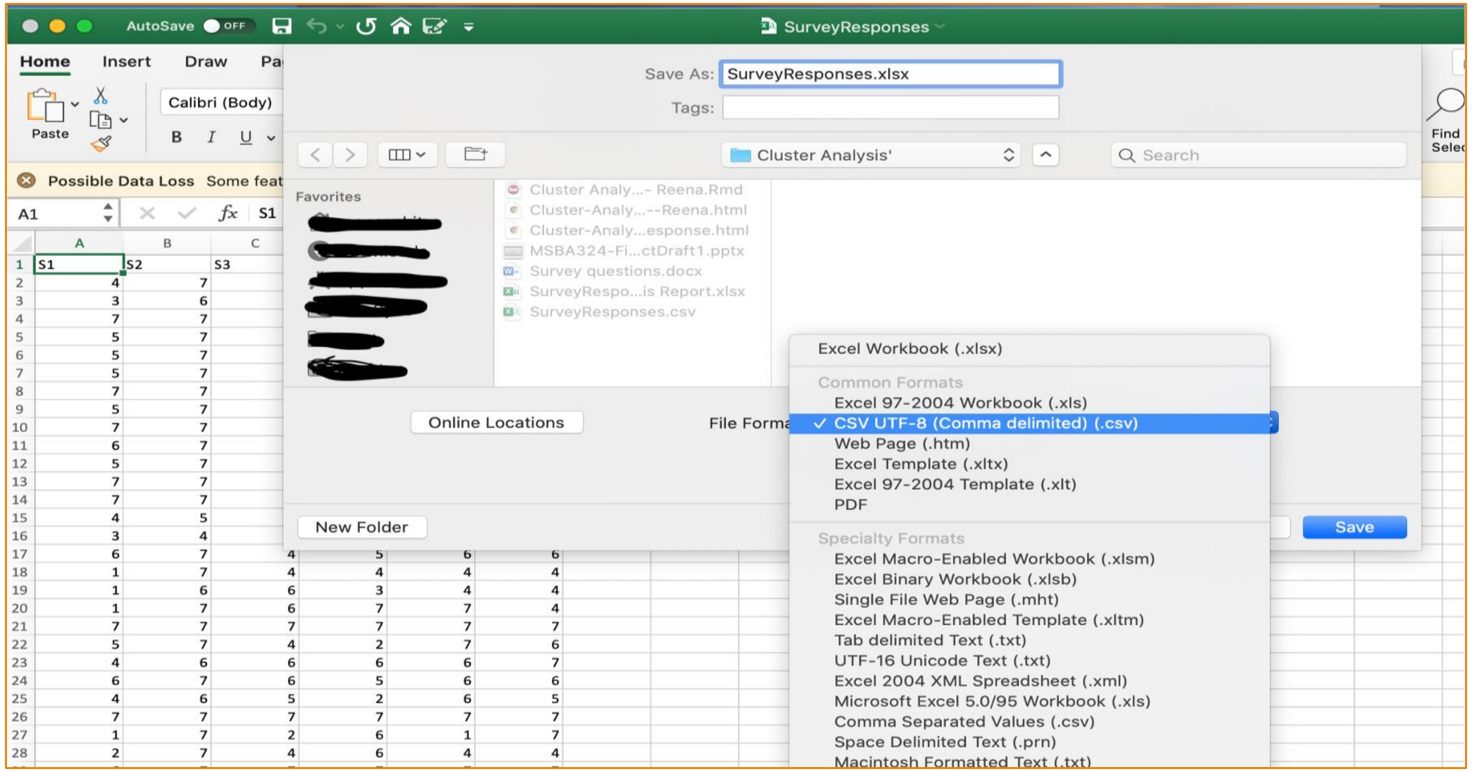
- We then created a word cloud dividing into positive and negative sentiment.

```
> PayPal_clean %>% inner_join(get_sentiments ( "bing" )) %>%
+ count ( word, sentiment, sort = TRUE ) %>%
+ acast ( word ~ sentiment, value.var = "n", fill = 0 ) %>%
+ comparison.cloud ( color = c ( "red", "blue" ), max.words = 100)
Joining, by = "word"
```

Software

Cluster analysis for Market Segmentation : Data Collected through Market Survey

- Our first step was to convert survey data from Microsoft Excel to Comma Separated Values (csv) format.



Software

Cluster analysis for Market Segmentation : Data Preparation (used R Console with Mac OS, snapshots look different)

- Next, we read the data into the data frame called **MobilePaymentSurvey** then computed matrix of distances using Euclidean method and invoked Ward.D2 method of clustering

```
> # Read the data in the data frame called MobilePaymentSurvey
> MobilePaymentSurvey = read.csv("SurveyResponses.csv", header = TRUE)
> #head(MobilePaymentSurvey) # To Check the first five rows
> # Compute matrix of distances using euclidean method
> distaceMatrix = dist(MobilePaymentSurvey,method ='euclidean')
> # Invoke Ward.D2 method of clustering
> tree = hclust(distaceMatrix, method = "ward.D2")
```

Software

Cluster analysis for Market Segmentation : Agglomerative Hierarchical Cluster Dendrogram

- In order to view potential segments, we plotted cluster dendrogram.

```
>
> plot(tree)
>
```

- In order to identify 4 clusters, we cut the tree vertically and added red boxed using hclust function.

```
>
> # Add colored boxes around clusters (groups) you identify
> plot(hclust(dist(distaceMatrix))) # To it adds a rectangle to dendrogram,need toplot the
dendrogram _first
> x <- rect.hclust(tree, k=4, border="red",)
>
```

Software

Cluster analysis for Market Segmentation : Members in Each Cluster

- We ran the code to see which respondent showed preferences in which cluster- 1, 2, 3, 4

```
>
> # Cut tree into 4 clusters
> clusternumber <- cutree (tree, k = 4 ) # cut tree into k = 4
> clusternumber
 [1] 1 1 2 2 3 2 2 2 2 2 2 2 2 1 3 2 3 3 1 2 3 1 2 3 2 3 3 2 1 3 1 2 1 1 2 2 2 2 2 3 1 2 3 2 1
[46] 1 2 2 3 3 2 1 2 2 2 3 1 3 4 1 2 2 3 3 2 4 1 2 1 2 2
>
```

Software

Cluster analysis for Market Segmentation : Cluster Members Analysis

- Next, we created the subset of MobilePaymentSurvey data using “subset” function for each cluster.

  - Members for cluster 1

```
> #Create subset of MobilePaymentSurvey data using “subset” function
> c1 = subset ( MobilePaymentSurvey, clusternumber == 1)
> c1
   S1 S2 S3 S4 S5 S6
1   4  7  7  5  5  6
2   3  6  6  5  6  6
14  4  5  6  6  6  6
19  1  7  6  7  7  4
22  4  6  6  6  6  7
29  5  5  7  7  7  7
31  3  7  6  6  6  7
33  4  7  7  4  4  7
34  3  6  6  7  7  6
41  3  7  7  7  4  7
45  4  7  5  7  7  6
46  4  7  7  4  7  7
52  4  7  5  7  7  7
57  4  7  6  6  5  7
60  1  7  7  7  7  7
67  4  7  5  7  7  5
69  4  6  7  7  7  6
>
```

Software

Cluster analysis for Market Segmentation : Cluster Members Analysis

- Members for cluster 2: Consisting of major portion of total respondents and require our attention

```
> #For c2
> c2 = subset ( MobilePaymentSurvey, clusternumber == 2)
> c2
   S1 S2 S3 S4 S5 S6
3   7  7  7  7  7  7
4   5  7  6  6  6  5
6   5  7  7  7  4  6
7   7  7  7  7  6  7
8   5  7  5  6  3  7
9   7  7  7  7  7  7
10  6  7  7  4  6  6
11  5  7  7  7  7  7
12  7  7  7  7  7  7
13  7  7  7  7  7  6
16  6  7  4  5  6  6
20  7  7  7  7  7  7
23  6  7  6  5  6  6
25  7  7  7  7  7  7
28  6  7  7  7  7  7
32  7  7  7  4  4  7
35  7  7  5  4  6  6
36  5  7  6  5  1  3
37  5  7  4  6  6  6
38  6  7  7  7  4  6
39  5  7  6  6  6  6
42  5  7  6  4  5  6
44  6  7  7  7  7  7
47  7  7  6  5  7  7
48  6  7  6  7  7  7
51  7  7  6  5  5  7
53  6  7  4  6  7  7
54  5  7  6  6  6  4
55  6  6  6  6  5  5
61  5  7  6  7  6  7
62  7  7  7  7  4  3
65  6  7  5  7  7  7
68  7  7  7  7  7  7
70  7  7  7  7  7  7
71  7  7  7  7  4  7
>
```

Software

Cluster analysis for Market Segmentation : Cluster Members Analysis

- Members for cluster 3 – there are fewer respondents than cluster 2 and cluster 1 falling into this cluster

```
> #For c3
> c3 = subset ( MobilePaymentSurvey, clusternumber == 3)
> c3
   S1 S2 S3 S4 S5 S6
5   5  7  6  2  7  2
15  3  4  7  6  7  5
17  1  7  4  4  4  4
18  1  6  6  3  4  4
21  5  7  4  2  7  6
24  4  6  5  2  6  5
26  1  7  2  6  1  7
27  2  7  4  6  4  4
30  2  6  6  4  7  4
40  2  5  7  1  5  6
43  1  7  2  4  4  4
49  4  7  1  3  5  4
50  3  7  3  2  7  7
56  1  7  7  4  5  5
58  3  3  5  5  7  5
63  4  7  1  1  4  1
64  4  4  4  5  5  5
>
```

Software

Cluster analysis for Market Segmentation : Cluster Members Analysis

- Members for cluster 4 – there are only 2 respondents falling into this cluster and probably doesn't require much of our attention while making marketing decisions.

```
>
> #For c4
> c4 = subset ( MobilePaymentSurvey, clusternumber == 4)
> c4
   S1 S2 S3 S4 S5 S6
59  2  1  3  2  3  3
66  1  1  1  1  1  1
>
```

Software

Cluster analysis for Market Segmentation : Cluster Averages

- In order to gain insights, we ran the code colMeans() function to get the summary statistics of each column for all four clusters - c1, c2, c3, c4

```
>
> # To compute mean(average) of each column for each cluster
> colMeans(c1)
      S1       S2       S3       S4       S5       S6 
3.470588 6.529412 6.235294 6.176471 6.176471 6.352941 
> colMeans(c2)
      S1       S2       S3       S4       S5       S6 
6.142857 6.971429 6.257143 6.171429 5.828571 6.285714 
> colMeans(c3)
      S1       S2       S3       S4       S5       S6 
2.705882 6.117647 4.352941 3.529412 5.235294 4.588235 
> colMeans(c4)
 S1  S2  S3  S4  S5  S6 
1.5 1.0 2.0 1.5 2.0 2.0 
>
>
```

Model Results

Results – Sentiment Analysis and Word Cloud

- The top 5 terms used to express negative sentiments are "bust", “emergency", “hate", “vomiting", “sad“.
- The top 5 terms used to express positive sentiments are “like", “reputation", “appreciated", “enough", “thank“.
- A few strong negative words such as “scam”, “fraud”, “cheap”, “debt” are used in the tweets to exhibit anger or complaints against PayPal.
- The highest number of sentiments are expressed in “very positive” category whereas the smallest number is expressed in the “negative” category.

Model Results

Results - Cluster Analysis

Table 1 summarizes the results of cluster analysis in form of mean values for each response under the four clusters. Table 2 consists of the details of all question statements that respondents responded to.

In table 1, significantly high and low rated responses are turned bold.

Table 1

Cluster	S1	S2	S3	S4	S5	S6
1	3.47	**6.53**	**6.24**	**6.18**	**6.18**	**6.35**
2	**6.14**	**6.97**	**6.26**	**6.17**	**5.83**	**6.29**
3	2.71	**6.12**	4.35	3.53	**5.24**	4.59
4	**1.50**	**1.00**	2.00	**1.50**	2.00	2.00

Table 2

Survey statements	
S1	I prefer to use a mobile device for all my in-store purchases. (eg: Apple pay supported by iPhone)
S2	For mobile payments, I value safety of transactions the most.
S3	For mobile payments, convenience (eg: one-touch pay, QR code) is my priority.
S4	For mobile payments, saving money (low or no transaction fee) is my priority.
S5	For mobile payments, I prefer to receive the best deals and offers from merchants
S6	For mobile payments, excellent customer service (eg: call center support) is my priority.

More details of these results are available in the notes section.

A 7-point Likert Scale was used (7 “Strongly Agree” to 1 “Strongly Disagree”).

Visualization

Sentiment Analysis - PayPal

- Tweets by Sentiment (Very Negative, Negative, Neutral, Positive, Very Positive)

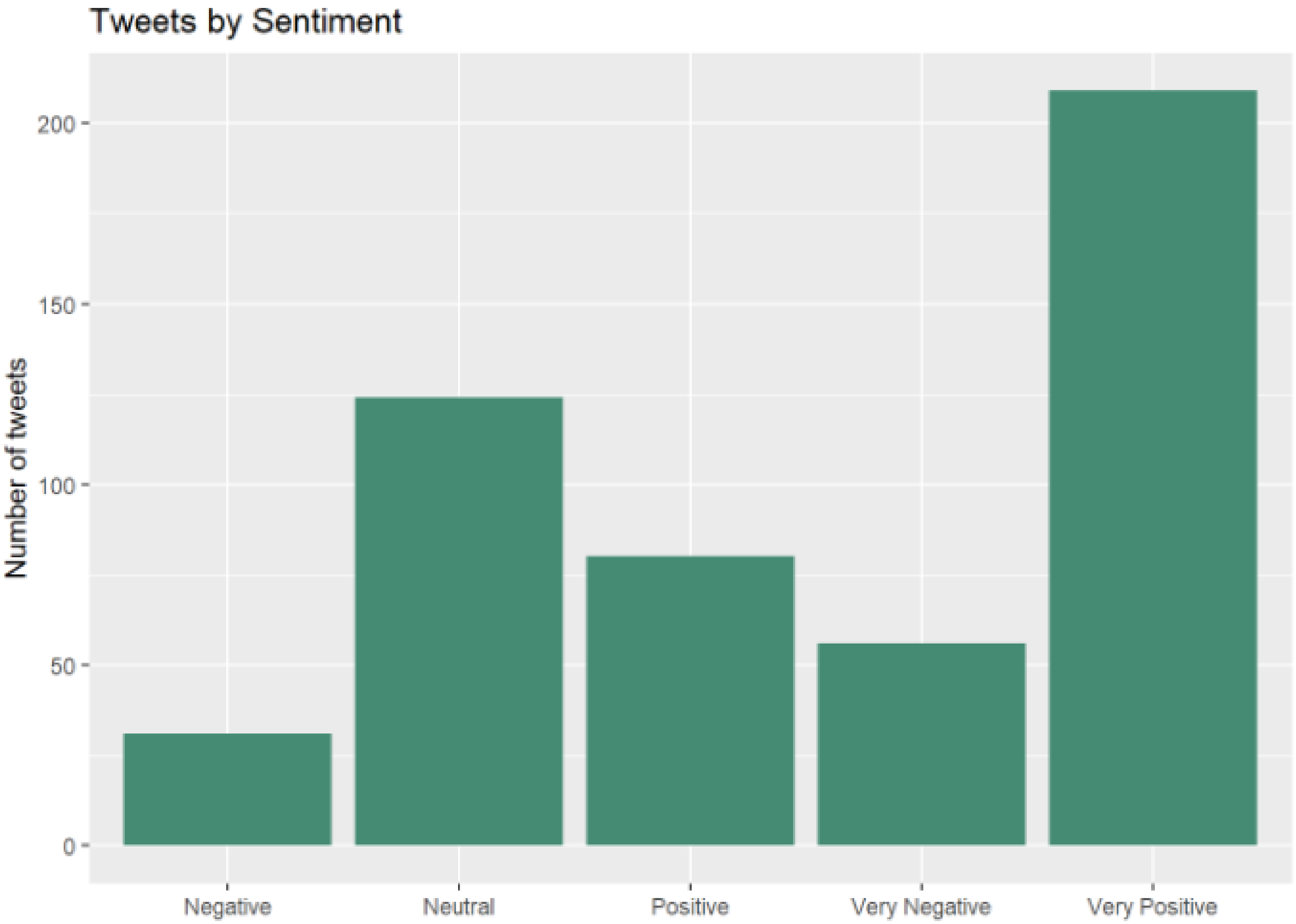


Figure 1

Visualization

Sentiment Analysis-PayPal

- ❑ Most frequently negative words used for PayPal
- ❑ Most frequently Positive words used for PayPal

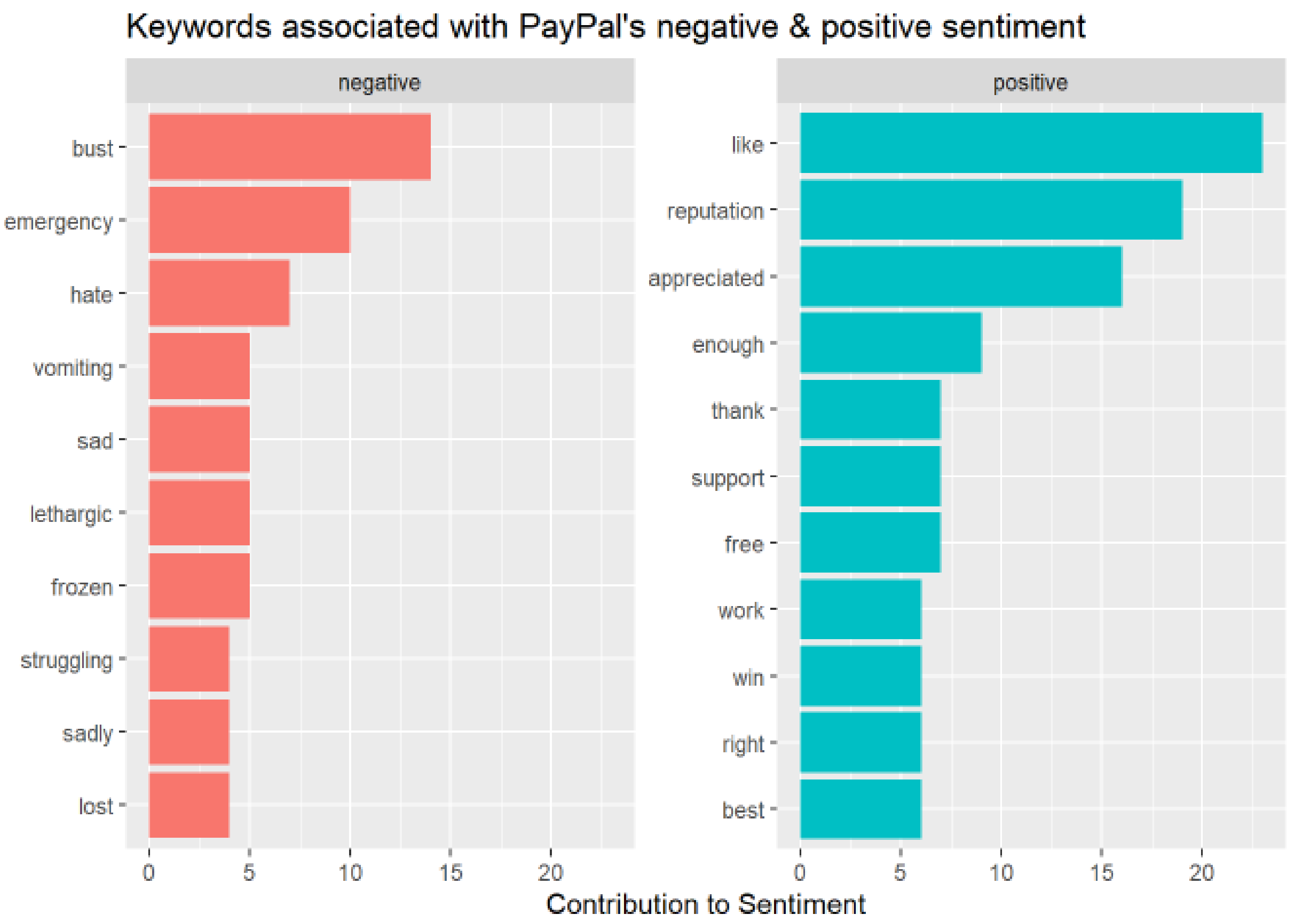


Figure 2

Visualization

Sentiment Analysis-PayPal

- ❑ Word Cloud with negative and positive words for PayPal

Figure 3

Visualization

Cluster Analysis: Cluster Dendrogram

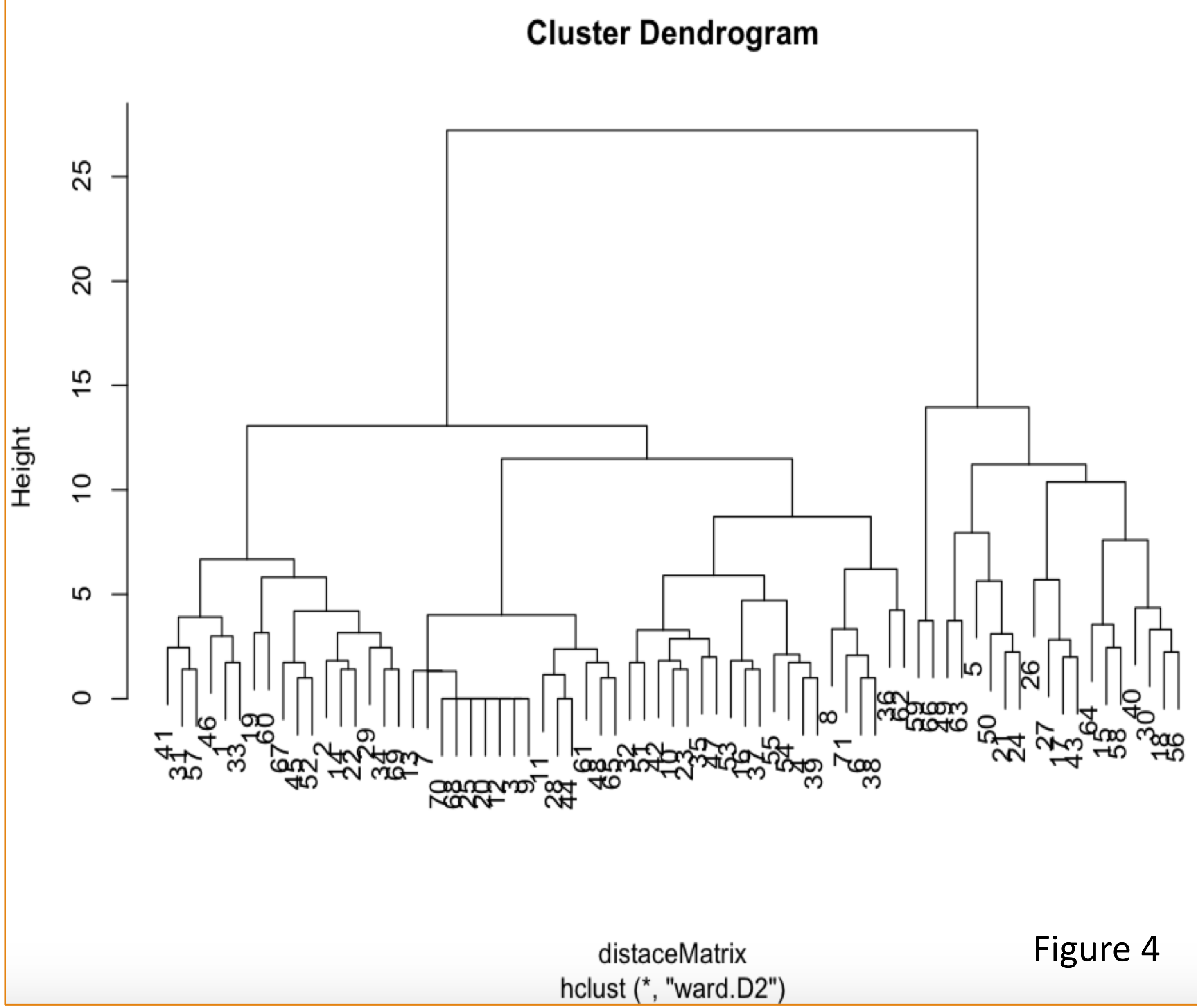


Figure 4

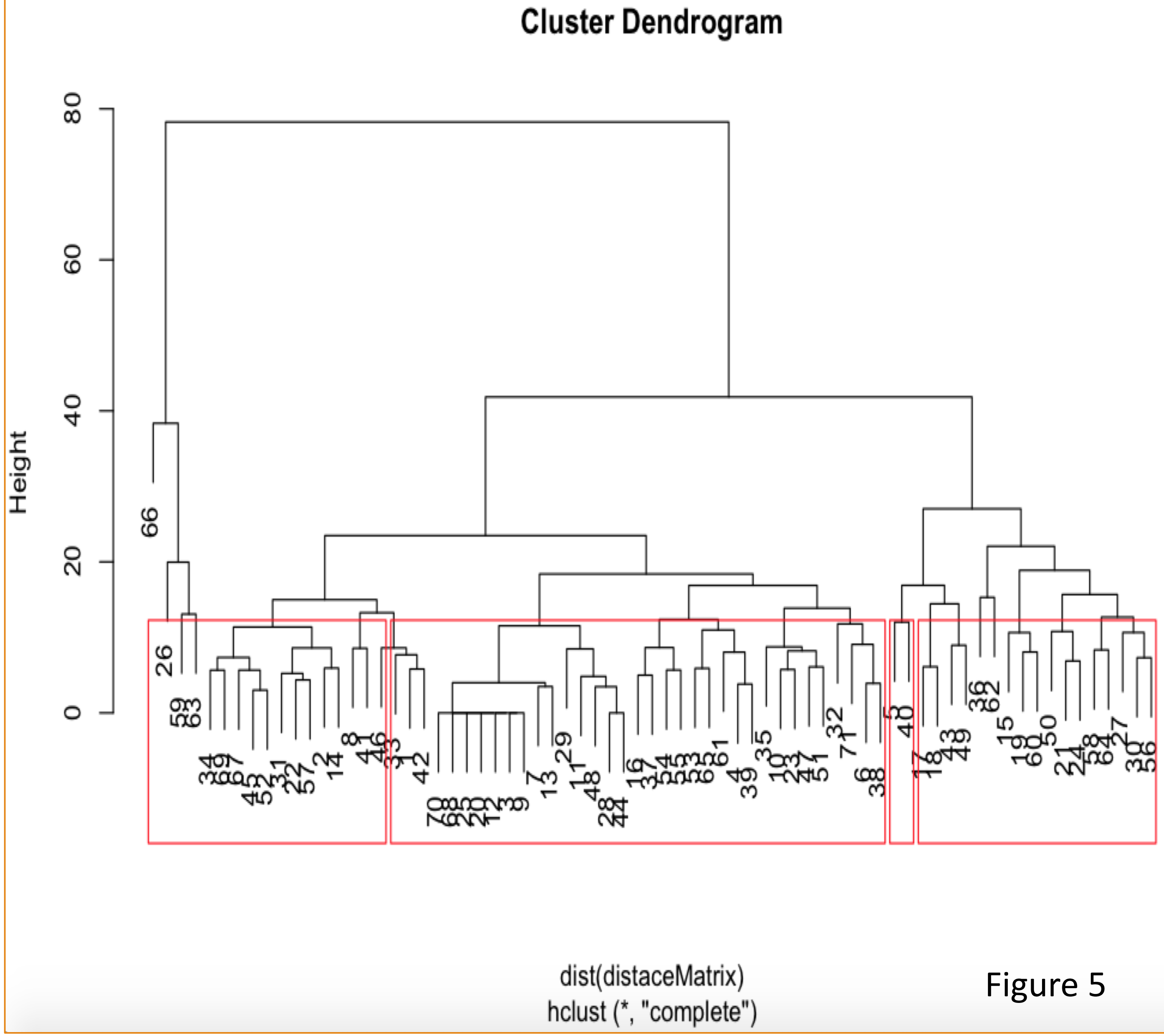


Figure 5

Visualization

Cluster Analysis: Cluster Dendrogram

- Potential clusters in dendrogram

Result Interpretation

Sentiment Analysis Result Interpretation:

- Overall, people exhibited more positive sentiments as compared to the negative and neutral sentiments. Collectively, number of “very positive” and “positive” tweets are more than other sentiments. The number of neutral sentiments is also significant (second highest). It can be interpreted as the presence of expressions that might not be associated with any sentiments or preferences for PayPal.
- We identified a significant frequency of negative and very negative emotion expressions (figures 1 and 3 - sentiment analysis bar chart and word cloud). The frequency is not insignificant, it is towards the higher side.
- Negative words for PayPal are used more frequently than positive words in the Word Cloud which is a little alarming. The top three words – like, reputation, and appreciated seems positive but cannot be considered as very positive (intense positive emotions such as expression of joy and ecstasy). On the other hand, the top negative words are extremely negative – bust, emergency, and hate. People are expressing moderately positive and extremely negative emotions for PayPal on Twitter.

Cluster Analysis Result interpretation:

- Four clusters are identified. Cluster 2 consists the major portion of total respondents and respondents in this cluster rated all statements in the questionnaire very high. They are the highest users of mobile payment services and prefer safety, convenience, saving, rewards, and excellent customer service.
- Cluster 1 is second largest cluster with 5 high ratings. They prefer safety, convenience, saving, rewards, and excellent customer service.
- Cluster 3 rated only two statements high. They prefer safety and rewards and don’t care about other features.
- Cluster 4 is the smallest cluster with only two respondents with very low ratings on all the statements in the questionnaire. This cluster will be excluded from our final cluster set which will include only 3 clusters.

Situation Comparison – Google Pay

Sentiment Analysis using Text Mining : Data Collection for Google Pay

- The first step is to extract Google Pay data from Twitter and import it in data frame format into R.

```
> searchTerm <- "GooglePay OR GooglePayMobilePayment OR GooglePayOnlineMobilePayment" #set words to get twitter data
> trendingTweetsGooglePay = searchTwitter(searchTerm,n=500, lang="en", since="2020-01-01", until="2020-11-24")
> GooglePay <- twListToDF(trendingTweetsGooglePay)
> attach(GooglePay)
> head(GooglePay$text,3)
[1] "Google - Financial Service Provider Of Tomorrow!\n#GooglePay
[2] "RT @googledevs: Hello @Angular devs!\n\nIt's even easier for
[3] "RT @TheTaiwanTimes: Google - Financial Service Provider Of Tc
```

- Number of rows and columns.

```
> dim(GooglePay)
[1] 500  16
```

Situation Comparison – Google Pay

Sentiment Analysis using Text Mining : Data Cleaning

- The next step is to clean the text data. Below code is used to remove http elements and punctuation, convert text to lowercase, add id for each tweet, remove stop words.

```
> GooglePay$stripped_text<- gsub("http.*","", GooglePay$text) # remove http elements manually
> GooglePay$stripped_text<- gsub("https.*","", GooglePay$stripped_text) # remove https elements manually
> GooglePay_clean <- GooglePay %>% dplyr::select(stripped_text) %>% unnest_tokens(word,stripped_text)
> GooglePay_keywords <- GooglePay_clean %>% anti_join(stop_words) # remove stop words from your list of words
Joining, by = "word"
```

Situation Comparison – Google Pay

Sentiment Analysis using Text Mining : Sentiments for Google Pay

- To understand and visualize the sentiments for PayPal mobile payment, the below code is used. It displays the sentiments in five categories: "Very Positive", "Positive", "Neutral", "Very Negative", "Negative".

```
> encodeSentiment <- function(x) {
+   if(x <= -0.5){
+     "Very Negative"
+   }else if(x > -0.5 & x < 0){
+     "Negative"
+   }else if(x > 0 & x < 0.5){
+     "Positive"
+   }else if(x >= 0.5){
+     "Very Positive"
+   }else {
+     "Neutral"
+   }
+ }
> tweetSentiments <- get_sentiment (GooglePay$text, method = "syuzhet") #syuzhet method for getting sentiment
> tweets <- cbind(GooglePay, tweetSentiments) #combining columns
> tweets$sentiment <- sapply(tweets$tweetSentiments,encodeSentiment) #apply encodesentiment function
> ggplot(tweets, aes(sentiment)) + #plot the graph using ggplot
+ geom_bar(fill = "aquamarine4") + #use aquamarine4(any color) to fill the bar
+ theme(legend.position="none", #No legend position is mentioned
+ axis.title.x = element_blank()) + #hide the x axis title
+ ylab("Number of tweets") + #put the label on y axis
+ ggtitle("Tweets by Sentiment") #title of the graph
```

Situation Comparison – Google Pay

Sentiment Analysis using Text Mining : most frequent positive and negative sentiments expressed for Google Pay and the Word Cloud

- To understand how frequently the positive and negative terms are used for Google Pay, we used code and created a bar chart showing positive and negative sentiments frequency.

```
> GooglePay_bing <- GooglePay_clean %>% inner_join ( get_sentiments ( "bing" )) %>% count ( word,
+ sentiment, sort = TRUE ) %>% ungroup ( )
Joining, by = "word"
> GooglePay_bing %>% group_by ( sentiment ) %>% top_n ( 10 ) %>% ungroup ( ) %>%
+ mutate ( word = reorder ( word, n )) %>% ggplot ( aes ( word, n, fill = sentiment )) +
+ geom_col ( show.legend = FALSE ) + facet_wrap ( ~sentiment, scales = "free_y" ) +
+ labs ( title = "Keywords associated with GooglePay's negative & positive sentiment", y = "Contribution to Sentiment",
+ x = NULL ) + coord_flip()
Selecting by n
```

- Code for word cloud classifying sentiments into positive and negative.

```
> GooglePay_clean %>% inner_join(get_sentiments ( "bing" )) %>%
+ count ( word, sentiment, sort = TRUE ) %>%
+ acast ( word ~ sentiment, value.var = "n", fill = 0 ) %>%
+ comparison.cloud ( color = c ( "red", "blue" ), max.words = 100)
Joining, by = "word"
```

Visualization

Sentiment Analysis – Google Pay

- Tweets by Sentiment (Very Negative, Negative, Neutral, Positive, Very Positive)

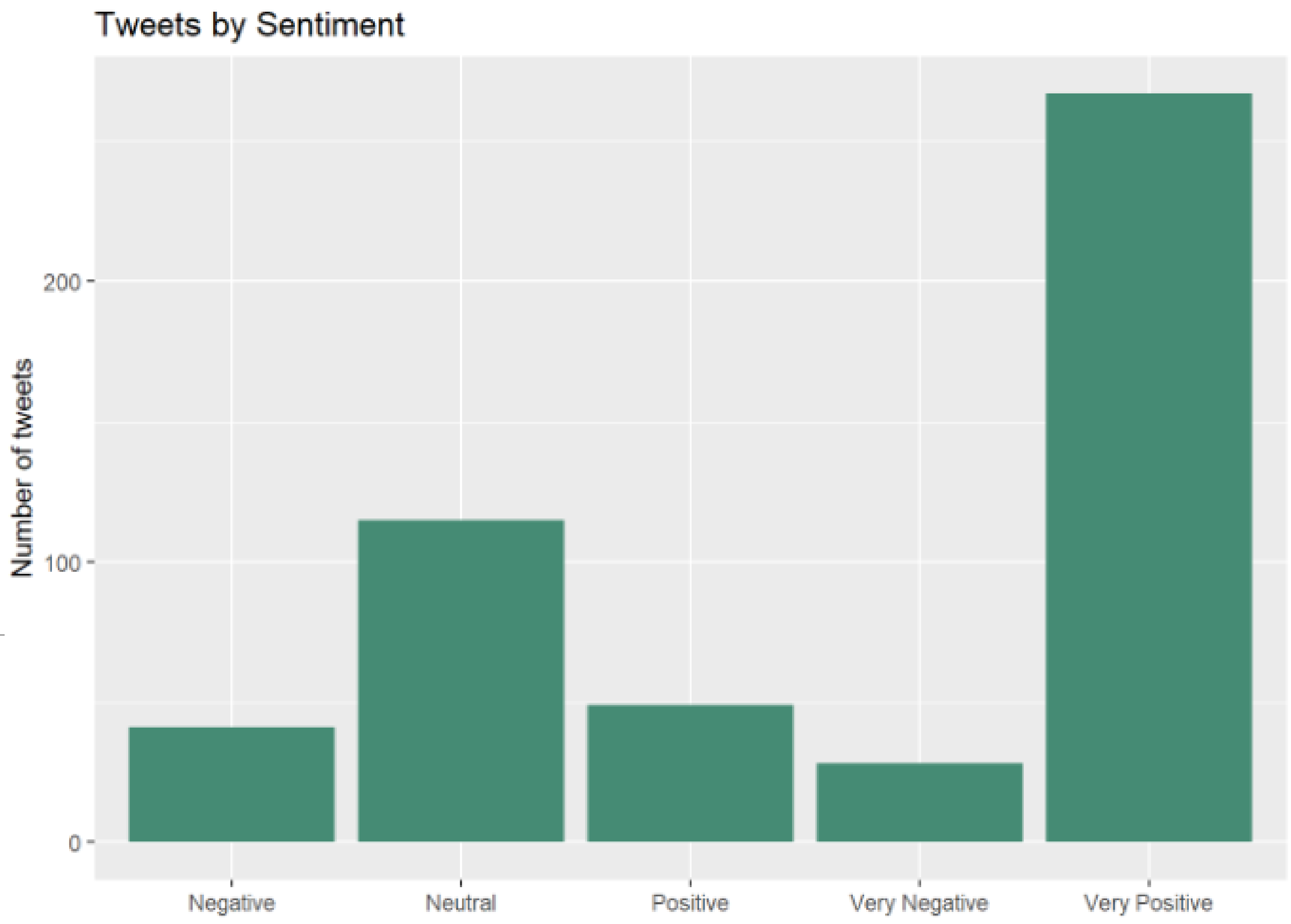


Figure 6

Visualization

Sentiment Analysis – Google Pay

- ❑ Most frequently negative words used for Google Pay
- ❑ Most frequently Positive words used for Google Pay

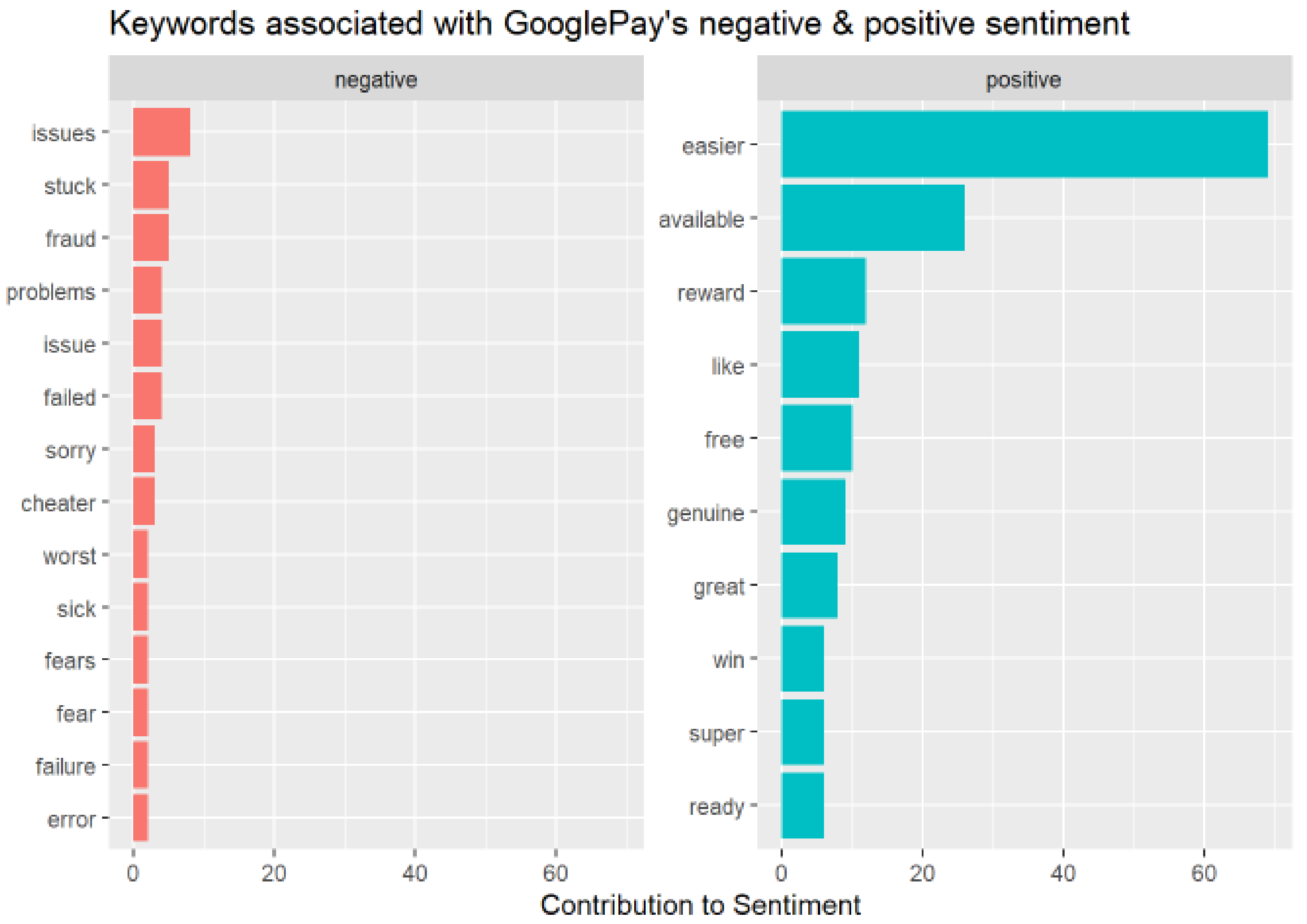


Figure 7

Visualization

Sentiment Analysis – Google Pay

- ❑ Word Cloud with negative and positive words for Google Pay

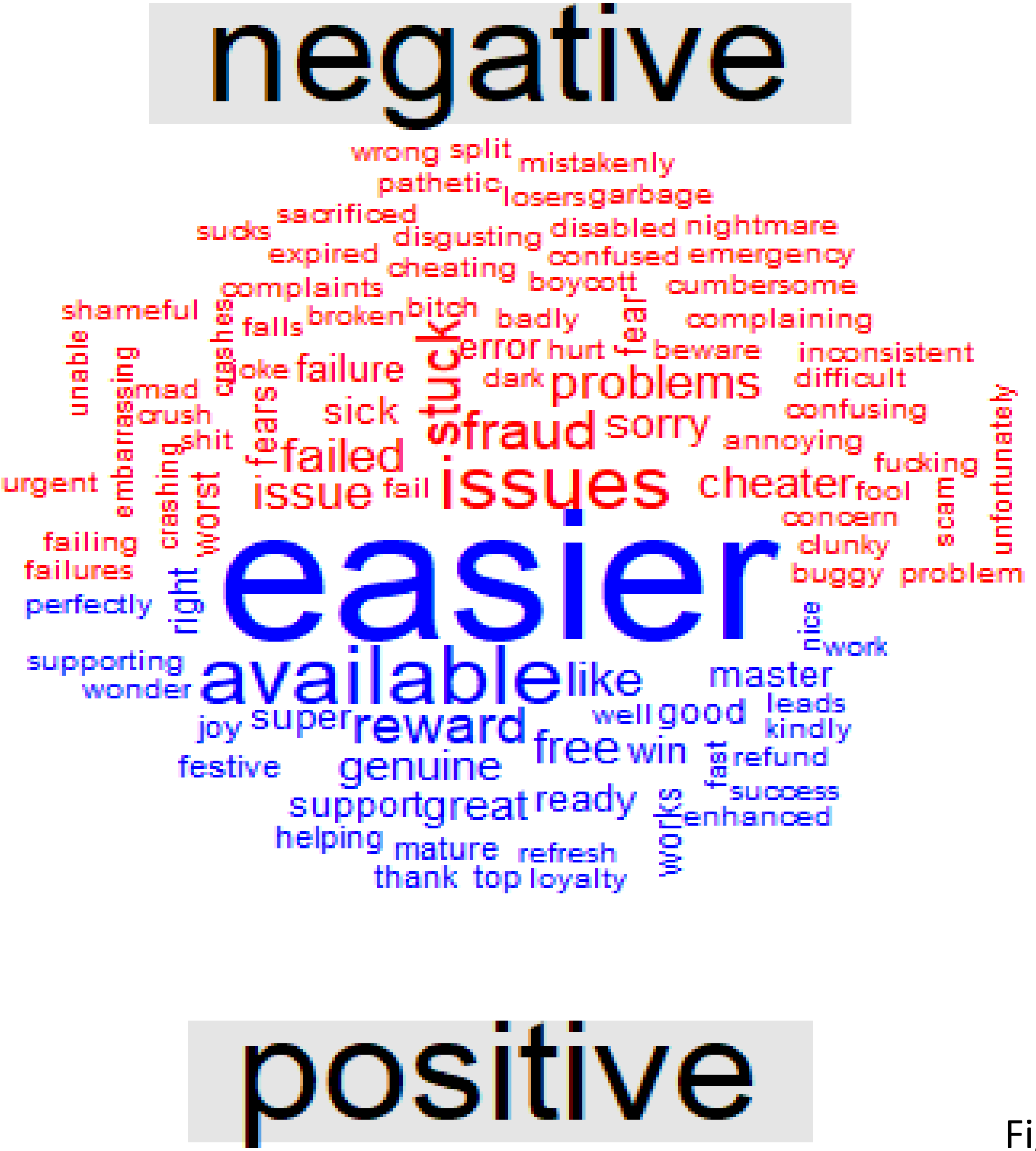


Figure 8

Google Pay - Model Results

Sentiment Analysis for Google Pay

- People exhibited more positive sentiments as compared to the negative and neutral sentiments. Collectively, number of “very positive” and “positive” tweets are more than other sentiments. The number of neutral sentiments is also significant (second highest).

- The top 5 terms used to express negative sentiments are “issue", “stuck", “fraud", “problem", “failed“.

- The top 5 terms used to express positive sentiments are “easier", “available", “record", “like", “free“.

- The negative words are more in number than positive words in the Word Cloud.

- A few more frequent negative words such as “stuck”, “inconsistent”, “fraud”, “annoying” are used in the tweets to show anger or complaints against Google Pay.

Situation Comparison – Brand and Performance Comparison

Comparison Criteria	PayPal	Google Pay
1. Primary service – value proposition	**PayPal** services enable companies and individuals to send money and to accept payments without revealing any financial details.	**Google Pay** is mostly used for making purchases at thousands of online stores and businesses.
2. Market share by top websites	**PayPal** is leading in Top 10K Sites, Top 100K Sites, Top 1M Sites and The Entire Web. Total number of websites using PayPal - 1,322,386	Despite its recent growth, **Google Pay** is still behind PayPal in all market share segments. Total number of websites using Google Pay - 63,748
3. Websites categories	**PayPal** has better usage coverage in more website categories. Including Computers Electronics & Technology, Lifestyle, Arts & Entertainment, Games and 20 other categories.	**Google Pay** hasn't got a lead over PayPal in any website category.
4. Geography	**PayPal** is leading in most countries, including United States, United Kingdom, Germany, France and 158 other countries.	**Google Pay** hasn't got a lead over PayPal in any country.
5. Number of active users worldwide	**PayPal** has 361 million active user accounts.	**Google Pay** is expected to reach 100 million user accounts in 2020.
6. Mobile payment priority	**PayPal** is not one of the top priorities for mobile payments.	**Google Pay** ranked as number 2 for most preferred mobile payment services.

Situation Comparison - using Sentiment Analysis

- For Google Pay, people exhibited more positive sentiments as compared to the negative and neutral sentiments. Collectively, number of “very positive” and “positive” tweets are more than other sentiments. The number of neutral sentiments is also significant (second highest). This pattern is quite similar to what we saw in PayPal analysis except for the frequency of negative sentiments which was significantly higher in PayPal’s case.

- The top 5 terms used to express negative sentiments about Google Pay are “issue", “stuck", “fraud", “problem", “failed“. Comparatively, top negative expressions used for PayPal are stronger - "bust", “emergency", “hate", “vomiting", and “sad“.

- The top 5 terms used to express positive sentiments for Google Pay are “easier", “available", “record", “like", “free“. For PayPal, they were - “like", “reputation", “appreciated", “enough", “thank“. They are quite similar.

- Just like PayPal, the number of negative words for Google Pay are more than positive words in the Word Cloud.

- As compared to PayPal, the frequency of negative emotions is significantly low for Google Pay. This shows that people are expressing more negative sentiments for PayPal compared to Google Pay.

Situation Comparison – Clustering and Customer Segmentation

How Google Pay in the same situation managed to solve their problem using a similar method

In this research, we performed cluster analysis to identify key customer segments for PayPal to help the company enter mobile payment market successfully. Google Pay has used a similar approach to segment its mobile payment customers into three distinct categories – 1. those who prefer speed, 2. those who prefer rewards and save money, 3. those who want to manage their money effectively by keeping track of their spending.

To attract the customers in these three segments, Google Pay has recently launched 'Google Pay app' to combine three important features. These three features are three tabs in the app. Google director of product management Josh Woodward describes this as:

"One is the ability to pay friends and businesses really fast. The second is to explore offers and rewards, so you can save money at shops. And the third is getting insights about your spending so you can stay on top of your money."

Since Google Pay is ranked number 2 for mobile payments and using this approach, we believe that creating customer clusters and segments based upon their priorities will help PayPal to gain mobile payment market share successfully.

Conclusion

Model 1, sentiment analysis is focused on studying what sentiments people exhibit about PayPal on social media. This is part of our *brand monitoring* strategy. We found that in most part, people express positive emotions about PayPal. However, a significant number of people also expressed negative and very negative sentiments along with extremely strong negative words. This is not a good sign for PayPal.

Model 2, cluster analysis helped us to identify three actionable market segments for PayPal. Cluster 1 and 2 have shown quite similar priorities such as safety, convenience, saving money, best offers/deals, and customer service. Cluster 3 only cared about the safety of the transactions and receiving best offers/deals. Based upon our study of these clusters, we conclude that cluster 2 is going to be PayPal's *Priority 1 customer segment*, cluster 2 will be *Priority 2*, and cluster 3 will be *Priority 3 customer segment*. We decided to exclude cluster 4 from our final set of segments due to the small number of respondents in the cluster and not highly significant responses.

Conclusion

We compared PayPal and Google Pay at three levels – brand and business comparison, social media sentiment comparison, and customer segment comparison. Overall, people have expressed more negative sentiments for PayPal as compared to Google Pay. PayPal's brand and business, globally, seems to be more successful but it is struggling to enter mobile payment market. We also found that Google pay has also adopted similar customer segmentation approach (segmenting customers based upon their priorities) and achieved great success in the market. Our market research for clustering is also based upon identifying people's priorities for mobile payments. These clusters helped us to finally create three actionable market segments for PayPal.

With our study, we achieved our success criteria. Long term: PayPal will use these segments to effectively design and market its services which will help the company to increase its revenue by $2.7 billion (~15% increase) in 1 year. Short term: we successfully identified social media sentiments for PayPal compared to the sentiments expressed for its competitor Google Pay. Supported by market research, we also identified top three, high priority customer segments for PayPal's mobile payment service.

Recommendations

- Very high expression of negative emotions is not good for company's image and overall bottom-line. PayPal should improve its reputation by launching a major social media marketing campaign in January 2021. The theme of the campaign - transaction safety, convenience, and rewards.

- Through our market research we identified the top three priorities for mobile payments – safety, convenience, and rewards (best offers/deals). Create an app to combine these three features into three tabs in the app. Google Pay's app is quite similar to this.

- The app should have built-in authentication, transaction encryption, and fraud protection systems. Allow the app users to choose the privacy settings that are right for them.

- For convenience allow users to keep credit cards, debit cards, transit passes, and more on their phone for quick, and easy access.

- A significant number of people seem to like rewards. Offer them more cashback rewards when they activate offers from their favorite businesses and make everyday purchases using PayPal's mobile payment app.

- Advertise PayPal's new app on Facebook, Twitter, and YouTube to reach 50 million users by June 2021. By December 2021, PayPal's new app will have 35 million new subscribers which will help the company to increase its revenue by $2.7 billion.

Bibliography

- *Online payments by type in the United States 2020*. (2020). Statista. https://www.statista.com/forecasts/997125/online-payments-by-type-in-the-us
- *Online payments by brand in the United States 2020*. (2020). Statista. https://www.statista.com/forecasts/997132/online-payments-by-brand-in-the-us
- *PayPal: annual revenue 2019*. (2020). Statista. https://www.statista.com/statistics/382619/paypal-annual-revenue/

Library, H. B. (n.d.). *Statista | Statistia | Baker Library | Bloomberg Center | Harvard Business School*. Retrieved December 3, 2020, from https://www.library.hbs.edu/Find/Databases/Statista

- *About SimilarTech*. (n.d.). https://www.similartech.com/about
- Statista. (2020). *PayPal-Statistics and Facts*. Statista. https://www.statista.com/topics/2411/paypal/
- Statista. (2020). *Digital wallet users by company*. Statista. https://www.statista.com/statistics/722213/user-base-of-leading-digital-wallets-nfc/
- *Google Pay VS PayPal - Payment Technologies Market Share Comparison*. (n.d.). SimilarTech. https://www.similartech.com/compare/google-pay-vs-paypal
- *Online payments by brand in the United States 2020*. (2020). Statista. https://www.statista.com/forecasts/997132/online-payments-by-brand-in-the-us
- Google Pay gets a major redesign. (2020). *TechCrunch*. https://social.techcrunch.com/2020/11/18/google-pay-gets-a-major-redesign/