

# SOCIAL MEDIA ANALYTICS FOR CANADIAN BANKS (USING TWITTER)

Student: Chris Tan (303428)

Instructor: Matthew Tenney

Course #: CSDA I050FI8SI

# CONTENTS

| Subject                             | Page  |
|-------------------------------------|-------|
| Abstract and scope                  | 3     |
| Research questions                  | 4     |
| Research method                     | 5     |
| Literature review                   | 6     |
| Addressing the research question #1 | 7-10  |
| Addressing the research question #2 | 11-12 |
| Addressing the research question #3 | 13-16 |
| Addressing the research question #4 | 17-21 |
| Conclusion                          | 22    |
|                                     |       |
|                                     |       |
|                                     |       |
|                                     |       |
|                                     |       |

## ABSTRACT AND SCOPE

**Abstract:** This is for the fulfillment of the York University's Advanced Analytics Course Capstone Project. The aim of this project is to uncover insights from the social media space through programmatic means.

**Project Scope:** Here are the boundaries of the project:

1. **Social media channel:** Twitter
2. **Social media scope:** Major Canadian Financial Institutions (FI) like BMO, CIBC, RBC, Scotiabank, TD
3. **Comparison** of the following insights across the above FIs: Sentiment Analysis (polarity and categorical); Word Cloud (conversation drivers); Key-word dendrogram (blend of sentiment and conversation drivers); Network Analysis (demographics and product segmentation). Paraphrases of these insights are given in the "Research Questions" section below

# RESEARCH QUESTIONS

Based on the exploratory data analysis (EDA) from Sprint #1, the research questions are revised as follows:

1. Which bank has the most favourable / unfavourable trending opinion?

*Comments and revision:* About 1,623 tweets have been collected since July 7th, 2019, with close 5,500 terms. The collection will increase in the next several weeks. It should be feasible to answer this research question. The main drawback is for the low count of CIBC tweets (40 tweets) versus that of Scotia Bank (661 tweets). The wide difference will skew the analysis, especially that of CIBC's

2. What are the current financial products being discussed?

*Comments and revision:* The EDA shows that frequent terms related to banking products are generic ones, for example, "stock", "charges", "account". Unless we have a much more collection of tweets, it will be difficult to objectively address this research question

3. What are the current emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) towards each bank?

*Comments and revision:* Not shown in this Sprint#1 report as the codes are still experimental, I managed to "see" these emotional terms at the "AllBanks" level. Again, due to the low tweet count for CIBC, it may be difficult to pin down the sentiments, especially for these 8 sentiment categories.

4. What are the current sentiments towards trending financial product segments / categories (and the general network of terms being tweeted)?

*Comments and revision:* As stated above, frequent terms related to banking products are generic ones, hence it will be difficult to assess sentiments towards product segments. Network of terms is certainly a possibility

# RESEARCH METHOD

Here is the general approach I adopted for this project. See the R codes for the detail steps:

## Sprint #1:

1. Data Preparation
2. Exploratory Data Analysis

## Sprint #2:

3. Social Media Analytics for Canadian Banks

## Sprint #3:

4. Addressing the Research Questions

# LITERATURE REVIEW

Here is a summary of the literatures that are relevant to this project

1. <https://ai.googleblog.com/2006/08/all-our-n-gram-are-belong-to-you.html> extracted on 14 July 2019, which describes example of N-grams analysis, which determines the likelihood of next word or character given N previous terms, by Google

2. [http://cs229.stanford.edu/proj2015/044\\_report.pdf](http://cs229.stanford.edu/proj2015/044_report.pdf) extracted on 14 July 2019, entitled “Detecting Sarcasm in Text: An Obvious Solution to a Trivial Problem”. Unlike humans, it is very difficult for machines to detect, let alone understand, sarcasm. This paper attempts “to design a machine learning algorithm for sarcasm detection in text”

3. <https://github.com/mjockers/syuzhet> extracted on 14 July 2019, which provides documentation on the R package, syuzhet, for sentiment analysis. It is based on concepts of “fabula” and “syuzhet” where the former is “the raw material of a story (chronology)” and the latter is “the way a story is organized (technique of the narrative)”. As such, the syuzhet R package helps reveals “emotional shifts that serve as proxies for the narrative movement between conflict and conflict resolution”

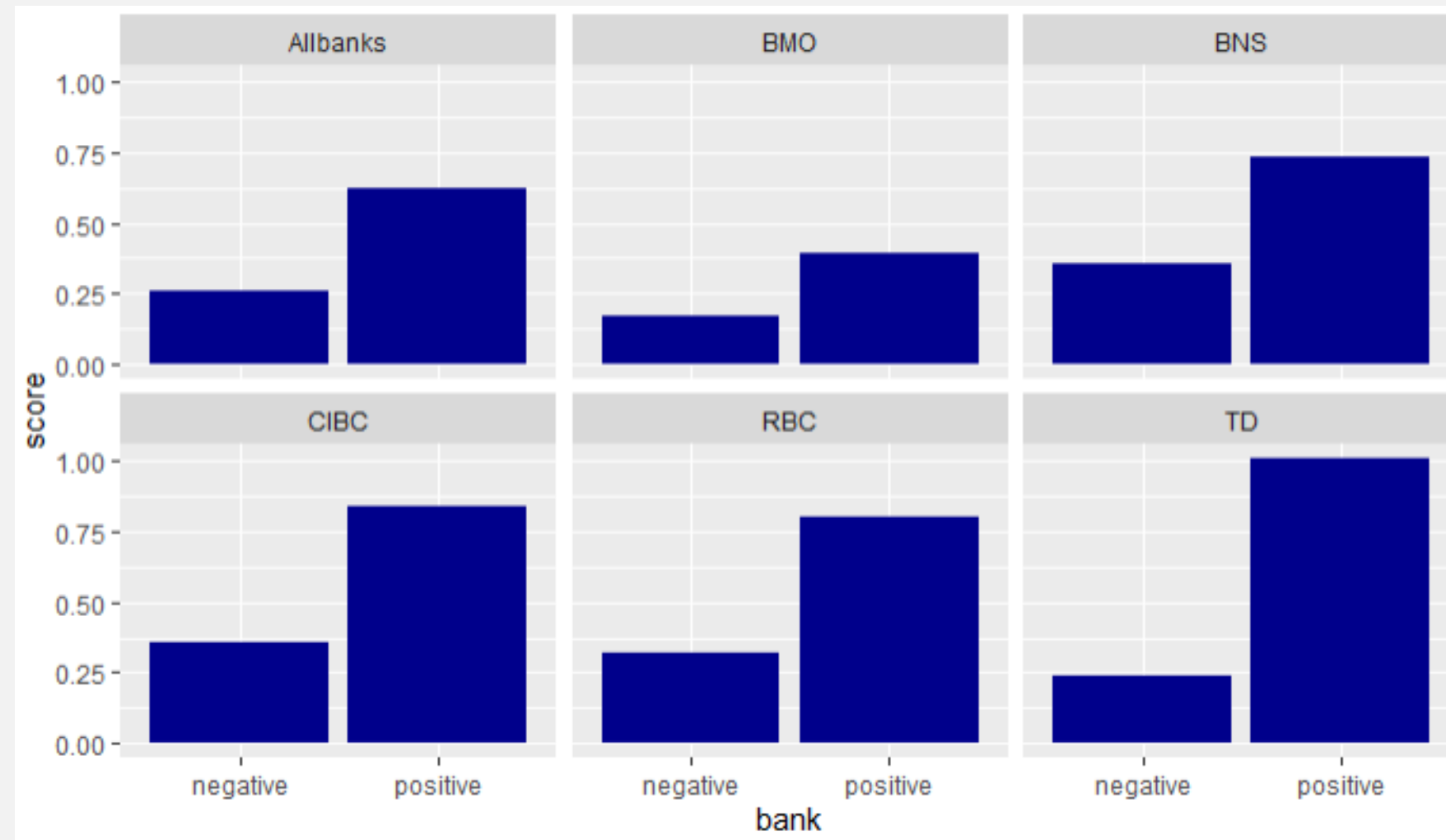
4. <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html> extracted on 14 July 2019. This article describes mining of online opinions, and then determine whether the opinions are positive or negative (opinion polarity)

5. <http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.html> extracted on 14 July 2019. This article helps take sentiment analysis to the next step. That is, besides sentiment polarity (positive or negative), we now have the ability to categorise sentiments into the eight basic human emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust)

6. <https://igraph.org/r/doc/aaa-igraph-package.html> extracted on 7 July 2019, which documents how to implement “follower graph” network analysis using the R igraph package

# ADDRESSING THE RESEARCH QUESTIONS#1

## I. Which bank has the most favourable / unfavourable trending opinion?



It is apparent that TD has the most favourable trending opinion. Inspection of the most frequent terms used for TD suggest TD's favourable opinion is helped by the following frequent terms: hiring, fit, like, money, love (See Figure 2)

Figure 1: Trending Opinion for Canadian Banks

# ADDRESSING THE RESEARCH QUESTIONS#I

## I. Which bank has the most favourable / unfavourable trending opinion?

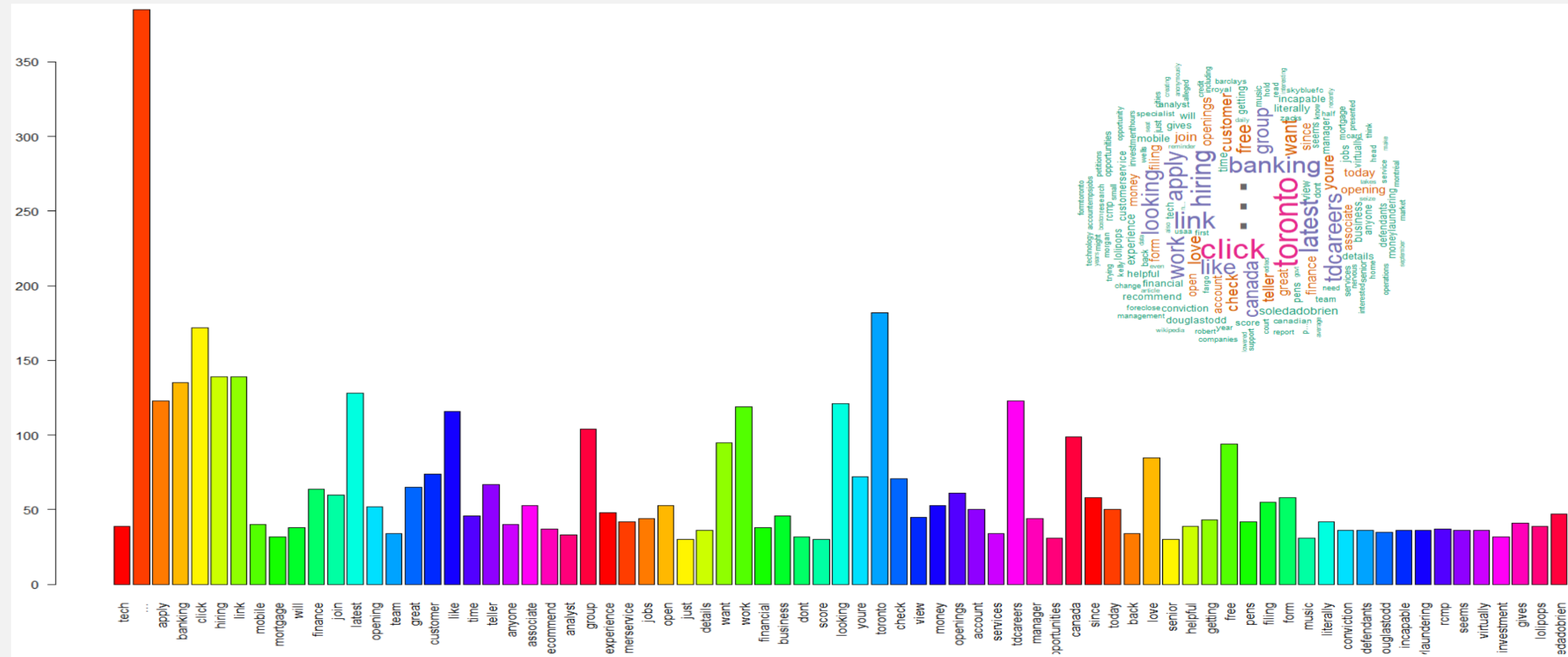


Figure 2: Most frequent terms used and wordcloud for TD



# ADDRESSING THE RESEARCH QUESTIONS#I

## I. Which bank has the most favourable / unfavourable trending opinion?

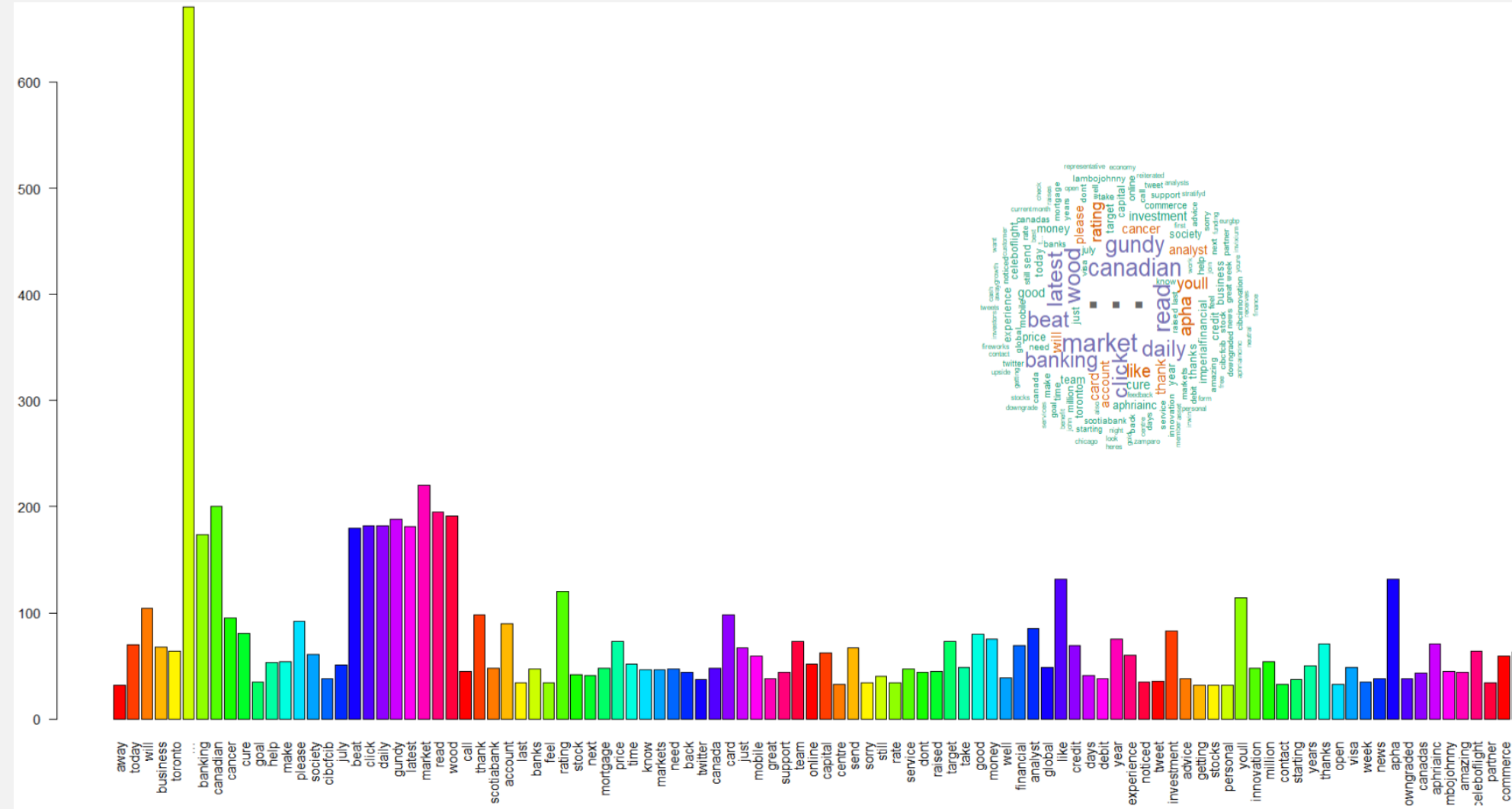


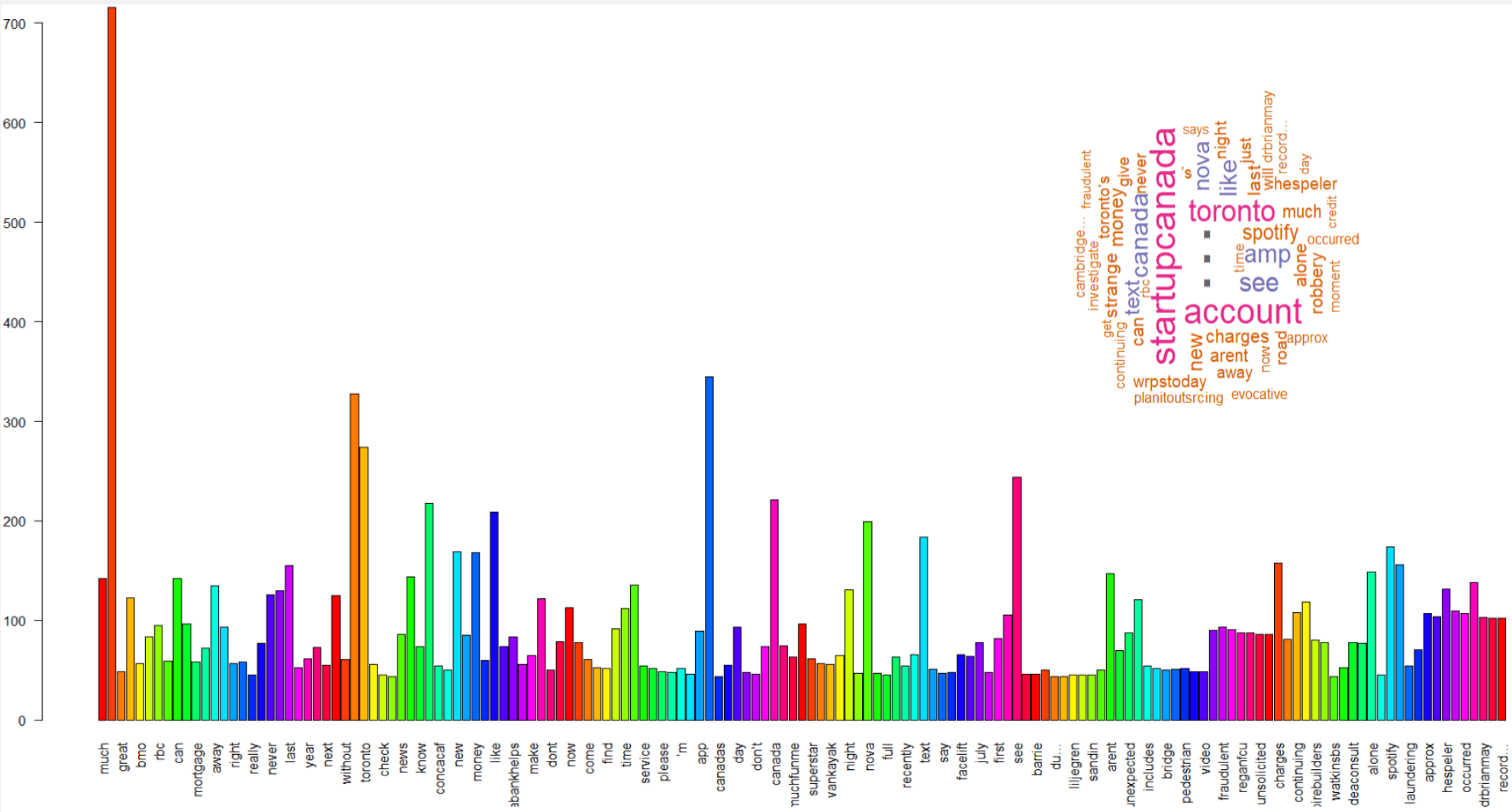
Figure 3: Most frequent terms used and word cloud for CIBC

The most unfavourable trending opinions are towards CIBC and Scotia Bank (BNS). Inspection of the most frequent terms used for CIBC and BNS suggest the impact of the following negative terms:

- For CIBC: away, last, dont, sorry (Figure 3)
- For BNS: robbery, hasnt, strange (Figure 4)

## ADDRESSING THE RESEARCH QUESTIONS#1

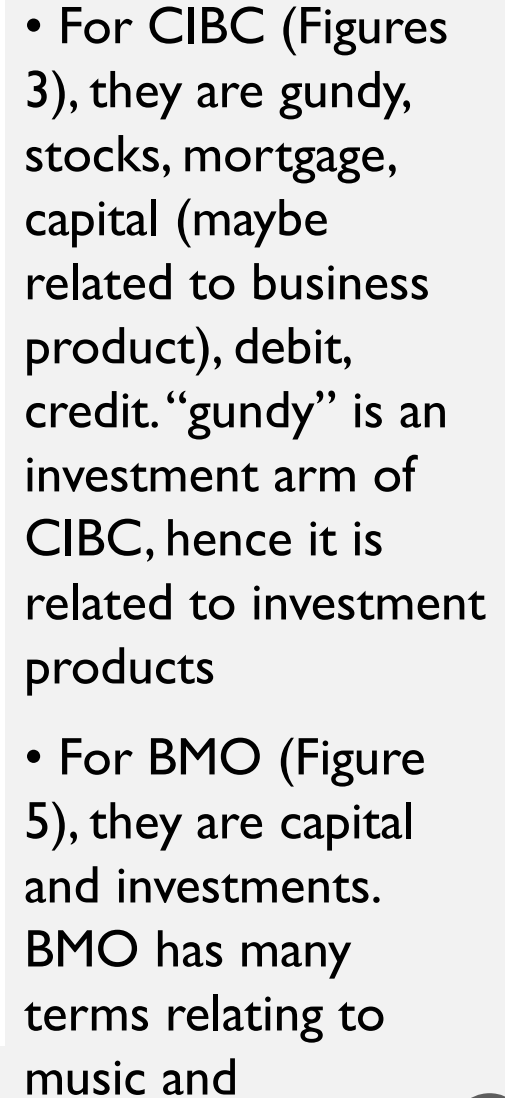
I. Which bank has the most favourable / unfavourable trending opinion?



The mitigating factor is that the positive opinion of both BNS and CIBC are higher than the average for Allbanks. Even though, four of the five banks have higher than average positive opinion, both CIBC and BNS have higher than average opinion - it is because of the enormous BMO tweet count (more than 13,400, about 45% of total tweets of all banks), and BMO has a low positive rating

### Figure 4: Most frequent terms used and wordcloud for BNS

## 2. What are the current financial products being discussed?



# ADDRESSING THE RESEARCH QUESTIONS#2

## 2. What are the current financial products being discussed?

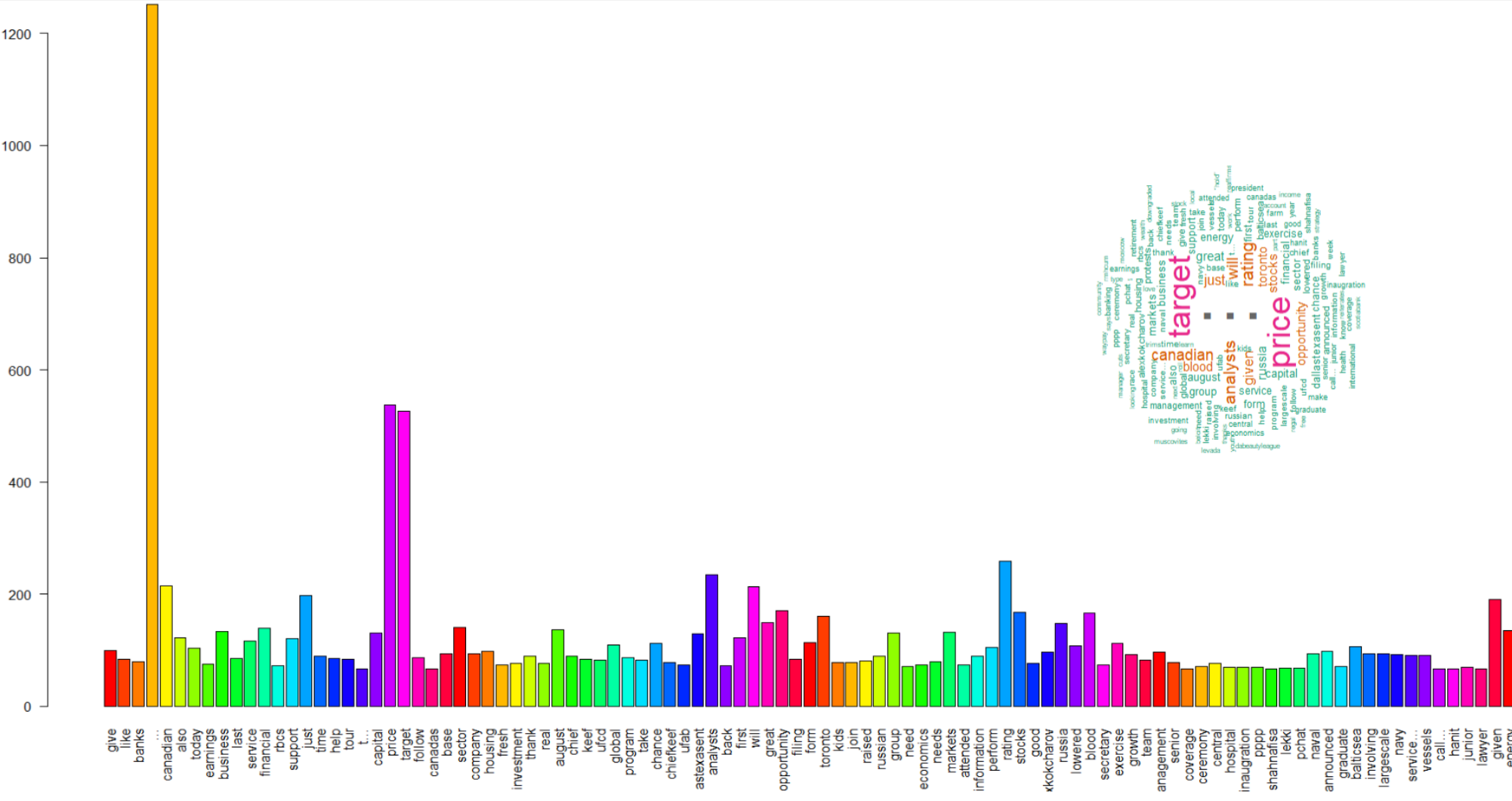


Figure 6: Most frequent terms and wordcloud for RBC

- For BNS (Figure 4), they are credit and business. Like BMO, who sponsors the BMO Field, BNS sponsors the Arena; hence the events at the Arena may cloud the terms relating to financial products
- For RBC (Figure 6), capital, investments, stocks. There is an interest term "energy" in the VWord cloud. Perhaps, RBC has growing interest in the "energy" sector
- For TD (Figure 2), mortgage. There are many job postings in TD tweets, hence, these may cloud the related financial product terms

# ADDRESSING THE RESEARCH QUESTIONS#3

3.What are the current emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) towards each bank?

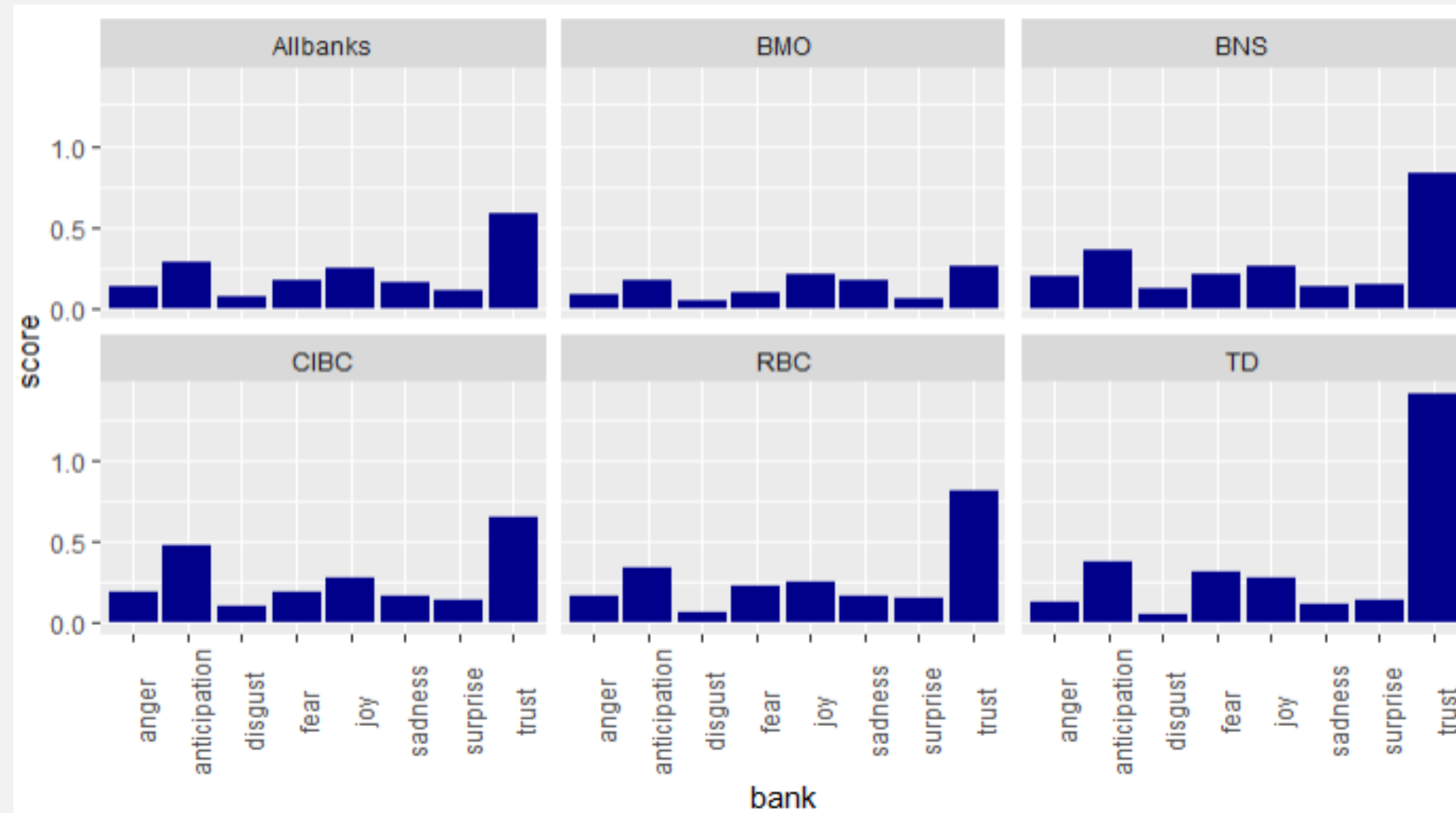


Figure 7: Sentiment Scores for Canadian Banks

- **Anger**: Most Anger is towards BNS and least is towards BMO
- **Anticipation**: Most Anticipation is towards CIBC and least is towards BNS. Anticipation is typically related to earnings release
- **Disgust**: Most is BNS and the least is BMO
- **Fear**: Most is TD, and the least is BMO

# ADDRESSING THE RESEARCH QUESTIONS#3

3.What are the current emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) towards each bank?

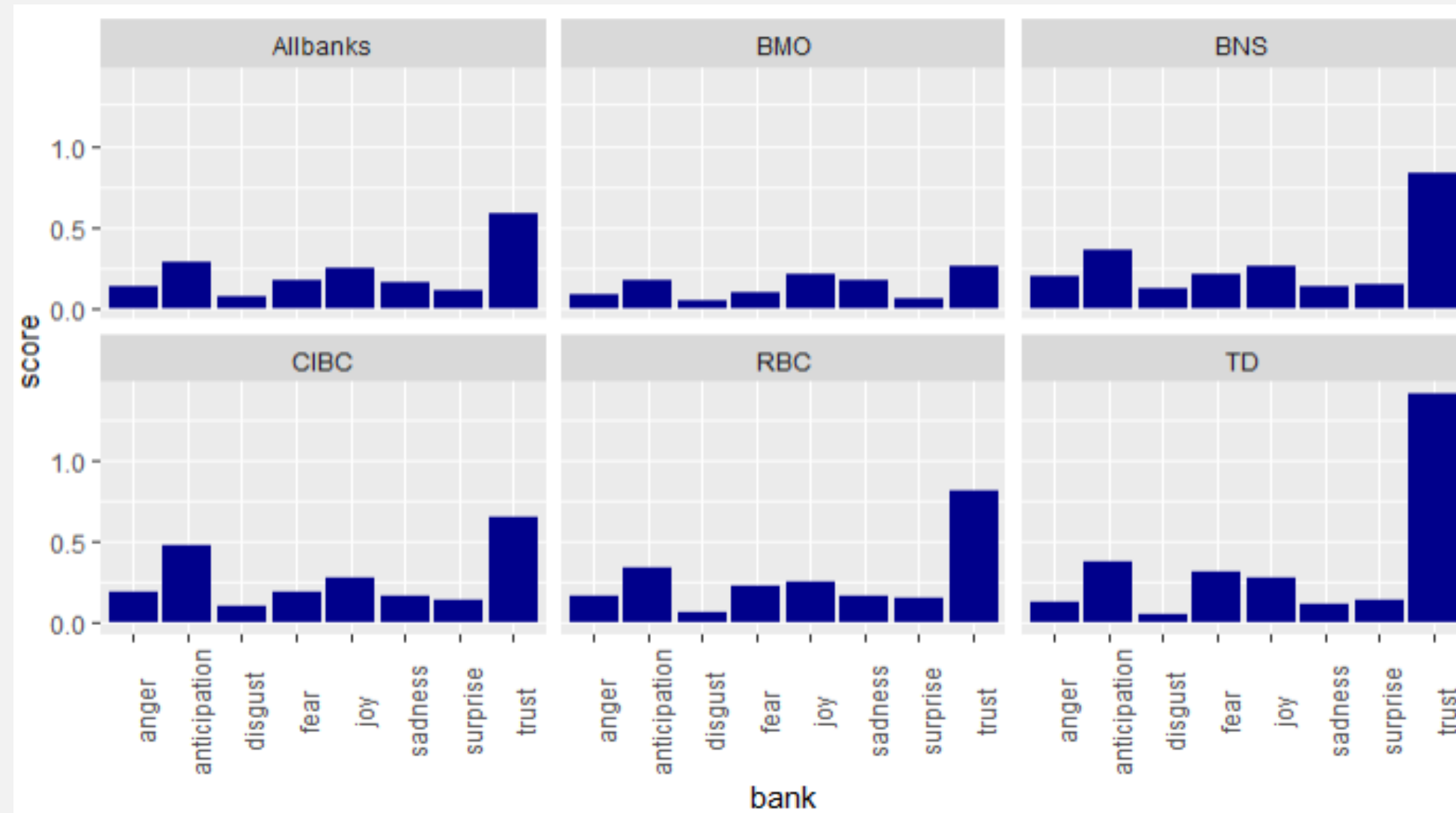
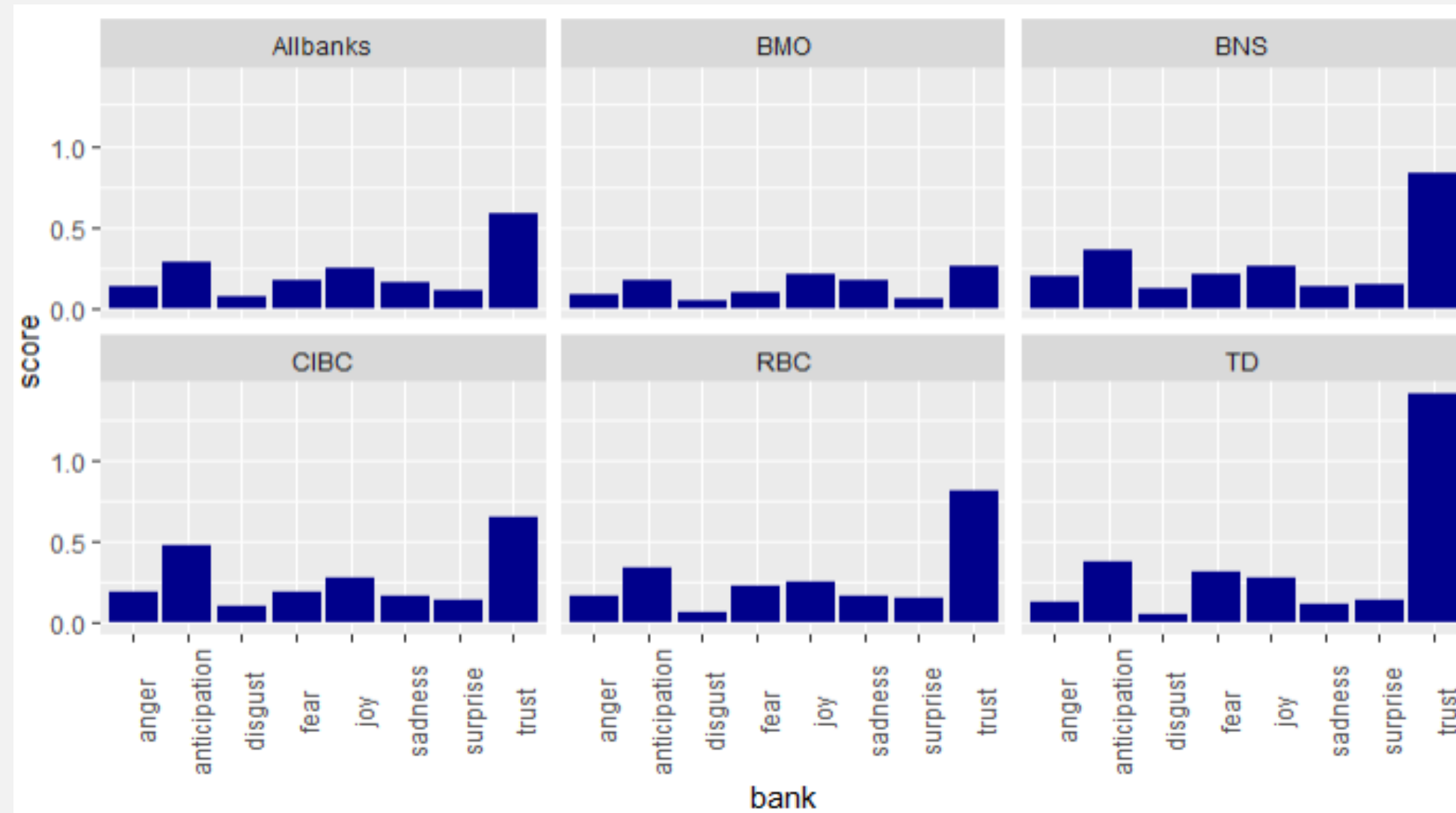


Figure 7: Sentiment Scores for Canadian Banks

- **Joy**: Most is TD, and the least is BMO. For TD, base on Figure 8, where existence network-of-terms "lollipops-gives-pens-literally-free" suggest there maybe a joyful promotional event going on
- **Sadness**: Most is CIBC, and the least is TD. For CIBC, it may be attributed to sad discussion of impact of cancer (CIBC is a major sponsor of the Run For The Cure for cancer). Not surprising, more joyful tweets in TD (see preceding note) will suppress the sadness

# ADDRESSING THE RESEARCH QUESTIONS#3

3.What are the current emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) towards each bank?



- **Surprise:** Most is RBC, least is BMO
- **Trust:** Most is TD, least is BMO. The initial hunch is that people are tweeting TD Canada "Trust". But inspection of the most frequent term used for TD in Figure 2 showed "trust" is not in the frequent term list. So this results does not appear to be fluke.

Figure 7: Sentiment Scores for Canadian Banks



# ADDRESSING THE RESEARCH QUESTIONS#3

3.What are the current emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) towards each bank?

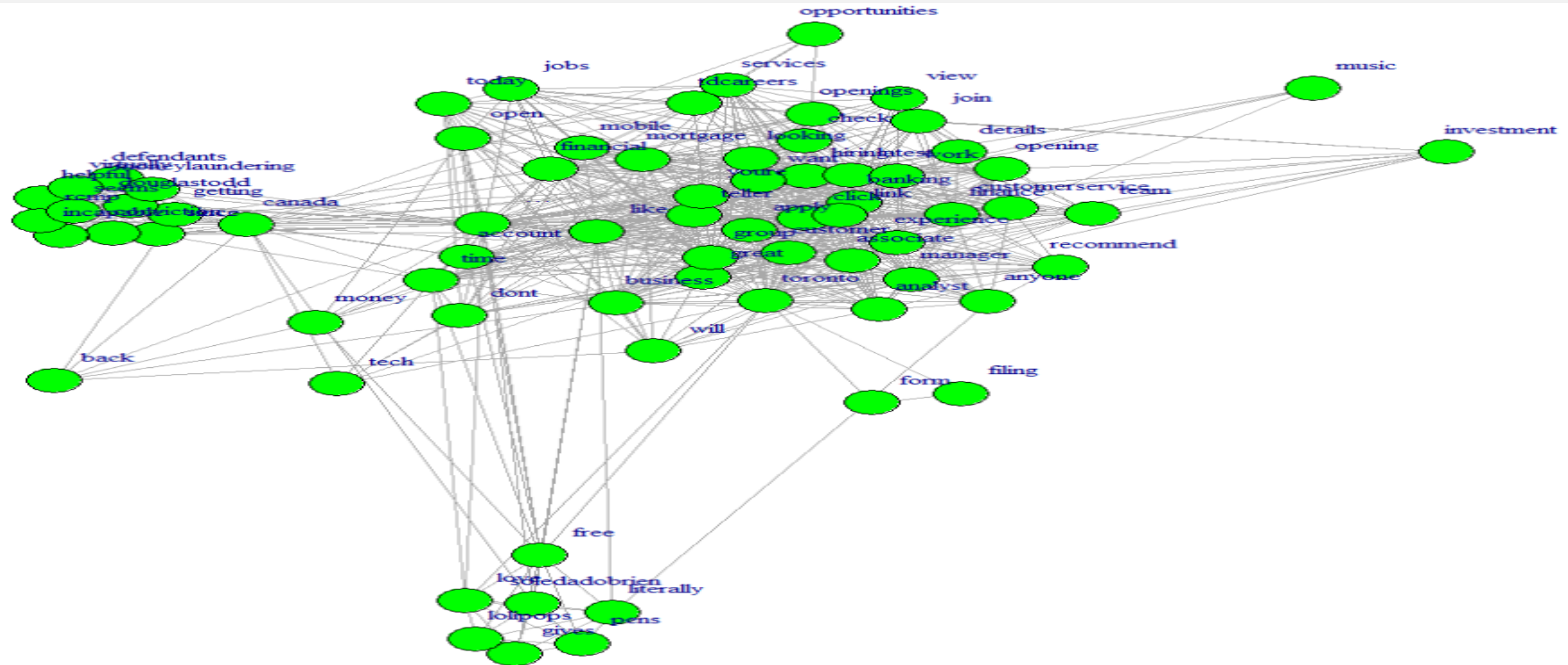


Figure 8: Network of terms for TD



# ADDRESSING THE RESEARCH QUESTIONS#4

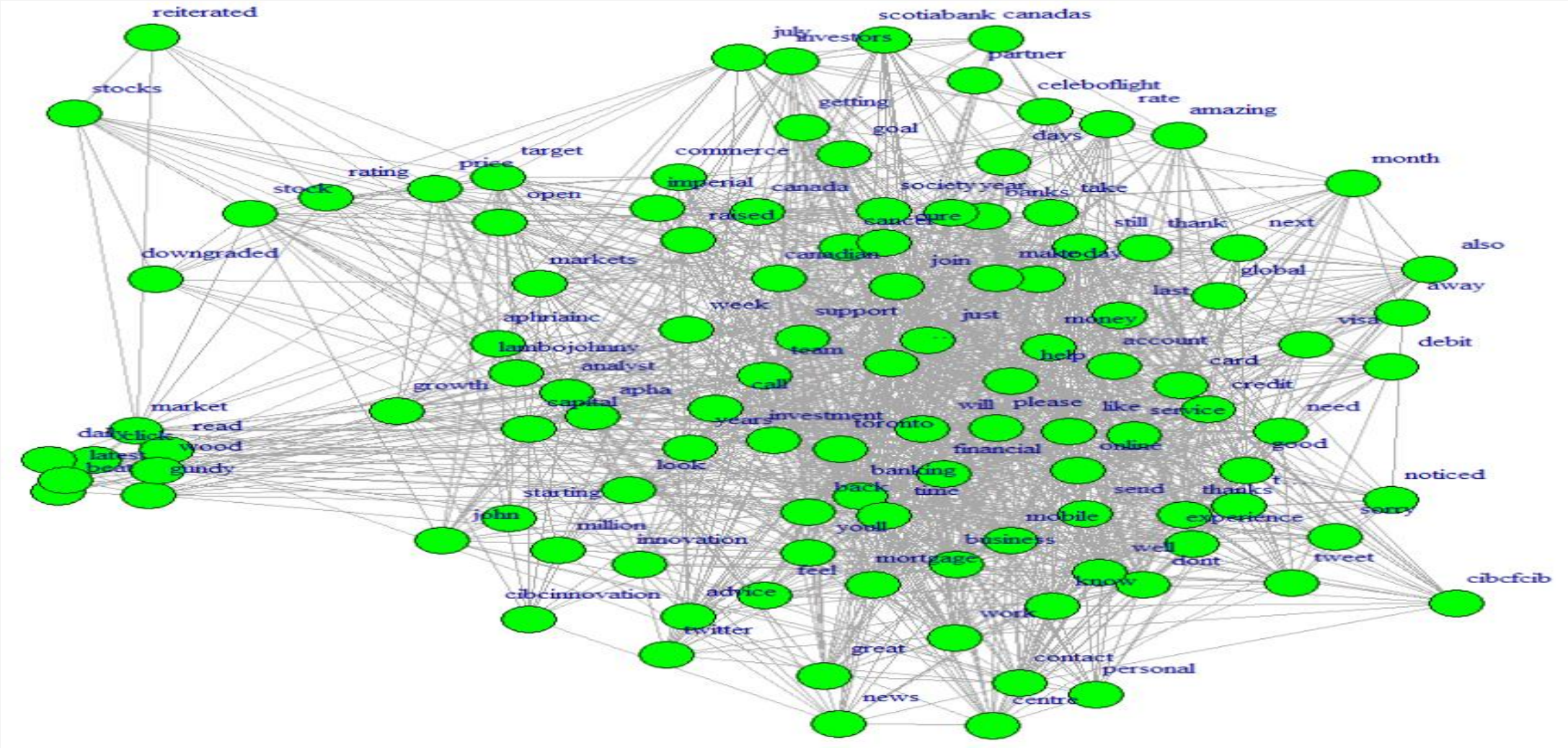
4. What are the current sentiments towards trending financial product segments / categories (and the general network of terms being tweeted)?

- Cluster of good-experience (although there is a sub cluster of sorry-noticed-tweet)
- Cluster of mortgage-mobile-online-back-great

Figure 9: Cluster dendrogram of terms for [CIBC](#)

## ADDRESSING THE RESEARCH QUESTIONS#4

4. What are the current sentiments towards trending financial product segments / categories (and the general network of terms being tweeted)?



**Figure 9: Cluster dendrogram of terms for CIBC**

- Nothing significant here
- CIBC has the least tweets for all the five banks. It does sponsor the Run For The Cure (RFTC) for cancer, but the RFTC tweets does significantly cloud the analysis

4. What are the current sentiments towards trending financial product segments / categories (and the general network of terms being tweeted)?

- Nothing significant for both Cluster dendrogram and Network of terms diagram for the following banks
  - **BMO**
  - **TD**
- See visualisations within the R code or in this file “Social-Media-Analytics-for-Canadian-Banks-Sprint-02--v12.pdf” for details)

# ADDRESSING THE RESEARCH QUESTIONS#4

4. What are the current sentiments towards trending financial product segments / categories (and the general network of terms being tweeted)?

- Cluster of july-credit-mortgage-since-hasn't
- Nothing significant based on network of terms diagram (not shown)

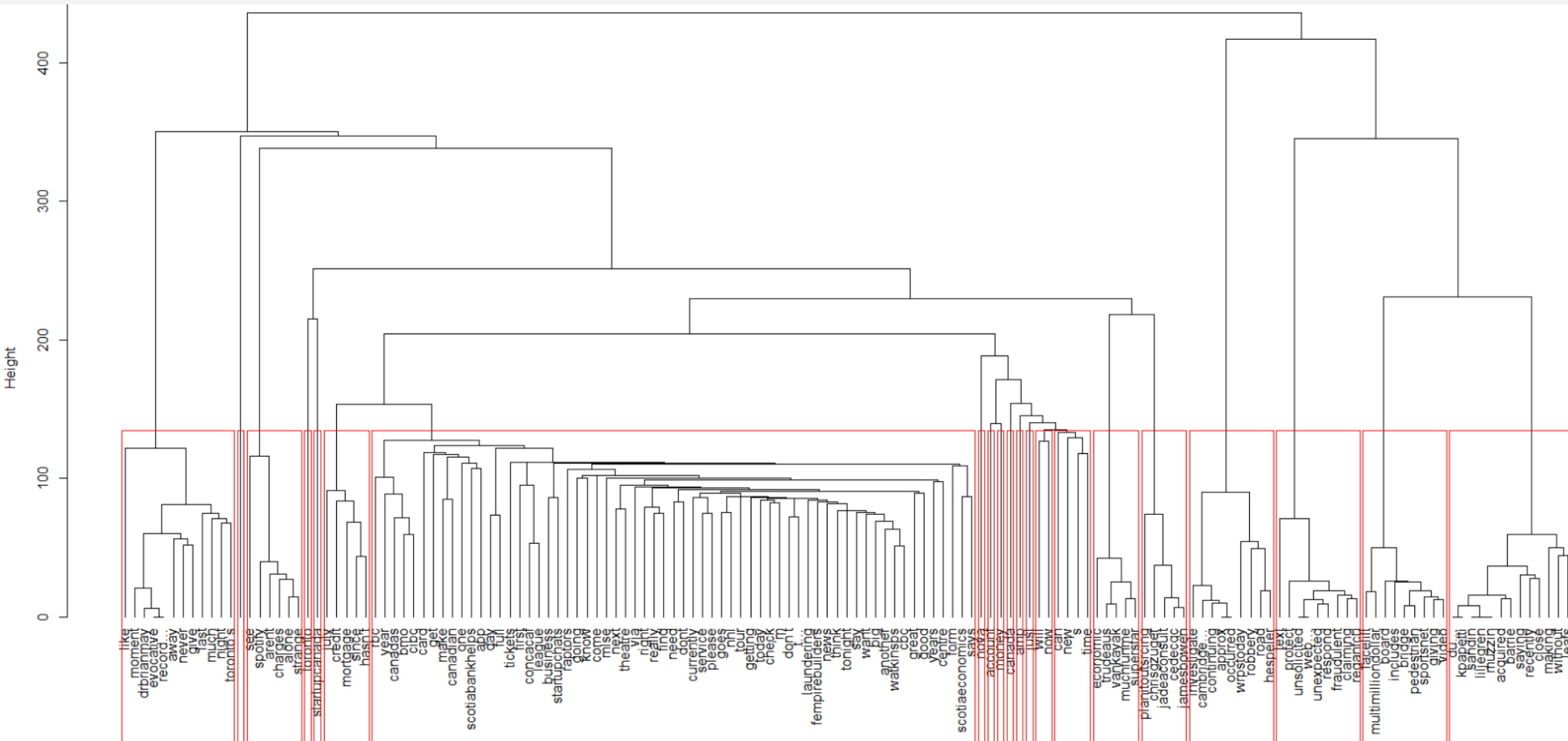
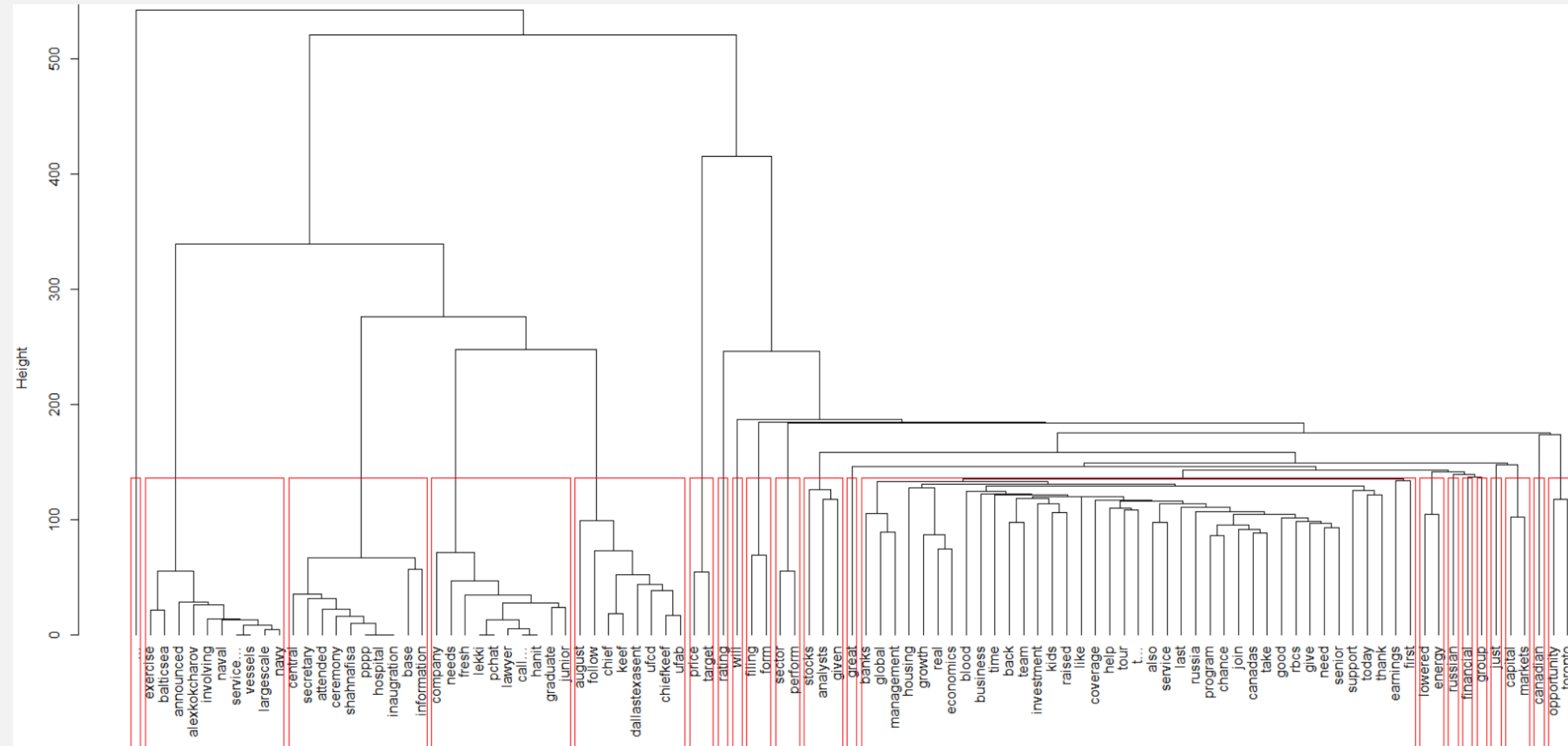


Figure 11: Cluster dendrogram of terms for **Scotia Bank (BNS)**

# ADDRESSING THE RESEARCH QUESTIONS#4

4. What are the current sentiments towards trending financial product segments / categories (and the general network of terms being tweeted)?



- Cluster of back-team-investment-kids-raised-like
- RBC has many tweets about job opportunities that would have clouded this analysis
- Nothing significant based on network of terms diagram (not shown)

Figure 12: Cluster dendrogram of terms for [RBC](#)



# CONCLUSION

This project outlines a Twitter analytical method to uncover the following insights for the Canadian banks

- Which bank has the most favourable / unfavourable trending opinion?
- What are the current financial products being discussed?
- What are the current emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) towards each bank?
- What are the current sentiments towards trending financial product segments / categories (and the general network of terms being tweeted)?