MATH513 Practical Presentation

10570155, 10696253, 10701983

10/12/2020

Introduction

- Samsung and Apple
- Flagship phones chosen
  - S20FE
  - iPhone12
  - S20

Tools Utilised

- Rstudio
- RTweet
- Twitter Developer API
- GitHub


Research

Choosing Twitter for Analysis

- Open API Access compared to others
- Almost all data is public
- Advanced filtering and queries
- Generous Rate limiting

Hashtags

- @SamsungMobile - https://twitter.com/SamsungMobile
- @Apple - https://twitter.com/Apple
- @tim_cook - https://twitter.com/tim_cook

Data Cleaning and Feature Engineering

Data Cleaning

- Duplicate tweet and user observations were removed
- Tweet text and user bios were cleaned
  - Removed links, hash-tags, emojis, and user mentions

Feature Engineering

- Users were marked as potential bots
- User country was extracted from the location of their profile
- Tweets were marked as potential spam
- Hash-tags were extracted from the tweet text
- Product features were extracted from the tweet text
  - Display, Battery, Camera, Price, and 5G Capability
- An overall sentiment score was calculated for each tweet

Summary of Collected Data

Total Tweets: 73690 after data cleaning

Total Features: `5 (Display, Battery, Camera, Price, and 5G)`

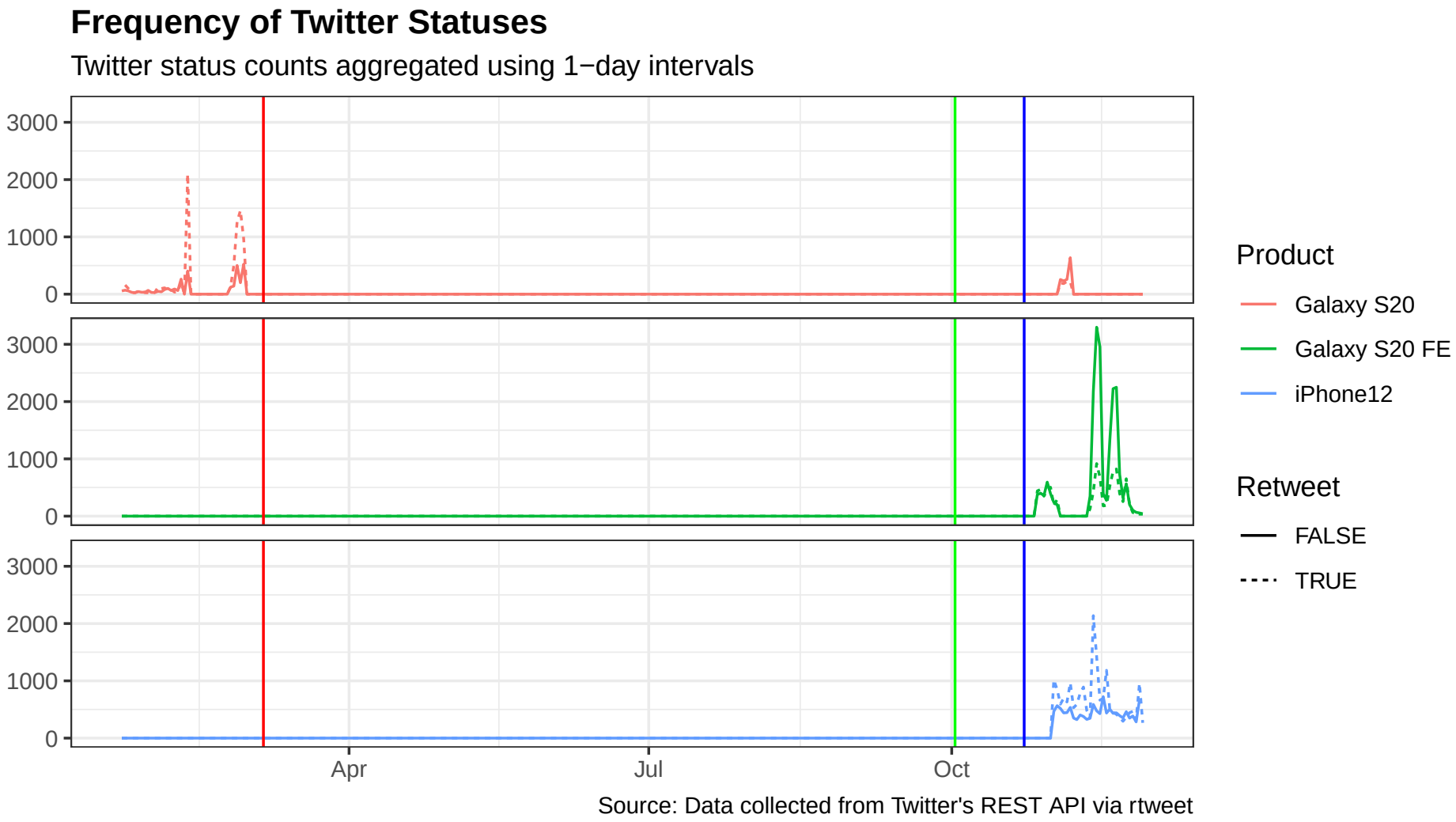
Table 1: Summary of Tweet Data

Product	Number of Tweets	% Spam Tweets	% Feature Tweets
Galaxy S20	13147	3%	20%
Galaxy S20 FE	28923	19%	19%
iPhone12	31620	13%	7%

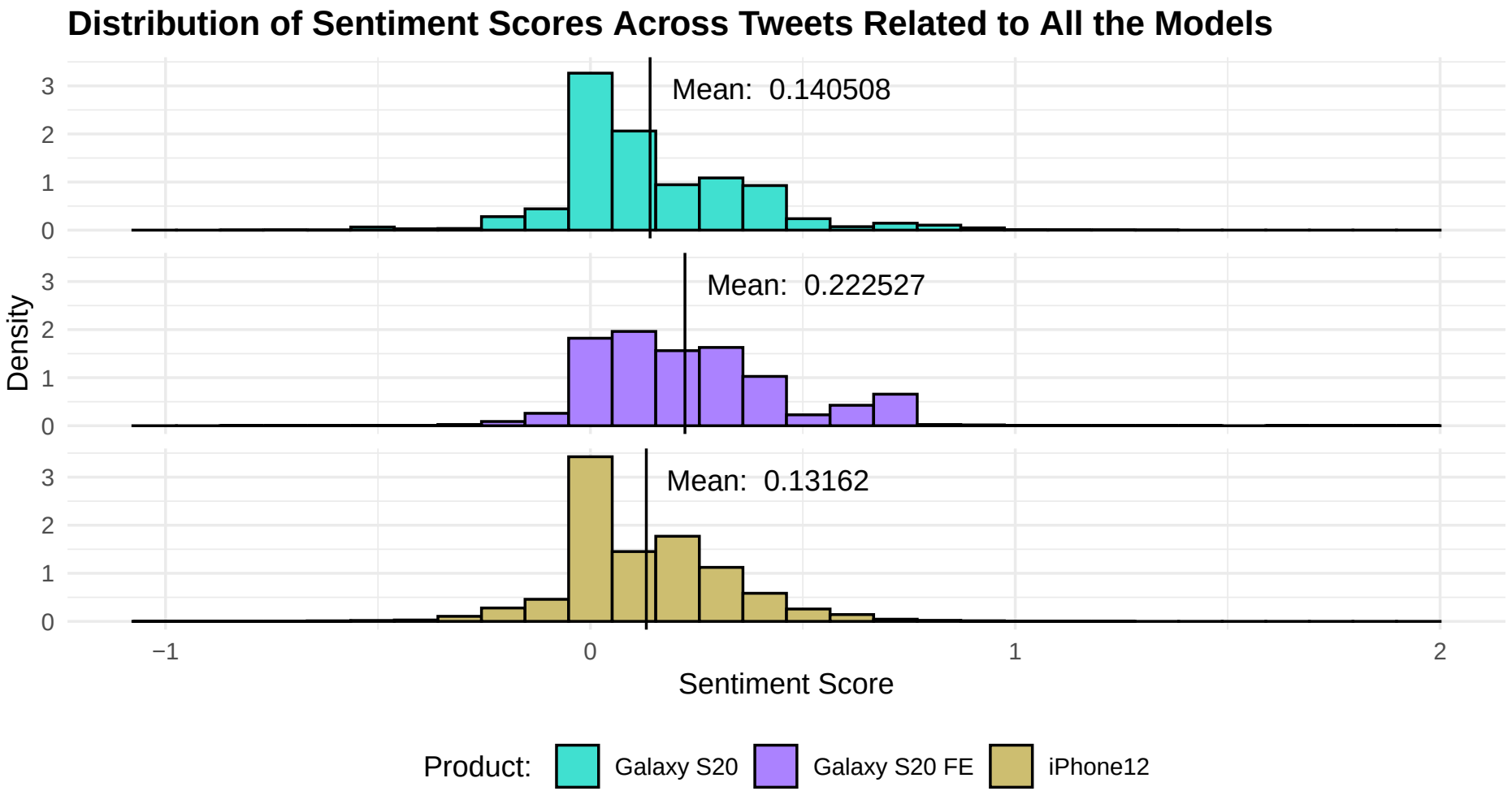
Table 2: Summary of User Data

Number of Users	% Bot Users	Unique Countries
35051	>1%	163

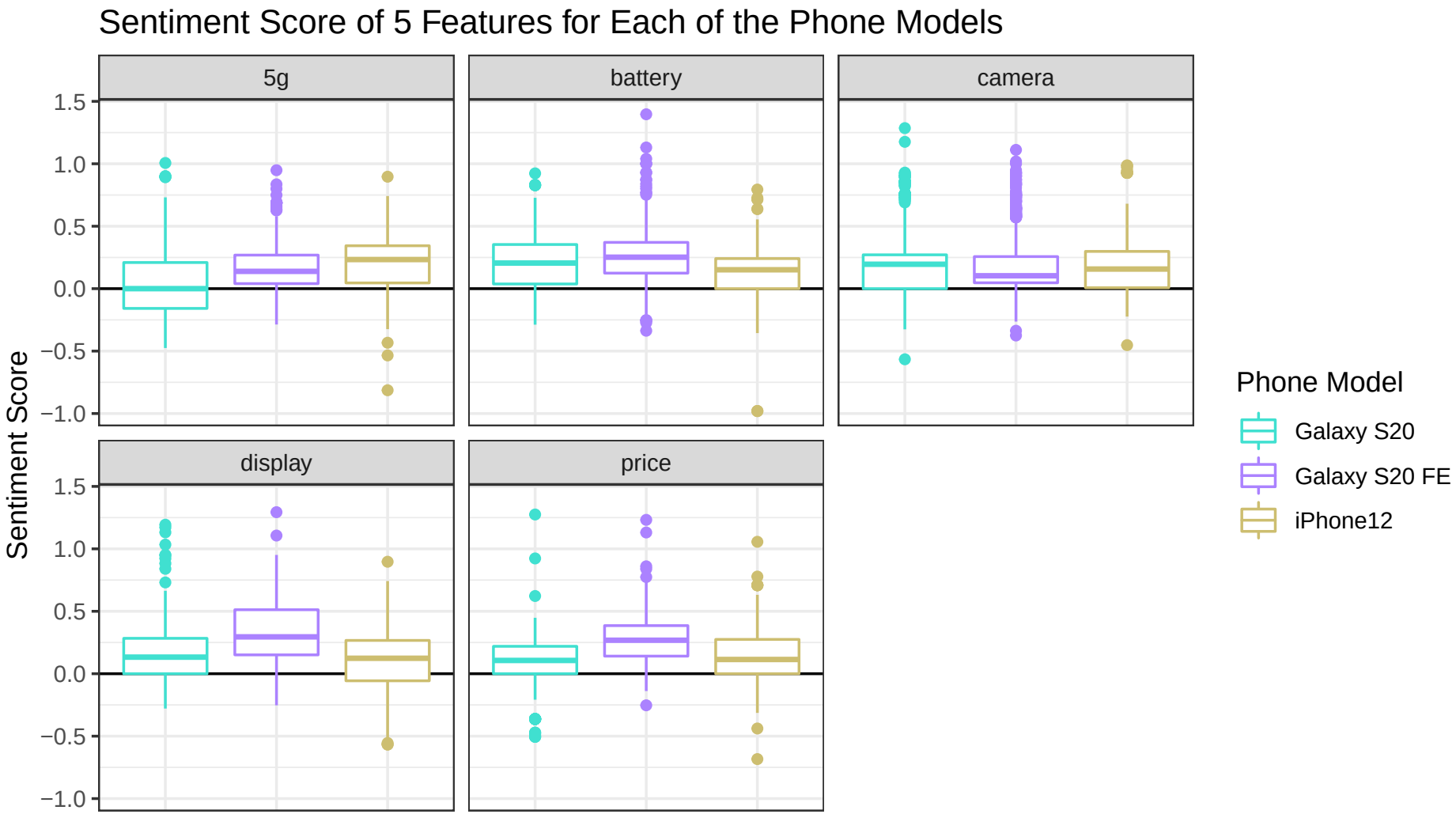
Time Periods for Data Collection



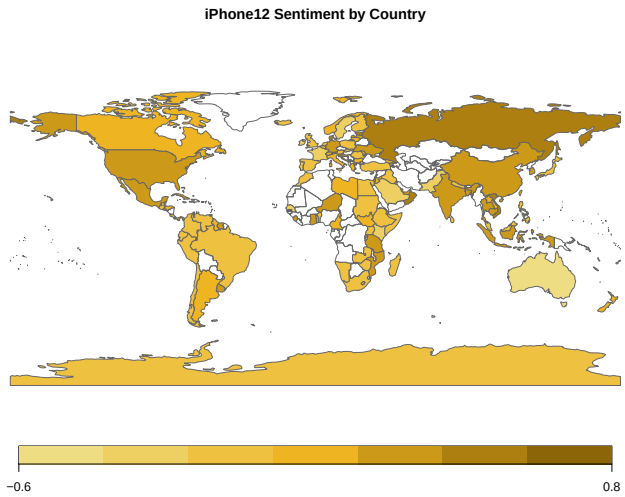
Results - Sentiment Analysis - All Tweets

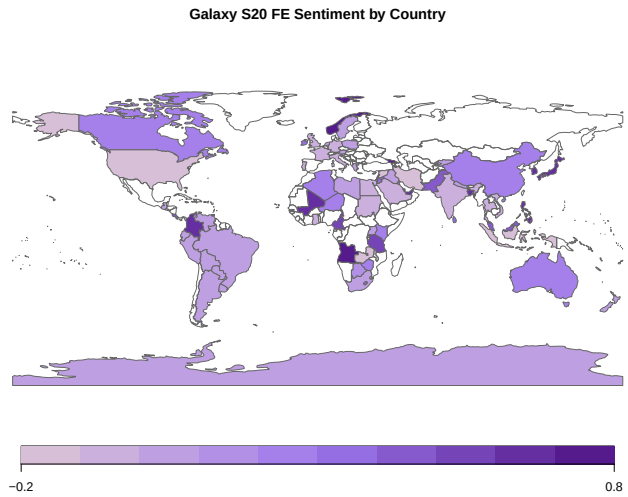


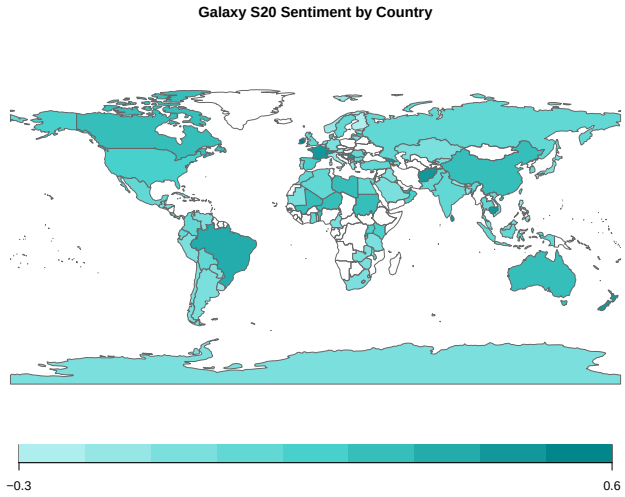
Results - Sentiment Analysis - Features



Results - Sentiment Globally







Statistical Test - T-Test

Improvements & Further Study

Improvements

- Google Maps API
- Look at mentions of apple in samsung and vice versa

Conclusions

- Twitter data provides up to date information for companies to analyse for customer feedback
- Data can provide useful information to guide product teams when analysed correctly

References

- Ahmed, Wasim (2019). *Using Twitter as a data source: an overview of social media research tools* Available at: https://blogs.lse.ac.uk/impactofsocialsciences/2019/06/18/using-twitter-as-a-data-source-an-overview-of-social-media-research-tools-2019/ (Accessed: 07 December 2020)
- Dalla Valle, Luciana (2020). *MATH513 Lecture and Tutorial Code* Available at: https://dle.plymouth.ac.uk/course/view.php?id=49628 (Accessed: 01 October 2020)
- Fuchs, Matti (2018) *Doing your first sentiment analysis in R with Sentimentr* Available at: https://towardsdatascience.com/doing-your-first-sentiment-analysis-in-r-with-sentimentr-167855445132 (Accessed: 06 December 2020)
- Rinker, Tyler (2020). *R Documentation - sentiment_by* Available at: https://www.rdocumentation.org/packages/sentimentr/versions/2.7.1/topics/sentiment_by (Accessed: 06 December 2020)
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.