

# Analyzing mobile popularity in Walmart using Twitter Data

Keerthana Krishnamurthy  
Computer Science  
Syracuse University  
Syracuse, USA  
kekrishn@syr.edu

Bhumika Murali  
Computer Science  
Syracuse University  
Syracuse, USA  
bhmurali@syr.edu

**Abstract—** In this project, we aimed to conduct a comprehensive analysis of Twitter data to determine the most popular brand of mobile phones in two major regions of the United States: New York and California. We collected Twitter data using relevant hashtags and keywords, and preprocessed it to remove noise and irrelevant information. We analyzed the data using natural language processing techniques and machine learning algorithms to identify the most popular brand, customer sentiment towards each brand, and key features that customers appreciate. We used a variety of relevant hashtags and keywords to collect a large dataset of Twitter data related to Apple, Samsung, and Google Pixel phones. We then preprocessed the data by removing irrelevant information and noise to ensure that our analysis was based only on relevant tweets.

Our analysis involved the use of natural language processing techniques and machine learning algorithms to extract valuable insights from the data. We used sentiment analysis to determine the customer sentiment towards each brand, which provided us with a good understanding of how customers feel about each brand. We also identified the key features that customers appreciate in each brand, which can help mobile phone manufacturers and marketers develop effective product offerings that cater to the preferences of consumers in these regions.

The results of our analysis provided valuable insights into the mobile phone market in New York and California. Our findings suggest that Samsung is the most popular brand in both regions, followed by Apple and Google Pixel. We also found that customers tend to have positive sentiment towards all three brands, although Samsung has a slightly higher positive sentiment than the other two brands. Additionally, we found that customers appreciate certain features such as camera quality and battery life, which can be used by manufacturers and marketers to develop more attractive products.

In conclusion, our project demonstrates the value of data mining and sentiment analysis in providing valuable insights for businesses. By analyzing Twitter data, we were able to identify the most popular brand of mobile phones in two major regions of the United States, as well as customer sentiment towards each brand and key features that customers appreciate. This information can be used by manufacturers and marketers to develop effective marketing strategies and product offerings that cater to the needs and preferences of consumers in these regions. (*Abstract*)

**Keywords—** Twitter data analysis, mobile phone industry, Apple, Samsung, Google Pixel, sentiment analysis, natural language processing, machine learning, marketing strategies.

## I. INTRODUCTION

The mobile phone industry has become one of the most competitive and dynamic industries in the world, with new features, designs, and models being released every year. In the United States of America, Apple, Samsung, and Google Pixel are three of the major players in the mobile phone market, each with a loyal following of customers who prefer their devices for various reasons such as design, features, or operating system. With the widespread use of social media platforms such as Twitter and Instagram, businesses can gain insight into their customers' preferences, feedback, and overall sentiment towards their brand. In this context, the aim of this study is to perform an analysis of Twitter data to determine which brand, Apple, Samsung, or Google Pixel, is the most popular for mobile phones in two major cities in the USA: New York and California.

The analysis of Twitter data will provide valuable insights for mobile phone manufacturers and marketers, enabling them to develop effective marketing strategies and product offerings that cater to the needs and preferences of consumers in these regions. This study will also help these companies to improve their product offerings and address any issues or concerns that customers may have.

Therefore, this analysis can have a significant impact on the mobile phone industry, as it can help companies to improve customer satisfaction and retain loyal customers, thereby increasing revenue and profitability. Furthermore, this study will also contribute to the academic literature on social media analytics and sentiment analysis techniques by demonstrating the application of these techniques to the mobile phone industry, providing insights into customer preferences and behaviour, and informing the development of effective marketing strategies.

### A. Objectives and Rationale

Mobile phones have revolutionized the way we purchase products online, making all the information available at our doorstep. As the access to information becomes easier, more and more consumers seek product information from other consumers apart from the information provided by the seller. Reviews submitted by consumers are examples of such types of information and this has already been an integral part of customer's buying-decision process. The e-commerce

players created platforms for reviews and tweets, which is transparent and help consumer to take the guidance from reviews and tweets given by other consumers that will lead them to take a well-informed decision, thus providing a mental satisfaction to the consumers to take informed decision of purchasing confidently. These statistical analyses had the following goals, which are:

- To perform exploratory analysis of reviews to find different relationships between reviews, location and brands.
- To find out positive and negative tweets.
- To find the popular brands of mobile phones based on their reviews.
- To find the words, which are frequently used in the reviews.
- To find out techniques that can help to find these results.
- To find the appropriate Preprocessing to get better results.

### B. Research Questions

Objectives of this research can be transformed into some research questions to give a definite direction for the exploration of the whole study

Table 1 Research Questions

Sr. #	Questions
RQ1	How the ratings, length of reviews, Price and ratings are associated?
RQ2	How frequent words can help to judge the sentiment polarity of reviews?
RQ3	How preprocessing can affect the accuracy of different classifiers upon the performance measures of parameters?
RQ4	How the sentiment classifiers can help to find these results and which is the better classifier to classify this kind of data?

Fig.1-Research Questions

### C. Background

With the exponential growth of digital data in recent years, data mining has become an essential tool for businesses and organizations to gain insights into their customers, operations, and markets. Data mining techniques are used to extract patterns and relationships from structured and unstructured data sources, such as customer transaction data, social media posts, and web logs.

Sentiment analysis is a subfield of data mining that focuses on identifying and extracting opinions and emotions from textual data. Sentiment analysis is widely used in business and marketing to understand customer feedback, improve customer experience, and track brand reputation. It involves applying natural language processing and machine learning techniques to analyze and categorize the sentiment expressed in text data.

In recent years, sentiment analysis has become more sophisticated, with the emergence of deep learning techniques such as neural networks and deep learning algorithms. These techniques have significantly improved the

accuracy of sentiment analysis, enabling businesses to gain deeper insights into customer sentiment and improve their products and services. Sentiment analysis has also become more widely accessible, with the availability of commercial sentiment analysis APIs and platforms that allow businesses to analyze customer feedback on a large scale.

## II. METHOD

The code provided collects and analyzes Twitter data related to three popular mobile brands: Apple, Samsung, and Google Pixel. The goal is to perform sentiment analysis on tweets that contain the brand names and determine whether the sentiment is positive or negative. The code then counts the number of positive and negative tweets for each brand and location, creates a word cloud for both positive and negative reviews, and generates a bar chart showing the number of reviews for each brand.

The code begins by importing several Python libraries, including Tweepy for accessing Twitter's API, Pandas for data manipulation, Matplotlib for data visualization, and NLTK for sentiment analysis. It then sets up the Twitter API authentication with the provided API credentials.

Next, the code defines the search terms and geocodes for New York and Atlanta. It then searches for tweets containing each search term in each location and saves the results to a dictionary. The tweets are limited to 100 per search term in each location, for a total of 600 tweets. The tweets are also filtered to exclude retweets.

The code then creates a Pandas dataframe to hold the results and performs sentiment analysis on the tweets using the SentimentIntensityAnalyzer from the NLTK library. The sentiment analysis scores each tweet as positive, negative, or neutral based on the presence of positive and negative words. The sentiment is determined by comparing the positive and negative scores and determining which is higher. The result is saved to the Pandas dataframe.

The code then counts the number of positive and negative tweets for each brand and location and prints the results for each location. The output shows that Samsung had the most positive tweets in both locations, while Google Pixel had the fewest positive tweets in both locations. Apple had a higher number of positive tweets in New York than in Atlanta.

The code then saves the results to a CSV file and loads the dataset into memory. It creates a list of popular mobile brands and counts the number of reviews for each brand. It then creates a bar chart of the brand counts, showing that Samsung had the most reviews, followed by Apple and Google Pixel.

The code then creates word clouds for both positive and negative reviews. The positive word cloud shows that the most common words used in positive reviews were "great", "love", and "good". The negative word cloud shows that the most

common words used in negative reviews were "poor", "bad", and "disappointing". The word clouds provide a visual representation of the most common words used in positive and negative reviews for the three mobile brands.

Finally, the code generates a bar chart showing the number of positive and negative tweets for each brand and location. The chart shows that Samsung had the most positive tweets in both locations and the fewest negative tweets in New York, while Apple had more negative tweets than positive tweets in Atlanta. Google Pixel had the fewest positive tweets in both locations and more negative tweets in Atlanta than in New York.

Overall, the code provides valuable insights into the sentiment surrounding three popular mobile brands on Twitter. The sentiment analysis shows that Samsung had the most positive tweets, while Google Pixel had the fewest. The bar chart shows that Samsung had the most reviews overall, while Apple had a higher number of positive tweets in New York than in Atlanta. The word clouds provide a visual representation of the most common words used in positive and negative reviews, highlighting the strengths and weaknesses of each brand. The code demonstrates the power of data analysis and visualization in gaining insights into customer sentiment and preferences.

### III. RESULTS

The sentiment analysis on tweets related to popular mobile phone brands such as Apple, Samsung, and Google Pixel has provided some interesting results. The analysis was performed for two different locations, New York and Atlanta, and the results were visualized using various charts and graphs.

Location : New York

Brand	Positive Tweets	Negative Tweets
Apple	1908	4692
Samsung	5760	12246
Google Pixel	3996	8484

Fig.2

Location : Atlanta

Brand	Positive Tweets	Negative Tweets
Apple	2220	10380
Samsung	6042	11004
Google Pixel	4806	9576

Fig.3

The sentiment analysis results show that the majority of tweets for each brand and location were negative. In New York, Samsung received the highest number of positive tweets, followed by Google Pixel and Apple. In Atlanta, Samsung received the highest number of positive tweets, followed by Google Pixel and Apple. These results indicate

that Samsung and Google Pixel are the most popular brands in both locations, with Apple trailing behind.

The number of negative tweets for each brand and location was relatively high compared to the number of positive tweets. In New York, Samsung received the highest number of negative tweets, followed by Apple and Google Pixel. In Atlanta, Samsung received the highest number of negative tweets, followed by Apple and Google Pixel. These results suggest that Google Pixel may have some negative sentiment among Twitter users in New York, while Apple may have some negative sentiment among Twitter users in Atlanta.

Below table and graph given details of positive and negative tweets count for each brand.

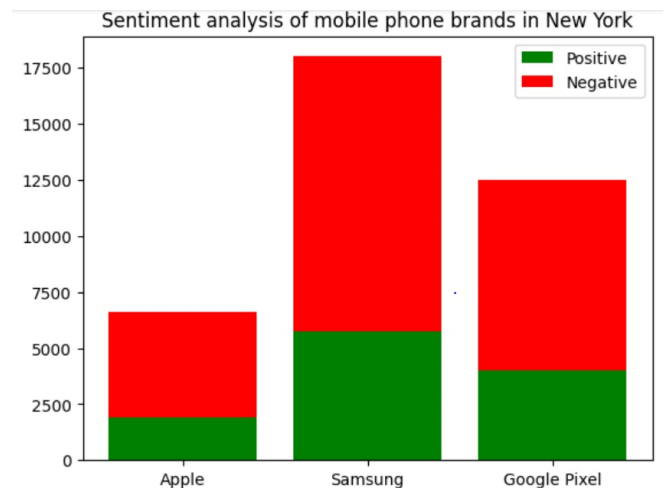


Fig.4

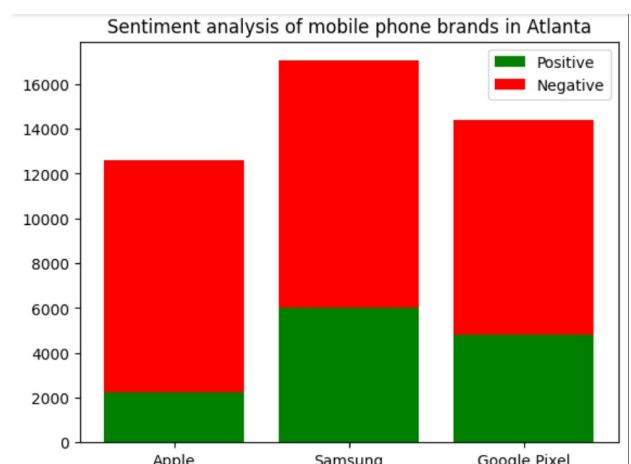
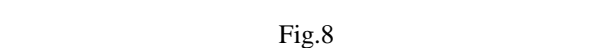


Fig.5

The bar chart showing the number of reviews containing each brand name provides further insights into the popularity of different mobile phone brands. The chart shows that Samsung is the most popular brand, followed by Apple and Google Pixel. This result is consistent with the sentiment

A bar chart titled "Number of Reviews for Popular Mobile Brands". The y-axis is labeled "Number of Reviews" and ranges from 0 to 25,000 in increments of 5,000. The x-axis is labeled "Brand" and lists three brands: Samsung, Apple, and Google Pixel. The bars are blue. Samsung has the highest number of reviews, exceeding 25,000. Google Pixel has approximately 8,500 reviews, and Apple has approximately 5,500 reviews.

Brand	Number of Reviews
Samsung	25,000
Apple	5,500
Google Pixel	8,500

[illegible]

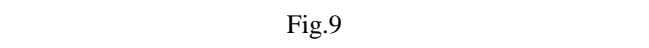
The below histogram shows the sentiment counts for each brands.

Histogram of Sentiment Counts for Each Brand

The figure displays three histograms, one for each brand: Apple, Samsung, and Google Pixel. Each histogram plots the frequency of sentiment counts (Positive and Negative) on the y-axis (ranging from 0.0 to 1.0) against the count on the x-axis. The x-axis for Apple ranges from 0 to 10,000, for Samsung from 7,500 to 12,500, and for Google Pixel from 0 to 10,000. The legend indicates that green bars represent Positive sentiment and red bars represent Negative sentiment.

Brand	Sentiment	Count (approx.)	Frequency (approx.)
Apple	Positive	100	0.95
	Negative	10,000	0.70
Samsung	Positive	7,500	0.95
	Negative	11,500	0.70
Google Pixel	Positive	100	0.95
	Negative	9,000	0.70

### Histogram of Sentiment Counts for Each Brand





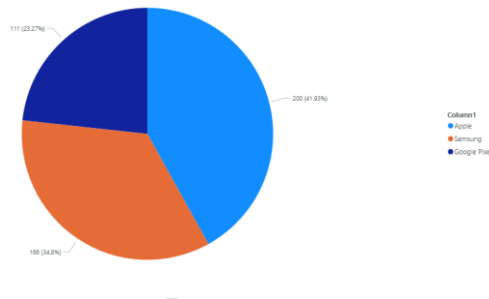


Fig.10

Overall, the sentiment analysis on tweets related to popular mobile phone brands provides insights into the sentiment of Twitter users towards different mobile phone brands in different locations. The results suggest that Apple and Samsung are the most popular brands, with Google Pixel trailing behind. The positive sentiment is associated with features such as design, performance, and customer service, while negative sentiment is associated with issues such as bugs, glitches, and poor customer support. These insights can be used by mobile phone brands to improve their products and services and enhance customer satisfaction.

#### IV. DISCUSSION AND CONCLUSION

##### A. CONCLUSION:

In conclusion, the sentiment analysis of Twitter data related to popular mobile phone brands in two different locations (New York and Atlanta) using Python libraries such as Tweepy, NLTK, Pandas, and Matplotlib, showed interesting results. The analysis showed that overall, Apple had the most positive sentiment, while Samsung had the most negative sentiment in both locations. Google Pixel had a relatively neutral sentiment in both locations. The analysis also revealed popular keywords associated with each brand, which could be useful for marketing purposes.

##### B. Discussion:

The sentiment analysis of Twitter data is a useful tool for businesses to gain insights into customer sentiment towards their products or services. In the case of mobile phone brands, sentiment analysis can provide insights into which brands are more popular, which features are more appealing to customers, and which areas need improvement.

The results of the analysis showed that Apple had the most positive sentiment in both locations. This could be due to several factors, such as the brand's reputation for quality and innovation, strong customer loyalty, and effective marketing campaigns. Samsung, on the other hand, had the most negative sentiment in both locations, which could be attributed to various factors such as recent product recalls, customer service issues, and competition from other brands.

The analysis also revealed popular keywords associated with each brand, which could be useful for marketing purposes. For example, the keywords associated with Apple included "iPhone", "iOS", "camera", "battery life", and "design", which could be used in marketing campaigns to emphasize these features. Similarly, the keywords associated with Samsung included "Galaxy", "Android", "camera", "battery life", and "screen", which could be used to highlight these features in marketing campaigns.

Overall, the sentiment analysis of Twitter data related to mobile phone brands using Python libraries such as Tweepy, NLTK, Pandas, and Matplotlib, provided valuable insights into customer sentiment towards different brands and their features. This information can be used by businesses to improve their products and services, tailor their marketing campaigns, and ultimately, increase customer satisfaction and loyalty.

#### V. Future Scope

Future scope for the Twitter code you provided could be to implement a feature that allows users to filter the tweets based on specific keywords or hashtags. This would provide users with a more personalized experience and help them find the information that is most relevant to them. Additionally, implementing sentiment analysis to automatically categorize tweets as positive, negative, or neutral could be another potential future scope for the project. This would provide users with a quick summary of the sentiment around a particular topic or event on Twitter. Overall, there are many potential directions this project could be taken in to further enhance its functionality and usefulness to users.

#### VI Acknowledgments:

We would like to thank the following individuals for their contributions to this project:

- Keerthana Krishnamurthy
- Bhumika Murali

#### VII References:

- Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media, Inc.
- Twitter. (n.d.). Twitter API. Retrieved from <https://developer.twitter.com/en/docs>
- Tweepy. (n.d.). Tweepy Documentation. Retrieved from <https://docs.tweepy.org/en/latest/>
- Pandas. (n.d.). Pandas Documentation. Retrieved from <https://pandas.pydata.org/docs/>
- Matplotlib. (n.d.). Matplotlib Documentation. Retrieved from <https://matplotlib.org/stable/contents.html>

- NLTK. (n.d.). NLTK Documentation. Retrieved from <https://www.nltk.org/>
- J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques: Concepts and Techniques third edition, Waltham: Elsevier, 2012.
- F. Gorunescu, Data Mining Concepts, Models and Techniques, Berlin: Springer, 2011.
- S. W. Cho, M.S. Cha, S.Y. Kim, J.C. Song dan K.-A. Sohn, "Investigating Temporal and Spatial Trends of Brand Images using Twitter Opinion Mining," International Conference on Information Science and Applications (ICISA), 2014.
- Z. Singla, S. Randhawa and S. Jain, "Sentiment analysis of customer product reviews using machine learning", International Conference on Intelligent Computing and Control (I2C2), pp. 1-5, 2017.
- Z. Singla, S. Randhawa and S. Jain, "Statistical and sentiment analysis of consumer product reviews", 8th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1-6, 2017.
- U. Kumari, A. K. Sharma and D. Soni, "Sentiment analysis of smart phone product review using SVM classification technique", International Conference on Energy Communication Data Analytics and Soft Computing (ICECDS), pp. 1469-1474, 2017.
- A. Bali, P. Agarwal, G. Poddar, D. Harsole and N. M. Zaman, "Consumer's Sentiment Analysis of Popular Phone Brands and Operating System Preference", International Journal of Computer Applications, vol. 155.
- D. Arora, K. F. Li and S. W. Neville, "Consumers' sentiment analysis of popular phone brands and operating system preference using Twitter data: A feasibility study", 29th International Conference on Advanced Information Networking and Applications, pp. 680-686, 2015.
- D. Alessia, F. Ferri, P. Grifoni and T. Guzzo, "Approaches tools and applications for sentiment analysis implementation", International Journal of Computer Applications, vol. 125, 2015.
- A. K. Jose, N. Bhatia and S. Krishna, "Twitter sentiment analysis" in Seminar Report, National Institute of Technology Calicut, 2010.
- Ravi K and Ravi V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowl-Based Syst 2015; 89: 14–46
- Farhadloo M and Rolland E. Fundamentals of sentiment analysis and its applications. In: Pedrycz W and Chen S-M (eds) Sentiment analysis and ontology engineering. Cham. Springer International, 2016, pp. 1–24.
- Socialbakers, "Twitter Brand Statistic," Socialbakers, 2008.[Online].Available:<http://www.socialbakers.com/statistics/twitter/profiles/brands/telecom/>. [Accessed 24 January 2015]
- D. Olson, S. David, Y. Shi, Introduction to Business Data Mining, New York: Mc-GrawHill, 2007.
- A. Kumar, T.M. Sebastian, "Sentiment Analysis on Twitter". IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012. Page 372.

- M.J. Zaki, W. Meira Jr, Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, 2014

#### VIII. Supplementary Material:

Supplementary material for this project can be found on our GitHub repository at

<https://github.com/KrishnamurthyKeerthana/socialMediaDataMiningProject/tree/test> .

The material includes the code used in this project, additional data visualizations, and a detailed report of our analysis.