Market value product summary using Sentiment Analysis

Amritbani Sondhi
Joy Shalom
Sachin Shivashankara
Akshay Manoharan

Social Media Mining

Spring 2018

Introduction

- Perform Sentiment Analysis using data collected by twitter to perceive the summary between any two entities.
- Rate their attributes
  - positive, negative and neutral
  - view the distribution of these attributes to comprehend them in a stronger way
  - geographic map, word cloud

What we do?

- Analyze the sentiments of tweets to compare two major multinational technology companies
- Examples: Samsung and Apple

Packages used

- Tweepy: for accessing twitter data
- Pandas: for using Panda dataframe
- Numpy: for data manipulation
- NLTK: for Sentiment Analysis
- TextBlob: for Natural Language Processing
- Bokeh: for Interactive Data Visualizations with bokeh server

Sentiment Analysis

- sometimes known as Opinion Mining or Emotion AI
- refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information
- *Textblob* python package is used to perform sentiment analysis

Sentiment Analysis

Implementation

- *Textblob* is used to classify the polarity of a tweet as positive, negative or neutral based on conceptualizing the entire text
- a number is assigned to each polarity of tweet.
  - positive tweets are given '1'
  - negative tweets '-1'
  - neutral tweets '0'
- associated each tweet to their polarity and added this information as a column in the data-frame
- an *attribute list* is defined
- *NLTK* package is used
  - to tokenize every tweet
  - to compare if these tokens are in the attribute list
- tweets that don't have any of these attributes are then removed from the data frame

Sentiment Analysis

	Tweet	SA	words	product	location
@evleaks @leon4449 @samsungmobile @samsungmobileus @samsung all that hate on the #notch got your ass copying too much notches. lmfao. i guess your not as innovative as i thought.		-1	[innovation]	samsung	united states
@samsungmobile when you'll release a good phone with replaceable battery? i'm still using samsung galaxy s5. :d big screens, powerful processors but awful battery lifes, heavy powerbanks...		-1	[battery, battery]	samsung	turkey
@samsungmobile facing problem in j7 pro.\nat starting i think it is a area problems but i face network issue in it.\ni call many time call centre and also visit service centre but don't resolve it.\nsamsung only cheats.		1	[service]	samsung	india
@mrwhosetheboss @samsungmobileus @samsunguk @samsungus @samsungmobile @huaweiphones @huaweimobile @huaweimobileuk @huaweideviceusa huawei photos are sharper compared to the samsung,but i wouldn't be surprised though cuz huawei has been making cameras for years as opposed to samsung.\nback in the early 2000s you couldn't find a laptop that didn't have a huawei webcam on it		1	[camera]	samsung	seychelles
all the samsung users are trying,with that shity android os the phone's screen will now break and you will fix it with almost half of the money you bought the phone .🤔 apple forever 🤷🏾‍♀️		-1	[os, display, price]	apple	nigeria
@techninjaspeaks we just need them to get their camera on point with samsung & apple & oneplus might take over		0	[camera]	apple	united states
@sportswalkthe @richeisen the first gen of most of their products are meh, the first ipod blew up, literally. the average iphone has a samsung chip and a sony camera, all they've done recently is bough superior product companies and implement them in apple products.		1	[camera]	apple	united states

Figure: Dataframe retrieved from NLTK

Sankey Diagram

- a visualization used to depict a flow from one set of values to another
- best used to represnt many-to-many mapping between two domains (e.g., samsung and apple)
- used to associate words with their polarity
- efficient way to reproduce the results of sentiment analysis

Sankey Diagram

Implementation

- A complete list of all instances of positive words are available in a dictionary.
- Set operation is performed in order to acquire only a unique set of positive words and pass it as an input to the sankey diagram.
- For linking them, source and target list has to be defined programmatically.
- *Plotly*
  - Used for the implementation, where composing, editing and sharing the visualization is easier
  - It gives an interactive platform which shows data information while hovering

Sankey Diagram

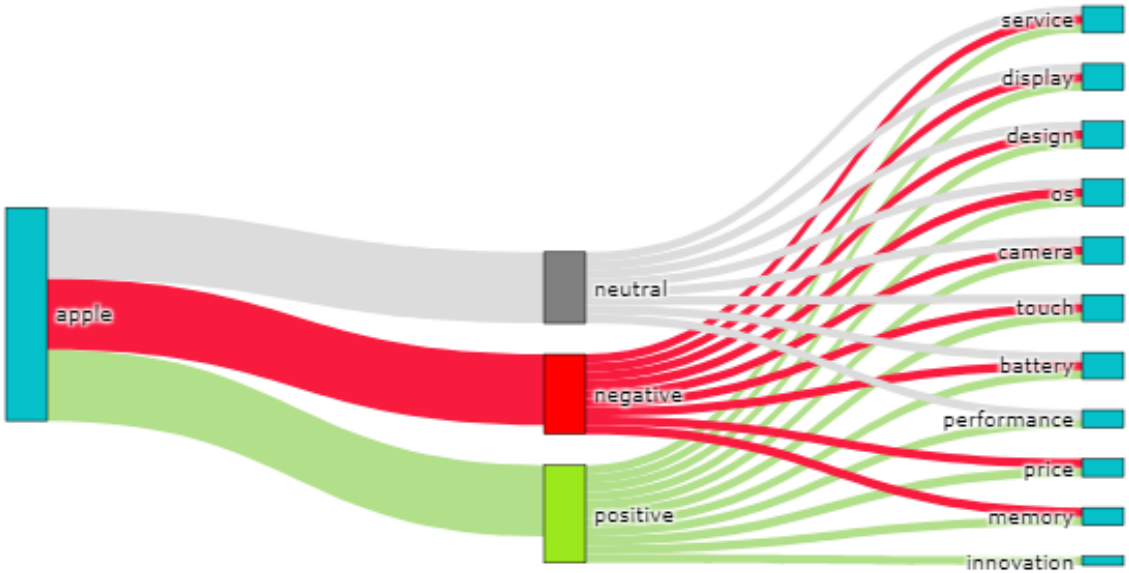

Figure: Sankey Diagram: Apple

Sankey Diagram

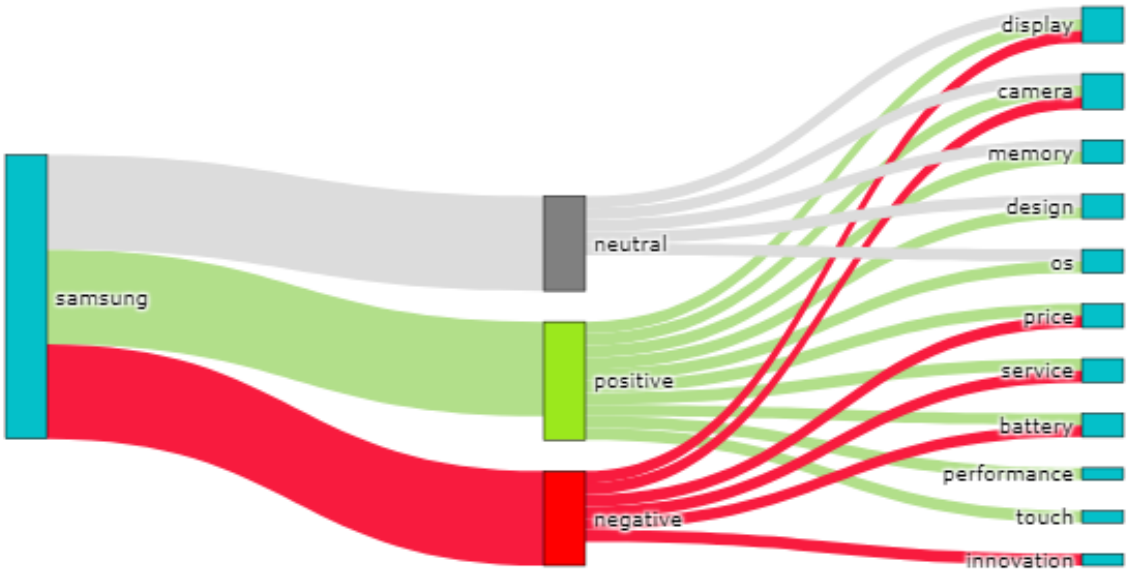

Figure: Sankey Diagram: Samsung

Word Cloud

- Word cloud offers insight on which words are largely associated with a particular brand or entity
- The size of the word represents the frequency of its usage in tweets
- The visualization from the word cloud is easy to interpret and shows how each attribute is compared to other attributes
- It is convenient in associating a set of attributes to a brand

Word Cloud

Implementation

- It is plotted for each sentiment polarity such as positive, negative and neutral for both Samsung and Apple
- A list of attributes for each polarity is given at its input
- Based on the frequency of occurrence of these attributes the size of that attribute is decided by the word cloud, higher the frequency larger the size of the attribute

Word Cloud

Positive Word Clouds

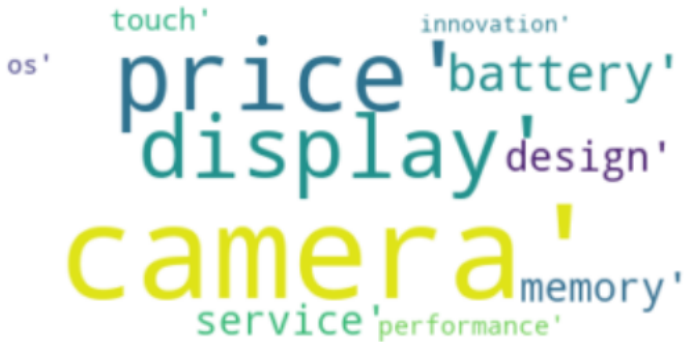

Figure: Positive Word Cloud: Samsung

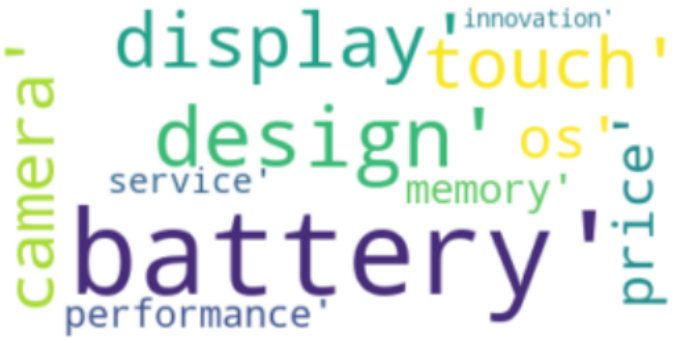

Figure: Positive Word Cloud: Apple

Word Cloud

Negative Word Clouds

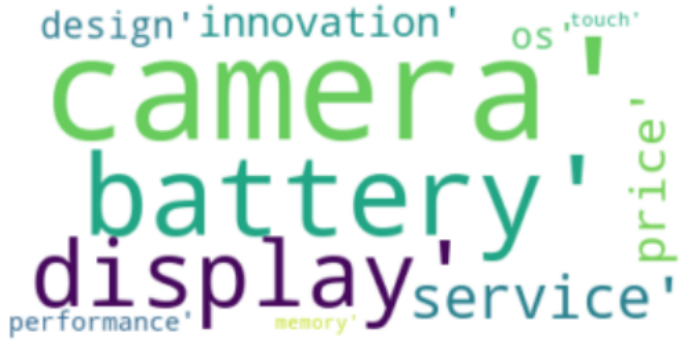

Figure: Sankey Diagram: Samsung

Figure: Sankey Diagram: Apple

Word Cloud

Neutral Word Clouds

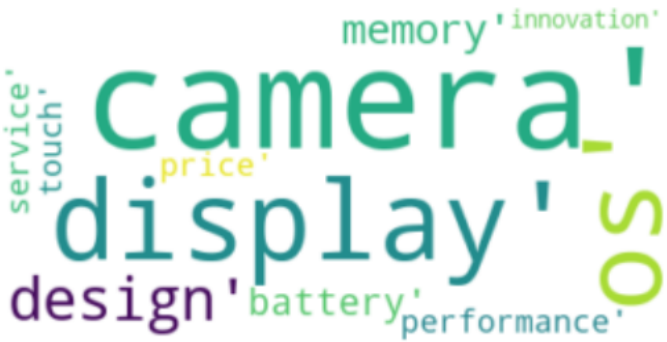

Figure: Neutral Word Cloud: Samsung

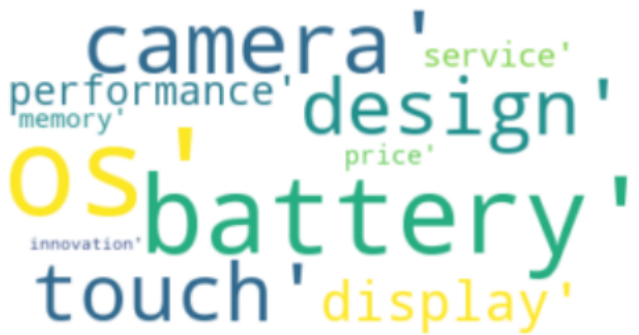

Figure: Neutral Word Cloud: Apple

Choropleth Map

- Choropleth map is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map
- It is used to see what nature of outlook a certain geographical area have on the entities compared
- Tweets are grouped as positive, negative and neutral tweets and are associated with a location
- The geographical areas are then given distinct colors based on the dominating sentiment polarity

Choropleth Map

Implementation

- A .csv file which contains all states and its relevant state code is loaded
- Every country from the initial dataframe is compared with this list
- A new column is added where country code was present
- For every country, the most dominant polarity is chosen based on the count of reviews over the region

Choropleth Map

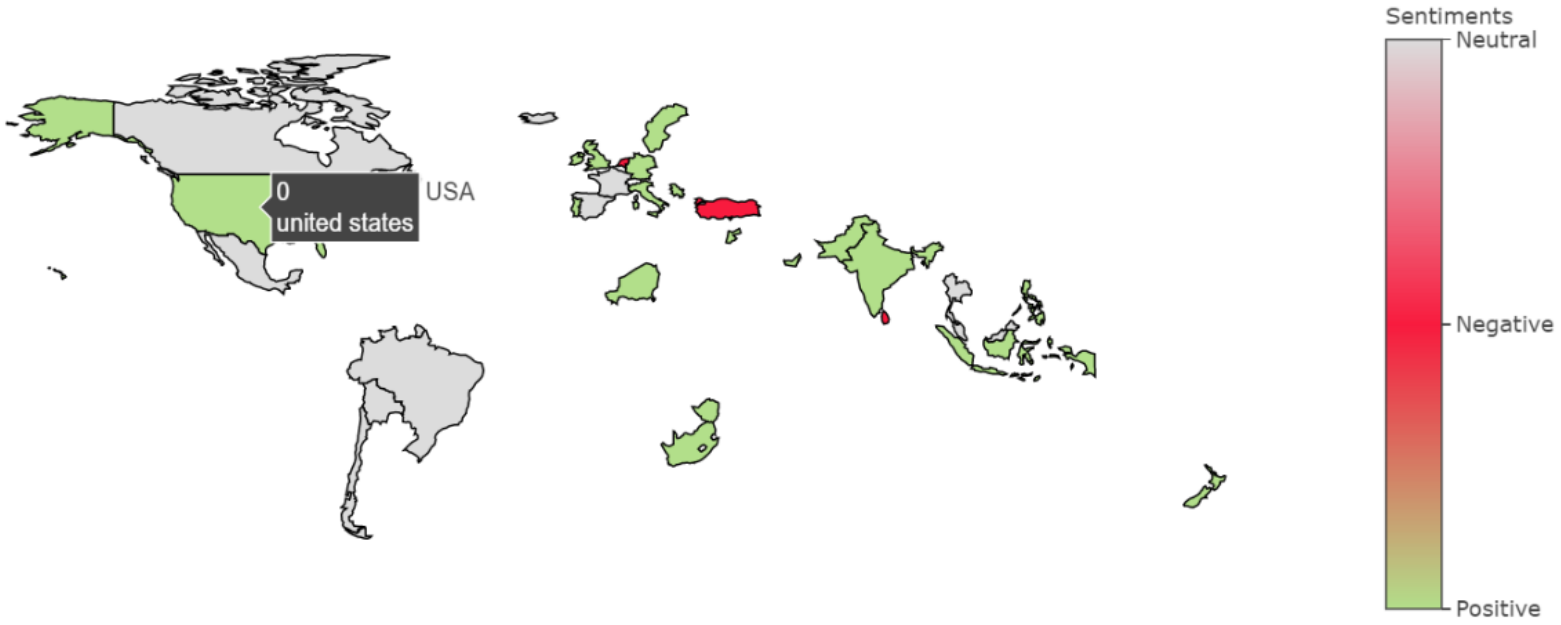

Figure: Choropleth Map: Samsung

Choropleth Map

Figure: Choropleth Map: Apple

User Interactions with the Server

- *Bokeh* is used to create the UI implementation of the project
- It creates model objects in python and converts them into JSON format
- The UI consists of the following widgets:
  - TextHandler: to input the products to compare
  - Button: to generate analytical plots
  - Gridplot: for layouts
- As *plotly* and *matplotlib* plots cannot be converted to bokeh figures and displayed, for better analysis on a button click event an image is plotted on the server and an interactive plot is launched on a new tab of the browser

User Interactions with the Server

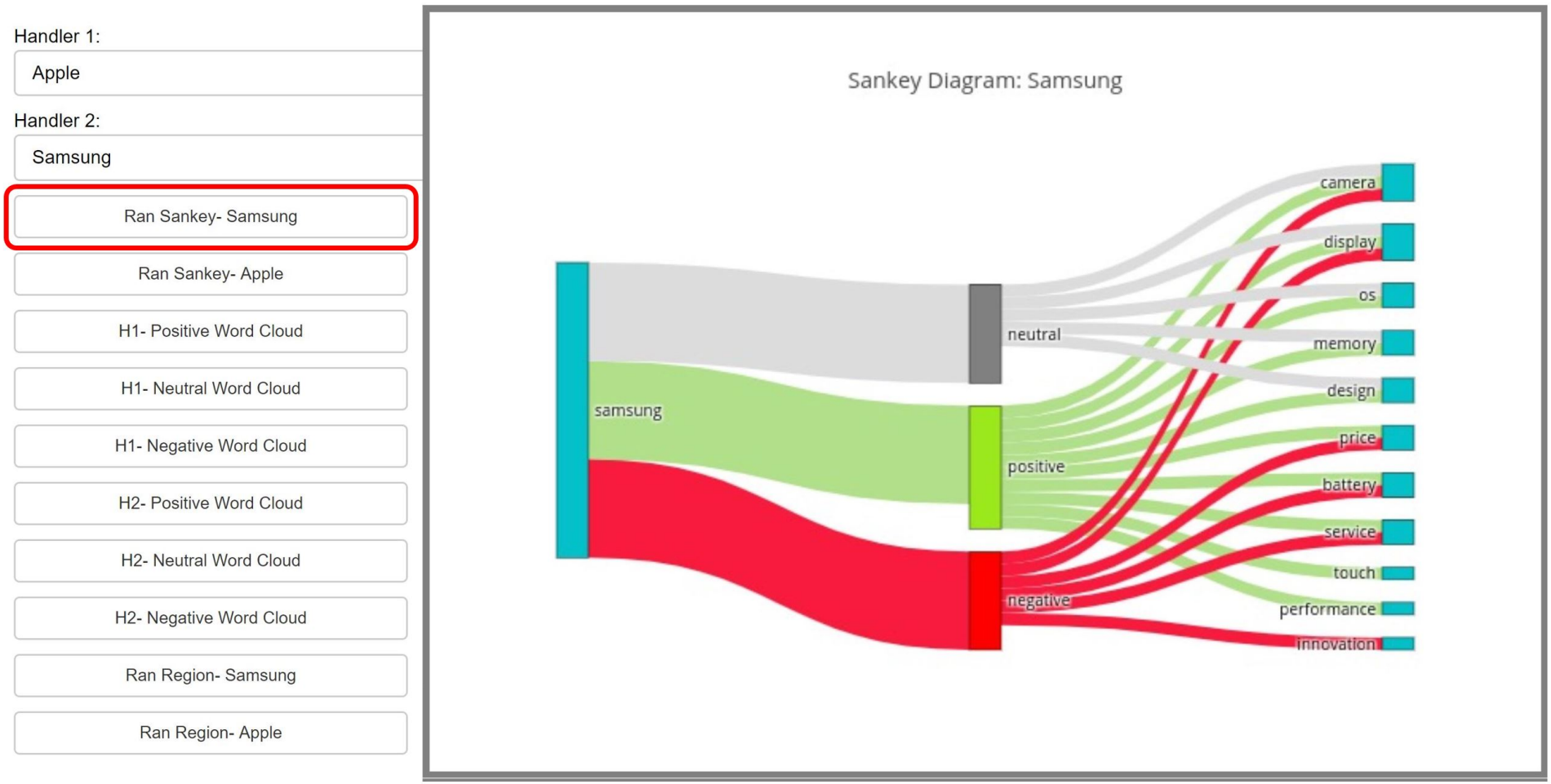

Figure: User Interface

Observations

Samsung

- Positive outlook towards price, design and OS
- Neutral outlook towards display
- Negative outlook towards battery and innovation
- Choropleth map concludes that 'Turkey, Netherlands and Sri Lanka' has more negative outlook compared to other countries

Observations

Apple

- Positive outlook towards design, camera and display
- Neutral outlook towards battery
- Negative outlook towards service
- Choropleth map concludes that ‘Nigeria, Finland, Denmark, Ireland, Saudi Arabia and India” has more negative outlook compared to other countries