**Social Media Analytics on Aerated Drinks**

Hai Ngoc Nguyen – BJ19BDS003

Rithwik Chhugani – BS18BDS008

Gautam Sadarangani – BS18BDS015

**Background and Introduction:**

With the advancement in technology and the need to connect the world, the use of social networks has increased dramatically over the last decade. Platforms such as Facebook, Instagram and Twitter have become popular in helping people get connected and communicate over the internet. With a great number of active users on a daily basis, these platforms have established a great opportunity for commercial brands to connect with their potential customers. By accessing and analyzing consumers’ data collected on these platforms, companies can derive insightful information of not only themselves but also their competitors for the business development.

In this project, we are going to collect and analyze the activities of 3 soft drink brands, namely Coca Cola, Pepsi and Sprite on 3 commonly used social networks, namely Facebook, Twitter and Instagram.

**Problem Statement:**

**Objectives:**

* Understand how to collect data from 3 mentioned platform and the implementation of API if applicable.
* Evaluate the quality of collected data.
* Understand how active each brand is on these platforms.
* Understand the pattern of how posts are distributed.
* Get interaction statistics (like, share, comment, …) of posts on each platform.
* Understand the popularity of each brands based on interaction on posts.
* Understand customers perspective towards brands by analyzing comments.

**The Project:**

**Instagram Part:** <https://drive.google.com/file/d/1DhnRTaId4PnRdN7bbHGm22fgerewD_qE/view?usp=sharing>

**Twitter Part:** <https://drive.google.com/file/d/15gTkUXz9LR08CIbA-TBRK5C6NUF_tf3p/view?usp=sharing>

**Facebook Part:** <https://drive.google.com/drive/folders/1Xp3YmiJ0sYxD1xhveFNn6BDPkPvsf3cF?usp=sharing>

**Main Flow:**

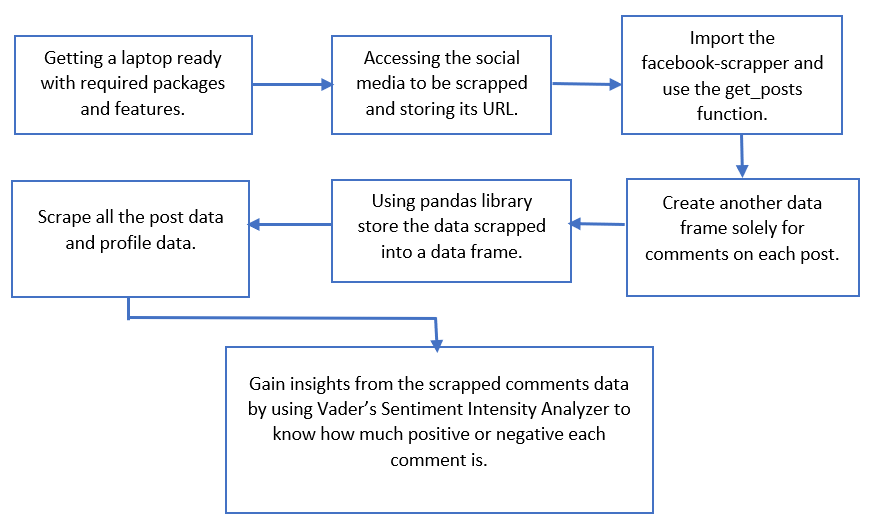
**Instagram:**

**Twitter:**

Diagram

Description automatically generated

**Facebook:**



**The Users:**

* Organizations publishing brand trust analysis and customer engagements reports.
* For individuals to understand the personality of a brand.
* Can be used by government bodies to investigate the effect of tweets on the public by these brands and take actions if needed. This will also help them look deeper into the sort of activities the brand is involved in.
* Competitors can use this technique to understand how well the brand does marketing and ways to keep customers engaged with the brand.
* Since it has geographical information as well you can analyse the presence of the brand in a particular country by specifying the geolocation and the radius. This can help countries to analyse interests of their population.

**Why should we use sentiment analysis?**

* **Invaluable Marketing:**  
  Using sentiment analysis companies and product owners use can use sentiment analysis to know the demand and supply of their products through comments and feedback from the customers.
* **Identifying key emotional triggers:**  
  In psychology and other medical treatment institutions, sentiment analysis can be used to detect whether the individuals’ emotion is normal or abnormal, and based on the data record they can decide person health.
* **Politics:**  
  In the political field, candidates to be elected can use sentiment analysis to predict their political status, to measure people’s acceptance. It can also be used to predict election results for electoral board commissions.
* **Education:**  
  Universities and other higher institutes like colleges can use sentiment analysis to know their student’s feedback and comment, therefore they can take consideration to revise or improve their education curriculum.

**Ethical Issues:**

The following approved and public APIs are used to gather information about the brands:

* [Tweepy](https://docs.tweepy.org/en/latest/api.html)
* [Twint](https://github.com/twintproject/twint)
* Instagramy
* Facebook-scraper

1. **For Instagram:**

* Officially, Instagram strictly prohibits any form of automated data scraping. The risks are:
* Content can be stolen and taken advantage of.
* Sending too many requests can burden the website traffic and possibly crash the site.
* The inner algorithms or data model/structures could be revealed and manipulated.

1. **For Facebook:**

* Facebook-scraper is an external IP not utilizing the Facebook’s well known Graph API hence too much scraping at once could lead to blocking of you IP address.

1. **For Twitter:**

Though scraped by default, the project does not utilise PII (Personally Identifiable Information) to perform network analysis or infer from the social relationship the users or the brand share with others. The project solely processes and draws inference from textual tweets leaving out all images, videos and other content sharing formats.

This project abides by the **Python Web Scrapping Code of Conduct** and DOES NOT,

* Download copies of document that are clearly not public.
* Share downloaded content illegally.
* Share what you cannot.
* Overload the server.
* Scrape the same data repeatedly.
* Violate rules mentioned in the robots.txt file.

**Pre Conditions:**

* Python version 3.7+
* Jupyter notebook installation
* Installing all used packages and dependencies
* Active Twitter developer account
* Authentication strings to fetch data through the API.
* Having all required JSON files downloaded or download using the command given in the notebook.
* Active platforms accounts and developer accounts if applicable.
* Instagram scrape needs session ID
* Active Facebook account to obtain Page ID of the pages to be scrapped.

**Post Conditions:**

* Graphs with different analysis should be printed.
* Summary statistics of different columns where mentioned
* Transformer analysis and sentiment analysis for different aspects should be printed.
* Collected data must be valid.
* Data are organized for easy and quick analysis.
* Insights are drawn with graph and chart presentations.
* Sentiment analysis must give comprehensive insights and results.

**Conclusion:**

From the analysis of 3 brands on three different platforms it is evident that Pepsi is a more consumer interactive brand on all the platforms.

**References:**

<https://pypi.org/project/facebook-scraper/>

<https://www.geeksforgeeks.org/facebook-sentiment-analysis-using-python/>

<https://www.analyticsvidhya.com/blog/2020/11/text-cleaning-nltk-library/>