Name: Janvi D. Marathe **DOP:**

Roll No: 36 **DOS:**

Division: **Score: Sign:**

Experiment No 1: 1 Study various - i) Social Media platforms (Facebook, twitter, YouTube etc) ii) Social Media analytics tools (Facebook insights, google analytics net lyticetc) iii) Social Media Analytics techniques and engagement metrics (page level, post level, member level) iv) Applications of Social media analytics for business.

i)

Social Media:

Social media facilitates the sharing of ideas and information through virtual networks. From Facebook and Instagram to Twitter and YouTube, social media covers a broad universe of apps and platforms that allow users to share content, interact online, and build communities. More than 4.7 billion people use social media, equal to roughly 60% of the world's population.Social media is a digital technology that facilitates the sharing of text and multimedia through virtual networks and communities.

Types of social media:

1. Traditional social networking sites- Most of us are familiar with social networking sites like Facebook, Twitter, LinkedIn, and TikTok. These platforms help us connect with friends, family, and brands. They encourage knowledge-sharing and are all about personal, human-to-human interaction.

A social networking site is a Jill of all trades. Users can share thoughts, curate content, upload photos and videos, form groups based on interests, and participate in lively discussions. They're built around the user and everything that's important to them and their social circles.

2. Social review sites

What's one of the first things you do when planning a trip or buying a new product? If you're anything like us, you'll head straight to the reviews.

Review sites like Yelp and TripAdvisor display reviews from community members for all sorts of locations and experiences. This eliminates a lot of the guesswork that goes into booking a restaurant or hotel. Not sure it's the right thing for you? Check out the reviews and you'll know.

3. Image and video sharing sites

Visual content like images, infographics, and illustrations capture our hearts, eyes and imaginations. Social media platforms like Instagram, Imgur, and Snapchat are designed to amplify the power of image sharing. (Or these days, video sharing.)

Users create, curate, and share unique content that sparks conversation and speaks for itself. A picture or video can be worth a thousand words to your business.

4. Video hosting sites

YouTube revolutionized the way we watch, create, and think about video. It transformed the medium into something accessible. Recent improvements in tech and connectivity helped video go the rest of the way.

Video hosting platforms like YouTube and Vimeo help creators put together content and share it to a platform optimized for streaming. This accessibility makes video a super important medium.

5. Community blogs

Sometimes an image or post isn't complex enough for the message you've got to share, but not everyone on the internet wants to run a blog from a self-hosted website. That's a lot of work.

Shared blogging platforms like Medium and Tumblr give people a space to express their thoughts and help connect them with readers.

These community blog sites provide an audience while allowing plenty of room for customization and self expression.

6. Discussion sites

While most of us have seen many a heated discussion happen on Facebook, discussion sites like Reddit and Quora are specifically designed to spark a conversation.

Anyone is free to ask a question or make a statement, and this attracts people with shared interests and curiosities. However, unlike Facebook and Instagram, users tend to give out less identifiable information. Anonymity is powerful when it comes to people opening up and getting real.

7. Sharing economy networks

Sites like Airbnb and Rover aren't just a cool place to find cheap holiday rentals or a pet sitter. Sharing economy networks bring people who've got something they want to share together with the people who need it.

These communities provide opportunities that won't exist otherwise by pooling resources on a large scale that wouldn't be possible without tech.

ii)

Purpose of Social Media analytics tools:

Social media analytics tools are used to monitor, measure, and analyze data from social media platforms such as Facebook, Twitter, Instagram, and LinkedIn. The purpose of these tools is to help businesses and organizations gain insights into their social media performance, their audience, and their competitors.

Some of the key purposes of social media analytics tools are:

Monitoring social media mentions: These tools help businesses track mentions of their brand or product on social media platforms. By monitoring social media mentions, businesses can respond

to customer inquiries, manage their online reputation, and identify opportunities to engage with their audience.

Measuring social media performance: Social media analytics tools provide businesses with data on their social media engagement, such as likes, shares, and comments. This information helps businesses measure the success of their social media marketing efforts and identify areas for improvement.

Identifying key influencers: Social media analytics tools can identify key influencers in a particular industry or niche. This information helps businesses build relationships with influencers and leverage their audience to reach new customers.

Understanding audience demographics: Social media analytics tools provide data on the demographics of a business's social media audience, including age, gender, and location. This information helps businesses tailor their social media content to their target audience.

Competitive analysis: Social media analytics tools can track the social media performance of a business's competitors. This information helps businesses understand their position in the market and identify opportunities to differentiate themselves from their competitors.

List of Social Media analytics tools with description:

Facebook insights:

Facebook Insights is a free analytics tool provided by Facebook for businesses that have a Facebook Page. It provides businesses with valuable data and insights into the performance of their Facebook Page and the engagement of their audience. Here are some of the key features of Facebook Insights:

Page Performance: Facebook Insights provides businesses with an overview of their Page performance, including data on Page views, likes, reach, and engagement.

Audience Insights: Facebook Insights allows businesses to analyze their audience demographics, such as age, gender, and location. This information helps businesses tailor their content to their target audience.

Post Performance: Facebook Insights provides businesses with data on the performance of their individual posts, including reach, engagement, and reactions.

Video Performance: Facebook Insights provides businesses with data on the performance of their videos, including views, engagement, and retention.

Page Comparisons: Facebook Insights allows businesses to compare their Page performance to that of their competitors, helping them identify opportunities for improvement and differentiation.

Ad Performance: Facebook Insights provides businesses with data on the performance of their Facebook Ads, including impressions, clicks, and conversions.

Overall, Facebook Insights is a powerful tool that helps businesses track their social media performance on Facebook, analyze their audience, and improve their social media strategy.

Google analytics:

Google Analytics is a free web analytics service provided by Google that allows businesses to track and analyze website traffic and user behavior. It provides businesses with valuable insights into their website's performance and user engagement. Here are some of the key features of Google Analytics:

1. Real-Time Analytics: Google Analytics provides real-time data on website traffic, allowing businesses to see how users are interacting with their website in real-time.

2. Audience Insights: Google Analytics allows businesses to analyze their website's audience demographics, such as age, gender, and location. This information helps businesses understand their target audience and tailor their website content to their users' interests.

3. Traffic Sources: Google Analytics provides businesses with data on the sources of their website traffic, including search engines, social media, and referral websites. This information helps businesses understand where their traffic is coming from and how to optimize their marketing efforts.

4. Behavior Analysis: Google Analytics allows businesses to analyze user behavior on their website, including pageviews, time spent on site, bounce rate, and conversion rate. This information helps businesses optimize their website design and improve user engagement.

5. E-commerce Tracking: Google Analytics provides businesses with data on e-commerce transactions, including product sales, revenue, and conversion rates. This information helps businesses understand their online sales performance and optimize their e-commerce strategy.

Overall, Google Analytics is a powerful tool that helps businesses track and analyze their website performance, user behavior, and e-commerce transactions. It provides businesses with valuable insights that can help them optimize their website design, improve user engagement, and drive online sales.

Twitter Analytics

Twitter Analytics is a free analytics tool provided by Twitter for businesses and individuals with a Twitter account. It provides users with valuable data and insights into the performance of their tweets and their audience engagement. Here are some of the key features of Twitter Analytics:

1. Tweet Performance: Twitter Analytics provides businesses with data on the performance of their individual tweets, including impressions, engagement rate, and click-through rate. This information helps businesses identify which tweets are performing well and which ones need improvement.

2. Audience Insights: Twitter Analytics allows businesses to analyze their audience demographics, such as age, gender, location, and interests. This information helps businesses understand their target audience and tailor their tweets to their users' interests.

3. Follower Growth: Twitter Analytics provides businesses with data on their follower growth, including the number of followers gained and lost over time. This information helps businesses track their social media growth and identify areas for improvement.

4. Video Performance: Twitter Analytics provides businesses with data on the performance of their videos on Twitter, including views, engagement, and completion rates.

5. Ad Performance: Twitter Analytics provides businesses with data on the performance of their Twitter Ads, including impressions, clicks, and conversions.

Overall, Twitter Analytics is a valuable tool that helps businesses track their social media performance on Twitter, analyze their audience, and improve their social media strategy.

Netlytic:

Netlytic is a social media analytics tool that helps businesses and researchers analyze large volumes of text data from social media platforms such as Twitter, Facebook, and Instagram. It is a cloud-based software that uses natural language processing (NLP) and machine learning algorithms to provide users with valuable insights into social media conversations. Here are some of the key features of Netlytic:

1. Data Collection: Netlytic allows users to collect data from social media platforms by specifying search terms, hashtags, or usernames. The tool can collect data from multiple sources and in multiple languages.

2. Data Analysis: Netlytic uses NLP and machine learning algorithms to analyze the collected data and identify key themes, topics, and sentiment. The tool provides users with visualizations of the data, including word clouds, topic networks, and sentiment analysis charts.

3. Social Network Analysis: Netlytic allows users to analyze the social network structure of social media conversations, identifying key influencers and communities.

4. Automated Reporting: Netlytic provides users with automated reports on social media conversations, allowing businesses and researchers to quickly and easily analyze large volumes of data.

5. Integration: Netlytic can be integrated with other software and platforms, including Microsoft Excel, Google Sheets, and Slack.

Overall, Netlytic is a powerful tool for businesses and researchers looking to analyze social media conversations and identify key insights. It provides users with valuable data visualizations and automated reporting, making it easy to analyze large volumes of data quickly and efficiently.

Youtube Analytics:

YouTube Analytics is a free analytics tool provided by YouTube for creators and businesses with a YouTube channel. It provides users with valuable data and insights into the performance of their videos and audience engagement. Here are some of the key features of YouTube Analytics:

1. Watch Time: YouTube Analytics provides creators with data on their watch time, including the total number of minutes users have watched their videos. This information helps creators understand how engaging their videos are and how long users are watching their content.

2. Demographics: YouTube Analytics allows creators to analyze their audience demographics, such as age, gender, and location. This information helps creators understand their target audience and tailor their videos to their users' interests.

3. Traffic Sources: YouTube Analytics provides creators with data on the sources of their traffic, including YouTube search, suggested videos, and external websites. This information helps creators understand where their traffic is coming from and how to optimize their video promotion efforts.

4. Engagement: YouTube Analytics provides creators with data on their engagement metrics, including likes, comments, and shares. This information helps creators identify which videos are resonating with their audience and which ones need improvement.

5. Revenue: YouTube Analytics allows creators to analyze their revenue streams, including advertising revenue, merchandise sales, and fan funding. This information helps creators understand their earnings potential and optimize their monetization strategy.

Overall, YouTube Analytics is a valuable tool for creators and businesses looking to track their video performance and audience engagement on YouTube. It provides users with valuable data and insights that can help them optimize their video strategy, improve user engagement, and increase revenue.

iii) Applications of Social media analytics for business

Social media analytics can be incredibly valuable for businesses in a variety of ways. Here are some of the key applications of social media analytics for business:

1. Audience Insights: Social media analytics can provide businesses with valuable insights into their audience, including their demographics, interests, and behaviors. This information can help businesses understand their target audience and tailor their marketing efforts to better reach and engage their customers.

2. Competitive Analysis: Social media analytics can also be used to analyze the social media performance of competitors, identifying areas of opportunity and potential threats in the market.

3. Campaign Performance: Social media analytics can provide businesses with data on the performance of their social media campaigns, including engagement rates, conversion rates, and click-through rates. This information can help businesses optimize their campaigns for maximum impact.

4. Brand Reputation: Social media analytics can help businesses monitor their brand reputation on social media, identifying potential crises and addressing negative feedback before it spreads.

5. Product Development: Social media analytics can provide businesses with insights into customer feedback and preferences, which can be used to inform product development and improve customer satisfaction.

6. Influencer Marketing: Social media analytics can also help businesses identify influencers with relevant audiences and track the performance of influencer campaigns.

Overall, social media analytics can provide businesses with valuable data and insights that can inform marketing strategy, improve customer engagement, and drive business growth.

iv) Describe your problem statement in minimum 250 words

AMUL

Amul is a dairy cooperative in India that produces a wide range of dairy products, including milk, butter, cheese, and ice cream. Social media analytics can be valuable for Amul in a variety of ways, including:

1. Product Promotion: Amul can use social media analytics to identify popular products and customer preferences, allowing them to target their social media promotions more effectively. By analyzing customer feedback and engagement on social media, Amul can identify the products that resonate most with their customers and adjust their marketing strategy accordingly.

2. Brand Reputation: Social media analytics can help Amul monitor their brand reputation on social media, allowing them to quickly identify and address any negative feedback or potential crises. By monitoring social media conversations around their brand, Amul can stay ahead of any issues that may arise and maintain a positive reputation with their customers.

3. Competitor Analysis: Social media analytics can also be used by Amul to monitor the social media performance of their competitors, identifying areas of opportunity and potential threats in the market. By analyzing the social media strategies of their competitors, Amul can adjust their own strategy to stay ahead of the competition.

4. Customer Engagement: Social media analytics can help Amul identify their most engaged customers and tailor their social media communications to better meet their needs. By analyzing customer feedback and engagement on social media, Amul can identify the topics and content that resonate most with their customers and adjust their social media strategy accordingly.

Overall, social media analytics can provide Amul with valuable insights into their customers, competitors, and brand reputation, allowing them to optimize their marketing strategy and improve customer engagement.

Name: Janvi D. Marathe	**DOP:**
Roll No: 36	**DOS:**
Division:	**Score:** **Sign:**

Experiment No 2: Select the social media platforms of your choice (Twitter, Facebook, LinkedIn, YouTube, Web blogs etc) and collect social media data for business.

Different ways of collecting data:
There are several ways of collecting data, including:

1. Surveys: Surveys are a popular method of data collection that involve asking questions to a group of people. Surveys can be conducted in person, over the phone, by mail, or online.

2. Interviews: Interviews involve asking questions of individuals or groups in person, over the phone, or through video conferencing. Interviews can be structured (asking a set of predetermined questions) or unstructured (allowing for more open-ended responses).

3. Focus Groups: Focus groups involve bringing together a group of people to discuss a particular topic or product. The group is moderated by a researcher who asks questions and facilitates discussion.

4. Observation: Observation involves observing people or events in their natural environment. This method can be useful for gathering data on behavior or interactions.

5. Experiments: Experiments involve manipulating one or more variables to observe the effect on an outcome. This method is often used in scientific research.

6. Social Media Monitoring: Social media monitoring involves collecting data from social media platforms like Twitter, Facebook, and Instagram. This method can be used to gather insights on customer opinions, preferences, and behaviors.

7. Web Analytics: Web analytics involve collecting data on website traffic, user behavior, and engagement. This method can be used to gather insights on customer behavior and improve website performance.

Overall, the choice of data collection method depends on the research question, the target population, and the resources available. Each method has its own advantages and limitations, and researchers must carefully consider which method is most appropriate for their study.

Web Scraping:

Web scraping is a technique used to extract data from websites. It involves writing software code to automate the process of gathering data from a website, typically by extracting information from HTML pages. Web scraping can be used to extract a variety of data, such as text, images, and other media.

There are several tools and libraries available for web scraping, including Beautiful Soup, Scrapy, and Selenium. These tools allow users to automate the process of data extraction, making it faster and more efficient than manual methods.

However, it's important to note that web scraping can sometimes be in violation of website terms of service and can potentially infringe on copyright laws. As a result, it's important to ensure that web scraping is done ethically and legally. This typically involves obtaining permission from website owners before scraping their data and ensuring that the data is used in compliance with applicable laws and regulations.

Use of Web Scraping:

Web scraping has many potential applications across various industries, including:

1. Market research: Web scraping can be used to gather data on consumer preferences, behavior, and opinions, allowing businesses to make more informed decisions about product development, marketing, and pricing strategies.

2. Competitive analysis: Web scraping can be used to gather data on competitor products, pricing, and marketing strategies, allowing businesses to stay ahead of the competition.

3. Lead generation: Web scraping can be used to gather data on potential customers, such as their contact information and interests, allowing businesses to generate leads for sales and marketing purposes.

4. Content creation: Web scraping can be used to gather data on industry trends, news, and other relevant information, allowing businesses to create content that is timely and relevant to their audience.

5. Academic research: Web scraping can be used in academic research to gather data on a variety of topics, such as social media sentiment, political opinions, and public health trends.

Overall, web scraping can be a powerful tool for gathering data and gaining insights into a wide range of topics. However, it's important to ensure that web scraping is done ethically and legally, and in compliance with applicable laws and regulations.

Different Categories Of Web Scraper:
There are several categories of web scrapers, including:

1. HTML Parser/Web Scraping Libraries: These are software libraries that provide a set of tools to parse and extract data from HTML pages. Popular examples include Beautiful Soup, lxml, and PyQuery.

2. Headless Browsers: These are automated web browsers that can navigate web pages and extract data from them. Examples include PhantomJS and Puppeteer.

3. Web Crawlers: These are programs that automatically traverse the web and collect data from multiple pages. Examples include Scrapy, Heritrix, and Nutch.

4. Visual Scrapers: These are graphical user interface tools that allow users to interact with web pages and extract data using a point-and-click interface. Examples include Parsehub, Octoparse, and Data Miner.

5. APIs: Some websites offer APIs (Application Programming Interfaces) that allow developers to programmatically access and extract data from the website.

6. Custom Web Scrapers: These are custom-built web scrapers that are designed to extract specific data from a particular website or set of websites.

Overall, the choice of web scraper depends on the specific needs and requirements of the project, as well as the skills and expertise of the developer. It's important to choose a scraper that is appropriate for the task at hand and that can effectively and efficiently extract the required data.

Web Scraping Tools And Features:
There are many web scraping tools available that offer a wide range of features. Some of the most popular web scraping tools and their features include:

1. Beautiful Soup: A Python library that is widely used for web scraping. It offers features like parsing HTML and XML, navigating the parse tree, and searching for specific elements on a webpage.

2. Scrapy: A Python-based web crawling framework that allows users to extract data from websites with ease. It offers features like support for handling cookies and sessions, support for multiple web crawlers, and automatic throttling and retrying of requests.

3. Selenium: A browser automation tool that allows users to automate browser actions, such as clicking buttons and filling in forms. It offers features like support for multiple programming languages, support for different browsers, and the ability to interact with web elements using JavaScript.

4. ParseHub: A visual web scraping tool that allows users to create scraping projects using a point-and-click interface. It offers features like automatic pagination, automatic click handling, and automatic data extraction.

5. Octoparse: Another visual web scraping tool that allows users to extract data from websites without coding. It offers features like support for multiple pages, support for dynamic web pages, and automatic data extraction.

6. Apify: A platform that allows users to build web scrapers and automate workflows. It offers features like a web-based editor, support for multiple programming languages, and built-in data storage and management.

Overall, the choice of web scraping tool depends on the specific needs and requirements of the project, as well as the skills and expertise of the developer. It's important to choose a tool that is appropriate for the task at hand and that can effectively and efficiently extract the required data.

Legal and Ethical Considerations In Web Scraping:

Web scraping can be a powerful tool for gathering data, but it's important to be aware of the legal and ethical considerations involved. Some key legal and ethical considerations in web scraping include:

1. Terms of Service: Websites may have terms of service that prohibit web scraping or restrict the use of their data. It's important to review these terms of service carefully and ensure that the web scraping is done in compliance with them.

2. Copyright and Intellectual Property: The data that is scraped may be protected by copyright or other intellectual property laws. It's important to ensure that the web scraping does not infringe on any copyrights or intellectual property rights.

3. Privacy: Web scraping may involve the collection of personal data, such as names and email addresses. It's important to ensure that the web scraping does not violate any privacy laws or regulations.

4. Data Security: Web scraping may involve the collection of sensitive data, such as financial information or login credentials. It's important to ensure that the web scraping is done securely and that the data is stored and transmitted in a secure manner.

5. Fair Use: In some cases, web scraping may be considered fair use under copyright law, particularly if the data is used for research or educational purposes. However, it's important to consult with a legal professional to ensure that the web scraping is done in compliance with fair use guidelines.

Overall, it's important to be aware of the legal and ethical considerations involved in web scraping and to ensure that the web scraping is done in compliance with applicable laws and regulations. It's also important to be transparent about the web scraping and to obtain consent from users if necessary.

Data Sources For Your Problem Statement:

1. **Google search results**

#	Title	URL	Description	Search term searchQuery.term	Domain searchQuery.domain	Page searchQuery.page	Total results resultsTotal
1	GCMMF :: Amul - The Taste of India | Bread Spread Producers ...	https://amul.com/	Amul MilkButterBread SpreadsCheesePaneerDahiCheese SauceBeverage RangeAmul Protein ProductsIce CreamGheeMilk...	amul	google.com	1	36000000
2	Amul - Wikipedia	https://en.wikipedia.org/wiki/...	Anand Milk Union Limited (abbreviated as Amul) is an Indian state government-owned dairy-based cooperative society officiall...	amul	google.com	1	36000000
3	Amul.coop (@Amul_Coop) · Twitter	https://twitter.com/Amul_Coo...		amul	google.com	1	36000000
4	Amul - Home	https://www.facebook.com/a...	Amul. 1627545 likes · 90736 talking about this. Amul is the flagship brand of GCMMF, India's largest food products marketing...	amul	google.com	1	36000000
5	Amul - The Taste of India (@amul_india) ...	https://www.instagram.com/a...	#Amul Topical: Apple's first store in India opens in Mumbai! Watch Chef Monika Manji Patel prepare Achari Toasts Dhokla &...	amul	google.com	1	36000000

2. **YOUTUBE**

#	Title	URL	Number of views viewCount	Date	Number of likes likes	Channel name channelName	Channel URL channelUrl	Number of subscribers numberOfSubscribers	duration
1	Chef Meenu Vig, Episode 5158 #SimpleHomemadeRecipes Facebook Live	https://www.youtube.com/wat...	60	15 hours ago	undefined	Amul TV	https://www.youtube.com/cha...	496000	14:36
2	Chef Ruchi Saxena, Episode 5157 #SimpleHomemadeRecipes Facebook Live	https://www.youtube.com/wat...	72	16 hours ago	undefined	Amul TV	https://www.youtube.com/cha...	496000	56 minutes
3	Chef Shweta Sharma, Episode 5156 #SimpleHomemadeRecipes Facebook Live	https://www.youtube.com/wat...	100	19 hours ago	undefined	Amul TV	https://www.youtube.com/cha...	496000	6:16
4	Chef Jamna Somani, Episode 5155 #SimpleHomemadeRecipes Facebook Live	https://www.youtube.com/wat...	106	20 hours ago	undefined	Amul TV	https://www.youtube.com/cha...	496000	5:15
5	Chef Arya Kolarkar, Episode 5154 #SimpleHomemadeRecipes Facebook Live	https://www.youtube.com/wat...	103	1 day ago	undefined	Amul TV	https://www.youtube.com/cha...	496000	22 minutes
6	Chef Swagata Guha, Episode 5153 #SimpleHomemadeRecipes Facebook Live	https://www.youtube.com/wat...	59	1 day ago	undefined	Amul TV	https://www.youtube.com/cha...	496000	29 minutes

Data Description:

Raw Data in csv format

Sample Collected Data:

```
import requests
from bs4 import BeautifulSoup


url = 'https://amul.com/'
reqs = requests.get(url)
soup = BeautifulSoup(reqs.text, 'html.parser')

urls = []
for link in soup.find_all('a'):
    print(link.get('href'))
```

/products/freshcream.php
/products/mithai-mate.php
/products/happy_treats.php
/products/amul-pro.php
/products/amul-cookies.php
/products/peanut-spread.php
/products/amul-puffles.php
/products/mithai-range.php
/products/roti_softener.php
/products/amul-panchamrit-info.php
/products/sour.php
/products/cattle-feed.php
/m/recipes
/m/gcmmf
/m/about-us
/files/pdf/Branch_Wise-GSTN.pdf
/m/integrated-management-system-policy
/files/pdf/annual_sustainability_report_2704.pdf
https://idadairyconference.com/brochure/
/m/organisation
/m/locate-us
/m/48th-annual-general-body-meeting
/m/awards
/m/exports

Date	HS Code	Description	Destination	Port of Loading	Unit	Quantity	Value (INR)	Per Unit (INR)
Nov 22 2016	04012000	AMUL UHT MILK (TAAZA) 12X1 LTR TRAY (1490 TRAY) (17880 LITRES) FAT:3.3% QC NO :263960 DT.21.11.2016	Philippines	Sabarmati ICD	KGS	18,452	994,753	54
Nov 22 2016	04059020	AMUL GHEE 12X1 LTR TIN (50 CARTON) QC NO: 263934 DT.21.11.2016	Philippines	Sabarmati ICD	KGS	543	221,095	407
Nov 22 2016	04039010	AMUL BUTTERMILK 200G	Chile	Tughlakabad	PAC	162	3,846	24
Nov 21 2016	04052000	MALAI PANEER CUBE (AMUL) 1 KG	Singapore	Banglore Air Cargo	KGS	408	78,950	194
Nov 21 2016	22029090	AMUL COOL KESAR DRINK (200ML X 30PC)	Togo	Kanakpura - Jaipur ICD	NOS	150	2,966	20
Nov 21 2016	22029090	AMUL KOOL COCO DRINK (200ML X 30PC)	Togo	Kanakpura - Jaipur ICD	NOS	150	4,448	30
Nov 21 2016	04059020	AMUL GHEE (1LTR X 12PC)	Togo	Kanakpura - Jaipur ICD	NOS	240	100,432	418

Name: Janvi D. Marathe	**DOP:**
Roll No:36	**DOS:**
Division:	**Score:** **Sign:**

Experiment No 3: Perform data cleaning (preprocess, filter)

What is data cleaning?

Data cleaning — also known as data cleansing or data scrubbing — is the process of modifying or removing data that's inaccurate, duplicate, incomplete, incorrectly formatted, or corrupted within a dataset.

While deleting data is part of the process, the ultimate goal of data cleaning is to make a dataset as accurate as possible. This might require fixing spelling and syntax errors, identifying and deleting duplicate data points, correcting mistakes like mislabelled or empty fields, and standardizing how data is entered or combined from multiple sources.

Why is data cleaning important?

Cleaning data is important because it will ensure you have data of the highest quality. This will not only prevent errors — it will prevent customer and employee frustration, increase productivity, and improve data analysis and decision-making.

This makes sense. Without cleaning data first, the dataset is more likely to be inaccurate, unorganized, and incomplete. Any data analysis will therefore be more difficult, less clear, and less accurate — and so will the decisions based on that data analysis.

Social media contains different types of data: information about user profiles, statistics (number of likes or number of followers), verbatims, and other media content. Quantitative data is very convenient for an analysis using statistical and numerical methods, but unstructured data such as user comments is much more challenging. On social media, unstructured data is related to text, images, videos, and sound and we will mostly deal with textual data. Then, the data has to be cleaned and normalized. Only after all these steps can we delve into the analysis.

Social media Data type and encoding

Comments and conversation are textual data that we retrieve as strings. In brief, a string is a sequence of characters represented by code points. Every string in Python is seen as a Unicode covering the numbers from 0 through 0x10FFFF (1,114,111 decimal). Then, the sequence has to be represented as a set of bytes (values from 0 to 255) in memory. The rules for translating a Unicode string into a sequence of bytes are called encoding. Encoding plays a very important role in natural language processing because people use more and more characters such as emojis or emoticons, which replace whole words and express emotions. Moreover, in many languages, there are accents that go beyond the regular English alphabet. In order to deal with all the processing problems that might be caused by these, we have to use the right encoding, because comparing two strings with different encodings is actually like comparing apples and oranges. The most common one is UTF-8, used by default in Python 3, which can handle any type of character. As a rule of thumb, always normalize your data to Unicode UTF-8.

Pre-processing and text normalization

Preprocessing is one of the most important parts of the analysis process. It reformats the unstructured data into uniform, standardized form. The characters, words, and sentences identified at this stage are the fundamental units passed to all further processing stages. The quality of the preprocessing has a big impact on the final result on the whole process.

There are several stages of the process: from simple text cleaning by removing white spaces, punctuation, HTML tags and special characters up to more sophisticated normalization techniques such as tokenization, stemming or lemmatization. In general, the main aim is to keep all the characters and words that are important for the analysis and, at the same time, get rid of all others, and the text corpus should be maintained in one uniform format.

```
import re, itertools

import nltk

from nltk.corpus import stopwords
```

When dealing with raw text, we usually have a set of words including many details we are not interested in, such as whitespace, line breaks, and blank lines. Moreover, many words contain capital letters so programming languages misinterpret for example, “go” and “Go” as two different words. In order to handle such distinctions, we can convert all words to lowercase format with the following steps:

1. Perform basic text mining cleaning.
2. Remove all whitespaces:

```
verbatim = verbatim.strip()
```

Many text processing tasks can be done via pattern matching. We can find words containing a character and replace it with another one or just remove it. Regular expressions give us a powerful and flexible method for describing the character patterns we are interested in. They are commonly used in cleaning punctuation, HTML tags, and URLs paths.

3. Remove punctuation:

```
verbatim = re.sub(r'[^ws]','',verbatim)
```

4. Remove HTML tags:

```
verbatim = re.sub('<[^<]+?>', '', verbatim)
```

5. Remove URLs:

```
verbatim = re.sub(r'^https?://.*[rn]*', '', verbatim, flags=re.MULTILINE)
```

Depending on the quality of the text corpus, sometimes there is a need to implement some corrections. This refers to the text sources such as Twitter or forums, where emotions can play a role and the comments contain multiple letters words for example, “happppppy” instead of “happy”

6. Standardize words (remove multiple letters):

```
verbatim = ''.join(''.join(s)[:2] for _, s initertools.groupby(verbatim))
```

After removal of punctuation or white spaces, words can be attached. This happens especially when deleting the periods at the end of the sentences. The corpus might look like: “the brown dog is lostEverybody is looking for him”. So there is a need to split “lostEverybody” into two separate words.

7. Split attached words:

```
verbatim = ' '.join(re.findall('[A-Z][^A-Z]*', verbatim))
```

Stop words are basically a set of commonly used words in any language: mainly determiners, prepositions, and coordinating conjunctions. By removing the words that are very commonly used in a given language, we can focus only on the important words instead, and improve the accuracy of the text processing.

8. Convert text to lowercase, lower():

```
verbatim = verbatim.lower()
```

9. Stop word removal:

```
verbatim = ' '.join([word for word in verbatim.split() if word not in (stopwords.words('english'))])
```

10. Stemming and lemmatization: The main aim of stemming and lemmatization is to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form. Stemming reduces word forms to so-called stems, whereas lemmatization reduces word forms to linguistically valid lemmas. Some examples of stemming are cars -> car, men -> man, and went -> Go

Duplicate removal: Depending on the data source we might notice multiple duplicates in our dataset. The decision to remove duplicates should be based on the understanding of the domain. In most cases, duplicates come from errors in the data collection process and it is recommended to remove them in order to reduce bias in our analysis.

Tokenization: Tokenization is the process of breaking a text corpus up into words (most commonly), phrases, or other meaningful elements, which are then called tokens. The tokens become the basic units for further text processing.

Grammar checking: Grammar checking is mostly learning-based, a huge amount of proper text data is learned, and models are created for the purpose of grammar correction. There are many online tools that are available for grammar correction purposes. This is a very tricky cleaning technique because language style and structure can change from source to source (for example language on Twitter will not correspond with the language from published books). Wrongly correcting grammar can have negative effects on the analysis.

Spelling correction: In natural language, misspelled errors are encountered. Companies, such as Google and Microsoft have achieved a decent accuracy level in automated spell correction. One can use algorithms such as the Levenshtein Distances, Dictionary Lookup, and so on, or other modules and packages to fix these errors. Again take spell correction with a grain of salt, because false positives can affect the results.

Data filtering

Data filtering is the process of choosing a smaller part of your data set and using that subset for viewing or analysis. Filtering is generally (but not always) temporary – the complete data set is kept, but only part of it is used for the calculation.

Filtering may be used to:

- Look at results for a particular period of time.
- Calculate results for particular groups of interest.
- Exclude erroneous or "bad" observations from an analysis.
- Train and validate statistical models.

Filtering requires you to specify a rule or logic to identify the cases you want to include in your analysis. Filtering can also be referred to as “subsetting” data, or a data “drill-down”. In this article we illustrate a filtered data set and discuss how you might use filtering.

Example of filtering

The table below shows some of the rows of a data set from a survey about peoples’ preferred Cola. The survey data contains demographic information about the respondents as well as each person’s preferred cola and that person’s rating (out of 5) for each of six varieties of cola.

				Brand Ratings					
ID	Age	Gender	Preferred cola	Coca-Cola	Diet Coke	Coke Zero	Pepsi	Diet Pepsi	Pepsi Max
1	25 to 29	Female	Pepsi Max	2	5	2	3	1	4
2	45 to 49	Male	Pepsi Max	5	1	5	5	3	4
3	25 to 29	Female	Diet Coke	5	4	2	3	1	1
4	25 to 29	Female	Coca-Cola	4	2	2	2	2	2
5	55 to 64	Female	Diet Coke	3	4	3	3	4	2
6	55 to 64	Female	Diet Pepsi	3	3	3	3	4	4
7	50 to 54	Female	Coke Zero	2	3	5	2	2	2
8	35 to 39	Female	Coca-Cola	4	2	5	3	2	5
9	65 or more	Male	Diet Pepsi	5	5	3	5	5	3
10	45 to 49	Female	Coke Zero	4	4	4	5	5	3
11	45 to 49	Male	Coca-Cola	4	1	1	4	1	1
12	55 to 64	Male	Coca-Cola	5	2	2	5	2	2
13	55 to 64	Male	Coca-Cola	5	2	2	3	2	2
14	30 to 34	Male	Pepsi Max	3	2	5	3	3	5
15	65 or more	Female	Diet Pepsi	2	4	2	5	4	2

Filtering this data involves:

1. Coming up with a rule for the observations needed.
2. Selecting the observations that fit the rule.
3. Conducting the analysis using only the information contained in those selected observations.

For example, the table below shows the data filtered for Males only. The darker colored rows are kept in the analysis while the remaining rows are excluded. Results computed for Males are then calculated based on the highlighted rows (ID's 2, 9, 11, 12, 13, 14). If we want to know the average rating for Coca-Cola among males, we would compute that as (5 + 5 + 4 + 5 + 5 + 3) / 6 = 4.5.

				Brand Ratings					
ID	Age	Gender	Preferred cola	Coca-Cola	Diet Coke	Coke Zero	Pepsi	Diet Pepsi	Pepsi Max
1	25 to 29	Female	Pepsi Max	2	5	2	3	1	4
2	**45 to 49**	**Male**	**Pepsi Max**	**5**	**1**	**5**	**5**	**3**	**4**
3	25 to 29	Female	Diet Coke	5	4	2	3	1	1
4	25 to 29	Female	Coca-Cola	4	2	2	2	2	2
5	55 to 64	Female	Diet Coke	3	4	3	3	4	2
6	55 to 64	Female	Diet Pepsi	3	3	3	3	4	4
7	50 to 54	Female	Coke Zero	2	3	5	2	2	2
8	35 to 39	Female	Coca-Cola	4	2	5	3	2	5
9	**65 or more**	**Male**	**Diet Pepsi**	**5**	**5**	**3**	**5**	**5**	**3**
10	45 to 49	Female	Coke Zero	4	4	4	5	5	3
11	**45 to 49**	**Male**	**Coca-Cola**	**4**	**1**	**1**	**4**	**1**	**1**
12	**55 to 64**	**Male**	**Coca-Cola**	**5**	**2**	**2**	**5**	**2**	**2**
13	**55 to 64**	**Male**	**Coca-Cola**	**5**	**2**	**2**	**3**	**2**	**2**
14	**30 to 34**	**Male**	**Pepsi Max**	**3**	**2**	**5**	**3**	**3**	**5**
15	65 or more	Female	Diet Pepsi	2	4	2	5	4	2

Results for different groups

A basic need for most research is to obtain results for different groups in the data. One may want to ask about the prevalence of a disease within a demographic segment of the overall population, understand sales figures for the past 3 months, or view feedback given by customers who gave

your restaurant 1 star on Yelp. In each case, a logical rule defines whether each case in the sample is excluded or included.

From the example above, we may wish to compute the average rating for each beverage within for the Males in the sample. Such filtering transforms the results like this:

Unfiltered

	Average	Sample size
Coca-Cola	3.73	15
Diet Coke	2.93	15
Coke Zero	3.07	15
Pepsi	3.60	15
Diet Pepsi	2.73	15
Pepsi Max	2.80	15

Filtered to Males

	Average	Sample size
Coca-Cola	4.50	6
Diet Coke	2.17	6
Coke Zero	3.00	6
Pepsi	4.17	6
Diet Pepsi	2.67	6
Pepsi Max	2.83	6

Sometimes filtering is carried out implicitly. For example, in survey research, the columns of a *crosstab* correspond to a special case of filtering, where filtered results are computed separately for each column, and the results are displayed side-by-side.

One reason for filtering data is to remove observations that may contain errors or are undesirable for analysis. For example, you may want to remove respondents who did not complete the survey, respondents who raced through the survey and selected answers without paying attention to what they were answering ("speeders"), or cases where data entered manually has been entered with mistakes. In other areas of research, a multivariate technique may only be applicable to cases where there is complete information for all the variables that were measured, and so a filter may be constructed to remove cases where some observations are missing.

Snapshot of Data Cleaning Process

```
import numpy as np #linear algebra
import pandas as pd # a data processing and CSV I/O library

import warnings # current version of seaborn generates a bunch of warnings that will be ignore
warnings.filterwarnings('ignore')

# Data Visualization
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(style='white', color_codes=True)
data = pd.read_csv('amul.csv')
data.head(10)
```

	Date	HS Code	Description	Destination	Port of Loading	Unit	Quantity	Value (INR)	Per Unit (INR)
0	Nov222016	4012000.0	AMUL UHT MILK (TAAZA) 12X1 LTR TRAY (1490 TRAY...	Philippines	Sabarmati ICD	KGS	18,452	994,753	54
1	Nov222016	4059020.0	AMUL GHEE 12X1 LTR TIN (50 CARTON) QC NO: 263...	Philippines	Sabarmati ICD	KGS	543	221,095	407
2	Nov222016	4039010.0	AMUL BUTTERMILK 200G	Chile	Tughlakabad	PAC	162	3,846	24
3	Nov212016	4052000.0	MALAI PANEER CUBE (AMUL) 1 KG	Singapore	Banglore Air Cargo	KGS	408	78,950	194
4	Nov212016	22029090.0	AMUL COOL KESAR DRINK (200ML X 30PC)	Togo	Kanakpura - Jaipur ICD	NOS	150	2,966	20
5	Nov212016	22029090.0	AMUL KOOL COCO DRINK (200ML X 30PC)	Togo	Kanakpura - Jaipur ICD	NOS	150	4,448	30
6	Nov212016	4059020.0	AMUL GHEE (1LTR X 12PC)	Togo	Kanakpura - Jaipur ICD	NOS	240	100,432	418
7	Nov212016	4052000.0	MALAI PANEER (AMUL) 500GM	Singapore	Banglore Air Cargo	KGS	204	39,756	195
8	Nov212016	21069019.0	AMUL MASTI - 1LTR - WE INTEND TO CLAIM REWARDS...	Singapore	Chennai Sea	CTN	100	76,969	770
9	Nov212016	19019090.0	ALL OTHER PRPD FLR MEAL CONSMR. RETAILPKD AMUL...	Nigeria	Tughlakabad	PCS	600	270,145	450

Tokenization

```
import nltk
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
True
```

```
[22] from nltk.tokenize import word_tokenize
tokens=word_tokenize(data.Description[0])
print(tokens)
```

```
['AMUL', 'UHT', 'MILK', '(', 'TAAZA', ')', '12X1', 'LTR', 'TRAY', '(', '1490', 'TRAY', ')', '(', '17880', 'LITRES', ')', 'FAT:3.3', '%', 'QC', 'NO', ':263960', 'DT.21.11.2016']
```

Convert to Lowercase

```
data['Destination'] = data['Destination'].str.lower()
print(data['Destination'])
```

```
0          philippines
1          philippines
2                chile
3            singapore
4                 togo
5                 togo
6                 togo
7            singapore
8            singapore
9              nigeria
10                 NaN
11                togo
12                togo
13                togo
14  united arab emirates
15                oman
16           singapore
17              israel
18             vietnam
19             vietnam
20                oman
Name: Destination, dtype: object
```

Snapshot of Data Filtering Process

```
data[['Description','Destination']]
```

	Description	Destination
0	AMUL UHT MILK (TAAZA) 12X1 LTR TRAY (1490 TRAY...	Philippines
1	AMUL GHEE 12X1 LTR TIN (50 CARTON) QC NO: 263...	Philippines
2	AMUL BUTTERMILK 200G	Chile
3	MALAI PANEER CUBE (AMUL) 1 KG	Singapore
4	AMUL COOL KESAR DRINK (200ML X 30PC)	Togo
5	AMUL KOOL COCO DRINK (200ML X 30PC)	Togo
6	AMUL GHEE (1LTR X 12PC)	Togo
7	MALAI PANEER (AMUL) 500GM	Singapore
8	AMUL MASTI - 1LTR - WE INTEND TO CLAIM REWARDS...	Singapore
9	ALL OTHER PRPD FLR MEAL CONSMR. RETAILPKD AMUL...	Nigeria
10	NaN	NaN
11	AMUL COOL STRAWBERRY DRINK (200ML X 30PC)	Togo
12	AMUL COOL BADAM DRINK (200ML X 30PC)	Togo
13	AMUL KOOL CAFE DRINK (200ML X 30PC)	Togo
14	AMUL MALAI PANEER 40 X 200 GM	United Arab Emirates
15	AMUL MALAI PANEER PACE 12 X 1 KG	Oman

Name: Janvi D. Marathe **DOP:**

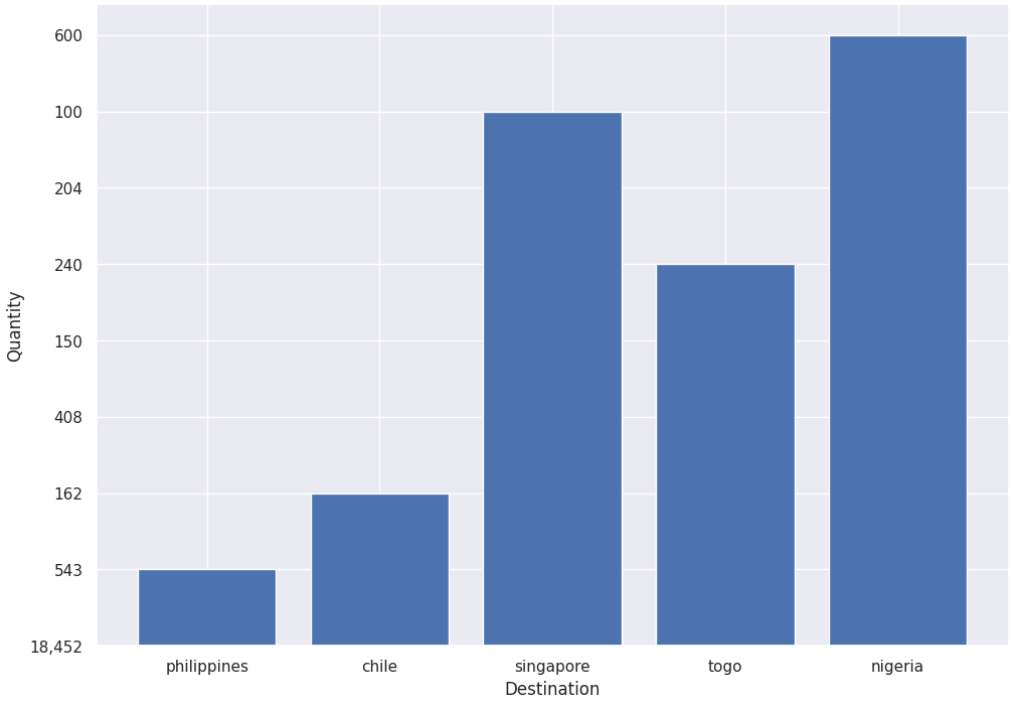
Roll No: 36 **DOS:**

Division: **Score:** **Sign:**

Experiment No 4: Perform exploratory data analysis and visualization of social media data for business.

Experiment can be directly printed from your IPython Notebook

```
from matplotlib import pyplot as plt
plt.xlabel("Destination")
plt.ylabel("Quantity")
name = data['Destination'].head(12)
price = data['Quantity'].head(12)
plt.bar(name[0:10], price[0:10])
```



```
freq={}
for i in data['Description']:
  if type(i) != str:
    continue
  for j in i.split(" "):
    if j not in freq.keys():
      freq[j]=1
    else:
      freq[j]=freq[j]+1
```

```
from wordcloud import WordCloud, STOPWORDS


# Create and generate a word cloud image:
wordcloud = WordCloud(width = 800, height = 800,
                background_color ='white',
                min_font_size = 10).generate_from_frequencies(freq)

# Display the generated image:
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

Name: Janvi D. Marathe **DOP:**

Roll No: **DOS:**

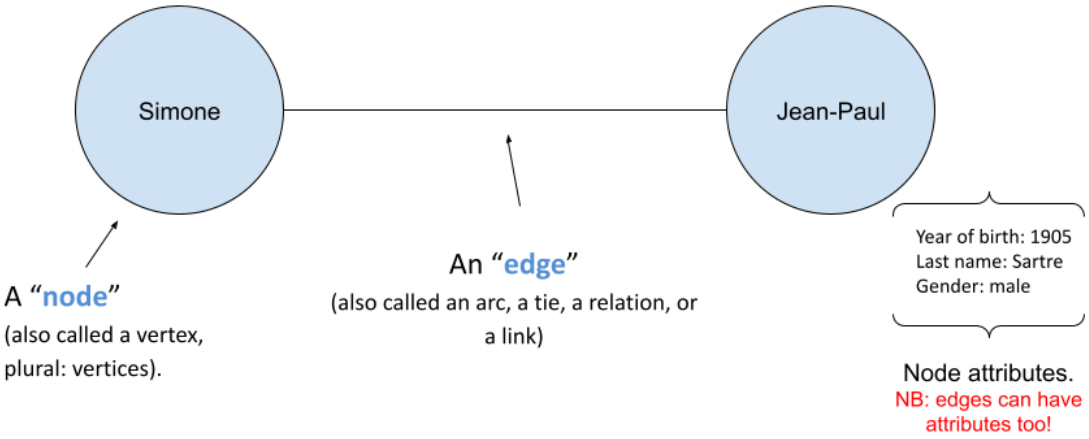
Division: **Score:** **Sign:**

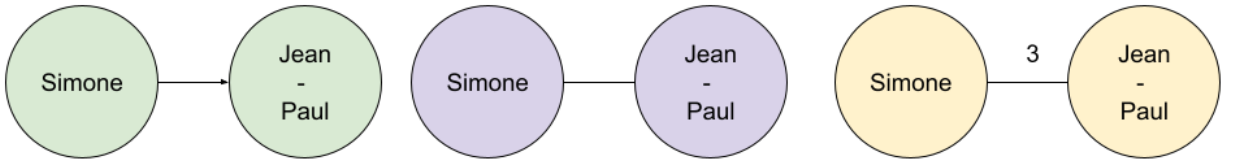
Experiment No 5: Perform social network data analytics to identify social media influencers for business

Social Network :

Social network data analytics can be a powerful tool for identifying social media influencers who can help businesses promote their products and services to a wider audience. Here are some steps businesses can take to identify social media influencers using social network data analytics:

1. Identify relevant social media platforms: Depending on the business and its target audience, certain social media platforms may be more effective than others. For example, Instagram may be more effective for fashion and beauty brands, while LinkedIn may be more effective for B2B companies. Identify the social media platforms where your target audience is most active.
2. Define the criteria for influencers: Determine the criteria for what makes an influencer relevant to your business. This could include factors such as the number of followers, engagement rates, content quality, and relevance to your brand.
3. Use social media listening tools: Use social media listening tools to monitor conversations on social media and identify influencers who are already talking about your brand or industry. These tools can help you track social media mentions, hashtags, and keywords related to your business.
4. Analyze social media data: Use social media analytics tools to analyze data on influencer activity, including the number of followers, engagement rates, and content performance. This data can help you identify influencers who have a strong following and are likely to be effective in promoting your brand.
5. Reach out to influencers: Once you have identified potential influencers, reach out to them to discuss potential partnerships or collaborations. Be sure to provide clear guidelines and expectations, and work with influencers to create content that aligns with your brand values and messaging.



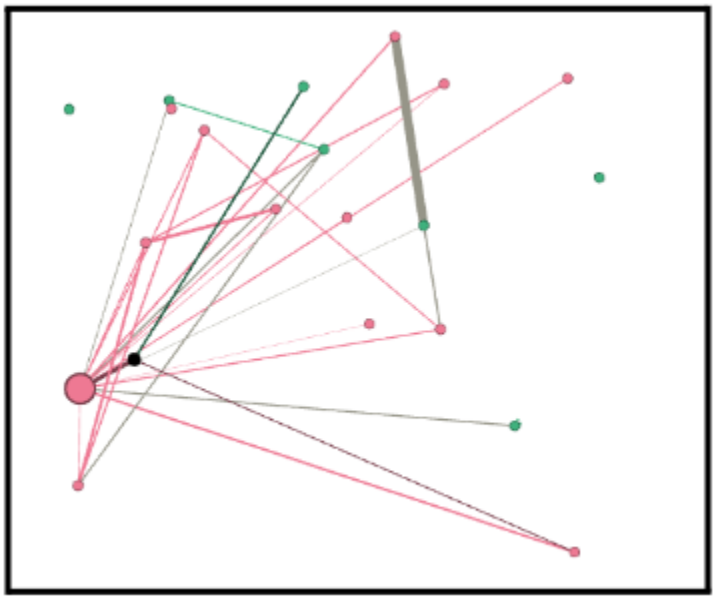


A directed network
(the direction of the relation matters).
Here, the arrow might mean "Simone writes a letter to Jean-Paul".

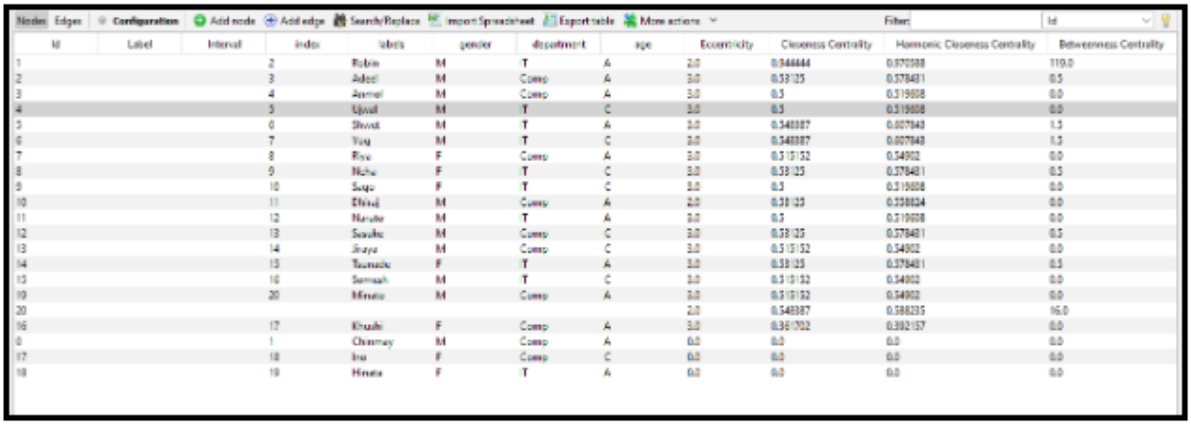
An undirected network
(the direction of the relation does not matter)
Here, the relation might mean "Simone and Jean-Paul are friends"

A weighted network
(the edges have a "strength" represented by a numerical value).
Here, the relation might mean "Simone and Jean-Paul have co-authored 3 books"

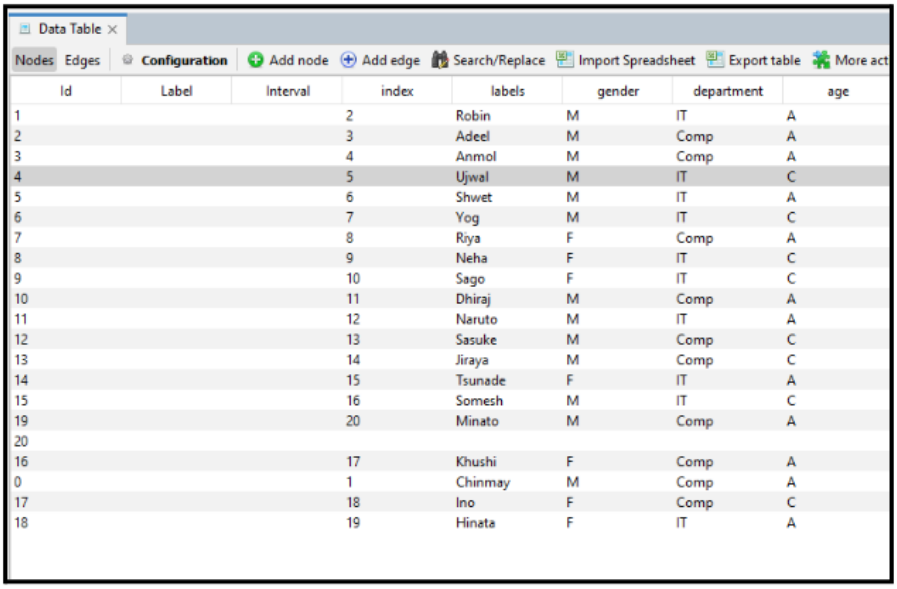
Graph Layouts:



Graph Parameters:



Snapshot of Nodes Table :

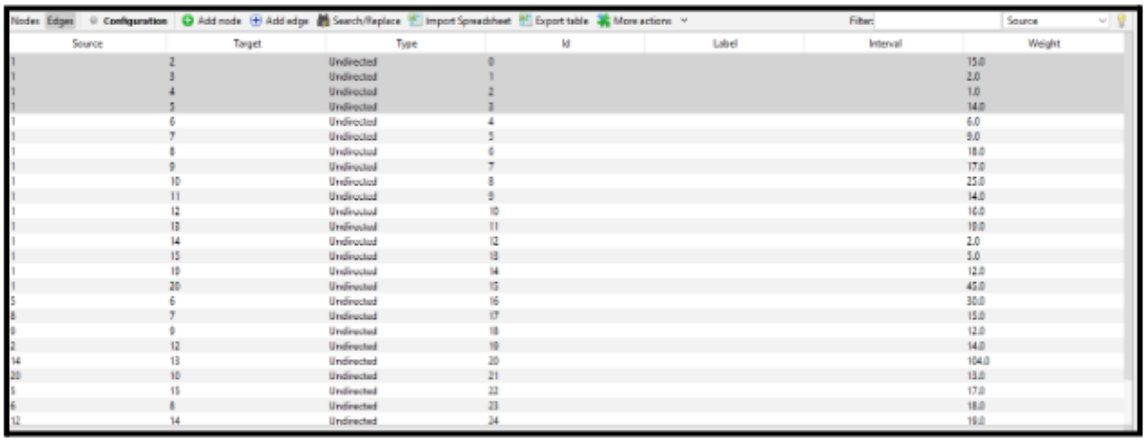


Data Table

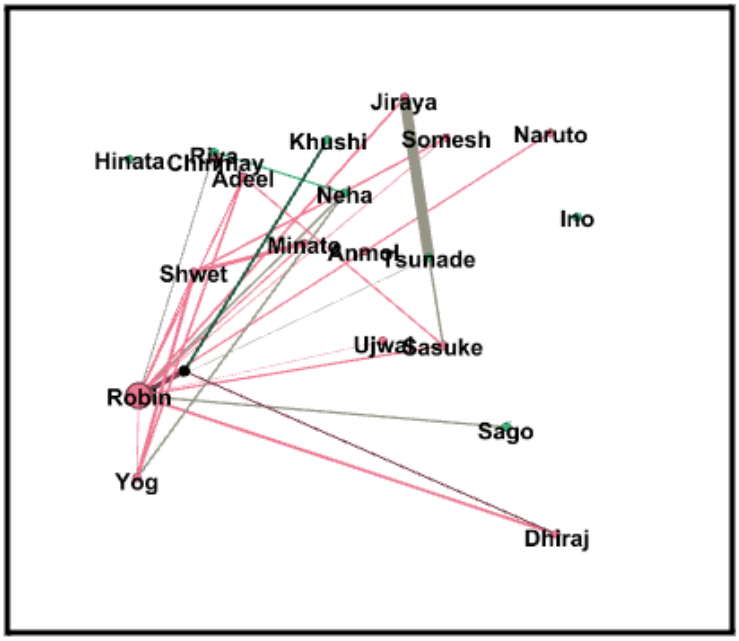
Nodes | Edges | Configuration | Add node | Add edge | Search/Replace | Import Spreadsheet | Export table | More act

Id	Label	Interval	index	labels	gender	department	age
1			2	Robin	M	IT	A
2			3	Adeel	M	Comp	A
3			4	Anmol	M	Comp	A
4			5	Ujwal	M	IT	C
5			6	Shwet	M	IT	A
6			7	Yog	M	IT	C
7			8	Riya	F	Comp	A
8			9	Neha	F	IT	C
9			10	Sago	F	IT	C
10			11	Dhiraj	M	Comp	A
11			12	Naruto	M	IT	A
12			13	Sasuke	M	Comp	C
13			14	Jiraya	M	Comp	C
14			15	Tsunade	F	IT	A
15			16	Somesh	M	IT	C
19			20	Minato	M	Comp	A
20							
16			17	Khushi	F	Comp	A
0			1	Chinmay	M	Comp	A
17			18	Ino	F	Comp	C
18			19	Hinata	F	IT	A

Snapshot of Edge Table :



Snapshot of Graph with most influential Node:



Name: Janvi D. Marathe **DOP:**

Roll No: 36 **DOS:**

Division: **Score:** **Sign:**

Experiment No 6: Perform content based analysis (topic, issue, trend, sentiment/opinion analysis, image) of social media data for business.

For topic and issue and sentiment analysis refer code file

Issue Analysis

```
[ ]  from nltk. corpus import stopwords
     stop = stopwords.words('english')
     new_tweet_df['renderedContent'].apply(lambda x: [item for item in x if item not in stop])
     new_tweet_df.shape
```

```
(13571, 62)
```

```
[ ]  #Remove unnecessary characters
     punct  =  ['%','/',':','\\','&amp','&',';','?']

     def remove_punctuations(text):
       for punctuation in punct:
         text = text.replace(punctuation,'')
       return text
```

```
[ ]  new_tweet_df['renderedContent'] = new_tweet_df['renderedContent'].apply(lambda x: remove_punctuations(x))
```

```
new_tweet_df = new_tweet_df.reset_index(drop=True)
new_tweet_df.head()
```

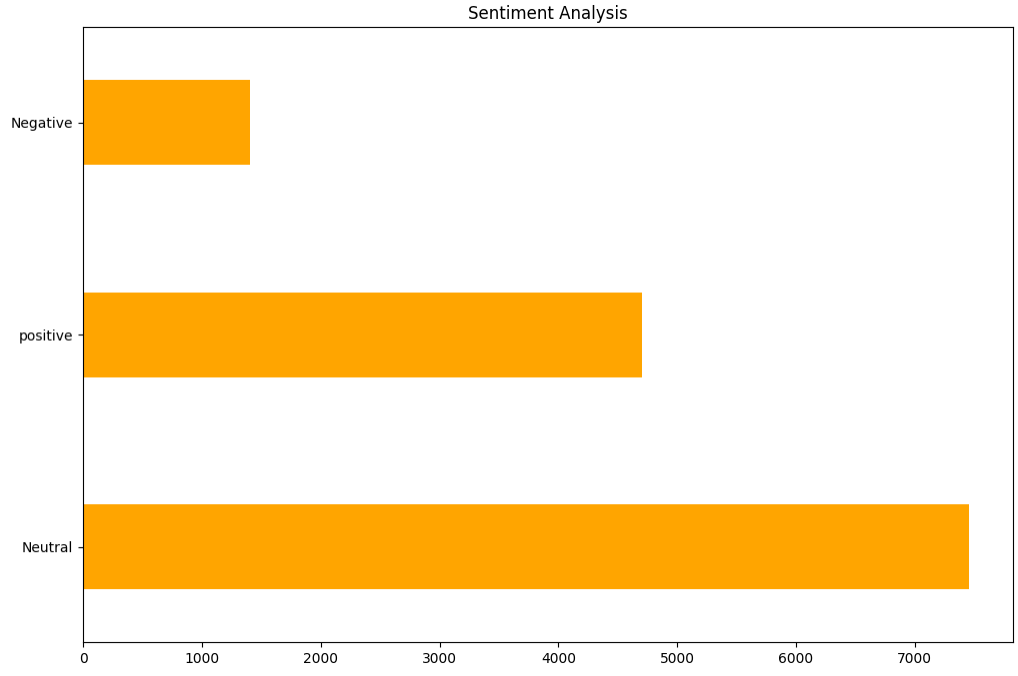
	_type	url	date	rawContent	renderedContent	id	user	replyCou
0	snscrape.modules.twitter.Tweet	https://twitter.com/gadgetboy_jp/status/160897...	2022-12-30 23:53:21+00:00	Google Pixel 7 vs Nothing Phone (1)\n\n スペック比較 \...	Google Pixel 7 vs Nothing Phone (1)\n\nスペック比較\...	1608974553925013507	{'_type': 'snscrape.modules.twitter.User', 'us...	
1	snscrape.modules.twitter.Tweet	https://twitter.com/navacute/status/1608960180...	2022-12-30 22:56:14+00:00	@broota_sumer @yabhishekhd Nothing phone 1 sup...	@broota_sumer @yabhishekhd Nothing phone 1 sup...	1608960180267683840	{'_type': 'snscrape.modules.twitter.User', 'us...	
2	snscrape.modules.twitter.Tweet	https://twitter.com/AmeerHworld/status/1608959...	2022-12-30 22:51:42+00:00	@yabhishekhd Nothing phone 1 🤩	@yabhishekhd Nothing phone 1 🤩	1608959038787813377	{'_type': 'snscrape.modules.twitter.User', 'us...	

```
textblob_df["textblob_sentiment"].value_counts()
```

```
Neutral     7457
positive    4706
Negative    1408
Name: textblob_sentiment, dtype: int64
```

```
textblob_df["textblob_sentiment"].value_counts().plot.barh(title = 'Sentiment Analysis',color='orange' , width=.4, figsize=(12,8),stacked = True)
```

```
<Axes: title={'center': 'Sentiment Analysis'}>
```



Topic Analysis

```
[ ] from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
    from sklearn.decomposition import NMF, LatentDirichletAllocation
    import numpy as np
```

```
[ ] vect =TfidfVectorizer(stop_words='english',max_features=1000)
    vect_text=vect.fit_transform(new_tweet_df['renderedContent'])
```

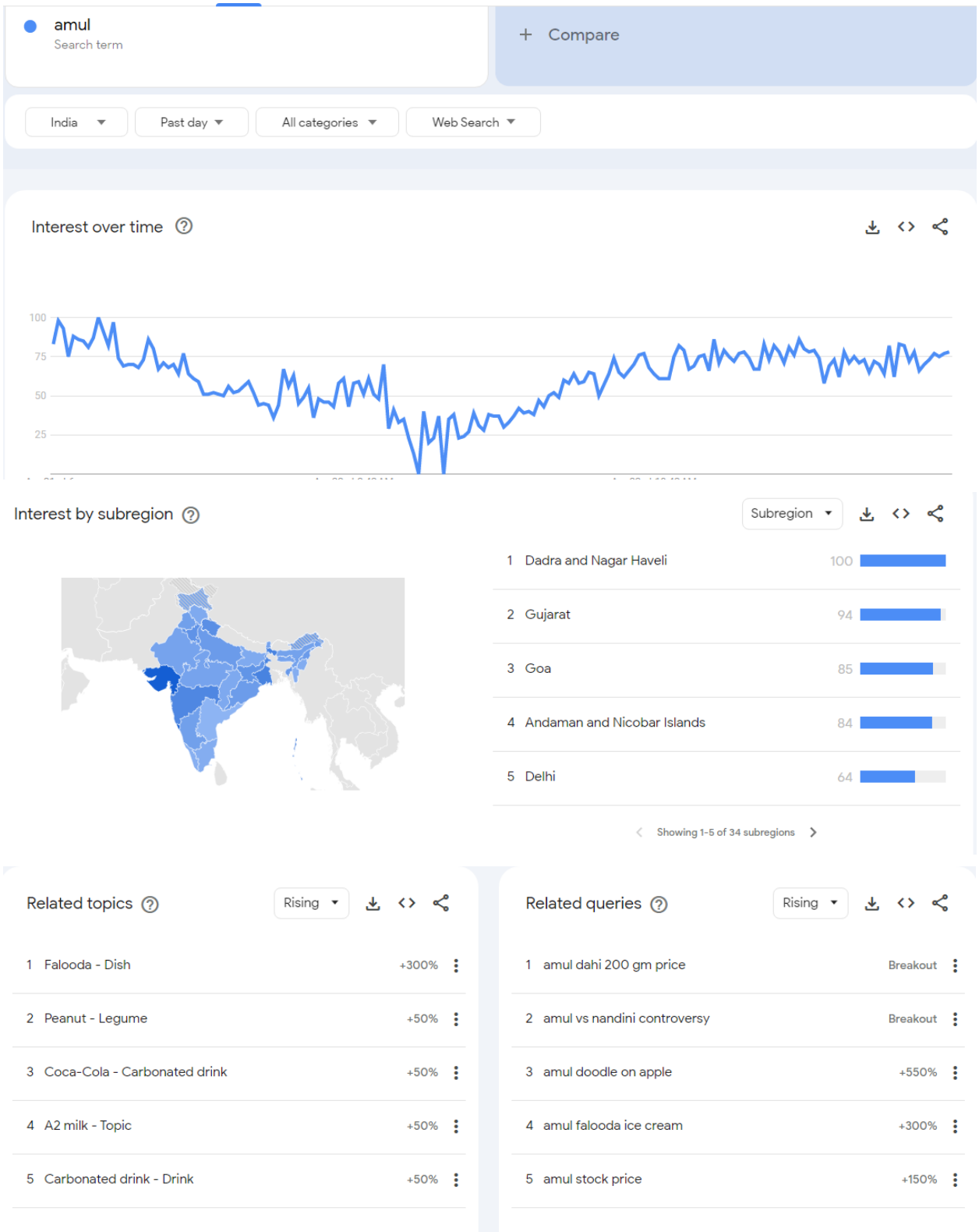
In natural language processing, Latent Dirichlet Allocation (LDA) is a generative statistical model that explains a set of observations through unobserved groups, and each group explains why some parts of the data are similar. The LDA is an example of a topic model. In this, observations (e.g., words) are collected into documents, and each word's presence is attributable to one of the document's topics. Each document will contain a small number of topics.

```
[ ] from sklearn.decomposition import LatentDirichletAllocation
    lda_model=LatentDirichletAllocation(n_components=7,
    learning_method='online',random_state=42,max_iter=1)
    lda_top=lda_model.fit_transform(vect_text)
```

```
vocab = vect.get_feature_names_out()
for i, comp in enumerate(lda_model.components_):
    vocab_comp = zip(vocab, comp)
    sorted_words = sorted(vocab_comp, key= lambda x:x[1], reverse=True)[:7]
    print("")
    print("Topic "+str(i)+": ")

    for t in sorted_words:
        print(t[0],end=" ")
        print("")
```

```
Topic 0:
nothingphone1
httpst
getpeid
ear
phone
buildingnothing
nothingos

Topic 1:
android
13
phone
pixel
pro
iphone
update

Topic 2:
phone
flipkart
buy
price
rs
available
details
```

For trend analysis add google trend snapshot



Name: Janvi D. Marathe	**DOP:**
Roll No: 36	**DOS:**
Division:	**Score:** **Sign:**

Experiment No 7: Develop a dashboard based on social media data.

Comparative Analysis of Twitter, Facebook and Youtube analytics dashboard

Theory:

1. Twitter Analytics: Twitter Analytics provides an in-depth analysis of your tweets, audience, and engagement on the platform. The dashboard offers insights into your account's performance over time, including impressions, engagement, top tweets, and followers.

Features:

- **Account summary:**Provides a quick overview of your account's performance, including tweets, impressions, and profile visits.
- **Tweet activity:**Offers detailed metrics on individual tweets, including impressions, engagement, and link clicks.
- **Audience insights:**Provides information on your followers' demographics, interests, and behavior.
- **Video insights:**If you've posted video content on Twitter, you can see metrics on video views, completion rates, and engagement.

2. Facebook Analytics: Facebook Analytics offers a comprehensive view of your page's performance, audience, and advertising campaigns. The dashboard offers insights into your page's activity, including engagement, reach, and demographic data.

Features:

- Page summary: Provides a quick overview of your page's performance, including page views, likes, and engagement.
- Audience insights: Offers detailed information on your page's audience, including demographics, interests, and behavior.
- Funnel insights: Shows you the steps users take on your page, including views, clicks, and conversions.
- Ads reporting: Provides detailed metrics on your advertising campaigns, including reach, frequency, and cost-per-result.

3. YouTube Analytics: YouTube Analytics offers a detailed view of your channel's performance, audience, and content. The dashboard offers insights into your video's activity, including views, engagement, and retention rates.

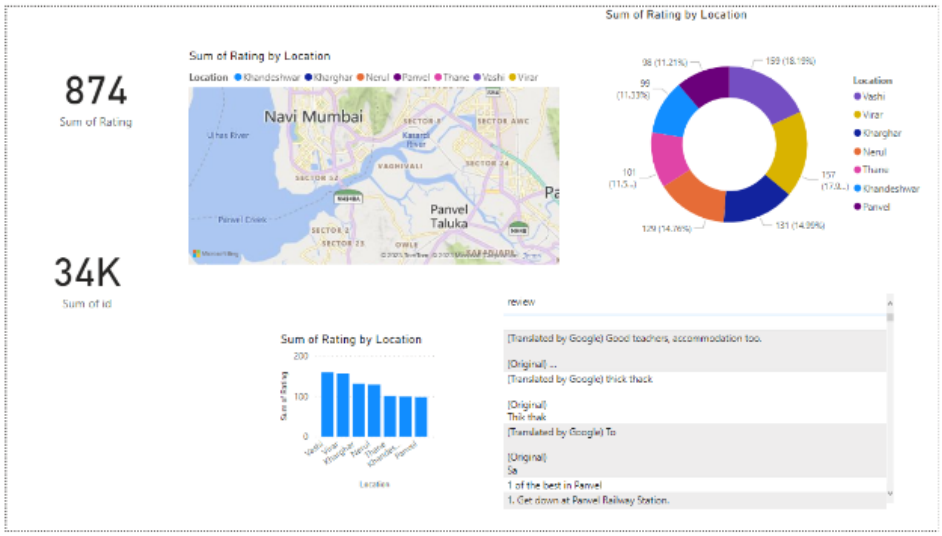
Features:

- **Overview:**Provides a quick overview of your channel's performance, including views, subscribers, and watch time.
- **Audience retention:**Shows how long viewers are watching your videos and which parts of your videos they're watching.
- **Traffic sources:**Offers insights into how viewers are discovering your videos, including search results, suggested videos, and external websites.
- **Ad performance:**Provides detailed metrics on your advertising campaigns, including impressions, views, and cost-per-view.

DATA:

	A	B	C
1	Rating	Location	
2	1	Panvel	
3	5	Virar	
4	3	Kharghar	
5	2	Vashi	
6	3	Vashi	
7	2	Khandeshwar	
8	1	Khandeshwar	
9	4	Kharghar	
10	1	Panvel	
11	4	Thane	
12	3	Virar	
13	1	Vashi	
14	3	Vashi	
15	1	Khandeshwar	
16	5	Nerul	
17	5	Virar	
18	3	Thane	

DASHBOARD :



Name: Janvi D. Marathe	**DOP:**
Roll No: 36	**DOS:**
Division:	**Score:** **Sign:**

Experiment No 8: Design the creative content for promotion of your business on social media platform

The Importance Of Social Media Marketing

Social media allows businesses to be in direct contact with target customers. Social media marketing is inexpensive and gives brands instant reach to billions of active users. For these reasons, social media has proven to be one of the most effective channels to promote products online. Social media has become the most influential and important virtual space where the platform is not only used for social networking but is also a great way of digitally advertising your brand and your products. Social media's power is commendable as you get to reach a large number of people within seconds of posting an ad, helping you reduce your costs, and making your ads reach out to your potential audience through these social media advertisements.

With over 80% of consumers reporting that social media—especially influencer content—significantly impacts buying decisions, marketers across industries are driving the evolution of social media marketing (SMM) from a stand-alone tool to a multipronged source of marketing intelligence on an increasingly important—and growing—audience.

Social Media Promotion

Promotion or marketing is said as a set of activities that make a product or service familiar or aware to the user. It is also an activity that attracts or influences the user to buy one. Nowadays, the majority of the promotion is done through various social media platforms and websites or even through cellphone-based applications.

Sometimes it so happens that companies promote their product or brand or service only through social media. Promotion through social media becomes more powerful and successful. Social media promotion helps companies to reach new customers and helps to engage with existing customers.

1. Upload content outside of business hours

It’s not advisable to promote when your audience is busy working and is inactive on social media. Your content will attract more eyeballs when they are at home before or after work. You can easily schedule posts at the optimal time with tools. Although, some potential customers might be on social networking sites while at work (e.g. during lunch). They are more likely to use it while commuting or after having dinner with their family. Reaching the right audience will also play a vital role in word-of-mouth promotion. It will increase the chances of them noticing your product, browsing your site, and even sharing your offerings with others. Create a posting schedule, and stick to it. If you don’t know how to start, check our ultimate guide to social media posting in 1 hour a day.

2. Publish high-quality & authentic content

Posting high-quality, authentic, and valuable content sheds a positive light on your brand. It encourages potential leads to visit your social media profile and website for more of your content. If you are finding it hard to come up with relevant and valuable content all the time, you can try out other ways of creating content. You can hire an agency, freelancers, or interns to help you create content. If you have any blog posts or videos, consider repurposing them into smaller bite-sized pieces. For example, consider creating visual quotes from a recent blog post.

3. Reward your most loyal fans

The best way to attract more followers is to listen to their needs and engage with them regularly. Look for users who actively provide suggestions or feedback. Encourage them to participate in various product discussions, and reward them for their input. Rewarding your customers by keeping them in the loop can increase your sales. You can host giveaways and contests to better engage with your customers. It will also help you get user-generated content, reviews, and free promotion.

4. Make use of promotional keywords

People tend to get attracted to keywords like “Free,” “offers”, “discount,” “giveaways,” and “sale.” If you incorporate these words appropriately in your posts, they can easily tempt your audience and bring you more sales opportunities. Although, make sure the content on your social media matches that of your website. Such promotional offers usually work best during holidays and other big events. While these keywords can supercharge your sales, be careful not to overdo it.

Also, if you feel aligning your content strategies according to the holidays and events is overwhelming, use social media posting tools to simplify your task.

5. Focus on driving traffic to your website

If you create content and nobody sees it, what's the benefit of all that effort? That's where promotion comes in! Test out different posts to see what resonates best with your audience. Is it images, videos, carousels, or stories? Maximize the number of viewers of your content to grow your website traffic. Consider collaborating with other influencers in your niche. You can even hold a social media contest that encourages your audience and their friends to follow you. Remember, testing new ideas is essential, so continue experimenting.

6. Keep up with the latest trends

It is crucial to be relevant and stay on top of the latest trends on social media channels. By being an early adopter, you can find ways to stand out and promote your brand more efficiently. Some examples are: Live video contents, Instagram stories, Messaging chatbots, 360° videos. It is also important to keep up-to-date with algorithm changes on each platform. It is a difficult task, but the benefits are worth your time. The use of these tactics is mandatory to create a winning strategy.

7. Create a hashtag to engage with customers

Customer testimonials play a very significant role in social media. Yet, searching for recommendations that fit into a tweet or social media post can be a difficult task. Consider creating a hashtag precisely for that purpose. For example, if you are running a recycling business, you can create the hashtag #CleanTheEarth. Inspire your customers to share their stories and interactions with your brand by using the hashtag. You can then repost their content as social proof of your brand impact.

8. Be innovative with videos on Instagram

Videos shared on Instagram are guaranteed to get more attention. Videos tell stories and showcase your new products, branding, and customer testimonials. Thus, effectively capturing and retaining the attention of the users. You can go live and take up a tutorial or Q&A session. You can post reels and IGTV videos to share a series of DIY or testimonials. Use Instagram stories to keep your presence even more active throughout the day.

9. Include social media icons on your website

Adding social media icons to your website brings more followers to your social media handles. They make it easier for customers to find and follow you. Make your icons stand out by

incorporating your brand colors into them. This consolidates them well into your existing website theme. Adding these social media icons will even increase the chances of your website content being shared. It eliminates that additional step of copying, pasting, and manually sharing the links to friends and family. Now they can directly share interesting blogs or offerings from your website to their social media accounts.

10. Generate content as Infographics

If you have a long piece of content, consider turning it into an infographic. Design it with a tool like Canva or Visme, and then link it to your blog post or profile, where customers can learn more about it. Many people find it easier to consume content in visual form. Infographics are a simple way of adding value. They are also easily shareable, making this a simple yet effective social media strategy.

11. Design a memorable character

Does your business have a logo, mascot, or perhaps a pet representing your business? You can use that as a recurring character on your social media. It will help supporters to recognize and connect with your brand. Also, it would help customers relate with you better, creating a sense of familiarity.

12. Follow Us on Social: Cross-Promote on Other Channels

Some of the easiest cross-promotion for cross channels is occasionally reminding your followers that you have other social media accounts. You don’t need to be shy about asking people to follow you on social media, or on other networks. This can be as simple as a link to a given social media page accompanied with a call-to-action asking people to follow the account. For a more subtle approach on cross-promotion, use space in your bio to mention that you’re on other networks.

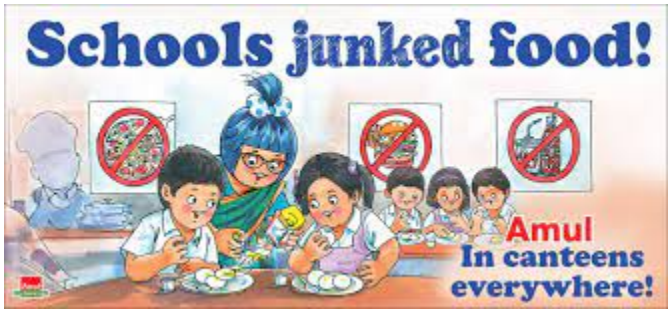
13. Provide Social-Only Discounts

In traditional advertising, you see weekly deals in the newspaper or receive discounts in email. Social is no different. These discounts are served to a certain segment of your audience and end after some time. The simplest way is to offer a code in a post. Experiment around with social-only discounts. Try doing flash deals with limited quantities in Instagram Stories or posting an Offer on Facebook.

14. Promote Your Product in Social Media Communities

Joining social media groups/communities relevant to your niche can also serve as a good way to spread awareness about your product. Your product promotion efforts in social media groups will implicitly bring better results since members are already interested in a similar product. But it is really important that you connect with group members well, converse in their interests, and be transparent when providing details of your product. If you pay attention to these points, social media communities will prove to be a good source for getting quality leads for your product.

Snapshot of creative content for promotion of your business on social media platform.



Name: Janvi D. Marathe	**DOP:**
Roll No: 36	**DOS:**
Division:	**Score:** **Sign:**

Experiment No 9 : Analyze competitor activities using social media data related to your business

A social media competitor analysis is an analysis of your competition on social media to find out what their strengths and weaknesses are, and how those strengths and weaknesses compare to your own.

It's a process of benchmarking your own results against the heavy-hitters in your industry, so you can identify opportunities for growth as well as strategies that aren't performing as well as they should.

A social media competitive analysis, specifically, will help you:

- Identify who your competitors are on social media
- Know which social platforms they're on
- Know how they're using those platforms
- Understand how well their social strategy is working
- Benchmark your social results against the competition
- Identify social threats to your business
- Find gaps in your own social media marketing strategy

Social media competitor analysis tools:

1. Unbox Social

Unbox Social is a multi-faceted tool, which offers many different features, including competition tracking. With the help of the competition tracking feature, you can track your competitors' activity, and know what kind of content that is working in the industry.

You can also capture your share of voice against that of your competitors in the industry with the help of Unbox Social. The tool allows you to monitor your performance against that of your competitors. It also helps you capture the interactions on your content against that of your competitors.

To dig a little deeper, with Unbox Social, you will get access to insights on what the audiences are saying about your competitors. The tool will also help you get an idea about the key themes around your competitors' conversations. Further, you can find out which of your competitors' posts are receiving a good response and build a content strategy accordingly.

2. BuzzSumo

BuzzSumo offers a competition research tool. This competition tracking tool lets you look at your competitors' most shared content by simply keying in the competitor domain. It also helps analyze what is working for them with the help of filters such as 'content format' and 'networks used to get traffic'.

The BuzzSumo tool helps you track mentions and discover trends.

3. Ahrefs

Ahrefs is a tool that helps you with everything related to SEO. Ahrefs comes with a range of features and tools that help you benchmark your SEO strategy against competitors. This makes it one of the most useful tools for social media competitor analysis. The tool offers solutions like site explorer, keywords explorer, site audit, rank tracker, and content explorer.

4. SimilarWeb

SimilarWeb is another Competition Analysis tool you can consider using. You can use SimilarWeb to gain comprehensive insights into your competition in the digital media marketing arena.

With the help of Competitive Benchmarking that the tool offers, you can compare the performance of your website as well as the app.

List down your business competitors and compare your business with competitors business

AMUL has various competitors across India :

Mother Dairy: Mother Dairy is a subsidiary of the National Dairy Development Board and one of the leading dairy brands in India. It sells dairy products, juices, and frozen snacks.
Britannia Industries: Britannia Industries is a well-known Indian food company with a wide range of products ranging from dairy products to snacks.
Nestle India: Nestle India is a subsidiary of Nestle SA and is one of the leading players in the food and beverage industry in India.
Gokul Refoils & Solvent: Gokul Refoils & Solvent is a Gujarat-based dairy products company, which produces a number of products such as ghee, paneer, and other dairy products.
Kwality Ltd: Kwality Ltd is a well-established dairy products company in India, which produces a range of dairy products.

Parag Milk Foods: Parag Milk Foods is a well-known dairy products company in India, which produces a range of dairy products.
Hatsun Agro Products: Hatsun Agro Products is a Chennai-based dairy products company, which produces a variety of products such as milk, cheese, and other dairy products.
Vadilal Industries: Vadilal Industries is an Indian dairy products company, which produces a range of dairy products.
Mahanand Dairy: Mahanand Dairy is one of the largest dairy companies in India, which produces a range of dairy products.

What changes need to be made in your business based on competitor analysis?
Low-cost products
One of the best features of the Amul marketing strategy is its low product pricing. Its pricing strategy caters to the daily consumption habits of the average Indian consumer. This makes Amul's products, such as butter, milk, ghee, curd, and more, affordable.
Digital marketing efforts
Albeit an old brand, Amul has not limited its efforts to the traditional style of marketing and advertising. It has expanded into the digital arena and advertised its products via Facebook, Twitter, and Instagram. Amul connects with its customers online and even resolves their queries and complaints through these platforms.
Marketing mix
The marketing mix is the combination of factors a company can control in order to influence its customers to purchase its products and services. It usually includes the four Ps: product, price, place, and promotion.
Amul does not have premium offerings. It, instead, focuses solely on the mass market. On a customer-wise targeting structure, Amul's products can be categorized under three broad labels, which are as follows:

Kids: Amul Kool and Amul Pro Kids
Youth: Amul cheese spread and Amul milk
Health-conscious customers: Amul Shakti and Amul Lite

Name: Janvi D. Marathe	**DOP:**
Roll No: 36	**DOS:**
Division:	**Score:** **Sign:**

Experiment No 10: Develop social media analytics models of your business

Depending on the business objectives, social media analytics can take four different forms, namely, descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics.

1. Descriptive Analytics (Is Reactive in Nature)

Descriptive SMA tackles the questions of "what happened and/or what is happening?" Descriptive analytics gather and describe social media data in the form of reports, visualizations, and clustering to understand a well-defined business problem or opportunity. Social media user comments analysis, for instance, falls into the descriptive analytics category. Comment analysis can be used to understand users' sentiments or identify emerging trends by clustering themes and topics.

Comments and posts can be grouped together, for instance, for the purpose of sentiment analysis (as offered, for instance, by our very own SentiOne platform). Any time you gather a lot of similar data points in order to analyse them looking for patterns, sentiments, and/or trends, you're dealing with descriptive analytics.

Currently, descriptive analytics accounts for the majority of social media analytics landscape.

2. Diagnostic Analytics (Is also Reactive in Nature)

Diagnostic SMA analytics looks into the questions of "why something happened?" For example, while descriptive analytics can provide an overview of your social media marketing campaign's performances (posts, mentions, followers, fans, page views, reviews, pins, etc); diagnostic analytics can distill this data into a single view to see what worked in your past campaigns and what didn't.

Diagnostic analysis focuses on the numbers: like counts, follower numbers, pageviews, reviews, shares, what have you. This type of analytics focuses on the performance of posts and campaigns

and attempts to discern what made them successful. By comparing the performance of different campaigns, trends and consumer preferences can be discerned.

Enablers of diagnostics analytics include inferential statistics, behavioral analytics, correlations & retrospective analysis and outcome being cause and effect analysis of a business issues.

3. Predictive Analytics (Is Proactive in Nature)

Predictive analytics involves analyzing large amounts of accumulated social media data to predict a future event. Thus, it deals with the question of "what will happen and/or why will it happen?" For example, an intention expressed over social media (such as buy, sell, recommend, quit, desire, or wish) can be mined to predict a future event (such as a purchase). Alternatively, businesses can predict sales figures based on historical visits (or in-links) to a corporate website.

Predictive analytics attempt to predict trends, events, and shifts based on existing data. It can range from simple things, like predicting possible visits to a location based upon posts expressing that intention, to forecasting entire trends and phenomena based upon mentions. Social listening tools can help identify upcoming trends and shifts in consumer behavior by analyzing large volumes of social media data and indicating the shifting popularity of keywords.

4. Prescriptive Analytics (Is also Proactive in Nature)

While predictive analytics help to predict the future, prescriptive analytics suggest the best action to take when handling a scenario. For example, if you have groups of social media users that display certain patterns of buying behavior, how can you optimize your offering to each group? Like predictive analytics, prescriptive analytics has not yet found its way into social media data. The main enablers of prescriptive analytics include optimization and simulation modeling, multi-criteria decision modeling, expert systems, and group support systems.

Prescriptive analysis is the analysis of data with the intention of providing the best way to proceed at any given moment. This can be applied to situations ranging from handling social media crises and incidents ("how well does this type of apology track with our target audience?") to purchase preferences ("we've identified this group of customers – how do we optimise our sales process to their habits?"). Although it's an incredibly useful form of analysis, it requires a lot of data in order to truly show its potential.

For the business problem statement:

Analysis of AMUL Dairy Company .Anand Milk Union Limited is an Indian state government-owned dairy-based cooperative society officially named the Gujarat Milk Marketing Federation based in Anand, Gujarat. It is under the ownership of Gujarat Cooperative Milk Marketing Federation Limited, Department of Cooperation, Government of Gujarat.

Using Social Media Analytics tool we will try to analyze the production of various AMUL products all over INDIA and its comparison with its competitors.

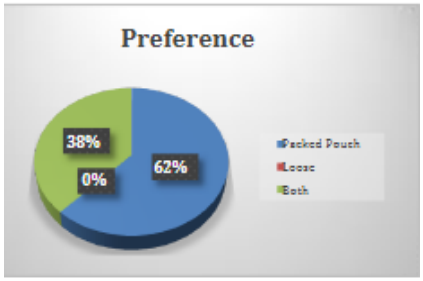
1. Descriptive Analytics

Our analysis describes the case study of the dairy cooperative, AMUL, in western India, which has developed a successful business model in the large emerging economy. It was primarily responsible, through its innovative practices, for India to become the world's largest producer of milk. The paper draws various insights from AMUL 's experience that can be useful to cooperatives around the world as well as to companies involved in doing business in broad emerging markets such as India and China. Most of these economies have underdeveloped markets and fragmented supply bases. Market failures are high for many of these small producers. On the other hand, the size of both markets and suppliers is very large. As a result, firms that identify appropriate business strategies that take these characteristics into account are more likely to succeed in these markets. The following is a key lesson from AMUL 's success: companies in such environments must simultaneously build markets and suppliers to synchronize demand and supply planning, establish or become part of the producer network (i.e., cooperatives in this case) to achieve economies of scale, concentrate on operational efficiency to achieve cost leadership in order to allow for a low price strategy. In addition, a central focus is needed to bring together a diverse element and a long-term approach. Different industrial sectors may be at different stages of development in emerging economies. For certain industries, any of the environmental features alluded to above may not be preserved. A subset of approaches adopted by AMUL would still be very useful, however. Therefore, companies that plan to target large undeveloped markets or intend to take advantage of a broad but limited supplier base will still benefit.

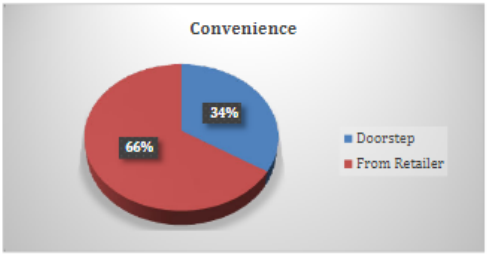
2. Diagnostic Analytics

The collected data were not easily understandable, so we like to analyze the collected data in a systematic manner and interpreted with simple method. The analysis and interpretation of the data involves the analyzing of the collected data and interpretation it with pictorial representation such as bar charts, pie charts and others.

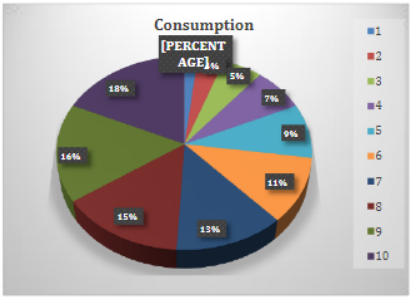
Which dairy product do you prefer?



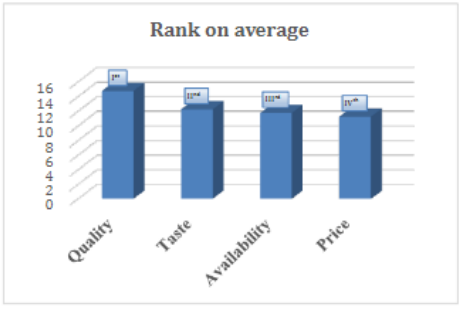
Do you get milk pouch at



What is total consumption of milk in litre/day?



What do you like in these products most "Quality, Availability, Price or Taste"?



3. Predictive Analytics

It is found that Amul has a strong brand value in the market and it increases rapidly through its advertising. Amul focus on Advertising , Quality and customer service which makes it a chief brand in dairy industry. Competitors like mother dairy focus on availability , thus they attracts only existing customers and narrowing their captured market. Amul should provide doorstep facility which is preferred from survey. The demand of Amul products is higher than its supply which impose customers to buy competitors' products.

4. Prescriptive Analytics

It is found that Amul has a strong brand value in the market and it increases rapidly through its advertising. Amul focus on Advertising , Quality and customer service which makes it a chief brand in dairy industry. Competitors like mother dairy focus on availability , thus they attracts only existing customers and narrowing their captured market. Amul should provide doorstep facility which is preferred from survey. The demand of Amul products is higher than its supply which impose customers to buy competitors' products.