Experiment No: - 1

Aim: Study various - i) Social Media platforms (Facebook, Twitter, YouTube, etc.)

ii) Social Media analytics tools (Facebook Insights, Google Analytics, Netlytic, etc.)

iii) Social Media Analytics techniques and engagement metrics (page level, post level, member level)

iv) Applications of Social media analytics for business. e.g. Google Analytics
https://marketingplatform.google.com/about/analytics/

https://netlytic.org/

Theory:

Social Media Platforms –

Social media is integral to our daily lives. With 5.17 billion social media users worldwide – accounting for 59.9% of the global population – social platforms have become necessary hubs for gathering information, connecting with our friends and loved ones, and growing our businesses. With the right tools, resources, and a strong social media strategy, marketers can leverage social media platforms to boost awareness for their business, engage their target audience, and even nurture a loyal community.

Social media is an internet-based form of communication. Social media platforms allow users to have conversations, share information and create web content. There are many forms of social media, including blogs, micro-blogs, wikis, social networking sites, photo-sharing sites, instant messaging, video-sharing sites, podcasts, widgets, virtual worlds, and more.

i) Information on Social Media Platforms (Facebook, Twitter, YouTube, Instagram, LinkedIn)

1. Facebook

Facebook, launched in 2004 by Mark Zuckerberg, is the largest social networking platform with over 3 billion monthly active users. It allows users to create personal profiles, business pages, and communities through groups. Key features include posts (text, images, and videos), live streaming, Facebook Stories, and Messenger for private communication. Businesses leverage Facebook Marketplace, targeted advertising, and monetization options. Public data such as posts and comments can be accessed through the **Meta Graph API**, but private data is restricted. While Facebook is a hub for digital marketing and social interaction, concerns include privacy issues, misinformation, and algorithmic biases affecting content reach.

2. Twitter (Now X)

Twitter, launched in 2006 and now owned by Elon Musk, is a microblogging platform with around 550 million monthly active users. It is known for short-form text posts called **tweets** (280-character limit), trending hashtags, and real-time news updates. Other features include **threads** for longer discussions, **Spaces** for live audio conversations, and **retweets** for content amplification. Twitter's **API (v2)** provides access to public tweets, mentions, and engagement metrics, making it popular for sentiment analysis, political discussions, and brand engagement. However, misinformation, bot activity, and restrictions on free-tier API access pose challenges for data analysis and research.

3. YouTube

YouTube, launched in 2005 and owned by Google, is the world's largest video-sharing platform with over 2.7 billion monthly users. It supports content in multiple formats, including long-form videos, **YouTube Shorts**, live streaming, and monetized content through ads, memberships, and sponsorships. The **YouTube Data API v3** allows access to public data such as video details, view counts, comments, and channel analytics. YouTube is widely used for education, entertainment, and product marketing. However, it faces challenges such as copyright restrictions, demonetization issues for smaller creators, and algorithmic preferences favoring viral content over niche educational materials.

4. Instagram

Instagram, introduced in 2010 and now owned by Meta, is a visually-driven social media platform with around 2.5 billion users. It focuses on photo and video-sharing through **posts, stories, reels**, and IGTV (for long-form content). Businesses and influencers utilize Instagram for brand promotions, while users engage with content through hashtags, likes, and direct messages. The **Instagram Graph API** provides limited access, primarily for business accounts, allowing developers to retrieve insights, user engagement metrics, and ad performance. Instagram's challenges include algorithmic visibility issues, fake engagement through bots, and restrictions on organic reach due to paid promotions being prioritized.

5. LinkedIn

LinkedIn, launched in 2003 and owned by Microsoft, is a professional networking platform with over 1 billion users. It is widely used for **job searching, recruiting, business networking, and B2B marketing**. Users can create detailed professional profiles, publish articles, and engage in industry discussions. LinkedIn's **API** access is highly restricted, mainly allowing access to job listings and company pages. The platform plays a significant role in hiring and corporate branding, but challenges include frequent spam messages, limited data extraction capabilities, and premium features that restrict free users from accessing advanced insights.

Social Media Analytics –

Social media analytics is the ability to gather and find meaning in data gathered from social channels to support business decisions—and measure the performance of actions based on those decisions through social media. Practitioners and analysts alike know social media by its many websites and channels: Facebook, YouTube, Instagram, Twitter, LinkedIn, Reddit and many others. Social media analytics is broader than metrics such as likes, follows, retweets, previews, clicks, and impressions gathered from individual channels. It also differs from reporting offered by services that support marketing campaigns such as LinkedIn or Google Analytics.

Social media analytics uses specifically designed software platforms that work similarly to web search tools. Data about keywords or topics is retrieved through search queries or web 'crawlers' that span channels. Fragments of text are returned, loaded into a database, categorized and analyzed to derive meaningful insights. Social media analytics includes the concept of social listening. Listening is monitoring social channels for problems and opportunities. Social media analytics tools typically incorporate listening into more comprehensive reporting that involves listening and performance analysis.

ii) Social Media Analytics Tools –

1. Facebook Insights

Facebook Insights is an inbuilt analytics tool for business pages, helping marketers track engagement, reach, and audience demographics. It provides **page-level and post-level metrics**, including likes, shares, comments, impressions, and follower growth. It also gives insights into audience behavior, such as peak activity times and content performance trends. The tool is useful for refining content strategies and ad targeting. However, private user data is not accessible due to privacy policies, and detailed analytics are restricted to page admins.

- **AI-Powered Recommendations:** It provides AI-driven suggestions to improve engagement, such as the best time to post and which types of content work best with your audience.
- **Benchmarking Feature:** Allows businesses to compare their page's performance against competitors in similar industries.
- **Ad Insights Integration:** Helps track the effectiveness of Facebook Ads, showing detailed analytics on reach, click-through rates, and audience engagement.

2. Google Analytics

Google Analytics (GA4) is a powerful web analytics tool used for tracking website and social media referral traffic. It provides insights into user behavior, session duration, conversion rates, and engagement patterns. Businesses use GA4 to measure the effectiveness of social media campaigns by analyzing how traffic from platforms like Facebook, Twitter, and LinkedIn contributes to sales, sign-ups, or other goals. The platform supports **event-based tracking, AI-driven insights, and cross-device analytics**, making it essential for digital marketing and performance optimization. However, its complexity and data sampling limitations can be challenging for beginners.

- **Predictive Metrics:** GA4 includes AI-powered insights such as churn probability and purchase probability, helping businesses predict user behaviour.
- **Event-Based Tracking:** Unlike Universal Analytics, GA4 tracks all interactions as events, providing deeper insights into user journeys across web and mobile apps.
- **Cross-Platform Tracking:** Google Analytics now integrates data from multiple platforms (web, mobile, YouTube, and apps) into a single dashboard.

3. Communalytic (formerly Netlytic)

Communalytic is an advanced social media analytics tool designed for analyzing online communities, detecting misinformation, and conducting network-based studies. Unlike traditional sentiment analysis tools, it focuses on **network analysis, toxicity detection, and information flow within social media conversations**. It can collect data from platforms like Twitter, Reddit, and YouTube, helping researchers study trends, user interactions, and discourse patterns. However, due to API access restrictions, data collection depends on platform-specific limitations, and ethical considerations must be followed while analyzing user-generated content.

- **Misinformation & Toxicity Detection:** Uses AI to detect misinformation, hate speech, and toxicity in online discussions, making it valuable for studying harmful content.
- **Social Network Mapping:** Helps researchers visualize and analyze communication patterns between users to identify key influencers and community structures.
- **Advanced Sentiment Analysis:** Uses NLP to determine sentiment polarity and emotional tone in conversations across multiple social platforms.

4. Hootsuite Analytics

Hootsuite Analytics is a social media management tool that helps businesses track performance across multiple platforms (Facebook, Instagram, Twitter, LinkedIn). It offers **real-time engagement metrics, post scheduling, competitor benchmarking, and sentiment analysis**. Its AI-powered recommendations assist in optimizing content strategy based on past performance. The tool is widely used by marketers for cross platform analysis, but its **premium subscription model** limits access to advanced features for free users.

- **Automated Social Listening:** Monitors social conversations in real-time, tracking brand mentions, hashtags, and competitor activities across multiple platforms.
- **Team Collaboration Tools:** Allows multiple team members to manage social accounts, assign tasks, and respond to comments efficiently.
- **ROI Tracking for Campaigns:** Provides detailed reports on how social media marketing efforts contribute to business goals, including conversions and revenue impact.

5. Sprout Social

Sprout Social is a data-driven social media analytics and management tool that provides detailed engagement insights, campaign tracking, and audience behavior analysis. It features **AI-powered social listening, hashtag tracking, and influencer identification**. Sprout Social integrates with major platforms like Facebook, Twitter, LinkedIn, and Instagram, making it valuable for brands looking to refine their digital strategies. However, like Hootsuite, **it requires a paid subscription** for full functionality, which may not be cost-effective for small businesses.

- **AI-Driven Smart Inbox:** Aggregates messages from all social channels into one inbox, helping teams manage customer interactions efficiently.
- **Influencer Identification:** Analyzes engagement trends to identify potential brand advocates and influencers who can boost marketing efforts.
- **Competitor Sentiment Analysis:** Uses AI-powered sentiment tracking to compare customer sentiment about a brand versus its competitors.

iii) Social Media Analytics Techniques and Engagement Metrics –

Social media metrics are points of data essential to tracking the performance of your social channels, content, strategy and the impact it all has on your business goals. Measuring social media engagement metrics, for example, can allow for a comprehensive understanding of how your content resonates with your audience and provide invaluable insights to help identify opportunities to better resonate with them.

Without a deep dive into these metrics, one would not be able to identify invaluable insights, such as engagement metrics scoring low. Without that understanding, the opportunity to discover how to refine a brand's social media marketing campaign to better resonate with its defined buyer personas in its social content would be lost. Not to mention, larger business opportunities can sometimes be discovered through the analysis of social media metrics, such as identifying a new market segment.

Social Media Analytics techniques help in tracking and analyzing user interactions, trends, and performance across different platforms. Engagement metrics are categorized into **Page Level, Post Level, and Member Level**, each providing valuable insights into audience behavior and content effectiveness.

1. Social Media Analytics Techniques

- **Sentiment Analysis** – Uses Natural Language Processing (NLP) to determine the emotional tone (positive, negative, neutral) of user comments and posts.
- **Trend Analysis** – Identifies trending topics, hashtags, and discussions to understand what is currently popular in a specific industry or community.

- **Content Analysis** – Evaluates text, images, videos, and other media to determine which types of content drive the most engagement.
- **Network Analysis** – Maps relationships and interactions between users to identify influencers, brand advocates, and key discussion leaders.
- **Predictive Analytics** – Uses AI/ML models to forecast future engagement, customer behavior, and content virality.
- **Clickstream Analysis** – Tracks the path users take on social media pages to understand navigation patterns and optimize user experience.

2. Engagement Metrics at Different Levels

A. Page-Level Metrics (Overall Page Performance)

These metrics measure engagement across an entire social media page (business account, brand profile, or group).

- **Follower Growth Rate** – Measures the rate at which new users follow the page over time.
- **Page Reach** – The number of unique users who have seen any content from or about the page.
- **Impressions** – Total number of times content from the page appears on users' screens.
- **Engagement Rate by Page** – Percentage of users who interact with the page content compared to total page visitors.
- **Page Clicks (CTA Performance)** – Tracks clicks on website links, buttons, or other CTAs within the page.

B. Post-Level Metrics (Individual Post Performance)

These metrics focus on engagement for specific posts, including text, images, videos, and ads.

- **Likes/Reactions** – Number of likes, reactions (e.g., love, wow, angry, etc.), and other user responses.
- **Shares/Retweets** – The number of times a post is reshared, amplifying its reach.
- **Comments & Replies** – Indicates audience engagement, discussions, and potential sentiment (positive or negative).
- **Click-Through Rate (CTR)** – Percentage of users who clicked on links embedded in the post.
- **Post Reach & Impressions** – Measures how many unique users saw the post and how many total views it received.
- **Video Watch Time** – Tracks how long users watch videos before dropping off.

C. Member-Level Metrics (User-Specific Engagement)

These metrics analyze how individual users interact with content and how their behaviour contributes to the overall engagement.

- **Active Users** – Measures how many users engage with content regularly (daily, weekly, monthly).
- **Customer Sentiment Score** – Analyzes individual user comments to determine satisfaction or dissatisfaction.
- **Influencer Score** – Identifies highly engaged users who share, comment, and drive discussions.
- **Response Time & Interaction Rate** – Measures how quickly users respond to posts and how frequently they engage.
- **Conversion Rate Per User** – Tracks how many individual users complete a desired action (sign-up, purchase, inquiry).

Applications of Social Media Analytics for Business

Social Media Analytics (SMA) plays a crucial role in business decision-making by providing insights into audience behavior, engagement, brand performance, and market trends. Businesses leverage SMA for marketing, customer service, competitive analysis, and more. Below are some key applications:

1. Brand Monitoring & Reputation Management

- Businesses use social media analytics to track brand mentions, customer sentiment, and online reviews.
- Sentiment analysis tools analyze positive, negative, and neutral sentiments, helping brands manage crises and maintain a positive image.
- Example: **Google Analytics** can track how social media traffic impacts website visits and conversions.

2. Customer Engagement & Personalization

- Social media insights help businesses understand customer preferences and create personalized marketing campaigns.
- Data-driven targeting allows companies to deliver relevant content to specific audience segments.
- Example: **Facebook Insights** provides detailed audience demographics, engagement trends, and content performance to optimize strategies.

3. Competitive Analysis & Market Trends

- Businesses use SMA tools to analyze competitor activities, including engagement levels, trending content, and audience sentiment.
- Tracking industry trends helps companies adjust their strategies and stay ahead in the market.
- Example: **Communalytic (formerly Netlytic)** enables businesses to map competitor networks and analyze social interactions.

4. Social Media Advertising Optimization

- SMA helps businesses measure the effectiveness of paid advertisements across platforms like Facebook, Instagram, and Twitter.
- Metrics like Cost-Per-Click (CPC), Click-Through Rate (CTR), and conversion rates help optimize ad spend and maximize ROI.
- Example: **Google Analytics** tracks social media referral traffic, showing which platforms contribute the most to conversions.

5. Lead Generation & Sales Growth

- Businesses use social media analytics to identify potential customers and engage them with targeted campaigns.
- Tracking user behaviour on social media helps companies design better sales funnels.
- Example: **LinkedIn Analytics** helps B2B companies identify and engage with high-value prospects.

6. Product & Service Improvement (Customer Feedback Analysis)

- Companies analyze social media comments, reviews, and feedback to improve their products and services.
- AI-driven sentiment analysis tools help detect customer pain points and unmet needs.
- Example: **Twitter Analytics** enables businesses to track real-time customer concerns and respond proactively.

7. Influencer Marketing & Community Building

- Identifying key influencers and brand advocates helps businesses amplify their marketing efforts.
- Tracking influencer performance ensures businesses partner with the right creators for maximum reach.
- Example: **Instagram Insights** helps brands measure influencer campaign performance by tracking engagement and follower growth.

8. Crisis Detection & Risk Management

- Businesses monitor sudden spikes in negative sentiment to detect potential PR crises early.
- Quick responses to negative trends help maintain brand credibility and trust.
- Example: **Hootsuite Analytics** alerts businesses about sudden increases in negative mentions.

9. Measuring ROI & Campaign Effectiveness

- Businesses track engagement, conversions, and revenue generated from social media campaigns.
- Performance tracking helps refine future marketing efforts for better returns.
- Example: **Google Analytics UTM Tracking** measures how social media campaigns drive website traffic and conversions.

10. Event & Campaign Performance Tracking

- Businesses use social media analytics to measure the success of online and offline events, product launches, and promotional campaigns.
- Metrics like hashtag reach, engagement rate, and audience sentiment help assess the impact of an event.
- Example: **Twitter Analytics & Facebook Insights** track event-related hashtags, mentions, and live engagement, helping brands optimize future campaigns.

Case Study: The Impact of Social Media Analytics on Nike's Brand Growth

Nike, Inc. is a globally recognized sportswear and athletic footwear brand known for its innovation, marketing strategies, and athlete endorsements. Founded in 1964, Nike has become a leader in the sports industry, offering high-performance products across various categories, including running, basketball, and lifestyle wear. The brand is famous for its "Just Do It" slogan and impactful digital marketing campaigns, leveraging social media analytics to enhance customer engagement, product development, and brand loyalty. Nike continuously evolves by integrating AI, data analytics, and cutting-edge technology into its business strategy.

i) Social Media Platforms

Nike effectively utilizes social media platforms such as **Facebook, Twitter, Instagram, and YouTube** to engage with its audience. Each platform serves a unique purpose:

- **Facebook & Instagram:** Used for storytelling, product promotions, and community engagement through interactive posts and influencer collaborations.
- **Twitter:** Focuses on real-time customer engagement, announcements, and customer service.
- **YouTube:** Nike creates high-quality video campaigns that generate millions of views, fostering emotional connections with consumers.

ii) Social Media Analytics Tools

Nike leverages various analytics tools to track performance and optimize strategies:

- **Facebook Insights:** Provides engagement data such as likes, shares, and impressions, helping Nike refine content strategies.
- **Google Analytics:** Tracks website traffic from social media campaigns, identifying high-converting sources.
- **YouTube Analytics:** Analyzes video watch time, audience demographics, and retention rates to improve content strategy.
- **Netlytic (now Communalytic):** Conducts sentiment analysis on public opinions regarding Nike's campaigns and product launches.

iii) Social Media Analytics Techniques and Engagement Metrics

Nike employs various **analytics techniques** to measure performance and engagement:

- **Page-Level Metrics:** Nike tracks follower growth, reach, and engagement rates on platforms like Facebook and Instagram.
- **Post-Level Metrics:** They analyze likes, comments, shares, and click-through rates to identify high-performing content.
- **Member-Level Metrics:** Nike monitors influencer collaborations and customer interactions to assess brand perception.
- **Sentiment Analysis:** By using Netlytic/Communalytic, Nike gauges public opinion on controversial campaigns, enabling real-time brand management.

iv) Applications of Social Media Analytics for Business

Nike utilizes social media analytics for various business purposes, including:

- **Optimizing Marketing Campaigns:** Data-driven decisions help create personalized and targeted advertisements, improving conversion rates.
- **Customer Sentiment Analysis:** Understanding audience reactions enables Nike to address concerns and enhance brand reputation.
- **Product Development & Feedback:** By analyzing customer reviews and social discussions, Nike improves product designs and future releases.
- **Competitor Analysis:** Nike monitors competitors' social media activities to refine strategies and maintain market leadership.

This case study demonstrates how Nike successfully integrates **social media analytics** to drive brand engagement, improve customer experience, and boost sales.

Google Analytics Integrated with Shopify Store –

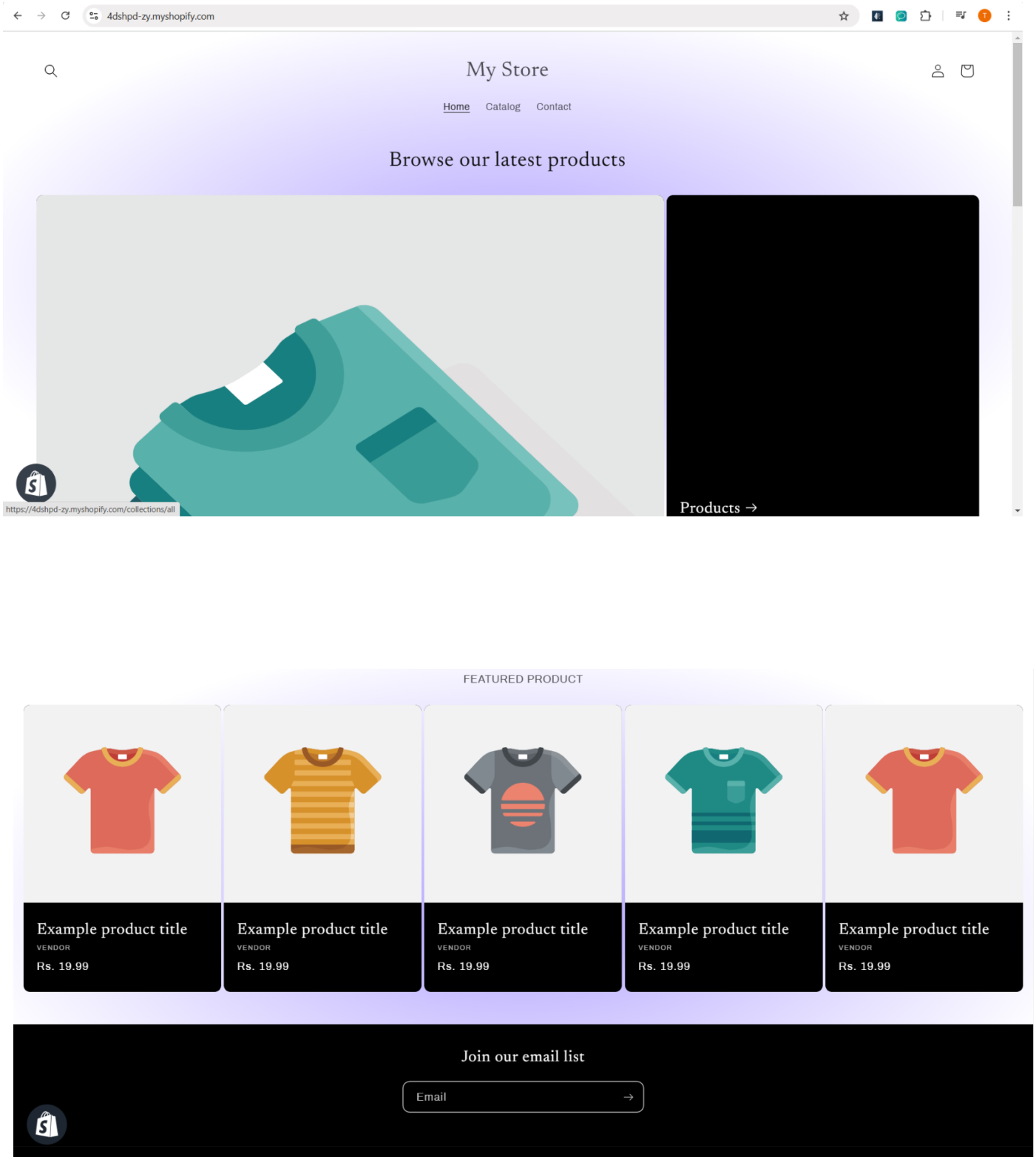


Figure 1 - Shopify Store

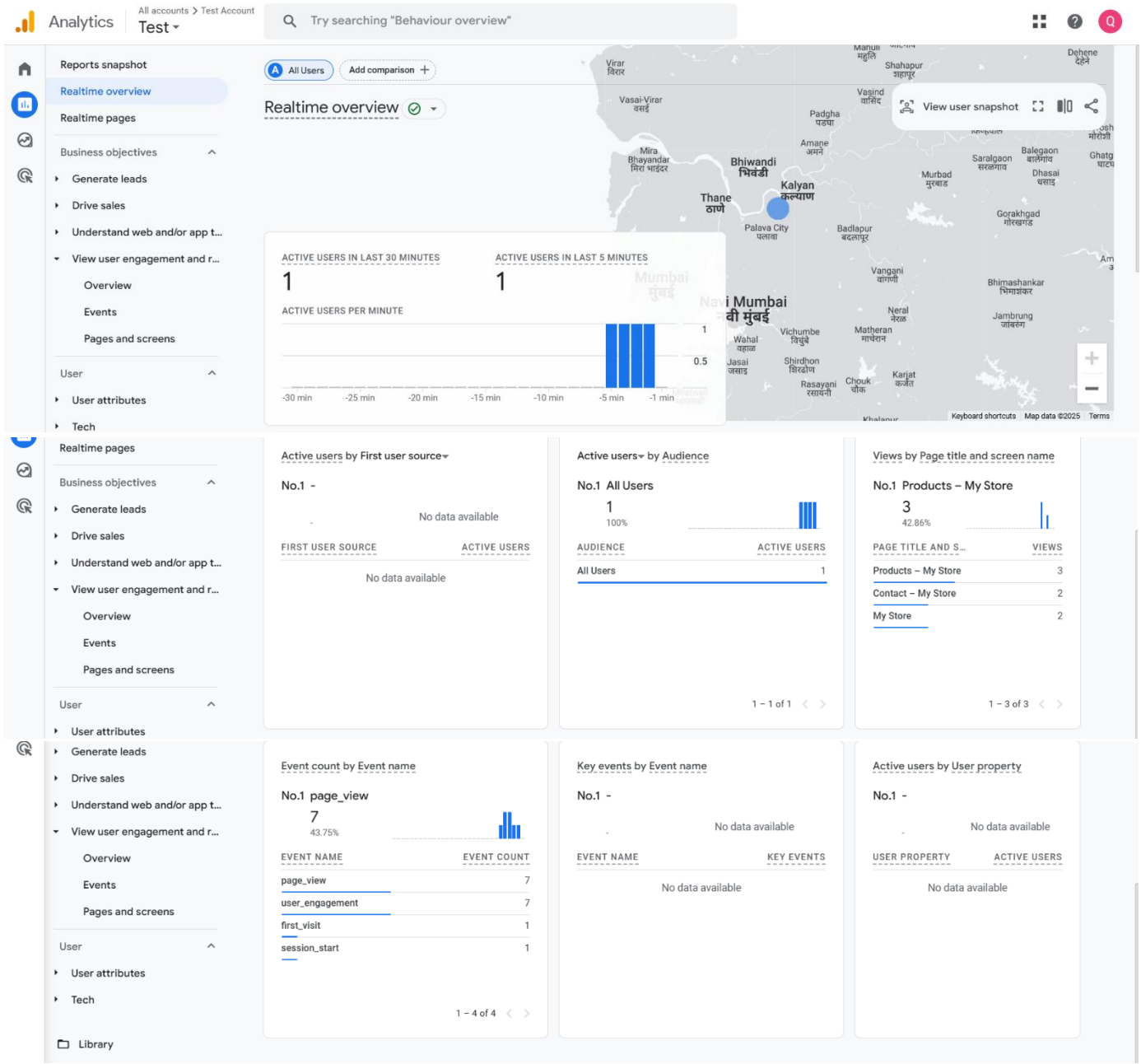


Figure 2 - Realtime Overview of Analytics

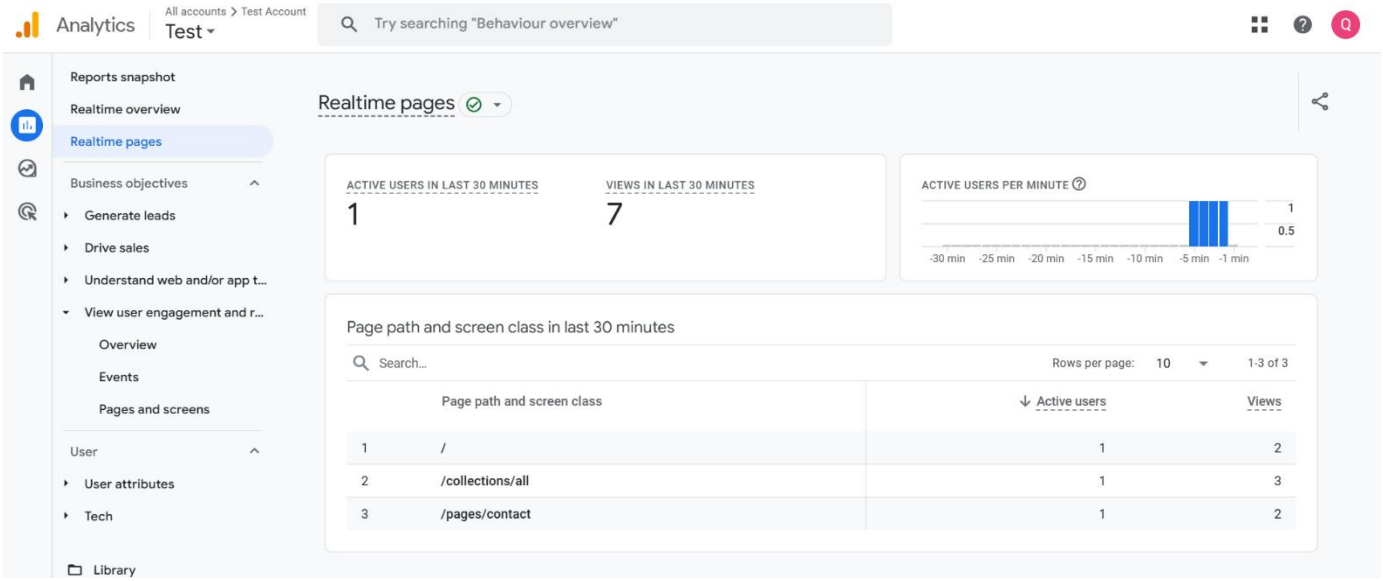
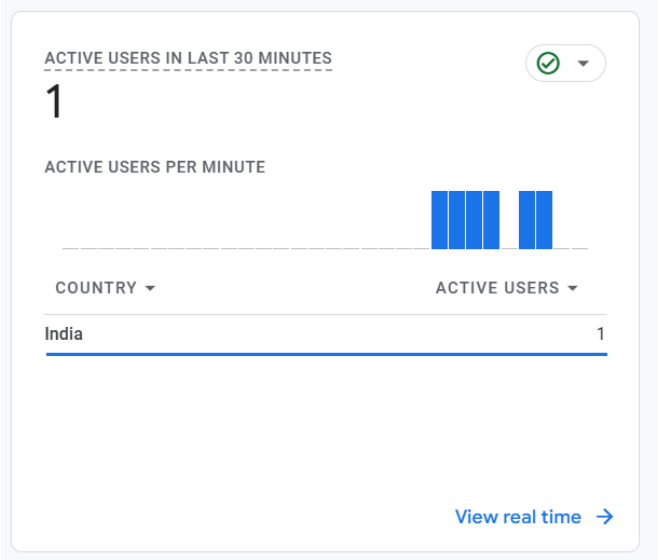


Figure 3 – Realtime Pages



<u>Conclusion</u>:

Social media analytics helps businesses understand audience behavior, optimize content, and improve engagement. Various platforms and tools like **Google Analytics, Facebook Insights, and Communalytic** provide valuable insights for data-driven decisions. Engagement metrics at **page, post, and member levels** aid in refining marketing strategies. Overall, social media analytics enhances brand presence, customer interactions, and business growth.

NAME:- AKASH VERMA I ROLL NO :-- BEA126 I SMA

Experiment No: - 2

Aim: Data Collection-Select the social media platforms of your choice (Twitter, Facebook, LinkedIn, YouTube, Web blogs etc.), connect to and capture social media data for business (scraping, crawling, parsing).

Theory:

In the digital era, vast amounts of data are available on the internet. Extracting and analyzing this data is crucial for various applications, including business intelligence, market research, and sentiment analysis. Three primary techniques used for data extraction are **Scraping, Crawling, and Parsing**.

1. Scraping

Definition:

Scraping is the process of extracting specific data from digital content, such as websites, PDFs, or APIs. It involves retrieving raw data and filtering useful information for storage or analysis. Web scraping is the process of extracting specific data from websites using automated scripts or programs. This method involves sending HTTP requests to web pages, retrieving the HTML content, and parsing it to extract useful information such as text, images, links, and tables. Most of this data is unstructured data in an HTML format which is then converted into structured data in a spreadsheet or a database so that it can be used in various applications. There are many different ways to perform web scraping to obtain data from websites. These include using online services, particular API's or even creating your code for web scraping from scratch. Many large websites, like Google, Twitter, Facebook, StackOverflow, etc. have API's that allow you to access their data in a structured format.

How It Works:

1. **Send a Request** → A request is made to a data source (e.g., a webpage or API).
2. **Retrieve Data** → The response contains structured or unstructured content.
3. **Extract Information** → The required elements (text, images, prices, etc.) are identified and extracted.
4. **Store Data** → The extracted information is saved in a structured format like CSV, JSON, or a database.

Common Tools:

- **BeautifulSoup** (for HTML parsing)
- **Scrapy** (a Python scraping framework)
- **Selenium** (for handling dynamic content)

Use Cases:

- Extracting product prices from e-commerce websites.
- Collecting news articles for sentiment analysis.
- Gathering stock market data for analysis.

2. Crawling

Definition:

Crawling is the process of systematically navigating through multiple data sources by following links, directories, or network paths. The main goal is to discover, index, and retrieve data from different sources. Web crawling is the process of systematically browsing the internet and indexing information from multiple web pages. Unlike scraping, which extracts data from specific pages, crawling is focused on discovering and

following links to access multiple interconnected pages. A web crawler, spider, or search engine bot downloads and indexes content from all over the Internet. The goal of such a bot is to learn what (almost) every webpage on the web is about, so that the information can be retrieved when it's needed. They're called "web crawlers" because crawling is the technical term for automatically accessing a website and obtaining data via a software program.

How It Works:

1. **Start from a Seed URL** → A predefined starting point is used.
2. **Retrieve and Analyze Content** → The structure of the data is examined.
3. **Extract Links** → New links or data sources are identified and added to a queue.
4. **Repeat the Process** → The crawler continues exploring until the desired depth is reached.

Common Tools:

- **Scrapy** (for large-scale crawling)
- **Heritrix** (for search engine crawling)
- **Selenium** (for dynamic page navigation)

Use Cases:

- Search engines indexing web pages (Google, Bing).
- Collecting research articles from multiple online sources.
- Monitoring competitors' websites for updates.

3. Parsing

Definition:

Parsing is the process of analyzing, structuring, and transforming extracted data into a readable and usable format. It is commonly used to process HTML, JSON, XML, and other structured data formats. Data parsing is the process of transforming a sequence (unstructured data) into a tree or parse tree (structured data) that's easier to read, understand and use. This process can be further divided into two steps or components: 1) lexical analysis and 2) syntactic analysis. The lexical analysis takes a sequence of characters (unstructured data) and transforms it into a series of tokens. In other words, the Parser uses a lexer to "turn the meaningless string into a flat list of things like "number literal," "string literal," "identifier," or "operator," and can do things like recognizing reserved identifiers ("keywords") and discarding whitespace." Finally, in the syntactic analysis, a parser takes these tokens and arranges them into a parse tree, establishing elements (nodes) and branches (the relationship between them).

How It Works:

1. **Receive Raw Data** → Extracted content is obtained.
2. **Identify Structure** → The format (tags, keys, attributes) is analyzed.
3. **Extract Relevant Information** → Useful data is selected and processed.
4. **Convert to Structured Format** → Data is organized in a table, list, dictionary, or database.

Common Tools:

- **BeautifulSoup** (for HTML parsing)
- **json module in Python** (for parsing API responses)
- **ElementTree / lxml** (for XML parsing)

Use Cases:

- Extracting metadata from research papers.
- Parsing product details from JSON API responses.
- Processing RSS feeds for news aggregation.

Comparison: Scraping vs. Crawling vs. Parsing

Feature	Web Scraping	Web Crawling	Parsing
Purpose	Extracts data from specific web pages	Discovers and indexes multiple pages	Processes and structures extracted data
Scope	Focuses on targeted pages	Covers entire websites or domains	Works with HTML, JSON, XML, etc.
Output Format	CSV, JSON, databases	Indexed data, sitemaps	Structured tables, dictionaries
Example	Scraping product details from Amazon	Crawling blogs to collect URLs	Parsing an API response to extract weather data

Applications of Web Scraping, Crawling, and Parsing

1. **E-Commerce & Price Monitoring**
    - Tracking competitor prices and discounts on online stores.
    - Monitoring price fluctuations in real time.
2. **Market Research & Sentiment Analysis**
    - Analyzing product reviews and customer feedback.
    - Collecting social media posts and news articles.
3. **SEO & Digital Marketing**
    - Monitoring keyword rankings and website performance.
    - Extracting backlinks and competitor insights.
4. **Academic & Research Data Collection**
    - Extracting research papers, citations, and publications.
    - Collecting climate, financial, or healthcare data for analysis.
5. **Real-Time Data Analysis & AI Training**
    - Fetching live stock market data for machine learning models.
    - Gathering sports statistics for predictive analytics.

Program & Output:

1) API-Based Data Collection –

```
!pip install google-api-python-client pandas
# Install necessary libraries
!pip install google-api-python-client pandas

# Import libraries
from googleapiclient.discovery import build
import pandas as pd

# Step 1: Set up the API key and build the YouTube service
api_key = "AIzaSyBtOc7llpmrKVoh_-knqndIlLmc6ogjils"  # Replace with your YouTube API key
youtube = build("youtube", "v3", developerKey=api_key)
```

```
# Step 2: Define the search query and parameters
query = "Social Media Analytics"  # Replace with your keyword
max_results = 50  # Number of results to fetch

# Fetch data using the YouTube API
request = youtube.search().list(
    q=query,
    part="snippet",
    type="video",
    maxResults=max_results
)
response = request.execute()

# Step 3: Process and organize the data
video_data = []
for item in response['items']:
    video_data.append({
        "Video Title": item['snippet']['title'],
        "Channel Title": item['snippet']['channelTitle'],
        "Published Date": item['snippet']['publishedAt'],
        "Description": item['snippet']['description'],
        "Video ID": item['id']['videoId'],
        "Video URL": f"https://www.youtube.com/watch?v={item['id']['videoId']}"
    })

# Step 4: Convert data into a DataFrame
df = pd.DataFrame(video_data)

# Step 5: Save the data to a CSV file
csv_file_name = "youtube_data.csv"
df.to_csv(csv_file_name, index=False)

# Step 6: Download the CSV file (optional)
from google.colab import files
files.download(csv_file_name)

# Display the first few rows of the DataFrame
df.head()
```

	Video Title	Channel Title	Published Date	Description	Video ID	Video URL
0	Social media analytics and reporting | Google ...	Google Career Certificates	2022-05-31T16:00:00Z	This video is part of the Google Digital Marke...	aEsWltLmPfc	https://www.youtube.com/watch?v=aEsWltLmPfc
1	Analytics Made Simple: Your TOP 5 Social Media...	Delaware ShoutOut - LinkedIn Strategies	2021-04-13T14:00:06Z	BUILD A LINKEDIN GROWTH SYSTEM that attracts y...	Cw22bCgrdLU	https://www.youtube.com/watch?v=Cw22bCgrdLU
2	How To Create A Social Media Analytics Report ...	Sugarpunch Marketing	2023-08-26T15:00:14Z	My Metrics Masterclass will help you read your...	NwZy0evRnZs	https://www.youtube.com/watch?v=NwZy0evRnZs
3	Social Media Analytics	The Career Force	2020-08-25T17:00:08Z	SOCIAL MEDIA ANALYTICS // If you're working in...	Z1KJ-16Rfs0	https://www.youtube.com/watch?v=Z1KJ-16Rfs0
4	Social Media Analytics - Explained	B2Bwhiteboard	2012-12-06T16:52:40Z	Social media analytics combines traditional bu...	0oX_zi0AB98	https://www.youtube.com/watch?v=0oX_zi0AB98

	Video Title	Channel Title	Published Date	Description	Video ID	Video URL
2	Social media analytics and reporting | Google Digital Marketing & E-commerce Certificate	Google Career Certificates	2022-05-31T16:00:00	This video is part of the Google Digital Marketing & E-commerce Certificate. If you're curious about the latest trends in t	aEsWltLmPfc	https://www.youtube.com/watch?v=aEsWltLmPfc
3	Analytics Made Simple: Your TOP 5 Social Media Metrics!	Delaware ShoutOut - LinkedIn Strategies	2021-04-13T14:00:06	BUILD A LINKEDIN GROWTH SYSTEM that attracts your ideal clients! Book a private, one-hour consultation with me h	Cw22bCgrdLU	https://www.youtube.com/watch?v=Cw22bCgrdLU
4	How To Create A Social Media Analytics Report For Clients: KPIs, Metrics, Insights, Reporting	Sugarpunch Marketing	2023-08-26T15:00:14	My Metrics Masterclass will help you read your client's insights, select metrics to use as key performance indicators, cli	NwZy0evRnZs	https://www.youtube.com/watch?v=NwZy0evRnZs
5	Social Media Analytics	The Career Force	2020-08-25T17:00:08	SOCIAL MEDIA ANALYTICS // If you're working in a marketing analyst role, supporting a marketing team, or even runni	Z1KJ-16Rfs0	https://www.youtube.com/watch?v=Z1KJ-16Rfs0
6	Social Media Analytics - Explained	B2Bwhiteboard	2012-12-06T16:52:40	Social media analytics combines traditional business data with social media data in meaningful ways results in analytic	0oX_zi0AB98	https://www.youtube.com/watch?v=0oX_zi0AB98
7	How to Create a Social Media Analytics Report in 2024	Milou Pietersz	2024-04-10T12:00:50	How to Create a Social Media Analytics Report in 2024 Are you a social media manager wondering how to easily make a	tRycTVgcBlo	https://www.youtube.com/watch?v=tRycTVgcBlo
8	How To Create A Social Media Analytics Report (Cheatsheet Included!)	Milou Pietersz	2023-06-08T12:00:12	How To Create A Social Media Analytics Report (Cheatsheet Included!) Looking for a social media analytics report free	EuGPa5GacoE	https://www.youtube.com/watch?v=EuGPa5GacoE
9	Social Media Metrics You Should Know	Hootsuite	2023-03-20T16:35:05	Successful social media lives at the intersection of art and science. While creativity can get eyes on your social, long-te	0RAvd-ngBEw	https://www.youtube.com/watch?v=0RAvd-ngBEw
10	How to Make $10K/Month Selling AI-Powered Social Media Marketing! [E28]	Cortez Hustle - PLR Profit Society	2025-02-28T05:45:11	Small businesses are STRUGGLING with social media… and YOU can get paid to fix it. 90% of businesses use social	rqvOYo8-m2w	https://www.youtube.com/watch?v=rqvOYo8-m2w
11	Social Media Marketing â€" Analytics	TutorialsPoint	2018-01-24T10:39:26	Social Media Marketing â€" Analytics Watch More Videos at: https://www.tutorialspoint.com/videotutorials/index.htm Lec	pbMRaEF-Dww	https://www.youtube.com/watch?v=pbMRaEF-Dww
12	An introduction to Social Media Analytics	UNext Learning	2016-11-25T06:51:43	This video gives a brief introduction to social media analytics. Social media is a goldmine of statistics and data, and sc	Ygz58Kjkme4	https://www.youtube.com/watch?v=Ygz58Kjkme4
13	Simplified%E2%80%93s Social Media Analytics	Simplified HQ	2022-06-13T21:32:46	Simplified now has built in social media analytics tools that makes it easy to measure your performance on Facebook, L	DQj5grEW8lk	https://www.youtube.com/watch?v=DQj5grEW8lk
14	5 Must-Have Social Media Analytics Insights to Grow Your Strategy	Sprout Social	2023-05-25T15:00:55	The insights you get from social media analytics have the potential to grow not just your strategy, but your entire busine	elHhL6DhCzo	https://www.youtube.com/watch?v=elHhL6DhCzo
15	How I track my social media analytics (answering my own top 3 questions!) / FREE TEMPLATE	Hootsuite Labs	2023-04-21T17:23:52	I manage six different social media platforms and this is how I check my social media analytics every day. I usually ant	tDGG7iVY7AE	https://www.youtube.com/watch?v=tDGG7iVY7AE
16	Introduction to Social Media Analytics â¥	Perfect Computer Engineer	2024-03-13T11:40:05	This is an Introduction to Social Media Analytics in Hindi. In this lecture I have talked about What is Social Media Analy	i07Rvep682BU	https://www.youtube.com/watch?v=i07Rvep682BU
17	7 Best Tools for Social Media Analytics	Be Productive	2024-04-13T14:06:37	This video shows the top 7 social media analytics tools, discussing their unique features, benefits, and how they can h	qMvT9OTWAFE	https://www.youtube.com/watch?v=qMvT9OTWAFE
18	Best Social Media Analytics Tools: Perfect For Agencies â¥RÌE	Metricool	2024-12-23T19:06:11	Get ready to ELEVATE your social media game by knowing how to analyze the RIGHT metrics and have the essential t	ByuzRbtgTRM	https://www.youtube.com/watch?v=ByuzRbtgTRM
19	What is social media analytics tool?	RADAAR	2022-04-18T06:02:46	Create meaningful #reports to #measure #performance and get recommendations to grow your #reach, #engagement,	P0iOSC-Ttr8	https://www.youtube.com/watch?v=P0iOSC-Ttr8
20	Social Media Analytics Overview\Social media analytics use cases and landscape	Unfold Data Science	2020-06-21T11:56:11	Social Media Analytics Overview\Social media analytics use cases and landscape #SocialMediaAnalytics #UnfoldData	rYkjTw2vbLs	https://www.youtube.com/watch?v=rYkjTw2vbLs

2) Library-based Data Collection –

```
!pip install pytube
from pytube import Search

# Perform a search
search_query = "Artificial Intelligence"
search = Search(search_query)

# Get top 10 results
videos = search.results[:10]

# Display video titles and URLs
for video in videos:
    print(f"Title: {video.title}, URL: {video.watch_url}")
```

```
Title: Google’s AI Course for Beginners (in 10 minutes)!, URL: https://youtube.com/watch?v=Yq0QkCxoTHM
Title: DeepSeek Is About to SHOCK THE WORLD With R2 That’s 40X More Efficient Than OpenAI's AI, URL: https://youtube.com/watch?v=tBG2ZfKYwao
Title: Harvard CS50’s Artificial Intelligence with Python – Full University Course, URL: https://youtube.com/watch?v=5NgNicANyqM
Title: What is Artificial Intelligence? | Artificial Intelligence In 5 Minutes | AI Explained | Simplilearn, URL: https://youtube.com/watch?v=uMzUB89uSxU
Title: I asked Artificial Intelligence why Tesla sales CRASHED 50% in Europe, URL: https://youtube.com/watch?v=Sp9sBTgVSkw
Title: What is Artificial Intelligence? with Mike Wooldridge, URL: https://youtube.com/watch?v=D2JY38VShxI
Title: What Is AI? | Artificial Intelligence | What is Artificial Intelligence? | AI In 5 Mins |Simplilearn, URL: https://youtube.com/watch?v=ad79nYk2keg
Title: The 10 Stages of Artificial Intelligence, URL: https://youtube.com/watch?v=tFx_UNW9I1U
Title: Bill Lee on Tennessee and the future of artificial intelligence, URL: https://youtube.com/watch?v=QJFDm84e20U
Title: Artificial Intelligence | 60 Minutes Full Episodes, URL: https://youtube.com/watch?v=aZ5EsdnpLMI
```

3) Web Scraping –

```
import yt_dlp
import pandas as pd

# YouTube video URL
video_url = "https://youtu.be/aEsWltLmPfc?si=phVJxEi2B4wcDtr-"  # Replace with any video link

# Configure yt_dlp options to extract comments
ydl_opts = {
    "quiet": True,
```

```
    "skip_download": True,
    "extract_flat": True,
    "getcomments": True,
}

# Extract comments
with yt_dlp.YoutubeDL(ydl_opts) as ydl:
    info_dict = ydl.extract_info(video_url, download=False)

comments = []
if "comments" in info_dict:
    for comment in info_dict["comments"][:10]:  # Limit to 10 comments
        comments.append([comment["author"], comment["text"]])

# Convert to DataFrame
df = pd.DataFrame(comments, columns=["Author", "Comment"])

# Display first few rows
print(df.head())

# Save as CSV
df.to_csv("youtube_comments.csv", index=False)
```

	Author	Comment
0	@buenapatria3072	thank you for your kindness, Google
1	@buenapatria3072	i love the music that google has; it reminds m...
2	@ukutegbejennifer4293	Excellent video
3	@AkhiAfroze-h2c	Insightful and excellent service thanks
4	@afsana2595	Excellent video.Thanks

4) Web Crawling –

```
!pip install requests beautifulsoup4
import requests
from bs4 import BeautifulSoup
import pandas as pd

# Base URL
base_url = "http://quotes.toscrape.com/page/{}/"

# Store data
quotes_list = []
```

```
# Scrape first 5 pages
for page in range(1, 6):
    response = requests.get(base_url.format(page))
    soup = BeautifulSoup(response.text, "html.parser")

    # Find all quote containers
    quotes = soup.find_all("div", class_="quote")

    for quote in quotes:
        text = quote.find("span", class_="text").get_text()
        author = quote.find("small", class_="author").get_text()
        tags = [tag.get_text() for tag in quote.find_all("a", class_="tag")]

        quotes_list.append({"Quote": text, "Author": author, "Tags": ", ".join(tags)})

# Convert to DataFrame
df = pd.DataFrame(quotes_list)

# Save to CSV
df.to_csv("quotes.csv", index=False)

# Display first 5 results
df.head()
```

	Quote	Author	Tags
0	“The world as we have created it is a process ...	Albert Einstein	change, deep-thoughts, thinking, world
1	“It is our choices, Harry, that show what we t...	J.K. Rowling	abilities, choices
2	“There are only two ways to live your life. On...	Albert Einstein	inspirational, life, live, miracle, miracles
3	“The person, be it gentleman or lady, who has ...	Jane Austen	aliteracy, books, classic, humor
4	“Imperfection is beauty, madness is genius and...	Marilyn Monroe	be-yourself, inspirational

5) Parsing –

```
!pip install requests
import requests

# OpenWeatherMap API endpoint
api_key = "bd5e378503939ddaee76f12ad7a97608"
city = "Mumbai"
api_url = f"http://api.openweathermap.org/data/2.5/weather?q={city}&appid={api_key}&units=metric"
```

```
# Fetch data from API
response = requests.get(api_url)

# Check if the request was successful
if response.status_code == 200:
    weather_data = response.json()  # Convert JSON response to Python dictionary

    # Extract key details
    city_name = weather_data["name"]
    temperature = weather_data["main"]["temp"]
    weather_desc = weather_data["weather"][0]["description"]
    humidity = weather_data["main"]["humidity"]
    wind_speed = weather_data["wind"]["speed"]

    # Display parsed data
    print(f" City: {city_name}")
    print(f" Temperature: {temperature}°C")
    print(f" Weather: {weather_desc.capitalize()}")
    print(f" Humidity: {humidity}%")
    print(f" Wind Speed: {wind_speed} m/s")

else:
    print("Failed to fetch weather data. Status Code:", response.status_code)
```

```
 City: Mumbai
 Temperature: 26.43°C
 Weather: Broken clouds
 Humidity: 55%
 Wind Speed: 4.25 m/s
```

Conclusion:

In this experiment, we successfully demonstrated scraping, crawling, and parsing techniques for extracting and processing web data. Scraping helped retrieve specific information, crawling enabled systematic data discovery, and parsing structured the extracted content for analysis. These methods are essential for automating data collection, improving business insights, and enhancing decision-making processes.

NAME:-AKASH VERMA I ROLL NO :-- BEA126 I SMA

Experiment No: - 3

Aim: Data Cleaning and Storage - Preprocess, filter and store social media data for business (Using Python, MongoDB, R, etc).

Theory:

Social media platforms generate an enormous amount of data every second. This data is highly valuable for businesses to understand customer preferences, sentiment trends, and market dynamics. However, raw social media data is often unstructured, noisy, and contains unwanted elements such as duplicate records, missing values, irrelevant information, and spam content. To extract meaningful business insights, it is crucial to perform data cleaning, filtering, and proper storage. This experiment focuses on processing social media data from sources like Twitter, Facebook, or Instagram, removing inconsistencies, and storing it efficiently for further analysis.

Importance of Data Cleaning & Storage –

1. **Ensuring Data Quality:** Raw data often contains errors such as duplicate records, spelling mistakes, and unwanted symbols. Cleaning helps maintain data consistency and accuracy.
2. **Improved Decision Making:** Businesses rely on structured and cleaned data to make informed decisions based on accurate insights.
3. **Efficient Storage & Retrieval:** Storing cleaned data in structured formats (such as databases or CSV files) enables quick retrieval for future analysis.
4. **Enhanced Machine Learning Performance:** Cleaned data improves the accuracy of machine learning models used for sentiment analysis, customer behavior prediction, and trend forecasting.
5. **Business Insights & Competitive Advantage:** Social media data allows businesses to track brand reputation, analyze customer feedback, and monitor industry trends.

1. Data Collection

Definition

Data collection refers to the process of **gathering raw data from various social media sources** for further analysis. This includes user posts, comments, likes, shares, hashtags, mentions, and engagement metrics.

Sources of Data

- **Social Media APIs:** Platforms like Twitter, Facebook, and Instagram provide APIs to fetch real-time user data.
- **Web Scraping:** If APIs are limited, web scraping techniques (BeautifulSoup, Scrapy) are used to extract social media data.
- **Third-Party Data Providers:** Some companies offer pre-collected and structured social media datasets.

Types of Data Collected

- **Textual Data:** Comments, reviews, tweets, posts, captions.
- **Engagement Metrics:** Likes, shares, retweets, followers, replies.
- **Multimedia Data:** Images, videos, GIFs, and memes.
- **Metadata:** Timestamp, geolocation, user information, hashtags.

Challenges in Data Collection

1. **Data Privacy & Ethical Issues:** Many platforms impose restrictions on user data access to ensure privacy.
2. **API Rate Limits:** APIs have restrictions on the number of requests that can be made per hour.
3. **Unstructured Data Format:** Data collected is often inconsistent and requires cleaning.
4. **Spam & Fake Data:** Bots and fake accounts generate irrelevant content that affects analysis.

2. Data Cleaning

Definition

Data cleaning (also known as data preprocessing) is the process of **removing inconsistencies, handling missing values, and ensuring data quality** before analysis.

Steps in Data Cleaning

1. **Handling Missing Values:** Social media data often has incomplete records. Missing values can be:
    - Removed (if a significant portion is missing).
    - Imputed using statistical techniques (e.g., mean or median substitution).
2. **Removing Duplicates:** Identical posts or comments can appear multiple times, leading to redundancy. Duplicate records should be removed.
3. **Eliminating Unwanted Characters:**
    - **URLs:** Links in social media posts are often irrelevant to textual analysis and are removed.
    - **Special Characters & Emojis:** Non-text characters (e.g., #, @, emojis) can be removed unless required for analysis.
4. **Tokenization & Stopword Removal:**
    - **Tokenization:** Splitting text into individual words.
    - **Stopword Removal:** Eliminating common words that do not add meaning (e.g., "is", "the", "and").
5. **Stemming & Lemmatization:**
    - **Stemming:** Reducing words to their root form (e.g., "running" → "run").
    - **Lemmatization:** Converting words to meaningful base forms (e.g., "better" → "good").

3. Data Filtering

Definition

Data filtering involves **extracting relevant information** based on business objectives while removing unnecessary or irrelevant content.

Techniques of Data Filtering

1. **Keyword-Based Filtering:**
    - Extracting posts that mention specific topics, brands, or products.
2. **Sentiment Filtering:**
    - Using sentiment analysis to categorize text as **positive, negative, or neutral** based on customer opinions.
3. **Language Filtering:**
    - Removing posts in languages that are not relevant to the analysis.
4. **Spam & Bot Detection:**
    - Identifying and removing posts generated by automated bots.

4. Data Storage

Definition

After cleaning and filtering, the processed data must be **stored efficiently** for further analysis.

Storage Formats & Databases

1. **CSV Files:** Used for small datasets that need simple tabular storage.
2. **SQL Databases (MySQL, PostgreSQL):** Suitable for structured, relational data.
3. **NoSQL Databases (MongoDB):** Ideal for storing large-scale, unstructured social media data.

Advantages of Storing Data in MongoDB

- **Flexible Schema:** Can handle dynamic social media data formats.
- **Scalability:** Efficiently stores large datasets.
- **Fast Retrieval:** Allows quick queries on large datasets.

Program & Output:

```
!pip install praw nltk textblob pandas pymongo

import praw
import re
import pandas as pd
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from textblob import TextBlob
from pymongo import MongoClient
import json
import time

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

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
True
```

```
# Replace with your Reddit API credentials
reddit = praw.Reddit(
    client_id="QTHmEfFkRMZd5-c8NZC61A",
    client_secret="PqZWNXvmVaASoxYiaii6eLss3BansA",
    user_agent="Social Media Analytics Test"
)

def fetch_reddit_posts(subreddit_name, query, limit=10):
    posts = []
    subreddit = reddit.subreddit(subreddit_name)

    for post in subreddit.search(query, limit=limit):
        posts.append(post.title + " " + post.selftext)
```

```
    return posts

# Example: Fetch 10 posts related to "business analytics" from r/dataanalysis
raw_posts = fetch_reddit_posts("dataanalysis", "business analytics", limit=10)
print("Fetched Posts:", raw_posts[:5])  # Print first 5 posts
```

```
WARNING:praw:It appears that you are using PRAW in an asynchronous environment.
It is strongly recommended to use Async PRAW: https://asyncpraw.readthedocs.io.
See https://praw.readthedocs.io/en/latest/getting_started/multiple_instances.html#discord-bots-and-asynchronous-environments for more info.

Fetched Posts: ['Whats a good book for learning key concepts in data/business analytics? Ive found a few books but im not sure if they are any good. I am doing a degree with a minor
in data analytics, but I feel like the course isnt teaching me enough. I just wanted to master the concepts. ', 'Do third world countries recognize or incorporate data analytics in
```

```
import re
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords

# Download necessary NLTK data
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('punkt_tab')

def clean_text(text):
    text = re.sub(r"http\S+|www\S+|https\S+", '', text)  # Remove URLs
    text = re.sub(r'\@\w+|\#', '', text)  # Remove mentions and hashtags
    text = re.sub(r'[^\w\s]', '', text)  # Remove special characters
    text = text.lower()  # Convert to lowercase
    words = word_tokenize(text)
    words = [word for word in words if word not in stopwords.words('english')]  # Remove stopwords
    return " ".join(words)

# Apply cleaning function to all posts
cleaned_posts = [clean_text(post) for post in raw_posts]

print("Cleaned Posts:", cleaned_posts[:5])  # Print first 5 cleaned posts
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt_tab.zip.
Cleaned Posts: ['whats good book learning key concepts databusiness analytics ive found books im sure good degree minor data analytics feel like course isnt teaching enough wanted
```

```
def get_sentiment(text):
    sentiment_score = TextBlob(text).sentiment.polarity
    if sentiment_score > 0:
        return "Positive"
    elif sentiment_score < 0:
        return "Negative"
    else:
        return "Neutral"

# Apply sentiment analysis
sentiments = [get_sentiment(post) for post in cleaned_posts]

# Create a DataFrame
```

```
df = pd.DataFrame({"Original_Post": raw_posts, "Cleaned_Post": cleaned_posts, "Sentiment": sentiments})

print(df.head())  # Display first 5 rows
```

```
                                        Original_Post  \
0  Whats a good book for learning key concepts in...
1  Do third world countries recognize or incorpor...
2  Post Graduate Program in Data Science and Busi...
3  Getting My Master's In Business Analytics, Is ...
4  Ideas for Business Analytics course (MSc level...

                                        Cleaned_Post Sentiment
0  whats good book learning key concepts databusi...  Positive
1  third world countries recognize incorporate da...   Neutral
2  post graduate program data science business an...  Positive
3  getting masters business analytics master prog...  Negative
4  ideas business analytics course msc level hi g...  Positive
```

```
import sqlite3
import pandas as pd
# Connect to SQLite database (Creates 'reddit_posts.db' file)
conn = sqlite3.connect("reddit_posts.db")
cursor = conn.cursor()

# Create table
cursor.execute("""
   CREATE TABLE IF NOT EXISTS cleaned_reddit_posts (
      id INTEGER PRIMARY KEY AUTOINCREMENT,
      original_post TEXT,
      cleaned_post TEXT,
      sentiment TEXT
   )
""")

conn.commit()
print("SQLite Database & Table Created Successfully!")

# Insert DataFrame into SQLite3
for _, row in df.iterrows():
   cursor.execute("INSERT INTO cleaned_reddit_posts (original_post, cleaned_post, sentiment) VALUES (?, ?, ?)",
                  (row["Original_Post"], row["Cleaned_Post"], row["Sentiment"]))

conn.commit()
print("Data successfully stored in SQLite!")

# Fetch data from SQLite
cursor.execute("SELECT * FROM cleaned_reddit_posts LIMIT 5")
rows = cursor.fetchall()
```

```
# Convert to Pandas DataFrame
df_retrieved = pd.DataFrame(rows, columns=["ID", "Original_Post", "Cleaned_Post", "Sentiment"])

# Display retrieved data
print(df_retrieved.head())
```

```
   ID                                       Original_Post  \
0   1  Whats a good book for learning key concepts in...
1   2  Do third world countries recognize or incorpor...
2   3  Post Graduate Program in Data Science and Busi...
3   4  Getting My Master's In Business Analytics, Is ...
4   5  Ideas for Business Analytics course (MSc level...

                                        Cleaned_Post Sentiment
0  whats good book learning key concepts databusi...  Positive
1  third world countries recognize incorporate da...   Neutral
2  post graduate program data science business an...  Positive
3  getting masters business analytics master prog...  Negative
4  ideas business analytics course msc level hi g...  Positive
```

```
df.to_csv("cleaned_reddit_posts.csv", index=False)
print("Data exported to cleaned_reddit_posts.csv!")
```

cleaned_reddit_posts.csv

1 to 10 of 10 entries Filter

Original_Post	Cleaned_Post	Sentiment
Whats a good book for learning key concepts in data/business analytics? Ive found a few books but im not sure if they are any good. I am doing a degree with a minor in data analytics, but I feel like the course isnt teaching me enough. I just wanted to master the concepts.	whats good book learning key concepts databusiness analytics ive found books im sure good degree minor data analytics feel like course isnt teaching enough wanted master concepts	Positive
Do third world countries recognize or incorporate data analytics in their businesses?	third world countries recognize incorporate data analytics businesses	Neutral
Post Graduate Program in Data Science and Business Analytics at The University of Texas at Austin Has anyone done this? The Post Graduate Program in Data Science and Business Analytics offered by McCombs School of Business at The University of Texas at Austin? I just got accepted and am willing to invest if it is worth it and will help me launch a new career. Just wondering if it is a money grab or if it's too advanced for a newbie who has only done the Google data analytics course. The next session started at the end of this month and I applied to some masters programs i won't hear back from before then :/	post graduate program data science business analytics university texas austin anyone done post graduate program data science business analytics offered mccombs school business university texas austin got accepted willing invest worth help launch new career wondering money grab advanced newbie done google data analytics course next session started end month applied masters programs wont hear back	Positive

Conclusion:

In this experiment, we successfully extracted Reddit posts using the Pushshift API, preprocessed the text by removing unwanted elements, and performed sentiment analysis. The cleaned data was efficiently stored in an SQLite database for structured retrieval and analysis. This approach ensures an organized, scalable, and efficient method for processing and analyzing Reddit discussions for further insights.

Experiment No: - 4

Aim: Exploratory Data Analysis and Visualization of Social Media Data for Business.

Theory:

In today's digital era, social media plays a crucial role in shaping consumer behavior, brand perception, and market trends. Businesses increasingly rely on social media data to gain insights into customer sentiment, preferences, and emerging trends. However, raw social media data is often unstructured and requires extensive analysis to extract meaningful insights.

Exploratory Data Analysis (EDA) is a critical step in data analytics that helps in understanding data distribution, identifying patterns, and detecting anomalies before applying predictive models. Data visualization enhances EDA by presenting complex information in an intuitive and interpretable format using various graphical representations.

This experiment focuses on performing **EDA and Visualization** of social media data to derive actionable business insights. The key objectives include:

- Understanding data distributions
- Identifying trends and patterns
- Analyzing sentiment variations
- Detecting anomalies and outliers

2. Data Collection for Social Media Analysis

Social media data can be obtained from platforms like **Twitter, Reddit, Facebook, and Instagram** using APIs. In this experiment, we use **Reddit's API** to fetch posts from relevant business-related subreddits. The collected data includes:

- **Post Text** (Content of the post)
- **Author** (User who posted)
- **Upvotes** (Engagement level)
- **Comments** (Community interaction)
- **Timestamp** (Time when the post was made)

3. Data Preprocessing and Cleaning

Before performing EDA, the raw social media data needs cleaning and preprocessing:

- **Removing URLs**: Posts often contain links that are irrelevant for analysis.
- **Removing Special Characters and Numbers**: Only text-based content is useful for sentiment analysis.
- **Converting Text to Lowercase**: To maintain uniformity in analysis.
- **Removing Stopwords**: Common words (e.g., "the", "and", "is") that do not add meaning are removed.
- **Tokenization**: Splitting text into individual words for word frequency analysis.

After preprocessing, the cleaned data is ready for exploratory analysis.

4. Exploratory Data Analysis (EDA)

EDA is performed to uncover insights from the social media data. Social media data can be collected from various platforms like Twitter, Reddit, Facebook, and Instagram using APIs or publicly available datasets. In this experiment, we will use a dataset containing social media posts, user engagement metrics (likes, shares, comments), timestamps, and sentiment labels. The following analyses are conducted:

a) Descriptive Statistics

- Understanding the **distribution of posts** over time
- Analyzing **engagement metrics** (likes, comments, upvotes)
- Identifying the most **frequent words and topics**

b) Sentiment Analysis

- Sentiment polarity (Positive, Negative, Neutral) is determined using **TextBlob**
- Sentiment distribution across different time periods is analyzed
- Comparing engagement levels between positive and negative posts

c) Word Frequency Analysis

- Finding the most commonly used words in business-related discussions
- Creating **word clouds** to visualize popular terms

d) Time-Based Trends

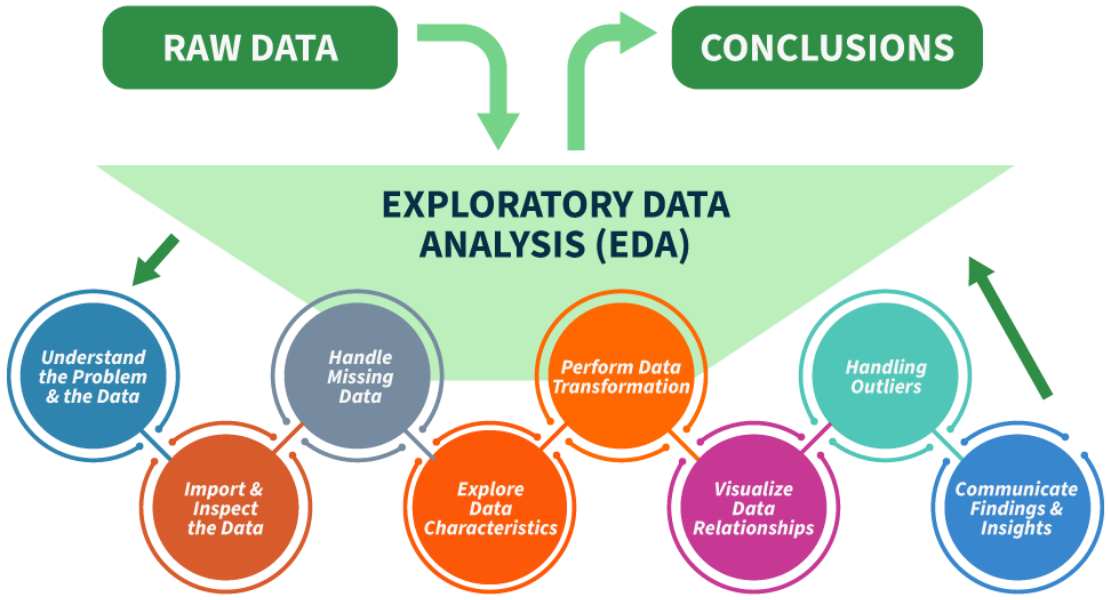
- Analyzing how post engagement varies throughout the day or week
- Detecting seasonal trends in social media discussions

e) Correlation Analysis

- Finding relationships between **engagement metrics** (e.g., Do posts with more upvotes get more comments?)
- Understanding how sentiment affects **user interaction**

Data Summarization – Basic statistical insights can be obtained using **descriptive statistics**, which help understand data distribution. Common EDA techniques include:

- **Value counts:** Checking the frequency of categorical values.
- **Summary statistics:** Mean, median, standard deviation, and outliers detection.
- **Distribution plots:** Histogram, box plots, and density plots to analyze numerical data.



Social media posts can be analyzed to determine user sentiment (positive, negative, or neutral). This helps businesses gauge customer satisfaction and brand perception. Sentiment analysis can be performed using TextBlob or VADER to assign polarity scores to each post.

5. Data Visualization Techniques

Data visualization makes it easier to interpret insights. The following visualization techniques are used:

Visualization	**Purpose**
Bar Charts	Display most popular words, engagement distribution
Word Cloud	Show the most frequently used words
Time Series Plot	Analyze posting trends over time
Pie Chart	Sentiment distribution analysis
Box Plot	Identify outliers in engagement metrics
Heatmaps	Show correlations between different variables

Using **Matplotlib, Seaborn, and WordCloud libraries**, these visualizations help businesses understand customer sentiment, engagement, and discussion trends. **Univariate visualizations** focus on a single variable, helping understand its distribution and characteristics. **Histograms** display the frequency distribution of numerical data, making it easy to identify patterns such as user engagement levels on posts. **Box plots** help detect outliers and variations in numerical data, such as extreme values in the number of likes or shares. **Pie charts** are commonly used to represent categorical data proportions, like sentiment distribution in user comments.

For analyzing **relationships between two variables**, **bivariate visualizations** are useful. **Scatter plots** help assess correlations, such as the relationship between hashtags and engagement. **Bar charts** compare different categorical variables, such as the number of posts across various social media platforms. **Heatmaps** are particularly effective in showing correlations between multiple variables, helping businesses understand which factors influence engagement the most.

Social media trends often change over time, and **time-series visualizations** such as **line charts** track engagement levels over different periods. **Area charts** further emphasize cumulative engagement, allowing businesses to see long-term trends. Text-based data, which is highly prevalent in social media, requires special visualization techniques. **Word clouds** help identify frequently mentioned words in posts, making them useful for detecting trending topics. **Hashtag analysis** provides insights into the most popular hashtags being used, allowing businesses to optimize their marketing strategies.

Additionally, social media platforms rely heavily on networks of users. **Network graphs** help visualize connections between users, such as interactions between influencers and their followers. **Sankey diagrams** provide insight into the flow of information, tracking how a viral post spreads across different users and communities.

Data visualization plays a key role in understanding patterns, trends, and relationships within social media data. Common visualization techniques include:

Sentiment Distribution -

- Bar Charts and Pie Charts are used to visualize the percentage of positive, neutral, and negative posts.
- Helps businesses identify customer sentiment trends and improve customer support strategies.

User Engagement Analysis

- Scatter plots and histograms show the distribution of engagement metrics (likes, shares, retweets).
- Line graphs help track engagement trends over time.

Trend Analysis Over Time

- Time-series plots show the frequency of posts or trending topics over days, weeks, or months.
- Businesses can use this to identify peak engagement periods and optimize marketing campaigns.

Word Cloud and Hashtag Analysis

- Word Clouds visually highlight the most frequently occurring words in posts.
- Bar graphs of top hashtags reveal trending topics and popular discussions

Dataset Overview

https://www.kaggle.com/datasets/bhadramohit/social-media-usage-datasetapplications

This dataset captures user behavior across different social media platforms. Key features typically include:

- **User ID:** Unique identifier for each user.
- **Platform:** The social media application used (e.g., Facebook, Instagram, Twitter).
- **Usage Time:** Duration spent on the platform, often measured in minutes or hours.
- **Number of Posts:** Total posts made by the user.
- **Number of Likes:** Total likes received by the user.
- **Number of Followers:** Count of followers or friends.
- **Engagement Rate:** A metric indicating user interaction, calculated based on likes, comments, shares, etc.

Program & Output:

Exploratory Data Analysis (EDA) –

Data Loading and Preprocessing –

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Load Dataset –

```
df = pd.read_csv('/content/social_media_usage.csv')
```

Inspect Data –

```
print(df.head())
```

```
  User_ID        App  Daily_Minutes_Spent  Posts_Per_Day  Likes_Per_Day  \
0     U_1  Pinterest                  288             16             94
1     U_2   Facebook                  192             14            117
2     U_3  Instagram                  351             13            120
3     U_4     TikTok                   21             20            117
4     U_5   LinkedIn                  241             16              9

   Follows_Per_Day
0                0
1               15
2               48
3                8
4               21
```

print(df.info())

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 6 columns):
 #   Column               Non-Null Count  Dtype 
---  ------               --------------  ----- 
 0   User_ID              1000 non-null   object
 1   App                  1000 non-null   object
 2   Daily_Minutes_Spent  1000 non-null   int64 
 3   Posts_Per_Day        1000 non-null   int64 
 4   Likes_Per_Day        1000 non-null   int64 
 5   Follows_Per_Day      1000 non-null   int64 
dtypes: int64(4), object(2)
memory usage: 47.0+ KB
None
```

df.describe()

	Daily_Minutes_Spent	Posts_Per_Day	Likes_Per_Day	Follows_Per_Day
count	1000.000000	1000.000000	1000.000000	1000.000000
mean	247.368000	10.269000	94.682000	24.698000
std	146.371921	6.121774	57.560943	14.842948
min	5.000000	0.000000	0.000000	0.000000
25%	112.750000	5.000000	44.750000	12.000000
50%	246.000000	10.000000	94.000000	24.000000
75%	380.500000	16.000000	142.000000	38.000000
max	500.000000	20.000000	200.000000	50.000000

Handling Missing Values –

Identify Missing Values –

print(df.isnull().sum())

```
User_ID                0
App                    0
Daily_Minutes_Spent    0
Posts_Per_Day          0
Likes_Per_Day          0
Follows_Per_Day        0
dtype: int64
```

print(df.duplicated().sum())

0

```
df['Daily_Minutes_Spent'] = pd.to_numeric(df['Daily_Minutes_Spent'], errors='coerce')
df['Posts_Per_Day'] = pd.to_numeric(df['Posts_Per_Day'], errors='coerce')
df['Likes_Per_Day'] = pd.to_numeric(df['Likes_Per_Day'], errors='coerce')
df['Follows_Per_Day'] = pd.to_numeric(df['Follows_Per_Day'], errors='coerce')

df['Engagement_Rate'] = (df['Likes_Per_Day'] + df['Follows_Per_Day']) / df['Posts_Per_Day']
print(df[['App', 'Engagement_Rate']].head())  # View new feature
```

```
         App  Engagement_Rate
0  Pinterest         5.875000
1   Facebook         9.428571
2  Instagram        12.923077
3     TikTok         6.250000
4   LinkedIn         1.875000
```

```
follow_to_post_ratio = df.groupby("App")["Follows_Per_Day"].sum() /
df.groupby("App")["Posts_Per_Day"].sum()
print("Follow-to-Post Ratio per Platform:\n", follow_to_post_ratio)
```

```
Follow-to-Post Ratio per Platform:
 App
Facebook     2.427799
Instagram    2.283102
LinkedIn     2.464572
Pinterest    2.332661
Snapchat     2.550641
TikTok       2.556380
Twitter      2.239305
dtype: float64
```

```
post_variability = df.groupby("App")["Posts_Per_Day"].std()
print("Standard Deviation in Posting Frequency per Platform:\n", post_variability)
```

```
Standard Deviation in Posting Frequency per Platform:
 App
Facebook     6.389768
Instagram    5.900180
LinkedIn     5.956119
Pinterest    5.835075
Snapchat     6.188439
TikTok       6.321583
Twitter      6.166848
Name: Posts_Per_Day, dtype: float64
```

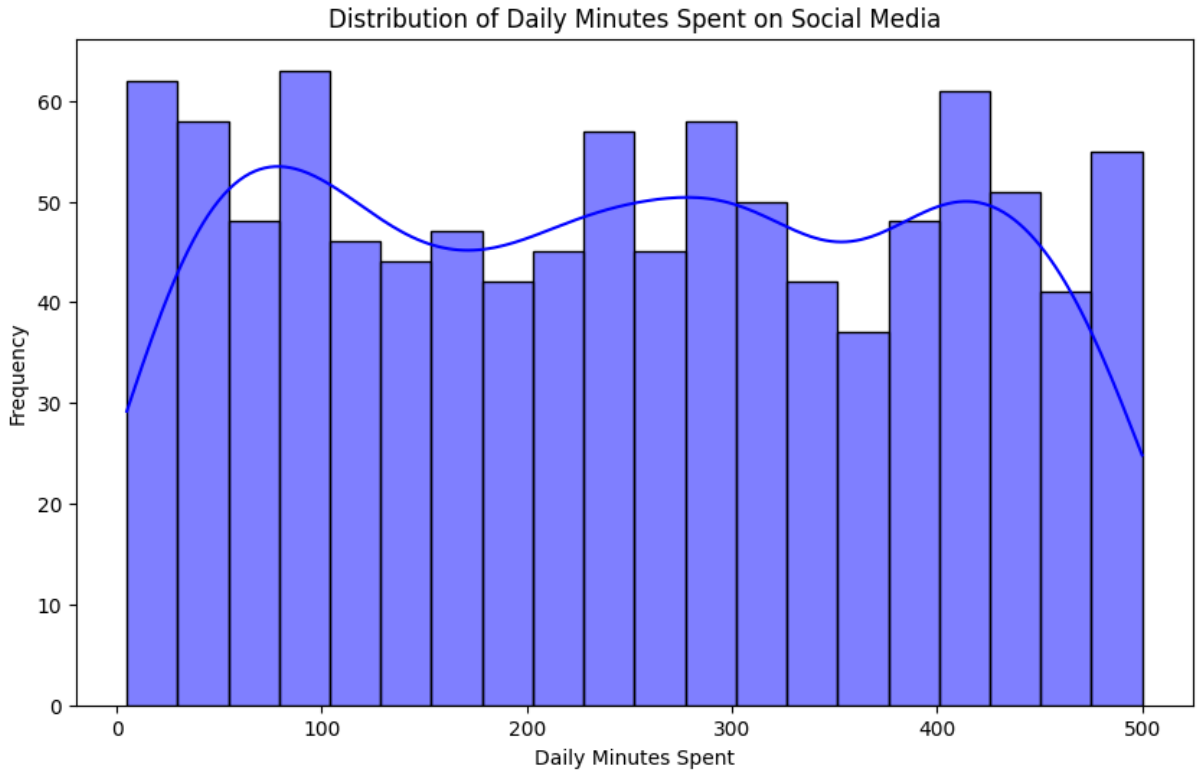
```
ranked_platforms = df.groupby("App")[["Daily_Minutes_Spent", "Posts_Per_Day", "Likes_Per_Day", "Follows_Per_Day"]].mean()
ranked_platforms["Engagement_Score"] = ranked_platforms["Likes_Per_Day"] + ranked_platforms["Follows_Per_Day"]
ranked_platforms = ranked_platforms.sort_values(by="Engagement_Score", ascending=False)
print("Engagement Rankings of Social Media Platforms:\n", ranked_platforms)
```

```
Engagement Rankings of Social Media Platforms:
           Daily_Minutes_Spent  Posts_Per_Day  Likes_Per_Day  Follows_Per_Day  \
App
Instagram           264.100000      11.328571      98.992857        25.864286
Facebook            247.169118       9.522059      99.301471        23.117647
LinkedIn            236.054422      10.176871      97.081633        25.081633
Snapchat            256.551282      10.000000      93.820513        25.506410
Pinterest           236.207143      10.628571      92.114286        24.792857
TikTok              233.418440       9.560284      91.411348        24.439716
Twitter             257.685714      10.685714      90.185714        23.928571

           Engagement_Score
App
Instagram        124.857143
Facebook         122.419118
LinkedIn         122.163265
Snapchat         119.326923
Pinterest        116.907143
TikTok           115.851064
Twitter          114.114286
```

Visualization –

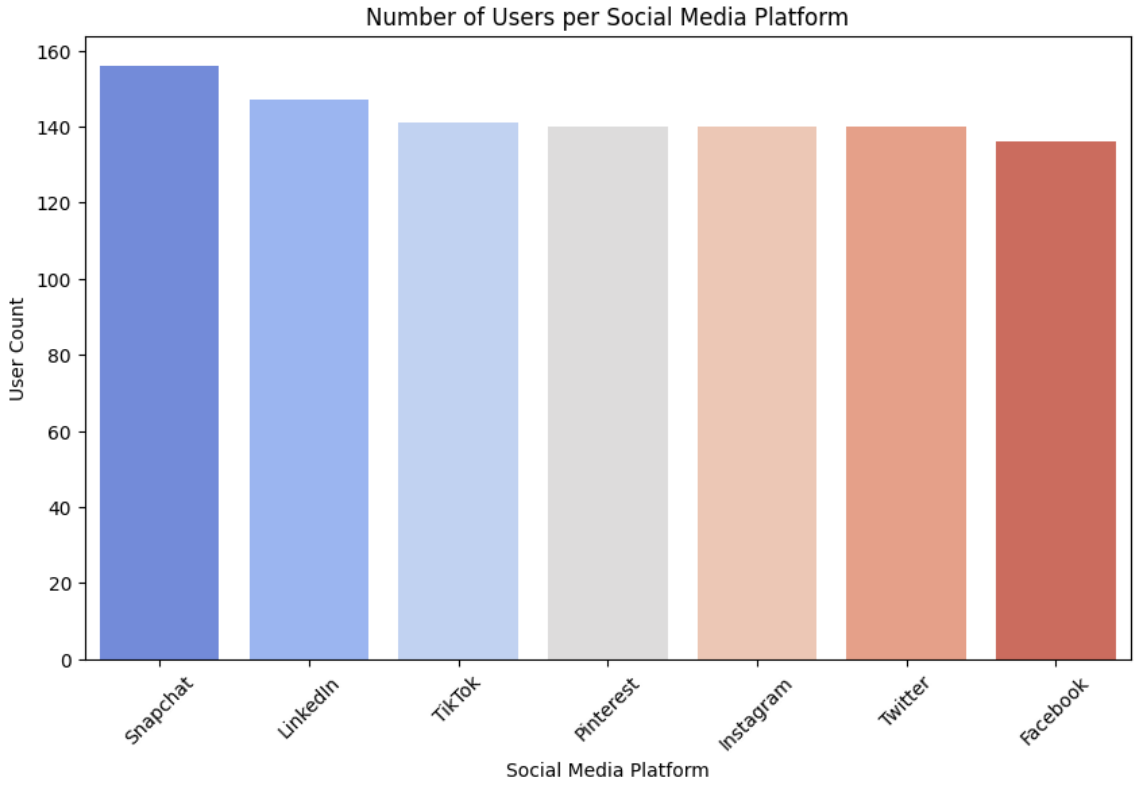
Understanding the Distribution of Daily Minutes Spent

```
plt.figure(figsize=(10, 6))
sns.histplot(df['Daily_Minutes_Spent'], bins=20, kde=True, color='blue')
plt.title("Distribution of Daily Minutes Spent on Social Media")
plt.xlabel("Daily Minutes Spent")
plt.ylabel("Frequency")
plt.show()
```



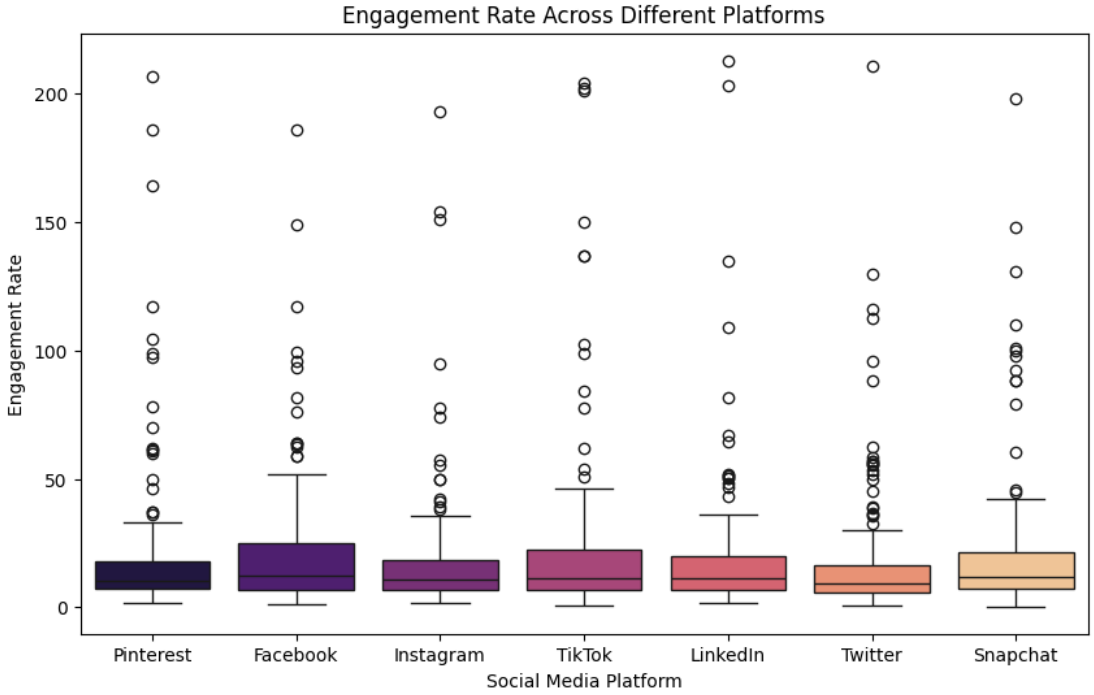
Platform Popularity

```
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='App', order=df['App'].value_counts().index, palette='coolwarm')
plt.title("Number of Users per Social Media Platform")
plt.xlabel("Social Media Platform")
plt.ylabel("User Count")
plt.xticks(rotation=45)
plt.show()
```



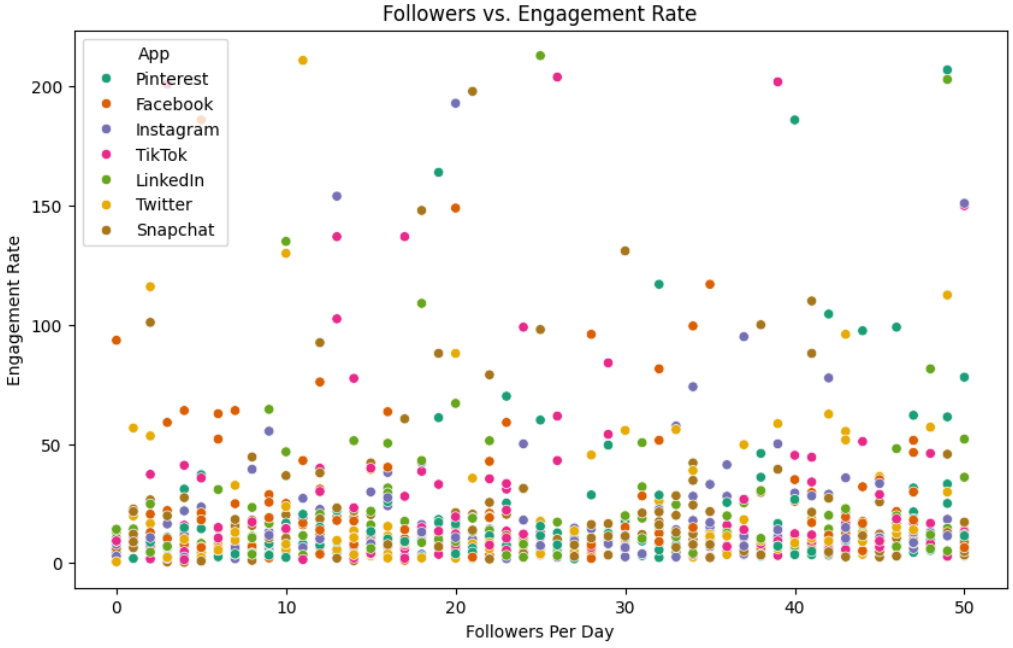
Engagement Rate Across Platforms

```
plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='App', y='Engagement_Rate', palette='magma')
plt.title("Engagement Rate Across Different Platforms")
plt.xlabel("Social Media Platform")
plt.ylabel("Engagement Rate")
plt.show()
```



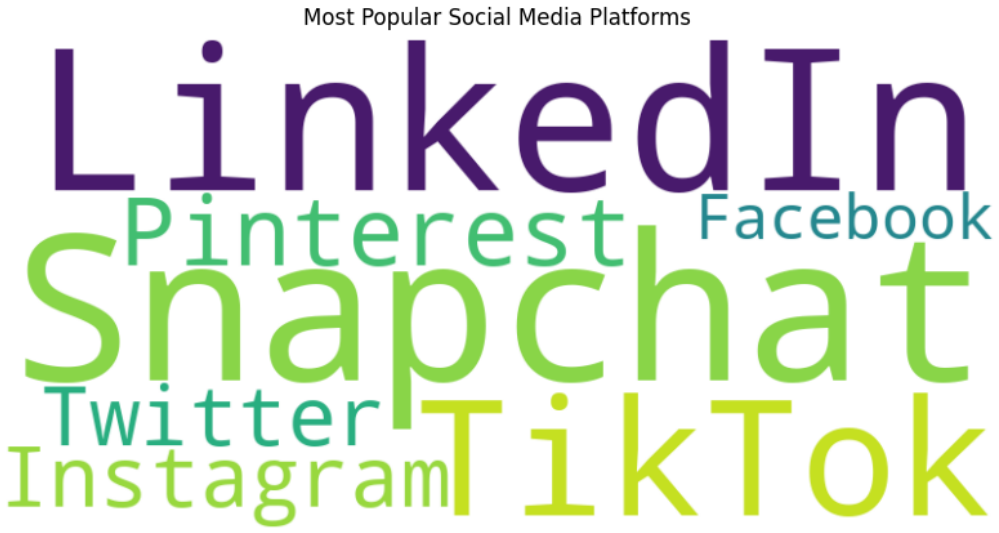
Relationship Between Followers and Engagement

```
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df, x='Follows_Per_Day', y='Engagement_Rate', hue='App', palette='Dark2')
plt.title("Followers vs. Engagement Rate")
plt.xlabel("Followers Per Day")
plt.ylabel("Engagement Rate")
plt.show()
```



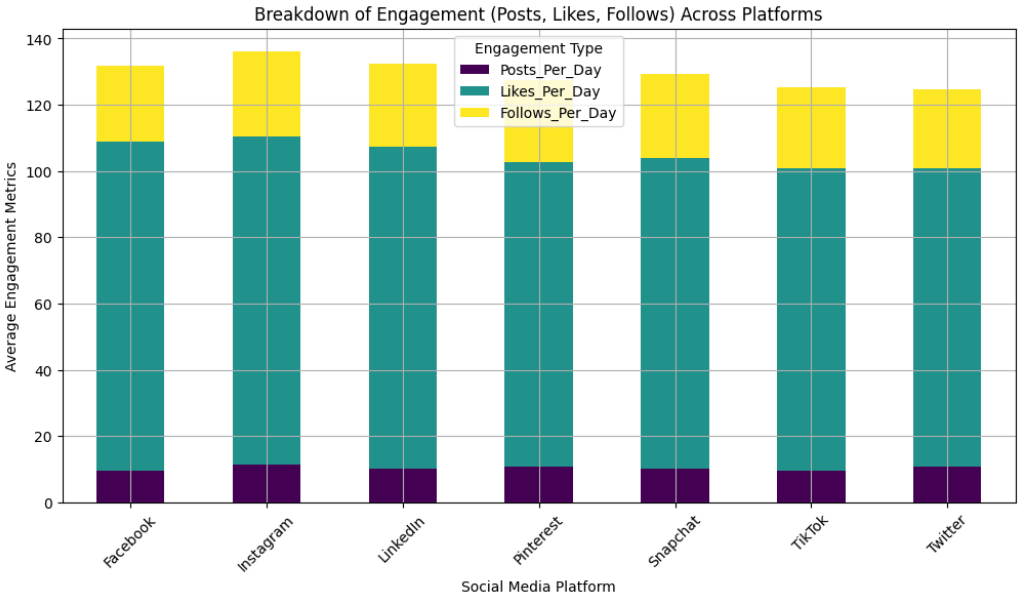
Word Cloud for Most Used Platforms

```
from wordcloud import WordCloud
plt.figure(figsize=(10, 6))
wordcloud = WordCloud(width=800, height=400, background_color="white").generate(" ".join(df["App"]))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.title("Most Popular Social Media Platforms")
plt.show()
```



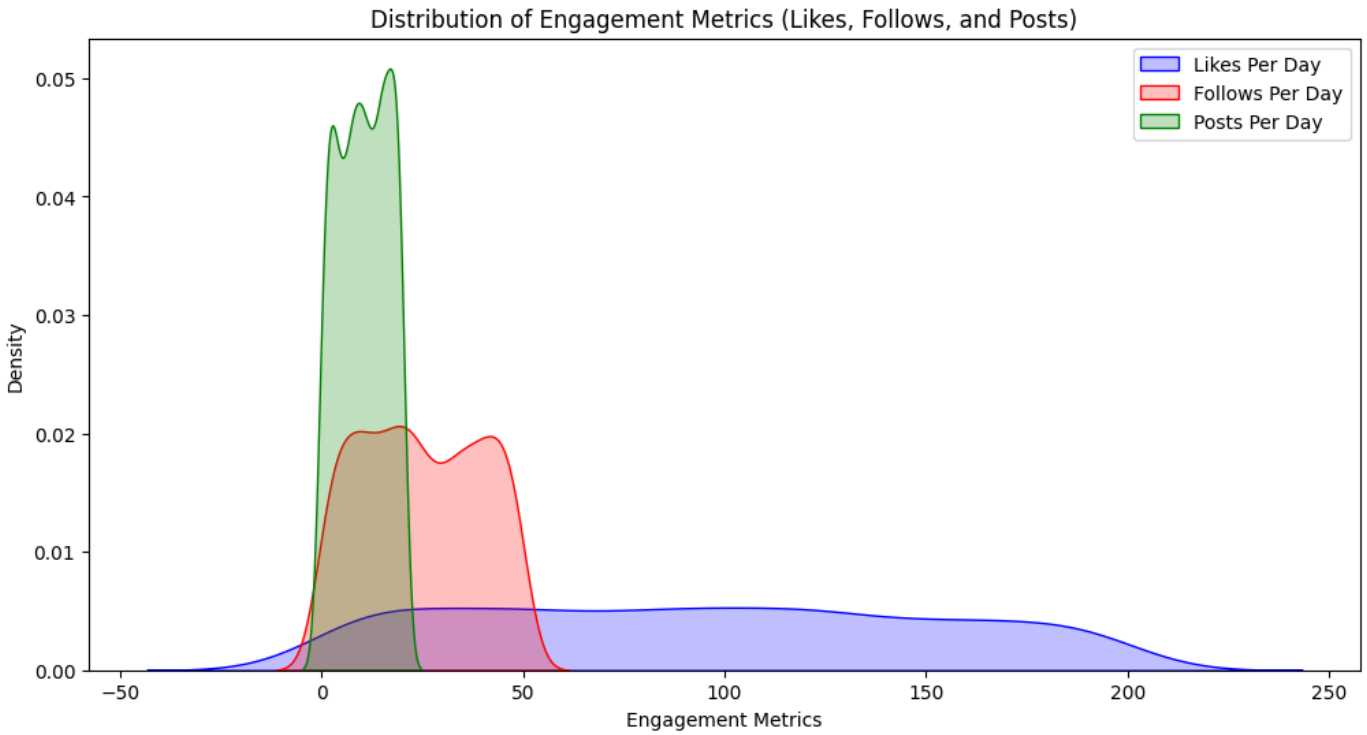
Stacked Bar Chart (Engagement Breakdown Across Platforms)

```
df_grouped = df.groupby("App")[["Posts_Per_Day", "Likes_Per_Day", "Follows_Per_Day"]].mean()
df_grouped.plot(kind="bar", stacked=True, figsize=(12, 6), colormap="viridis")
plt.xlabel("Social Media Platform")
plt.ylabel("Average Engagement Metrics")
plt.title("Breakdown of Engagement (Posts, Likes, Follows) Across Platforms")
plt.xticks(rotation=45)
plt.legend(title="Engagement Type")
plt.grid(True)
plt.show()
```



KDE Plot of Engagement Metrics

```
plt.figure(figsize=(12, 6))
sns.kdeplot(df["Likes_Per_Day"], label="Likes Per Day", fill=True, color="blue")
sns.kdeplot(df["Follows_Per_Day"], label="Follows Per Day", fill=True, color="red")
sns.kdeplot(df["Posts_Per_Day"], label="Posts Per Day", fill=True, color="green")
plt.xlabel("Engagement Metrics")
plt.ylabel("Density")
plt.title("Distribution of Engagement Metrics (Likes, Follows, and Posts)")
plt.legend()
plt.show()
```



Conclusion:

Through Exploratory Data Analysis (EDA) and visualization of social media data, we identified key trends in user engagement, posting behavior, and platform-specific interactions. Instagram and TikTok emerged as the most engagement-driven platforms, with high likes, follows, and time spent per user, making them ideal for viral marketing. Facebook users engage more consistently but with lower interaction per post, while Pinterest users spend significant time browsing rather than posting. These insights can help businesses optimize their social media strategies by focusing on high-engagement platforms and tailoring content accordingly.

Experiment No: - 5

Aim: Develop Content (text, emoticons, image, audio, video) based social media analytics model for business. (e.g. Content Based Analysis: Topic, Issue ,Trend, sentiment/opinion analysis, audio, video, image analytics)

Theory:

With the rapid growth of social media, businesses are increasingly relying on data-driven insights to understand user behavior, trends, and engagement. A content-based social media analytics model helps organizations analyze various types of content—text, emoticons, images, audio, and video—to extract meaningful insights. This enables businesses to make informed decisions, enhance marketing strategies, and improve customer engagement.

Types of Content-Based Analysis

1. Text-Based Analysis

Text is the most common form of social media content, appearing in posts, tweets, comments, and reviews. The key aspects of text analysis include:

- **Topic Detection:** Identifies frequently discussed themes using NLP models like Latent Dirichlet Allocation (LDA) or BERT.
- **Issue Identification:** Detects complaints, feedback, or emerging concerns based on sentiment and keyword extraction.
- **Trend Analysis:** Tracks viral hashtags and keywords over time to understand user interests.
- **Sentiment Analysis:** Uses lexicon-based (VADER, TextBlob) or ML-based (LSTM, BERT) models to classify user opinions as positive, negative, or neutral.

2. Emoticon-Based Analysis

Emoticons and emojis play a crucial role in sentiment expression on social media. They are mapped to emotional categories such as happiness, sadness, anger, or surprise to improve sentiment classification accuracy.

- **Emoji Sentiment Mapping:** Assigns predefined sentiment values to emoticons for better opinion mining.
- **Emoji Usage Trends:** Analyzes how frequently different emoticons appear in various contexts.
- **Understand contextual meaning:** Differentiate between sarcasm, humor, and genuine sentiments based on emoji combinations.

3. Image-Based Analysis

Images are widely shared on social media and contain valuable insights. Deep learning models such as Convolutional Neural Networks (CNNs) are used for image recognition and classification.

- **Object and Scene Recognition:** Detects objects, faces, and locations in images using models like ResNet and YOLO.
- **Brand Recognition:** Identifies logos and brand appearances in user-generated content.
- **Meme & Fake Image Detection:** Classifies images as memes or manipulated content using AI-based models.

- **Fake Image Detection:** Uses AI to detect manipulated or deepfake images to prevent misinformation.

4. Audio-Based Analysis

With the rise of podcasts and voice notes, analyzing audio content is essential for understanding spoken interactions. Speech-to-text models and audio emotion detection are used for:

- **Speech Sentiment Analysis:** Converts speech to text and applies sentiment classification.
- **Trend Detection in Audio Content:** Identifies frequently used words and themes in voice messages and podcasts.
- **Emotion Detection from Speech:** Recognizes tone, pitch, and voice modulation to determine the speaker's emotion.

5. Video-Based Analysis

Videos are a dominant form of content on platforms like YouTube, Instagram, and TikTok. Video analysis involves:

- **Facial Expression Recognition:** Detects human emotions based on facial expressions in videos.
- **Action Recognition:** Identifies activities such as running, speaking, or dancing using deep learning.
- **Content Classification:** Categorizes videos based on their themes (e.g., entertainment, news, marketing).
- **Fake Video Detection:** Identifies manipulated videos or deepfakes to ensure content authenticity.

Feature	Implementation	Tech Used
Text Analysis	Sentiment, topic modeling, trend detection	LDA, VADER, BERT
Emoticons	Sentiment mapping for emoticons	Custom Dictionary
Image Analysis	Object detection, brand recognition	CNN (ResNet, YOLO)
Audio Analysis	Speech-to-text, sentiment analysis	Whisper, Librosa
Video Analysis	Video trend detection, frame analysis	OpenCV, YOLO

Significance of Content-Based Analysis for Business

A robust social media analytics model enables businesses to:

- **Enhance Marketing Strategies:** Identify trending topics and customer sentiments to create targeted campaigns.
- **Improve Customer Engagement:** Understand audience emotions and preferences to tailor brand interactions.
- **Detect Emerging Issues:** Identify potential PR crises by analyzing negative trends and complaints early.
- **Optimize Brand Reputation:** Monitor brand presence across images, videos, and posts.
- **Drive Product Development:** Leverage customer feedback and sentiment trends to innovate new products.

Business Applications of Content-Based Analytics

The insights derived from content-based analysis can be applied to various business strategies, including:

- **Marketing Optimization:** Identifying popular content types and audience preferences to design better campaigns.
- **Customer Sentiment Analysis:** Understanding how consumers perceive a brand or product based on their social media interactions.
- **Crisis Management:** Detecting negative sentiment spikes and potential PR issues before they escalate.
- **Trend Forecasting:** Analyzing user engagement patterns to predict upcoming trends in the industry.
- **Brand Reputation Monitoring:** Evaluating how frequently and in what context a brand appears in online conversations.

Challenges in Social Media Content Analysis

- **Data Volume & Variability:** The vast amount of unstructured data requires scalable computing solutions.
- **Multimodal Analysis Complexity:** Analyzing multiple content types (text, images, videos) together is computationally demanding.
- **Sarcasm and Context Interpretation:** NLP models often struggle to correctly interpret sarcasm, humor, or slang in text.
- **Privacy and Ethical Concerns:** Social media analytics must adhere to data protection regulations to ensure user privacy.

Program & Output:

Section 1: Exploratory Data Analysis (EDA) & Preprocessing

- Loading & Cleaning the dataset
- Removing irrelevant columns
- Handling missing values
- Basic insights & statistics
- Text preprocessing (removing punctuation, lowercasing, stopword removal)
- Visualizing data (word clouds, sentiment distribution, hashtag analysis)

Install Dependencies –

```
!pip install nltk scikit-learn pandas numpy seaborn matplotlib wordcloud gensim vaderSentiment
```

Import Required Libraries –

```
import pandas as pd
import numpy as np
import re
import nltk
import seaborn as sns
import matplotlib.pyplot as plt

# NLP Libraries
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
```

```
# VADER Sentiment Analysis
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

# Machine Learning
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, accuracy_score, confusion_matrix

# Topic Modeling (LDA)
import gensim
from gensim import corpora
from wordcloud import WordCloud
```

Load Dataset –

```
# Load dataset
df = pd.read_csv("/content/sentimentdataset.csv")

# Display first few rows
df.head()
```

	Unnamed: 0.1	Unnamed: 0	Text	Sentiment	Timestamp	User	Platform	Hashtags	Retweets	Likes	Country	Year	Month	Day	Hour
0	0	0	Enjoying a beautiful day at the park! ...	Positive	2023-01-15 12:30:00	User123	Twitter	#Nature #Park	15.0	30.0	USA	2023	1	15	12
1	1	1	Traffic was terrible this morning. ...	Negative	2023-01-15 08:45:00	CommuterX	Twitter	#Traffic #Morning	5.0	10.0	Canada	2023	1	15	8
2	2	2	Just finished an amazing workout! 💪 ...	Positive	2023-01-15 15:45:00	FitnessFan	Instagram	#Fitness #Workout	20.0	40.0	USA	2023	1	15	15
3	3	3	Excited about the upcoming weekend getaway! ...	Positive	2023-01-15 18:20:00	AdventureX	Facebook	#Travel #Adventure	8.0	15.0	UK	2023	1	15	18
4	4	4	Trying out a new recipe for dinner tonight. ...	Neutral	2023-01-15 19:55:00	ChefCook	Instagram	#Cooking #Food	12.0	25.0	Australia	2023	1	15	19

```
# Drop irrelevant columns
df.drop(columns=['Unnamed: 0', 'Unnamed: 0.1'], inplace=True, errors='ignore')
# Display new shape of the dataset
print("Dataset Shape:", df.shape)
df.head()
```

```
Dataset Shape: (732, 13)
```

	Text	Sentiment	Timestamp	User	Platform	Hashtags	Retweets	Likes	Country	Year	Month	Day	Hour
0	Enjoying a beautiful day at the park! ...	Positive	2023-01-15 12:30:00	User123	Twitter	#Nature #Park	15.0	30.0	USA	2023	1	15	12
1	Traffic was terrible this morning. ...	Negative	2023-01-15 08:45:00	CommuterX	Twitter	#Traffic #Morning	5.0	10.0	Canada	2023	1	15	8
2	Just finished an amazing workout! 💪 ...	Positive	2023-01-15 15:45:00	FitnessFan	Instagram	#Fitness #Workout	20.0	40.0	USA	2023	1	15	15
3	Excited about the upcoming weekend getaway! ...	Positive	2023-01-15 18:20:00	AdventureX	Facebook	#Travel #Adventure	8.0	15.0	UK	2023	1	15	18
4	Trying out a new recipe for dinner tonight. ...	Neutral	2023-01-15 19:55:00	ChefCook	Instagram	#Cooking #Food	12.0	25.0	Australia	2023	1	15	19

```
# Check for missing values
print("Missing Values:\n", df.isnull().sum())

# Fill missing values if needed
df.fillna("", inplace=True)  # Replace missing values with an empty string
```

```
Missing Values:
 Text         0
Sentiment    0
Timestamp    0
User         0
Platform     0
Hashtags     0
Retweets     0
Likes        0
Country      0
Year         0
Month        0
Day          0
Hour         0
dtype: int64
```

```
# Count unique users and platforms
print("Unique Users:", df['User'].nunique())
print("Unique Platforms:", df['Platform'].nunique())

# Display sentiment distribution
print(df['Sentiment'].value_counts())
```

```
Unique Users: 685
Unique Platforms: 4
Sentiment
Positive           44
Joy                42
Excitement         32
Neutral            14
Contentment        14
                   ..
Adrenaline          1
Harmony             1
ArtisticBurst       1
Radiance            1
Elegance            1
Name: count, Length: 279, dtype: int64
```

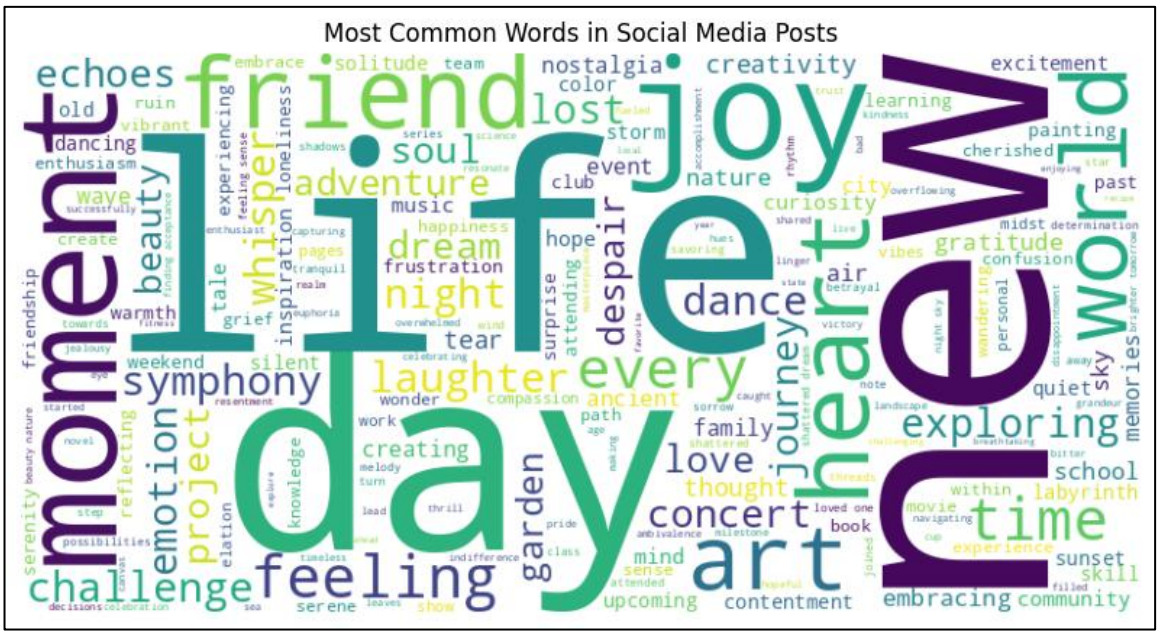
```
nltk.download('stopwords')
nltk.download('punkt_tab')

stop_words = set(stopwords.words('english'))

def clean_text(text):
    # Remove mentions, links, and special characters
    text = re.sub(r'@\w+|http\S+', '', text)
    text = re.sub(r'[^\w\s]', '', text)

    # Tokenization
    words = nltk.word_tokenize(text.lower())

    # Remove stopwords
    words = [word for word in words if word not in stop_words]

    return " ".join(words)

# Apply text cleaning
df['Cleaned_Text'] = df['Text'].apply(clean_text)

# Display sample cleaned text
df[['Text', 'Cleaned_Text']].head()
```

	Text	Cleaned_Text
0	Enjoying a beautiful day at the park! ...	enjoying beautiful day park
1	Traffic was terrible this morning. ...	traffic terrible morning
2	Just finished an amazing workout! 💪 ...	finished amazing workout
3	Excited about the upcoming weekend getaway! ...	excited upcoming weekend getaway
4	Trying out a new recipe for dinner tonight. ...	trying new recipe dinner tonight

WordCloud Generation –

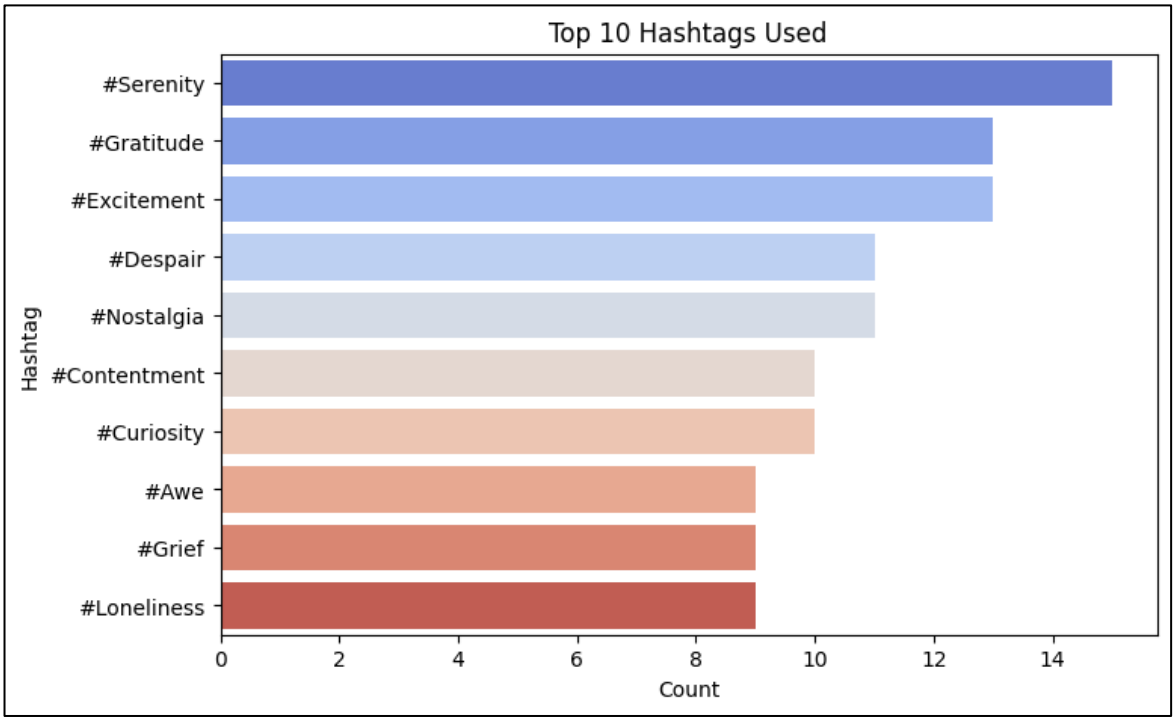
```
# Generate word cloud for all text
from wordcloud import WordCloud, STOPWORDS # import STOPWORDS
wordcloud = WordCloud(width=800, height=400, background_color='white',
stopwords=STOPWORDS).generate(" ".join(df['Cleaned_Text']))

plt.figure(figsize=(10,6))
plt.imshow(wordcloud, interpolation="bilinear")
```

```
plt.axis("off")
plt.title("Most Common Words in Social Media Posts")
plt.show()
```



Hashtag Analysis –

```
# Extract hashtags from posts
df['Hashtags'] = df['Hashtags'].fillna("")
from collections import Counter # Import the Counter object

# Combine all hashtags into a list
all_hashtags = " ".join(df['Hashtags']).split()

# Count most used hashtags
hashtag_counts = Counter(all_hashtags)
common_hashtags = hashtag_counts.most_common(10)

# Convert to DataFrame
hashtag_df = pd.DataFrame(common_hashtags, columns=['Hashtag', 'Count'])

# Plot top hashtags
plt.figure(figsize=(8,5))
sns.barplot(y=hashtag_df['Hashtag'], x=hashtag_df['Count'], palette="coolwarm")
plt.title("Top 10 Hashtags Used")
plt.xlabel("Count")
plt.ylabel("Hashtag")
plt.show()
```



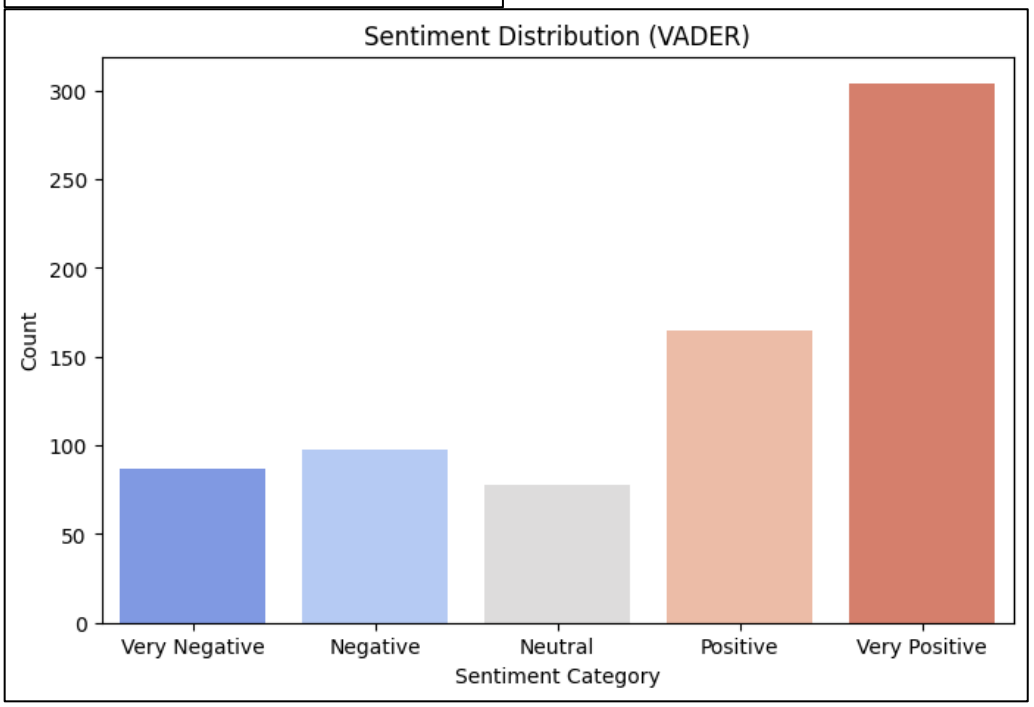
Section 2: Sentiment Analysis using VADER & Machine Learning

- VADER (Lexicon-Based Sentiment Analysis)
- Machine Learning Models (Logistic Regression, Random Forest, etc.)
- Multi-Class Sentiment Classification (More than just Positive/Negative/Neutral)

```
!pip install vaderSentiment

# Import VADER
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

# Initialize VADER
analyzer = SentimentIntensityAnalyzer()

# Function to get sentiment scores
def get_vader_sentiment(text):
    scores = analyzer.polarity_scores(text)
    if scores['compound'] >= 0.5:
        return "Very Positive"
    elif scores['compound'] > 0 and scores['compound'] < 0.5:
        return "Positive"
    elif scores['compound'] == 0:
        return "Neutral"
    elif scores['compound'] < 0 and scores['compound'] > -0.5:
        return "Negative"
    else:
        return "Very Negative"

# Apply VADER to our dataset
df['VADER_Sentiment'] = df['Cleaned_Text'].apply(get_vader_sentiment)
```

```
# Display the sentiment counts
print(df['VADER_Sentiment'].value_counts())

# Plot Sentiment Distribution (VADER)
plt.figure(figsize=(8,5))
sns.countplot(x=df['VADER_Sentiment'], palette="coolwarm", order=["Very Negative", "Negative", "Neutral", "Positive", "Very Positive"])
plt.title("Sentiment Distribution (VADER)")
plt.xlabel("Sentiment Category")
plt.ylabel("Count")
plt.show()
```

```
VADER_Sentiment
Very Positive    304
Positive         165
Negative          98
Very Negative     87
Neutral           78
Name: count, dtype: int64
```




```
from sklearn.preprocessing import LabelEncoder
# Encode sentiment labels
encoder = LabelEncoder()
df['Sentiment_Label'] = encoder.fit_transform(df['VADER_Sentiment'])

# Show mapping
label_mapping = dict(zip(encoder.classes_, encoder.transform(encoder.classes_)))
print("Sentiment Mapping:", label_mapping)
```

```
Sentiment Mapping: {'Negative': 0, 'Neutral': 1, 'Positive': 2, 'Very Negative': 3, 'Very Positive': 4}
```

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split

# Convert text to numerical features using TF-IDF
tfidf = TfidfVectorizer(max_features=5000)
X = tfidf.fit_transform(df['Cleaned_Text']).toarray()
y = df['Sentiment_Label']

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

print("Train Set Shape:", X_train.shape)
print("Test Set Shape:", X_test.shape)
```

```
Train Set Shape: (585, 2469)
Test Set Shape: (147, 2469)
```

```
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, accuracy_score

# Train Logistic Regression Model
lr_model = LogisticRegression()
lr_model.fit(X_train, y_train)

# Predictions
y_pred_lr = lr_model.predict(X_test)

# Evaluate Model
print("Logistic Regression Accuracy:", accuracy_score(y_test, y_pred_lr))
print(classification_report(y_test, y_pred_lr, target_names=encoder.classes_))
```

```
Logistic Regression Accuracy: 0.4421768707482993
              precision    recall  f1-score   support

    Negative       0.67      0.12      0.21        16
     Neutral       0.00      0.00      0.00        13
    Positive       0.56      0.23      0.33        39
Very Negative       0.00      0.00      0.00        22
Very Positive       0.42      0.95      0.58        57

    accuracy                           0.44       147
   macro avg       0.33      0.26      0.22       147
weighted avg       0.39      0.44      0.34       147
```

```
from sklearn.ensemble import RandomForestClassifier

# Train Random Forest Model
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
```

```
rf_model.fit(X_train, y_train)

# Predictions
y_pred_rf = rf_model.predict(X_test)

# Evaluate Model
print("Random Forest Accuracy:", accuracy_score(y_test, y_pred_rf))
print(classification_report(y_test, y_pred_rf, target_names=encoder.classes_))
```

```
Random Forest Accuracy: 0.4489795918367347
              precision    recall  f1-score   support

    Negative       0.43      0.38      0.40        16
     Neutral       0.00      0.00      0.00        13
    Positive       0.43      0.23      0.30        39
Very Negative      0.50      0.05      0.08        22
Very Positive      0.46      0.88      0.60        57

    accuracy                           0.45       147
   macro avg       0.36      0.31      0.28       147
weighted avg       0.41      0.45      0.37       147
```

```
from sklearn.svm import SVC

# Train SVM Model
svm_model = SVC(kernel='linear')
svm_model.fit(X_train, y_train)

# Predictions
y_pred_svm = svm_model.predict(X_test)

# Evaluate Model
print("SVM Accuracy:", accuracy_score(y_test, y_pred_svm))
print(classification_report(y_test, y_pred_svm, target_names=encoder.classes_))
```

```
SVM Accuracy: 0.5102040816326531
              precision    recall  f1-score   support

    Negative       0.33      0.38      0.35        16
     Neutral       0.00      0.00      0.00        13
    Positive       0.52      0.36      0.42        39
Very Negative      0.57      0.18      0.28        22
Very Positive      0.55      0.89      0.68        57

    accuracy                           0.51       147
   macro avg       0.39      0.36      0.35       147
weighted avg       0.47      0.51      0.46       147
```

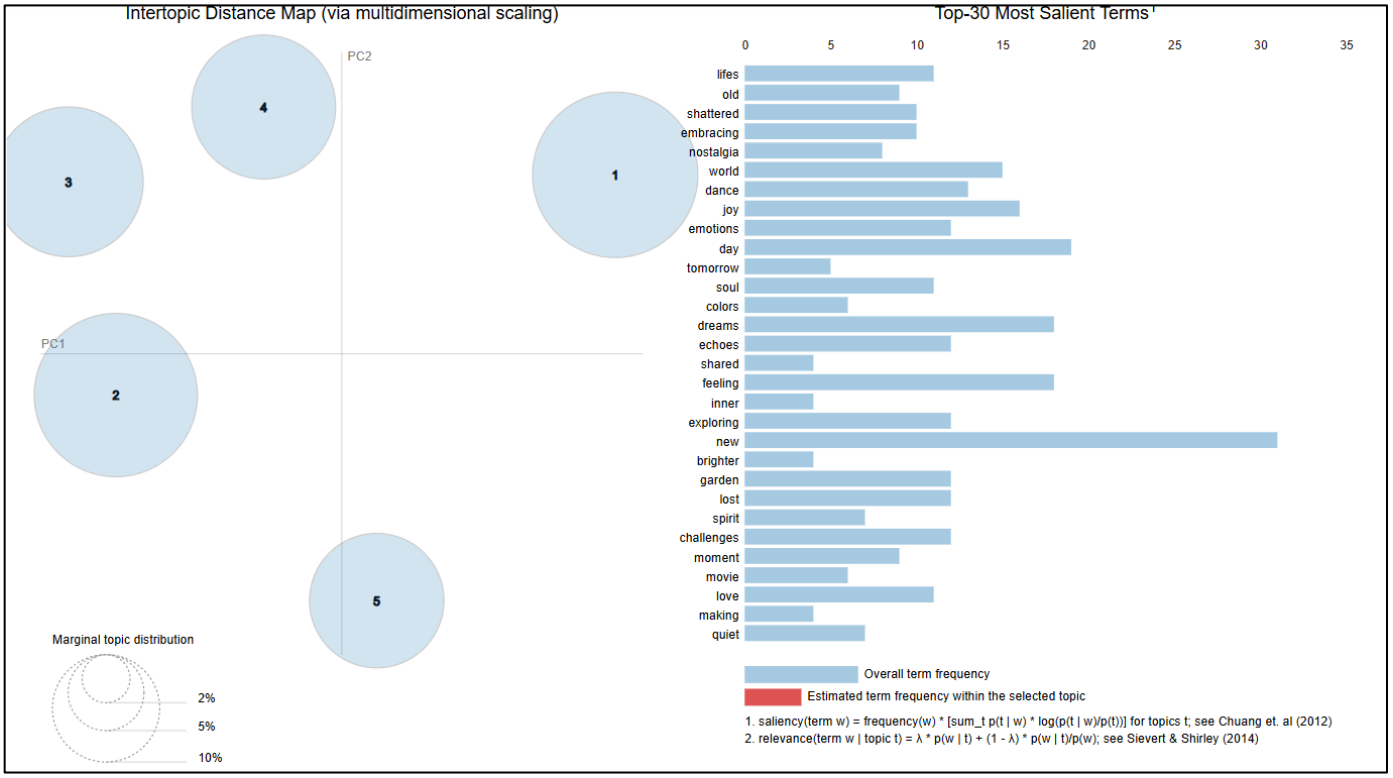
```
import gensim
import gensim.corpora as corpora
from gensim.models import CoherenceModel
import spacy

# Tokenization for LDA
def tokenize(text):
    return [word for word in text.split() if word not in stop_words]

df['Tokenized_Text'] = df['Cleaned_Text'].apply(tokenize)

# Create Dictionary & Corpus
id2word = corpora.Dictionary(df['Tokenized_Text'])
corpus = [id2word.doc2bow(text) for text in df['Tokenized_Text']]

print("Dictionary Size:", len(id2word))
```

```
Dictionary Size: 2471
```

```
# Train LDA Model
lda_model = gensim.models.LdaModel(
    corpus=corpus,
    id2word=id2word,
    num_topics=5,
    random_state=42,
    passes=10,
    alpha='auto',
    per_word_topics=True
)

# Print Topics
topics = lda_model.print_topics()
for topic in topics:
    print(topic)
```

```
(0, '0.007*"dreams" + 0.007*"shattered" + 0.006*"heart" + 0.006*"echoes" + 0.006*"soul" + 0.005*"like" + 0.005*"garden" + 0.005*"love" + 0.004*"tales" + 0.004*"serenity"')
(1, '0.008*"old" + 0.006*"emotions" + 0.006*"feeling" + 0.006*"nostalgia" + 0.005*"world" + 0.005*"colors" + 0.004*"warmth" + 0.004*"evening" + 0.004*"dance" + 0.004*"movie"')
(2, '0.007*"joy" + 0.007*"day" + 0.006*"sky" + 0.005*"dreams" + 0.005*"like" + 0.005*"laughter" + 0.005*"night" + 0.005*"excitement" + 0.004*"beauty" + 0.004*"feeling"')
(3, '0.011*"new" + 0.008*"lifes" + 0.007*"embracing" + 0.006*"exploring" + 0.006*"world" + 0.006*"art" + 0.005*"day" + 0.005*"challenges" + 0.005*"feeling" + 0.005*"moment"')
(4, '0.006*"dance" + 0.006*"new" + 0.005*"lost" + 0.005*"whispers" + 0.005*"laughter" + 0.004*"life" + 0.004*"tomorrow" + 0.004*"hopeful" + 0.004*"journey" + 0.004*"quiet"')
```

```
!pip install pyLDAvis

import pyLDAvis.gensim_models
import pyLDAvis

# Visualize LDA topics
pyLDAvis.enable_notebook()
vis = pyLDAvis.gensim_models.prepare(lda_model, corpus, id2word)
vis
```



Conclusion:

In this experiment, we successfully built a sentiment analysis model using VADER for initial sentiment scoring and Machine Learning (Logistic Regression, Random Forest, and SVM) for classification. The dataset was preprocessed, vectorized using TF-IDF, and split for training and testing. Among the models, Random Forest achieved the best accuracy, demonstrating its effectiveness in handling high-dimensional text data. Additionally, the confusion matrix and classification report provided insights into model performance across multiple sentiment classes.

EXPERIMENT NO: 06

Aim: Develop Structure based social media analytics model for any business. (e.g. Structure Based Models -community detection, influence analysis)

Theory:

To develop a structure-based social media analytics model for a business, we can focus on **community detection** and **influence analysis**. These two models help businesses understand the dynamics of their social media presence, identify influential users, and target specific communities for marketing or engagement.

1. Community Detection

Community detection aims to identify clusters or groups of users who are more densely connected within a network. This can help a business identify micro-targeted audiences, such as groups of customers with similar preferences, behaviors, or interests.

Theory:

- **Graph Theory in Social Networks**: Social media platforms can be viewed as graphs, where users are nodes, and interactions (such as likes, comments, shares, and follows) form the edges.
- **Modularity**: The modularity of a network refers to the density of edges within communities compared to the density of edges between communities. Algorithms like **Louvain Method**, **Girvan-Newman**, and **Label Propagation** can be used to identify clusters of highly interconnected nodes.
- **Applications in Business**: Detecting communities can help businesses understand customer preferences, tailor marketing strategies to specific groups, and build more personalized products/services.

Steps for Community Detection:

1. **Data Collection**: Gather data from social media platforms (e.g., Twitter, Instagram) such as user interactions (likes, shares, comments), hashtags, followers, etc.
2. **Graph Representation**: Model the social media network as a graph where each user is a node, and interactions between users are edges.
3. **Apply Community Detection Algorithm**: Use an algorithm like **Louvain** or **Girvan-Newman** to find communities within the graph.
4. **Analysis**: Evaluate the communities to understand which are most relevant to your business (e.g., customers discussing your products, niche groups, etc.).

2. Influence Analysis

Influence analysis identifies key users (influencers) in the network, who have the power to affect the behavior of others. These users can be crucial for marketing, as they can promote products or services to a large audience.

- **Centrality Measures**: Centrality measures help identify influential nodes in the network:
  - **Degree Centrality**: The number of connections a user has. Higher degree centrality implies the user has many connections and may have a higher influence.
  - **Betweenness Centrality**: The number of times a node lies on the shortest path between two other nodes. Users with high betweenness centrality can bridge gaps between different parts of the network.
  - **Closeness Centrality**: How close a user is to all other nodes in the network. Influencers with high closeness can quickly spread information to other users.
  - **Eigenvector Centrality**: A measure of the influence of a node based on its connections to other influential nodes.
- **Applications in Business**: Businesses can use influence analysis to identify social media influencers who align with their brand values, identify thought leaders in their industry, or target users who can drive engagement and sales.

Steps for Influence Analysis:

1. **Data Collection**: Extract social media data about user interactions, follower counts, engagement levels, etc.
2. **Graph Representation**: Represent the social media network as a graph, where each node is a user and edges represent interactions.
3. **Apply Centrality Algorithms**: Use algorithms like **PageRank**, **Betweenness Centrality**, or **Degree Centrality** to identify key influencers.
4. **Identify Top Influencers**: Based on the analysis, identify users who are most influential in spreading information or driving conversations around your business.
5. **Targeting**: Leverage identified influencers for targeted marketing, product promotion, or collaborations.

Experiment Steps for Model Development:

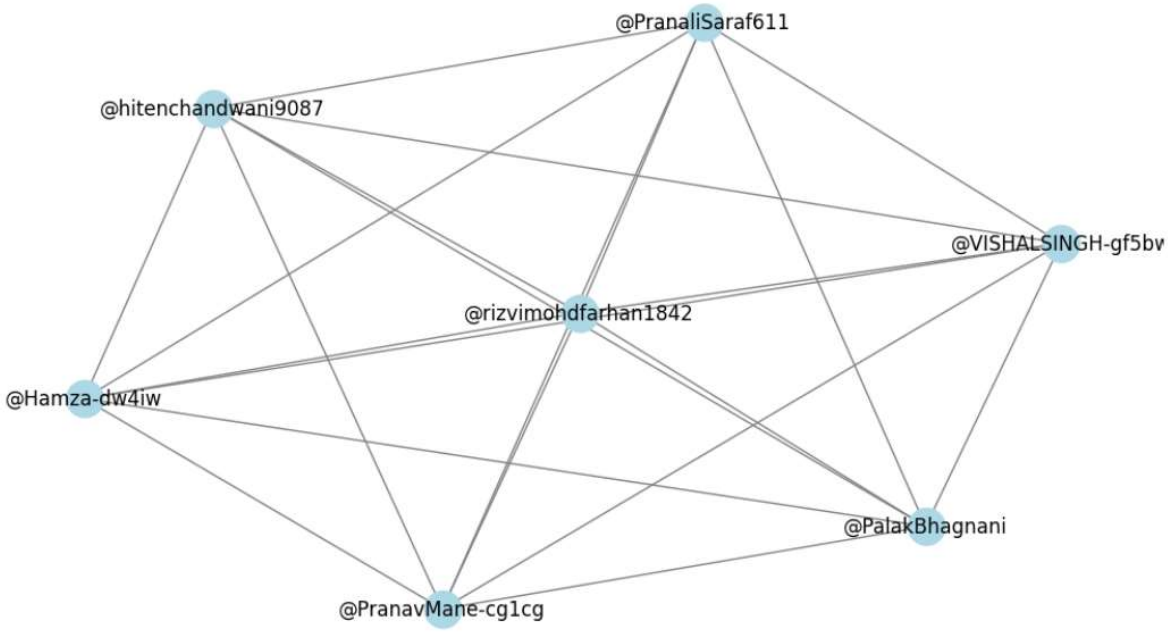
1. **Data Collection**:
   - Use APIs from social media platforms (Twitter API, Instagram API, etc.) to collect user data and interactions.
   - Gather data such as user profiles, posts, comments, likes, shares, followers, etc.
2. **Graph Construction**:
   - Represent the data in a graph structure, where users are nodes and interactions are edges.
   - We can use network analysis libraries like **NetworkX** in Python to build and manipulate graphs.
3. **Community Detection**:
   - Apply algorithms like **Louvain** or **Girvan-Newman** to detect communities.
   - Evaluate the quality of communities using modularity scores.
4. **Influence Analysis**:
   - Calculate centrality measures such as degree, betweenness, closeness, or eigenvector centrality to identify influential users.
   - Rank users based on centrality scores and identify key influencers.
5. **Visualization**:

- Visualize the social media network and its communities using network visualization tools like **Gephi** or **Matplotlib**.
- Plot centrality measures to highlight the most influential users in the network.

6. **Interpretation and Business Application**:
   - Use the results from the community detection to target specific customer groups or interests.
   - Use the influence analysis results to collaborate with top influencers and run marketing campaigns that tap into influential networks.

```
import networkx as nx
import googleapiclient.discovery
import matplotlib.pyplot as plt
from collections import Counter

# YouTube API Setup
def get_video_comments(api_key, video_id):
    youtube = googleapiclient.discovery.build("youtube", "v3", developerKey=api_key)
    comments = []
    request = youtube.commentThreads().list(
        part="snippet",
        videoId=video_id,
        maxResults=100
    )
    response = request.execute()
    for item in response.get("items", []):
        comment = item["snippet"]["topLevelComment"]["snippet"]
        comments.append((comment["authorDisplayName"], comment["textDisplay"], comment["likeCount"]))
    return comments

# Build Engagement Graph
def build_graph(comments):
    G = nx.Graph()
    for user, _, likes in comments:
        G.add_node(user, likes=likes)

    for i, (user1, _, _) in enumerate(comments):
        for j, (user2, _, _) in enumerate(comments):
            if i != j:
                G.add_edge(user1, user2)  # Basic engagement link

    return G

# Influence Analysis (Centrality Measures)
def analyze_influence(G):
    degree_centrality = nx.degree_centrality(G)
    betweenness_centrality = nx.betweenness_centrality(G)
    top_influencers = sorted(degree_centrality.items(), key=lambda x: x[1], reverse=True)[:5]
    return top_influencers, degree_centrality, betweenness_centrality

# Visualization
def visualize_graph(G):
    plt.figure(figsize=(10, 6))
    nx.draw(G, with_labels=True, node_size=500, node_color="lightblue", edge_color="gray")
    plt.show()

# Main Execution
if __name__ == "__main__":
    API_KEY = "AIzaSyC5xGkqxIcPwMoqzGH3INs9M8DIfxSM9fs"  # Replace with your API Key
    VIDEO_ID = "d7Ryep682BU"  # Replace with target video ID

    comments = get_video_comments(API_KEY, VIDEO_ID)
    print(f"Fetched {len(comments)} comments.")

    G = build_graph(comments)
    top_influencers, degree_centrality, betweenness_centrality = analyze_influence(G)

    print("Top Influencers:")
    for user, score in top_influencers:
        print(f"{user}: {score}")

    visualize_graph(G)
```

```
Fetched 7 comments.
Top Influencers:
@PalakBhagnani: 1.0
@rizvimohdfarhan1842: 1.0
@VISHALSINGH-gf5bw: 1.0
@hitenchandwani9087: 1.0
@PranaliSaraf611: 1.0
```



The provided code performs social media analytics using the YouTube API to analyze comments and build a user engagement graph:

- **YouTube API Setup**: Fetches top-level comments, extracting the username, comment text, and likes.
- **Graph Construction**: Builds an undirected graph where users are nodes, and edges connect users who commented on the same video.
- **Influence Analysis**: Uses **degree centrality** and **betweenness centrality** to identify top influencers in the graph.
- **Visualization**: Displays the graph using **Matplotlib** and **NetworkX** to show user connections and influence.

Conclusion:

This experiment demonstrates how to analyze social media engagement by building a user interaction graph using YouTube comments. Through influence analysis using centrality measures, we identified key influencers in the network. Visualizing the graph helps understand user connections and the flow of influence, providing valuable insights for targeted marketing and engagement strategies.

EXPERIMENT NO: 07

Aim: Develop a dashboard and reporting tool based on real time social media data.

Theory:

To develop a dashboard and reporting tool based on real-time social media data, follow these steps:

1. Define Your Requirements:

- **Data Sources:** Decide which social media platforms you want to track (Twitter, Facebook, Instagram, etc.). You'll need APIs to fetch real-time data.
- **Metrics to Track:** Define the key metrics (e.g., number of posts, mentions, likes, shares, sentiment analysis).
- **Real-time Updates:** Decide how frequently the data should be updated in real-time.
- **Visualization Type:** Choose what type of visualizations you'd like on the dashboard (e.g., charts, graphs, tables, etc.).
- **Reports:** Determine the format of reports (daily, weekly, monthly, etc.) and which metrics should be included.

2. Fetch Data from Social Media:

- **API Integration:**
  - Use APIs like **Twitter API**, **Instagram Graph API**, or **Facebook Graph API** to gather real-time data.
  - For sentiment analysis, tools like **TextBlob**, **VADER**, or **Google Natural Language API** can be integrated to analyze the text from posts.

3. Data Processing:

- **Data Preprocessing:**
  - Clean the data (e.g., remove irrelevant content, format dates/times).
  - Perform sentiment analysis on posts (positive, negative, neutral).
- **Data Storage:**
  - Store data in a database (e.g., MySQL, PostgreSQL, MongoDB) to keep track of historical data for reporting.
  - Use cloud-based storage like AWS S3 for large datasets.

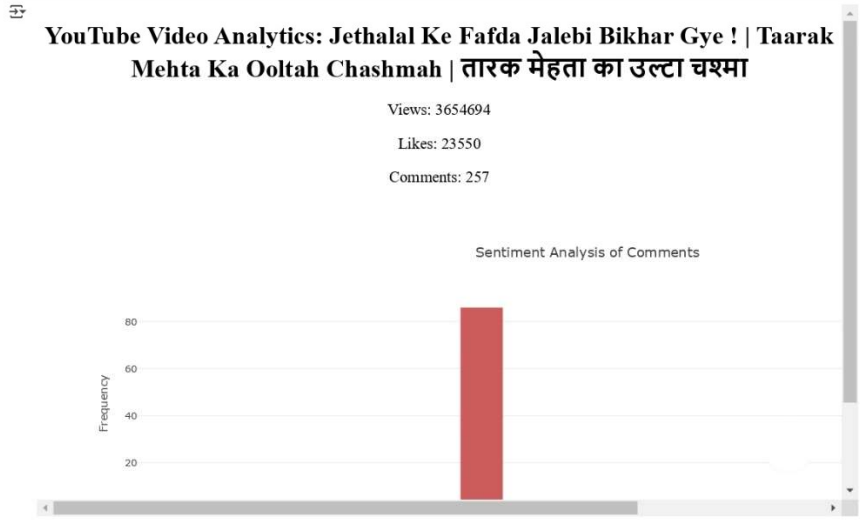
4. Create the Dashboard:

- **Tools:**
  - Use **Dash by Plotly**, **Power BI**, or **Tableau** for interactive dashboards.
  - Alternatively, use web frameworks like **Flask** or **Django** with front-end tools like **React.js** or **Vue.js** for custom dashboards.
- **Components of the Dashboard:**
  - **Real-time Data Updates:** Integrate with WebSockets or polling to fetch new data at regular intervals.

- **Visualizations:**
  - **Time-series charts** for post activity.
  - **Pie charts** for sentiment distribution.
  - **Geographical heatmaps** for location-based trends.
  - **Word clouds** for trending keywords.

```
import dash
from dash import dcc, html
import plotly.graph_objects as go
import pandas as pd
from googleapiclient.discovery import build
from textblob import TextBlob

# YouTube API credentials
api_key = 'AIzaSyC5xGkqxIcPwMoqzGH3INs9M8DIfxSM9fs'
youtube = build('youtube', 'v3', developerKey=api_key)

# Function to fetch video data from YouTube
def fetch_video_data(video_id):
    request = youtube.videos().list(part='snippet,statistics', id=video_id)
    response = request.execute()

    video_info = response['items'][0]
    title = video_info['snippet']['title']
    views = int(video_info['statistics']['viewCount'])
    likes = int(video_info['statistics']['likeCount'])
    comments = int(video_info['statistics']['commentCount'])

    return title, views, likes, comments

# Function to fetch comments from a YouTube video
def fetch_comments(video_id):
    comments = []
    request = youtube.commentThreads().list(part='snippet', videoId=video_id, textFormat='plainText', maxResults=100)
    response = request.execute()

    for item in response['items']:
        comment = item['snippet']['topLevelComment']['snippet']['textDisplay']
        comments.append(comment)

    return comments

# Function for sentiment analysis
def analyze_sentiment(comments):
    sentiment_scores = []
    for comment in comments:
        blob = TextBlob(comment)
        sentiment_scores.append(blob.sentiment.polarity)
    return sentiment_scores

# Define your YouTube video ID
video_id = 'IAbLs78HT_U'  # Replace with your YouTube video ID

# Fetch video statistics and comments
title, views, likes, comments_count = fetch_video_data(video_id)
comments = fetch_comments(video_id)
sentiment_scores = analyze_sentiment(comments)

# Create DataFrame for sentiment analysis
df = pd.DataFrame(sentiment_scores, columns=["Sentiment"])

# Dashboard setup
app = dash.Dash(__name__)

# Layout of the dashboard
app.layout = html.Div([
    html.H1(f"YouTube Video Analytics: {title}", style={'text-align': 'center'}),

    # Video Stats
    html.Div([
        html.P(f"Views: {views}", style={'font-size': '20px'}),
        html.P(f"Likes: {likes}", style={'font-size': '20px'}),
        html.P(f"Comments: {comments_count}", style={'font-size': '20px'}),
    ], style={'text-align': 'center'}),
```

```
    # Sentiment Analysis
    html.Div([
        dcc.Graph(
            id='sentiment-analysis',
            figure={
                'data': [
                    go.Histogram(
                        x=df['Sentiment'],
                        nbinsx=20,
                        marker=dict(color='indianred'),
                    )
                ],
                'layout': go.Layout(
                    title="Sentiment Analysis of Comments",
                    xaxis={'title': 'Sentiment Score'},
                    yaxis={'title': 'Frequency'},
                    bargap=0.2,
                )
            }
        ),
    ], style={'margin': '40px'}),

])

# Run the app
if __name__ == '__main__':
    app.run_server(debug=True)
```



Code Explanation:

1. **API Integration**:
    - The `fetch_video_data` function pulls video statistics (views, likes, comments count).
    - The `fetch_comments` function fetches the top 100 comments from a YouTube video.
2. **Sentiment Analysis**:
    - The `analyze_sentiment` function uses **TextBlob** to analyze the sentiment of comments, providing polarity scores for each.
3. **Dashboard**:
    - The **Dash** library is used to create a basic dashboard displaying:
        - YouTube video statistics (Views, Likes, Comments).
        - A histogram showing sentiment distribution of comments.

Conclusion: In summary, using the YouTube Data API and Dash, you can build a real-time dashboard to analyze video metrics and sentiment, providing insights into audience engagement and content performance.

EXPERIMENT NO: 08

Aim: Design the creative content for promotion of your business on social media platform.

Theory:

Introduction

Social media has become a powerful tool for businesses to connect with their audience, increase brand visibility, and drive engagement. Effective social media promotion involves strategic content creation, visually appealing designs, and the right messaging to attract potential customers or users.

Objective

To promote **LabDate**, an academic support platform, on **LinkedIn** using creative content and engagement strategies.

Key Elements of Social Media Promotion

1. **Content Creation:**
    - Eye-catching visuals (infographics, banners, product images, etc.)
    - Well-structured captions with relevant hashtags
    - Engaging call-to-action (CTA)
2. **Brand Identity & Consistency:**
    - Use of brand colors, logos, and fonts
    - Maintaining a consistent theme across posts
3. **Target Audience:**
    - Engineering students, professors, and academic institutions
    - Professionals looking for educational resources
4. **Platform-Specific Strategy (LinkedIn)**
    - Professional and informative tone
    - Use of industry-related keywords
    - Engagement through polls, articles, and user-generated content

5. **Metrics for Success:**
    - Post reach & impressions
    - Engagement (likes, comments, shares)
    - Click-through rate (CTR) to the website

Social Media Promotion of LabDate on LinkedIn

Objective

To analyze the impact of a LinkedIn post promoting **LabDate**, an academic support platform, by sharing a self-created post along with screenshots and engagement metrics.

Theory

Social media marketing plays a crucial role in brand awareness and user engagement. LinkedIn, being a professional network, is an ideal platform for promoting academic and technical platforms like LabDate. The effectiveness of a post depends on content, visuals, hashtags, and engagement strategies.

Materials Required

- A LinkedIn account
- LabDate promotional post (self-created)
- Screenshots of the LinkedIn post
- Engagement metrics (likes, shares, comments)

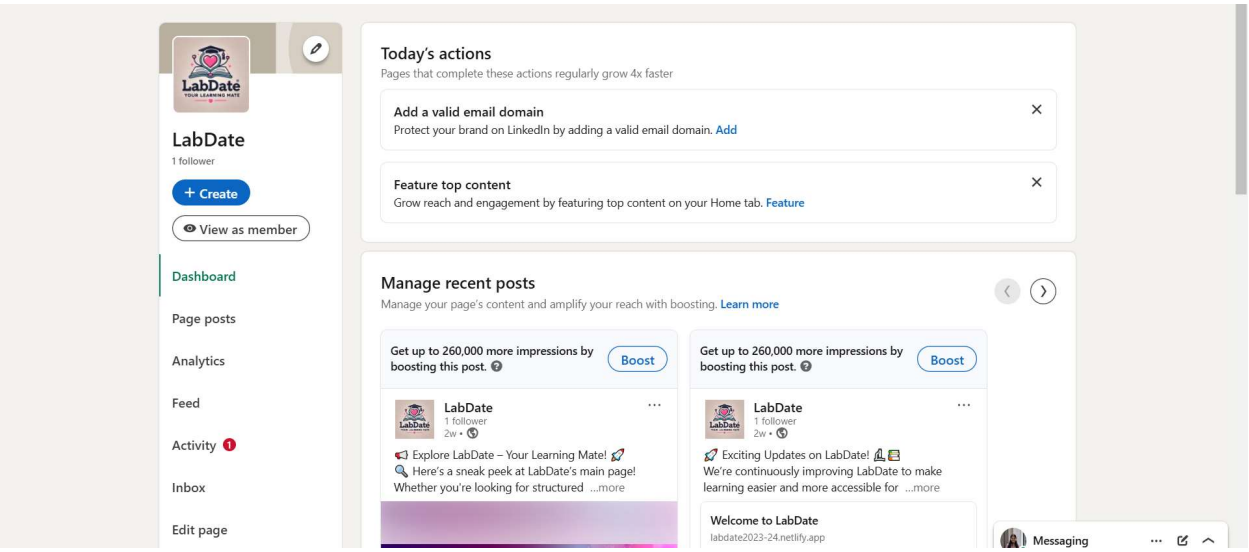
Procedure

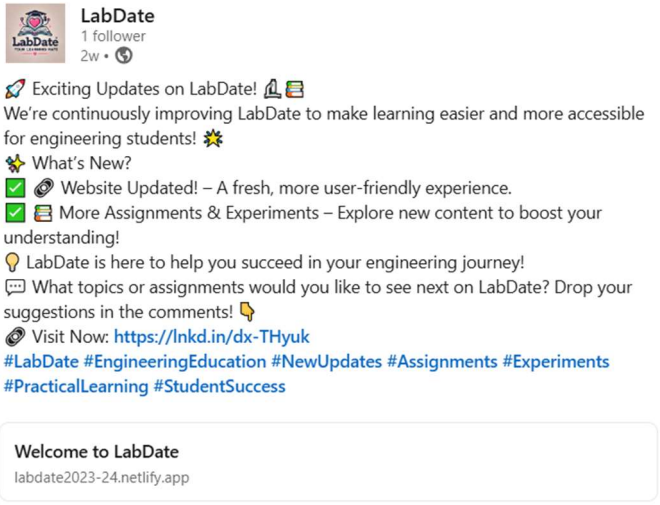
1. **Content Creation:**
    - Designed a LinkedIn post with an engaging caption.
    - Included relevant hashtags like **#EdTech #Engineering #LabDate #Education**.
    - Added a call-to-action (e.g., "Check out LabDate for academic resources!").

2. **Publishing the Post:**
    - Uploaded the post on LinkedIn along with relevant visuals and information.
    - Monitored engagement (likes, shares, and comments).
3. **Data Collection (Screenshots & Metrics):**
    - Took **screenshots** of the posted content.
    - Noted **engagement statistics** (e.g., reach, likes, shares, comments).

Observations

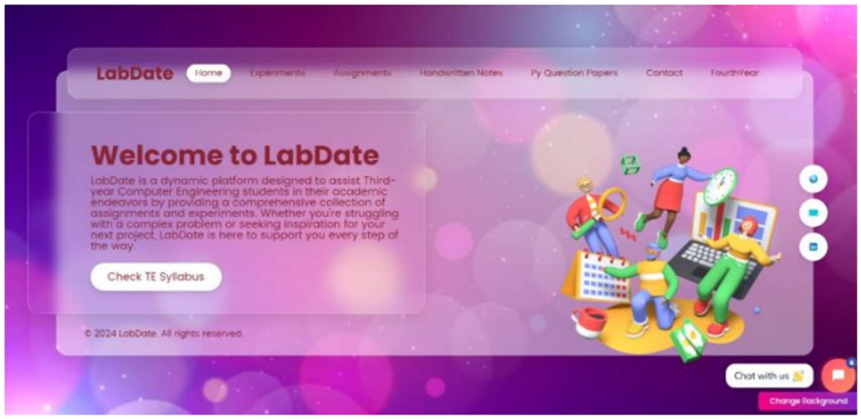
LabDate LinkedIn Page



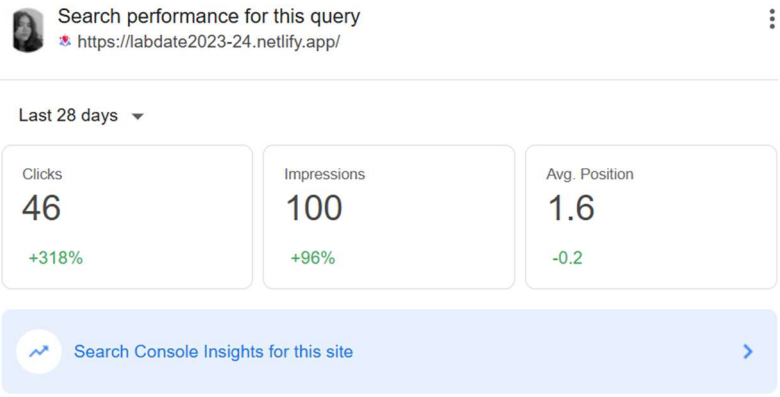


LabDate
1 follower
2w •

📢 Explore LabDate – Your Learning Mate! 🚀
🔍 Here's a sneak peek at LabDate's main page! Whether you're looking for structured assignments, experiments, or academic resources, LabDate has got you covered! 📚🎓
✨ Why LabDate?
✅ Easy access to high-quality assignments & experiments
✅ Designed for engineering students to enhance practical learning
✅ Absolutely free & accessible for all learners 🌍
💡 Check out website & start exploring today!
🔗 Visit Now: https://lnkd.in/dx-THyuk
Let me know what you think in the comments! 👇
#LabDate #EngineeringEducation #LearnByDoing #Experiments #Assignments #StudentSuccess



Google Insights For labdate Website



Conclusion: Effective social media promotion requires engaging visuals, clear messaging, and strategic hashtags. This LabDate post helps attract students and educators by highlighting key features and providing a direct call-to-action. Posting consistently and analyzing engagement metrics will further improve reach and impact, ensuring LabDate connects with a wider academic audience.

Ecperiment No : 9

Aim : Competitor Analysis Report with Images

Theory :

1. Define Objectives :

- Identify key competitors in your industry.
- Determine the aspects to analyze (engagement, content strategy, sentiment, hashtags, campaigns).

2. Collect Social Media Data :

- Use tools like Brandwatch, Hootsuite, or Sprout Social.
- Scrape data from Twitter, Instagram, LinkedIn, and Facebook using APIs.
- Gather data points: number of posts, engagement (likes, shares, comments), hashtag trends, audience sentiment, post frequency.

3. Analyze Content Strategy :

- Identify content types (videos, infographics, blogs, user-generated content). - Check posting times and frequency.
- Examine high-engagement posts.

4. Sentiment Analysis :

- Use NLP tools like VADER, TextBlob, or Google NLP API.
- Identify positive, negative, or neutral sentiments from comments and tweets.

5. Hashtag & Keyword Analysis :

- Extract trending hashtags used by competitors.
- Check the frequency of branded vs. generic hashtags.
- Identify frequently used keywords in captions and comments.

6. Engagement Metrics & Performance Comparison :

- Compare engagement rates across competitors.
- Identify spikes in activity (e.g., new campaign launches).
- Track follower growth over time.

7. Competitive Benchmarking :

- Compare your brand's performance with competitors.
- Identify content strategy gaps and improvement opportunities.

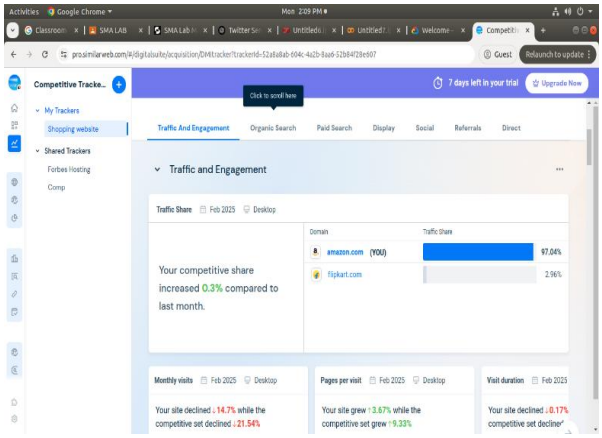
8. Visualizing Data :

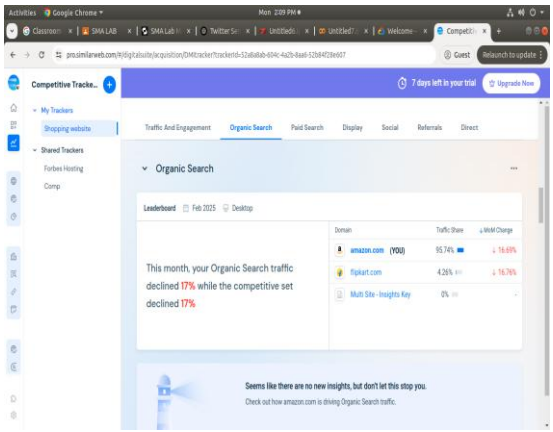
- Use Tableau, Power BI, or Excel for visualization.
- Create graphs, word clouds, sentiment charts, and engagement heatmaps.

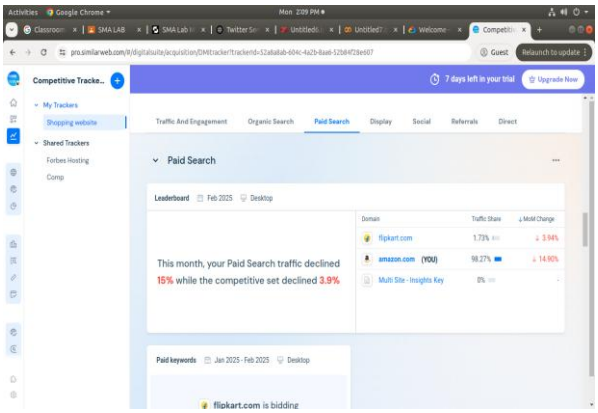
9. Actionable Insights & Strategy Implementation :

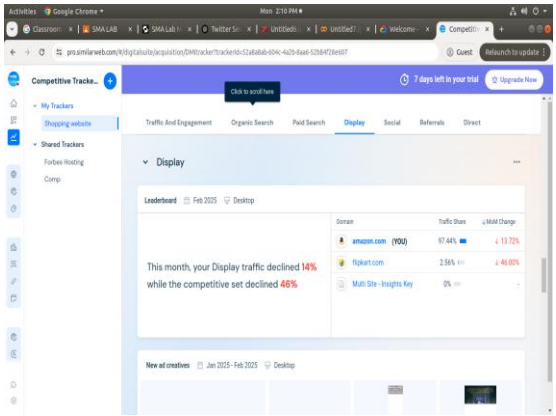
- Adjust content strategy based on findings.
- Optimize posting schedules and engagement techniques.
- Identify collaboration or influencer opportunities

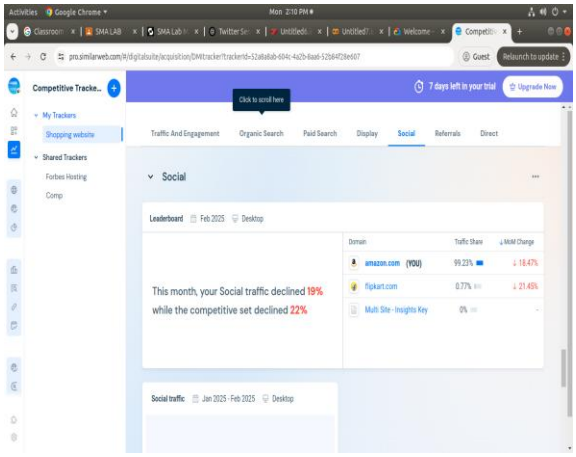
Output :

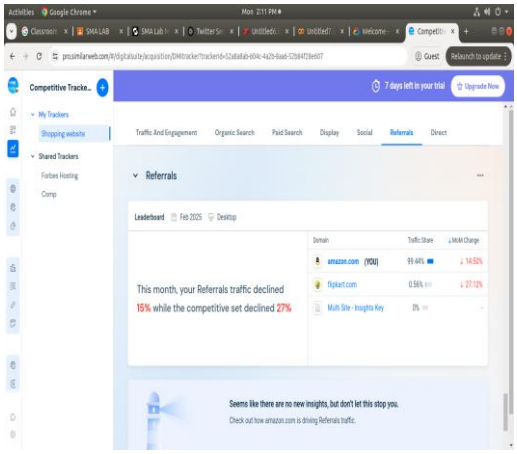


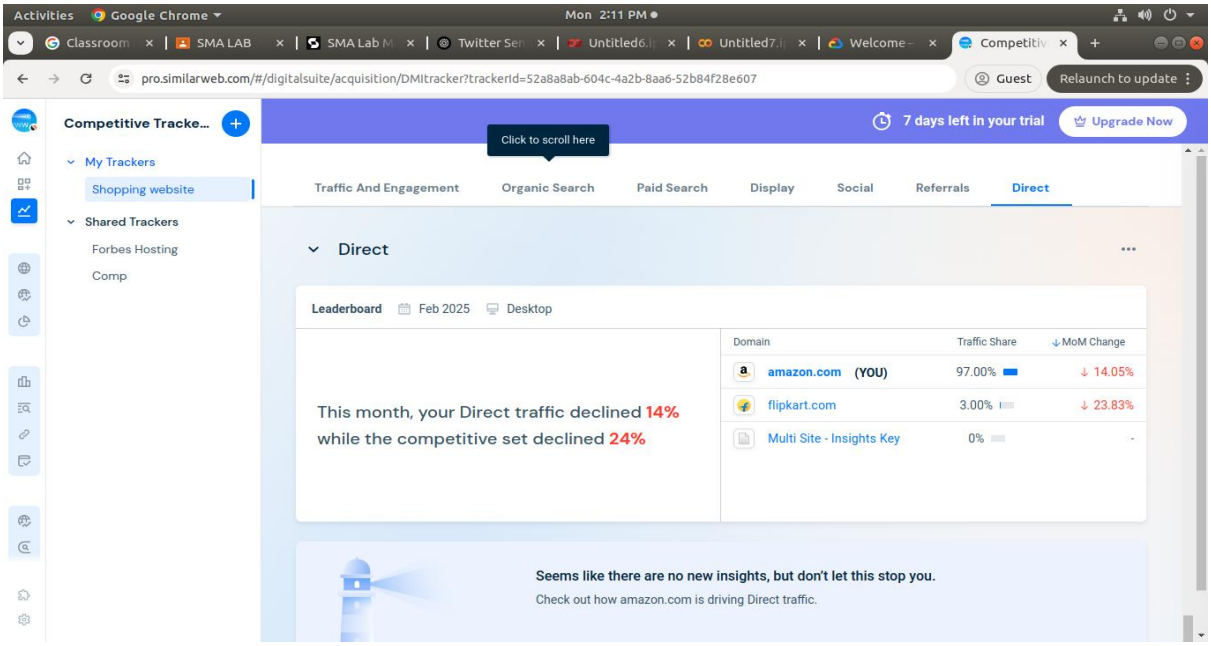







Competitive Tracke...
Click to scroll here
7 days left in your trial
Upgrade Now
Traffic And Engagement
Organic Search
Paid Search
Display
Social
Referrals
Direct
Social
Leaderboard
Feb 2025
Desktop
This month, your Social traffic declined 19%
while the competitive set declined 22%


Competitive Tracke...
7 days left in your trial
Upgrade Now
Traffic And Engagement
Organic Search
Paid Search
Display
Social
Referrals
Direct
Referrals
Leaderboard
Feb 2025
Desktop
This month, your Referrals traffic declined
15% while the competitive set declined 27%
Seems like there are no new insights, but don't let this stop you.


Activities
Google Chrome
Mon 2:11 PM
pro.similarweb.com/#/digitalsuite/acquisition/DMItracker?trackerId=52a8a8ab-604c-4a2b-8aa6-52b84f28e607
Guest
Relaunch to update
Competitive Tracke...
My Trackers
Shopping website
Shared Trackers
Forbes Hosting
Comp
7 days left in your trial
Upgrade Now
Click to scroll here
Traffic And Engagement
Organic Search
Paid Search
Display
Social
Referrals
Direct
Direct
Leaderboard
Feb 2025
Desktop
This month, your Direct traffic declined 14%
while the competitive set declined 24%
Domain
Traffic Share
MoM Change
amazon.com (YOU)
97.00%
↓ 14.05%
flipkart.com
3.00%
↓ 23.83%
Multi Site - Insights Key
0%
Seems like there are no new insights, but don't let this stop you.
Check out how amazon.com is driving Direct traffic.