

LAB MANUAL
PART A
(PART A: TO BE REFERRED BY STUDENTS)

Experiment No-02

A.1 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).

A.2	Lab Objective	To collect and store social media data.
A.3	Lab Outcome	Collect, monitor, store and track social media data.

A-4 Prerequisite

Knowledge of web scraping, crawling, Python, social media tool, Power BI

A.5 Theory:

PART B
(PART B: TO BE COMPLETED BY STUDENTS)

Roll No: C36	Name: Sanskruti Kadam
Class: BE-Comps	Batch: C2
Date of Experiment:	Date of Submission:
Grade:	

B.1. Study the fundamentals of social media platform and social media tools:

- Students can use twitter API/Facebook Graph API/Octoparse to collect social media data.
- Decide a trend, topic to collect social media Data like Union Budget 2023, Political issue, Election, etc.

B.2 Input and Output:

Web scraping through Octoparse:

The screenshot displays the Octoparse web scraping tool interface. The main window shows a browser view of the Amazon.in website, specifically the 'Deal of the Day' section for men's clothing. The interface includes a sidebar with navigation options like 'Upgrader', 'New', 'Task List', 'Templates', 'Tools', 'Posting', 'Automation', and 'Data Service'. The main content area shows product listings with details like price, discount, and delivery options. A 'Data Preview' table is visible at the bottom, showing extracted data for various products. The table has columns for 'No.', 'Image', 'asizebaseplus', 'asizebasepl...', 'aliconalt', 'asizebase', 'abadgetext', 'alinknormal...', 'Prs', and 'Actions'. The data rows show products like 'Allen Solly', 'Qube By Fort Collins', and 'Van Heusen'.

No.	Image	asizebaseplus	asizebasepl...	aliconalt	asizebase	abadgetext	alinknormal...	Prs	Actions
1									
2	https://m.media-am...	Allen Solly	Men's Cotton Crew ...	4.0 out of 5 stars	24,005	Best seller	https://www.amazon...	₹599	
3	https://m.media-am...	Qube By Fort Collins	Men's Parka Coat	4.1 out of 5 stars	1,728	Deal of the Day	https://www.amazon...	₹639	
4	https://m.media-am...	Qube By Fort Collins	mens Jacket	4.0 out of 5 stars	920	Deal of the Day	https://www.amazon...	₹749	
5	https://m.media-am...	Qube By Fort Collins	Men's Cotton Crew ...	4.0 out of 5 stars	10,820	Deal of the Day	https://www.amazon...	₹499	

Amazon.in

Buy Allen Solly Men Cas...

Amazon.in

56

Running

[Loop Item] executing loop item #8

Duplicates: 1 lines

Time Spent: 19s

Avg. Speed: 147 lines/min

Pause

Stop

Task Overview

Data List

Event Log

Recent Runs

#	Image	asizebaseplus	asizebaseplus1	aliconalt	asizebase	abadgetext	alinknormal_URL	Price	asizebase2
6	https://m.media-a...	Puma	Mens Jacket	5.0 out of 5 stars	1		https://www.amaz...	₹1,899	M.R.P.
7	https://m.media-a...	Qube By Fort Collins	Mens Jacket	4.1 out of 5 stars	35	Deal of the Day	https://www.amaz...	₹892	M.R.P.
8	https://m.media-a...	U.S. POLO ASSN.	Men Beige Colour ...	5.0 out of 5 stars	3		https://www.amaz...	₹3,464	M.R.P.
9	https://m.media-a...	Qube By Fort Collins	Fort Collins Men's ...	4.1 out of 5 stars	863	Deal of the Day	https://www.amaz...	₹569	M.R.P.
10									Price and c
11	https://m.media-a...	Amazon Brand - Sy...	Men's Cotton Rich ...	4.0 out of 5 stars	386	Deal of the Day	https://www.amaz...	₹649	M.R.P.
12	https://m.media-a...	Puma	men warmcell pad...	5.0 out of 5 stars	1		https://www.amaz...	₹3,199	M.R.P.
13	https://m.media-a...	Van Heusen	Men Pullover Swea...	4.2 out of 5 stars	7		https://www.amaz...	₹1,799	M.R.P.
14	https://m.media-a...	Qube By Fort Collins	Men's Nylon Short...	4.1 out of 5 stars	756		https://www.amaz...	₹829	M.R.P.
15	https://m.media-a...	Amazon Brand - Sy...	Mens Midweight B...	5.0 out of 5 stars	1		https://www.amaz...	₹1,799	M.R.P.
16	https://m.media-a...	Van Heusen	Men Sweatshirt - U...	4.0 out of 5 stars	435		https://www.amaz...	₹1,329	M.R.P.

< 1 2 3 > Go to Page

5 https://m.media-am... Qube By Fort Collins Men's Cotton Crew ... 4.0 out of 5 stars 10,820 Deal of the Day https://www.amazon... ₹499

6 https://m.media-am... Qube By Fort Collins Men's Quilted Jacket 4.0 out of 5 stars 1,703 Deal of the Day https://www.amazon... ₹910

Apply

34°C

Smoke

Search

16:46

05-02-2024

Amazon.in

Buy Allen Solly Men Cas...

Amazon.in

245

Running

Wait 2 second(s)

Duplicates: 4 lines

Time Spent: 40s

Avg. Speed: 359 lines/min

Pause

Stop

Task Overview

Data List

Event Log

Recent Runs

#	Image	asizebaseplus	asizebaseplus1	aliconalt	asizebase	abadgetext	alinknormal_URL	Price	asizebase2
1	https://m.media-a...	Red Tape	Men's Teal Graphic...	4.4 out of 5 stars	6		https://www.amaz...	₹929	M.R.P.
2	https://m.media-a...	Puma	Men Sweatshirt	5.0 out of 5 stars	2		https://www.amaz...	₹1,999	M.R.P.
3	https://m.media-a...	Qube By Fort Collins	Men Cotton Hood...	4.0 out of 5 stars	586		https://www.amaz...	₹686	M.R.P.
4	https://m.media-a...	U.S. POLO ASSN.	Men Olive Solid St...	5.0 out of 5 stars	1		https://www.amaz...	₹3,040	M.R.P.
5	https://m.media-a...	Marks & Spencer	Men Sweater	5.0 out of 5 stars	1		https://www.amaz...	₹2,000	M.R.P.

< 1 ... 9 10 11 12 13 > Go to Page

5 https://m.media-am... Qube By Fort Collins Men's Cotton Crew ... 4.0 out of 5 stars 10,820 Deal of the Day https://www.amazon... ₹499

6 https://m.media-am... Qube By Fort Collins Men's Quilted Jacket 4.0 out of 5 stars 1,703 Deal of the Day https://www.amazon... ₹910

Apply

34°C

Smoke

Search

16:46

05-02-2024

amazon.in Delivering to Mumbai 400070 [Update location](#)

Search Amazon.in

*Inclusive of bank discount

Revamp your home in style

Cushion covers, bedsheet & more

Figurines, vases and more

Home storage

Lighting solutions

[Explore all](#)

Up to 60% off | Styles for men

Clothing

Footwear

Watches

Bags & wallets

End of season sale

Starting ₹99 | All your home improvement needs

Spin mops, wipes & more

Bathroom hardware & accessories

Hammers, screwdrivers & more

Extension boards, plug & more

[Explore all](#)

Up to 60% off | Styles for women

Women's Clothing

Footwear+Handbags

Automotive essentials | Up to 60% off

Cleaning accessories

Tyre & rim care

Get the perfect screen size | TVs Starting ₹6,999

Budget TVs | Up to 60%

4K TVs | Up to 24 months

United Kingdom

0.58%

United Arab Emirates

0.46%

Canada

0.43%

[See more countries →](#)

Traffic Sources

Worldwide Desktop Only Dec 2023

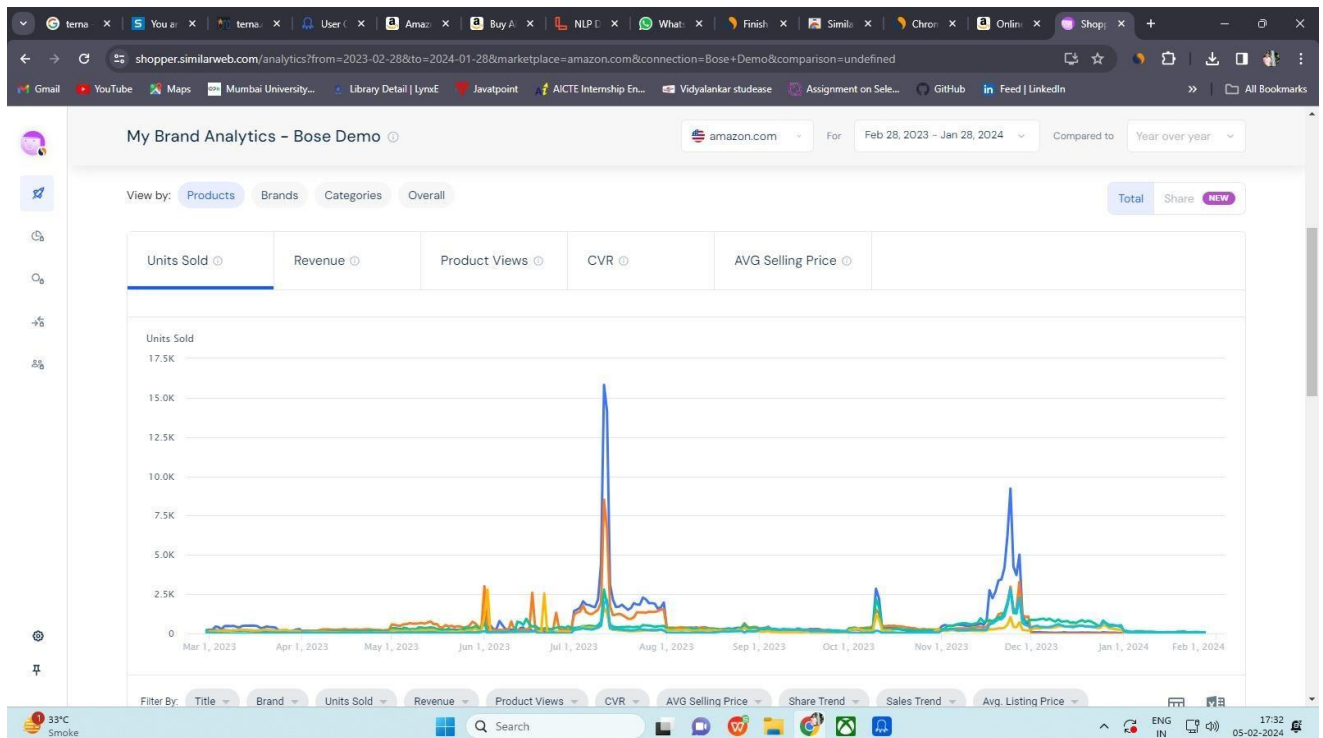
Marketing Mix

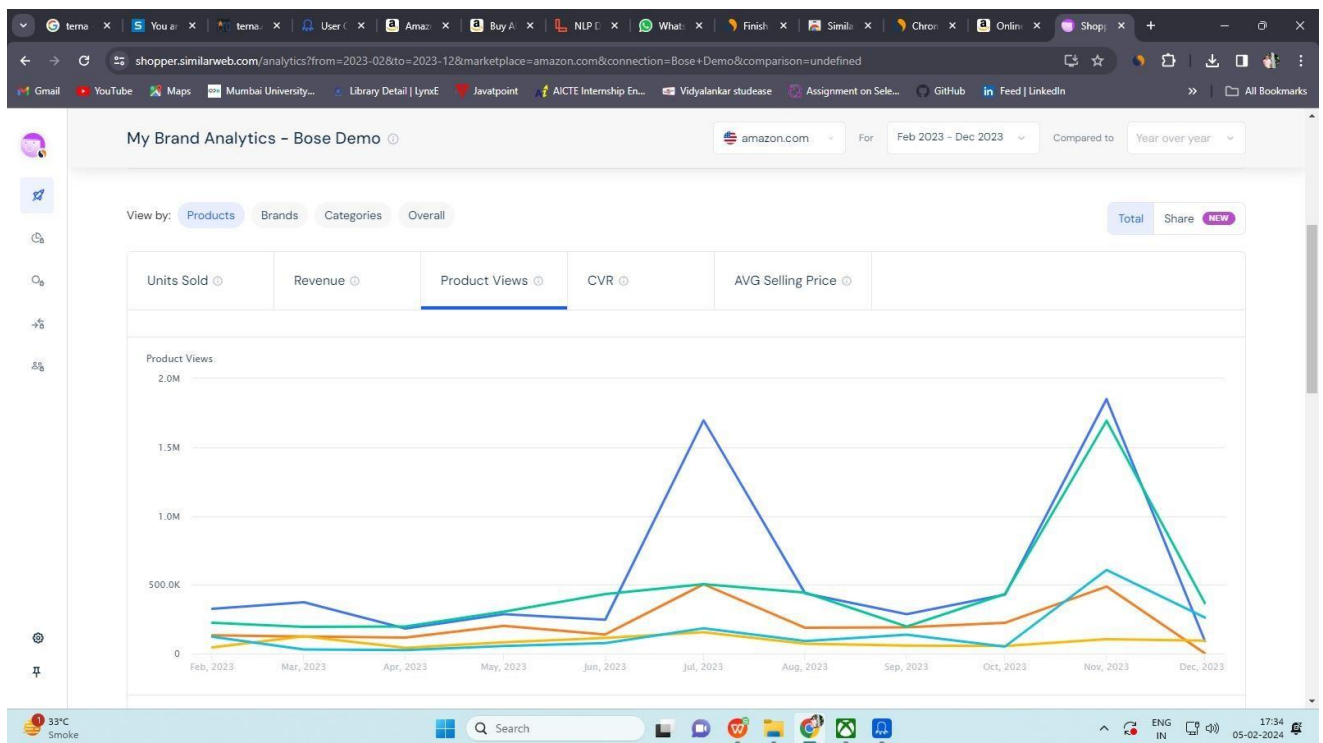
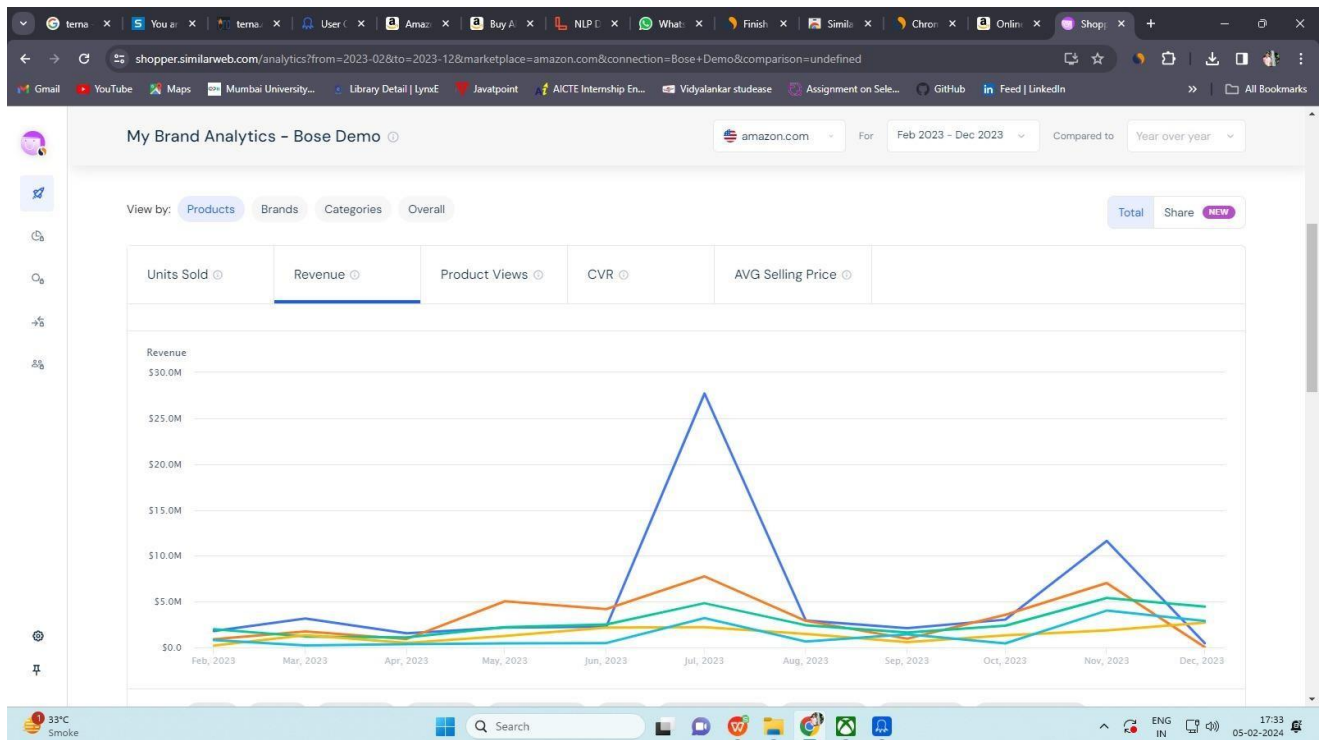
Source

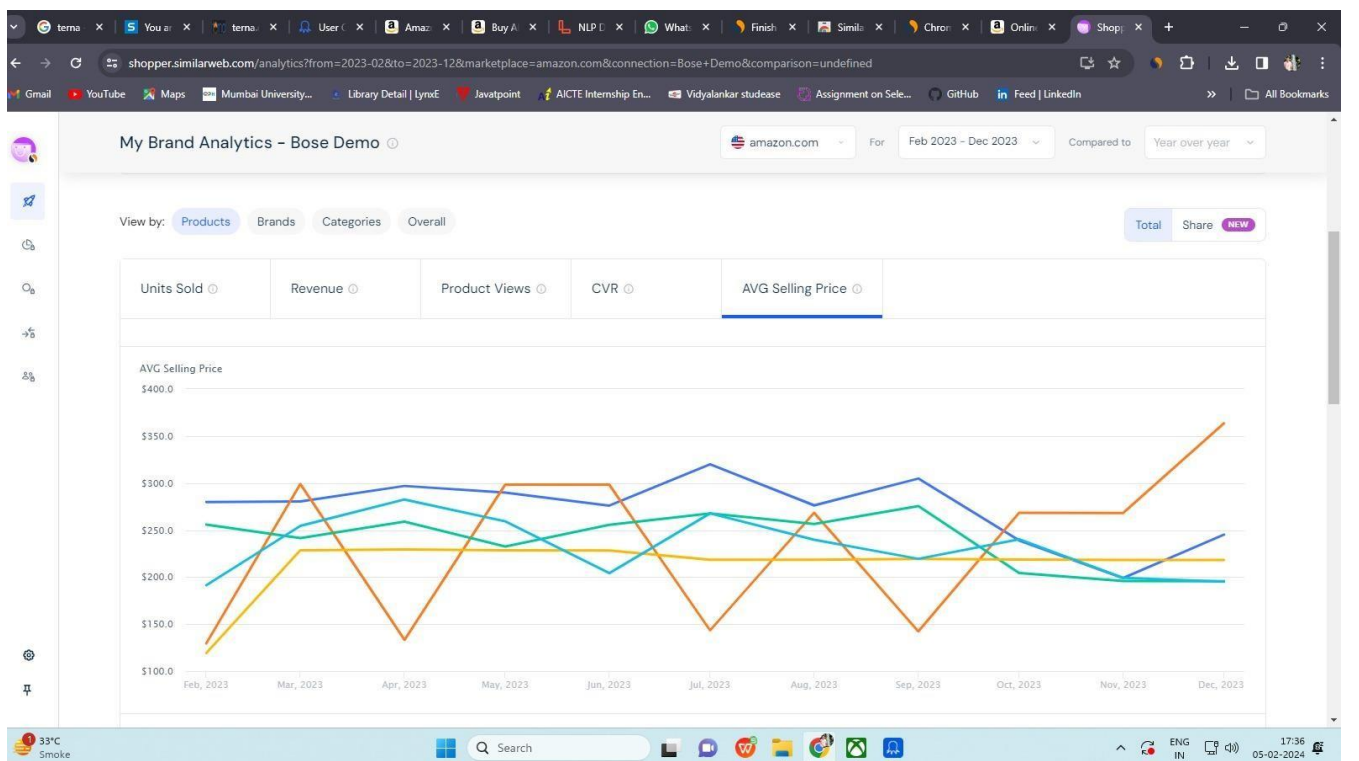
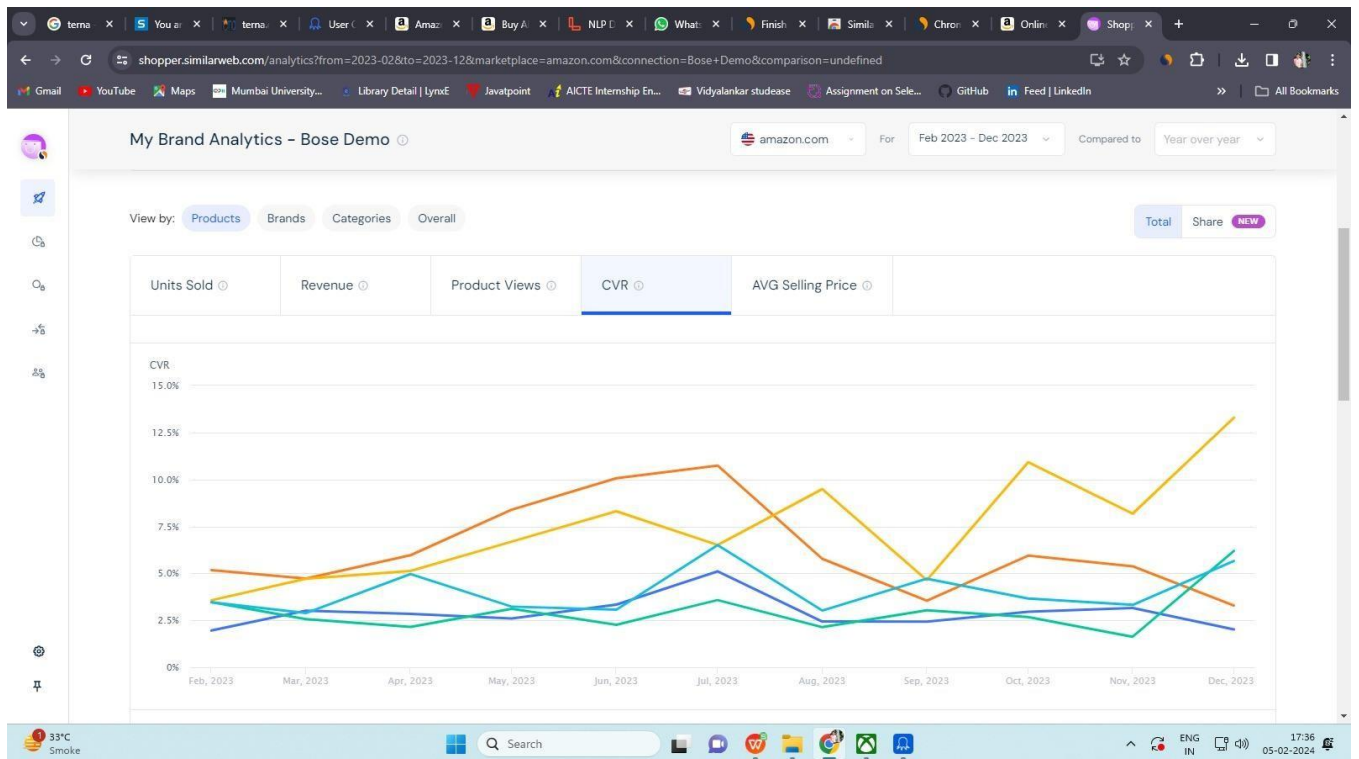
- Direct 55.25%
- Search 34.98%
- Social 4.28%
- Referrals 3.95%
- Email 1.14%
- Display Ads 0.41%

Want to see more insights?

[Go to Similarweb](#)







Web parsing through Lexalytics:

The screenshot shows the Lexalytics NLP Demo website. The header includes the Lexalytics logo, navigation links (Platform, Solutions, Tech, More, NLP Demo), and buttons for 'Contact Us', 'Get a Quote', and 'Self Assessment'. The main content area is divided into two steps:

- 1. Select Industry Pack**: A row of seven icons representing different industries: General, BusinessVOE, Pharma, Retail (selected), Airlines, Restaurants, and Hotels.
- 2. Select a text sample**: Three text boxes containing sample text. The first box is about Walmart, the second is about a clothing department, and the third is about a shoe department. The third box is selected.

A red button labeled '3. Show Analysis' is located below the text samples. The bottom of the page shows a Windows taskbar with various application icons and system information.

The screenshot shows the Lexalytics NLP Demo website after the analysis. The left sidebar contains the same selection steps as the previous screenshot, but with 'Retail' selected in the first step and 'Sample Three' selected in the second step. The main content area displays the analysis results:

JSON View (toggle switch)

Document (selected), Topics, Themes, Entities

Summary

There needs to be greater organization and space in the shoe department... This department is often very cluttered and it is difficult to try on shoes... I also had to wait for assistance at the register for a long time; however the sales associate was very apologetic for the extensive wait time; however the sales associate was very apologetic for the extensive wait time and checked me out expeditiously when I made my purchase...

Sentiment Phrase	Sentiment Score
professional	0.75
assistance	0.49
happy	0.40
difficult	-0.30

The bottom of the page shows a Windows taskbar with various application icons and system information.

terna - Google Search

You are signed in as T

terna.qualcampus.com

User Center | Octopars

Amazon.in

Buy Allen Solly Men C

NLP Demo - Lexalytics

lexalytics.com/nlp-demo/

YouTubeMapsGmail

Lexalytics

PlatformSolutionsTechMoreNLP Demo

Contact Us

1. Select Industry Pack ?
Retail

2. Select a text sample ?
Sample Three

There needs to be greater organization and space in the shoe department. This department is often very cluttered and it is difficult to try on shoes. I also had to wait for assistance at the register for a long time; however the sales associate was very apologetic for the extensive wait time and checked me out expeditiously when I made my purchase. I am happy to have this store in close proximity to my home since the Insight Retail's store at Shannon Mall closed; however I would like to see a greater emphasis placed on the quality of merchandise for professional individuals. There also needs to be more organization and space in the shoe department. Please consider carrying more of a selection of Michael Kors shoes for women.

JSON View

DocumentTopicsThemesEntities

Topic	Type	Sentiment
Product-Clothes_&Accessories-General	query	☹️
Purchase-Promotions-Sales-General	query	☹️
Store-General-Orderliness	query	☹️
Product-Availability_&Choice	query	☹️
Product-General-Satisfaction_With_Product-Positive	query	☹️

Upcoming Earnings

Search

ENG IN 17:00 05-02-2024

terna - Google Search

You are signed in as T

terna.qualcampus.com

User Center | Octopars

Amazon.in

Buy Allen Solly Men C

NLP Demo - Lexalytics

lexalytics.com/nlp-demo/

YouTubeMapsGmail

Lexalytics

PlatformSolutionsTechMoreNLP Demo

Contact Us

1. Select Industry Pack ?
Retail

2. Select a text sample ?
Sample Three

There needs to be greater organization and space in the shoe department. This department is often very cluttered and it is difficult to try on shoes. I also had to wait for assistance at the register for a long time; however the sales associate was very apologetic for the extensive wait time and checked me out expeditiously when I made my purchase. I am happy to have this store in close proximity to my home since the Insight Retail's store at Shannon Mall closed; however I would like to see a greater emphasis placed on the quality of merchandise for professional individuals. There also needs to be more organization and space in the shoe department. Please consider carrying more of a selection of Michael Kors shoes for women.

JSON View

DocumentTopicsThemesEntities

Theme	Sentiment
shoe department	☹️
extensive wait time	☹️
sales associate	☹️
close proximity	☹️
professional individuals	☹️

Upcoming Earnings

Search

ENG IN 17:01 05-02-2024

1. Select Industry Pack ?
Retail

2. Select a text sample ?
Sample Three

There needs to be greater organization and space in the shoe department, This department is often very cluttered and it is difficult to try on shoes, I also had to wait for assistance at the register for a long time; however the sales associate was very apologetic for the extensive wait time and checked me out expeditiously when I made my purchase. I am happy to have this store in close proximity to my home since the Insight Retail's store at Shannon Mall closed; however I would like to see a greater emphasis placed on the quality of merchandise for professional individuals, There also needs to be more organization and space in the shoe department, Please consider carrying more of a selection of Michael Kors shoes for women.

JSON View

Document ? Topics ? Themes ? Entities ?

Entity	Type	Sentiment
Insight Retail	Company	😊
Shannon Mall	Company	😊
Michael Kors	Person	😊

B.3 Observations and learning:

We have hence observed how we can collect data from social media sites for business analysis using web scrapping tools like Octaparse, Web crawling tools like Similarweb, Web parsing tools like Lexalytics. All these tools have the ability to capture tabular data from a website.

B.4 Conclusion:

Hence, we can conclude that we have successfully performed web scrapping, web crawling, web parsing on social media sites and gained insights from it.

B.5 Question of Curiosity

Q1. Why it is important for business managers to understand and mine social media data?

Ans: Understanding and mining social media data is crucial for business managers for several reasons:

1. **Customer Insights:** Social media platforms are rich sources of information about customer preferences, behaviors, and sentiments. Analyzing social media data allows business managers to gain valuable insights into customer opinions, needs, and trends. This information can be used to tailor products, services, and marketing strategies to better meet customer expectations.

2. **Competitive Analysis:** Monitoring social media data helps businesses keep track of their competitors' activities, customer feedback, and market trends. This knowledge allows managers to identify opportunities and threats in the market, helping them make informed strategic decisions and stay ahead of the competition.
3. **Brand Perception:** Social media provides a platform for customers to express their opinions about a brand publicly. By mining social media data, business managers can gauge the public perception of their brand and identify areas for improvement. This information is valuable for maintaining and enhancing brand reputation.
4. **Marketing Effectiveness:** Social media data can be used to evaluate the effectiveness of marketing campaigns. By analyzing engagement metrics, sentiment analysis, and other relevant data, managers can assess the impact of their marketing efforts and optimize strategies for better results.
5. **Customer Service:** Social media platforms are often used by customers to seek support or share their experiences with a product or service. Business managers can use social media data to identify common issues, address customer concerns, and improve overall customer satisfaction.
6. **Innovation and Product Development:** Social media data can provide valuable insights into emerging trends and consumer preferences. This information can guide businesses in the development of new products or the enhancement of existing ones, ensuring that offerings align with market demands.
7. **Risk Management:** By monitoring social media data, business managers can quickly identify and address potential PR crises or negative sentiment surrounding their brand. This proactive approach to risk management allows for timely intervention and damage control.
8. **Targeted Advertising:** Social media platforms offer advanced targeting options for advertising. Analyzing social media data helps businesses understand their target audience better, allowing them to create more personalized and effective advertising campaigns.

9. **Decision-making Support:** Social media data can serve as a valuable source of information for making data-driven decisions. Business managers can rely on insights derived from social media analytics to make informed choices in areas such as product development, marketing strategies, and resource allocation.

Q2. What is social media analytics, and how it is different from traditional business analytics?

Ans: Social media analytics refers to the process of collecting, analyzing, and interpreting data from social media platforms to gain insights into user behavior, sentiments, and trends. It involves the use of specialized tools and techniques to extract valuable information from the vast amount of data generated on social media.

Here are key differences between social media analytics and traditional business analytics:

1. Data Source:

- **Social Media Analytics:** Focuses specifically on data generated from social media platforms such as Facebook, Twitter, Instagram, LinkedIn, etc. It involves analyzing user interactions, mentions, likes, shares, comments, and other social media-specific metrics.
- **Traditional Business Analytics:** Encompasses a broader range of data sources, including internal sources (like sales, financial, and operational data) and external sources (like market research, customer surveys, etc.). Traditional business analytics looks at a comprehensive set of data related to various aspects of business operations.

2. Data Variety:

- **Social Media Analytics:** Involves unstructured and semi-structured data, such as text, images, videos, and user-generated content. The analysis may include sentiment analysis, topic modeling, and other techniques specific to social media data.
- **Traditional Business Analytics:** Typically deals with structured data, such as databases and spreadsheets, and may involve more structured analysis methods like reporting, data mining, and statistical analysis.

3. Metrics and KPIs:

- **Social Media Analytics:** Focuses on metrics like engagement rates, sentiment analysis, follower growth, and social reach. Key performance indicators (KPIs) often include likes, shares, comments, and click-through rates.
- **Traditional Business Analytics:** Utilizes a wide range of metrics depending on the business goals, such as revenue, profit margins, customer acquisition cost, customer retention rates, and other key financial and operational indicators.

4. Real-time Analysis:

- **Social Media Analytics:** Often requires real-time or near-real-time analysis to capture and respond to current trends, customer sentiments, and emerging issues on social media platforms.
- **Traditional Business Analytics:** May involve periodic reporting and analysis, with a focus on historical data and trends over longer time periods.

5. Use Cases:

- **Social Media Analytics:** Primarily used for understanding customer sentiments, improving social media marketing strategies, managing brand reputation, and monitoring social trends.
- **Traditional Business Analytics:** Applied across various business functions, including finance, marketing, operations, supply chain management, and human resources, to optimize processes, identify inefficiencies, and support decision-making.

Q3. What ethical issues should be considered when mining social media data?

Ans: Mining social media data raises several ethical considerations due to the sensitive nature of personal information and the potential impact on individuals and communities. Here are some key ethical issues to consider:

1. Privacy Concerns:

- **Informed Consent:** Users often share personal information with the expectation of privacy. Ensuring that users are aware of how their data will be used and obtaining informed consent is crucial.
- **Anonymization:** Stripping personally identifiable information from data or aggregating it can help protect user privacy. However, complete anonymity is challenging to achieve, and re-identification risks must be considered.

2. Transparency and Openness:

- **Disclosure of Intentions:** It is essential to be transparent about the purpose of data collection and analysis. Users should know how their data will be used, who will have access to it, and for what purposes.
- **Openness about Algorithms:** If algorithms are used in the analysis, there should be transparency about how they work to avoid hidden biases or discriminatory outcomes.

3. User Control and Consent:

- **Opt-in vs. Opt-out:** Users should have the option to opt into data collection rather than being automatically included. Providing clear and accessible options for users to control their privacy settings is crucial.
- **Data Deletion:** Users should have the ability to request the deletion of their data, and businesses should comply with such requests promptly.

4. Data Security:

- **Protection Against Breaches:** Social media platforms and businesses must take measures to secure the data they collect. Unauthorized access or data breaches can have severe consequences for user privacy.
- **Secure Storage:** Ensuring that collected data is stored securely and employing encryption methods can help prevent unauthorized access.

5. Bias and Fairness:

- **Algorithmic Bias:** Social media data analysis may involve algorithms that can introduce biases. Ensuring fairness and minimizing discrimination in algorithmic outcomes is crucial.
- **Representativeness:** Ensuring that the data used for analysis is representative of diverse demographics is essential to avoid perpetuating biases.

6. Stigmatization and Discrimination:

- **Avoiding Harm:** Data analysis should not lead to the stigmatization or discrimination of individuals or groups. Ethical considerations should involve minimizing harm and negative consequences.
- **Sensitive Information:** Handling sensitive information, such as health conditions or political affiliations, with care to prevent potential harm or misuse.

7. Data Ownership and Accountability:

- **Ownership Clarity:** Clarifying who owns the social media data and who is accountable for its ethical use is important.
- **Responsible Data Sharing:** If data is shared with third parties, ensuring that they adhere to ethical standards and guidelines is crucial.

8. Community and Cultural Sensitivity:

- **Respecting Cultural Norms:** Recognizing and respecting cultural differences and norms when analyzing social media data from diverse communities is important to avoid cultural insensitivity.
- **Community Impact:** Understanding the potential impact of data mining on communities and taking steps to mitigate negative consequences.

Q4. Compare different social media analytics tools available in the market and explain their strengths and weakness.

Ans: Here are some commonly used social media analytics tools along with their general strengths and weaknesses:

1. Hootsuite:

Strengths:

- Comprehensive social media management platform.
- Supports scheduling and publishing posts across multiple platforms.
- Provides analytics on engagement, reach, and audience growth.

Weaknesses:

- Advanced analytics may require additional paid plans.
- User interface complexity for beginners.

2. Buffer:

Strengths:

- User-friendly interface.
- Scheduling and publishing capabilities.
- Basic analytics for tracking post performance.

Weaknesses:

- Limited advanced analytics compared to some other tools.
- May not cover as many social media platforms as other tools.

3. Sprout Social:

Strengths:

- Offers a unified inbox for managing messages from different platforms.
- Robust social listening capabilities.
- Provides detailed analytics and reporting.

Weaknesses:

- Higher cost compared to some other tools.
- Advanced features may have a steeper learning curve.

4. Google Analytics:

Strengths:

- Integrates with various platforms, including social media.
- Provides in-depth website analytics and conversion tracking.
- Can track traffic from social media sources.

Weaknesses:

- Primarily focused on website analytics, may lack some social media-specific features.
- Limited social listening capabilities.

5. Brandwatch:

Strengths:

- Advanced social listening and sentiment analysis.
- Provides insights into brand mentions and industry trends.
- Customizable dashboards for in-depth reporting.

Weaknesses:

- Higher cost, may be more suitable for larger enterprises.
- Complexity might be overwhelming for smaller businesses.

6. Talkwalker:

Strengths:

- Powerful social listening and analytics.
- Real-time monitoring of brand mentions.
- Competitive benchmarking and trend analysis.

Weaknesses:

- Complexity may be challenging for beginners.
- Pricing may be on the higher side for smaller businesses.

7. Sprinklr:

Strengths:

- Comprehensive social media management and analytics.
- Suitable for large enterprises with global social media presence.
- Provides insights into customer engagement and sentiment.

Weaknesses:

- High cost, making it less accessible for smaller businesses.
- Can be overwhelming with its extensive features.

8. Socialbakers:

Strengths:

- AI-driven analytics for performance prediction.
- Competitive benchmarking and influencer analysis.
- Detailed reporting and audience demographics.

Weaknesses:

- Pricing may be a barrier for smaller businesses.
- Some users may find the interface complex.