Aim: To Study various social media platforms, analytics tool & techniques, and Business Application

Software: Youtube Analytics, Google Analytics

Theory:

This experiment will help in developing a basic understanding of social media platforms. In addition, students will be exposed to various analytics tools and techniques widely used for analysis and visualization of performance metrics. Students will understand the importance of social media analysis in various businesses. The objectives are listed as follows:

i) Learn the basic interface of popular Social Media platforms (Facebook, twitter, youtube, etc.).

ii) Use widely available Social Media analytics tools (facebook insights, google analytics netlytic, etc.)

iii) Perform Social Media Analytics and learn various techniques and important engagement metrics using google analytics, netlytic, lexalytics, etc.) iv) Study the importance of social media analytics for any business.

Procedure:

Objective 1: Create a user account on any three social media platforms and explore its Features.

Objective 2: If you have already created pages/content on Instagram or Facebook, use the 'insights' feature to learn several statistics regarding how the content has been viewed or has performed over time. Try logging in to Google Analytics Demo Account from your personal Gmail account and analyze the performance of "Google Merchandise Store". Share snapshots of the same here!

Objective 3: (i) Access NLP Demo from lexalytics.com website and analyze available sample texts regarding article/reviews, feedback, etc. (ii) Login to netlytic.org and analyze the performance of any youtube video, keyword on twitter, google sheets, RSS feed, reddit data.

Conclusion:

Characteristics of Instagram:

1) Multiple filters can be used for entertainment purpose.
2) Stories Highlight can be important for memories.
3) Video posts can be used as most successful marketing tool.
4) Augmented reality based filters are used for superimposing futuristic, unique and interactive experiences.
5) IGTV is Instagram's standalone video platform to grow following and increase engagement.

Characteristics of Twitter:

1) Tweets are the most important feature of Twitter. It allows users to share their opinion/thoughts regarding a certain topic.
2) Users can pay a subscription fee to get their accounts verified.
3) Twitter's trending page gives the user realtime information about things which are happening in the world.
4) Twitter thread feature is used by users to give information about in multiple tweets or the users uses this feature to have conversation over it.
5) Retweets features are used by users to share a particular tweet from different user on their profile.

Importance of Social Media Analytics:

1. They help you understand your audience
2. They show you what your best social networks are
3. Social data can help you create better content
4. Better Business Insights and Decisions.

Aim: To Collect Data from Social Media platforms and popular websites of various businesses.

Software: Google Chrome Web Scrapper

Theory:

Since the beginning of the web, web scraping has been the main challenge for anyone who wanted to exploit the richness of information available on the Internet. In the very beginning, very few APIs were available and people used to copy the content of websites by just using copy-paste schema. Then, some programmatic tools were created to follow links (crawling) and extract the content from web pages (scraping). The information was structured by using text patterns (regex) or DOM (Document Object Model) parsing methods. More recently, the development of semantic analysis tools and artificial intelligence enabled alternative approaches, which are much more efficient and closer to human understanding and interpretation of website content. Search engines, especially Google, have been leaders in web scraping, as they go about crawling the entire web to index content from web pages and make it available through its search engine for the entire world. However, nowadays scraping is an essential component of many applications of unstructured web content for natural language processing and other requirements. Some of the features of web crawling and scraping are listed in TABLE I.

TABLE I: Web Crawling vs Web Scraping

Web Crawling	Web Scraping
It is used for indexing of web pages.	It is used for extracting specific information.
It crawls until it visits all pages of a website.	It visits only those pages of a website that are pre-defined.
It is mostly employed in a large scale.	It is done on both small and large scales.
Deduplication is an essential part of the process.	Data de-duplication is not a necessary part.
It only needs a crawl agent.	It needs a crawl agent and a parser.

Procedure:

Objective 1: Scrap data of any two products/business/services from websites such as amazon, flipkart, myntra, practo, etc., using Google chrome web scraper/Python.

Objective 3: Extract data from (any two) social media platforms on any two products/businesses/services/topics of your choice using Google chrome web scraper/Octoparse/ netlytic.org/Python.

Objective 2: Scrap data of any products/business/services from websites such as amazon, flipkart, myntra, practo, etc., using Octoparse/Python.

Conclusion:

In this experiment we have studied about difference between Web Crawling and Web Scraping. Also we done web scraping of different websites using google chrome extension.

Aim: To understand the importance of data cleaning by pre-processing, filtering, and visualization of data from the social media platform Twitter using python.

Software: Google Colab

Theory:

The analysis is divided into four parts:

1. Importation & cleaning
2. Visualisation with WordCloud
3. Obtaining Tweet's sources
4. Sentiment analysis

1. Importation & cleaning

In this section we are going to focus on the most important part of the analysis. In general rule the tweet are composed by several string that we have to clean before working correctly with the data. I have separated the importation of package into four parts. Usually numpy and pandas are part of our toolbox.

For the visualisation we use Seaborn, Matplotlib, Basemap and finall word_cloud. In order to clean our data (Text) and sentiment analysis the most common library is NLTK. NLTK is a leading platform Python programs to workin with human language data. It exists another Natural Language Toolkit (gensim) but in our case it is not necessary to use it.

The data with which are going to work is a list of tweets with the hashtag #goodmorning.

2. Visualisation with WordCloud

Throughout this part we are going to focus on the kind of data visualisation word cloud. It always interesting to do this kind of viz in order to have a global vision of the data. The visualisation are going to do with the column "text" and "country".

3. Tweet's source

In this third part we are going to check the source of the tweets. And by the source I mean the device and the location. The purpose is to see the repartition of the tweet by deveice. As usual the first is the cleaning. The kind of device is situated at the end in the column "source". With the following example we can see that the device is just before
"/a>".

4. Sentiment Analysis

Throughout last part we are going to do an sentiment analysis. The objective is to class by type th tweets. We are going to distingush 3 kind of tweets according to their polarity score. We will have the positive tweets, the neutral tweets and the negative tweets.

Conclusion:

In this way we have studied the importance of data cleaning and pre-processing and visualization of Twitter data analysis.

Aim: To perform exploratory data analysis and visualization of (i) financial data of any business (ii) social media data of any business.

Software: MS Power BI

Theory:

Exploratory Data Analysis (EDA) is a process of describing the data by means of statistical and visualization techniques in order to bring important aspects of that data into focus for further analysis. This involves inspecting the dataset from many angles, describing & summarizing it without making any assumptions about its contents. Just like everything in this world, data has its imperfections. Raw data is usually skewed, may have outliers, or too many missing values. A model built on such data results in sub-optimal performance. In hurry to get to the machine learning stage, some data professionals either entirely skip the exploratory data analysis process or do a very mediocre job. This is a mistake with many implications, that includes generating inaccurate models, generating accurate models but on the wrong data, not creating the right types of variables in data preparation, and using resources inefficiently.

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. Additionally, it provides an excellent way for employees or business owners to present data to non-technical audiences without confusion.

In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions.The importance of data visualization is simple: it helps people see, interact with, and better understand data. Whether simple or complex, the right visualization can bring everyone on the same page, regardless of their level of expertise.

Procedure:

1. Install Power BI Desktop (with official college email IDs only)
2. Import data from an excel sheet or any source of choice.
3. Transform the data (if required) then import.
4. Create different visualizations of the data.
5. Repeat steps 2 to 5 for (i) Financial data (ii) Social Media Data of Twitter.

Conclusion: In this experiment we perform data analysis and visualization on financial data of any business.

Aim: To Study sentiments analysis from Youtube comments

Software : Google Colab

Theory :

Sentiment analysis, also referred to as opinion mining, is an approach to natural language processing (NLP) that identifies the emotional tone behind a body of text. This is a popular way for organizations to determine and categorize opinions about a product, service or idea. Sentiment analysis involves the use of data mining, machine learning (ML), artificial intelligence and computational linguistics to mine text for sentiment and subjective information such as whether it is expressing positive, negative or neutral feelings.

Sentiment analysis systems help organizations gather insights into real-time customer sentiment, customer experience and brand reputation. Generally, these tools use text analytics to analyze online sources such as emails, blog posts, online reviews, customer support tickets, news articles, survey responses, case studies, web chats, tweets, forums and comments. Algorithms are used to implement rule-based, automatic or hybrid methods of scoring whether the customer is expressing positive words, negative words or neutral ones.

Procedure:

Sentiment analysis generally follows these steps:

Collect data. The text being analyzed is identified and collected. This involves using a web scraping bot or a scraping application programming interface.

Clean the data. The data is processed and cleaned to remove noise and parts of speech that don't have meaning relevant to the sentiment of the text. This includes contractions, such as I'm, and words that have little information such as is, articles such as the, punctuation, URLs, special characters and capital letters. This is referred to as standardizing.

Extract features. A machine learning algorithm automatically extracts text features to identify negative or positive sentiment. ML approaches used include the bag-of-words technique that tracks the occurrence of words in a text and the more nuanced word-embedding technique that uses neural networks to analyze words with similar meanings.

Pick an ML model. A sentiment analysis tool scores the text using a rule-based, automatic or hybrid ML model. Rule-based systems perform sentiment analysis based on predefined, lexicon-based rules and are often used in domains such as law and medicine where a high degree of precision and human control is needed. Automatic systems use ML and deep learning techniques to learn from data sets. A hybrid model combines both approaches and is generally thought to be the most accurate model.

These models offer different approaches to assigning sentiment scores to pieces of text. Sentiment classification. Once a model is picked and used to analyze a piece of text, it assigns a sentiment score to the text including positive, negative or neutral. Organizations can also decide to view the results of their analysis at different levels, including document level, which pertains mostly to professional reviews and coverage; sentence level for comments and customer reviews; and sub-sentence level, which identifies phrases or clauses within sentences.

Conclusion : In this particular experiment we lernt about the sentiment analysis from Youtube comment which refered to an option mining that identifies the emotional tone behind a body of text .

Aim : To Perfrom social network analysis of facebook data

Software : Python colab

Theory :

A social graph is a diagram that illustrates interconnections among people, groups and organizations in a social network. The term is also used to describe an individual's social network. When portrayed as a map, a social graph appears as a set of network nodes that are connected by lines.

The Social Network graph shows:

Entity-to-entity links: You see all the entities related to the main (hub) entity. However, the attributes that link the entities do not display on the graph but are accessible by using the Attribute Explorer in combination with the graph.

Relationship clusters: The Social Network graph is unique in that it displays the related entities in groups or clusters. This graph can help you see all the relationship clusters a particular entity belongs to and look for patterns in among the clusters and relationships.

You can expand the graph to show all the related entities for any entity. Each time you show all entities related to a particular entity, that entity node becomes the hub entity in a new relationship cluster.To maintain the integrity of each relationship cluster, an entity can be displayed on the graph multiple times in multiple relationship clusters. But each entity displays in each relationship cluster only once. To see every relationship cluster the entity is part of, select the entity by clicking on that node. The interior of the selected entity node changes to blue in each relationship cluster that the entity is part of.

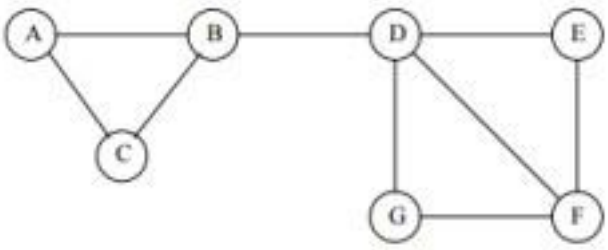


Figure: Example of a small social network

Procedure:

Clustering of Social-Network Graphs:

Clustering of the graph is considered as a way to identify communities. Clustering of graphs involves following steps:

Degree Centrality

The people most popular or more liked usually are the ones who have more friends. Degree centrality is a measure of the number of connections a particular node has in the network. It is based on the fact that important nodes have many connections. NetworkX has the function degree_centrality() to calculate the degree centrality of all the nodes of a network.

Eigenvector Centrality

It is not just how many individuals one is connected too, but the type of people one is connected with that can decide the importance of a node. In Delhi Roads whenever the traffic police capture a person

Betweenness Centrality

The Betweenness Centrality is the centrality of control. It represents the frequency at which a point occurs on the geodesic (shortest paths) that connected pair of points. It quantifies how many times a particular node comes in the shortest chosen path between two other nodes. The nodes with high betweenness centrality play a significant role in the communication/information flow within the network. The nodes with high betweenness centrality can have a strategic control and influence on others. An individual at such a strategic position can influence the whole group, by either withholding or coloring the information in transmission. Networkx has the function betweenness_centrality() to measure it for the network. It has options to select if we want betweenness values to be normalized or not, weights to be included in centrality calculation or not, and to include the endpoints in the shortest path counts or not.

Conclusion: In this experiment we learnt about Social network analysis of face book data

Aim: To Create a dashboard in PowerBI for Superstore Dataset.

Software: MS PowerBI

Theory:

Power BI offers interactive and dynamic features required for creating interactive dashboards. These dashboards, which are simply a collection of visuals, can be built with a deep level of interactivity and are accessible in various formats to consumers. Since they are usually a single page, Power BI dashboards need to be well-designed highlights of an entire story.

It is also important to note that Power BI dashboards are quite different from Power BI reports. For example:

Power BI reports are available on Power BI Desktop and Power BI service, while Power BI dashboards can only be found on Power BI service.

Reports can be multi-paged, while dashboards are single-paged highlights.

Procedure:

Importing Data

The first step in building a Power BI dashboard is to import the dataset that will be used to build the report. You can connect to a variety of data sources, including Excel worksheets, databases, the web, and cloud services.

Opening a Report from the Uploaded Data

Notice that from the workspace image, there are two file types of the same name. One is the dataset and the other is the dashboard.

Click on the Superstore.xlsx dashboard. On the blank canvas that pops up, click on the dataset name ‘Superstore.xlsx’.

Creating a Tile and ‘Pin to a Dashboard’

Power BI report service, just like the desktop version, includes a variety of page formatting options, including visuals, shapes, and images, that can help your report stand out. One of the most efficient ways to identify and communicate insights is to use Power BI to create visuals.

Conclusion : In this experiment we create a dashboard in powerBI for superfore dataset which offer interactive and dynamic feature required for creating interactive dashboard

Experiment 8

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

Software Required: Canva, Visme or free version of any social media ad making tool.

Results:

Spotify #Wrapped Campaign:

Spotify's Wrapped campaign brought customer data to life. Spotify used many channels, including different types of social media ads, to celebrate their users and music in general. Music carries a certain nostalgia; Spotify used this to humanise its brand, bringing personal taste and self-expression to the fore. It involved Subscribers receiving their personalised listening data from the past yearand billboards created accordingly.Users could also share their own data on social media directly from the Spotify app. This style of user-generated content also extended to artists sharing their success data over the year, talking to their huge social media following. Like Maroon 5 and Marianas Trench:

Apple #ShotOniPhone Campaign:

n 2015, Apple encouraged users to post photos they had shot on their iPhones as part of a contest. The prize: a chance for their work to be featured on one of 10,000 billboards across the world.In 2017, Apple opened their Instagram account to keep the #ShotOniPhone momentum going as sort of an organic advertising campaign intended to generate massive amounts of user-generated content. It worked! The hashtag #ShotOniPhone has over 16 million entries on Instagram alone. Apple cleverly made the most of their huge user base and invested in user-generated content to promote the technology and the Apple lifestyle. Very few of the photos taken with iPhones have iPhones in them, and yet they're obviously marketing the iPhone! it clearly delivers results. It's worth nothing, as well, that this is a kind of social media advertising that Apple doesn't put much budget or effort into – it's an ongoing organic exposure campaign that basically runs itself at this point. Nice!

Starbucks #UnicornFrappuccino Campaign:

As part of the Frappuccino Happy Hour rollout, Starbucks launched the Unicorn Frappuccino for one week only during April 2017. It was an Instagrammable succes that drove significant trafficIn this campaign, Starbucks did their user research and aimed their ads and organic posts at an audience renowned for taking photos of their food – Millennials. They designed a glittery pink and purple frozen drink that looked awesome in photos and made it available for a tiny period in select stores. Manufacturing scarcity and inciting the fear of missing out is one of the oldest tricks in the book. Combining this with social media can create the foundation for a viral sensation.As a result of this campaign, global same-store sales in America went up 3% for the second quarter in 2017. What's more, the Unicorn Frappuccino, and its hashtag, generated nearly 155,000 Instagram posts during that time period. .

Procedure to create ad using canva:

- Open and sign up to canva
- In the social media tab, choose Instagram post then click on 'create new Instagram poet'.
- A blank template appears then click on element and drag the necessary elements like circle, stalk images, text, etc.
- After a basic layout has been created make changes according to brand requirement and specifications.
- Once, the advertisement d has been created then download it to a local directory.
- For creating a advertisement on any social media platform canva will generate a template accordingly in which you can make changes to satisfy the needs of the client.

Outputs:

Facebook post

Instagram post

Conclusion

In this experiment, we analyzed different social media campaigns and created one for social media using canva.

Experiment 9

Aim: To analyze competitor activities using social media data.

Software Required: Competitor analysis web tools such as 'Similarweb.com', 'Semrush.com', etc.

Theory:

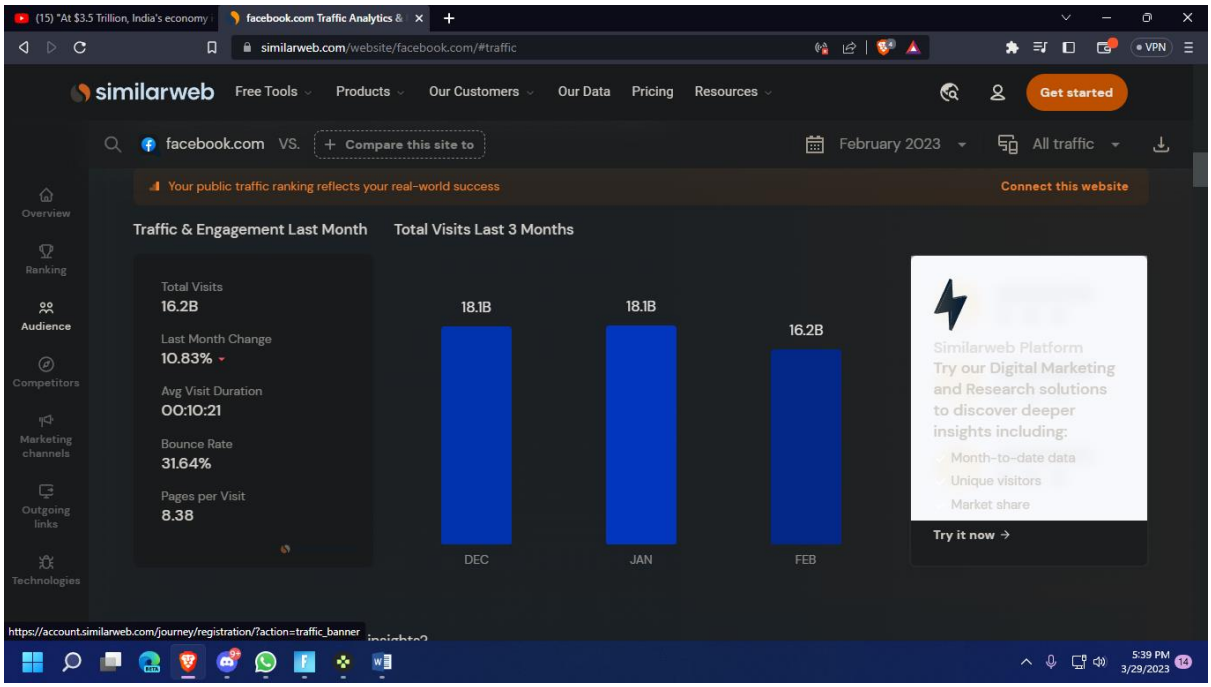
SimilarWeb is an web analytics company specializing in web traffic and performance. Similarweb develops tools that enable the analysis of the traffic and behavior of users on websites and apps.Similarweb ranks websites and apps based on traffic and engagement metrics. Its ranking is calculated according to the collected datasets and updated on a monthly basis with new data. The ranking system covers 210 categories of websites and apps in 190 countries and was designed to be an estimate of a website's popularity & growth potential. The company ranks websites based on traffic and engagement data, and ranks apps in the App Store (iOS/iPadOS) and Google Play Store based on installs & active user data. Analytics from SimilarWeb are comparable (albeit usually lower) than from Google Analytics.

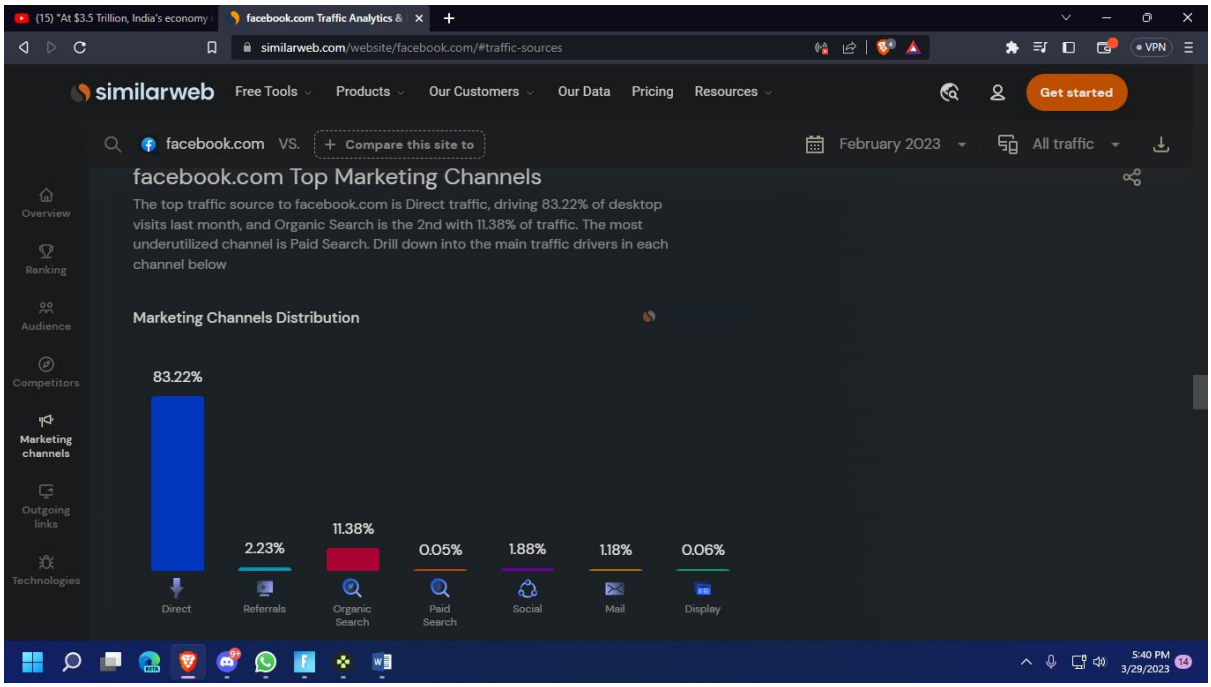
Procedure:

- When you open similerweb, type the name of the website for which you want to see website metrics.
- Once loaded, It will show you an overview and you can navigate through key metrics by sidebar which lists all of them accordingly.
- To compare websites metrics with other competitors navigate to the ']compititions
- ' Tab which will show websites top compitors and to compare with a specific compittitor click on the competition company name and it will give you metric by metric comparison between the two websites.

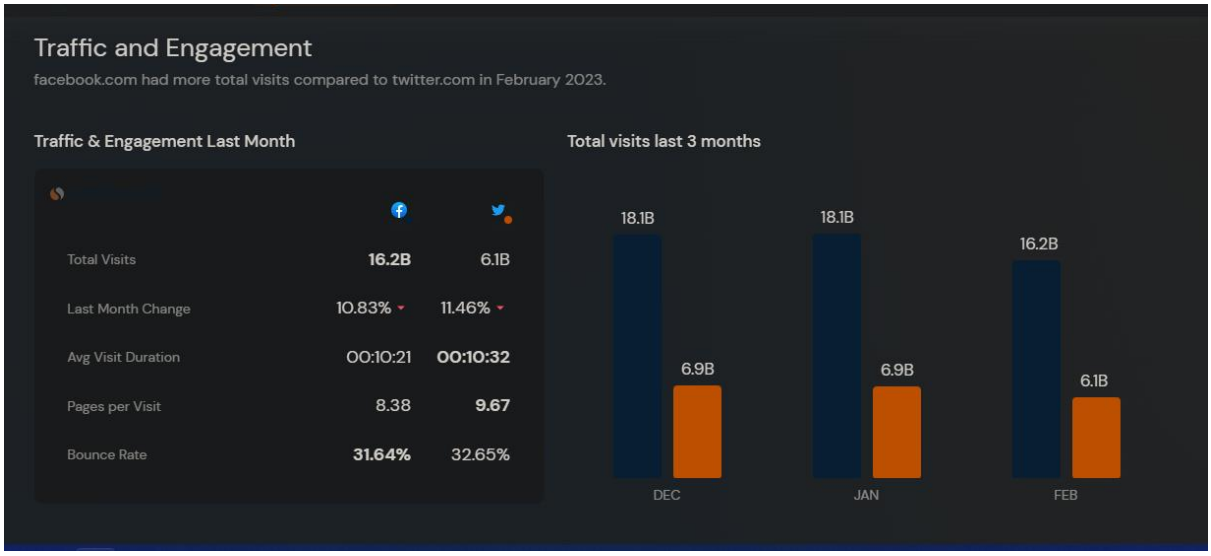
Output:

Website performance metrics(facebook)

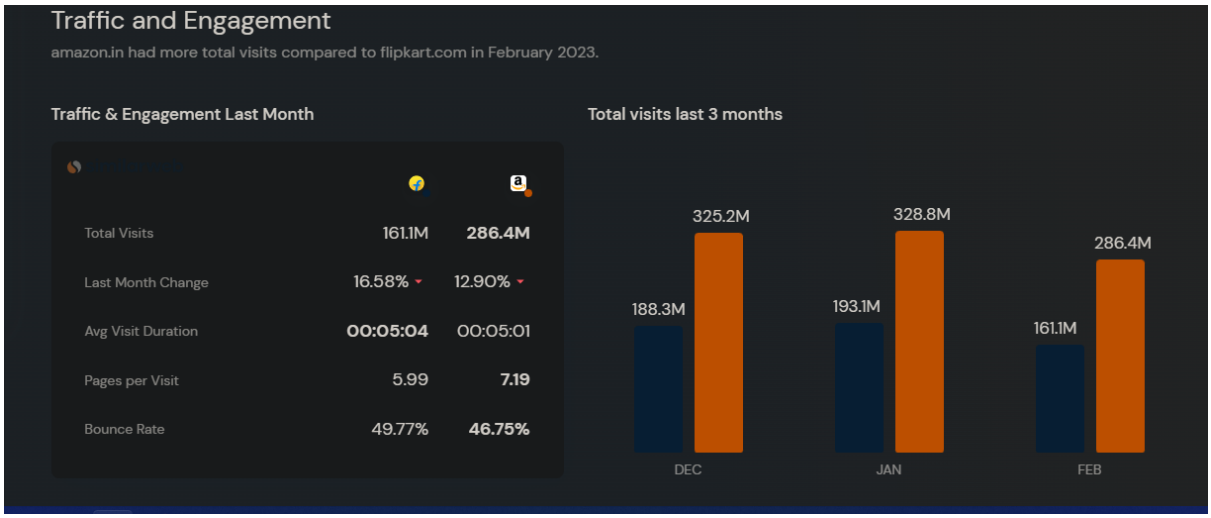

similarweb
Free Tools
Products
Our Customers
Our Data
Pricing
Resources
Get started
facebook.com VS.
Compare this site to
February 2023
All traffic
Your public traffic ranking reflects your real-world success
Connect this website
Overview
Ranking
Audience
Competitors
Marketing channels
Outgoing links
Technologies
Traffic & Engagement Last Month
Total Visits
16.2B
Last Month Change
10.83%
Avg Visit Duration
00:10:21
Bounce Rate
31.64%
Pages per Visit
8.38
Total Visits Last 3 Months
18.1B
18.1B
16.2B
DEC
JAN
FEB
Similarweb Platform
Try our Digital Marketing and Research solutions to discover deeper insights including:
Month-to-date data
Unique visitors
Market share
Try it now
https://account.similarweb.com/journey/registration/?action=traffic_banner


similarweb
Free Tools
Products
Our Customers
Our Data
Pricing
Resources
Get started
facebook.com VS.
Compare this site to
February 2023
All traffic
facebook.com Top Marketing Channels
The top traffic source to facebook.com is Direct traffic, driving 83.22% of desktop visits last month, and Organic Search is the 2nd with 11.38% of traffic. The most underutilized channel is Paid Search. Drill down into the main traffic drivers in each channel below
Overview
Ranking
Audience
Competitors
Marketing channels
Outgoing links
Technologies
Marketing Channels Distribution
83.22%
2.23%
11.38%
0.05%
1.88%
1.18%
0.06%
Direct
Referrals
Organic Search
Paid Search
Social
Mail
Display

Comparison between facebook and twitter:



Comparison between amazon and flipkart:



Conclusion

In this experiment, we used similerweb to compare and learn about various social media and e-commerce websites key analytics.

Experiment 10

Aim: To implement sentiment analysis using BERT.

Software Required: python, google colab

Theory:

Natural language processing (NLP) is one of the most cumbersome areas of artificial intelligence when it comes to data preprocessing.BERT stands for Bidirectional Encoder Representations from Transformers and it is a state-of-the-art machine learning model used for NLP tasks. Jacob Devlin and his colleagues developed BERT at Google in 2018. Devlin and his colleagues trained the BERT on English Wikipedia (2,500M words) and BooksCorpus (800M words) and achieved the best accuracies for some of the NLP tasks in 2018. There are two pre-trained general BERT variations: The base model is a 12-layer, 768-hidden, 12-heads, 110M parameter neural network architecture, whereas the large model is a 24-layer, 1024-hidden, 16-heads, 340M parameter neural network architecture.

Transformers provides thousands of pretrained models to perform tasks on texts such as classification, information extraction, question answering, summarization, translation, text generation, etc in 100+ languages. Its aim is to make cutting-edge NLP easier to use for everyone.

Procedure:

- We install necessary dependencies lioke transformers,huggingface
- Load the BERT Classifier and Tokenizer alıng with Input modules;
- Download the IMDB Reviews Data and create a processed dataset (this will take several operations;
- Configure the Loaded BERT model and Train for Fine-tuning.
- Make Predictions with the Fine-tuned Model.

Output:

Model training:

```
model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=3e-5, epsilon=1e-08, clipnorm=1.0),
              loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True),
              metrics=[tf.keras.metrics.SparseCategoricalAccuracy('accuracy')])

model.fit(train_data, epochs=2, validation_data=validation_data)
```

```
Epoch 1/2
206/Unknown - 10219s 49s/step - loss: 0.3970 - accuracy: 0.8192
```

Resulr:

```
This was an awesome movie. I watch it twice my time watching this beautiful movie if I have known it was this good
 Positive
One of the worst movies of all time. I cannot believe I wasted two hours of my life for this movie :
 Negative
```

Conclusion

In this experiment, we used BERT pretrained model to perfrom sentiment analysis.

Practical Evaluation Sheet

Experiment No: 8

Title: To design the creative content for promotion of your business on social media platform.

Student Name:	SHREYASH VIJAY MUKTALKAR
Roll No.:	19CE7018
Div./Batch	C/C3

CO’s	
PO’s Mapping:	
PSO Mapping:	

Evaluation:

Preparedness	Performance	oral	Additional Learning	Overall Grade

Faculty Sign with date

Practical Evaluation Sheet

Experiment No: 9

Title: To analyze competitor activities using social media data.

Student Name:	SHREYASH VIJAY MUKTALKAR
Roll No.:	19CE7018
Div./Batch	C/C3

CO’s	
PO’s Mapping:	
PSO Mapping:	

Evaluation:

Preparedness	Performance	oral	Additional Learning	Overall Grade

Faculty Sign with date

Practical Evaluation Sheet

Experiment No: 10

Title: To develop NLP model based sentiment classifier using the IMDB movie reviews dataset, Tensorflow and hugging face transformers.

Student Name:	SHREYASH VIJAY MUKTALKAR
Roll No.:	19CE7018
Div./Batch	C/C3

CO's	Implement penetration testing using forensics tools
PO's Mapping:	PO1, PO2, PO3, PO4, PO5, PO6, PO9, PO10,PO12
PSO Mapping:	PSO1, PSO2,PSO3

Evaluation:

Preparedness	Performance	oral	Additional Learning	Overall Grade

Faculty Sign with date