SMA PRACTICAL

<https://colab.research.google.com/drive/1yhQQnbSfVsMPDPk37hJiM92IdMa820ds#scrollTo=2nmJrQn3IpoC>

* **Exp 2**

**Scraping** is extracting data from websites, often using automated tools to collect text, images, links, etc.

**Crawling** is systematically browsing the web to discover and index pages, used by search engines like Google.

**Parsing** is analyzing structured data (like HTML) to extract specific information during scraping.

Step 1: The video\_id is the unique identifier for a YouTube video. You can extract it directly from a YouTube video URL.

Step 2: How to Get api\_key: To use the YouTube Data API, you need to generate your own API key from Google Cloud Console.

🔧 Steps to Get Your API Key: Go to: [https://console.cloud.google.com/](https://www.google.com/url?q=https%3A%2F%2Fconsole.cloud.google.com%2F)

Sign in with your Google account. Create a new project (or use an existing one). Go to APIs & Services > Library. Search for YouTube Data API v3 and enable it. Then go to APIs & Services > Credentials. Click "Create Credentials" > API key. Copy your generated API key.

Step 3: Final Working URL Construction: Once you have both values, plug them into the URL

* **Exp 3**

This experiment demonstrates how to scrape a web article and perform bigram frequency analysis using Python. The focus is on extracting meaningful text content from a webpage, processing the textual data, and visualizing the most frequent word pairs (bigrams) to understand key themes.

Web Scraping: The requests library is used to fetch the HTML content of a Medium article. BeautifulSoup parses the HTML to extract the article's title and body text from the tags.

Data Storage: The cleaned and combined article text, including the title, is saved into a local text file (medium\_article.txt) for future reference or processing.

Text Vectorization: Using CountVectorizer from sklearn, bigrams (two-word phrases) are extracted from the article. Common English stopwords are removed to ensure relevance.

Bigram Analysis & Visualization: The frequency of each bigram is calculated, and the top 8 most frequent ones are identified. These bigrams are then visualized using a horizontal bar chart with matplotlib, providing insight into the most emphasized topics or phrases within the article.

* **Exp 4**

Exploratory Data Analysis (EDA) is an approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations. An EDA is a thorough examination meant to uncover the underlying structure of a data set and is important for a company because it exposes trends, patterns, and relationships that are not readily apparent.

The four types of EDA are: Univariate non-graphical, Multivariate non-graphical, Univariate graphical, Multivariate graphical.

Techniques and Tools: There are a number of tools that are useful for EDA, but EDA is characterized more by the attitude taken than by particular techniques.

Typical graphical techniques used in EDA are:

Box plot, Histogram, Multi-vari chart, Run chart, Pareto chart, Scatter plot (2D/3D), Stem-and-leaf plot, Parallel coordinates, Odds ratio, Heat map, Bar chart,  
Horizon graph, Dimensionality reduction: Multidimensional scaling, Principal component analysis (PCA), Multilinear PCA, Iconography of correlations

* **Exp 5**

Content-based social media analytics involves analyzing the actual content (text, emojis, images, audio, and video) shared on social media platforms to gain insights for business decision-making. This approach helps businesses understand their audience, monitor brand perception, identify emerging trends, and improve engagement.

In this experiment, we use Brand24, a social media monitoring tool, to analyze various types of content shared online. It allows for tracking brand mentions, analyzing sentiment, identifying popular topics and hashtags, and recognizing influencers and visual elements (like logos) in media.

The analysis includes:

Topic Analysis: Understanding what topics people are discussing.

Issue Analysis: Detecting customer concerns or complaints.

Trend Analysis: Spotting new or growing conversations.

Sentiment Analysis: Determining public opinion (positive, negative, neutral).

Multimedia Analytics: Analyzing images, videos, and audio content to understand visual and audio engagement.

By using Brand24, businesses can develop a complete content-based analytics model to enhance their marketing strategies, improve customer relationships, and stay competitive in the market.

* **Exp 6**

Structure-based social media analytics focuses on understanding the relationships and interactions between users in a social network. It helps businesses uncover patterns like communities, user influence, and network dynamics. This model uses graph theory to represent users as nodes and their interactions as edges.

Using NetworkX, a Python library for network analysis, we can perform:

Degree Centrality: Identifies nodes with the most direct connections.

Betweenness Centrality: Finds nodes that act as bridges in the network.

Closeness Centrality: Measures how close a node is to all other nodes.

Eigenvector Centrality: Evaluates a node’s influence based on connections to other influential nodes.

These measures help businesses detect influential users and communities, allowing for better targeting and engagement strategies.

* **Exp 7**

Power BI is a business intelligence and data visualization tool developed by Microsoft. It allows users to connect to real-time data sources, transform data, and create interactive dashboards and reports. In the context of social media analytics, Power BI helps visualize trends, audience engagement, and platform performance across channels like YouTube, Twitter, and Instagram.

Using Power BI, businesses can build dashboards from structured data (e.g., Excel or JSON files) and apply filters, slicers, and visuals (such as charts, tables, and cards) to gain insights and make data-driven decisions.

Attributes to select:

line graph -> location / age and gender

* **Exp 8 CANVA**
* **Exp 9 GOOGLE TRENDS**

Competitor analysis is the process of identifying and evaluating key competitors to understand their strategies, strengths, and weaknesses. It helps businesses improve their positioning, identify market gaps, and stay ahead in the industry. Tools like Google Trends and SimilarWeb assist in analyzing search popularity, website traffic, audience behavior, and engagement metrics. This analysis supports informed decision-making and the development of more effective marketing strategies.

Google trends

* **EXP 10**

Social media text analytics is the process of extracting insights from customer reviews and comments to improve existing products or services. It helps businesses understand customer sentiment, identify trends, and make data-driven decisions.

Steps:

Data Collection: Gather text data from social media platforms using APIs or web scraping.

Text Preprocessing: Clean and normalize the text (remove URLs, punctuation, lowercase conversion, etc.).

Sentiment Analysis: Use NLP models (like VADER or ML classifiers) to classify reviews as positive, negative, or neutral.

Insight Generation: Analyze the sentiment trends to identify strengths, weaknesses, and customer expectations.

This method helps businesses continuously improve based on real customer feedback