

<b>Project Title</b>	<b>Comment2Likes: Estimating Video Likes using Comment Data</b>
<b>Technologies</b>	<b>Data Cleansing, EDA, Visualization, NLP</b>
<b>Domain</b>	<b>Data Science</b>

### Problem Statement:

The objective of the "Comment2Likes: Estimating Video Likes using Comment Data" project is to develop a predictive model that can estimate the number of likes a video will receive based on the information extracted from its comments. By leveraging the valuable insights present in user comments, this project aims to provide content creators, marketers, and platform administrators with a tool to gauge the potential popularity of their videos and optimize their content strategies accordingly.

### About Data:

#### Context

This dataset contains two files for analyzing the relationship between the popularity of a certain video, the most relevant/liked comments of said video and like counts

### File Descriptions

#### **videos-stats.csv:**

This file contains some basic information about each video, such as the title, likes, views, keyword, and comment count.

#### **comments.csv:**

For each video in **videos-stats.csv**, **comments.csv** contains the top ten most relevant comments as well as said comments' sentiments and likes.

### Column Descriptions

#### **videos-stats.csv:**

- **Title:** Video Title.
- **Video ID:** The Video Identifier.
- **Published At:** The date the video was published in YYYY-MM-DD.
- **Keyword:** The keyword associated with the video.

- **Likes:** The number of likes the video received. If this value is -1, the likes are not publicly visible.
- **Comments:** The number of comments the video has. If this value is -1, the video creator has disabled comments.
- **Views:** The number of views the video got.

**comments.csv:**

- **Video ID:** The Video Identifier.
- **Comment:** The comment text.
- **Likes:** The number of likes the comment received.
- **Sentiment:** The sentiment of the comment. A value of 0 represents a negative sentiment, while values of 1 or 2 represent neutral and positive sentiments respectively.

**Dataset:**

Dataset\_Link: [Data Link](#)

**Problem to be answered:**

Create a machine learning model that allows content creators, marketers, and platform administrators to gain insights into the potential popularity of their videos based on comment information. This predictive capability can empower decision-making processes, content optimization, and marketing strategies, leading to improved engagement and success on video-sharing platforms.

**Note:**

After completion of all the task you need to create a PowerPoint presentation

That should contain the:

1. Problem Statement
2. Tools Used
3. Approaches
4. EDA Insights

### **Project Evaluation metrics:**

- Project evaluation will be done in the live session and have to showcase the approaches done to complete the project
- You are supposed to write a code in a modular fashion (in functional blocks)
- Maintainable: It can be maintained, even as your codebase grows.
- Portable: It works the same in every environment (operating system)
- You have to maintain your code on GitHub.(Mandatory)
- You have to keep your GitHub repo public so that anyone can check your code.(Mandatory)
- Proper readme file you have to maintain for any project development(Mandatory)
- You should include basic workflow and execution of the entire project in the readme file on GitHub
- Follow the coding standards: <https://www.python.org/dev/peps/pep-0008/>

