**Title: YouTube Trending Video Analytics**

YouTube is among the most widely used platforms for content consumption especially in France where communication plus entertainment rely on digital media. Since this topic was chosen, content creators, marketers, and analysts can make better decisions through a comprehension of trending video patterns. A great variety of data points came from the dataset. On account of this, the dataset was ideal for those people who wanted to practice the real-world data cleaning, translation, sentiment analysis, SQL querying, also data visualization by using tools such as Python and Tableau.

This project explores what makes a video trend on YouTube in the French region. The original dataset came with a lot of challenges including text in French, special characters, and inconsistent formatting which made it tough to analyze right away. To make sense of the data, I cleaned it thoroughly, translated non-English titles into English, and organized it in a way that allowed for meaningful analysis. This helped uncover valuable insights into how viewers in France interact with content, what types of videos perform best, and how long they stay popular. The project also looks at what key factors might influence a video's popularity and how long it stays on the trending list.

**The project set out to explore several important questions:**

* Which types of video content tend to get the most views while trending?
* How long do videos typically stay on the trending list in France?
* Is there a connection between the number of views and how long a video remains trending?
* What kind of sentiment is most common in trending video titles?
* Which channels or creators consistently perform the best in terms of views and engagement?
* What data cleaning steps are needed to make the dataset usable for analysis?

**Tools and Technologies Used:**

* Python and excel for cleaning data, ex., **Standardize** the data if not Standardized, filling the **missing values**, removing **special character** and changing the **language to English**.
* SQL to calculate the **average number of views** for each video category to find out which categories perform best.
* **In Tableau,**

Categories ranked by average view - Bar Chart

Average trending days over time - Line Chart

Total Views per Country - Map

Popular Categories by Country - Side-By-Side Bar Chart

I got the dataset from Kaggle, from a public collection called **“Trending YouTube Video Statistics.”** For this project, I specifically used the **France (FR)** dataset. It includes detailed information on thousands of YouTube videos that were trending in France like their titles, number of views, likes, dislikes, tags, when they were published, and more.

**Challenges faced:**

* Many French video titles had accented characters, which caused issues during processing in SQL and Python. We cleaned them using regex while making sure the original meaning stayed intact.
* There was no column named for region in the dataset and I made a separate column for this dataset with region, then merged the region column with the actual dataset.
* Since majority of the data was in French, I had to translate it for sentiment analysis, but some automatic translations weren’t always accurate.
* Due to the large dataset, the data was not clean and I had to come back to clean it every time moved forward with my project. So, I faced difficulty in tableau not able to visualize properly and SQL not accepting the dataset with different language.

**What did I do with the data?**

* Imported the dataset from an Excel file using **Python (Pandas)**.
* Performed **data cleaning**, including removal of special characters, translation of non-English (French) titles into **English**, elimination of duplicates and null values.
* Standardized text columns to ensure consistency (e.g., removing accents and foreign scripts).
* Created a **region column** by generating 40,725 rows with French regions, randomly repeated.
* Used **sentiment analysis** on video titles to categorize them into **positive, neutral, or negative**.
* Saved all cleaned and translated data into new **Excel files** for visualization and analysis.

**The SQL queries I used?**

Used SQL to rank categories by average views and identify top channels; created bar, pie, histogram, line, and scatter plots to visualize trends, sentiments, and relationships in the data.

**Conclusion:**

This project offered meaningful insights into how YouTube videos trend in France. It became clear that titles with a positive tone are more likely to appear on the trending list. Videos that stayed trending for a longer time also tended to gather more views, showing that visibility over time plays a big role in popularity. Overall, the analysis helped us understand how factors like sentiment and how long a video remains trending can impact its success.

To make the project even better, using actual regional data instead of randomly assigned locations would give a clearer picture of how trends vary across different parts of France. The sentiment analysis could also be improved to better capture the tone and meaning behind video titles. Adding some predictive features to forecast which videos might trend in the future would make the analysis more powerful. Including real-time data and looking at other engagement metrics like watch time could offer deeper insights. Lastly, building an interactive dashboard would make it easier for anyone to explore and understand the results.