YouTube Trends Analysis Dashboard: Final Project Document

***Group Members**: Shruthi, Shishir, Nujoum, Sakshi, Nadifa*

Project Goals

The YouTube Trends Analysis Dashboard aims to address challenges faced by Alex Carter, an emerging content creator, in understanding the dynamics of audience engagement and content performance. The primary objectives include:

- Identifying actionable insights to improve Alex's content strategy.
- Analysing trends in content categories, engagement metrics, and regional differences.
- Helping Alex align his content with high-performing trends to boost visibility and audience engagement.

The intended users of this dashboard are Alex Carter and other YouTube creators who aim to refine their strategies for increased engagement, regional penetration, and trending visibility. The dashboard provides data-driven recommendations to optimize content categories, tags, and publishing strategies, enhancing creators' chances of success.

Team Roles

Shishir: Managed data preprocessing, including cleaning and validating YouTube trending data on both python and Prep-builder and worked on BIS Design documentation.

Shruthi: Handled Text-preprocessing for word clouds, designed Content Analysis dashboard and worked on BIS Design and final documentation.

Nujoum: Supported Text-preprocessing for sentiment Analysis, and designed Geographic Insights dashboard and worked on BIS Design and final documentation.

Sakshi: Designed Overview dashboard and responsible for formatting BIS Design and final documentation.

Nadifa: Worked on Engagement Analysis dashboard, responsible for collating and finalizing the design of all dashboards and worked on BIS Design documentation.

Data Sources and Preparation

We collected the data from Kaggle (Link for reference: YouTube-New Dataset). The dataset was collected using YouTube API.

Steps for Data Preparation:

1. Appended data for all regions and made required changes on datatypes for required column.
2. Removed Null values from the data.
3. Basic text pre-processing on text columns like description, tags, channel, title etc.
4. Additional Metric Calculations like Engagement Ratio, Days of the week etc.
5. Performed required scaling and other pre-processing techniques to calculate sentiment of title, description and tags combined for sentiment and world cloud Analysis.
6. Channel Name Analysis using text mining techniques like TF-DIF
7. Finalized the data and exported in .csv format.

Dashboard interactions and findings

The dashboards created provide a comprehensive analysis of YouTube data, allowing users to interact with various visualizations to uncover meaningful insights.

The **Overview Dashboard** provides a high-level summary of YouTube trends and performance metrics. It is designed to help viewers quickly analyze engagement patterns, regional performance, and content distribution. The dashboard includes a combination of visualizations such as a bar chart showcasing **Engagement Ratios by Region**, a line chart tracking **Total Views** over time, and a pie chart illustrating the **Category Distribution** of videos. Additionally, a dual-line chart for **Channel Performance Metrics** highlights the positive and negative engagement rates across various categories. The layout is interactive, allowing users to filter data by categories or regions, making it a comprehensive tool for understanding trends and audience behavior. This dashboard serves as a one-stop solution for gaining insights into the key drivers of video engagement and content success.

The **Geographic Insights** dashboard enables users to explore the distribution of top-performing content categories across regions. The treemap highlights how "Music" consistently dominates in the United States and Great Britain, while India showcases a broader mix of interests. Regional sentiment trends over time further reveal how audience sentiment varies across geographies, with stable trends in Canada contrasted by fluctuations in India and the United States. Seasonal trends in engagement are also prominently featured, helping users identify months like December as high-engagement periods for specific categories, such as leisure content.

The **Content Analysis** dashboard delves deeper into the factors influencing content performance. Users like Alex can analyze the impact of channel names, comparing generic and unique channels to determine how naming conventions affect the like-to-view ratio. The title formatting analysis offers insights into how specific elements, such as exclamation marks or capitalization, influence viewership. A word cloud visualization provides an engaging representation of popular tags, such as "new", "funny", and "latest", giving actionable guidance for optimizing content discoverability.

Moving to the **Engagement Analysis** dashboard, users can uncover trends in engagement metrics, such as views and comments, across days of the week. Saturday emerges as the most effective day for content releases, particularly for music, which demonstrates consistent high performance regardless of the day. The dashboard also showcases category-wise performance, where pop culture content dominates viewership, while comedy excels in generating comments, indicating its potential for fostering audience interaction.

The interactive features across these dashboards allow users to filter by region, category, and time period, creating a dynamic experience for uncovering tailored insights. For instance, filtering the sentiment trends by category or region highlights subtle but impactful differences in audience preferences. The dashboards collectively provide a powerful tool for decision-making, blending visuals with data-driven narratives to guide content strategies effectively.

Insights from Dashboard Interactions

Based on the dashboard findings, Alex should focus on the following trends to optimize content strategy and engagement:

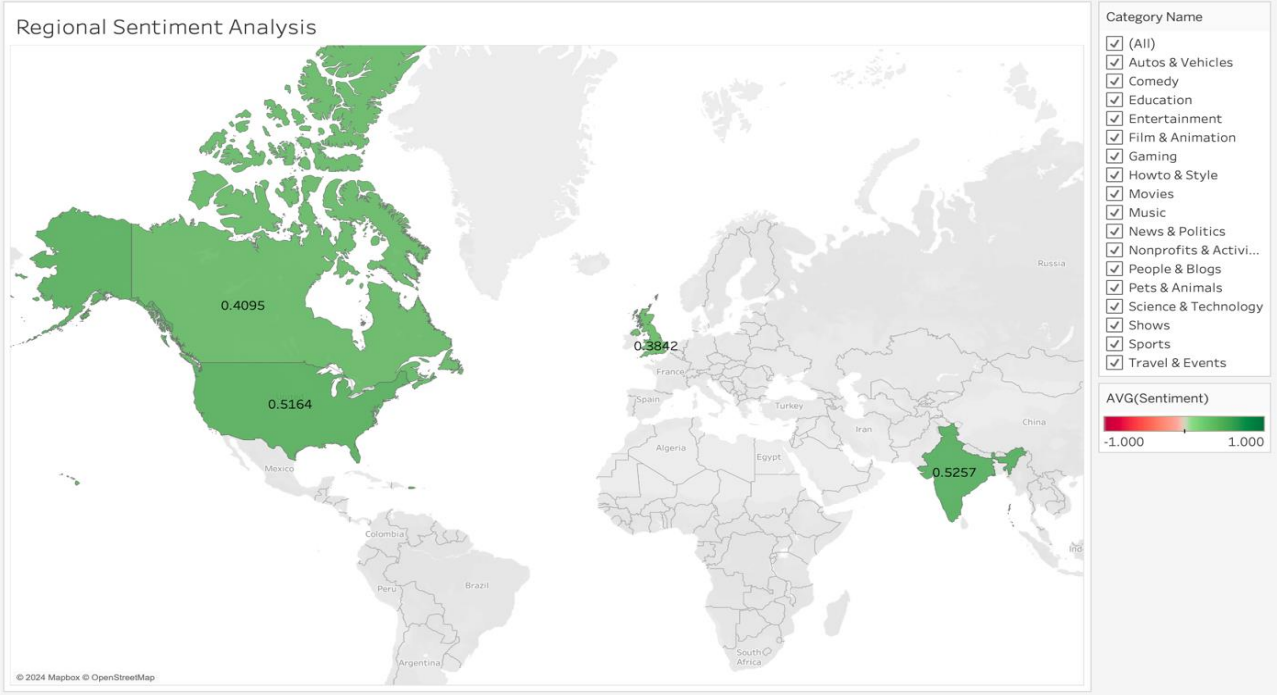
1. **Category-Specific Trends**:
    - **Pop Culture Hub Dominance**: Categories like music, movies, and sports drive the majority of views and engagement. Within these, **music** stands out as a consistent performer across all regions. Alex should prioritize these areas when targeting mass viewership and engagement.
    - **Comedy's Engagement Potential**: Although it may not dominate in views, comedy generates the highest number of comments, reflecting strong audience interaction. This makes it a valuable category for building community and fostering discussions.
    - **Leisure's Steady Performance**: Leisure-related content, including entertainment, how-to, and style, is the second most popular category in terms of views and engagement. Alex should continue to invest in this space for sustained growth.
2. **Seasonal Trends**:
    - **High Engagement Periods**: December is a peak engagement month for leisure content, while May stands out as a high-performance month for pop culture categories. Alex can align content launches with these periods to maximize reach and impact.
    - **Steady Trends for Travel Content**: While other categories show fluctuations, travel content exhibits rising sentiment over time, indicating growing interest. This could be a niche area to explore further.
3. **Regional Preferences**:
    - The **United States** and **Great Britain** are major drivers of engagement for pop culture categories, particularly music. Alex should consider region-specific campaigns to amplify content in these areas.
    - **India's Fluctuating Sentiment**: India shows more volatility in sentiment for categories like shows and travel. Alex can explore tailored strategies to stabilize and boost engagement in this region.
4. **Audience Behavior**:
    - **Saturday as the Best Day**: Across categories, Saturday records the highest views, making it the optimal day for content releases. This is particularly true for music, which performs well regardless of the day.
    - **Channel Naming and Title Formatting**: Content from unique channels shows slightly less performance than generic ones, but title formatting plays a critical role. Shorter titles or those using punctuation effectively (e.g., exclamation marks) tend to perform better.
5. **Tag Usage and Content Discovery**:
    - Tags like "comedy," "music," and "trailer" are highly popular and drive discoverability. Alex should ensure relevant tags are strategically used to maximize visibility.

6. **Geographic Sentiment Insights**:
    - Regional sentiment analysis suggests a stable trend in Canada but highlights spikes and drops in sentiment in countries like India and the U.S., particularly for niche categories like shows and travel. Understanding these trends can help tailor content to better resonate with audiences in specific locations.

By focusing on these trends, Alex can fine-tune their content strategy to capitalize on high-performing categories, optimize publishing schedules, and create regionally tailored campaigns that align with audience behavior and preferences.

Changes from Original Design Document

1. **Substitution of the Parallel Plot**: Due to Tableau's lack of support for parallel plots, a dual-axis chart was implemented as a practical alternative. This adaptation preserves the ability to compare key metrics effectively while ensuring the insights remain clear and accessible.
2. **Category Grouping for Clarity**: Categories were consolidated into broader groups to minimize visual clutter in stacked bar charts. This grouping not only enhances readability but also facilitates the identification of high-performing categories, such as "Pop Culture Hub" and "Leisure," making the insights more digestible and actionable.
3. **Focus on Regional Sentiment Trends**: The choropleth map showing average video sentiment across regions was replaced with a "Regional Sentiment Trend Over Time" analysis. The map's uniform positive sentiment across regions offered limited insights. In contrast, the trend analysis provides a dynamic view of how sentiments evolve over time in different regions, offering Alex more specific guidance on categories with increasing engagement.



4. **Language Tree Map Elimination**: The planned language distribution visualization was removed due to the challenges of analyzing and integrating non-English data into the broader analysis. Focusing on English-speaking regions simplified the process and aligned with Alex's target audience. By removing this redundant visualization, the analysis remains streamlined and focused on actionable insights.