Portfolio Project Social Media Engagement Summary

1. Acquiring and Reviewing Data

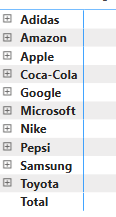
Data Set found at <https://www.kaggle.com/datasets/subashmaster0411/social-media-engagement-dataset>

* 1. This machine-generated dataset simulates social media engagement data across various metrics, including likes, shares, comments, impressions, sentiment scores, toxicity, and engagement growth. It is designed for analysis and visualization of trends, buzz frequency, public sentiment, and user behavior on digital platforms.
  2. While this dataset is designed for this purpose, there is still data transformation that needs to occur prior to visualization.

1. Data Transformation (Planning)
   1. The field language is populated in a shorthand that won’t mean anything to an end user, this needs to be mapped into plain language.
   2. Timestamp has day and time in one field, we may miss trends by looking at information this way, so I’ll separate time and date into separate fields to see if there is value to trending the two fields separately, it’s possible to do this without splitting the field in Power BI but I prefer to do it this way.
   3. Similarly, location has city cand country in the same field, this may make mapping visualization more specific than is helpful, so I’ll separate that out as well.
   4. the columns for hashtags, the column for keywords and the column for mentions allow for multiples in a single column. I’ll separate these out into lists and count the occurrence of each and sort them from most recurring to least.
2. Data Transformation (Execution)
   1. While I could manually clean and align this data if this were for a client, I’d want to make this as user friendly and easily repeatable as possible, so I’ll write a script to conduct the transformations detailed above.
3. The detailed script for how I transformed this is available at <https://github.com/Holley-Michelle/SlytherScript/blob/main/Social%20Media%20Engagement%20Script(Update).ipynb>
4. Load Data to Power BI
   1. Script results in 2 types of files
      1. Original data frame with the new columns
      2. Summary count files for Keyword, Mentions, and Hashtags

Dashboard Design

1. When designing a dashboard, it’s important to establish the question of intended audience, or “who is this for?” for the purposes of this project I’m going to assume the intended audience is a management company with all the brands listed as clients.
2. Looking at the data in a summary view (see below), it appears our analysis would benefit by being grouped by industry, which is not field in the original data file, so I’ll create it and then go and amend the script to create it in the future.



Shoe/Sport Brands

* Adidas
* Nike

Soda Brands

* Pepsi
* Coca-Cola

Tech Brands

* Amazon *(grouping my product this data set only includes tech items which is why it’s been placed in this category)*
* Apple
* Microsoft
* Samsung
* Google

Car Brands

* Toyota

Now that I’ve done that I’ve built several high-level visuals that will give us some insight

1. Average Engagement by Industry and Brand
2. Total Impressions by Industry and Campaign (drill down to see Brand, Campaign, and Campaign Phase)
3. Average Sentiment Score by Industry Over Time (hover to view Avg Sentiment Score in 2024 vs 2025 by Industry the Avg Change and the ranking across industries given)

Admittedly we may want to see this visual by brand not industry but having all shown at the same time introduces noise to the visual, instead I’ve built filters that will allow users to filter all the visuals by brand and/or by industry).

\*Please note this dashboard is still in development this is just the initial pass of the data