

Project Description

Fetch Engagement Data and Analyze Social Media Post Performance

Team Name: Jinxed Coders

Objective:

This project demonstrates the integration of modern tools like DataStax Astra DB, Langflow, and GPT to analyze and gain insights from social media engagement data. The primary goal is to create a small-scale workflow to simulate, analyze, and derive actionable insights on social media Post performance.

Tasks Performed:

1. Fetch Engagement Data -

Data Simulation:

Simulate and store social media engagement data. We have used Facebook engagement data for this project. The dataset contains the following attributes:

- Page_T_Likes - The total number of likes on the Facebook page.
- Type - Type of post made.
- Category - The classification of the post's content.
- Month - The month when the post was made.
- Weekday - The day of the week the post was published.
- Hour - The hour the post was published.
- Paid - Indicates whether the post was promoted/boosted (1 for paid, 0 for organic).
- T_Reach - The total number of unique users who saw the post.
- T_Impression - The total number of times the post was displayed on users' screens.
- Engaged_Users - The number of unique users who interacted with the post.
- Consumers - Users who clicked on the post, such as viewing photos or playing videos.
- Consumption - The total number of clicks or actions taken on the post.
- LP_Impression - The cumulative number of impressions the post received over its lifetime.
- LP_Reach - The cumulative number of unique users reached by the post over its lifetime.
- LP_Engage_With_Post - The total number of interactions (likes, shares, comments) on the post over its lifetime.
- comment - The total number of comments received on the post.
- like - The total number of likes received on the post.
- share - The total number of times the post was shared by users.
- T_Interactions - The sum of all interactions on the post, including likes, shares, and comments.

```

Page_T_Likes,Type,Category,Month,Weekday,Hour,Paid,T_Reach,T_Impression,Engaged_Users,Consumers,Consumption,LP_Impression,LP_Reach,LP_Engage_With_Post,comment,like,share,T_Interactions
139441,Photo,2,12,4,3,0,2752,5091,178,109,159,3078,1640,119,4,79,17,100
139441,Status,2,12,3,10,0,10460,19057,1457,1361,1674,11710,6112,1108,5,130,29,164
139441,Photo,3,12,3,3,0,2413,4373,177,113,154,2812,1503,132,0,66,14,80
139441,Photo,2,12,2,10,1,50128,87991,2211,790,1119,61027,32048,1386,58,1572,147,1777
139441,Photo,2,12,2,3,0,7244,13594,671,410,580,6228,3200,396,19,325,49,393
139441,Status,2,12,1,9,0,10472,20849,1191,1073,1389,16034,7852,1016,1,152,33,186
139441,Photo,3,12,1,3,1,11692,19479,481,265,364,15432,9328,379,3,249,27,279
139441,Photo,3,12,7,9,1,13720,24137,537,232,305,19728,11056,422,0,325,14,339
139441,Status,2,12,7,3,0,11844,22538,1530,1407,1692,15220,7912,1250,0,161,31,192
139441,Photo,3,12,6,10,0,4694,8668,280,183,250,4309,2324,199,3,113,26,142
139441,Status,2,12,5,10,0,21744,42334,4258,4100,4540,37849,18952,3798,0,233,19,252
139441,Photo,3,12,5,10,0,3113,5500,300,137,145,3007,3174,155,0,88,10,106

```

Data Storage in DataStax Astra DB:

The dataset is stored in a cloud-native NoSQL database, leveraging Astra DB's scalability and fast query capabilities. This dataset originally in csv was converted to JSON for further use. The data was then converted to embeddings using NV-Embed Model and stored in vector database.

2. Analyze Post Performance - Use Langflow to query Astra DB and give insights for social media performance.

- **Langflow Workflow:** There 2 components in our Langflow workflow -
 - Retriever Flow - This part takes input from the user, searches for the query in the dataset stored in Astra DB, parses the retrieved data, structures it to fit the prompt, and then utilizes the Groq API to generate the output.
 - Load Data Flow - This part is about RAG. Loads data from a local file and embed it into the vector database by splitting the text into chunk size of 512 and chunk overlap of 50.

3. Provide Insights - Leverage GPT integration in Langflow to generate actionable insights based on analyzed data.

- **GPT Integration in Langflow:**
 - GPT is used to translate engagement metrics into meaningful insights.
 - Example Prompts for GPT:
 - “Give a comparison of video, photo and status Post performance without considering category”
- **Example Output Insights:** Based on the analysis, the performance of video, photo, and status posts without considering category is as follows:
 - a. Video posts have the highest engaged users but not the highest total interactions.
 - b. Status posts have the highest likes, shares, comments, and total interactions.
 - c. Photo posts have the lowest performance in all categories.

Workflow Summary

1. Data Simulation and Storage:

Simulated social media data is stored in Astra DB using its efficient NoSQL storage capabilities.

2. Analysis with Langflow:

Langflow is used to create an intuitive workflow for querying Astra DB and processing data.

3. **Insights Generation:**

GPT in Langflow generates insights, making the data actionable and easy to interpret.

This project showcases the potential of combining cutting-edge tools to streamline data analysis and deliver insights, making it a practical solution for social media performance monitoring.