Facebook Marketplace Dataset

The Facebook Live Sellers in Thailand dataset contains information about the Facebook pages of 10 Thai fashion and cosmetics retail sellers. Below is a description of the dataset:

1. Title: Facebook Live Sellers in Thailand Dataset

2. Source: The dataset is sourced from the UCI Machine Learning Repository.

3. Data Type: The dataset is in a tabular format, typically stored in a CSV (Comma Separated Values) file.

4. Number of Instances: There are a total of 7050 instances (rows) in the dataset.

5. Number of Attributes: The dataset initially consists of 16 attributes (columns). After removing redundant columns, there are 14 attributes remaining.

6. Attribute Information:

- status_id: Unique identifier for each status post.
- status_published: Date and time when the status post was published.
- status_type: Nature of the status post (e.g., video, photo, status, link).
- num_reactions: Number of reactions (e.g., likes, loves, wow, haha, sad, angry) received on the status post.
- num_comments: Number of comments received on the status post.
- num_shares: Number of shares received on the status post.
- Additional numerical and categorical attributes related to engagement metrics and status post features.

7. Missing Values: The dataset may contain missing values, which need to be handled during data preprocessing.

Questions

1. Is there a correlation between the number of reactions (num_reactions) and other engagement metrics such as comments (num_comments) and shares (num_shares)? If so, what is the strength and direction of this correlation?
2. Is there a correlation between the number of reactions (num_reactions) and other engagement metrics such as comments (num_comments) and shares (num_shares)? If so, what is the strength and direction of this correlation?
3. Use the columns status_type, num_reactions, num_comments, num_shares, num_likes, num_loves, num_wows, num_hahas, num_sads, and num_angrys to train a K-Means clustering model on the Facebook Live Sellers dataset.
4. Use the elbow method to find the optimum number of clusters.
5. What is the count of different types of posts in the dataset?
6. What is the average value of num_reaction, num_comments, num_shares for each post type?