Insights to develop a content strategy for Infosys based on past marketing data

EDA / Descriptive Statistics Introduction:

Exploratory Data Analysis (EDA) and descriptive statistics are two powerful tools that data analysts use to understand a dataset before diving deeper into statistical modelling or machine learning. They work together to summarize the data, identify patterns and trends, and uncover potential issues. Descriptive statistics provide a numerical summary of the data. It helps get a basic understanding of the data set by calculating measures of central tendency (like mean, median) and spread (like standard deviation, variance).

Overall design strategy:

1. Read the CSV file using the panda's library.
2. Display the first few lines of the CSV file to understand its structure and content.
3. Perform Exploratory Data Analysis (EDA) to uncover trends and correlations between creative attributes and performance metrics.
4. Identify which visual and text elements are most influential in driving engagement.
5. Visualize the results using appropriate plots and charts.

Steps:

1. Load the CSV file and display the first few lines.
2. Perform data cleaning and preprocessing if necessary.
3. Analyze the distribution of performance metrics (views, likes, comments, shares, total engagements, engagement rate).
4. Investigate the relationship between creative attributes (e.g., company name, post topic, logos, number of persons) and performance metrics.
5. Visualize the findings using plots and charts.

Data Overview:

The CSV file contains various columns related to Twitter posts, such as:

date: The date of the post.

views: The number of views the post received.

likes: The number of likes the post received.

comments: The number of comments on the post.

shares: The number of shares the post received.

total engagements: The total number of engagements (likes, comments, shares).

engagement rate %: The engagement rate percentage.

company name: The name of the company that made the post.

post topic: The topic of the post.

post url: The URL of the post.

Additional columns related to logos and persons in the post

Read and Display the CSV File: The first few lines of the CSV file were displayed, showing the structure and content of the data.

Data Cleaning and Preprocessing: The date column was converted to datetime, and missing values were filled with 0.

Summary Statistics for Performance Metrics: Summary statistics (count, mean, std, min, 25%, 50%, 75%, max) for the performance metrics (views, likes, comments, shares, total engagements, engagement rate) were displayed.

Visualizations:

Distribution of Performance Metrics: Histograms with KDE plots were created to visualize the distribution of views, likes, comments, shares, total engagements, and engagement rate.

Relationship Between Company Name and Performance Metrics: Bar plots were created to show the relationship between company name and performance metrics.

Relationship Between Post Topic and Performance Metrics: Bar plots were created to show the relationship between post topic and performance metrics.

Influence of Logos on Performance Metrics: Bar plots were created to show the influence of the presence of logos on performance metrics.

Influence of Number of Persons on Performance Metrics: Bar plots were created to show the influence of the number of persons on performance metrics.

Interpretation of Results:

Summary Statistics: The summary statistics provide an overview of the distribution of performance metrics. For example, the average number of views is approximately 2868, with a standard deviation of 4945, indicating high variability in the number of views.

Distribution Plots: The histograms with KDE plots help visualize the distribution of each performance metric, showing how the data is spread out and identifying any skewness or outliers.

Bar Plots: The bar plots help identify trends and correlations between creative attributes (company name, post topic, presence of logos, number of persons) and performance metrics. For example, you can see which companies or post topics tend to have higher engagement rates.

Strategy Recommendations:

Use Faces in Images: Include images with faces to increase relatability and engagement.

Optimal Text Length: Use longer, engaging text for posts to provide more information and engage readers.

Color Schemes: Use dominant colors that have shown higher engagement in previous posts.

Analyze Competitor Data: Benchmark findings with competitor data to understand industry trends and adapt strategies accordingly.