

# Exploratory Data Analysis Report

## Data Overview

- **Total Records:** 10,526
- **Columns:** 10
- **Key Columns:** platform, created\_time, text\_original, likes\_count, shares\_count, comments\_count, views\_count.
- The dataset contains engagement metrics (likes, comments, shares, views) with numeric values and timestamps for posts.
- Some columns are not required for further analysis, such as :account\_id,id,text\_additional

## Data Cleaning Process

- Removed columns: account\_id,id,text\_additional
- Converted created\_time into a datetime format for time-based analysis.
- Removed all non-ASCII characters from the 'platform' column using a regular expression. Filtered the rows, leaving only those where the value in the 'platform' column belongs to the list of known platforms: TikTok, YouTube, Facebook or Instagram.
- Create a new column 'hashtags' in the dataframe, where for each row in the column 'text\_original', using a regular expression, there are all hashtags (words starting with the symbol '#').

## Key Insights

### Likes:

- Average: 1414
- Maximum: 188611
- Few posts have extremely high engagement.

```
Likes_count:
Mean: 1414.38
Maximum: 188611
Minimum: 0
Median: 354
Standard deviation: 5989.21
```

### Comments:

- Average: 299
- Maximum: 80,415
- Similar skewed distribution as likes.

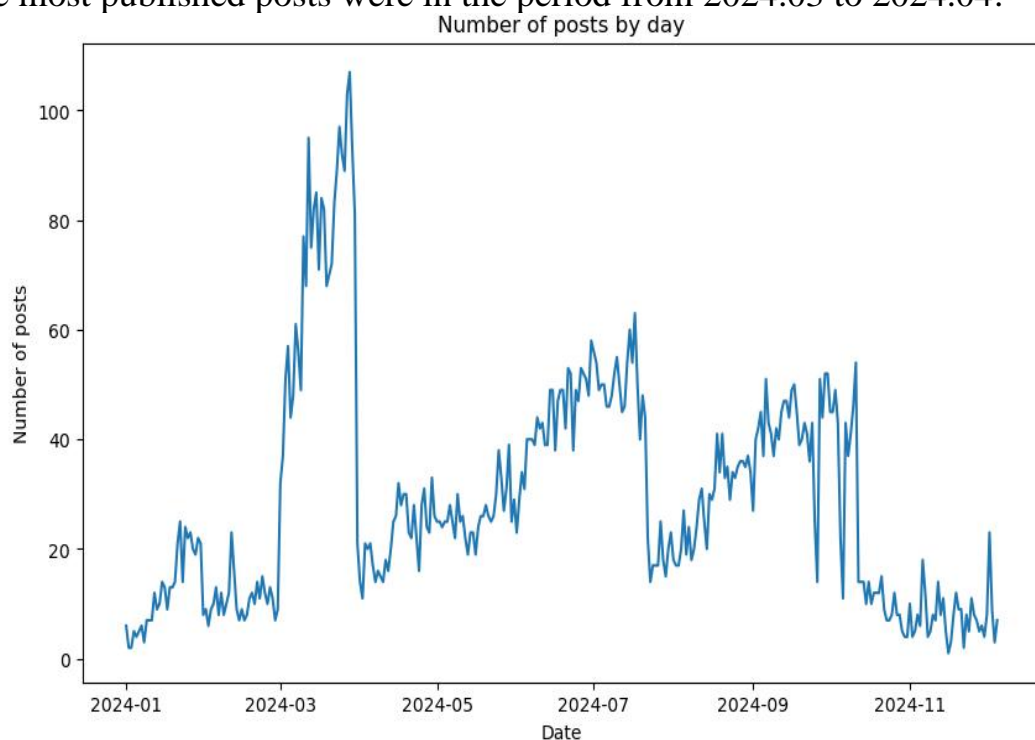
```
Comments_count:
Mean: 299.86
Maximum: 80415
Minimum: 0
Median: 93
Standard deviation: 1259.44
```

## Views:

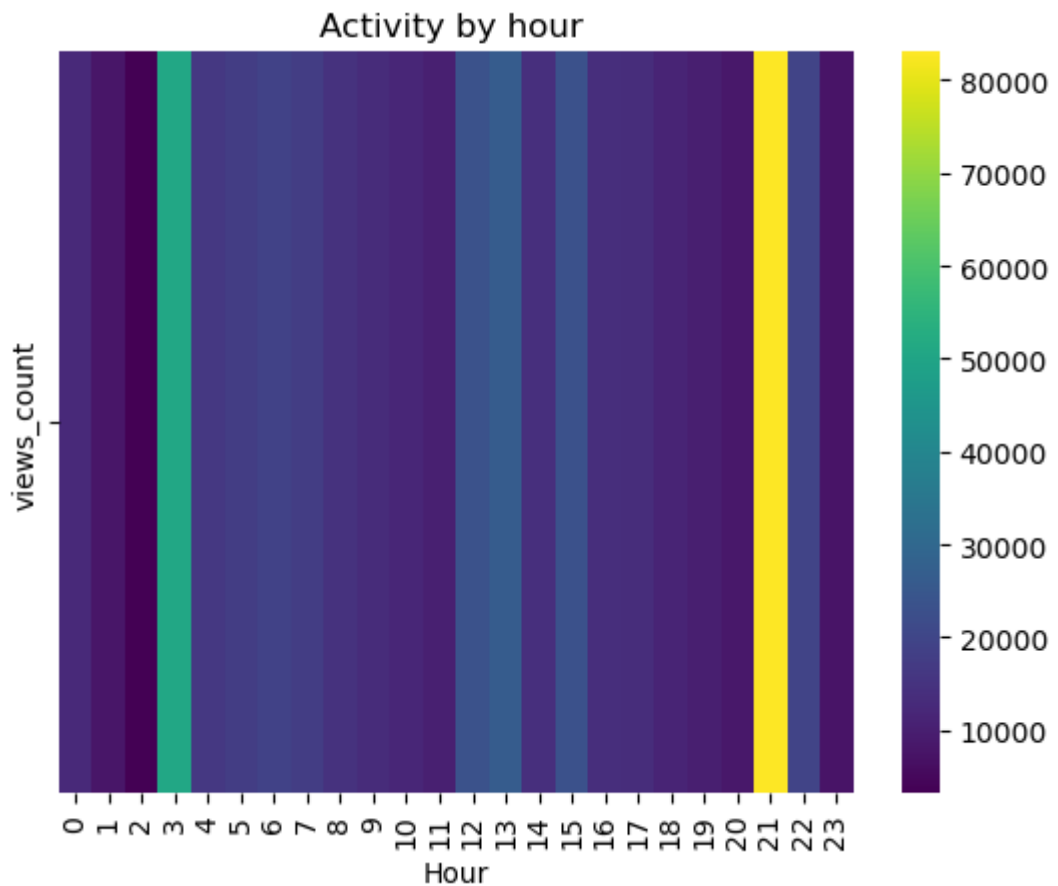
- Average: 17,973
- Maximum: 3,500,000
- Large variance suggests some posts have much broader reach.

```
Views_count:
Mean: 17973.53
Maximum: 3500000
Minimum: 3
Median: 3704
Standard deviation: 101249.23
```

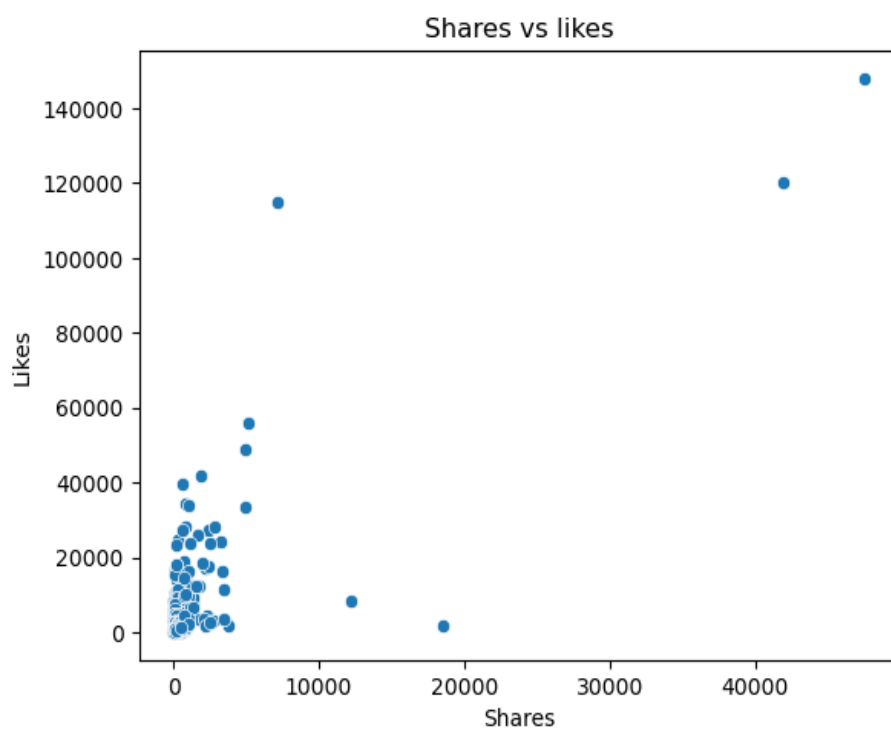
- In the morning there is more activity due to the hours.
- The most frequent hashtags and text features relate to religious.
- Posts span across 2024, showing potential for seasonal trends in engagement.
- The most published posts were in the period from 2024.03 to 2024.04.



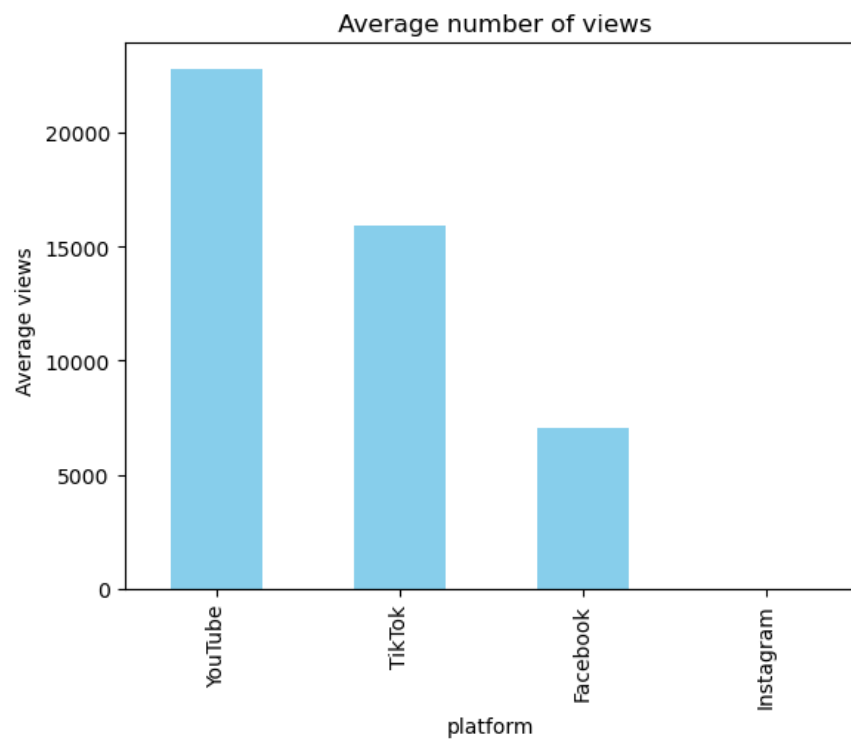
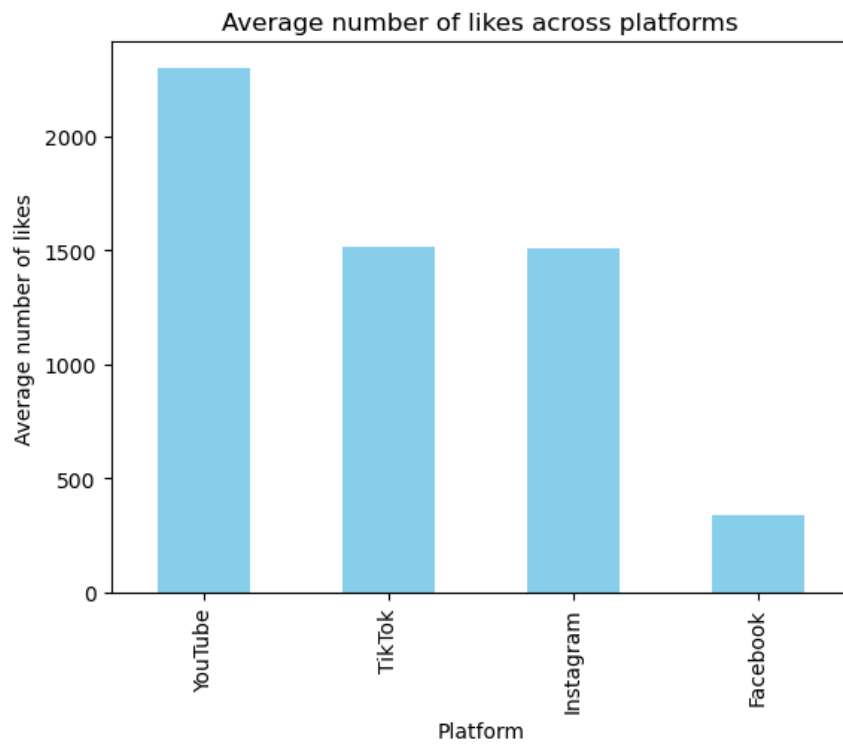
- The highest viewing activity is observed at 21:00 but also noticeably high activity at 3:00.



- Scatter Plot of shares\_count vs likes\_count: Indicates a positive correlation between shares and likes.



- In terms of the average number of likes and views, YouTube and TikTok are leading.



- But Facebook is leading in terms of comments.

