coursera
project
network

Sentiment Analysis in **Mobile Photography**

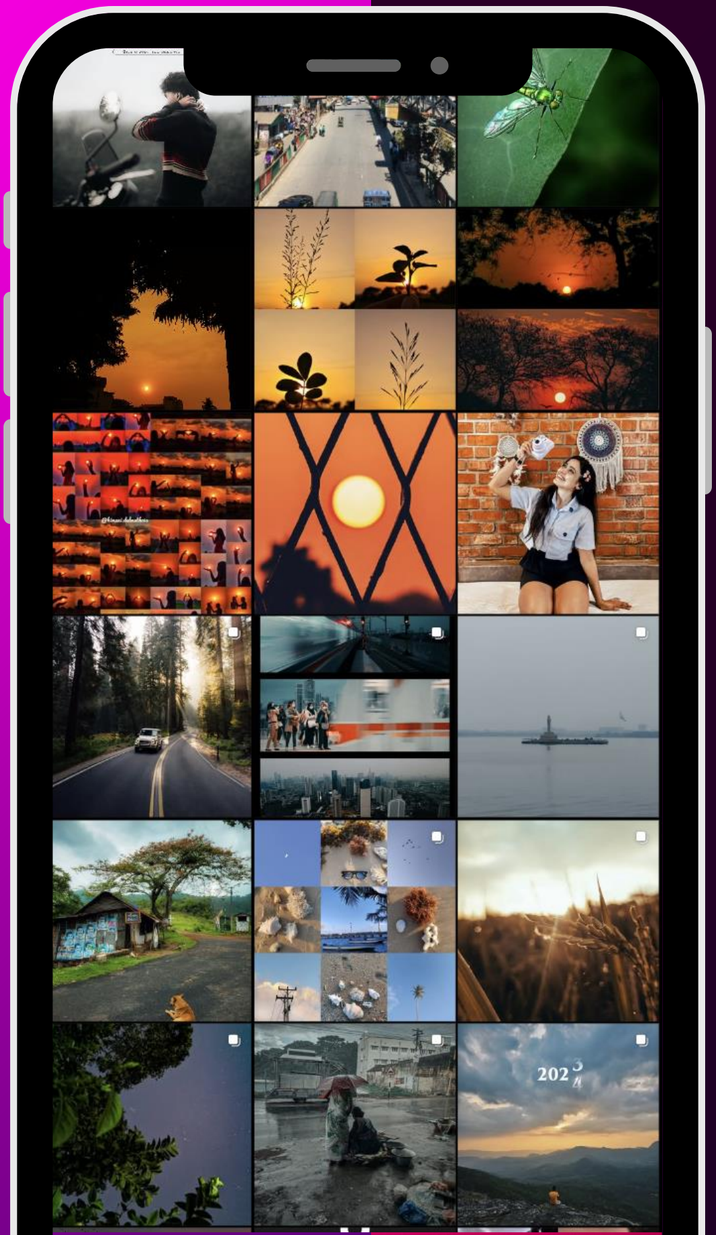
Disclaimer:

This research was conducted from December 23, 2024, to December 26, 2024, using a personal ‘Twitter / X’ account limited to only 100 requests. During this period, a total of 1,138 rows of data were collected, which was reduced to 1,113 rows after data cleaning.

Due to the limited dataset, the results of this research should not be considered professional references. However, it provides insights into engagement patterns on social media, specifically on ‘Twitter/X’, focusing on #mobilephotography content and user behavior.

Special thanks to Coursera for presenting this project challenge.

Warm regards,
Indonesia, December 28, 2024
Rendra



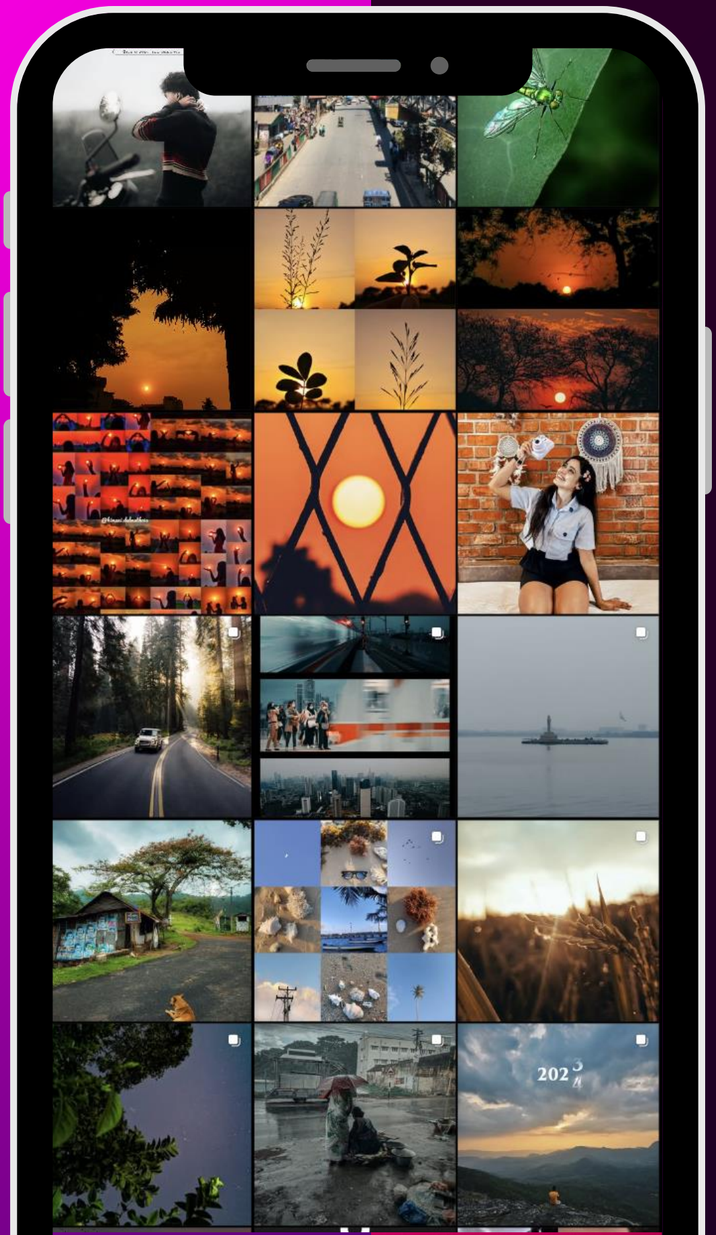
Background

Project Scenario

A social media marketing team specializing in promoting brands and products on a popular social media platform wanted to understand how smartphone users utilize their devices for photography. They aim to uncover insights such as what types of photos users like to post and the times they most frequently share content on twitter/X.

Using Python, clean, analyze, and visualize the data to provide the marketing team with data-driven recommendations for their clients. These insights will help improve social media marketing by aligning with customers' preferences for mobile photography styles and developing content to enhance engagement.

This project is a Coursera Self-Paced Project on :
https://www.coursera.org/learn/analyze-social-media-python/supplement/uKwPr/the-project-scenario



Background

References :

1. https://towardsdatascience.com/how-to-access-data-from-the-twitter-api-using-tweepy-python-e2d9e4d54978
2. https://developer.x.com/en/docs/x-api/tweets/search/api-reference/get-tweets-search-recent
3. Sentiment Analysis of Tweets using Unsupervised Learning Techniques and the K-Means Algorithm

Methods

1. Using the **Tweepy library** and **X developer tokens** (free subscriptions) to create dataset.
2. Collect Twitter/X data with the hashtags: **#mobilephotography**, **#shotoniphone**, and **#shotonsnapdragon**.
3. Gather data from **Monday, 23 December 2024 – Thursday, 26 December 2024**.
4. Clean the data by removing duplications, handling missing values, and dropping features with **NaN** values.
5. Duplications found in the **'tweet_id'** feature: after removing duplications, some are kept as users often upload content sequentially on social media.
6. Perform **Feature Engineering** to cluster data by day_name, hashtag, smartphone type, photo theme, and total engagement.
7. Perform Univariate, Bivariate & Multivariate Analysis using **Matplotlib** and **Seaborn libraries** to create visualizations.
8. Apply the **groupby()** function to identify patterns by time, photo theme, and total engagement.
9. Using **K-Means** & **PCA** to cluster the dataset

Exploratory Data Analysis

23 December 2024 – 26 December 2024

Features Type & List

Total Numerical Features	List of Numerical Features
11	'tweet_id', 'domain_id', 'entity_id', 'likes', 'retweets', 'replies', 'quotes', 'bookmark_count', 'impression_count', 'hour', 'total_engagement'

Total Object Features	List of Object Features
16	'tweet_text', 'created_at', 'domain_name', 'domain_description', 'entity_name', 'entity_description', 'datetime', 'date', 'day_name', 'hashtag', 'hashtag_first3', 'hashtag_last3', 'smartphone', 'photostyle_first', 'photostyle_last', 'phototheme'

Exploratory Data Analysis

23 December 2024 – 26 December 2024

```
summary_eda = {
    'Metric': [
        'total content',
        'total unique users',
        'total unique entity',
        'total likes',
        'total retweets',
        'total replies',
        'total quotes',
        'total bookmark_count',
        'total impression_count',
        'total_engagement',
        'total features'
    ],
    'Value': [
        df.shape[0],
        df.tweet_id.nunique(),
        df.entity_id.nunique(),
        df.likes.sum(),
        df.retweets.sum(),
        df.replies.sum(),
        df.quotes.sum(),
        df.bookmark_count.sum(),
        df.impression_count.sum(),
        df.total_engagement.sum(),
        df.columns.nunique()
    ]
}

df_eda = pd.DataFrame(summary_eda)

df_eda
```

	Metric	Value
0	total content	1,113.0
1	total unique users	68.0
2	total unique entity	164.0
3	total likes	5,967.0
4	total retweets	359.0
5	total replies	1,280.0
6	total quotes	15.0
7	total bookmark_count	140.0
8	total impression_count	83,717.0
9	total_engagement	91,478.0
10	total features	27.0

Numerical Features Descriptions

Total Content	Total Users (Unique)	Total Entity (Unique)
1,113	68 twitter accounts	164

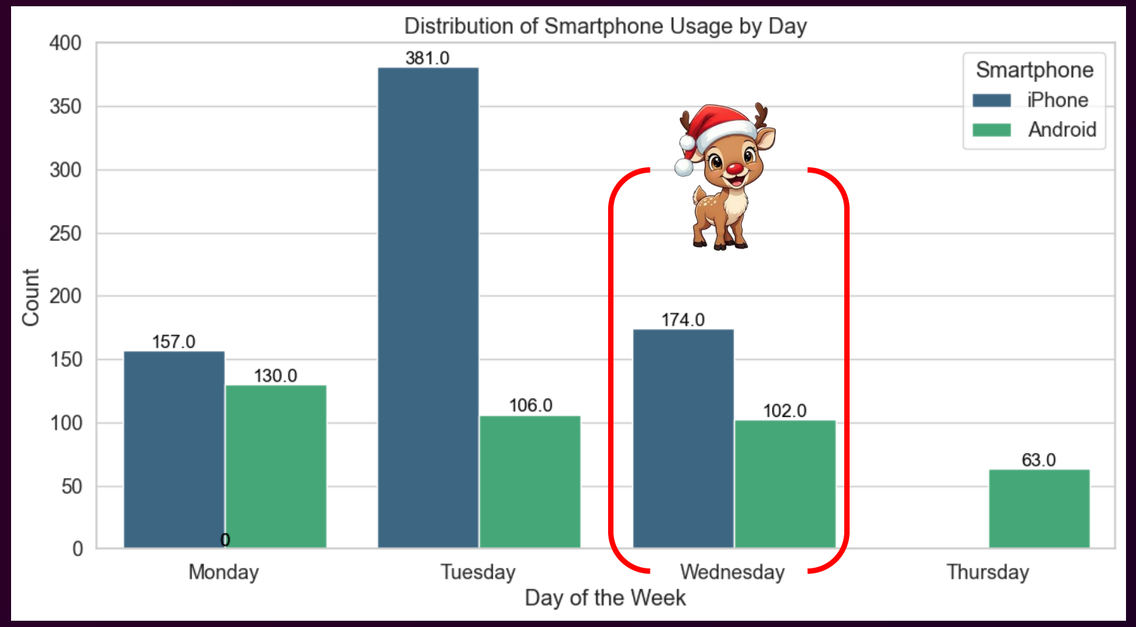
Total Likes	Total Retweets	Total Replies	Total Quotes
5,967	359	1,280	15

Total Bookmark	Total Impression	Total Engagement
140	83,717	91,478

Total Features
27

Insights :

Total Contents by Day of Phone Posting

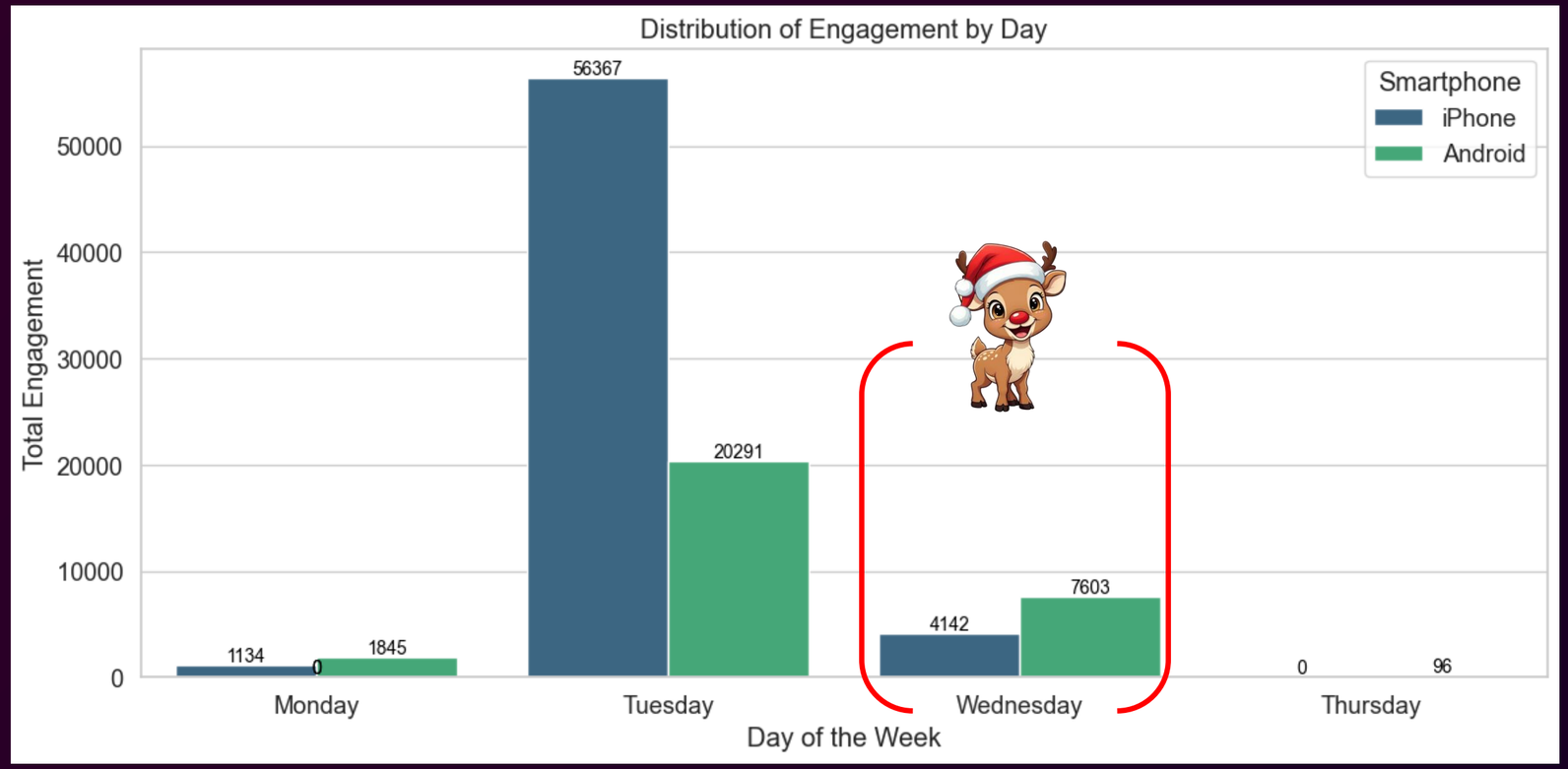


This plot shows the smartphone usage between iPhone and Android users during the research period from Monday (December 23, 2024) to Thursday (December 26, 2024). It can be observed that iPhone users tend to post more content on Twitter, except on Thursday.

There was a significant increase in content on Tuesday, which needs to be examined in terms of the type of content (photo) and the time when the spike occurred, as well as the engagement during this period.

Insights :

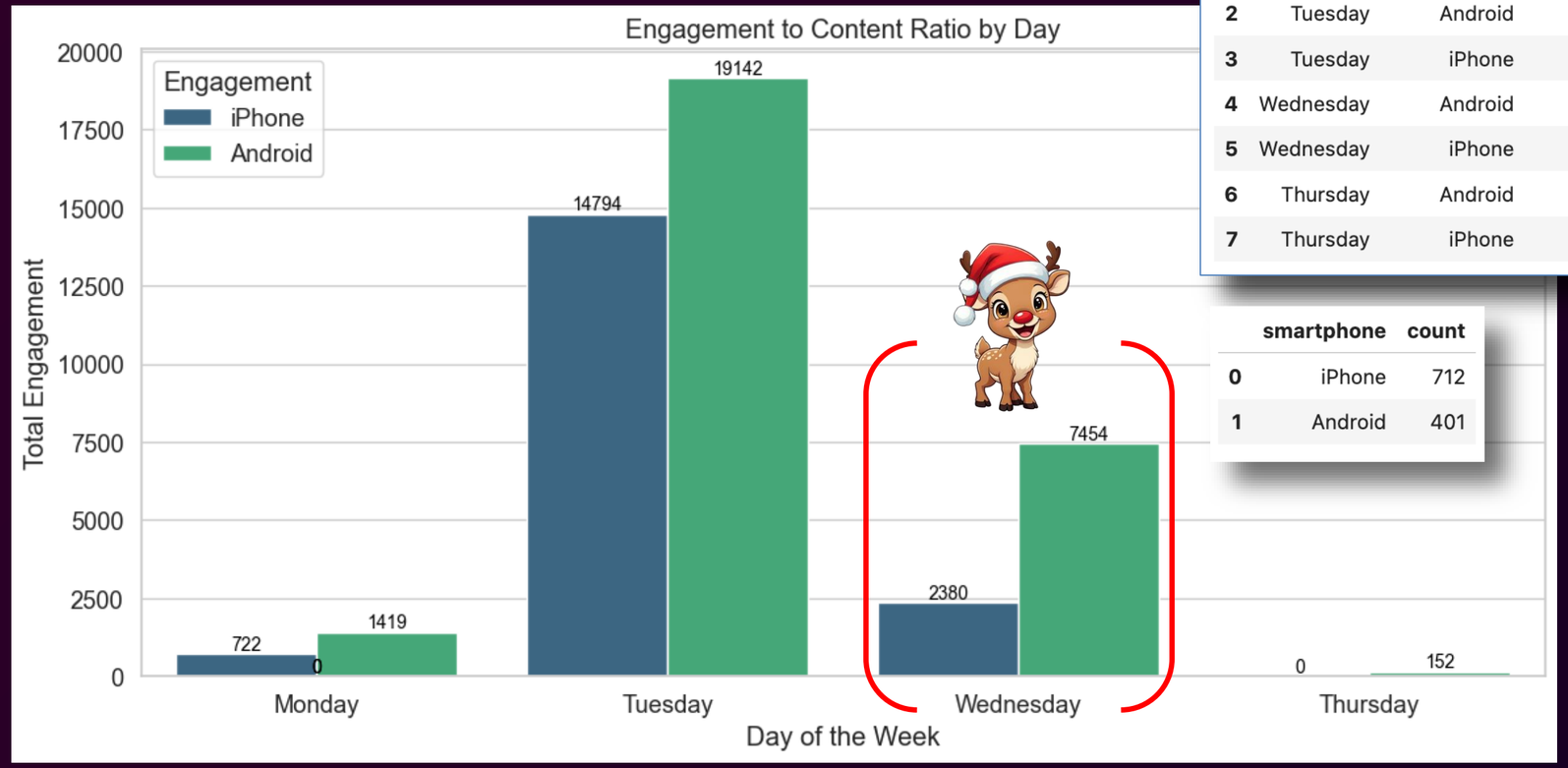
Total Engagement by Day of Phone Posting



Engagement for iphone users rise high on Tuesday compare to the other day, while Android users are more consitent creating engagement. Therefore, we will calculate the *content-to-engagement ratio* to examine the correlation between content effectiveness and engagement.

Insights :

Total Engagement Ratio by Day of Phone Posting



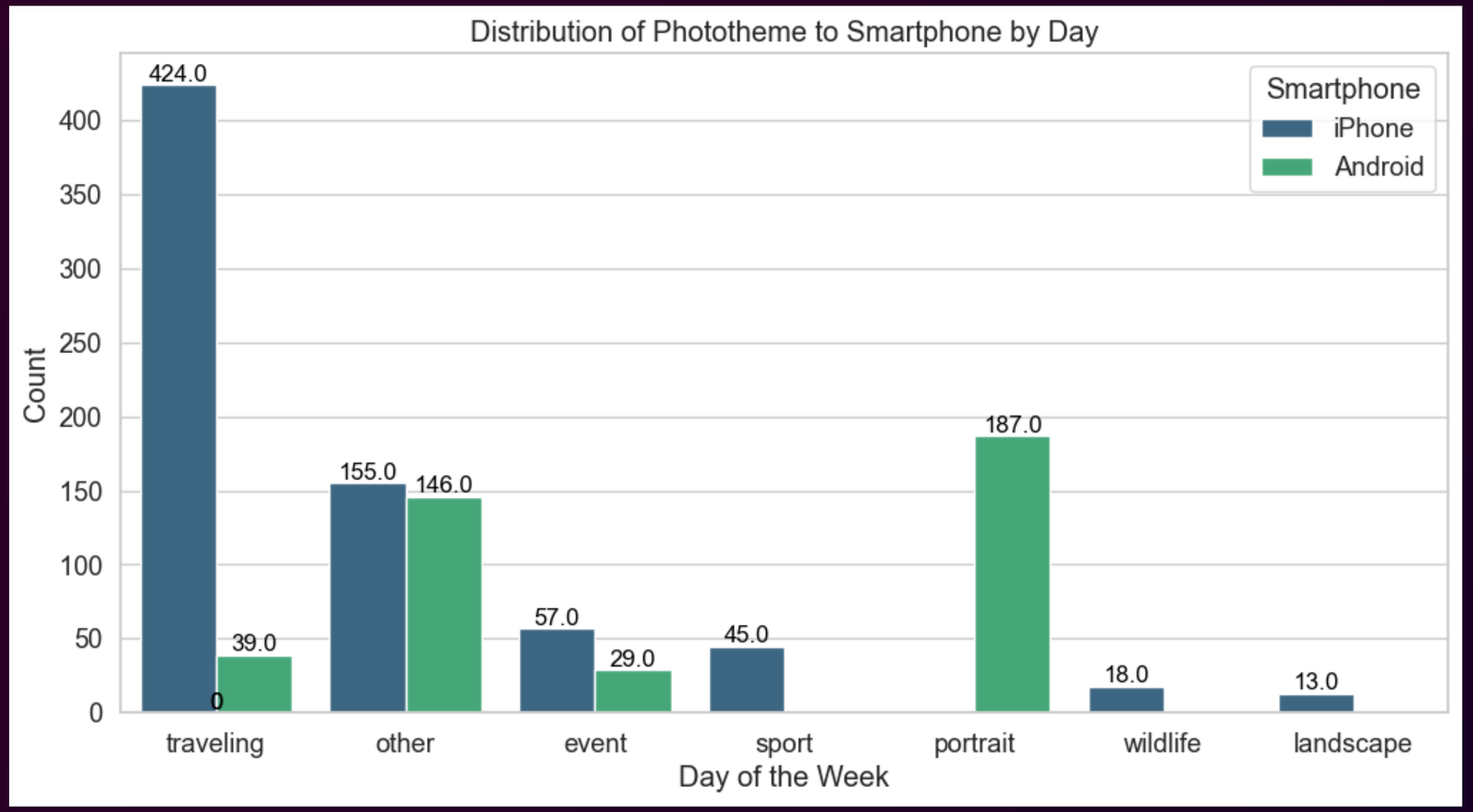
	day_name	smartphone	total_engagement	total_content	scale
0	Monday	Android	1,845.0	130	1,419.23
1	Monday	iPhone	1,134.0	157	722.29
2	Tuesday	Android	20,291.0	106	19,142.45
3	Tuesday	iPhone	56,367.0	381	14,794.49
4	Wednesday	Android	7,603.0	102	7,453.92
5	Wednesday	iPhone	4,142.0	174	2,380.46
6	Thursday	Android	96.0	63	152.38
7	Thursday	iPhone	0.0	0	0.0

	smartphone	count
0	iPhone	712
1	Android	401

In this plot, we can see that Android users are more consistent in generating engagement on social media. Although their number in the dataset is only 401 compared to 712 iPhone users, Android users are, in fact, more active in creating engagement and display greater consistency.

Insights :

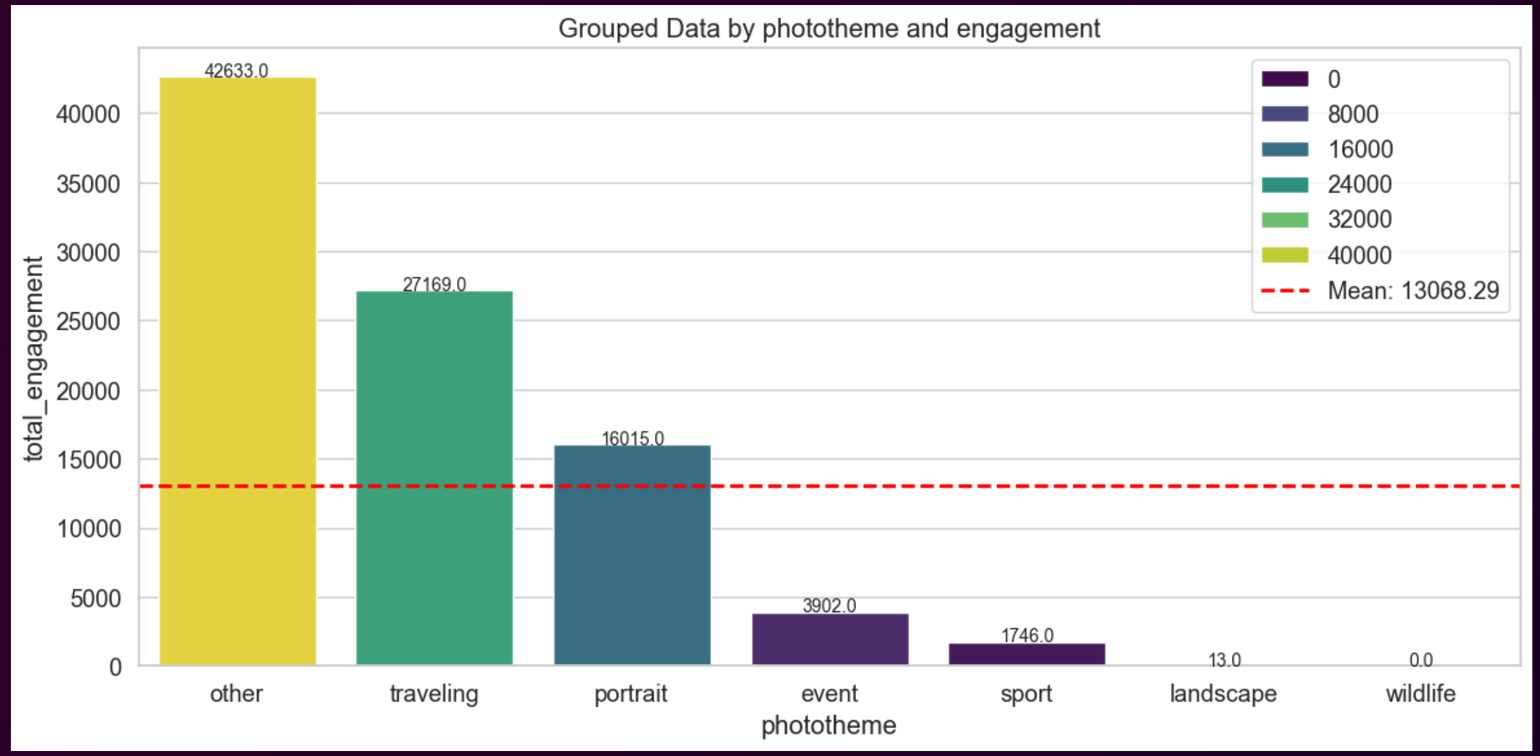
Analysis on Photo Theme on Smartphone Posting



iPhone users are spread across various types of mobile photography, with traveling being the most popular. Meanwhile, Android users seem to like casual photo or *prefer not to bother* with specific photography genres. We will now examine the engagement for each photo theme across both smartphone types.

Insights :

Analysis on Photo Theme by Engagement on Posting



phototheme	other	traveling	portrait	event	sport	landscape	wildlife
total_engagement	42,633.0	27,169.0	16,015.0	3,902.0	1,746.0	13.0	0.0

Before we delve further, we will examine the dominance of smartphone types on Twitter and the types of photos taken, both from iPhone and Android.

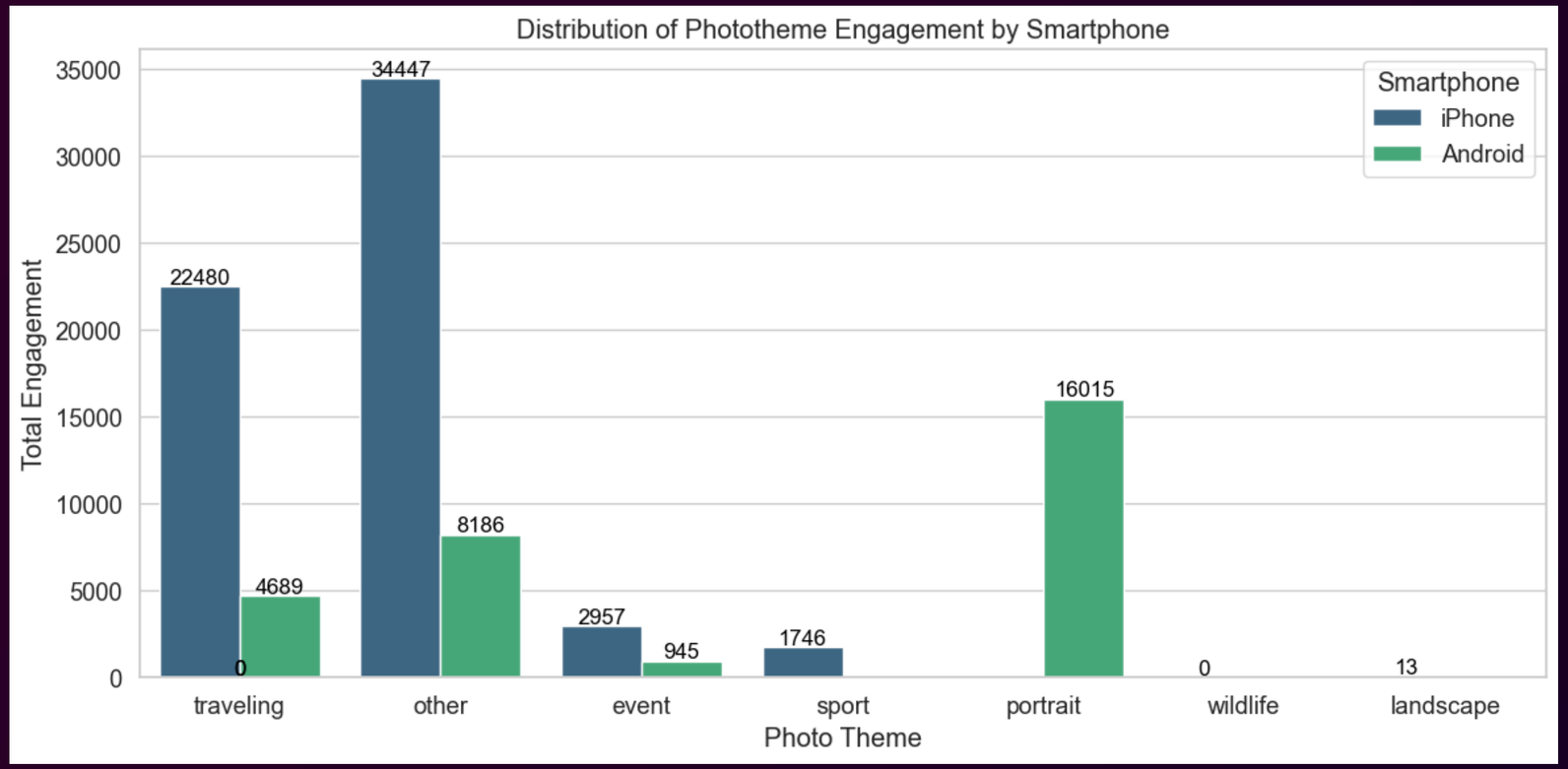
It can be seen that iPhone users are the most active on X in mobile photography, and supprisingly the most enggaged content type '**other**'.

'Other' is a post-category with various hashtag. Sometimes people just take selfie, cans of beer or wine or any photos and can't be categorized with any subjects photo.

The most probable reason is people just want to show picture without dazzling with specific topics of photo content.

Insights

Analysis on Photo Theme by Engagement on Smartphone Posting



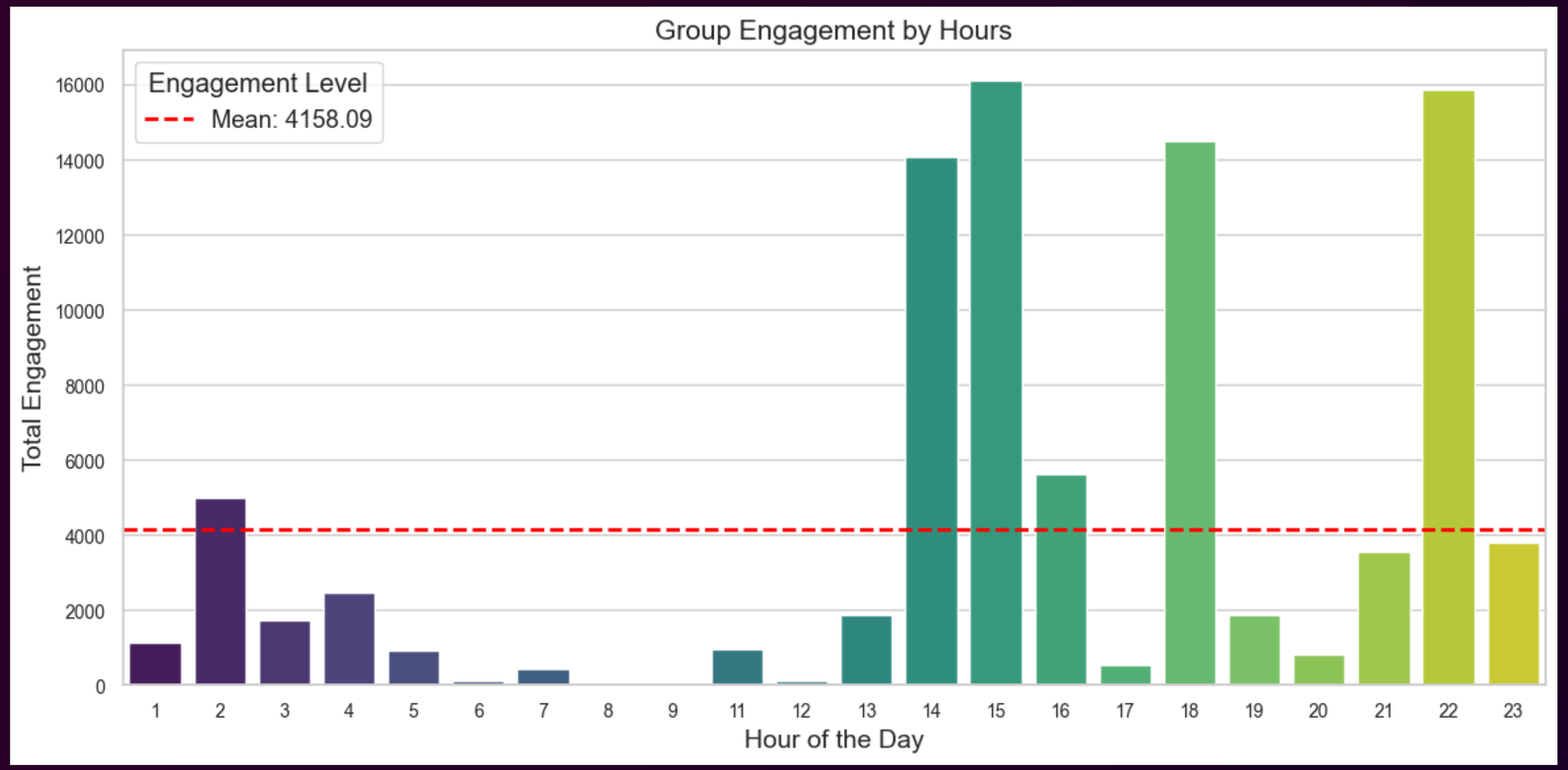
The photo content created by iPhone users generates significant engagement, especially in the traveling category. Even photos in the 'other' category create considerable engagement. The portrait photo type, exclusively created by Android users, generates the highest engagement at 16,015.

For ‘ther photo themes, iPhone content consistently generates at least three times more engagement compared to content from Android users.

Understandable most people doesnt want to dazzling with hashtag to specify their mobile photography content. They will use only #mobilephotography instead.

Insights

Analysis on Posting & Engagement by Hours

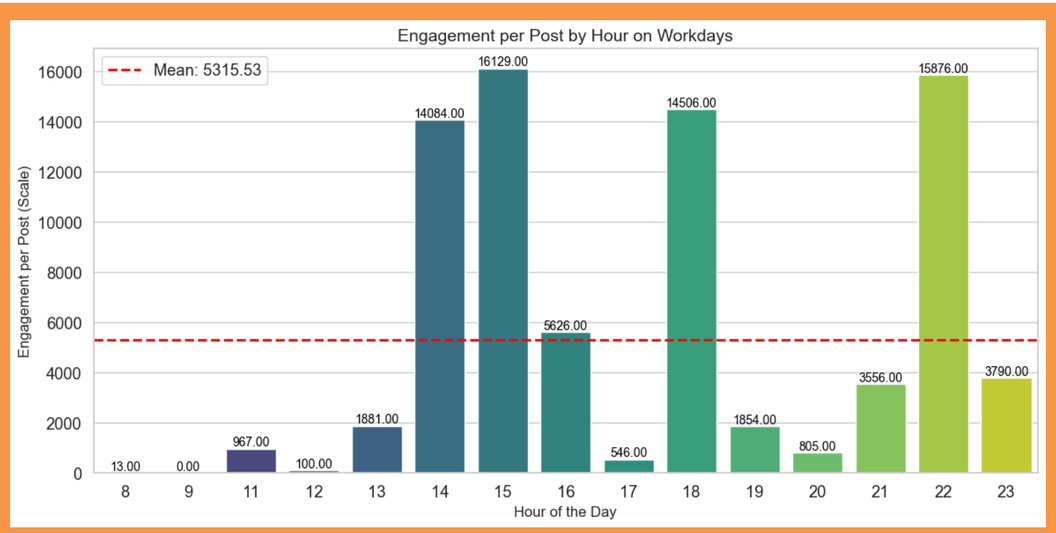


As observed, the highest engagement occurs around 14:00, 15:00, 18:00, and 22:00. Further analysis will focus on engagement during weekdays and the Christmas holiday to better understand user behavior on social media and identify periods that drive higher engagement.

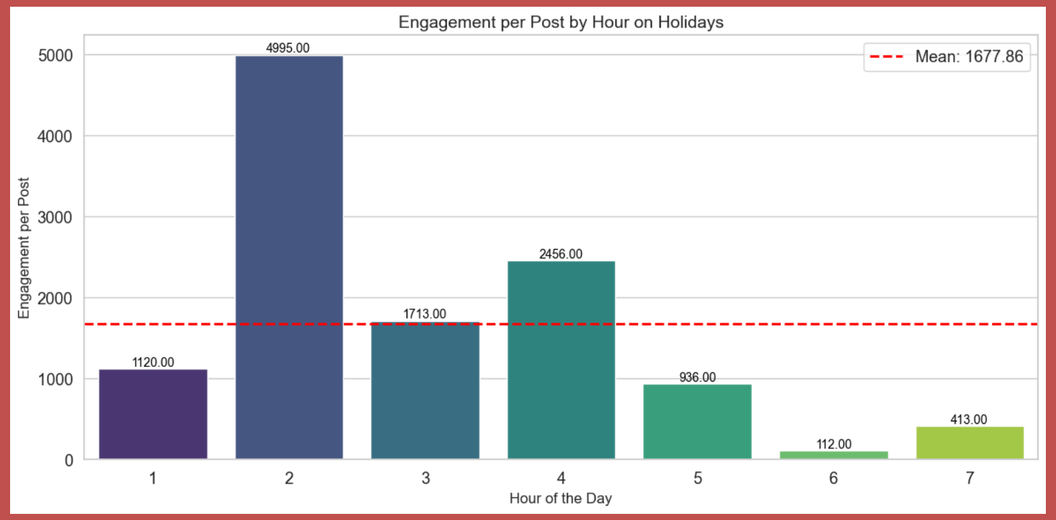
Insights

Analysis on Posting & Engagement by Hours on Working Days & Holiday

WORKINGDAY



HOLIDAY

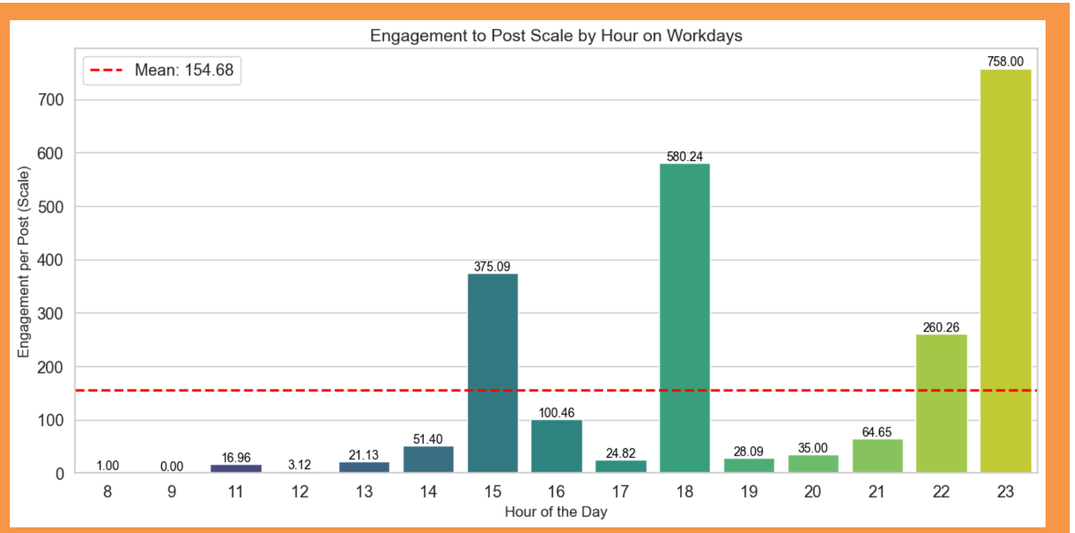


The bar plot illustrates that while posts occur almost every hour (except at 10 AM), user engagement—active interaction with social media content—is concentrated between 8 AM and 11 PM. The highest engagement peaks are observed at 2 PM, 3 PM, 6 PM, and 10 PM. In contrast, during the Christmas holiday, engagement shifts to the early morning hours, from 1 AM to 7 AM, reflecting a notable difference in user behavior during holiday periods.

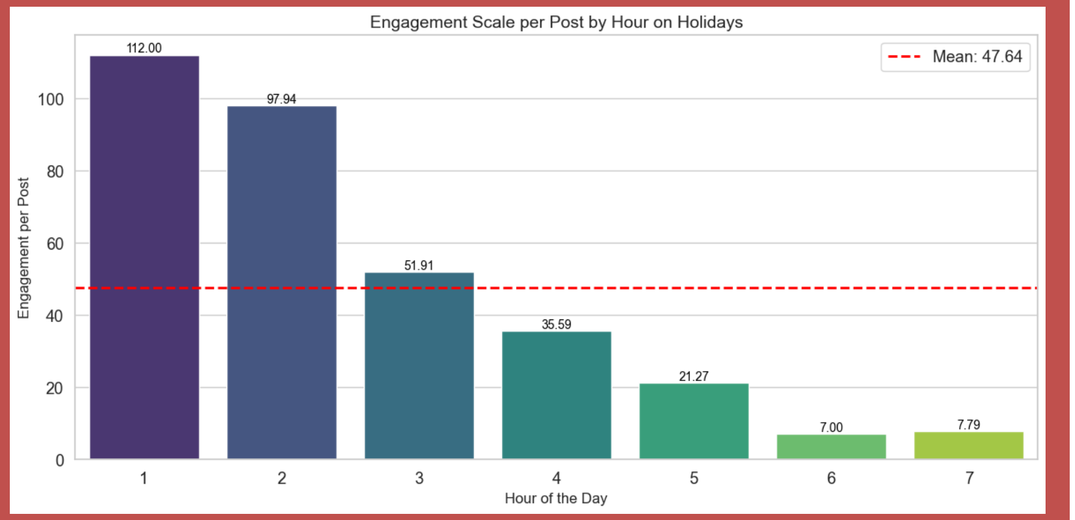
Insights

Analysis on Posting & Engagement by Hours on Working Days & Holiday

WORKINGDAY



HOLIDAY



The highest engagement scale on working days occurs at 3:00 PM, 6:00 PM, and 11:00 PM. On holidays, peak engagement is observed during the early hours of 1:00 AM and 2:00 AM.

Insights

Analysis Unsupported Learning K-Means & PCA

Cluster Analysis Explanation:

Cluster 0: Represents highly popular and widely viewed content, yet it lacks the ability to spark significant discussions or interactions.

Cluster 1: Represents content with potentially distinct characteristics, such as niche-focused topics or content aimed at fostering a sense of community and engagement.

Insights

Analysis Most Phototheme (Topics) by Hour

	hour	phototheme_cluster0	total_engagement_cluster0	phototheme_cluster1	total_engagement_cluster1
0	1	other	1,120.0	0	0.0
1	2	traveling	4,689.0	0	0.0
2	3	traveling	1,368.0	0	0.0
3	4	event	1,952.0	0	0.0
4	5	traveling	696.0	0	0.0
5	6	0	0.0	other	112.0
6	7	other	392.0	traveling	21.0
7	8	0	0.0	other	13.0
8	9	0	0.0	other	0.0
9	11	0	0.0	traveling	915.0
10	12	0	0.0	other	80.0
11	13	0	0.0	other	1,832.0
12	14	traveling	14,014.0	sport	60.0
13	15	other	16,129.0	traveling	0.0
14	16	other	2,454.0	0	0.0
15	17	other	322.0	0	0.0
16	18	other	14,462.0	other	44.0
17	19	traveling	1,530.0	traveling	324.0
18	20	portrait	805.0	0	0.0
19	21	traveling	3,556.0	0	0.0
20	22	portrait	14,706.0	0	0.0
21	23	other	3,790.0	0	0.0

Cluster Analysis Explanation:

Cluster 0: Represents highly popular and widely viewed content, yet it lacks the ability to spark significant discussions or interactions.

Cluster 1: Represents content with potentially distinct characteristics, such as niche-focused topics or content aimed at fostering a sense of community and engagement.

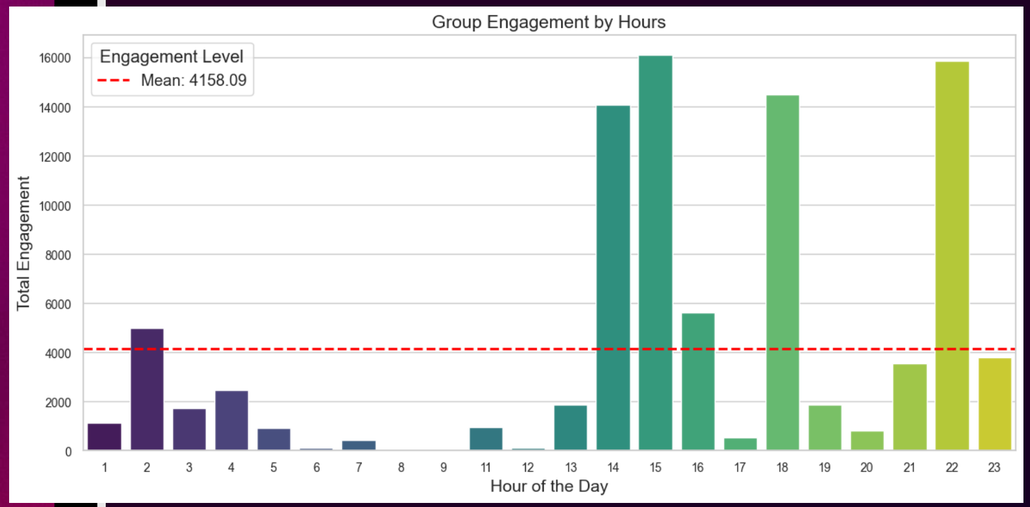
Cluster 0, despite having high engagement, primarily consists of passive interactions such as 'views.' In contrast, Cluster 1 exhibits more active engagement, where users are more likely to give likes, retweets, reposts, and participate in conversations.

Business Solution

The conclusion based on the data is that the best time to publish content is at 3:00 PM, as it is the most efficient for engagement.

01 15.00 exist in both cluster 0 & cluster 1

02 Content posted 15.00 on working day creates the highest engagement

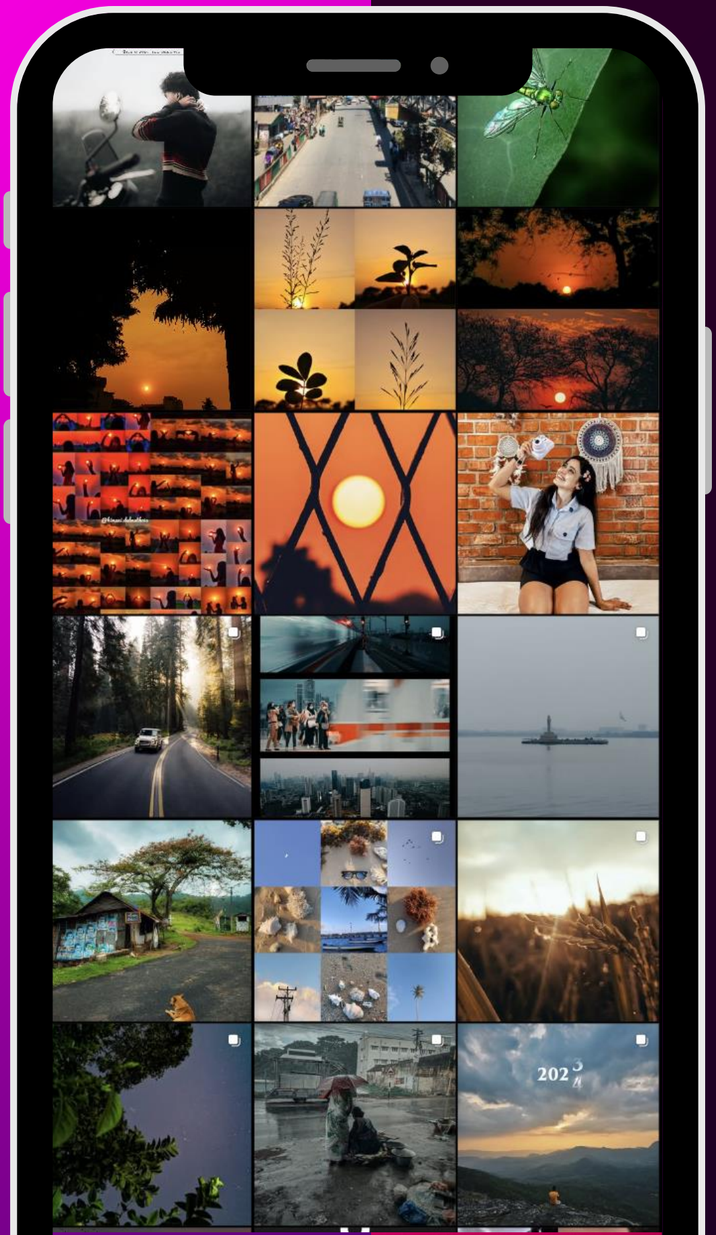



```
# Hours (intersection) among Clusters
common_hours = np.intersect1d(sorted_hours_clh0, sorted_hours_clh1)

print('Intersection Hours Among Clusters :',common_hours)
```

```
Intersection Hours Among Clusters : [ 7 14 15 18 19]
```

My Profile



An aviation professional retiree who transitioned into a data science enthusiast in 2023, they joined the Rakamin Academy Data Science online bootcamp from April 2023 to February 2024.

Then completed the Coursera Google Business Intelligence Professional Certificate in November 2024 and are now eager to further explore data science to gain a deeper understanding of coding, driven by a passion for AI and automation.

Link of my profile as follows :
https://www.coursera.org/learner/amarindra-ardinova-8231
https://www.rakamin.com/profile/amarindra-ardinova-7vymdzcnnlr2aotq

Thank You