EXPLORATORY DATA ANALYSIS ON FACEBOOK ACTIVITIES

TABLE OF CONTENTS

Introduction

Social media has transformed how individuals share their lives, interact with others, and express themselves. For many, platforms like Facebook have served as digital diaries, spaces for social interaction, and outlets for sharing personal milestones, opinions, and content. This project focuses on a detailed analysis of personal Facebook post data over the past decade (2014-2024) to uncover trends, patterns, and insights that reflect changes in engagement, activity levels, and possibly underlying shifts in life events or interests.

OBJECTIVES

This analysis will explore various aspects of Facebook post activity, such as:

1. Temporal Trends: Examine how the volume of posts varies over the years, with attention to fluctuations, peaks, and declines. This can reflect changes in platform usage patterns, engagement with online communities, and personal sharing habits.
2. Seasonality: Identify recurring patterns or cycles within the data. Certain times of the year, such as holidays, birthdays, or significant public events, might prompt increased activity.
3. Growth and Engagement Analysis: Determine if there is a noticeable growth or decline in post frequency over the years. Insights could reveal periods of heightened engagement and points of change, possibly aligned with life stages, professional shifts, or changes in interest levels.
4. Anomalies and Outliers: Investigate periods of unusually high or low

activity, which may correspond to unique events, lifestyle changes, or shifts in platform interaction style.

5. Impact of External Events: Contextualize spikes or drops in activity in light of major external factors. World events, policy changes in the platform, or new social media trends could influence how and when posts were made.

METHODOLOGY

Data for this analysis is derived from personal Facebook post history, specifically focusing on the post count by month over a ten-year period. This temporal dataset enables a time-series analysis, making it possible to visualize long-term trends, pinpoint recurring patterns, and assess the consistency of posting habits. Each data point represents the monthly

post count, allowing for a granular understanding of posting frequency over time.

SIGNIFICANCE

By examining a decade of social media interaction, this project offers insights into the evolving relationship with digital platforms, reflecting broader shifts in personal expression and social media's role in everyday life. The findings could inform future decisions about social media usage, providing a personal perspective on how digital habits have evolved and potentially inspiring mindful engagement with platforms like Facebook.

Through this analysis, the project will shed light on:

- Personal Digital Evolution: An understanding of how personal use of Facebook has evolved in frequency and consistency.

- Data-Driven Insights for Future Social Media Use: An awareness of the habits and preferences that drive engagement, which can guide more intentional social media use in the future.
- Reflection of Broader Social Media Trends: Although this is an individual dataset, the analysis may echo broader trends in social media engagement and user behavior.

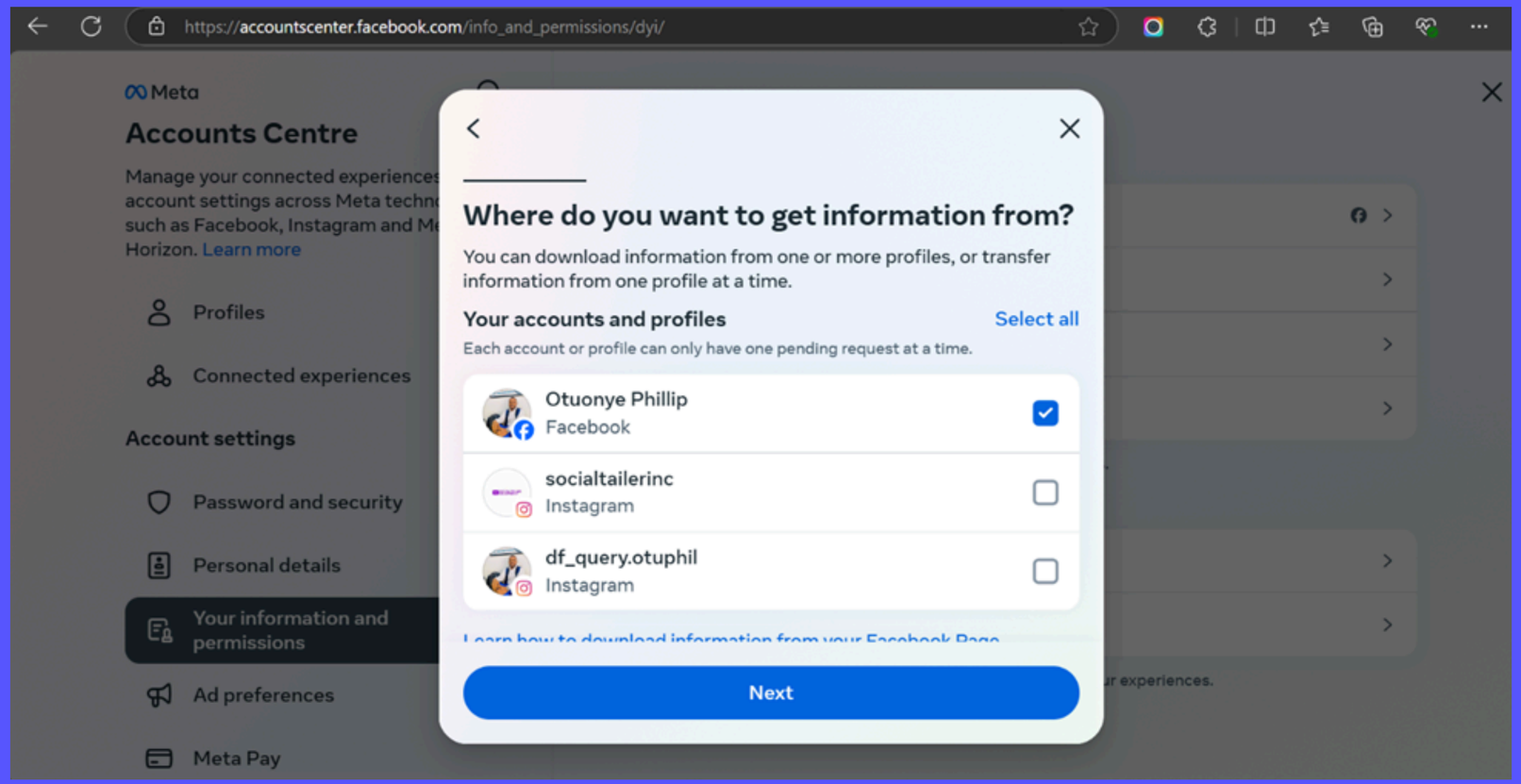
Expected Outcomes

The analysis aims to produce visualizations and interpretive insights that not only highlight posting patterns but also contextualize them within personal experiences and broader social media trends. Additionally, by observing how engagement has shifted over a decade, the project offers a lens through which to view both the personal and collective evolution of social media use.

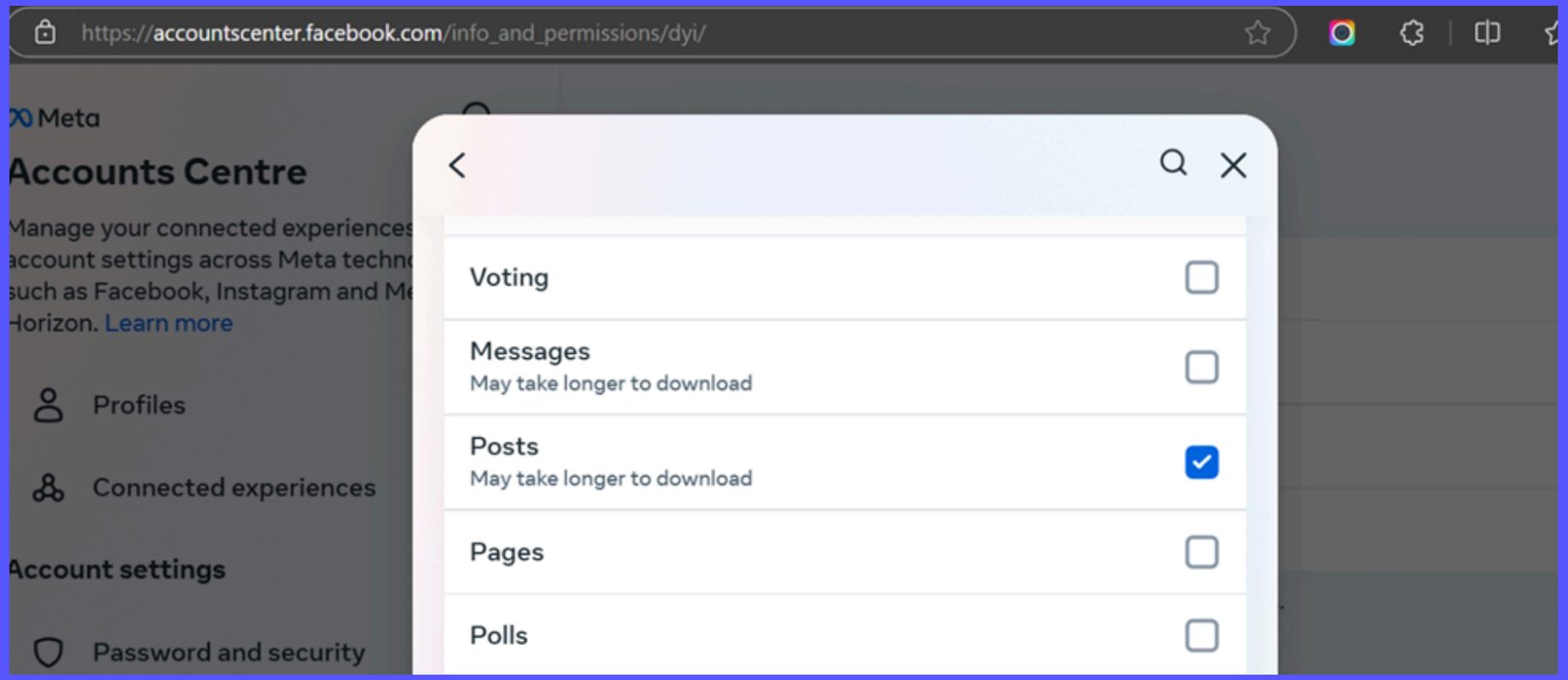
Data Preparation & Cleaning

8

Data collection: (1) Facebook



Posts



Data Migration

The downloaded Facebook files are then imported into Jupyter Notebook. This is done by importing Pandas and also the Facebook JSON file.

[9]:
```
# read Facebook json file into a dataframe
df = pd.read_json("file:///C:/Users/USER/Desktop/projects_kitchen/technology/posts/posts_on_other_pages_and_profiles.json")

df.head()
```

[9]:

	timestamp	media	label_values	fbid	ent_name
0	2016-12-13 07:15:23	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	1212277202182206	EntDirectedPost
1	2016-12-13 07:15:16	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	1516476981701861	EntDirectedPost
2	2016-12-13 07:15:10	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	10209525082296476	EntDirectedPost
3	2016-12-13 07:15:00	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	10154669123837348	EntDirectedPost
4	2016-12-12 07:27:19	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	1851108238437176	EntDirectedPost

Data Cleaning

Now it’s time to get my hands dirty.

Rename timestamp to date column

```
# Change date and rename timestamp column
df.rename(columns={'timestamp': 'date'}, inplace=True)

df.info()
df.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1190 entries, 0 to 1189
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   date          1190 non-null   datetime64[ns]
 1   media         1190 non-null   object
 2   label_values  1190 non-null   object
 3   fbid          1190 non-null   int64
 4   ent_name      1190 non-null   object
dtypes: datetime64[ns](1), int64(1), object(3)
memory usage: 46.6+ KB
```

	date	media	label_values	fbid	ent_name
0	2016-12-13 07:15:23	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	1212277202182206	EntDirectedPost
1	2016-12-13 07:15:16	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	1516476981701861	EntDirectedPost
2	2016-12-13 07:15:10	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	10209525082296476	EntDirectedPost
3	2016-12-13 07:15:00	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	10154669123837348	EntDirectedPost
4	2016-12-12 07:27:19	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	1851108238437176	EntDirectedPost

Shape of dataset

```
# Shape of dataset
print(df.shape)
df.tail()
```

```
(1190, 5)
```

	date	media	label_values	fbid	ent_name
1185	2020-12-04 18:11:23	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	1002870916864231	EntDirectedPost
1186	2020-12-04 17:43:50	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	10216899599307652	EntDirectedPost
1187	2020-12-03 10:45:36	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	10217065471699306	EntDirectedPost
1188	2020-11-28 09:54:32	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	10222143251322164	EntDirectedPost
1189	2020-11-26 13:32:00	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	1087782868314035	EntDirectedPost

Shape of dataset

```
# Shape of dataset
print(df.shape)
df.tail()
```

```
(1190, 5)
```

	date	media	label_values	fbid	ent_name
1185	2020-12-04 18:11:23	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	1002870916864231	EntDirectedPost
1186	2020-12-04 17:43:50	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	10216899599307652	EntDirectedPost
1187	2020-12-03 10:45:36	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	10217065471699306	EntDirectedPost
1188	2020-11-28 09:54:32	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	10222143251322164	EntDirectedPost
1189	2020-11-26 13:32:00	[]	[{'ent_field_name': 'Message', 'label': 'Messa...	1087782868314035	EntDirectedPost

Exploratory Data Analysis

FIGURE OUT MONTHLY POST

```
# Monthly post count
df= df.set_index('date')
# Set as Index the date column
post_counts= df['media'].resample('MS').size()
# Resample the media column y month and count the number of occurring posts
post_counts.head(300)
```

```
date
2015-11-01     1
2015-12-01    11
2016-01-01    11
2016-02-01     3
2016-03-01    20
              ..
2024-06-01    88
2024-07-01    64
2024-08-01    59
2024-09-01    68
2024-10-01     4
Freq: MS, Name: media, Length: 108, dtype: int64
```

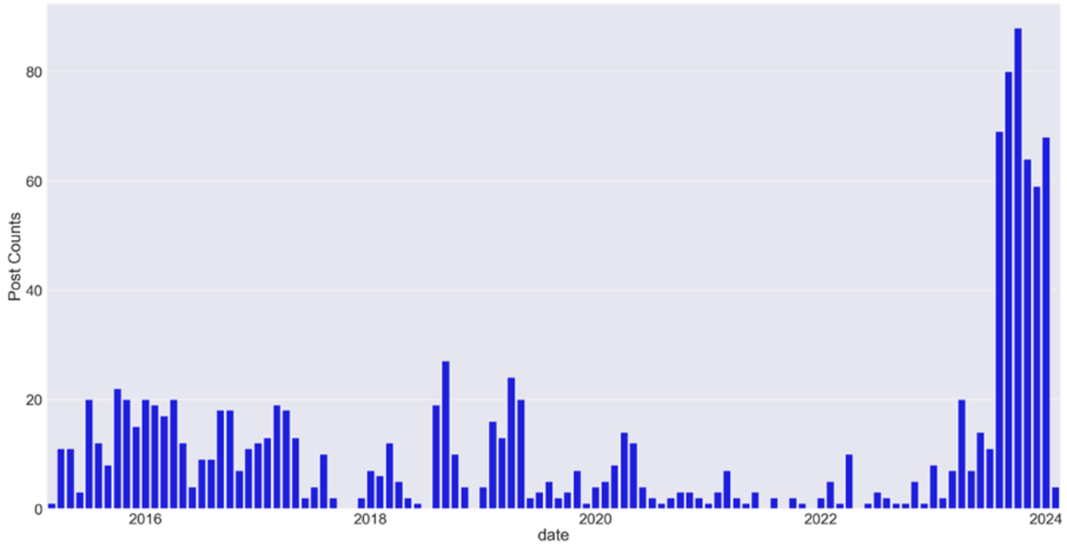
VISUALIZATION

```python
# Set figure and font size
sns.set(rc={'figure.figsize': (40, 20)})
sns.set(font_scale=3)

# Assuming `post_counts` and `x_labels` are defined earlier in your code
sns.barplot(x=x_labels, y=post_counts, color="blue")

# Only show x-axis labels for Jan 1 of every other year
tick_positions = np.arange(10, len(x_labels), step=24)

# Reformat date to display year only
plt.xticks(tick_positions, x_labels[tick_positions].strftime("%Y"))
plt.ylabel("Post Counts")
plt.show()
```



INSIGHT

Based on the time series data you provided, which shows the number of media posts (or interactions) recorded monthly from November 2015 to October 2024, you can derive several insights regarding trends, patterns, and potential areas for strategic planning. Here's a structured analysis:

Observed Insights

1. Overall Growth Trend:

The data appears to show varying levels of media activity over time. If you see an overa increase in post counts from the initial months to the later months, it indicates growth in engagement or content production.

2. Seasonal Patterns:

Monthly Fluctuations: Certain months seem to have notably higher counts than

others (for instance, the spike to 88 in June 2024). This could suggest seasonal trends where certain times of the year are more conducive to media production or engagement.

Recurrent Peaks and Valleys: If certain months consistently have higher or lower counts, that could indicate recurring events or seasonal factors influencing media activity. For example, spikes during holiday seasons or particular events.

3. Anomalies or Outliers:

Significant spikes or drops, such as the jump to 88 posts in June 2024 compared to the preceding months, warrant further investigation. It might correlate with a specific campaign, product launch, or external event.

The sharp drop to 4 in October 2024 suggests an abrupt change in behavior, which could be worth analyzing further to understand its causes.

4. Engagement Growth Over Time:

Comparing earlier periods (2015-2016) to more recent periods (2023-2024) will help assess whether the volume of media content is growing and if your engagement strategies are becoming more effective.

5. Recent Decrease:

The drop in media activity observed in October 2024 may indicate a seasonal lull, a strategy change, or could require investigation into whether there were internal or external factors impacting content production.

Suggested Actions Based on Insights

1. Analyze Causes of Peaks and Valleys:

Delve into the reasons behind spikes in media activity during certain months. Assess

whether they correspond to specific marketing campaigns, seasonal promotions, or other events.

2. Refine Content Strategy:

If certain months consistently perform better, consider scheduling content or campaigns during those peak times to maximize engagement.

3. Investigate Anomalies:

Conduct a deeper analysis into the sharp drop in October 2024 to understand if it was due to a strategic decision, resource constraints, or external factors (like a change in consumer behavior).

4. Forecasting:

Use this historical data to build predictive models for future months. Analyzing past trends can help anticipate future media activity and inform resource

allocation and planning.

5. Consider Audience Feedback:

Gather audience insights or feedback, especially around times of low engagement, to understand what types of content resonate and what might be missing.

6. Monitor Trends Continuously:

Implement regular reviews of media performance to adapt quickly to changing trends, ensuring that you remain aligned with audience interests and market demands.

Recommendation 21

Based on the insights derived from your time series data on monthly media posts, here are some specific recommendations to optimize media engagement, address anomalies, and plan future strategies effectively:

1. Leverage High-Engagement Periods:
   - Action: Focus media campaigns, promotions, and launches during historically high-engagement months, like June 2024, to maximize audience interaction and visibility.
   - Tactics: Plan out a content calendar around these peak months, aligning with possible holidays, special events, or historical patterns in audience interest.

2. Investigate and Address Anomalies (e.g., October 2024 Drop):
   - Action: Look into the significant decrease in post counts in October 2024 to understand underlying causes, whether internal (like resource allocation or policy changes) or external (such as audience interest shifts or platform changes).
   - Tactics: Conduct a retrospective analysis, consulting with relevant teams or looking at audience feedback to detect any possible gaps or sudden shifts that may have affected media output or audience engagement.
3. Strengthen Predictive Planning and Adaptation:
   - Action: Use historical data to develop forecasting models that anticipate months of high or low engagement. This enables proactive planning, better resource management, and more strategic content deployment.

- Tactics: Employ statistical or machine learning models to make data-driven predictions for the upcoming months, adjusting media strategy and resources as required.

4. Establish a Feedback Loop with the Audience:
   - Action: Gather insights from your audience around times of both high and low engagement to better understand preferences and improve future content.
   - Tactics: Use surveys, polls, or social listening tools during various times of the year to gain qualitative insights into what resonates most with your audience. Integrate this feedback into your content strategy, especially for months with low engagement.
5. Refine Seasonal Content Strategy:
   - Action: Develop content tailored to observed seasonal trends, leveraging insights

- on audience behavior and platform engagement patterns.
- Tactics: Create seasonal content that aligns with known spikes, such as holiday-themed posts or event-related media, and consider promotional offers or interactive content during peak engagement periods.

6. Regular Monitoring and Agile Adjustments:
   - Action: Implement a system for ongoing monitoring of monthly media activity, allowing for rapid response to changing engagement patterns or unexpected shifts.
   - Tactics: Schedule monthly or quarterly performance reviews with relevant teams to ensure that content strategy aligns with current trends, making adjustments in real-time to remain responsive to the audience.