Social Media Analysis

Analysis among 5 different social media apps used globally:
Instagram, youtube, bilibili, TikTok, RedNote

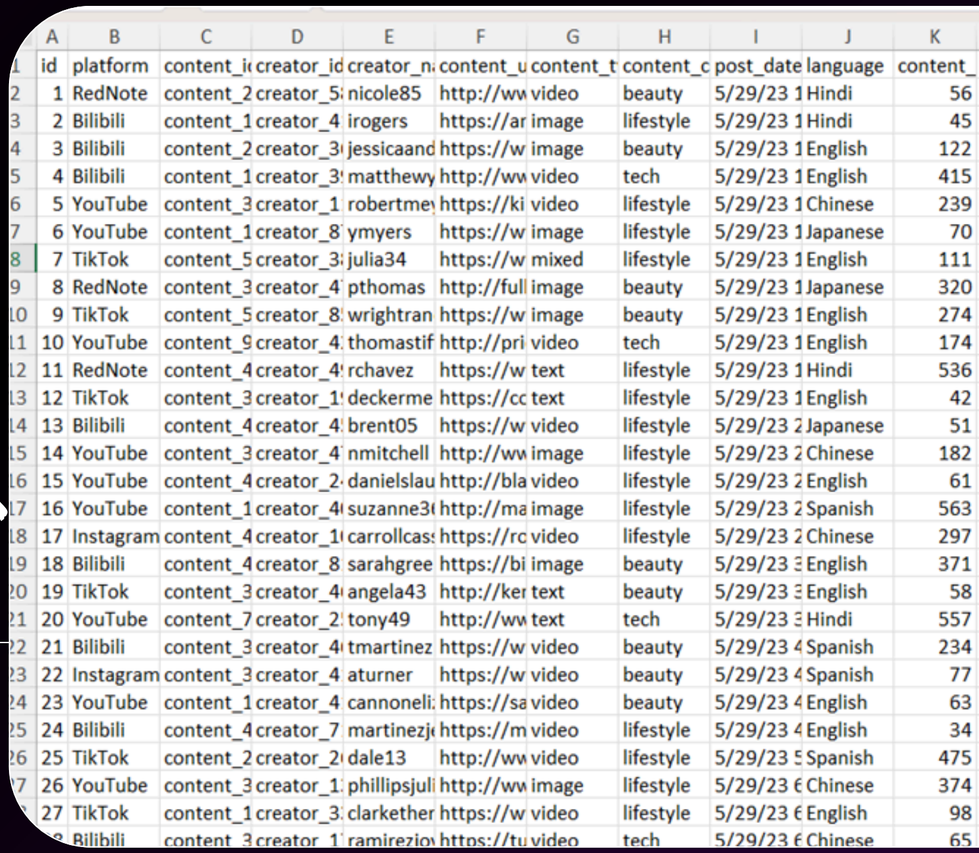
Turning Data into Insights

Using Excel & Python

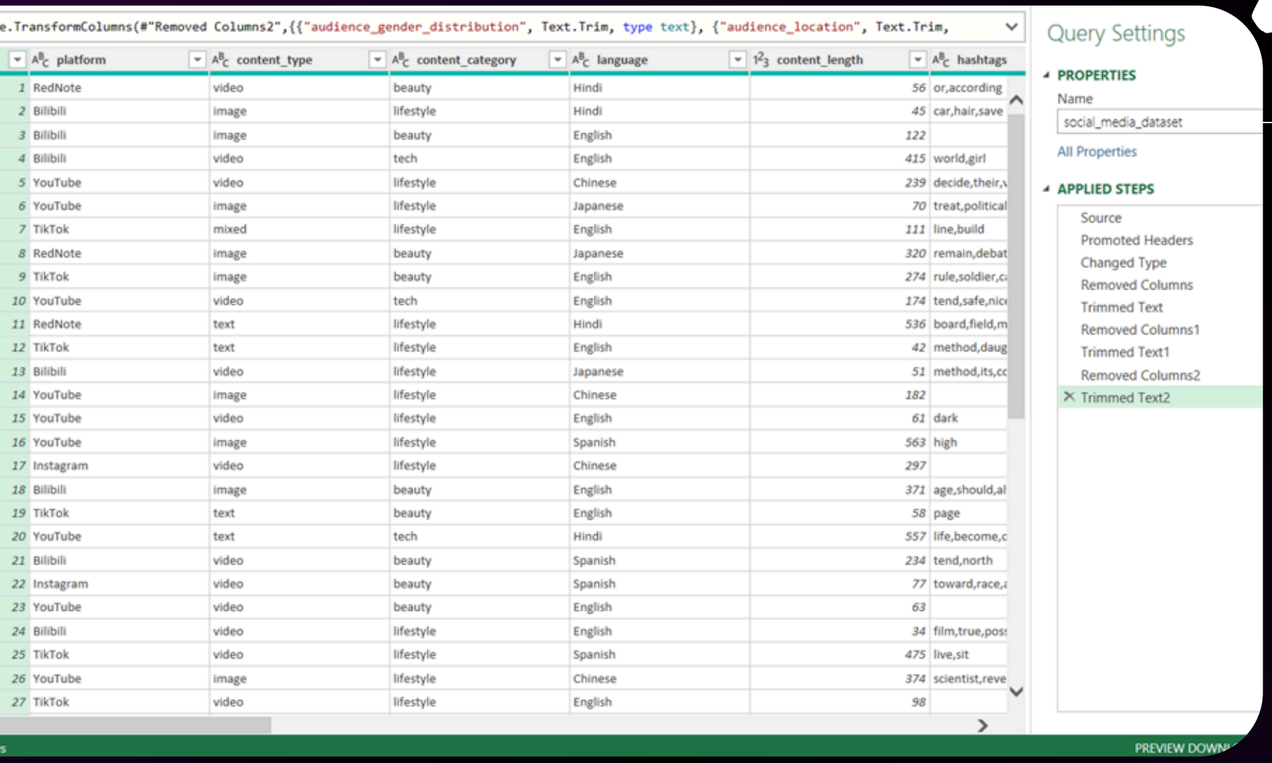
Key Questions Driving the Analysis

- What age group uses which platform the most and interacts the most with content?
- Most preferred content type on each platform
- Which platform has the highest engagement rate?
- Is sponsored content performing better or worse than organic content?
- Which content category performs best?
- What days or months have the highest posting activity or highest average views?
- Which sponsor category gets the highest engagement?

Data Loading in Excel (Power Query)



	A	B	C	D	E	F	G	H	I	J	K
1	id	platform	content_i	creator_id	creator_n	content_u	content_t	content_c	post_date	language	content_
2	1	RedNote	content_2	creator_5	nicole85	http://ww	video	beauty	5/29/23 1	Hindi	56
3	2	Bilibili	content_1	creator_4	irogers	https://ar	image	lifestyle	5/29/23 1	Hindi	45
4	3	Bilibili	content_2	creator_3	jessicaand	https://w	image	beauty	5/29/23 1	English	122
5	4	Bilibili	content_1	creator_3	matthewy	http://ww	video	tech	5/29/23 1	English	415
6	5	YouTube	content_3	creator_1	robertmey	https://ki	video	lifestyle	5/29/23 1	Chinese	239
7	6	YouTube	content_1	creator_8	ymyers	https://w	image	lifestyle	5/29/23 1	Japanese	70
8	7	TikTok	content_5	creator_3	julia34	https://w	mixed	lifestyle	5/29/23 1	English	111
9	8	RedNote	content_3	creator_4	pthomas	http://full	image	beauty	5/29/23 1	Japanese	320
10	9	TikTok	content_5	creator_8	wrightran	https://w	image	beauty	5/29/23 1	English	274
11	10	YouTube	content_9	creator_4	thomastif	http://pri	video	tech	5/29/23 1	English	174
12	11	RedNote	content_4	creator_4	rchavez	https://w	text	lifestyle	5/29/23 1	Hindi	536
13	12	TikTok	content_3	creator_1	deckerme	https://cc	text	lifestyle	5/29/23 1	English	42
14	13	Bilibili	content_4	creator_4	brent05	https://w	video	lifestyle	5/29/23 2	Japanese	51
15	14	YouTube	content_3	creator_4	nmitchell	http://ww	image	lifestyle	5/29/23 2	Chinese	182
16	15	YouTube	content_4	creator_2	danielslau	http://bla	video	lifestyle	5/29/23 2	English	61
17	16	YouTube	content_1	creator_4	suzanne3(	http://ma	image	lifestyle	5/29/23 2	Spanish	563
18	17	Instagram	content_4	creator_1	carrollcas	https://rc	video	lifestyle	5/29/23 2	Chinese	297
19	18	Bilibili	content_4	creator_8	sarahgree	https://bi	image	beauty	5/29/23 3	English	371
20	19	TikTok	content_3	creator_4	angela43	http://ker	text	beauty	5/29/23 3	English	58
21	20	YouTube	content_7	creator_2	tony49	http://ww	text	tech	5/29/23 3	Hindi	557
22	21	Bilibili	content_3	creator_4	tmartinez	https://w	video	beauty	5/29/23 4	Spanish	234
23	22	Instagram	content_3	creator_4	aturner	https://w	video	beauty	5/29/23 4	Spanish	77
24	23	YouTube	content_1	creator_4	cannoneli	https://sa	video	beauty	5/29/23 4	English	63
25	24	Bilibili	content_4	creator_7	martinezj	https://m	video	lifestyle	5/29/23 4	English	34
26	25	TikTok	content_2	creator_2	dale13	http://ww	video	lifestyle	5/29/23 5	Spanish	475
27	26	YouTube	content_3	creator_1	phillipsjuli	http://ww	image	lifestyle	5/29/23 6	Chinese	374
	27	TikTok	content_1	creator_3	clarkether	https://w	video	lifestyle	5/29/23 6	English	98
		Bilibili	content_3	creator_1	ramirezjo	https://tu	video	tech	5/29/23 6	Chinese	65



```
e.TransformColumns(#"Removed Columns2",{{"audience_gender_distribution", Text.Trim, type text}, {"audience_location", Text.Trim,
```

	platform	content_type	content_category	language	content_length	hashtags
1	RedNote	video	beauty	Hindi	56	or,according
2	Bilibili	image	lifestyle	Hindi	45	car,hair,save
3	Bilibili	image	beauty	English	122	
4	Bilibili	video	tech	English	415	world,girl
5	YouTube	video	lifestyle	Chinese	239	decide,their,
6	YouTube	image	lifestyle	Japanese	70	treat,political
7	TikTok	mixed	lifestyle	English	111	line,build
8	RedNote	image	beauty	Japanese	320	remain,debat
9	TikTok	image	beauty	English	274	rule,soldier,c
10	YouTube	video	tech	English	174	tend,safe,nic
11	RedNote	text	lifestyle	Hindi	536	board,field,m
12	TikTok	text	lifestyle	English	42	method,daug
13	Bilibili	video	lifestyle	Japanese	51	method,its,c
14	YouTube	image	lifestyle	Chinese	182	
15	YouTube	video	lifestyle	English	61	dark
16	YouTube	image	lifestyle	Spanish	563	high
17	Instagram	video	lifestyle	Chinese	297	
18	Bilibili	image	beauty	English	371	age,should,al
19	TikTok	text	beauty	English	58	page
20	YouTube	text	tech	Hindi	557	life,become,c
21	Bilibili	video	beauty	Spanish	234	tend,north
22	Instagram	video	beauty	Spanish	77	toward,race,a
23	YouTube	video	beauty	English	63	
24	Bilibili	video	lifestyle	English	34	film,true,poss
25	TikTok	video	lifestyle	Spanish	475	live,sit
26	YouTube	image	lifestyle	Chinese	374	scientist,reve
27	TikTok	video	lifestyle	English	98	

Query Settings

PROPERTIES

Name

social_media_dataset

All Properties

APPLIED STEPS

- Source
- Promoted Headers
- Changed Type
- Removed Columns
- Trimmed Text
- Removed Columns1
- Trimmed Text1
- Removed Columns2
- ✕ Trimmed Text2

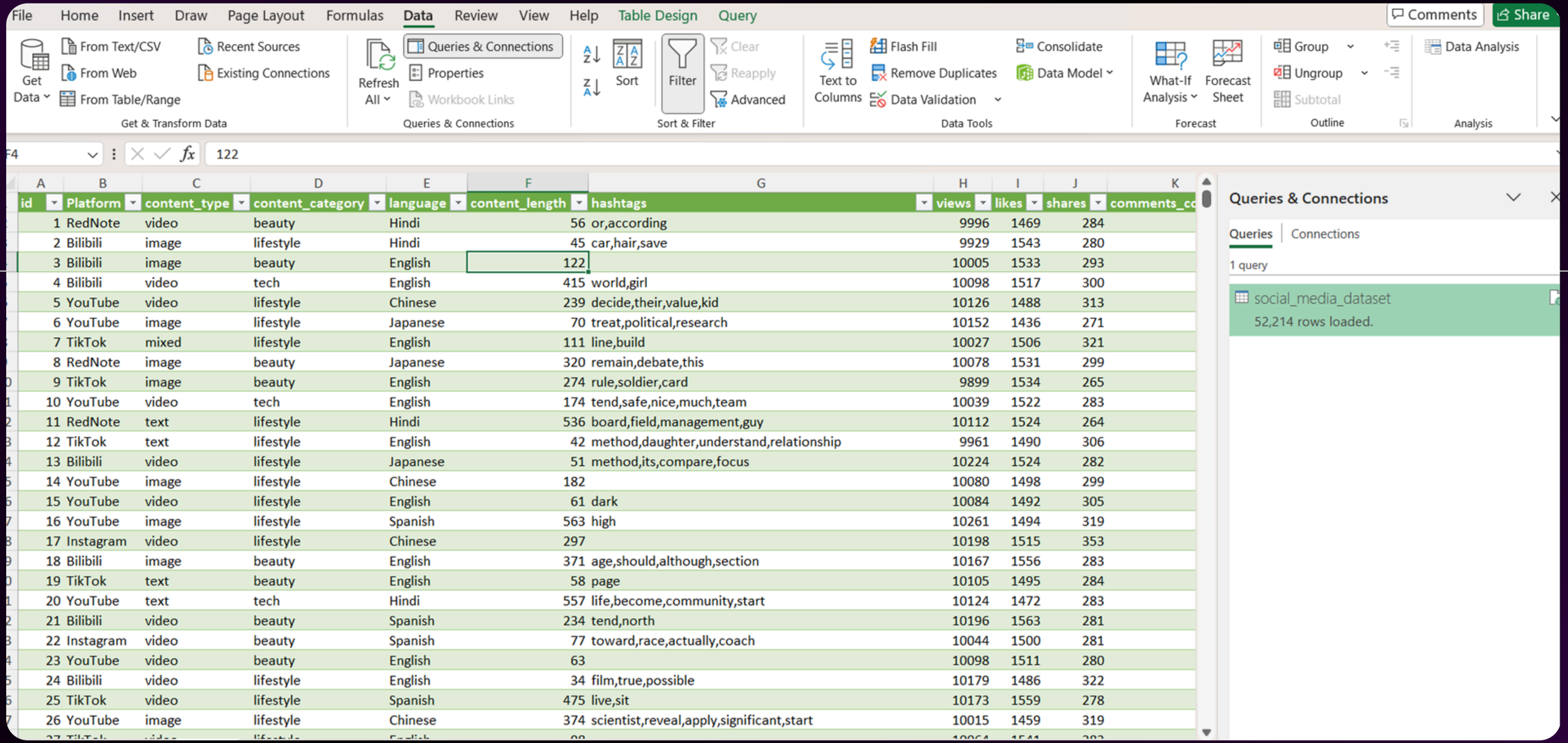
PREVIEW DOWN

Data Import and cleanse using Power Query

- Loaded raw social media dataset using Power Query
- Removed unnecessary columns
- Changed incorrect data types (date, text, numbers)
- Ensured data consistency before analysis

Data Cleaning & Structured Table

- Removed duplicate records
- Handled missing and blank values
- Standardized column names and formats
- Converted raw data into a clean analysis-ready table



id	Platform	content_type	content_category	language	content_length	hashtags	views	likes	shares	comments_c
1	RedNote	video	beauty	Hindi	56	or,according	9996	1469	284	
2	Bilibili	image	lifestyle	Hindi	45	car,hair,save	9929	1543	280	
3	Bilibili	image	beauty	English	122		10005	1533	293	
4	Bilibili	video	tech	English	415	world,girl	10098	1517	300	
5	YouTube	video	lifestyle	Chinese	239	decide,their,value,kid	10126	1488	313	
6	YouTube	image	lifestyle	Japanese	70	treat,political,research	10152	1436	271	
7	TikTok	mixed	lifestyle	English	111	line,build	10027	1506	321	
8	RedNote	image	beauty	Japanese	320	remain,debate,this	10078	1531	299	
9	TikTok	image	beauty	English	274	rule,soldier,card	9899	1534	265	
10	YouTube	video	tech	English	174	tend,safe,nice,much,team	10039	1522	283	
11	RedNote	text	lifestyle	Hindi	536	board,field,management,guy	10112	1524	264	
12	TikTok	text	lifestyle	English	42	method,daughter,understand,relationship	9961	1490	306	
13	Bilibili	video	lifestyle	Japanese	51	method,its,compare,focus	10224	1524	282	
14	YouTube	image	lifestyle	Chinese	182		10080	1498	299	
15	YouTube	video	lifestyle	English	61	dark	10084	1492	305	
16	YouTube	image	lifestyle	Spanish	563	high	10261	1494	319	
17	Instagram	video	lifestyle	Chinese	297		10198	1515	353	
18	Bilibili	image	beauty	English	371	age,should,although,section	10167	1556	283	
19	TikTok	text	beauty	English	58	page	10105	1495	284	
20	YouTube	text	tech	Hindi	557	life,become,community,start	10124	1472	283	
21	Bilibili	video	beauty	Spanish	234	tend,north	10196	1563	281	
22	Instagram	video	beauty	Spanish	77	toward,race,actually,coach	10044	1500	281	
23	YouTube	video	beauty	English	63		10098	1511	280	
24	Bilibili	video	lifestyle	English	34	film,true,possible	10179	1486	322	
25	TikTok	video	lifestyle	Spanish	475	live,sit	10173	1559	278	
26	YouTube	image	lifestyle	Chinese	374	scientist,reveal,apply,significant,start	10015	1459	319	

Pivot Table Analysis

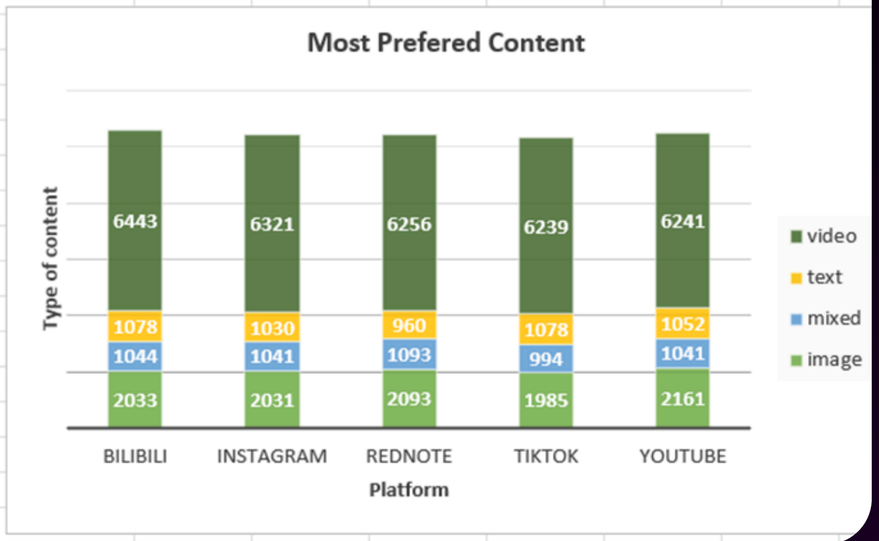
Count of Platform	Column Labels				
Row Labels	Bilibili	Instagram	RedNote	TikTok	YouTube
13-18	1610	1567	1555	1576	1544
19-25	3696	3662	3618	3675	3625
26-35	3149	3156	3140	3012	3243
36-50	1627	1507	1579	1492	1531
50+	516	531	510	541	552

Row Labels	Count of views
Not Sponsored	29900
Sponsored	22314

Count of views	Column Labels	
Row Labels	Not Sponsored	Sponsored
beauty	12097	8926
lifestyle	11870	8891
tech	5933	449

Count of likes	Column Labels		
Row Labels	beauty	lifestyle	tech
Bilibili	4251	4160	2187
Instagram	4197	4137	2089
RedNote	4221	4125	2056
TikTok	4130	4128	2038
YouTube	4224	4211	2060

Count of views	Column Labels			
Row Labels	image	mixed	text	video
Bilibili	2033	1044	1078	6443
Instagram	2031	1041	1030	6321
RedNote	2093	1093	960	6256
TikTok	1985	994	1078	6239
YouTube	2161	1041	1052	6241

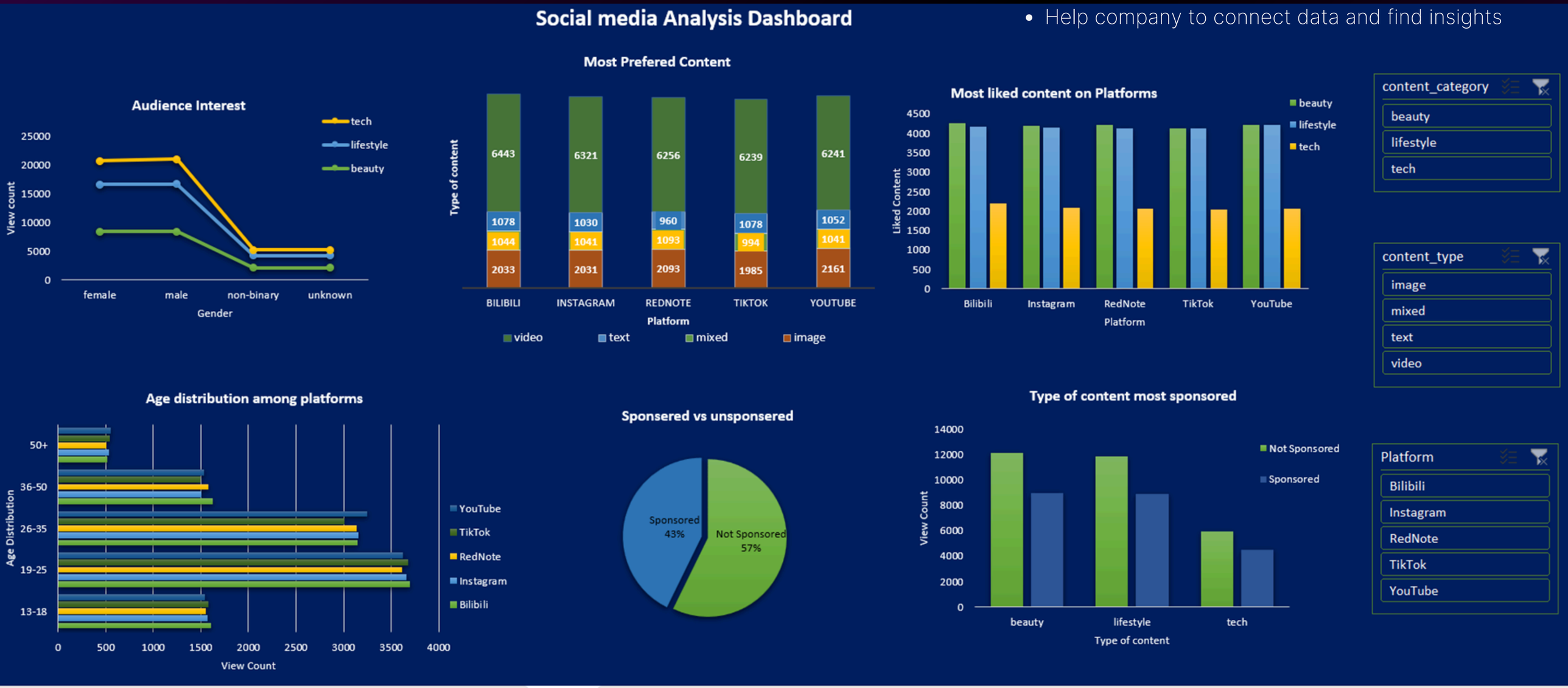


Analysis Using Pivot Tables

- Created pivot tables to summarize engagement metrics
- Analyzed performance by:
- Platform
- Post type
- Engaegement
- Identified high-performing and low-performing content on each platform

Interactive Dashboard (Excel)

- Built a dashboard using charts and pivots
- Visualized key:
  - Likes, Comments, Shares
  - Engagement trends according to age groups
- Easy to understand for non-technical users
- included slicers for clear differentiations
- Help company to connect data and find insights



Data Loading, Cleaning & Exploratory Data Analysis in Python

Insights & Data Understanding

- Loaded data using Pandas
- Checked:
  - Data types
  - Null values
  - Dataset shape
- Cleaned and removed unwanted columns
- Identified patterns in engagement data
- Compared metrics across platforms and post types
- Validated Excel insights using Python analysis

```
## IMPORT LIBRARIES
import pandas as pd
import matplotlib.pyplot as plt


## LOAD DATA

df = pd.read_csv("C:/Users/dimpl/Downloads/social_media_dataset.csv")
df.head()

## BASIC CHECKS

print(df.info())
print(df.describe())
print(df.isnull().sum())
print(df.columns.tolist())

## REMOVE UNNECESSARY COLUMNS

cols_to_drop = [
    "content_url", "content_description", "hashtags",
    "comments_text", "creator_name", "sponsor_name",
    "disclosure_location", "id"]
df = df.drop(columns=cols_to_drop, errors="ignore")
```

```
id                              0
platform                        0
content_id                      0
creator_id                      0
creator_name                    0
content_url                     0
content_type                    0
content_category                0
post_date                       0
language                        0
content_length                  0
content_description             0
hashtags                     8743
views                           0
likes                           0
shares                          0
comments_count                  0
comments_text                8688
follower_count                  0
is_sponsored                    0
disclosure_type                 0
sponsor_name                    0
sponsor_category                0
disclosure_location             0
audience_age_distribution       0
audience_gender_distribution    0
audience_location               0
dtype: int64
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 52214 entries, 0 to 52213
Data columns (total 27 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   id                            52214 non-null  int64
 1   platform                      52214 non-null  object
 2   content_id                    52214 non-null  object
 3   creator_id                    52214 non-null  object
 4   creator_name                  52214 non-null  object
 5   content_url                   52214 non-null  object
 6   content_type                  52214 non-null  object
 7   content_category              52214 non-null  object
 8   post_date                     52214 non-null  object
 9   language                      52214 non-null  object
 10  content_length                52214 non-null  int64
 11  content_description           52214 non-null  object
 12  hashtags                      43471 non-null  object
 13  views                         52214 non-null  int64
 14  likes                         52214 non-null  int64
 15  shares                        52214 non-null  int64
 16  comments_count                52214 non-null  int64
 17  comments_text                 43526 non-null  object
 18  follower_count                52214 non-null  int64
 19  is_sponsored                  52214 non-null  bool
 20  disclosure_type               52214 non-null  object
 21  sponsor_name                  52214 non-null  object
 22  sponsor_category              52214 non-null  object
 23  disclosure_location           52214 non-null  object
 24  audience_age_distribution     52214 non-null  object
 25  audience_gender_distribution  52214 non-null  object
 26  audience_location             52214 non-null  object
dtypes: bool(1), int64(7), object(19)
memory usage: 10.4+ MB
```

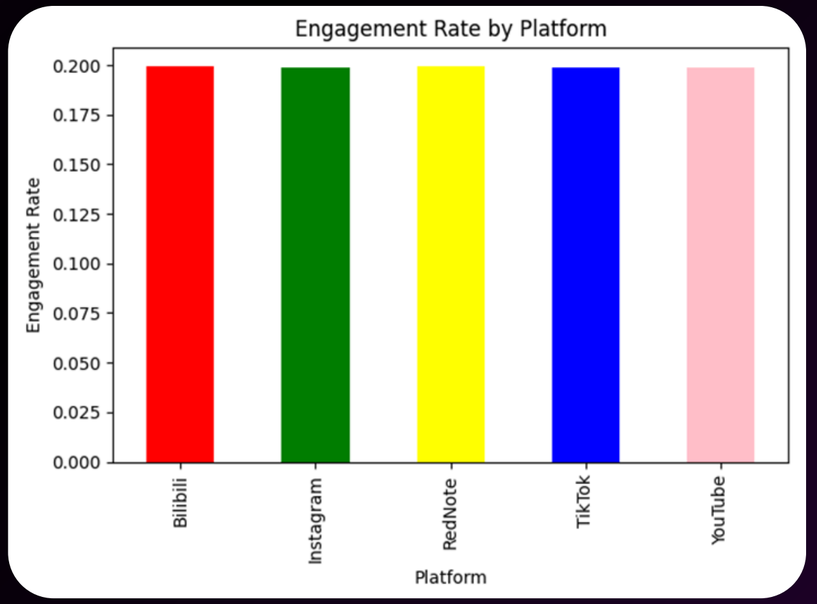
	id	content_length	...	comments_count	follower_count
count	52214.000000	52214.000000	...	52214.000000	52214.000000
mean	26107.500000	229.537787	...	200.015456	499884.514670
std	15073.027815	167.086525	...	14.093912	288044.593669
min	1.000000	10.000000	...	140.000000	1013.000000
25%	13054.250000	95.000000	...	190.000000	250817.750000
50%	26107.500000	174.000000	...	200.000000	498490.000000
75%	39160.750000	360.000000	...	210.000000	749818.750000
max	52214.000000	599.000000	...	258.000000	999998.000000

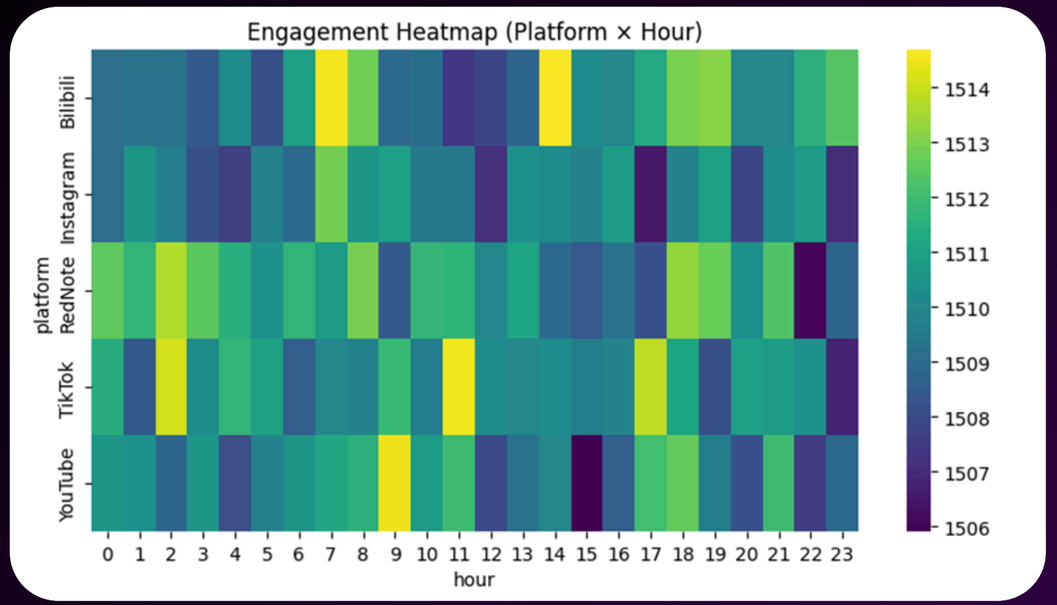
Visualizations Using Python

- Created charts using Python libraries
- Visualized:
  - Engagement rate
  - Engagement trends over time platform-wise
- Supported insights with code-based visuals

```
Engagement by platform
colors = ["red", "green", "yellow", "blue", "pink"]
platform_eng.plot(kind="bar" , color = colors)
plt.title("Engagement Rate by Platform")
plt.xlabel("Platform")
plt.ylabel("Engagement Rate")
plt.show()
```

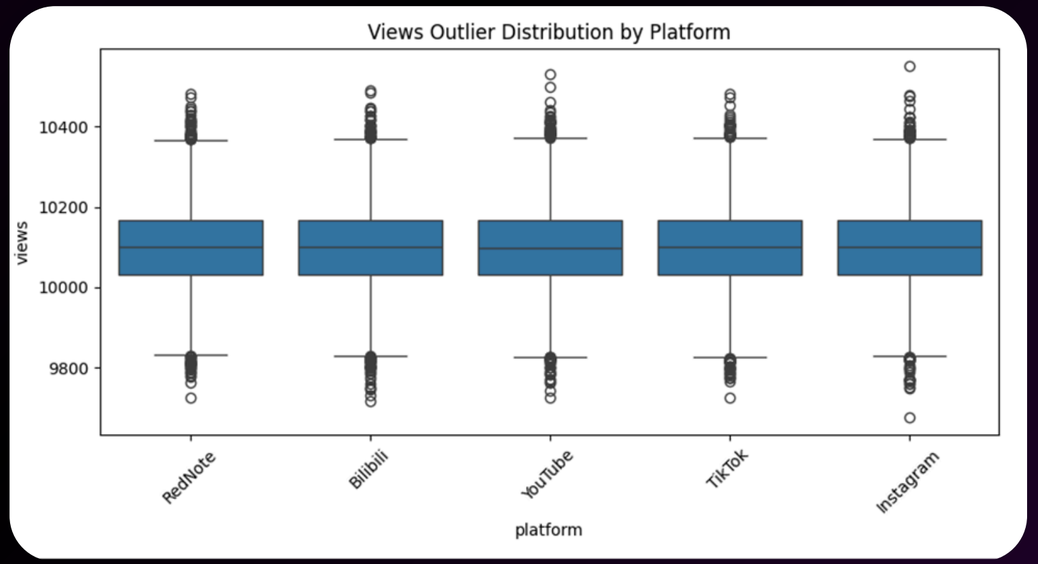
```
import seaborn as sns
import matplotlib.pyplot as plt

df["hour"] = df["post_date"].dt.hour
heat = df.pivot_table(values="likes", index="platform", columns="hour", aggfunc="mean")

plt.figure(figsize=(12,6))
sns.heatmap(heat, cmap="viridis")
plt.title("Engagement Heatmap (Platform × Hour)")
plt.show()
```



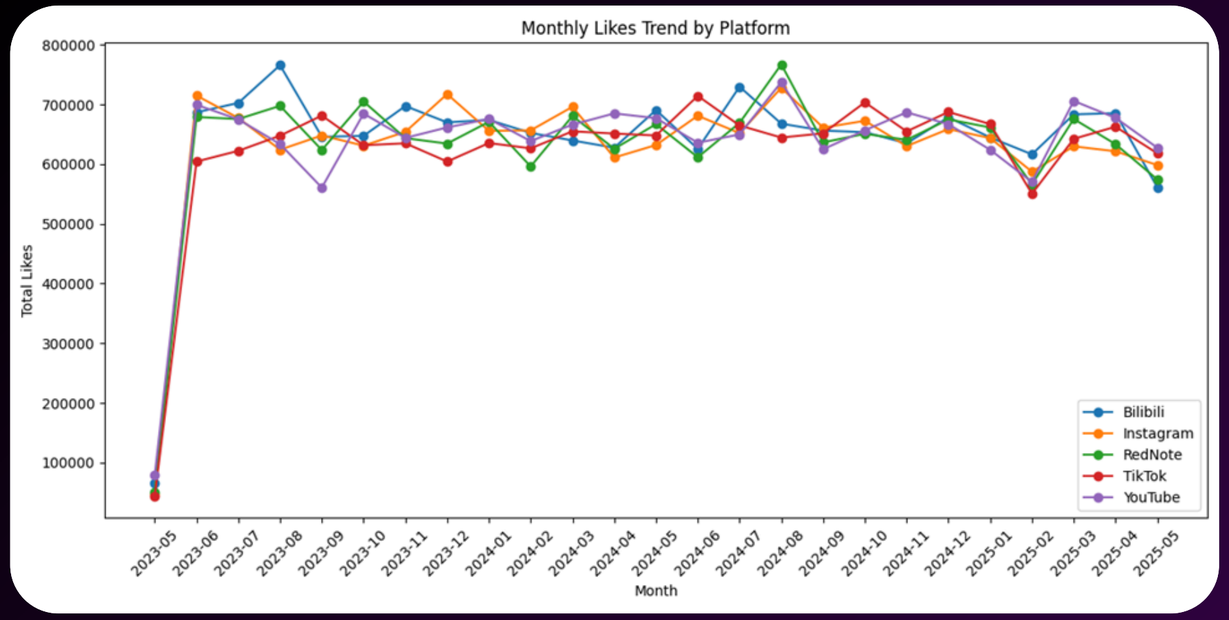


Charts & Visual Analysis

```
plt.figure(figsize=(10,5))
sns.boxplot(data=df, x="platform", y="views")
plt.title("Views Outlier Distribution by Platform")
plt.xticks(rotation=45)
plt.show()
```




```
df["month"] = df["post_date"].dt.to_period("M").astype(str)
trend = df.groupby(["month", "platform"])["likes"].sum().reset_index()

plt.figure(figsize=(12,6))

for p in trend["platform"].unique():
    subset = trend[trend["platform"] == p]
    plt.plot(subset["month"], subset["likes"], marker="o", label=p)

plt.xlabel("Month")
plt.ylabel("Total Likes")
plt.title("Monthly Likes Trend by Platform")

plt.xticks(rotation=45)
plt.legend()
plt.tight_layout()
plt.show()
```



What conclusions can we draw from this?

- The **26–35 age** group is the **most active across platforms** therefore, **making it the primary target audience** for digital marketing campaigns.
- **Beauty content consistently performs best**, indicating strong demand and higher ROI potential for brands in this category.
- **Instagram** records the **highest engagement rate**, making it the most effective platform for audience interaction.
- **Sponsored content** performs **weaker** than organic content, suggesting that brands should **focus on authentic** and **value-driven messaging.**
- **Video** content outperforms other formats, helping brands communicate **more effectively and drive higher** engagement.
- Platform-specific **peak usage times** highlight the importance of optimized posting schedules:
  - YouTube – 9 AM | Instagram – 7 AM | TikTok – 11 AM | RedNote – 2 AM | Bilibili – 7 AM
- **Beauty and Lifestyle** sponsor categories achieve the **highest engagement**, making them **ideal for sponsored** collaborations.

Business Recommendations

- Audience segmentation should focus on the 26–35 age group, as they consistently interact more across platforms which will help brands improve targeting efficiency.
- Platform-specific strategies are essential where Instagram should be used for engagement-led campaigns while others can support reach and awareness.
- Shift content strategy toward video-first formats to simplify messaging and increase user retention.
- Align posting schedules with peak usage times for each platform to increase visibility without additional ad spend.
- Adopt continuous performance monitoring to quickly identify underperforming content and optimize future campaigns.

Business Value

These actions help companies increase engagement efficiency, improve content relevance, and make smarter, data-backed marketing decisions.

Thank you

Khushi Sanhotra