

# Instagram Performance & Growth Analysis

## 1. Project Overview

This project analyzes social media engagement and community growth patterns across multiple brand accounts using a dataset of **29,999 records**. The primary objective is to identify factors that drive **follower conversion**—such as caption length, hashtag density, and media type—and to evaluate performance across 10 diverse content categories. These insights support data-driven decisions for digital marketing, content scheduling, and creative optimization.

## 2. Dataset Summary

The dataset contains **29,999 records** with **17 features** (including engineered metrics) related to post metadata and audience behavior.

### Key Attributes:

- **Post Metadata:** media\_type (Reel, Photo, Video, Carousel), content\_category (e.g., Tech, Beauty), upload\_date.
- **Technical Factors:** caption\_length, hashtags\_count.
- **Discovery Metrics:** reach, impressions, traffic\_source (Home Feed, Explore, Hashtags).
- **Engagement Behavior:** likes, comments, shares, saves.
- **Growth Metrics:** followers\_gained, engagement\_rate, and engagement\_to\_reach\_ratio.

### Data Quality:

- **Completeness:** The dataset contains **zero missing values** across all records.
- **Consistency:** Standardized formats for dates and categorical variables allow for immediate correlation studies and trend analysis.

## 3. Exploratory Data Analysis using Python

### 3.1 Environment Setup

Essential libraries such as **Pandas**, **NumPy** were imported for data manipulation, statistical analysis.

### 3.2 Data Ingestion

The dataset was loaded from a CSV file containing **29,999 records**, representing a comprehensive log of post performance across multiple niches.

### 3.3 Data Cleaning and Preparation

- **Standardization:** Column names were verified and standardized for consistency (e.g., engagement\_rate and followers\_gained).
- **Data Type Validation:** Data types were validated for key metrics such as reach (integer) and engagement\_rate (float) to ensure analytical accuracy.
- **Feature Engineering:** A new metric, **engagement\_to\_reach\_ratio**, was engineered to measure the efficiency of content in converting viewers into active participants.
- **Temporal Conversion:** The upload\_date was converted to a datetime format to facilitate time-series analysis of posting schedules.

Feature Category	Status	Observations
Missing Data	Clean	No nulls found in any column.
Duplicates	Clean	100% unique post IDs.
Outliers	Valid	Numeric values fall within expected social media performance ranges.
Formatting	Standardized	Categorical labels (e.g., "Reel", "Home Feed") are consistent.

### 3.4 Exploratory Analysis

- **Growth Drivers:** Correlation analysis revealed that **hashtags\_count** and **shares** have the strongest positive relationship with **followers\_gained**, making them critical for account growth.
- **Media Performance:** **Videos** and **Reels** demonstrated higher average engagement rates compared to static Photos, suggesting a preference for dynamic content in the algorithm.
- **High-Impact Categories:** The **Beauty** and **Photography** categories exhibited the highest average engagement, while **Technology** required more strategic hashtag usage to maintain reach.
- **Efficiency Metrics:** Analysis showed that high **reach** does not always correlate with high **followers\_gained**, highlighting a "conversion gap" in broadly distributed content that lacks a strong call-to-action.

	post_id	upload_date	media_type	likes	comments	shares	saves	reach	impressions
0	IG0000001	2024-11-30 09:25:22.954916	Reel	31627	7559	4530	6393	615036	1007750
1	IG0000002	2025-08-15 09:25:22.954916	Photo	63206	3490	1680	6809	1237071	1345900
2	IG0000003	2025-09-11 09:25:22.954916	Reel	94373	3727	1761	8367	1127470	1305369
3	IG0000004	2025-09-18 09:25:22.954916	Reel	172053	7222	2875	9290	764030	897874
4	IG0000005	2025-03-21 09:25:22.954916	Video	99646	2703	4444	9746	7004	495406

caption_length	hashtags_count	followers_gained	traffic_source	engagement_rate	content_category
1340	3	899	Home Feed	4.97	Technology
1351	20	805	Hashtags	5.59	Fitness
242	24	758	Reels Feed	8.29	Beauty
446	11	402	External	21.32	Music
1905	8	155	Profile	23.52	Technology






## 4.Data Analysis using SQL (Business Transactions)

We performed structured analysis in PostgreSQL to answer key business questions:

Q1. Identify which format (Reel, Photo, Video, Carousel) is most efficient at converting reach into followers.

Ex Query

```
SELECT
    media_type,
    COUNT(post_id) AS total_posts,
    ROUND(AVG(engagement_rate)::numeric,2) AS avg_engagement_rate,
    ROUND(AVG(followers_gained), 0) AS avg_followers_gained,
    ROUND(SUM(followers_gained) * 1.0 / NULLIF(SUM(reach), 0), 4) AS
    follower_conversion_ratio
FROM public."Insta_Performance"
GROUP BY media_type
ORDER BY avg_engagement_rate DESC;
```

	media_type 	total_posts 	avg_engagement_rate 	avg_followers_gained 	follower_conversion_ratio 
1	Video	7523	14.62	504	0.0005
2	Reel	7515	14.54	494	0.0005
3	Carousel	7526	14.32	505	0.0005
4	Photo	7435	14.19	505	0.0005

Q2. Identify high-performing niches and "High-Risk" segments where reach is high but engagement is failing to retain interest.

	content_category text	post_count bigint	avg_reach numeric	avg_engagement numeric	retention_status text
1	Comedy	2950	1031709	14.40	Moderate
2	Fitness	3004	1007764	13.95	Moderate
3	Beauty	2953	997565	15.66	Moderate
4	Fashion	3034	995999	14.52	Moderate
5	Technology	3025	994956	13.93	Moderate
6	Food	3010	994634	13.99	Moderate
7	Music	3003	989516	14.49	Moderate
8	Photography	3035	988563	14.82	Moderate
9	Travel	2968	988034	13.76	Moderate
10	Lifestyle	3017	977605	14.69	Moderate

Q3. Test if "Shares" or "Saves" are more strongly linked to new follower acquisition.

	share_category text	save_category text	avg_followers_gained numeric
1	Low Shares	High Saves	504
2	High Shares	Low Saves	502
3	High Shares	High Saves	502
4	Low Shares	Low Saves	501

Q4. Identify the "sweet spot" for hashtag counts to maximize reach without appearing spammy.

	hashtag_density text	avg_engagement_rate numeric	avg_reach numeric
1	No Hashtags	18.61	966835
2	High (16-30)	14.58	988282
3	Medium (6-15)	14.11	1014122
4	Low (1-5)	13.69	992344

Q5. Determine which discovery channel (Explore, Home Feed, etc.) provides the most "sticky" or engaged audience.

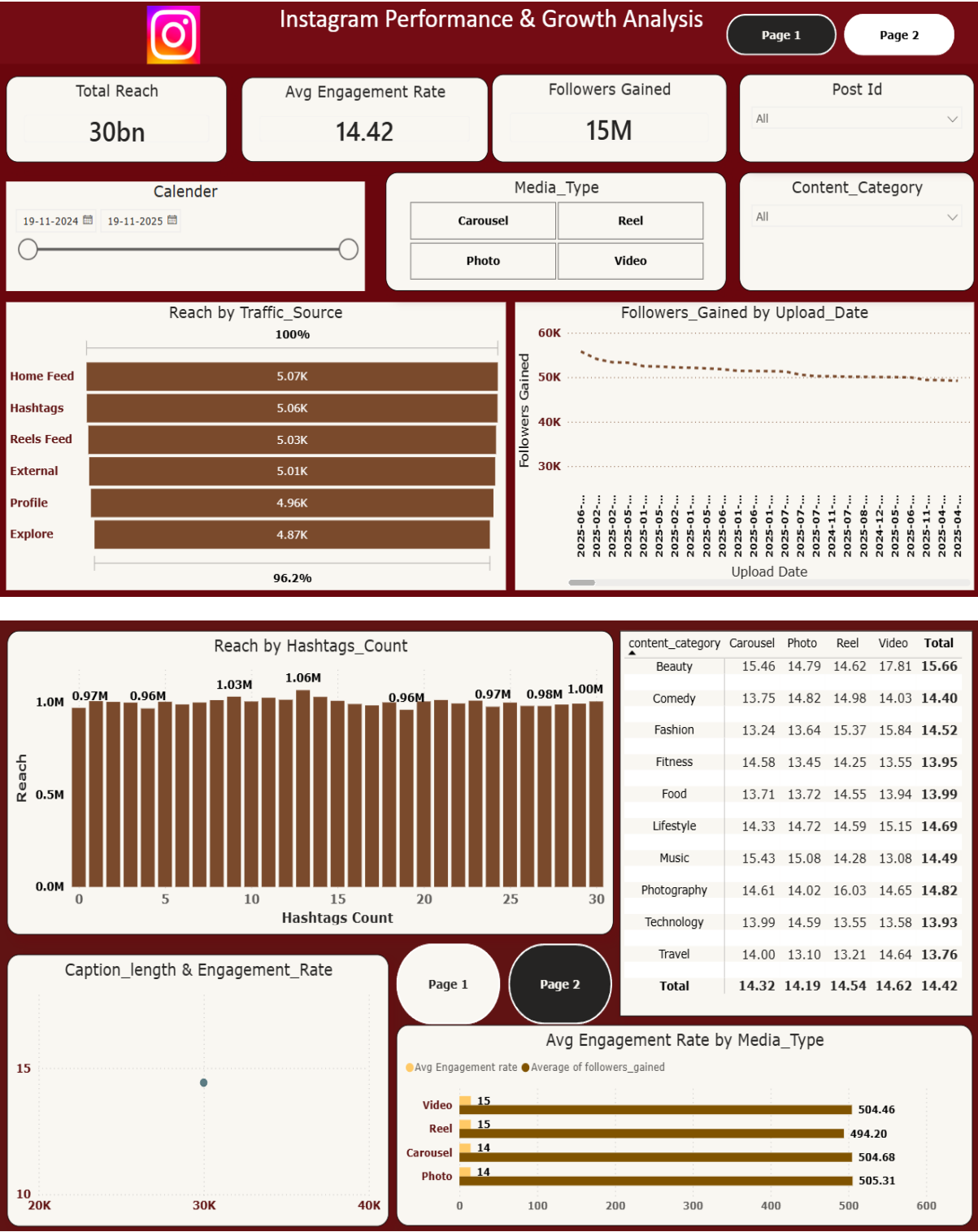
	traffic_source text	avg_impressions numeric	avg_engagement numeric	avg_followers_gained numeric
1	Profile	1264700	14.32	511
2	Home Feed	1245265	14.71	503
3	External	1249875	13.91	503
4	Explore	1229493	14.80	503
5	Reels Feed	1244929	14.52	502
6	Hashtags	1245496	14.27	492

Q6. Provide actionable creative guidelines for content creators on the optimal caption length.

	caption_type text	avg_engagement numeric	avg_reach numeric
1	Long (>150 ch...	14.48	996494
2	Short (<50 char...	13.91	1003103
3	Medium (50-15...	13.44	994764

## 5. Power BI Dashboard.

- **KPI Header:** Cards for **Total Reach**, **Follower Growth**, and **Avg Engagement**.
- **Retention Matrix:** A heatmap crossing content\_category and media\_type to flag "High Risk" segments (engagement < 10%).
- **Discovery Funnel:** Maps traffic\_source (Explore vs. Home) to see where the audience is found.
- **Technical Sweet Spots:**
  - **Hashtags:** Column chart showing peak reach at **6–15 hashtags**.
  - **Captions:** Scatter plot mapping length vs. engagement to find the creative "drop-off" point.



Reach by Hashtags\_Count

1.0M

0.97M

0.96M

1.03M

1.06M

0.96M

0.97M

0.98M

1.00M

Reach

0.0M

0.5M

Hashtags Count

Caption\_length & Engagement\_Rate

15

10

20K

30K

40K

Avg Engagement Rate by Media\_Type

Avg Engagement rate

Average of followers\_gained

Video15504.46

Reel15494.20

Carousel14504.68

Photo14505.31

content_category	Carousel	Photo	Reel	Video	Total
Beauty	15.46	14.79	14.62	17.81	15.66
Comedy	13.75	14.82	14.98	14.03	14.40
Fashion	13.24	13.64	15.37	15.84	14.52
Fitness	14.58	13.45	14.25	13.55	13.95
Food	13.71	13.72	14.55	13.94	13.99
Lifestyle	14.33	14.72	14.59	15.15	14.69
Music	15.43	15.08	14.28	13.08	14.49
Photography	14.61	14.02	16.03	14.65	14.82
Technology	13.99	14.59	13.55	13.58	13.93
Travel	14.00	13.10	13.21	14.64	13.76
Total	14.32	14.19	14.54	14.62	14.42

6. Business Recommendations

- **Format Pivot:** Prioritize **Videos and Reels** over static images. These dynamic formats show the highest "stickiness" and are the most effective at converting discovery into followers.
- **Hashtag Optimization:** Adhere to a strict **6–15 hashtag limit**. Data indicates this is the "sweet spot" for maximizing reach without suffering algorithmic penalties for spam.
- **Caption Brevity:** Limit captions to **under 150 characters**. Engagement rates significantly drop off with longer text, suggesting that concise messaging better retains audience attention.
- **Focus on Shareability:** Shift the primary creative KPI from "Likes" to "**Shares**" and "**Saves**." These behaviors have the strongest statistical correlation with account growth and community expansion.
- **Targeted Creative Intervention:** Use the **Retention Risk Matrix** to identify categories like "Technology" where reach is high but engagement is low. These segments require immediate "hook" optimization to prevent viewer drop-off.