

Instagram Post Scraper (Handover Manual & Walkthrough)

This document provides a comprehensive overview and user guide for the Instagram Post Scraper system. It is designed to assist marketing agencies in deriving social media strategies by automating the scraping, analysis, and organization of Instagram data.

1- System Overview

This system automates the process of scraping organic Instagram posts and posts from targeted accounts, extracting content and engagement data, enriching the data with AI analysis, and storing everything in Airtable for structured access. It is designed to help companies derive their social media marketing strategy, understand target audiences, and perform R&D based on public perception.

The system is built using n8n workflows, Apify actors, and Airtable as the database and user interface.

2- Core Components

- **Airtable Base**
 - Tables: Search Requests, Scraped Content.
 - Acts as both the input queue for search requests and the storage for scraped results and AI-driven insights.
 - **Apify Actors**
 - Instagram Scraper: Scrapes posts, metadata, and user profiles.
 - Instagram Comment Scraper: Collects comments on selected posts.
 - **n8n Workflow**
 - Orchestrates the scraping, enrichment, and storage processes.
 - Runs on demand via a webhook trigger.
 - **AI Modules**
 - Content Analysis: Analyzes images and videos to identify key insights, visual hooks, undeniable proof, and overall themes.
 - Comments Analysis: Summarizes comment sections, identifies pain points and common questions, and determines overall sentiment.
 - Post Review: Provides snippets on post strengths, areas for improvement, and key learnings.
-

3- Workflow Flowchart

- Airtable Search Request submitted via form →
 - n8n webhook triggers →
 - n8n picks request →
 - Apify Instagram Scraper fetches posts →
 - Posts stored in Scraped Content table →
 - (User can then toggle on analysis) →
 - Apify Comment Scraper fetches comments OR Gemini/GPT analyzes content →
 - AI analysis enriches both posts & comments →
 - Results are available in Airtable for review.
-

4- Airtable Tables

4.1 Search Requests (tblujKxHFWJiIOTYt)

- Fields: Search Type, Keyword, Instagram Account, Number of Posts, Duration, Run Flag.
- Role: Serves as the input queue where users submit requests to be processed by the n8n workflow.

4.2 Scraped Content (tblj4yZDgccxh3pjs)

- Fields: Author, Followers, Timestamp, Reactions, Comments, Shares, Virality Score, and AI-generated insights (Key Insights, Visual Hook, etc.).
 - Role: Stores all scraped Instagram post metadata and the results of the AI content and video analysis. Linked to the Scraped Comments table.
-

5- API & Authentication

- Airtable API: REST + JSON.
 - Auth: Bearer Token.
 - Apify API: REST + JSON.
 - Auth: API Token.
 - LLM APIs (e.g., Gemini, GPT): REST + JSON.
 - Auth: API Key.
 - Standard HTTP codes are used for error handling across all APIs.
-

6- Limitations & Error Handling

6.1 System Limitations

- AI-Generated Data: All AI analysis is produced by LLMs (e.g., GPT/Gemini). Outputs may not always be contextually accurate, so human review is mandatory.

- Dependency on Scraping Platform (Apify): The data quality, completeness, and accuracy of parameters like dates are dependent on Apify's scraping capabilities. Changes in Instagram's structure may break the scraper.
- Execution Limitations: Only one webhook execution can run at a time. Each run can perform only one type of operation: Search content, Analyze video, or Analyze comments.
- Account Search Requirement: For account-level searches, the exact username must be entered. The account "display name" will not work.
- Comments Analysis Threshold: Comments analysis will only be triggered if a post has at least 10 comments. Posts with fewer comments will not be analyzed.
- Video Analysis Latency: Video analysis relies on the Gemini LLM, which can be slow and may have a queue. While wait nodes are used, users may need to retry if no output is returned.
- Keyword vs. Hashtag Input: The scraper does not distinguish between keywords and hashtags. For example, "Nike" and "#Nike" will return the same results.
- Location-Based Search: Meta's algorithm does not share regions for posts, so location-based searches are not supported. A search for "Nike" will show posts from the region where the search is initiated.

6.2 Error Handling

- 2xx: Success.
- 4xx: User errors (e.g., invalid request, missing fields).
- 5xx: Server errors (retry logic in n8n is recommended).

7- Virality Score Formula

The scoring formula ranks Instagram posts by a custom Virality Score, calculated in Airtable.

- Formula inputs: Creator's follower count (tiered), engagement rate (reactions + comments + shares ÷ views), and recency (boost for recent posts).
- Example structure (pseudo):

$$\text{ViralityScore} = (\text{EngagementRate} * 0.6) + (\text{FollowerTierScore} * 0.3) + (\text{RecencyBoost} * 0.1)$$

This ensures high-quality, engaging content is prioritized, regardless of the creator's follower count.

8- AI Prompts (for OpenAI or Similar LLM)

- Content Analysis Prompt: "You are given an Instagram post's caption, engagement stats, and content. Analyze the post to identify the key insights, its visual hook (what makes it a scroll stopper), any undeniable proof, and the overall theme of the post."

- Comment Analysis Prompt: "You are given comments from an Instagram post. Summarize the overall comments, identify common pain points, list the most frequently asked questions, and determine the overall sentiment in the comment section."
- Post Review Prompt: "Review the provided post. What are its strengths and areas for improvement? What are the key takeaways or lessons that can be learned from this post?"

n8n Workflow Structure

- Trigger Node: Webhooks.
- Airtable Nodes: Fetch search requests, update status, and save results.
- Apify Nodes: Run the Instagram Scraper and Comment Scraper.
- Gemini/OpenAI Nodes: Analyze content and comments.
- HTTP Request Node: Used for communication between nodes and external services.
- Wait Node: Used to enforce delays, particularly for LLM analysis.
- Filter Node: Used to check conditions, such as the minimum comment threshold before analysis.

9- Maintenance Checklist

- Rotate Airtable and Apify tokens regularly.
- Monitor Apify actor changes (API updates may break scrapers).
- Review n8n logs weekly for failed runs.
- Clear old requests from the Search Requests table monthly.

10- Troubleshooting

- No posts scraped? Check if the Instagram account is private or if the Apify actor quota has expired.
- Comments missing? Verify the post has at least 10 comments for analysis to trigger.
- Rate limit errors? Add a Wait node in n8n between Airtable writes.
- No AI analysis output? Check the n8n execution log for errors. The LLM may be busy; try again later.

11- Future Improvements

- Integrate with data visualization tools (e.g., Google Data Studio).
- Expand the number of comments to be analyzed with batching.
- Add auto-tagging of posts by content theme.

Appendix

- n8n Workflow JSON (full export):
<https://dgtlbase.app.n8n.cloud/workflow/J9YrpKPB4G1VO8Fi>
- Screenshots of Airtable schema <https://airtable.com/appyo8jFcTlpZfc8c/api/docs>
- Prompt library for all AI modules: See Section 8 and the n8n workflow.
- Sample JS snippets for webhook automation within Airtable:

JavaScript

```
let table = base.getTable("Search Requests");
let record = await table.selectRecordAsync(recordId);

await fetch("YOUR_N8N_WEBHOOK_URL_HERE", {
  method: "POST",
  headers: { "Content-Type": "application/json" },
  body: JSON.stringify(record)
});

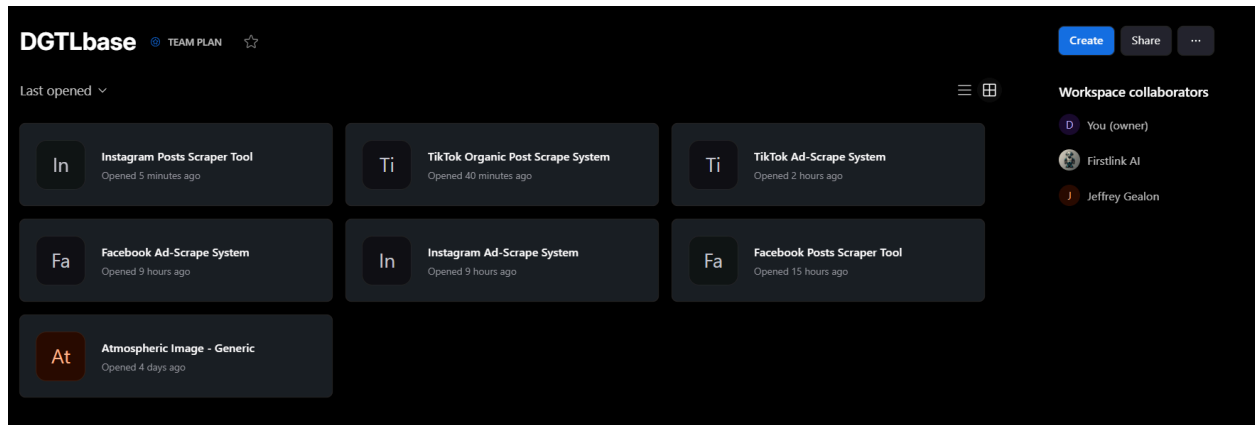
console.log("Sent to n8n:", record);
```

User Guide - Detailed Walkthrough

This guide provides a step-by-step walkthrough for team members on how to use the system.

A - Access and Roles

- Who needs access: 1) Marketing editors (Airtable Editor), 2) Ops engineer (n8n access), 3) Data reviewer (Airtable Viewer+comments).
- Action: Confirm access before running searches. Ensure required Airtable permissions are set.



B - Create a Keyword Search (step-by-step)

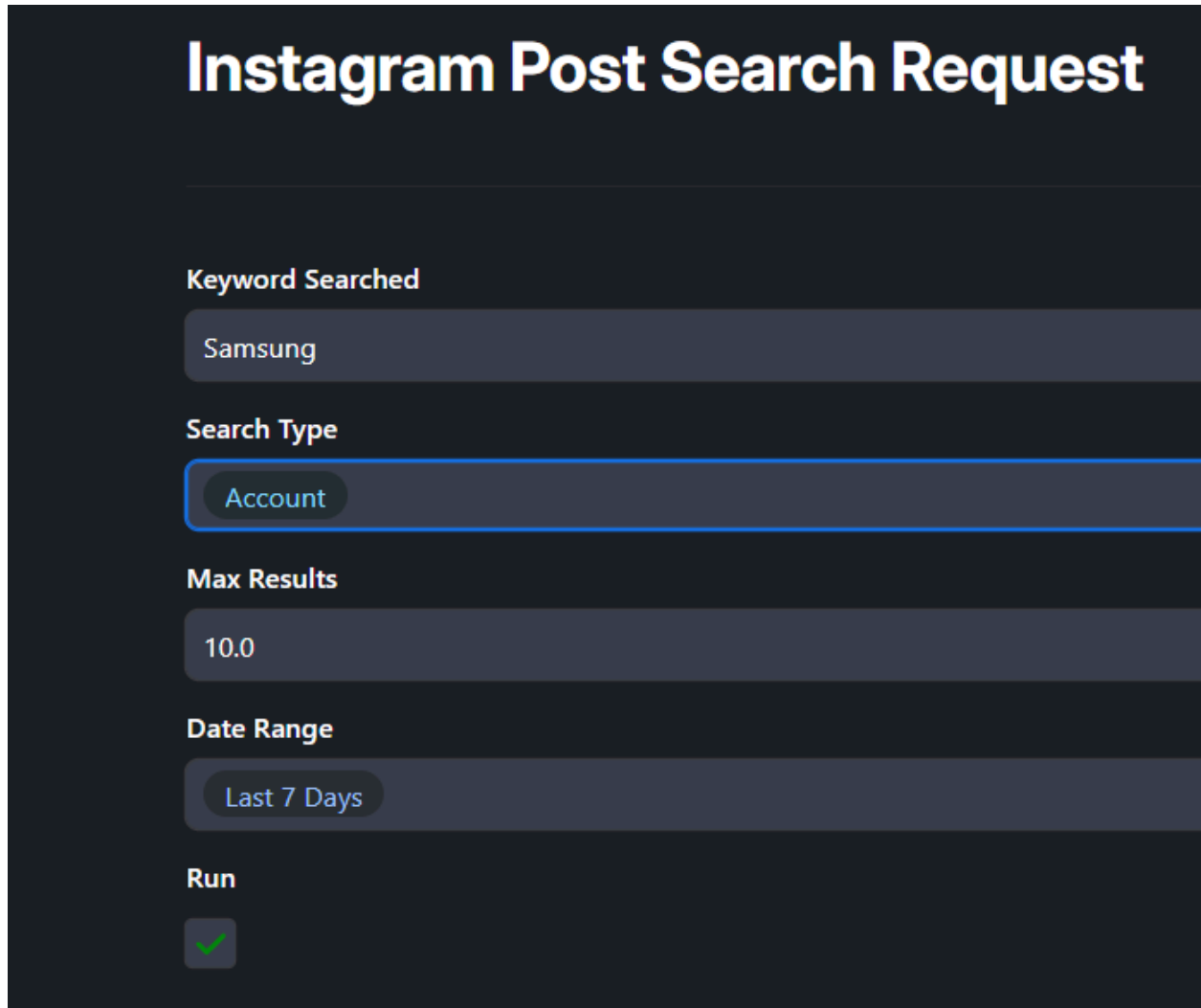
1. Open the Instagram search request form interface in Airtable.
2. Populate fields:
 - Search Type: Keyword Search
 - Keyword: "Samsung"
 - Number of Posts: 10
 - Duration: Last 7 Days
 - Run: TRUE
- Tips:
 - Avoid trailing spaces in keyword fields.
 - Use short keyword lists to reduce noise; run multiple targeted searches for variants.

The screenshot shows the "Instagram Post Search Request" form. On the left is a sidebar with navigation links: "Search Request History", "Scraped Content", "Post Content Analysis", "Comments Analysis", and "Ad-Review". The main form area contains the following fields:

- Keyword Searched**: A text input field containing "Samsung".
- Search Type**: A dropdown menu with "Keyword" selected.
- Max Results**: A text input field containing "10.0".
- Date Range**: A dropdown menu with "Last 7 Days" selected.
- Run**: A button with a green play icon.

C - Create an Account Search (step-by-step)

1. Choose Account Search in the `Search Type` field.
2. Enter the exact Instagram username (e.g., `Samsung`) in the `Instagram Account` field. Do not use the display name.
3. Set the `Number of Posts` and `Run` to TRUE.



The screenshot shows a dark-themed form titled "Instagram Post Search Request". It contains several input fields and a checkbox, all with a dark blue background and light blue text. The fields are: "Keyword Searched" with the value "Samsung", "Search Type" with the value "Account" (highlighted with a blue border), "Max Results" with the value "10.0", and "Date Range" with the value "Last 7 Days". At the bottom, there is a "Run" label and a checkbox that is checked, indicated by a green checkmark icon.

Instagram Post Search Request

Keyword Searched
Samsung

Search Type
Account

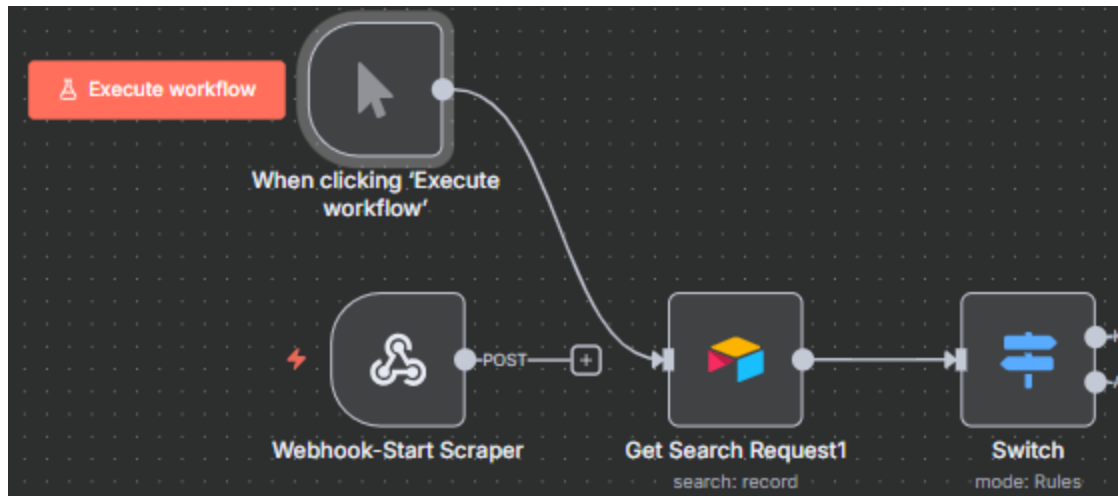
Max Results
10.0

Date Range
Last 7 Days

Run
☒

D - Run the Workflow (manual)

- If you need to force a run, open the n8n workflow.
- Click Execute Workflow or trigger the webhook (if configured).
- Action: Observe the execution status in n8n to ensure it is running correctly.



Personal / Instagram Scraper + Add tag Active ●

Editor Executions Evaluations

Executions No active executions ⓘ

☒ Auto refresh

Aug 30, 06:32:04 Succeeded in 532ms	
Aug 30, 03:57:58 Succeeded in 676ms	
Aug 30, 03:57:18 Succeeded in 22.351s	

Aug 30, 06:32:04
Succeeded in 532ms | ID#6196

E - Monitor Apify (optional)

- In the Apify console, use the run ID to inspect the raw dataset output, subtitles, or errors.
- Action: Check Apify run details and the dataset to verify returned items.

Instagram Hashtag Scraper - Run
< >

✓ Succeeded Actor succeeded with 5 results in the dataset

RESULTS	REQUESTS	PRICE	STARTED	DURATION
5	1 of 1 handled	\$0.0115	2025-08-30 02:45	10 s

Help Apify improve your experience with Instagram Hashtag Scraper
We need your permission to share all past and future runs of this Actor with the developer. This will help the developer improve the actor. You can revoke your consent anytime in [Settings > Privacy](#).

Yes, I agree
No, thanks

Output 5 Log Input Storage Live view Triggered integrations 0

Overview All fields

#	Text caption	Author ownerFullName	Author Username ownerUsername	Post URL url
1	Kids today will never know the pleasure of watching that 8-bit dunk cutscene 🏀 🎮 ...	John Toma ⚡ Gen X Nostalgia	that_80s_dude	https://www.instagram.com/p/DN58tZKk

F - Review Scrapped Content

1. Open the Scrapped Content table or the **Scrapped Content** interface in Airtable.
2. Open a record to confirm the data: **Caption**, **Views**, **Likes**, **Comments**, **Shares**, etc.
3. Look for the **Analyze Content** and **Analyze Comments** toggles.
 - Reviewer tips:
 - If a field is blank, check if the data was available for that post.
 - Compare the output to the source post in the Apify dataset if data looks missing or odd.

Instagram Post Scraper Tool > Scrapped Content									Group	Filter	Sort	Q
Search Keyword	Instagram User Name	Post Type	Post URL	Followers	Views	Likes	Caption	Duration	Published Date			
Apple	apple	Video	https://www.instagram.com/p/DN58tZKk	34,261,671.0	225,739	36,910	Dressed to impress...	–	2025-08-14 12:01am			
Apple	apple	Video	https://www.instagram.com/p/DN58tZKk	34,261,671.0	298,905	39,473	What are you doing down...	–	2025-08-12 12:07am			

G - Request Analysis (content/comments)

- Toggle **Analyze Content** or **Analyze Comments** to TRUE on a record to trigger the analysis.
- Note: For comments analysis, the post must have at least 10 comments.
- Action: Monitor n8n runs for AI processing completion.
- Timing guidance: LLM jobs may be queued; expect variable completion times depending on server load.

Instagram Post Scraper Tool Scraped Content										Group	Filter	Sort	Q	...
Name	Post Type	Post URL	Followers	Views	Likes	Caption	Duration	Published Date	Analyze Comment?	Analyze Content?				
nti	Sidecar	https://www.instagram.co...	1,348.0	–	3	As vezes não conseguimos...	–	2025-08-30 1:06am						
	Image	https://www.instagram.co...	36.0	–	–	🔥 আপনার ঘুকের স্বপ্নে সে...	–	2025-08-30 12:46am						
	Sidecar	https://www.instagram.co...	3,421.0	–	–	THANK YOU SO MUCH...	–	2025-08-30 12:24am						
	Image	https://www.instagram.co...	239.0	–	3	#cetaphil #crèmesolaire...	–	2025-08-30 12:27am						
	Sidecar	https://www.instagram.co...	10,114.0	–	–	🌸 Cleanse, Dream, Repeat...	–	2025-08-30 12:00am						
erine	Image	https://www.instagram.co...	524.0	–	–	🍷 بين احسن المرطبات ...	–	2025-08-29 11:04pm						
up	Image	https://www.instagram.co...	694.0	–	–	Sauve More To It Cosmetic...	–	2025-08-29 10:17pm						

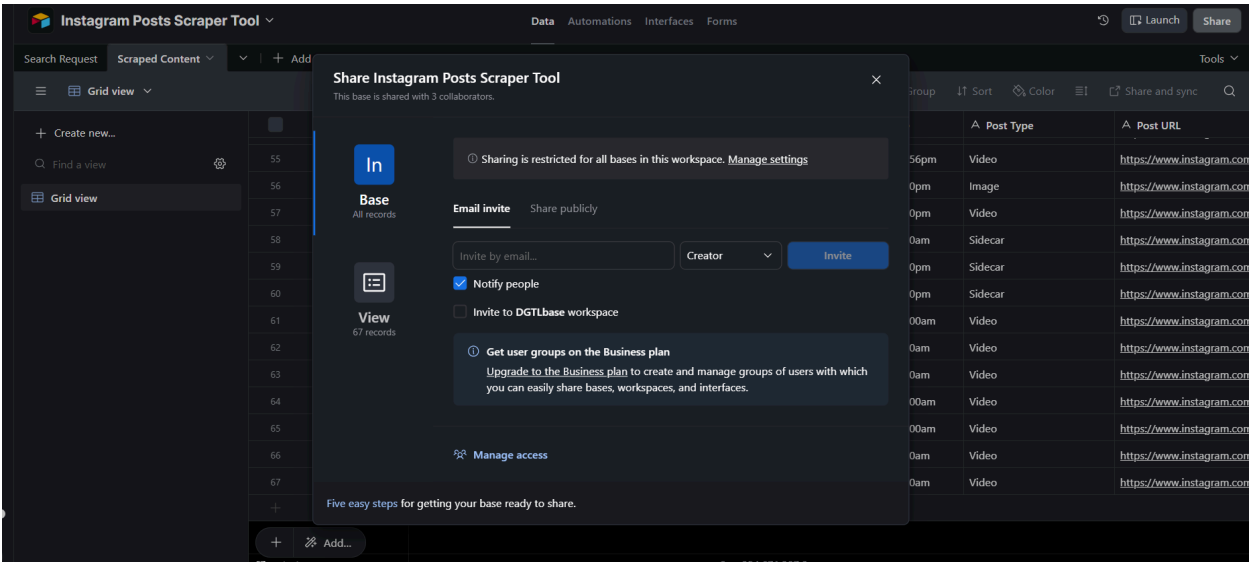
H - Interpret AI Outputs and Take Action

- Key fields produced by AI:
 - Virality Score: Use to prioritize content.
 - Key Insights: Summary of the post's value.
 - Visual Hook: What makes the post a "scroll stopper."
 - Undeniable Proof: Credible elements of the post.
 - Comments_Sentiment / Comments_Summary: Insights from the comment section.
 - Post Review: Strengths, weaknesses, and key takeaways.
- Suggested next steps based on outputs:
 - High Virality + positive insights: Add to a content replication pipeline or creative brief.
 - Negative sentiment: Route to product or support teams.
 - High number of questions: Collect into an FAQ.

Instagram Post Scraper Tool - Comments Analysis			
Search Keyword	Instagram User Name	Post URL	Comments Pain Point
Apple	apple	https://www.instagram.co...	Many users praise the vide
NBA	stevebrada	https://www.instagram.co...	LeBron James' career is...
Samsung	samsung	https://www.instagram.co...	Customers experience lon.
Samsung	samsung	https://www.instagram.co...	Display and 'green line'...
Samsung	samsung	https://www.instagram.co...	Warranties being denied f

Post URL	https://www.instagram.com/p/DNg9w1EIXSL/
Comments Summary	Most comments express frustration over Samsung's product quality, defective devices, poor customer service, and high repair costs, especially regarding phones and TVs. A few users share positive or neutral suggestions, but the overall tone is critical of support and reliability.
FAQ in Comments	Can Samsung address the 'green line' and display issues?. Why is customer support unresponsive or denying warranty service?. Does Samsung provide home sales visits (Delhi query)?. Why are trade-in values not honored as promised?
Comments Pain Point	Display and 'green line' issues recurring after updates or in premium devices.. High repair costs and denied warranty claims even for known manufacturing defects.. Trade-in and refund process dissatisfaction.. Poor or delayed support from customer care across regions.. Suggestions for new features and hardware customization.. Scattered positive remarks or unrelated comments.
Comments Trend	Numerous complaints about display/green line issues on Samsung phones.. Negative sentiment about after-sales service, repair costs, warranty denials, and delayed customer response.. Growing distrust and calls to switch brands or boycott Samsung.. Some feature requests and international feedback regarding localized functions.
Comments Sentiment	negative

I - Exporting / Sharing



J - Reviewer QA Checklist (short)

- Key Insights are accurate to the content and visuals.
- Factual claims are supported by "Undeniable Proof."
- Comments sentiment is consistent with a sample of comments.