

TikTok User Data Analysis

Project Report

Motivation

My motivation behind this project comes from a growing interest in understanding personal digital footprints and the insights they can offer. In an age where social media platforms like TikTok play a significant role in daily life, analyzing my interactions with these platforms can reveal patterns and behaviors that might otherwise go unnoticed. This project aims to provide a deeper understanding of personal usage patterns, preferences, and engagement on TikTok.

Data Source

The data for this project was sourced from TikTok, TikTok directly allows us to download our data as a machine readable .json file, which helps tremendously with the data collection, as the data is already kind of sorted. This feature gives us full access to the data we generate on the platform. The data provides comprehensive and structured representation of various activities, including watch history, likes, comments, shares, live stream views, and login details.

Data Analysis Techniques Used

The data analysis involved the following key stages and techniques:

Data Loading: The JSON file was loaded into Python for analysis, ensuring proper handling of the file's structure and encoding.

Data Parsing: Each category of data (watch sessions, comments, likes, shares, live streams, login history) was individually parsed and processed.

Statistical Analysis: Basic statistical methods were applied to compute totals, averages, and other relevant metrics.

Date-Time Analysis: Timestamps were used to determine patterns over time, such as the most active day of the week and watch session computing.

Text Analysis: Comments were analyzed for length and emoji usage.

Stages of Analysis

Watch Sessions: Analysis of total videos watched, session lengths, and active days.

Comments: Analysis of total comments, average length, and emoji usage.

Likes and Shares: Tracking of total counts, most active days, and first interactions.

Live Streams: Analysis of total views and timing of livestream interactions.

Login History: Insights into device usage, network preferences, and carrier information.

Findings

The analysis revealed interesting personal insight into my TikTok usage, without further ado, here they are:

Watchtime Analysis:

Total watch time (minutes): 12028

Watch sessions: 566

Average session length (minutes): 21

Longest watch session (minutes): 291

Earliest video watched: 2023-06-12 13:10:36

Last video watched: 2023-12-08 07:04:47

Most active weekday: Monday

Comments:

Total Comments: 110

Average Comment Length: 34 characters

Most used emoji: 🤔

Most used emoji count: 6

Likes:

Total likes: 6667

Day with most liked posts: 2021-05-14

Likes count on that day: 167

First liked video: <https://www.tiktokv.com/share/video/6825672473003003141/>

Date of first liked video: 2020-05-20 12:44:15

Shares:

Total shares: 94

Day with most shared posts: 2023-09-18

Shares count on that day: 5

First shared video: <https://www.tiktokv.com/share/video/7242030089142078746/>

Date of first shared video: 2023-06-11 01:36:02

Livestream Data:

Total Lives viewed: 29

Earliest watched live: 2023-01-04 04:33:02

Link: <https://vm.tiktok.com/ZM6Yne4dM/>

Latest watched live: 2023-09-28 13:59:43

Link: <https://vm.tiktok.com/ZM6YtTFfH/>

Data related to logins:

Times Logged In: 2169

Most Used Device Model: iPhone15,3

Most Used Network Type: Wi-Fi

Most Used Carrier: Turkcell

Comments on my findings:

First of all, whoa, that's a lot of videos watched. On top of that, its evident that TikTok only records your watch data back up to 6 months, and it looks like I watched 11 thousand videos in the last 6 months, which is 61 videos a day. Extrapolating that back to my first liked video 3 and a half years ago, that gives us around 78 thousand videos. Which is absolutely insane. I would never think I watched that many videos, but also, that's extrapolated data, I still trust it though.

And it looks like the average time I spend in a video is just above 1 minute so that means I spend about an hour on TikTok daily, which is higher than I anticipated.

Considering my average watch session is 21 minutes long, that would signal that I open the app around 3 times a day. And my longest watch session is 291 minutes long, again, whoa, I don't remember this happening, I guess I was stuck somewhere waiting for something, just doom scrolling, maybe at an airport?

And my most active day is Monday, which is interesting, I would assume it to be on the weekends but I guess I open the app more when im bored as opposed to when I have free time.

Looking at the comment data, there isn't much to see here, but when I look at the raw data, I barely remember writing those comments, it takes me back, these don't go back only 6 months. I don't really use emojis but its interesting to see the crying face is my most used one.

When we come to the like data, its also evident that this doesn't go back only 6 months, because my first liked video is in the beginning of 2021, and yes, I remember that video. And there was a day where I liked 167 videos? Looks like it was years ago but I have no clue what happened there. If we divide my total extrapolated watched video number to the number of likes I gave, we get %8.5. Which is very close to the industry standard where a video is liked %10 of the time, guess I like videos a bit less than average. Also that means our extrapolated data is somewhat true.

Looking at the share data, it looks like I shared 94 videos, but that's not right, I know I share way more, and that's right since the data only goes back 6 months. But we can still make an estimation on how many videos I share daily, and it comes to around half a video a day, which I would say is right. I only share ones I like a lot, prefer not to spam people.

Livestream data is bleak, I don't watch livestreams, these are the ones that TikTok tries to shove down my face, I just skip them, not much to see here.

And lastly looking at login data, this goes back all the way back to when I signed up, I wanted to look at where I sign in, I sometimes sign in on my pc aswell, but yeah, mostly my phone as can be seen from the data. Most used network type is Wi-Fi, which is good. And the most used carrier is turkcell, yep that's my carrier.

Limitations and Future Work Limitations

Data Scope: The analysis is limited to the data provided by TikTok, which might not capture the full spectrum of user interaction. For example, on certain types of data, it only goes back 6 months, which is way less than I'd expect to see.

Session Definition: The definition of a 'session' in watch history analysis is assumed as 1 hour between sessions, which might not perfectly align with actual usage patterns.

Personal data: I would for example like to go into data with messages exchanged with other users, but that's obviously very sensitive data.

Comparative Analysis: I would love to compare this data with others, see how it matches up with other peoples watch times, average session times and so on. The data has to be anonymized of course, without the .json file including comments and messages exchanged with other users, which it currently does.

Conclusion

This was definitely an interesting undertaking, it was fun to analyze the data and come up with these metrics. I feel like I achieved most of my goals, my only wish is that some of the data sets would go back more, and I also thought there would be data on your liked topics, like how the algorithm recommends you videos. I would have loved to see the keywords TikTok uses to recommend me videos, but still it proved to be an interesting project nonetheless and I had a lot of fun seeing how my data racked up. I was definitely shocked with some numbers, and some data ended up as I expected. I look forward to running this script again in a couple years to see how it changes.