Motivation

My Daily Habit:

I watch videos every day. It's a big part of my routine. I wanted to know more about what kinds of videos I watch most.

Different Kinds of Videos:

I watch all sorts of videos - long ones, YouTube shorts, and music videos. I was curious to see if I prefer one type over others at different times.

Looking at Changes Over Time:

I started this project to see if I watch more of one type of concent in some months than others. I thought this would be interesting to find out.

Consistency in What I Watch:

Even though I noticed I watch different things in different months, I wanted to see if there's a pattern. Like, do I always watch more music videos in Saturdays at 11 AM?

What This Means for Me:

By looking at my own video watching habits, I can learn a lot about what I like. It's also fun to see if there are any patterns in what I watch.

Data Source

Origin of the Data:

My project utilizes data directly from my personal YouTube history. This is a rich source of information that accurately reflects my viewing habits.

Data Extraction Process:

Initially, the data was in HTML format, which isn't directly suitable for analysis. To address this, I employed web scraping techniques to transform it into a more usable format, such as a CSV or Excel file.

Data Details:

The dataset includes several key pieces of information about each video I watched. This includes the video's title, its duration, the date when I watched it, and the exact time of viewing. These details are crucial for the comprehensive analysis I plan to conduct.

Data Analysis

Data Acquisition:

The first step in my analysis involved web scraping to extract my YouTube viewing history. This method provided me with raw data directly relevant to my watching habits.

Data Preparation with Pandas:

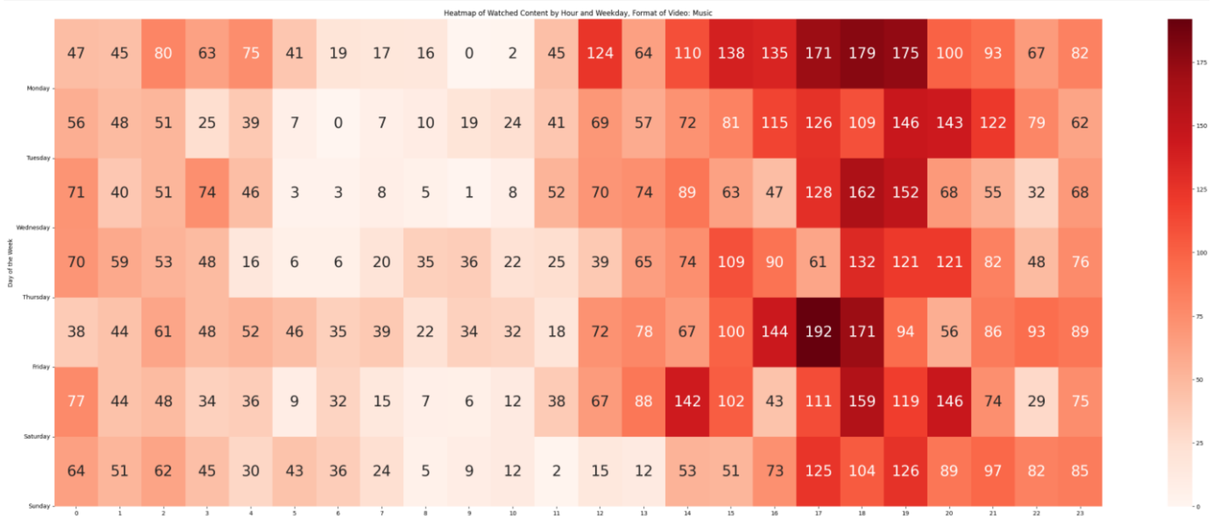
To make the data more suitable for analysis, I utilized the Pandas library in Python. This included editing existing fields and enhancing the dataset with additional information. For instance, I extracted the day of the week, the hour of the day, and the month in which each video was watched, enriching the dataset for a more detailed analysis.

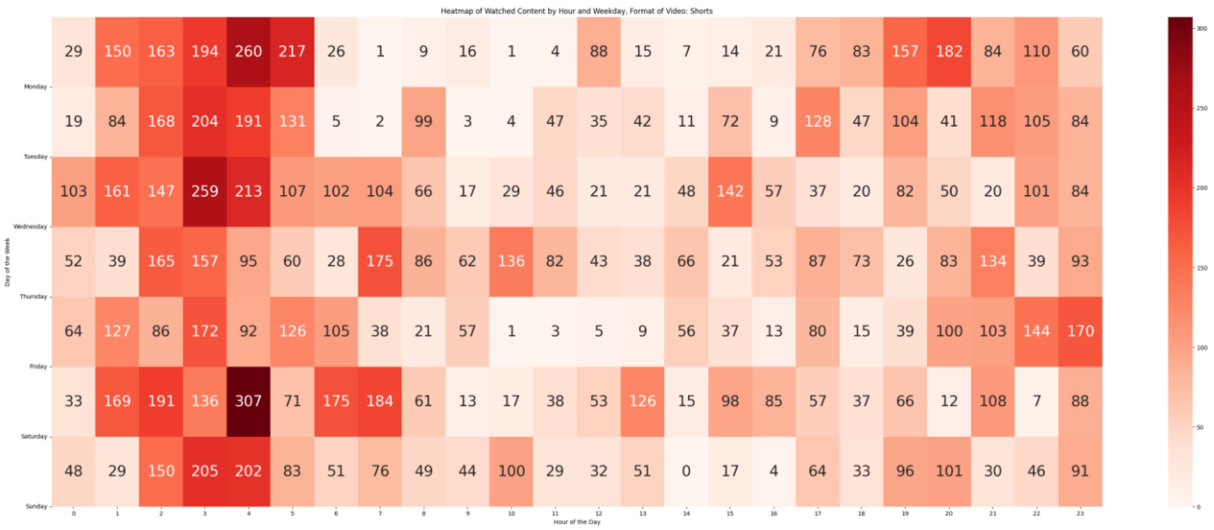
Visualization Techniques:

Instead of relying solely on numerical calculations, I employed visualization tools like Seaborn and Matplotlib. These tools helped me see the data in a more comprehensive and digestible format.

Heatmaps: I used these to identify patterns across different fields, such as the time of day and type of content.

Two heatmaps to demonstrate the difference between my Shorts, and YouTube music consumption pattern throughout 2023.

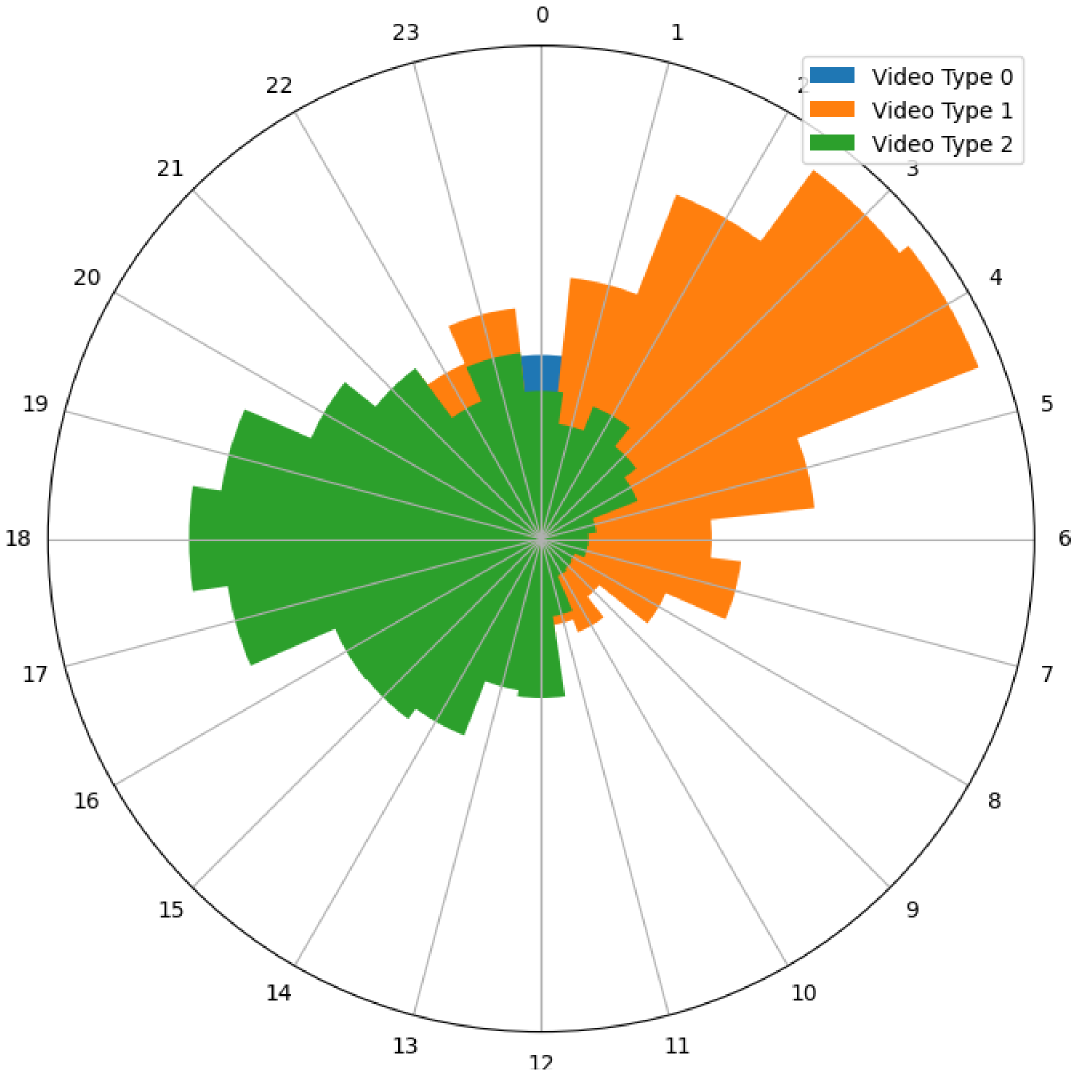




Correlation Matrix: This helped me understand the relationships between time, weekdays, and the content type.

Circular Time Chart: A unique way to visualize the connection between the time of viewing and the type of content.

0 demonstrates long videos, 1 short videos, 2 music videos.



Insights from Visualization:

The visual analysis revealed strong and predictable patterns in my video watching habits. This was a key insight, indicating regularity in my viewing preferences.

Machine Learning Application:

Despite the patterns, I noticed that the relationships in the data were not linear and correlations were moderate. To tackle this, I chose Decision Trees as my machine learning algorithm.

The goal was to predict the 'Video Type' (Long, Shorts, or Music) using 'Month', 'Hour', and 'Weekday' as inputs.

The model achieved an accuracy of around 77%.

Notably, it was particularly effective at predicting Short Videos(81%), suggesting that my consumption pattern for short videos is more consistent compared to longer formats.

It was not that effective at predicting Long Videos (66%), suggesting that Long video watching times are less consistent compared to others.

Findings: What I Learned About Myself

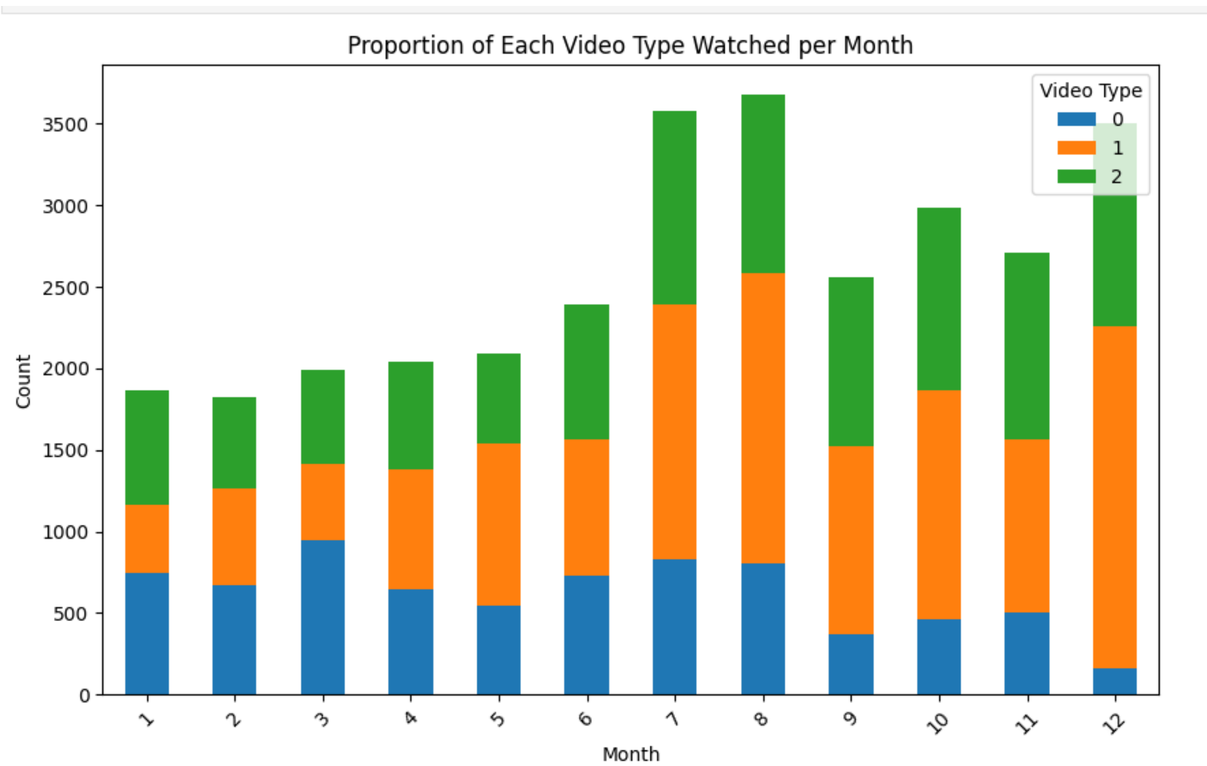
Consistency in Content Type:

One of the most striking output was the consistency in the types of videos I watched. Despite a wide variety of available content, I tended to gravitate towards certain types more regularly at specific days and hours.

The analysis showed clear patterns in my viewing habits related to the time of day and week. For example, I noticed a tendency to watch more music videos or YouTube shorts during certain hours or on specific days, suggesting a routine in my viewing behavior.

Seasonal Variations:

Interestingly, the data indicated seasonal changes in my video consumption. Certain months showed a distinct preference for specific types of content, reflecting how my interests or available time might vary throughout the year.



Insight into Personal Preferences:

This project provided me with a quantitative backing to what I intuitively knew about my preferences. It was fascinating to see how certain content types consistently captured my interest more than others.

Broader Perspective on Media Consumption:

Beyond personal insights, the project helped me understand how individual preferences can be quite nuanced and varied, offering a broader perspective on digital media consumption patterns.

Limitations and future work: What could be done better?

Limited Categorization of Long Videos:

A significant limitation is the broad categorization of long videos. Breaking down this category into more specific genres, such as education, entertainment, and news/informative content, could offer a more nuanced understanding. The current general category makes it challenging to discern distinct patterns in my long video watching habits, which are less consistent compared to other types.

Absence of Detailed Viewing Metrics:

Another limitation is the lack of certain metrics in the data. For instance, including a field like 'video watch duration percentage' would offer insights into how much of each video I actually watched. Additionally, a field indicating whether the video was watched on mobile or desktop could reveal interesting patterns in device-specific viewing habits.

Limited Time Frame of Data:

The dataset covers a specific time period and does not account for long-term changes in viewing preferences. A more extended data collection period could provide a better understanding of how my video consumption evolves over time.

Personal Stance on Project Expansion:

While these limitations highlight areas for potential improvement, I have decided not to pursue further expansion of this project. My objectives have been met, and I believe the insights gained serve the purpose I intended.