YouTube Watching History Analysis

A Data-Driven Look into Viewing Habits and Preferences

Emircan Tartar 00030810

Fall 24-25

Introduction: Motivation

- **Why did you choose this project?**
- With video content playing a central role in daily life, understanding my YouTube-watching habits provides insights into how I consume digital content.
- This analysis helps uncover patterns and trends in my behavior, providing valuable self-awareness and aiding in better time management.

Introduction: Goals

- **What are the main questions you aim to answer?**

1. **Peak Viewing Times:** What times of the day or week do I watch videos the most?
2. **Channel Preferences and Loyalty:** Which channels capture my interest, and how consistently do I return to them?

- **Why is understanding viewing habits important?**

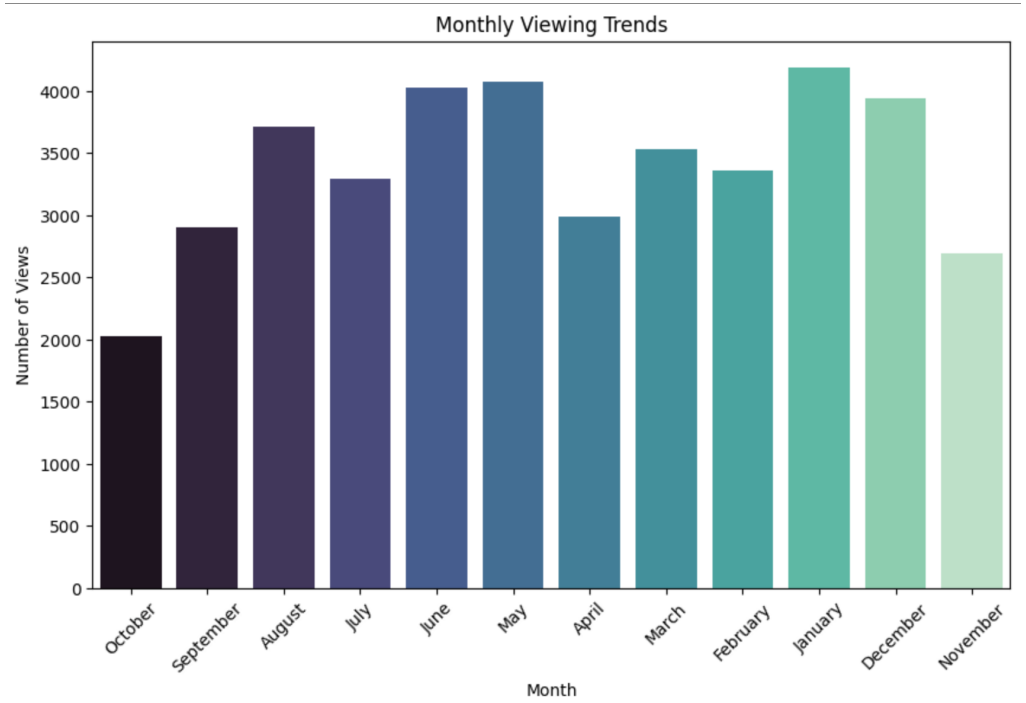
- Recognizing patterns in viewing habits can help optimize content consumption and productivity.
- It sheds light on how interests evolve over time and supports making informed decisions about media consumption.

Hypothesis

1. Viewing patterns will show a peak during evening hours and weekends, indicating leisure-time consumption.
2. I am viewing some channels more than others so that that take place most of the time that I view videos.

Dataset

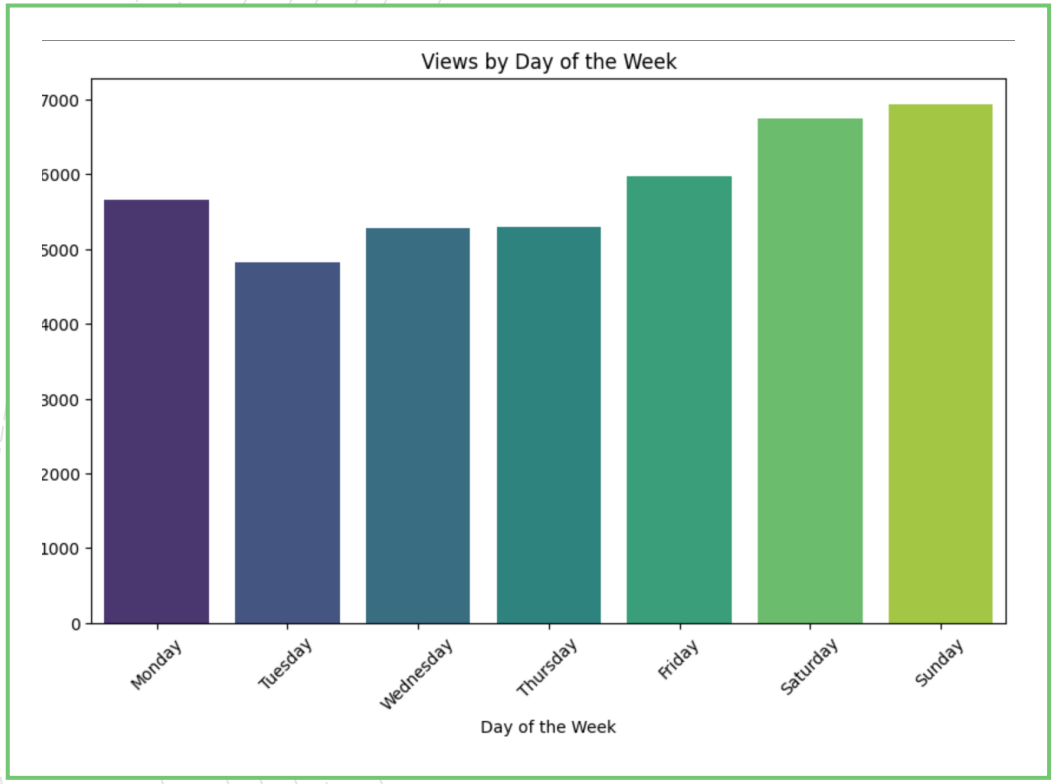
Dataset was directly taken from Google Takeout Youtube option with watch history chosen to continue further analysis.



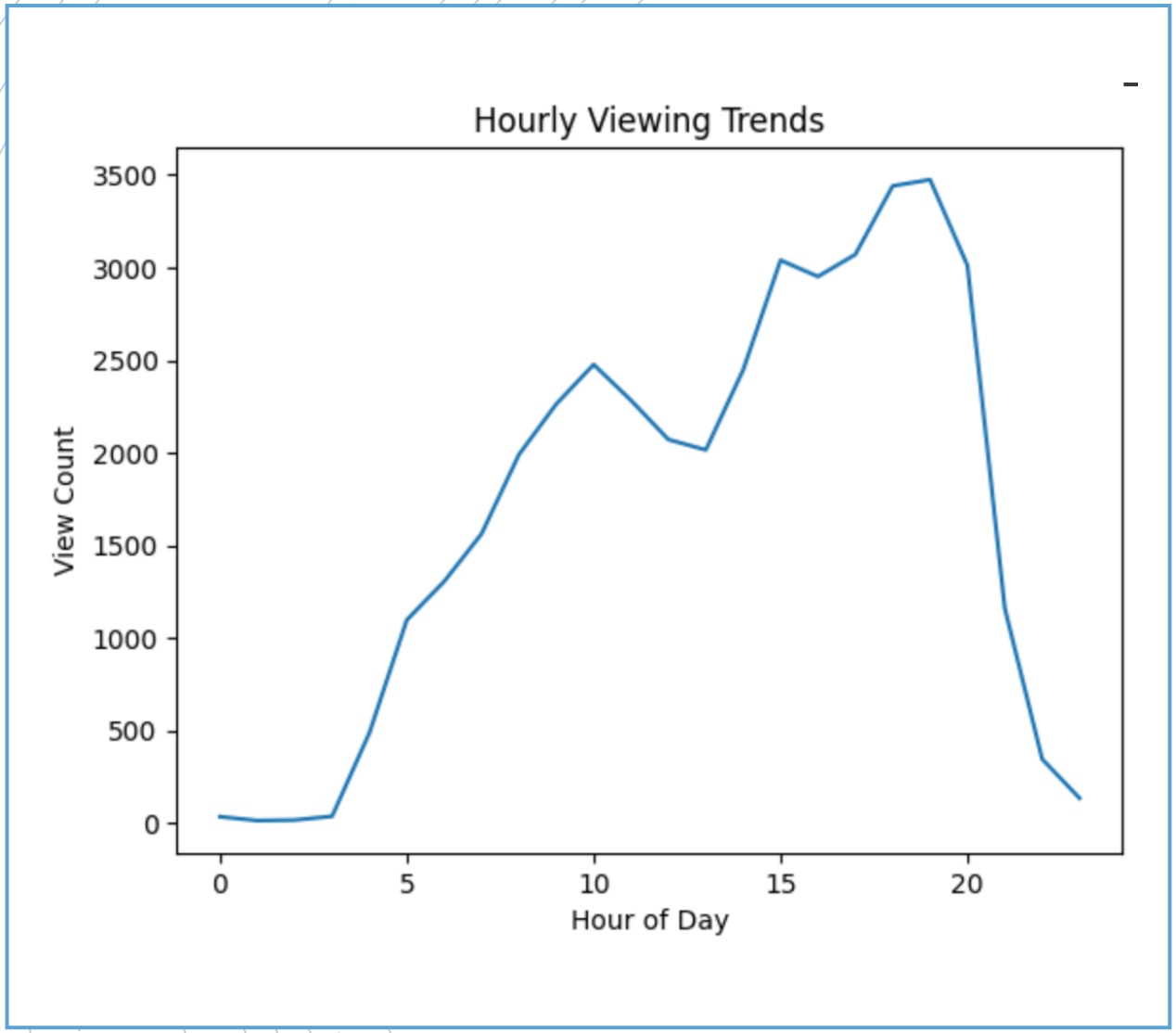
First started by doing a monthly analysis before going into details

- First founding was in the months I was more busy with school, my video views lowered; while in the months I was in school break like June- August and January-February, I was watching more video monthly.

Next step was daily analysis



- Again similar to monthly analysis, I was watching more videos on the days that there was no school.



As a last step of my EDA, I have looked in my hourly watching trend.

- **Although this graph gave me an idea about hour trends, data cleaning and more detailed analysis was needed for to have a comment about my hypothesis.**

```
[13]  1 # Extract date, hour, day of the week, and month
      2 youtube_data['date'] = youtube_data['watch_time'].dt.date
      3 youtube_data['hour'] = youtube_data['watch_time'].dt.hour
      4 youtube_data['day_of_week'] = youtube_data['watch_time'].dt.day_name()
      5 youtube_data['month'] = youtube_data['watch_time'].dt.month_name()
      6
      7 # Summarize viewing counts by day of the week
      8 views_by_day = youtube_data['day_of_week'].value_counts().reindex(
      9     ['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday']
     10 )
     11 print("Views by Day of the Week:\n", views_by_day)
     12
     13 # Summarize viewing counts by hour
     14 views_by_hour = youtube_data['hour'].value_counts().sort_index()
     15 print("Views by Hour:\n", views_by_hour)
     16
```

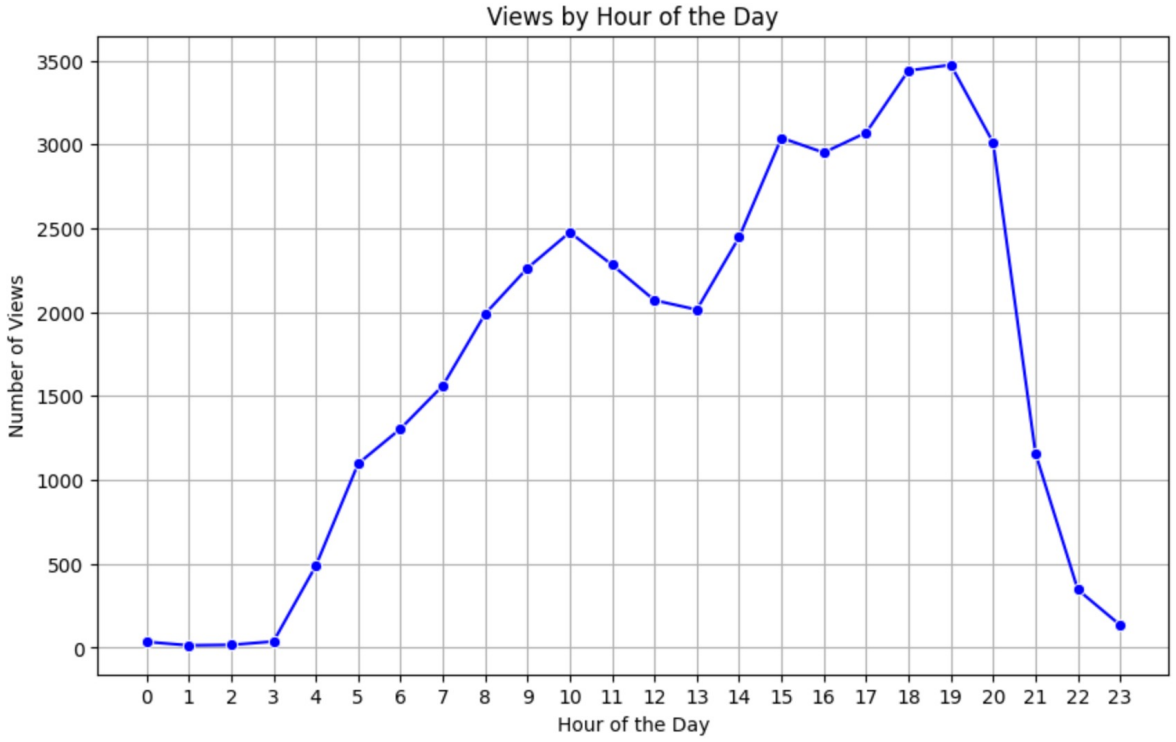
```
  1 # Check for missing values
  2 print(youtube_data.isnull().sum())
  3
  4 # Check for duplicate rows
  5 duplicate_count = youtube_data.duplicated().sum()
  6 print(f"Number of duplicate rows: {duplicate_count}")
  7
```

```
watch_time         0
video_title        0
video_url          9
channel_name    4820
dtype: int64
Number of duplicate rows: 0
```

```
[9]   1 # Convert watch_time to datetime
      2 youtube_data['watch_time'] = pd.to_datetime(youtube_data['watch_time'], errors='coerce')
      3
      4 # Check for invalid conversions
      5 print(youtube_data[youtube_data['watch_time'].isnull()])
      6
      7 # Drop rows with invalid dates (optional)
      8 youtube_data = youtube_data.dropna(subset=['watch_time'])
      9
```

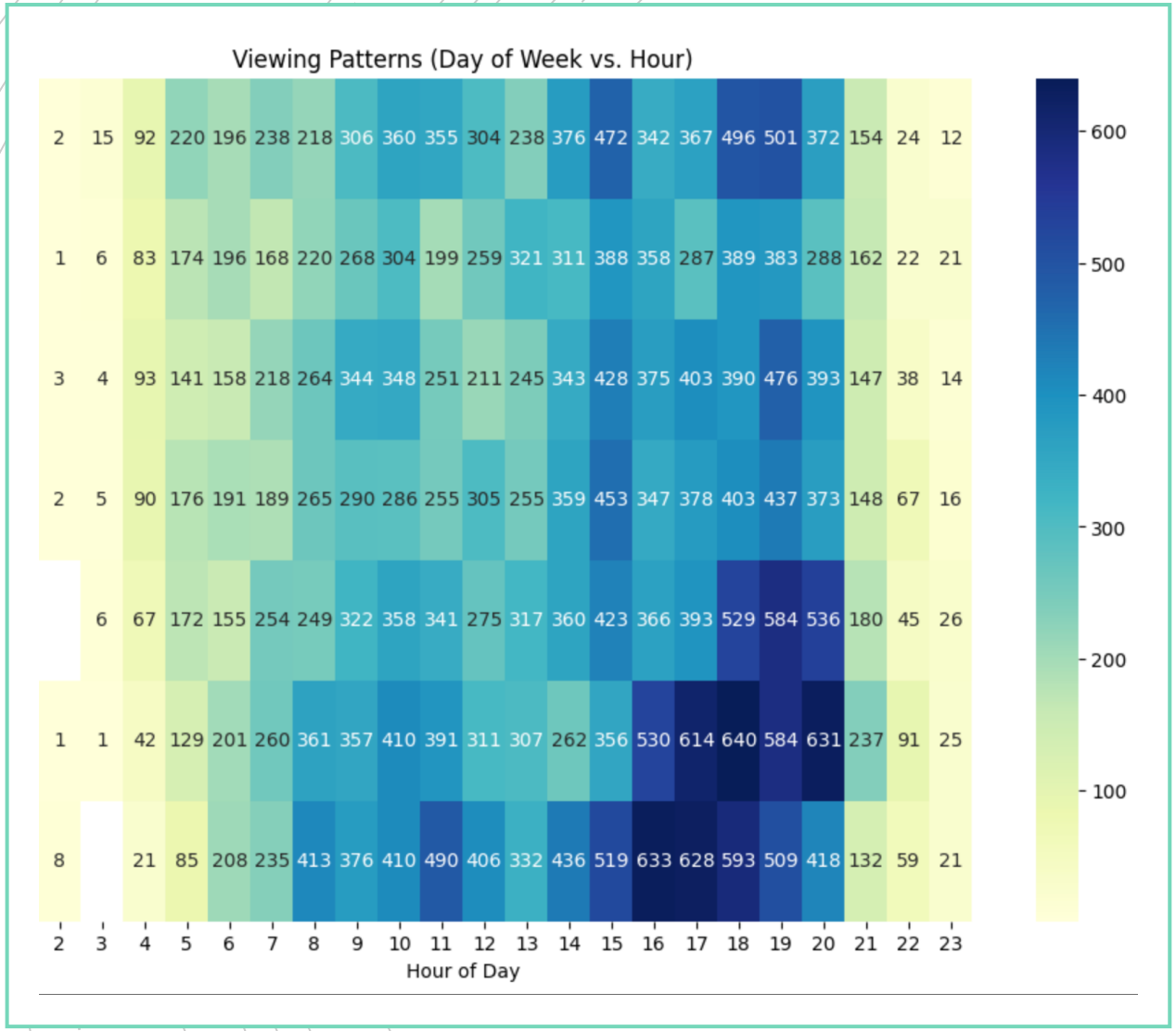
Doing the necessary adjustments

- Checked for missing values and duplicate rows, converted date and time to make it more useful and extracted these attributes from the dataset.



Detailed Analysis

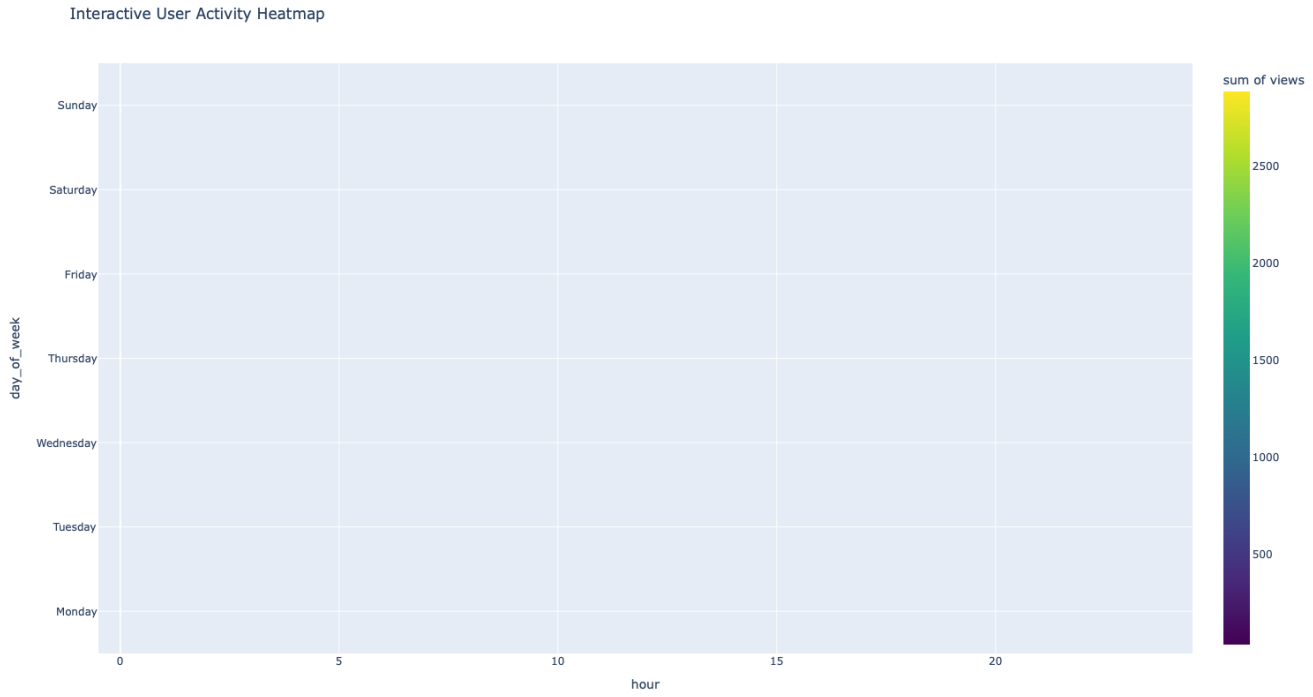
- With the cleaned dataset and more organized graphs, it was confirmed that in the hours that I was not in school, I was viewing more videos.

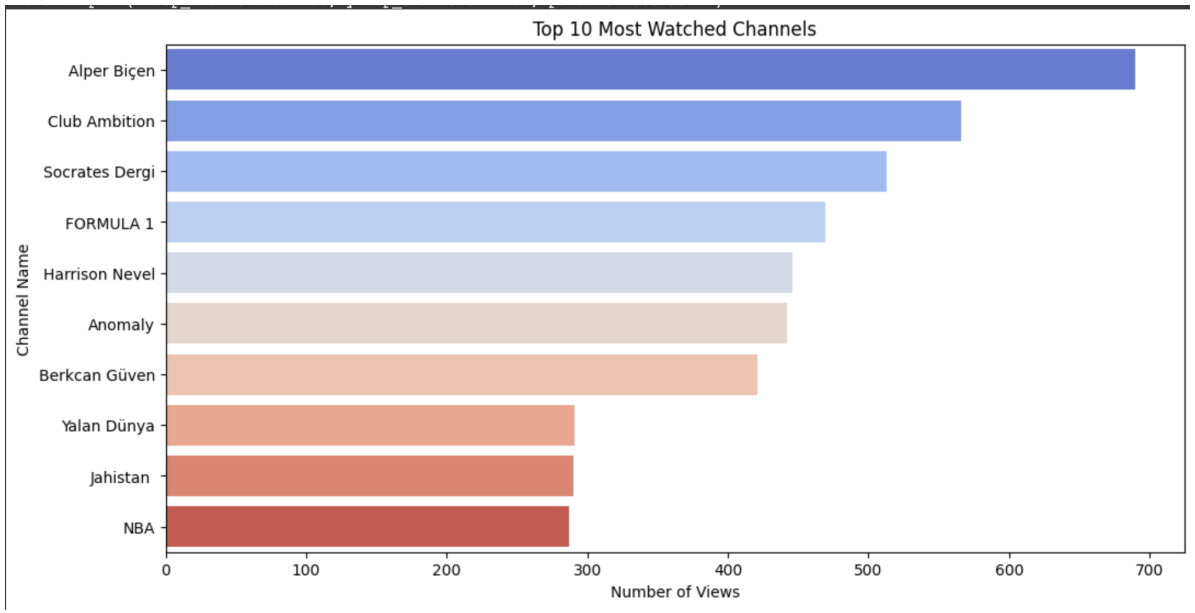


Complext Graphs for hourly and daily analysis

- With heatmaps I was able to comment on my hypothesis more as it was easier to see with colours.

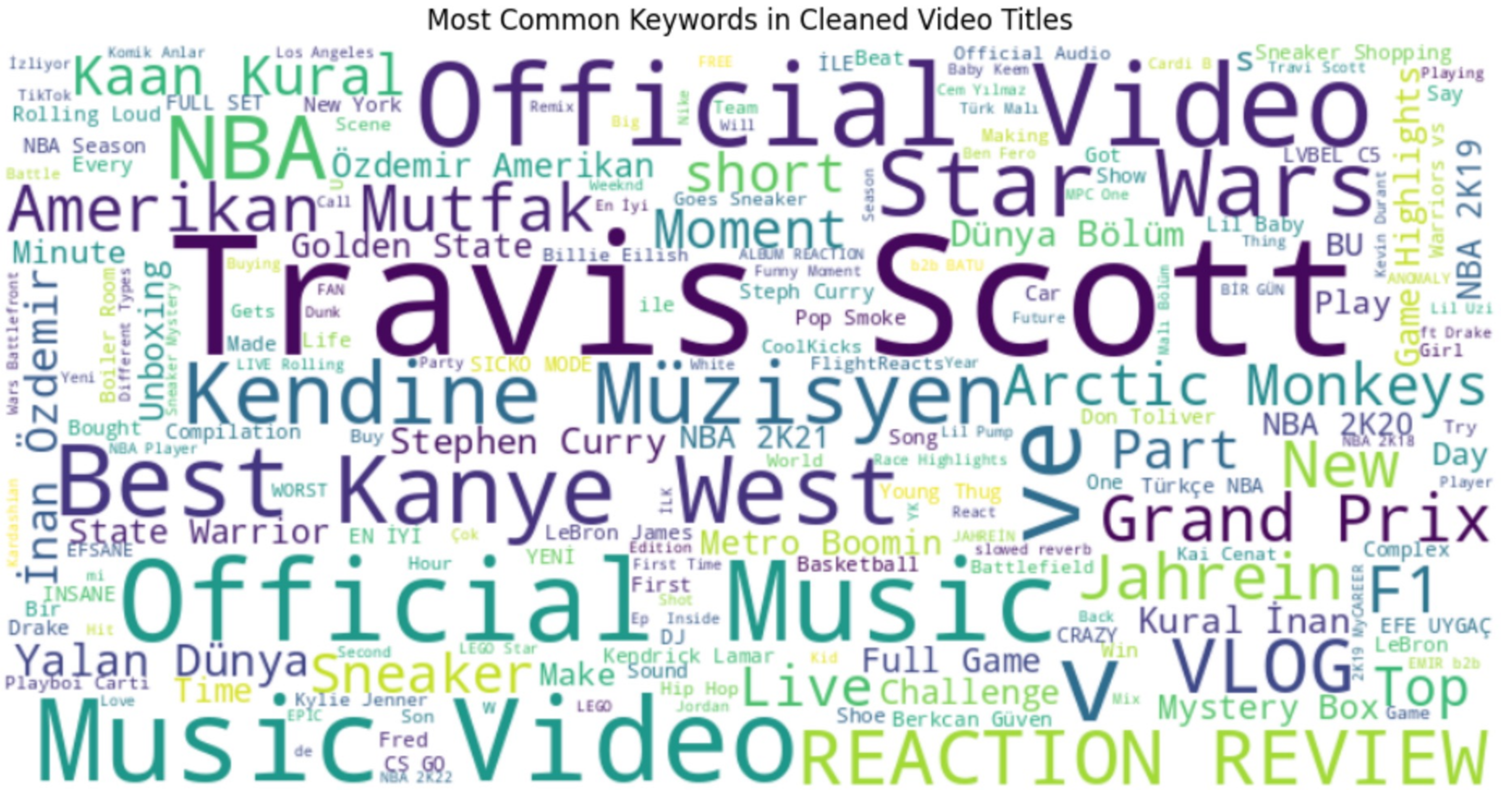
Here is an interactive graph! (html file is on github)





Finding most watched channels

- Like my hypothesis, most of my watch time was focused on some youtube channels that are on this graph.



Also here is all the keywords that are mentioned most in my cleaned and adjusted watching history!

Conclusion

- As a conclusion I have accepted my hypothesis as there is a specific pattern in my watching day and hours with also focusing on some specific channels.

Further Improvements

- Some of the data was lost because of the incognito mode in Youtube.
- These losses can be covered by machine learning techniques that will guess the watching time and watched channels.
- Also the times that my watching time peaked can be compared with my meal times, where another project can be about my eating and watching habit.