Analysis of My YouTube Viewing Habits

Uncovering patterns, trends, and insights from personal watch history

Batuhan Özdemir 28332

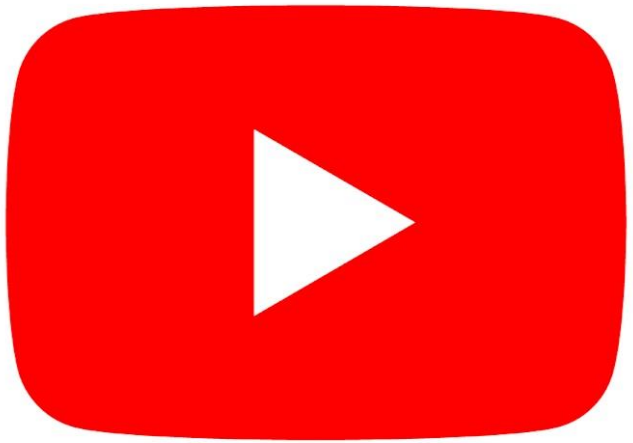
DSA210 Term Project

•**Why Analyze My YouTube Watch History?**
YouTube has become an integral part of my daily routine, offering both entertainment and learning opportunities. This project explores how I engage with the platform and uncovers meaningful patterns.

•**Objectives:**

1. Identify dominant content preferences (e.g., Music, Gaming, Education).
2. Understand viewing habits across time (hour of the day, day of the week).
3. Validate the hypothesis: *"I watch more YouTube outside of regular 9–17 working hours, especially close to midnight."*

•**End Goal:**
Gain actionable insights to optimize my digital consumption habits.



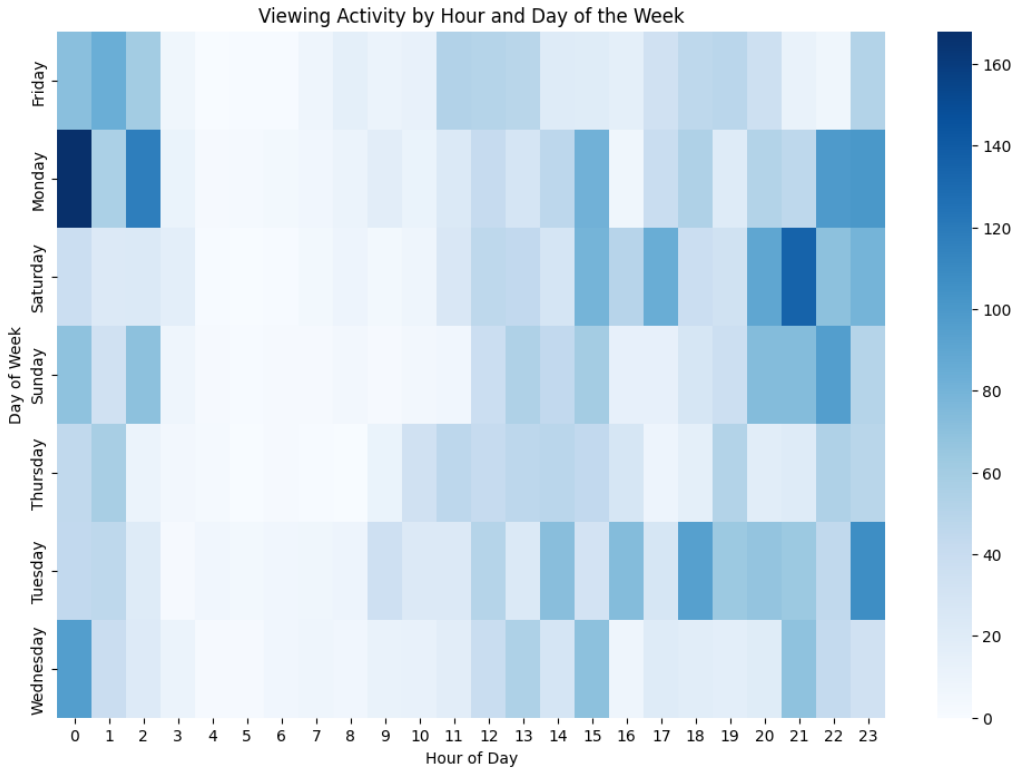
- **Data Source:** Google Takeout (YouTube watch history)
- **Key Features:**
  - Video Titles
  - Timestamps (Date and Time)
  - Channel Names
- **Preprocessing Steps:**
  - Parsed timestamps for hour and day analysis.
  - Categorized videos based on content.
  - Removed duplicates and irrelevant entries.

EDA

- **EDA Goals:**

1. Identify most-watched channels and categories.
2. Analyze viewing patterns by time and day.
3. Use visualizations (e.g., bar plots, heatmaps) for clear insights.

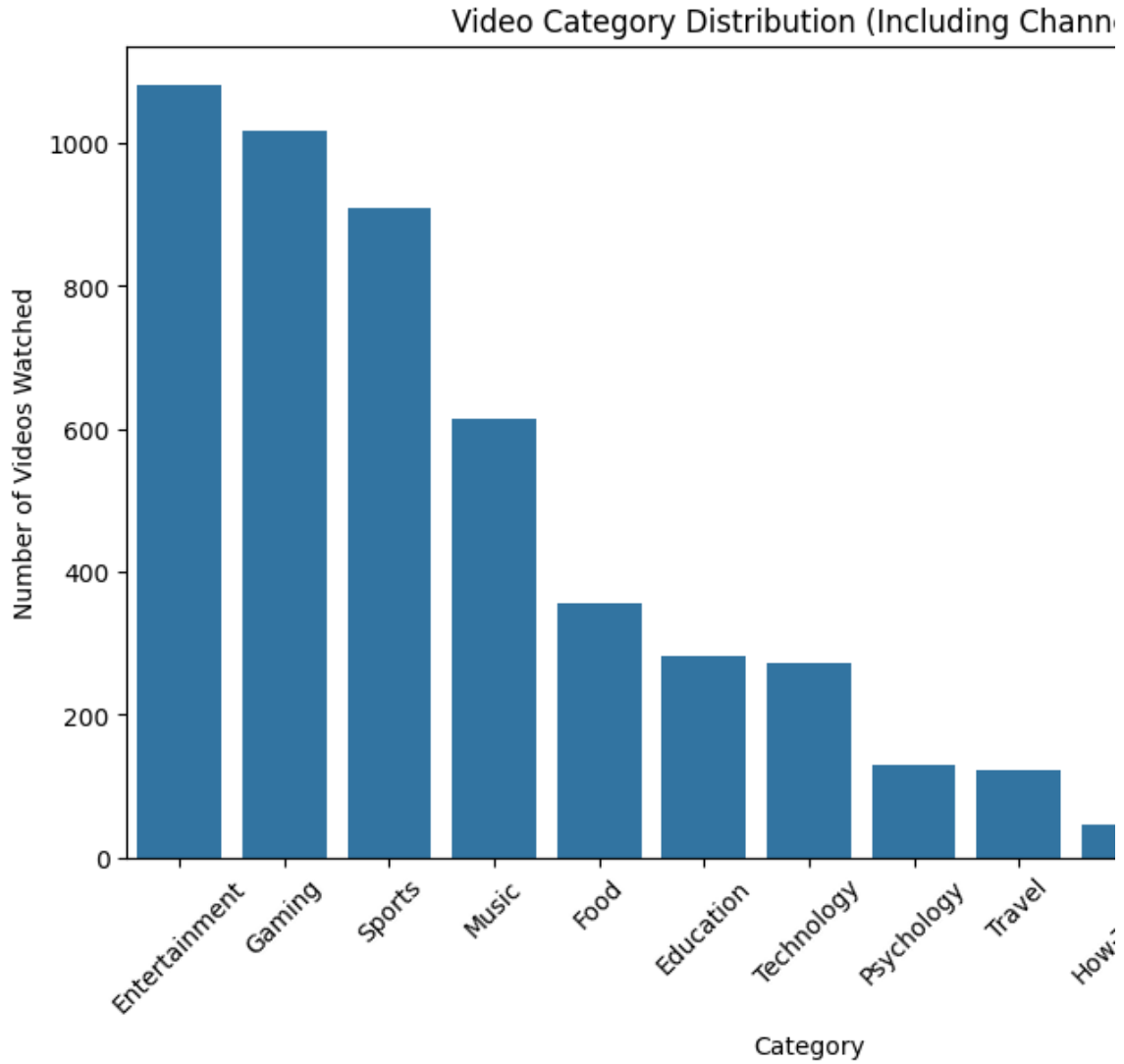
- **Visualizations:**
- Category distribution (bar plot).
- Hourly activity trends (bar plot).
- Weekly trends (day-of-week analysis).

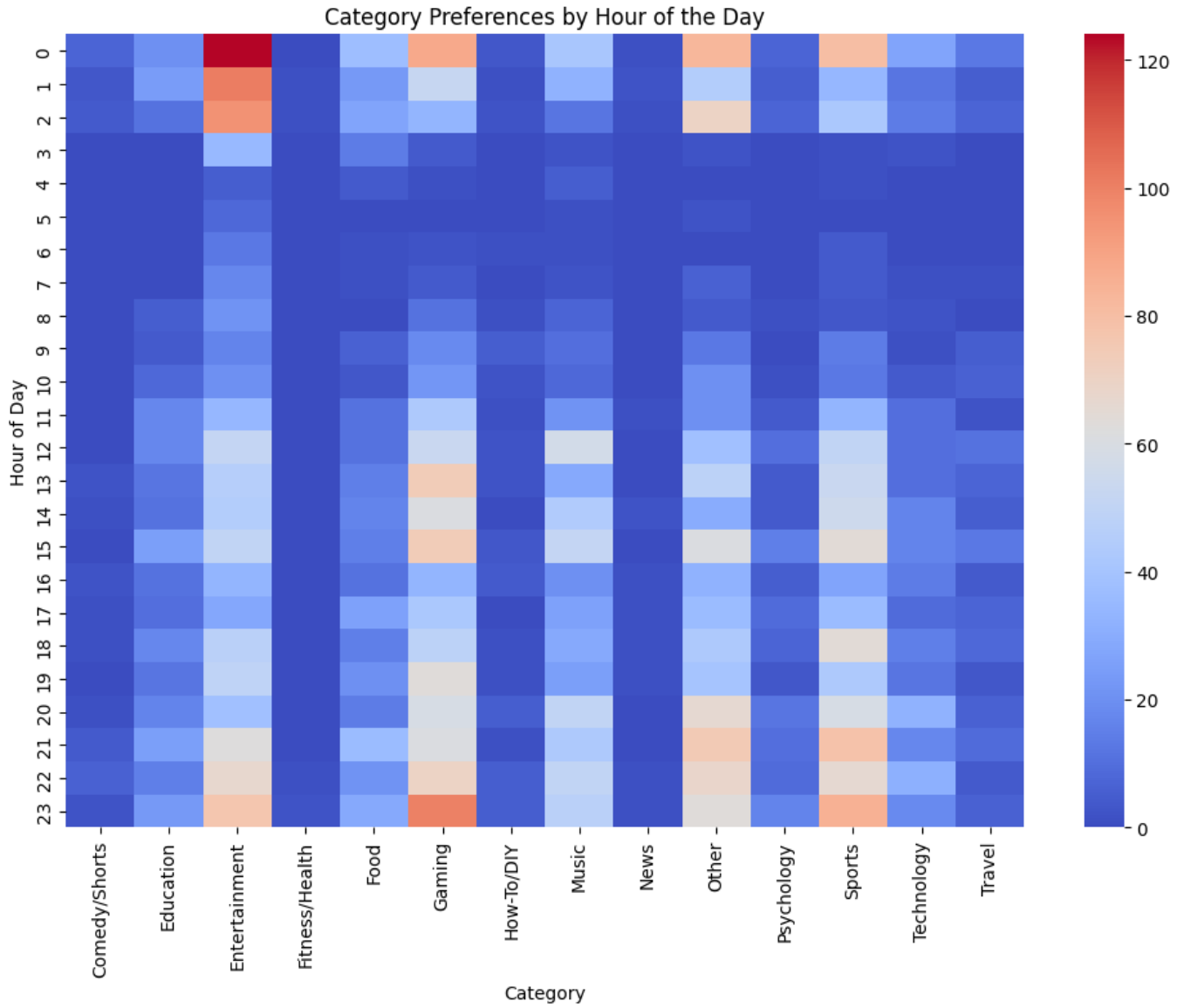


Word Cloud of Video Titles

Key Insights

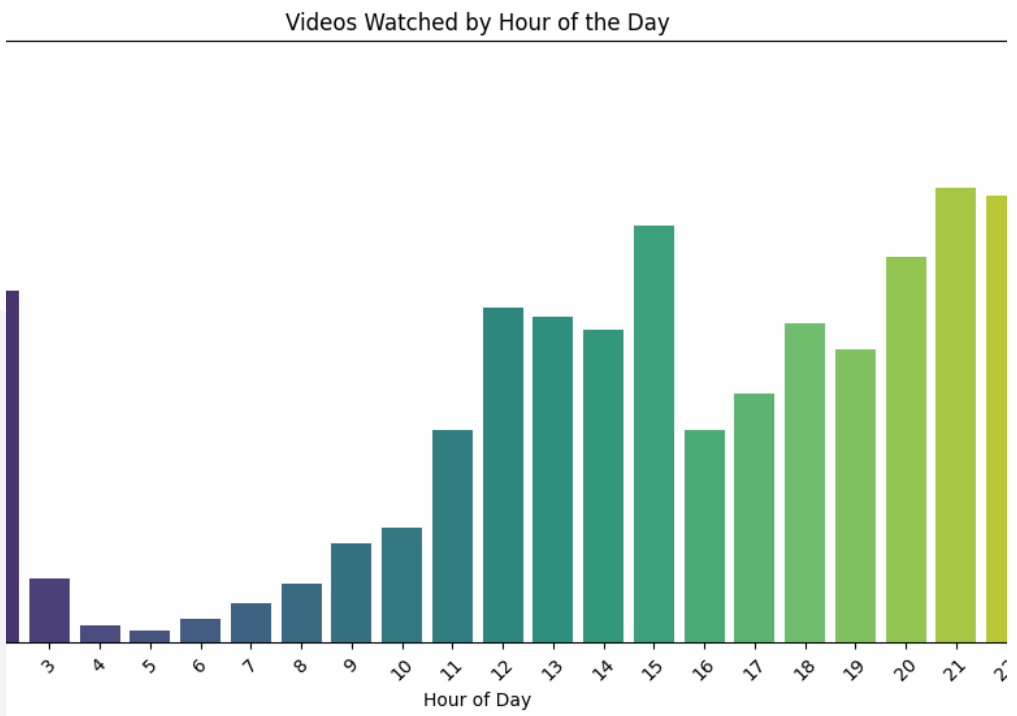
- **Content Preferences:**
  - Dominant categories: Sports, Gaming, Entertainment.
  - Influential channels: *Crossover Talks*, *Boiler Room*.
- **Time Patterns:**
  - Most activity occurs during non-working hours, particularly late at night.
  - Viewing peaks between 9 PM and midnight.
- **Weekday vs. Weekend Trends:**
  - Higher activity on weekends, reflecting increased free time.

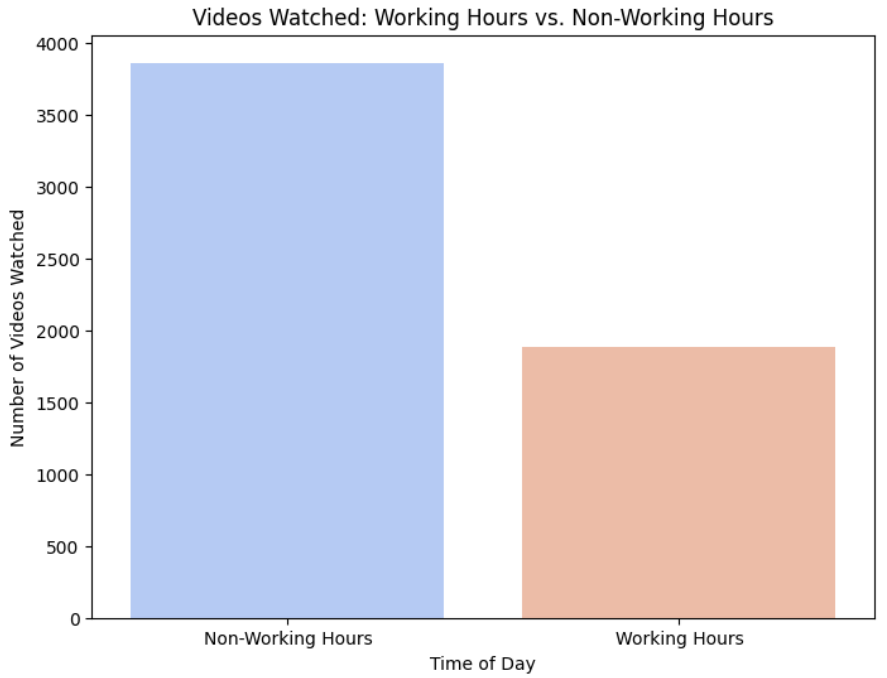



Category Preferences by Hour of the Day
Hour of Day
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Comedy/Shorts
Education
Entertainment
Fitness/Health
Food
Gaming
How-To/DIY
Music
News
Other
Psychology
Sports
Technology
Travel
Category
120
100
80
60
40
20
0

Hypothesis

- **Hypothesis:**
  "I watch more YouTube outside of regular 9–17 working hours, especially close to midnight."
- **Key Findings:**
  - Videos watched during working hours: **1882**.
  - Videos watched during non-working hours: **3863**.
  - Significant peaks in viewing activity during late-night hours.
- **Conclusion:**
  The data strongly supports the hypothesis.





Future Work

Expand the Dataset:

Include data from platforms like Spotify and Netflix.

Category-Specific Analysis:

Investigate trends within categories (e.g., preferred Music genres).

Predictive Modeling:

Use machine learning to forecast future preferences.

Content Balance Analysis:

Explore the balance between educational and entertainment content.

Conclusion

- **Key Findings:**
  - Dominant content: Music, Gaming, Entertainment.
  - Peak activity during non-working hours, particularly close to midnight.
  - Weekends associated with higher activity.
- **Takeaways:**
  - This analysis validates my viewing patterns and provides insights for better content consumption choices.
  - Data-driven methods help uncover meaningful trends in everyday behavior.