Introduction

YouTube is a significant part of my daily routine, serving as a source of both entertainment and learning. This project explores my YouTube watch history to identify patterns, validate hypotheses, and gain actionable insights into my viewing behavior. Using data analysis techniques, I was able to extract meaningful trends from my activity and evaluate how my interests align with my consumption habits.

Dataset Overview

- Source: Google Takeout (YouTube watch history)
- Key Features:
    - Video Titles
    - Timestamps (date and time)
    - Content Categories
    - Channels
- Preprocessing Steps:
    - Parsed timestamps to extract the hour of the day and day of the week.
    - Categorized videos into predefined content genres (e.g., Music, Gaming, Education).
    - Cleaned the dataset by removing duplicates and irrelevant entries.
- Tools used: Python, with libraries such as pandas, matplotlib, and seaborn for data processing and visualization.

Project Objectives

1. Content Analysis: Identify the most-watched categories and influential channels.
2. Time-Based Patterns: Analyze viewing activity across different times of the day and days of the week.
3. Hypothesis Validation: Test the hypothesis that most YouTube activity occurs outside of regular 9–17 working hours, particularly close to midnight.

Analysis Process

1. Exploratory Data Analysis (EDA):
    - Used bar charts to identify the distribution of videos across categories such as Music, Gaming, and Entertainment.
    - Heatmaps visualized viewing activity across hours of the day and highlighted category preferences at different times.

2. Time-Based Patterns:
   - Extracted hourly viewing trends from timestamps to analyze activity peaks.
   - Classified timestamps into working hours (9–17) and non-working hours to compare activity levels.
3. Channel Analysis:
   - Grouped videos by channel to identify the most influential content creators.
   - Word clouds visualized recurring channel names in the dataset.
4. Hypothesis Testing:
   - Segmented data into working and non-working hours.
   - Compared the number of videos watched in each time range to validate the hypothesis.

Key Findings

1. Content Preferences:
   - Most-watched categories include Music, Gaming, and Entertainment.
   - Channels like *Crossover Talks* and *Boiler Room* significantly shape my preferences.
2. Time Patterns:
   - Viewing activity peaks during non-working hours, especially late at night (9 PM to midnight).
   - Minimal activity occurs during working hours (9–17).
3. Weekday vs. Weekend Trends:
   - Higher activity on weekends aligns with increased free time and leisure.
4. Hypothesis Validation:
   - The hypothesis was confirmed, with 3863 videos watched during non-working hours compared to 1882 videos during working hours.

Visualizations

1. Category Distribution:
   - A bar chart highlighted the dominance of Music, Gaming, and Entertainment categories.
2. Hourly Activity Trends:

    - Heatmaps revealed the distribution of activity across hours and categories.
3. Weekend vs. Weekday Activity:
    - A bar chart compared viewing patterns on weekdays versus weekends.

How These Insights Were Achieved

1. Data Cleaning:
    - Removed unnecessary entries and ensured timestamps were correctly formatted.
2. Categorization:
    - Developed a comprehensive dictionary to classify video titles into predefined categories.
3. Visualization:
    - Created bar plots and heatmaps to effectively communicate trends and patterns.
4. Hypothesis Testing:
    - Segmented the data and compared counts across working and non-working hours to validate assumptions.

Future Work

- Dataset Expansion: Analyze data from other platforms like Spotify and Netflix to provide a holistic view of digital consumption.
- Predictive Modeling: Use machine learning to forecast future viewing trends and identify emerging interests.
- Category-Specific Analysis: Dive deeper into preferences within Music, Gaming, and other genres.
- Content Balance: Explore the ratio between entertainment and educational content to align habits with personal growth goals.

Conclusion

This analysis validates my YouTube viewing patterns, particularly the hypothesis that I watch significantly more content during non-working hours. The findings provide valuable insights into my digital consumption habits and set the stage for more intentional and balanced content choices. Using data-driven techniques, I've gained a deeper understanding of my preferences and patterns, laying a foundation for continued exploration and improvement.