Analysis Of My Youtube Data

This project delved into my YouTube usage data, initially collected through Google Takeout in a JSON format. After confirming that the dataset was complete, with no missing values, I focused on the "time" variable to explore how my YouTube habits varied throughout the day. The main goal was to see if there was a connection between the time of day and the amount of time I spent on YouTube.

Using statistical tests, I found interesting results. The chi-squared test, a method for analyzing categorical data, showed a very low p-value. This led me to reject the idea that there was no connection between the time of day and my YouTube usage. On the other hand, the t-test suggested there might not be a significant relationship.

I also tried machine learning models, especially decision tree regression, which predicted the season based on the time. The models had good accuracy, confirming the link between months and YouTube usage.

Visualizing the data played a crucial role. Monthly comparison charts highlighted differences between summer and non-summer months, emphasizing higher average usage during non-summer periods. Another graph displayed the hourly distribution, showing peak activity between 15:00 and 20:00, with increased usage towards the end of each month.

In summary, this project was a deep dive into understanding my YouTube habits. The rejection of the null hypothesis, backed by statistical tests and visualization, pointed to a clear connection between the time of day and my YouTube usage patterns. The main takeaway is that our statistical tests and visualizations provided consistent evidence for a link between time and YouTube activity. Future work could explore more aspects of user behavior or consider additional variables for a more nuanced understanding.