Group 5 Kayla Williams

Victoria Zbierowski

Jack Hildebrandt

**Project Proposal**

Problem

Social media and other entertainment consumption has an impact on users’ well-being, habits, and digital engagement. This project will analyze user behavior across demographics using the [Social Media and Entertainment Dataset](https://www.kaggle.com/datasets/ashaychoudhary/social-media-and-entertainment-dataset/data) from Kaggle. By applying clustering and predictive modeling, we aim to categorize users based on their habits and develop a recommendation system that predicts advertisement interaction trends, which will help advertisers target users more effectively.

Dataset

We will be using the [Social Media and Entertainment Dataset](https://www.kaggle.com/datasets/ashaychoudhary/social-media-and-entertainment-dataset/data) from Kaggle.

This dataset contains 40 different columns with a variety of data. Our dataset has discrete valued columns, continuous valued columns, and categorical data. This dataset is very informative and encompasses a variety of factors that can contribute to daily social media or entertainment use. There is a lot of potential for analysis on how excessive social media/entertainment use can impact someone’s well-being (quality sleep, social well-being, and engagement in physical activities) and how social media or entertainment can develop tech savviness. We can look at these different metrics across a variety of different countries to study the differences across regions. With over 300,000 data points, there are substantial amounts of data that can offer a more realistic view into the factors that can affect people’s day to day lives. Although our data is synthetic, there are plenty of avenues for analysis to draw conclusions from.

Methodology

1. Data Collection and Preprocessing
   1. Data Cleaning – Fill in missing data and format data uniformly
   2. Clustering – grouping users into different categories based on their demographics and habits
2. Recommendation Feature Design
   1. Predictive Modeling – predict what type of user group is likely to interact the most with different types of advertisement
3. Model Evaluation
   1. Accuracy – Verifying how accurate our model is
   2. Refining Model – go back and fine tune our model
4. Results