Introduction:

The role and use of social media has been a widely discussed topic in recent years. The rapid rise of platforms like TikTok, along with the continued popularity of Instagram, Facebook, X (formerly Twitter), and YouTube, has sparked much debate and controversy. This has led to numerous studies and policy actions aimed at understanding the effects of social media. This project aims to look at global social media trends, its effect on factors like sleep, how they are used, and how much they are used on a daily basis.

Introduce the Data:

The dataset that I am using for this project is the Social Media and Entertainment Dataset. The dataset was put together using web scraping and can be found on Kaggle (here). The dataset consists of the digital habits of 300,000 individuals from different countries, capturing their social media and entertainment platform interactions. It also comes with 40 features to analyze such as daily screen time, digital-well being, daily social media/entertainment time, work/study time, and much more.

Preprocessing:

When looking at the feature columns, for what questions I want the answers too, I don't need all 40 columns. One of the preprocessing steps that I did was to drop unneeded features. The next step was to check the integrity of the data, to make sure that there were no nulls or missing values for each row, lucky the dataset was all filled and there was no need to throw out or fill any null spots.