Franchesca Johnson

7/30/2023

Introduction

It's January 2020 and the first confirmed case of Covid-19 is in the U.S. Even though China has been dealing with the spread of this unknown virus for the past month, it hasn't really hit the Western part of the world, yet.

Slowly, more positive cases are starting to emerge throughout the west coast of the U.S. As the virus starts to spread worldwide affecting every country in different ways, it still hasn't completely affected the Americas. Around springtime of 2020, more and more positive cases of covid are emerging in our country. Employers and state officials are taking drastic steps to ensure the safety of their employees and the general population. Schools are closing and going virtual, daycares are shutting down, curfews are being enforced, jobs are becoming virtual and you're being told to stay home at all times.

You're not supposed to leave your house unless necessary, so what are you to do? I guess it's time to Netflix and Netflix. There was some chilling going on too, especially since there were so many covid babies. Either way, I'm going to be researching the most watched show/movies between 2020 to March 2022.

Research Questions

Other than the increase in births between that time and netflix's stock, what show or movie was most binged watched between that time?

What are the top 10 shows that were watched?

Who watched it (age group) and why?

Was Netflix watched more during certain times of the day?

Did location, gender, age, class have any influence as to what was watched?

Was the show new or an old series?

Did the program(s) match the current culture of that time? Are the shows/movies still in the top line up since new series have started to broadcast?

Will this data tell us anything about our culture?

Approach

First reviewing the data, will it be able to answer all my questions? Meaning is there enough information/variables will I need to eliminate certain questions immediately and potentially add new

ones based on the review? Once I can do a process of elimination, if necessary, then I'll see if I notice any additional information that stands out that may need clarification.

How your approach addresses (fully or partially) the problem.

I think it will partially address the problem based on how much data is there.

Data

Kanawattanachai, Prasert. "Netflix Daily Top 10." *Kaggle*, 12 Mar. 2022, www.kaggle.com/datasets/prasertk/netflix-daily-top-10-in-us.

Netflix movies and TV shows dataset - dataset by crawlfeeds. data.world. (2022, August 8). https://data.world/crawlfeeds/netflix-movies-and-tv-shows-dataset/workspace/file?filename=netflix_movies_and_tv_shows_sample_dataset_sample.csv

Mahmood, A. (2022, October 3). *Netflix Movies Dataset*. Kaggle. https://www.kaggle.com/datasets/anasmahmood000/netflix-movies-dataset

Required Packages

I'm using ggplot to plot a scatterplot, tidyverse to clean the data, and data explorer. The scatterplot will be a helpful tool, especially when doing the linear regression analysis. I started out with the summary function to get an overview of each variable and the min. max of each. From there I'll determine what questions I can answer based on the data and which one I can eliminate. I'm hoping new questions will emerge based on the information I see within the data.