

Response Summary:

1. Student Information *

First Name	Caroline
Last Name	Dixon
Major	Game Development
Course (e.g. CGT 270-001)	CGT 270-001
Term (e.g. F2019)	S2022

2. Email Address *

(University Email Address is required.)

dixon128@purdue.edu

3. Visualization Assignment *

- Lab Assignment

Generate

4. Identify appropriate data sources: is the data publicly available? What search methods were used? *

Data source 1	https://www.kaggle.com/shivamb/netflix-shows - aquired from this list -> https://public.tableau.com/en-us/s/resources?qt-overview_resources=1#qt-overview_resources
Data source 2	above data comes from https://www.netflix.com
Data source 3	above data comes from https://www.netflix.com

5. Data format: what format is the data in? Structured vs instructed? All text, a combination, multiple sources? Is it primary or secondary data? *

This is a primary source, as it contains all direct metadata from the netflix website, contained within text and number format.

6. Data types: what types of data are in the data? How are they stored? What is the access to the data (API, JSON, txt, csv, etc.)? What structure holds the data (data base, spreadsheet, etc.)? *

The data is in a.xlsx (Microsoft Excel) file format, and is in the form of a spreadsheet of mostly integers and strings.

Evaluate

7. Variables: list the data variables? What are the parameters? Give them names. What are the dependent variables and independent variables? *

Independent Variables: titles, cast, category, directors, type, title, countries

Dependent Variables (for titles): duration minutes, duration seasons, , date added, release year, rating, description, etc.

8. Audience & Assumptions: list any assumptions you have about the data. Who is your audience? *

My assumptions of the data are what kinds of film and tv shows would be included within this set - I'm already aware of Netflix's most popular shows on their platform like Stranger Things and Friends, I also assume that based on their demographics their primary genres are comedy and action. The audience for this data set would be the userbase of Netflix, which is a very broad demographic of people in this day and age.

Generate

9. What real life behavior does the data reflect? Does it show patterns of activity, regularity of events, a timeline, population data, etc? Explain. *

The data reflects the real life behavior of what kinds of shows people are watching on Netflix. Netflix removes shows that aren't doing well or that don't have enough demand to keep them on the platform, so

11. What are the weaknesses of the data source? Is it likely that the source will be available in the future? Is the data complete? What is the quality of the data? Is it specific to your needs for the current project? Is the data in the format you need? Are there missing data? Explain. *

The data can never be complete, because Netflix will always be updating and adding new shows and movies. If you download the file at any point, it will be outdated within a week.

12. What information is emphasized? What is the central focus of the data? Explain. *

The central focus of the data is on the individual titles and their qualities, which makes sense for Netflix metadata. The purpose of the data was to record the qualities of each title Netflix streams on their platform, so it makes sense that it would be organized this way.

13. At what level of granularity is the data provided? Is the data summarized, or do you have access to the raw data? Is the data categorized or is the data in a format that allows you to create your own categories, etc. Explain. *

The data is already categorized by type of metadata on the website (descriptions, minute durations, movie or tv show, etc), and is raw data.

14. What is the scope of the data? What topics can be covered using the data? Is there a time range/frame? Is the data for a specific area/discipline/demographic etc.? Explain. *

The data can cover the genres that Netflix streams, and the directors and actors in each film across the platform, which has a pretty broad demographic since Netflix is a popular well-known movie streaming platform and many people have switched to streaming overall.
