

Response Summary:

Mine Worksheet

Goal: to identify patterns, extreme and subtle features about the data

Objectives: Students will identify basic descriptors for the data, and categorize the data according to the specifications from the Parse Worksheet

Outcomes: Three (3) specific questions to be answered using the data

1. Student Information *

First Name	Nathan
Last Name	Rusk
Course (e.g. CGT 270-001)	CGT27000-09
Term (e.g. F2019)	F2021

2. Email Address *

nrusk@purdue.edu

3. Visualization Assignment *

- Lab Assignment

Analyze

4. Basic Descriptors: for each data component from the Parse Worksheet, identify basic descriptors (basic statistics). Explain *

Prime TV Shows Data set.csv

S.no - Average, max, min, Name of the show - String length, Year of release - Average, max, min, No of seasons available - Average, max, min, Language - String length, Genre - String length, IMDb rating - Average, max, min, Age of viewers - String length.

5. Categorize: consider what is similar and what is different? Categorize the data. Are the variables categorical (normal, ordinal, or rank). Are they quantitative (discrete or continuous)? Show categories. Explain. *

Prime TV Shows Data set.csv

S.no - Average, max, min, Name of the show - String length, Year of release - nominal, No of seasons available - Average, max, min, Language - nominal, Genre - nominal, IMDb rating - Average, max, min, Age of viewers - String length.

6. Temporal: is the data streaming data? How is it stored (all at one time, over several years in years, days, minutes, seconds)? Explain. *

The data is like a snapshot of what is currently on the streaming platform at the time of creating the data set. This particular set of data does not get updated, added, or removed from. It is stored all at one time, the time it is compiled into the excel sheet.

7. Range and Distribution: what is the distribution of the data? Few values, small size, evenly spread, sparse or dense? Explain. *

This is not a small data set either, as it has over 50 rows. The data somewhat sparse only in the IMDb Ratings category, as it has quite a lot missing. It is not extraordinarily dense as it only has a few variables per TV show. The ranges of the ratings are mostly on the higher end of the 1-10 spectrum, and the range of the S.no category range from 1 to 393, which is the amount of shows.

Evaluate

8. Questions and Assumptions: list at least 3 questions you plan to answer with the data or list the questions if they were provided. Must be complete sentences and end in a question mark. What assumptions are you making? *

Question 1	What is the average release year?
Question 2	What is the most common rating?
Question 3	What is the most common genre?
Assumptions	<ul style="list-style-type: none">-The average release year will not be the current year-Most movies or shows will not have the highest possible rating-The most common rating will likely not be that low