

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	Zhenghao
Last Name	Fang
Course (e.g. CGT 270-001)	CGT 270-009
Term (e.g. F2019)	F2021

2. Email Address *

fang254@purdue.edu

3. Visualization Assignment *

- Training Data

Analyze

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

Hollywoods Most Profitable Stories

String: Story Name, Media Name, Genre Basic description: Text length

Integer: Year, Audience Score, Rotten Tomato Score Basic description: number max min average

Float: Profitability, Worldwide Gross Basic description: number max min average

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. *

Genre is nominal. Film name, lead studio is normal and they are qualitative.

Audience score, profitability, worldwide gross year are quantitative. Year is continuous, since it could take on any value between the natural upper and lower. Audience score, profitability, and world wide gross are discrete, since they don't have "in-between" state.

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

Yes, the data is temporal. It stored the data over several years in years, from 2007 to 2011.

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

The data is small with 75 rows. The data is dense which contain lots of category of data, but it not evenly spread.

The audience score is range from 35 to 89

Profitability is range from 0.005 to 66.934

Rotten Tomato% is range from 3 to 96

World wide gross is range from 0.025 to 709.82

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 max and min score of the audience?
Question 2	What is the max and min profitability?
Question 3	Are any relationships that between the audience score and profitability?
Assumptions	The film with the higher audience score will has higher profitability.