

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 *

String: Film, Lead studio, Story, Genre. Basic descriptors: string length

Integer: Year, Rotten Tomato%, Audience score, Number of the theaters in US Opening Weekend, Box Office Average per US Cinema. Basic description: Number max min average

Float: Domestic Gross, Foreign Gross, Worldwide gross, Budget, Profitability, Net profit, Opening Weekend \$
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. *

Story and genre are nominal. They describe the different types of category.

Year is quantitative and continuous since it could natural upper and lower.

Rotten Tomato%, Audience score, Number of the theaters in US Opening Weekend, Box Office Average per US Cinema are also quantitative but discrete.

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 streaming. The data is over several years in year 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 large with 672 rows.

The data is sparse and not evenly spread, because there are lots of missing value.

Rotten Tomato% is range from 0 to 99%

Audience score is range from 19% to 96%

Number of the theaters in US Opening Weekend is range from 2 to 4468

Box Office Average per US Cinema is range from 1003\$ to 93230\$

Domestic Gross is range from 0 to 760.5\$

Foreign Gross is range from 0 to 2023\$

Worldwide gross is range from 0 to 2783.5\$

Budget is range from 0 to 300\$

Profitability is range from 0.2% to 983400%

Net profit is range from -111.01\$ to 2546.5\$

Opening Weekend \$ is range from 0 to 169.19\$

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 maximum of the Net profit?
Question 2	What is the maximum of the box office cinema profit?
Question 3	Which way is the profit of films mainly comes from?
Assumptions	For Hollywoods movies the profit is mainly from the cinema.