

## 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 \*

<b>First Name</b>	Zhenghao
<b>Last Name</b>	Fang
<b>Course</b> (e.g. CGT 270-001)	CGT 270-009
<b>Term</b> (e.g. F2019)	F2021

### 2. Email Address \*

fang254@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 \*

String: Film, Story, Genre Basic description: String length

Float: Profitability budget Worldwide gross Basic description: Number max min average

Integer: Year 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 the nominal

Profitability, budget, Worldwide gross, Year are quantitative.

Year is continuous and Profitability, budget, Worldwide gross are 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. It is stored 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 sparse since it has many voids, and not evenly spread.

Year is range from 2007 to 2011

Budget is from 3\$ to 227.8\$

Profitability is range from 1.5% to 897.8%

## 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? \***

<b>Question 1</b>	What is the average of the budget?
<b>Question 2</b>	What is the average of the profitability?
<b>Question 3</b>	What are the relationship between the budget and the profitability?
<b>Assumptions</b>	Mostly the movie with the lower budget will have higher profitability

---