URL to view Results	[Click Here]

Response Summary:

Acquire Worksheet

Goal: Identify appropriate data sources, analyze the data, identify data types, variables, list assumptions about the data
Objectives: Students will identify and acquire data from appropriate data sources
Outcomes: Data for the current visualization challenge

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

Generate

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

Data source 1	Yes, it's available. I searched on the tableau public data base https://public.tableau.com/s/resources?qt-overview_resources=1
Data source 2	Yes, it's available. I search from the open data source https://github.com/gchan/hollywood-budgets/blob/master/public/data/Most%20Profitable%20Hollywood%20Stories.csv
Data source	Yes, it's available. I search from the open data source

5. Data format: what format is the data in? Structured vs instructed? All text, a combination, multiple sources? Is it primary or secondary data? *
    The data format is structured. The data is a combination of text, figures and graphs. It's a secondary data.

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.)? *
    They are integers. They are stored by the char and bar graph. Can be access to xlsx and csv. The data base holds the data

Evaluate

7. Variables: list the data variables? What are the parameters? Give them names. What are the dependent variables and independent variables? *
    Films, Genre, Profitability, the rate from the audience, Year
    The dependent variables: profitability
    The independent variables: the rate from the audiences

8. Audience & Assumptions: list any assumptions you have about the data. Who is your audience? *
    May the people who watch the movie, the social media, the professional film critics.

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. *
    I think it is a pattern of activity. This data includes many different audience, such the movies critic website

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. *
    I think the data is too mess. It's hard to compare with each other and to see which story has the most profitability. Yes, it's kind of specific to our needs.

12. What information is emphasized? What is the central focus of the data? Explain. *
    The rate from the audience. It is the dependent variable. This the information that we need to analyze and compare.

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 summarized. The data is already organized, I can still add more variable that I want to research.

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 The Yes, there is a time range, because the economy and technology will change as the time change the which will effect the results.