

Acquire Worksheet Submission

Byrd Vis Lab <vlbyrd@purdue.edu>

Thu 8/26/2021 8:12 PM

To: Rusk, Nathan Kyle <nrusk@purdue.edu>



Acquire Worksheet

Thank you for completing this worksheet!
A copy of your responses are included below.

[Download as PDF](#)

<i>URL to view Results</i>	[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 *

<i>First Name</i>	Nathan
<i>Last Name</i>	Rusk
<i>Course</i> (e.g. CGT 270-001)	CGT 27000-009
<i>Term</i> (e.g. F2019)	F2021

2. Email Address*

nrusk@purdue.edu

3. Visualization Assignment *

- Training Data

Generate

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

<i>Data source 1</i>	Netflix Movies and TV Shows, Public data on Tableau, sample given on Tableau
<i>Data source 2</i>	Disney Plus Shows, Public data from Kaggle, found via search function on Kaggle
<i>Data source 3</i>	Amazon Prime TV Shows, Public data from Kaggle, found via search function on Kaggle

5. Data format: what format is the data in? Structured vs instructed? All text, a combination, multiple sources? Is it primary or secondary data? *

Structured. All text. Primary Data. Multiple sources as in information from the streaming services (Disney Plus) and information about the ratings from the

Data types: String, integer, floating point, and void. Stored in tables. Two csv files and one xlsx file. Data held in spreadsheets.

Evaluate

7. Variables: list the data variables? What are the parameters? Give them names. What are the dependent variables and independent variables? *

Data variables such as title, genre, year, type, id, rating, number of rating votes, number of seasons, director, actors, plot, writer, and awards.

Parameters such as year (2018), Title (The Boys), number of seasons (2).

Dependent Variables such as number of rating votes and type (movie or series). Independent variables such as rating and number of seasons.

8. Audience & Assumptions: list any assumptions you have about the data. Who is your audience? *

Can assume it will change, More will be added, ratings will change, and some will no longer be currently available so it'll be removed. The audience is the people interested in watching the content on Netflix, Disney Plus, and/or Amazon Prime. There will be a wide range of audience age demographic being as there are different maturity ratings.

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. *

Ratings show what people are interested in during the current time period. Number of seasons indicates interest. Maturity ratings and IMDB ratings can show age demographic and content popularity.

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. *

The data will frequently change as new content is added or existing content removed. The data is in an appropriate format. Some data can be void such as a movie not having anything to put in for number of seasons.

categorized in ways such as genres, maturity ratings, and whether it is a movie or a series.

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. *

You are able to see what type of content (movies or TV shows) is available. You can see the level of relevancy by looking at how many new shows or movies are being added. You can get an idea of the targeted majority audience by looking at the quantities of each maturity rating.
