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	Gabby
Last Name	Willard
Course *(e.g. CGT 270-001)*	CGT 27000-LC4
Term *(e.g. F2019)*	Fall 2021

2. Email Address *
gwillard@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	(Keywords: Titanic || Publicly Available) https://public.tableau.com/s/resources?qt-overview_resources=1
Data source 2	(Keywords: Titanic || Publicly Available) https://www.kaggle.com/warrenelder/titanic-passenger-nationalities
Data source 3	(Keywords: Titanic || Publicly Available) https://www.kaggle.com/khushimehta1729/titanic-crew-survived

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 format for all three data sources is structured. They are all secondary data. All of the data sources have a combination of text and numbers.

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.)? *

The access to all of the data sources are csv extensions. The structure of all of the data sources is a spreadsheet. The Titanic Passenger Nationality data has integers and strings. The Titanic Crew data has integers, strings, and a few void spaces. The Titanic Passenger List data has integers, floats, characters, strings, and a few void spaces.

Evaluate

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

The data variables for the Titanic Passenger List are pclass, survived, name, sex, age, sibsp, parch, ticket, fare, cabin, embarked, boat, body, and home.dest. The data variables for the Titanic Passenger Nationalities are PassengerId and Nationality. The data variables for the Titanic Crew are #, Name, Age, Embarked, Position, Crew_Type, and Survived.

The independent variables for the Titanic Passenger List are pclass, name, sex, age, sibsp, parch, ticket, fare, cabin, embarked, and home.dest. The dependent variables for the Titanic Passenger List are survived, boat, and body. The independent variable for the Titanic Passenger Nationalities is PassengerId. The dependent variable for the Titanic Passenger Nationalities is Nationality. The independent variables for the Titanic Crew are #, Name, Age, Embarked, Position, and Crew_Type. The dependent variable for the Titanic Crew is Survived.

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

My audience most likely consists of professional historians, history teachers, and their students learning about the Titanic. I noticed that most of the first-class passengers got a lifeboat, around half of the second-class passengers got a lifeboat, and only a small percentage of third-class passengers got a lifeboat. I noticed that the class with the most amount of people was third-class. A lot of the crew that worked as an engineer did not survive. The job amongst the crew that had the most people working was engineering.

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. *

It does prove that the first-class people were treated higher over the other two because they had the highest number of people who got lifeboats. It also shows that because the engineers were at one of the lowest levels of the boat, a majority of them didn't actually survive.

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. *

It is likely that all of the data sources will be available in the future. Not all of the data is complete in the Titanic Crew data and the Titanic Passenger List data. The quality of the data is detailed. The data is specific to my needs for the current project. The data is also in the format that I need, which is spreadsheets.

12. What information is emphasized? What is the central focus of the data? Explain. *

The central focus of the data is what kind of people on the Titanic were most likely to survive. There are many different details about all of the individuals that influenced the chance of whether they survived or not.

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 inside of the spreadsheets. There are no calculations so I would have to do that myself. The data is categorized by columns in the spreadsheets which makes it organized for storing the data. It does allow me to create my own categories.

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. *

Topics that can be covered by the data include class, age, sex, fare, cabin, boat, home destination, survived, nationality, and crew position. The time frame is from when the Titanic first set sail to after the crash. The data is for the specific demographic of what happened to the people on the Titanic.