

URL to view Results [\[Click Here\]](#)

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

2. Email Address *

gwillard@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 *

The variable PassengerId has a maximum of 1309, a minimum of 892, a median of 1100.5, and an average of 1100.5. That means that the largest passenger id is 1309, the smallest passenger id is 892, the number in the middle of the ascending list is rounded to 1100.5, and the average of all of the passenger ids is 1100.5 which is the same as the median.

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. *

PassengerId is an interval variable because it starts at the number 892 and not 0. This variable is discrete because the numbers are not changing. Nationality is a nominal variable because it contains categories that do not need to be placed in any specific order. This variable is also discrete because it just contains concrete information.

6. Temporal: is the data streaming data? How is it stored (all at one time, over several years in years, days, minutes, seconds)? Explain. *

The data is not streaming because the tragedy of the Titanic already happened. It was stored all at one time because this is an event in history.

7. Range and Distribution: what is the distribution of the data? Few values, small size, evenly spread, sparse or dense? Explain. *

The data is completely evenly spread because there are no missing values and there are distributed fairly along both sides of the means on both variables.

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	How many people are numbered in the dataset?
Question 2	Which is the location that was the most embarked from?
Question 3	What country had the most people on the Titanic?
Assumptions	There is no way to tell the identity of each passenger because the only corresponding information about them is a number and their nationality.
