

Acquire Worksheet Submission

Byrd Vis Lab <vlbyrd@purdue.edu>

Thu 8/26/2021 9:54 AM

To: Myers, Jack Anthony <myers436@purdue.edu>



Acquire Worksheet

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

[Download as PDF](#)

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	Jack
Last Name	Myers
Course (e.g. CGT 270-001)	CGT 270-009
Term (e.g. F2019)	F2021

2. Email Address *

4. Email Address

myers436@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	American Veterinary Medical Association
Data source 2	Kaggle
Data source 3	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? *

The data data format that is being used to present this data is in a structured data set. It lists each state in the United States and then lists various data about dogs and cats in that state. This data is all text. This data is primary as it comes straight from the American Veterinary Medical Association.

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.)? *

This data contains floating point numbers with mainly Integer data. This data is stored in excel, .xls. This data is held in a spreadsheet.

Evaluate

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

The dependent variable is the United States. The independent variable is the dogs and cats owned. The parameters of the data are the people living in the United States. Parameters include number of households, dog population, etc.

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

The audience of this data is anyone interested in knowing information about cat and dog ownership in the United States. Assumptions that I can make from the data is that more households in the United States own dogs over cats. Another assumption is that more people own a cat or dog over not owning a cat or dog in the U.S.

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. *

A pattern this data shows is that dogs are more popular than cats in U.S. households. You can also see that states with a higher population own more cats or dogs.

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. *

One weakness that this data has is the type of dog or cat the U.S. households own. It is also missing U.S. owned regions like Puerto Rico. Under one tab is claims "number of pet households" but it is only using cats and dogs as pets while many other types of pets exist. This data also seems to be missing the state Hawaii.

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

The central focus of this data is comparing dog vs cat ownership in the U.S. Dog owning households by the 1000 and cat owning household by the thousand is emphasized.

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. *

This data is summarized, I do not have access to the raw data. The data is categorized by location in the United States.

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 scope of this data is households in the United States. There is no given timeframe in the data set.