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	Senny
Last Name	Huang
Course *(e.g. CGT 270-001)*	CGT 270-009
Term *(e.g. F2019)*	F2021

2. Email Address *
huan1669@purdue.edu

3. Visualization Assignment *
- Assignment 1

Generate

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

Data source 1	Pokemon
Data source 2	Pokemon All Moves
Data source 3	Pokemon All Status Data

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

All three of the data sources are secondary data as they all took information from the pokemon database and for one of the sources bulbapedia which is another widely used pokemon database. The three data sources that I have chosen were all structured data sets meaning they already had categories for all of the data laid out before use.

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 three sources of data that I have chosen were all csv files which means they are all in spreadsheet structure. In each of my data sets, a common theme is the "type" for each pokemon and move (eg. Fire Blast is fire type; Charizard is Fire/Flying type).

Evaluate

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

KEY:
(I) - Independent Variable
(D) - Dependent Variable
The parameters for each of the datasets are:
Dataset 1: Pokemon
~ Pokedex Number (1-718) (I)
~ Name (I)
~ Type (I)
~ Health point (HP) values (I)
~ Attack values (I)
~ Defense values (I)
~ Special Attack values (I)
~ Special Defense values (I)
~ Speed values (I)
Dataset 2: Pokemon Move set
~ Name (I)
~ Type (I)
~ Category (I)
~ Power (I)
~ Accuracy (I)
~ Power Points (PP) (I)
~ Technical Machine (TM) (I)
~ Probability (%) (I)
~ Generation the move was introduced (I)
Dataset 3: Pokemon All Status Data
~ Pokedex Number (1-898) (I)
~ Pokemon Code (I)
~ Pokemon Serial Number (I)
~ Name (I)
~ Type (I)
~ Type 2 (I)
~ Color of Pokemon (I)
~ Ability 1 (I)
~ Ability 2 (I)

~ Ability 2 (I)
~ Generation (I)
~ Legendary Flag (I)
~ Mega Evolution Flag (I)
~ Height (I)
~ Weight (I)
~ Health point (HP) values (I)
~ Attack values (I)
~ Defense values (I)
~ Special Attack values (I)
~ Special Defense values (I)
~ Speed values (I)
~ Stat value totals (D)
~ Capture rate (I)
~ Base Egg Hatch Cycle (I)
~ Base Egg Hatch Steps (I)
~ Egg group 1 (I)
~ Egg group 2 (I)
~ Base EXP (I)
~ Effort HP (I)
~ Effort ATK (I)
~ Effort DEF (I)
~ Effort SP-ATK (I)
~ Effort SP-DEF (I)
~ Effort SPD (I)
~ Effort Total (D)

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

The assumption I have for the data is that when the data was compiled, there was only a finite amount for each source. This is clear with the amount of pokemon in each data set as one of my data sets only goes to generation 7 while another goes all the way to generation 8 of pokemon.
The audience for my data is more than likely either a competitive pokemon player, or a pokemon player who would like to know more about the game. Since I am looking at the data source, the audience metric of student can also be a possibility when it comes to who this data could be for.

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. *

The data I chose shows the static information about any specific pokemon (or move) up a specific generation. This is obvious with the marked number for each move and pokemon, as well with the various stats each metric is provided with.

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 of the biggest weaknesses for these data sources is definitely the lack of

variables that change the stats drastically. These variables include things such as game state, game version, status effects, and items that enhance/weaken a pokemon. Out of the three data sets, my third set covering overall pokemon status has the most indepth and highest quality of the three as not only does the set have a wider library of pokemon covered (up to generation 8), but it also incorporates new stat integers that were not present in the other sources such as the stat totals, capture rates, and even information for egg hatching. The data I have gathered from the three is plently however there is a good bit missing still such as pokemon weaknesses, match ups, changes, where to obtain specific pokemon, and optimal move sets for each pokemon.

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

The information that is emphasized for each one is definitely the name of the pokemon and moves.

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. *

Out of three data sets, the data set for Pokemon Status was the least granular, while the other two sets are somewhat low in detail and definitely could use more information. This is because the data set for pokemon status has the most extensive catalog not only in the amount of pokemon, but it also has the most when it comes to specific metrics for each pokemon. The other two sets are more general, and have a basic amount of information needed to understand what is needed from the data.

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 general scope for the data is to cover the general information about each pokemon without taking into account the different variables that could skew the results. There does not seem to be a specific time range for the data however the creation of each data set could be assumed for the amount of pokemon data for each set (the number each pokedex goes to). The data is most likely targeted towards people who are heavily interested in pokemon and want to know more about their pokemon as a whole. This is due to the fact that most of the information could be gathered from either the databases they originated from or the respective games they were from as well. Overall, the data does not seem to be 100% targeted however you would need a level of understanding to interpret the data in a way that is useful.