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

2. Email Address *
tcluff@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	Tableau's Pokemon Index. Found via the provided link. Publicly available.
Data source 2	Complete Pokemon Dataset found on Kaggle. Publicly available.
Data source 3	Egg Hatching Dataset found on Kaggle. Publicly available.

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 sets are structured data using a combination of text and numbers. Data source 1 used the public site Pokemondb.net as a source. Data source 2 used mainly pokemondb.net and serebii.net for some missing info on pokemondb.net. Data source 3 used bulbapedia (a respected pokemon wiki). All data tertiary data because it is pulled from the pokemon video games and then indexed on the websites.

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.)? *

All three data sets are stored as .csv files. All are relational databases; i.e spreadsheets.

Evaluate

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

Data source 1:
Sheet 1: # Name Type HP Attack Defense Special Attack Special Defense Speed
Sheet 2: Name Type Cat. Power Acc. PP TM Effect Prob. (%)
Sheet 3: Evolving from Evolving to Level Condition Evolution Type
Sheet 4: Attack Defense Effectiveness Multiplier
Data source 2:
pokedex_number, name, german_name, japanese_name, generation, status, species, type_number, type_1, type_2, height_m, weight_kg, abilities_number, ability_1, ability_2, ability_hidden, total_points, hp, attack, defense, sp_attack, sp_defense, speed, catch_rate, base_friendship, base_experience, growth_rate, egg_type_number, egg_type_1, egg_type_2, percentage_male, egg_cycles, against_normal, against_fire, against_water, against_electric, against_grass, against_ice, against_fight, against_poison, against_ground, against_flying, against_psychic, against_bug, against_rock, against_ghost, against_dragon, against_dark, against_steel, against_fairy
Data Source 3:
ID, Pokemon.Name, Gen.2.Steps, Gen.3.Steps, Gen.4.Steps, Gen.5.6.Steps, Gen.7.Steps,
Almost all of these data sources are attributes of specific Pokémon. For this type of data the Pokémon's Pokedex number (unique ID) is the independent variable and all attributes of the Pokémon (name, weight, types) are dependent.
The other data sources are moves and type match ups. For moves, the name of the move is the independent variable and the move's PP, type, power, etc. are the dependent variables. For type match ups, the two types being matched up are the independent variables and the effectiveness and multiplier are the dependent variables.

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

My assumptions about this data is that Pokémon with the same types would be similar. So all water types would have similar move sets and be effective against

similar. So all water types would have similar move sets and be effective against similar Pokémon. Because these datasets cover in-depth game mechanics, the intended audience is people who grew up playing Pokémon as a child and now want to do statistical analysis or train AI to play the games.

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. *

This data reflect a game design structure that dictates how Pokemon video games are played. It lays out a strategy to be good at the games and overcome the set obstacles created by the game designers or other players.

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 the data will be available in the future as the sources are well renowned, public databases that store information about the Pokémon games. Dataset 2 is complete up to Generation 8 (the newest released by Nintendo), dataset 3 is up until Generation 7, and dataset 1 is up until Generation 6. The quality for the data is good because it was pulled from public respected databases that use the code of the games as sources.

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

The first two datasets are focused on gathering as much data as possible about every single Pokémon and move. The third dataset is more specific and focuses on the steps required to hatch a Pokémon's egg for each Generation

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 the raw data that is found in the games structured into relational databases and formatted for ease of use. I could easily create my own categories if I wanted to.

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. *

Because the data covers almost every possible attribute of a Pokémon there are many relations that could be visualized. A few examples are "Are bigger Pokémon friendlier?", "Does it take longer to hatch a Pokémon in newer games?" and "Which Pokémon type is least effective against the average Pokémon?". As these datasets cover video games, the only time frame that could be used is the Generations of Pokémon (what series they are apart of). But this frame of reference is useful for seeing how the designers of Pokémon change how they create it over the years.