

Marco Noriega

Mario Banuelos

MATH 120

9th December 2025

### Final Project Written Draft

#### 1. Final Project GitHub Link: [GitHub Link](#)

My Final Project will be based on analyzing Pokemon.csv from Alberto Barradas. This dataset consists of Pokémons from Generations 1-6. I decided to choose a dataset involving Pokémons to analyze, as I grew up with Pokémons.

#### **Questions aimed to answer from my project:**

- What are the most common primary and secondary typings?/ What is the most common typing overall?
  - How much higher, on average, are the total base stats of legendary Pokémons in comparison to non-legends Pokémons?
  - How are the average total base stats changed among the different Pokémon generations?
  - Which Pokémons typings are correlated to be more offensive (higher attack/special attack/speed) and more defensive (higher defense/special defense/lower speed)?
2. The Pokémons dataset is from the following link: [Pokémon.csv](#). According to the author, the dataset consists of 721 unique Pokémons, but the dataset is 800 rows and 13 columns long, because it also contains the different forms of some Pokémons, such as Mega evolutions.

#### **The dataset consists of the following variables in columns:**

- #: Pokédex Number

- Name: Pokémon Name
  - Type 1: Primary Type of the Pokémon
  - Type 2: Secondary Type of the Pokémon
  - Total: Sum of all the base stats.
  - HP: Hit points
  - Attack: Attack stat
  - Defense: Defense stat
  - Sp. Attack: Special Attack stat
  - Sp. Defense: Special Defense stat
  - Speed: Speed stat
  - Generation: The generation the Pokémon was introduced in
  - Legendary: A boolean that indicates whether a Pokémon is legendary or not.
3. This paragraph is a combination of the methods I plan to use and the progress I made during my First Draft.

**Preliminary Progress:**

- Used the built-in function `drop_duplicates`, which removes identical rows from the dataset. This is my method in cleaning my dataset.
- Created bar charts to find the most common primary, secondary, and overall typings.
- Built a Pokémon class where each stat is an attribute to the class with method function “find” which returns the data of the Pokémon.
- Tried to use `merge` to join columns Type 1 and Type 2, but that didn’t work because `merge` is supposed to be used to join common keys from different dataframes.

- Instead, I ended up using concat from Pandas as concat joins two datasets from two different columns in the same dataset.

### **What is expected of me to use in the future:**

- Use line charts and functions to compare trends among legendary and non-legends Pokémons or the total base stats of the different generations.
- Visualizations are going to be my main method in evaluating my results.

### **Challenges:**

- Applying classes to datasets. We haven't had a notebook that requires the use of them in the context of datasets.
- Trying to find a use for loops. Loops are mainly used for bi-gram models, which are used for text analysis, not datasets.

4. Here are the steps that I will use to complete the final draft:

- The use of more visualizations. Visualizations are a reliable way of showing trends because they are both easy for viewers to understand the patterns, and are easy to code. This is because I can use previous notebooks that use a visualization that I want to use.
- The use of groupby and its aggregation functions, such as mean and count, to select specific columns.
- The use of functions as another method to compare the total base stats among Pokémons. Functions give decimal outputs, while visualizations often round to the nearest whole number.