**Term Project**

Harlan Wittlieff

Data Science, Bellevue University

DSC 530: Data Exploration and Analysis

Neerja Sahu

November 20th, 2021

1. Statistical/Hypothetical Question

Do physical characteristics (height, weight, and gender) impact Pokémon base stats?

1. Outcome of your EDA

The initial exploratory data analysis found some merit that physical characteristics impact Pokémon base stats. For example, during the difference in means testing between the Pokémon that are primarily male and the Pokémon that are primarily female, a p-value of 0.026 was seen. This indicates that the difference in means between the two subsets of Pokémon is significant. Additionally, significant results were seen from the regression analysis that predicted attack based on height\_m and weight\_kg. However, only 20% of the variability in the attack values could be accounted for due to height\_m and weight\_kg.

1. What do you feel was missed during the analysis?

A single metric that accounts for all of the Pokémon base sets would have been useful in this analysis. The initial EDA made comparisons individually when considering base stats (hp, attack, defense). Combining these variables into a single metric that physical characteristics could be investigated against would have enhanced the EDA.

1. Were there any variables you felt could have helped in the analysis?

This analysis used a subset of the Pokémon physical characteristics. Including type may have benefited the analysis. Additionally, several other base stats were excluded from the analysis to reduce the overall scope (speed for example), including these in future studies may also be beneficial.

1. Where there any assumptions made you felt were incorrect?

During this exploratory data analysis, I wanted all Pokémon data to impact the analysis, therefore no outliers were removed. An additional analysis with the outlier removed may warrant further investigation.

1. What challenges did you face, what did you not fully understand?

Some of the height\_m and weight\_kg values were missing for various Pokémon in my dataset. These values had to be individually researched and corrected.

Additionally, I attempted to use some coding methodology that was not featured in our course book. This required additional research and troubleshooting.