HEL-8048

# Justification of Decisions

# 1 Figure styles

The choice of figure styles in the project contributes to clarity ans accessibility of the data presented. For example, is the use of scatter plots particularly effective in visualizing the individual data points and highlight outliers. The use of transparency, enhances this by additionally show the density of Pokémon within a certain weight range. Histograms are useful for comparing categorical data, such as average weights across generations, making it easier to spot overall trends. This why the histogram also include a trend line, resulting from a linear regression of the data. Pie charts are used to make it easier to visually compare the weights and number of Pokémon across generations. Unlike histograms, which focus on individual data distributions, pie charts present generational data as relative categories of a whole, allowing for clearer comparisons. And lastly, a correlation matrix is made as a more complicated addition considering the variables weight, generation and type of Pokémon.

The consistent use of style, including colors, labels and axis titles ensures that the viewers can interpret the data quickly and accurately, only by looking at the figure.

The use of the python package, `matplotlib`, in the notebook allows a custom appearance, improving readability and comprehensions. It also allows the viewer to customize the plots according to their preferences, *e.g.* by changing colors in case of color blindness or other personal preferences.

# 2 Reporting

The notebook is frequently using markdown cells, to provide structured explanations and interpretations of the figures and analysis, all in one place. Clear sectioning enhances readability and ensures that viewers can navigate the report efficiently. The project also includes a `README.md` file to give through instructions on how to navigate the the different files.

# 3 Analysis

The data used in the project can be accessed through the Kaggle API, ensuring the data is sourced reliably and efficiently. Different statistical methods, such as averages, median values, standard deviations and regression is used during the analysis to get an easier understanding of the compared data. For the analysis several python libraries has been used. To make sure the notebook is runnable and reproducible, all packages are listed, with the used versions both in the `README.md` and `env.yml` file.