

Task 1

Welcome to your EDA assignment ! We hope you are excited to implement and test everything you have learnt up until now. The dataset which you'll use includes information about Pokemons.

I've got an interesting set of questions for you to get a basic understanding of pandas and data visualization libraries. GOOD LUCK!

Dataset : https://drive.google.com/drive/folders/1ZylvUuS7hFH_vGEzkhzG0VdsmShoknCN?usp=sharing

This assignment is expected to be completed in Google Colab and submitted as an ipynb file.!!

All these ques are to be answered in separate cell .

Ques1) Start by importing all important libraries . For eg, "import numpy as np"

Import numpy, matplotlib, seaborn, pandas and any others that you require.

Ques 2) read the csv file and assign it to a variable .

Ques 3) Display shape of dataframe .

Ques 4) Return an array containing names of all the columns.

Ques 5) Remove the column '#' and update the dataframe.

Ques 6) Print a list of all the unique Type-1 powers

Ques 7) How many pokemons have 'Mega' in their name?

Ques 8) What is the standard deviation of Sp. Def. in the dataset ?

Ques 9) What percentage (upto 3 decimal places) of pokemons are legendary ?

Now that you have imported the dataset and libraries, and answered some basic questions, you can move on to the visualizations and some intermediate statistics !

Ques 10) Discrete variables:

a. For each variable, plot the frequency of each unique value (histogram).

b. For each variable, identify the mode value.

Ques 11) Continuous variables:

(Define a function for parts 'a' and 'b' which will accept feature name as input and print/plot required quantities as output. This will save time)

a. For each variable, print mean, variance, skew, min, max, median, 25th percentile, 75th percentile, and inter-quartile range.

b. For each variable, plot box-and-whiskers plots.

Ques 12) Pairwise interactions :

a. Pick any two discrete-continuous pairs, and plot box-and-whiskers plot for the continuous variable side-by-side for each value of the discrete variable .

b. Plot a heatmap of correlation between all pairs of continuous variables. Use seaborn for this.