CS/CE 457/464 - Homework Assignment 2: Exploratory Data Analysis (EDA)

Due Date: Monday, September 9 at 11:59 pm

Purpose:
Demonstrate exploration of data via creation of statistical tables and visualizations using Python and tell interesting stories and insights around your analysis.

Points: 100
Part 1 (30 points)
Use `FIFA_Players_Data.csv` dataset

Below are some suggestions that might help you cleaning the data and make it more suitable for EDA

- "joined", "nation_position" and "nation_jersey_number" need imputation and missing values can be filled using appropriate values.
- "nation_position" value "GK" contains all empty values for pace, shooting, passing etc. columns. You can fill in values or completely drop GK rows. Choice is yours.
- Many values of "skill_ball_control" and "power_stamina" contain numbers with + or -. You need to process it and calculate the final number based on addition or subtraction (make sure you change the data type of column as integer after calculation)

Please provide reason/explanation for all the data cleaning and data imputation (filling missing values).

Part 2 (70 points)
You do not need to use all the columns/attributes for Part 2. Use your imaginations to come up with interesting Univariate, Bivariate and Multivariate analysis.

- Generate appropriate summary (count, mean, median or mode) tables using group keyword in pandas.
    - Include at least two tables analysis or results
- Generate appropriate visualizations for Univariate analysis
    - At least one bar chart
    - At least one histogram
- Generate appropriate visualizations for Bivariate analysis
    - At least one scatter plot (continuous vs continuous)
    - At least one visualization for (discrete vs continuous)
    - One correlation plot
- Generate one Multivariate visualization (more than two variables)

Include appropriate titles and labels for all the visualization and tables. **Interpret all the results. No points will be given without explanation.**
Deliverable: Submit a ipynb file containing your code, outputs and explanations. Include homework title, your name and your email on top of your ipynb code file.