Kicking Insights: Exploring the Correlation of Soccer Player Attributes and their Relation to Overall Player Ability

Table of Contents

Introduction

> Welcome to my data analysis project titled 'Kicking Insights: Exploring the Correlation of Soccer Player Attributes and their Impact on Player Ability'. The objective of this project is to examine the relationship among different soccer player metrics, including age, speed, height, stamina, and more. The investigation will categorize and prioritize these attributes according to their correlation levels, offering valuable insights into the attributes that have the stongest correlation to a player's overall performance rating. This information could potentially be useful for coaches in terms of strategic planning or tactical adjustments. It could also benefit scouts and recruiters by helping to identify players who are likely to make a positive impact on the team. Hopefully, the study can serve as a valuable tool for decision-makers in soccer teams, providing data-driven insights that enhance player recruitment, development, and overall team strategy.

> The dataset is sourced from kaggle.com and provides ratings of soccer player attributes, their names, salaries, and positions. The data is gathered from FIFA 23, a soccer simulation video game published by Electronic Arts. It contains data for almost 20,000 players, including ratings of a myriad of attributes, such as ball control, height, weight, stamina, speed, aggression, and many others. The link is provided below:
>
> https://www.kaggle.com/datasets/sanjeetsinghnaik/fifa-23-players-dataset

Data Wrangling and Cleaning

> The data for this project is held in a .csv file. I will use the pandas library to create a dataframe, and then perform cleaning and wrangling operations. I will check the dimensions of the dataset, check for duplicate rows, ensure there are no null values, and if needed, rename columns.

> Let's begin the analysis by importing the necessary packages we will need.

In [1]:
```
# Import necessary libraries.
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats import pearsonr
from statsmodels.stats.outliers_influence import variance_inflation_factor
```