Exploratory Data Analysis on Hockey Dataset

Name: James Ndungu

Program: Cyber Shujaa Data Science Program

Date: May 2025

1. Introduction

This report outlines the process and outcomes of performing Exploratory Data Analysis (EDA) on the Hockey.csv dataset using Python. The goal was to understand the structure of the data, clean it, and derive useful insights using visualization and statistical methods.

2. Task Completion

a. Data Loading and Inspection

The dataset was successfully loaded using pandas. The initial inspection showed that the dataset includes several numerical and categorical variables relevant to hockey player statistics.

b. Data Cleaning

- Null values were checked using .isnull().sum().
- Appropriate steps such as filling missing values and converting data types were carried out.
- Unnecessary columns were dropped to streamline the analysis.

c. Exploratory Data Analysis

- Descriptive statistics were computed using .describe().
- Data visualization was done using matplotlib and seaborn to highlight:
  * Distribution of key numerical variables.
  * Correlation between variables using a heatmap.
  * Player performance trends.

d. Data Export

The cleaned DataFrame was exported to a .csv file using df.to_csv('./Hockey.csv').

e. Code Practices

All variables were named descriptively, and the notebook contains appropriate comments and spacing for readability, following good coding practices.

f. Evidence of Completion

Screenshots showing the successful execution of key cells, data previews, and plots have been attached as required.

3. Conclusion

This lab helped in reinforcing key EDA concepts such as data cleaning, inspection, and visualization. It also emphasized the importance of good coding standards. The insights derived from the dataset could be valuable in making strategic decisions in hockey team management and performance evaluation.

Notebook Link:

[Insert your shared notebook link here - ensure it is set to 'Anyone with the link can view']