Question40:

40. Scenario: You are a data analyst working for a sports analytics company. The company has

collected data on various soccer players, including their names, ages, positions, number of goals

scored, and weekly salaries. Create dataset on your own and store in a CSV file.

Question: Develop a Python program to read the data from the CSV file into a pandas data frame,

to find the top 5 players with the highest number of goals scored and the top 5 players with the

highest salaries. Also calculate the average age of players and display the names of players who are

above the average age and visualize the distribution of players based on their positions using a bar

chart.

Answer:

import pandas as pd

import matplotlib.pyplot as plt

# Load the data

df = pd.read\_csv(r"C:\Users\jampa\Downloads\soccer\_players.csv")

# Top 5 players with the highest number of goals

top\_goals = df.sort\_values(by="Goals", ascending=False).head(5)

print("Top 5 Players by Goals:\n", top\_goals[['Name', 'Goals']], "\n")

# Top 5 players with the highest weekly salaries

top\_salary = df.sort\_values(by="WeeklySalary", ascending=False).head(5)

print("Top 5 Players by Salary:\n", top\_salary[['Name', 'WeeklySalary']], "\n")

# Average age calculation

avg\_age = df['Age'].mean()

print("Average Age of Players:", avg\_age)

# Players above average age

above\_avg\_age = df[df['Age'] > avg\_age]

print("\nPlayers Above Average Age:\n", above\_avg\_age[['Name', 'Age']], "\n")

# Visualize player distribution by position

position\_counts = df['Position'].value\_counts()

position\_counts.plot(kind='bar', color='skyblue', edgecolor='black')

plt.title("Player Distribution by Position")

plt.xlabel("Position")

plt.ylabel("Number of Players")

plt.grid(axis='y', linestyle='--', alpha=0.7)

plt.tight\_layout()

plt.show()

Output:



