**qwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnm**

|  |
| --- |
| IST 652  Mini-project using Python  ARYAN KAKADE |

* **The Data and its Source:** The dataset was taken from the Kaggle server. This dataset contains all the information about the soccer players that played in 2023 FIFA world cup. (https://www.kaggle.com/datasets/sanjeetsinghnaik/fifa-23-players-dataset?select=Fifa+23+Players+Data.csv)
* **About Dataset:** The data set consist of 5000 rows and 9 columns, where:

Player Name – Shows the name of the first initial and last name of the player.

Full Name – Shows the first and last name of the player.

Overall – Shows the rating of the player based on his current form.

Value – Shows the current market value in Euro.

Positions Played – Shows all the position the player has played in the past.

Best Position – Shows the best position the player is naturally used to play.

Nationality – Shows the county the player represents

Age – Age of the player.

Club Name – The name of the club the player represents.

* **Data Exploration:**

1) Better analysis of dataset using head and tail function.

2) Checked unique values, total count, top, frequency, std, mean, min, max using describe function.

3) Checked the data types of each column to manipulate Data accordingly.

4) Overview of total number of rows and columns using shape function to get more detailed insights about the data.

* **Data Cleaning:**

1) Limited the number of rows to 5,000 from 18,000 using nrows.

2) After exploring the data, decided to drop irrelevant columns from the datasets and further renaming them for better analysis.

3) Identifying the number of duplicate rows using df.duplicated function and dropping these values from the dataset.

4) Counted the total number of rows to check if the duplicate rows have been dropped or not.

5) Checked for the null values using isnull function and dropping these values from the dataset.

* **Data Comparison:**

1) Comparison of Clubs based on the Market Values of their players: We compared the club names based on their current market value. This comparison uses 'Current Marketvalue' as the unit of analysis. Computed the average market value and arranged it in descending order using the sored function, yielding the names of the clubs with the highest average market value.

2) Comparison of Clubs based on the Overall Ratings of their players: We compared the club names based on the overall ratings of the players. This comparison uses 'Overall Rating’ as the unit of analysis. Computed the average overall rating and arranged it in descending order using the sored function, yielding the names of the clubs with the highest average player rating.

* **Program Summary :**

1) Imported necessary libraries.

2) Connected Google colab with the Google drive.

3) Calling the data path.

4) Reading the csv file (dataset) using pd.read\_csv function.

5) Calling the dataset player\_df to check the dataset for basic overview or analysis.

6) Using head and tail function on dataset to get insights about the data.

7) Checked the data types of all the columns using df.datatypes function and checked all the column names using df.columns to further rename them.

8) Use of df.describe function to check all the attributes of the data like, unique values, min, max, mean and, standard deviation.

9) Dropping of irrelevant columns using df.drop function.

10) Counting the total number of rows and columns and checking for the duplicate values. There were total of 119 duplicate rows from 8 columns.

11) Dropped all the duplicate values.

12) Use of count function to verify the data cleaning process.

13) Finding the null values using isnull function and dropping them from the dataset.

14) Comparison of the data using pivot function and used aggregate function to get mean of the attribute on with comparison has been made.

15) Sorted the comparison result in descending order to give better analysis result.

* **Analysis Result:**

1) The analytical results demonstrate that the current market value of the players accurately describes the club's total financial value and the ranking or form of the club is based on the overall ratings of the players.