Assignment1.2

March 16, 2025

[31]:
```
# DSC630
# Armin Heldovac
# Assignment 1.2
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
file_path = '\\Users\\armin\Documents\\Homework\\DSC630\\fifa_eda_stats.csv'
df = pd.read_csv(file_path)

# Summary of data
# This dataset contains information about professional soccer players regarding␣
 ↪their personal details, attributes, skills, and contracts. This dataset can␣
 ↪be useful for analyzing player performance
# and the different skillsets across all of the leagues. The key aspects of the␣
 ↪dataset includes a player's personal information such as their name, age,␣
 ↪and nationality.
# It also shows their physical attributes such as height and weight. Another␣
 ↪aspect of the dataset is the skill metrics such as their dribbling and␣
 ↪shooting.
# A few questions I would like to answer regarding this dataset is How is the␣
 ↪age of the players distributed throughout the dataset?
# Also, Which nationalities have the most players represented within this␣
 ↪dataset?

# Display basic information about the dataset
print("Dataset Info:")
df.info()
print("\nSummary Statistics:")
print(df.describe())
```

```
Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18207 entries, 0 to 18206
Data columns (total 57 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
```

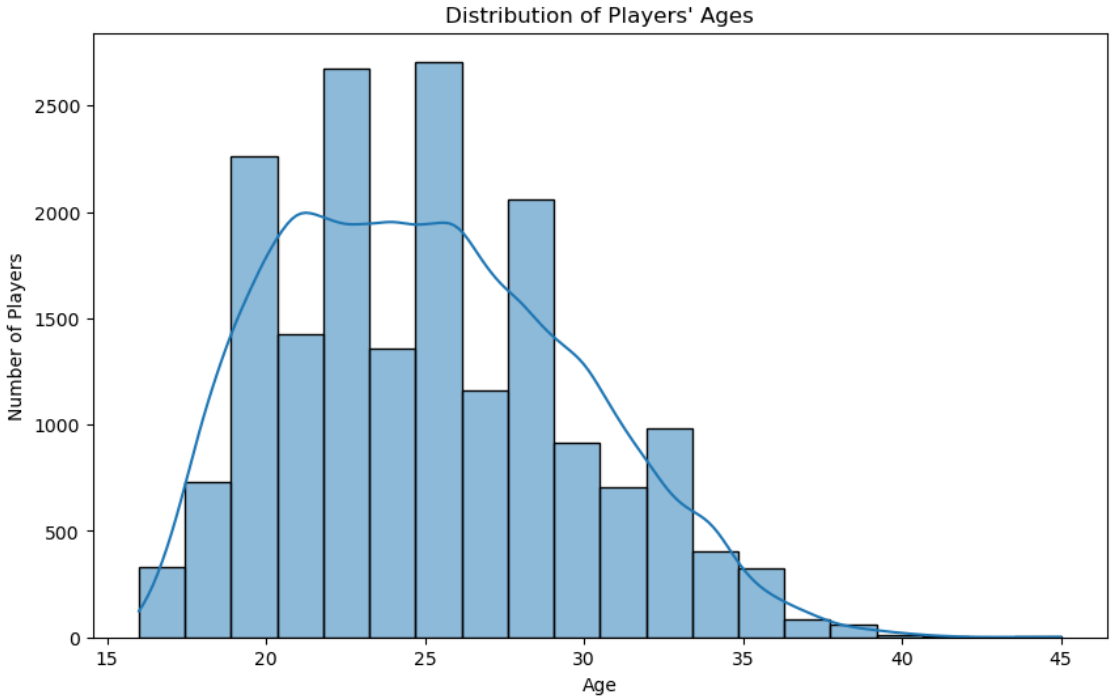
```
0   ID                        18207 non-null  int64
1   Name                      18207 non-null  object
2   Age                       18207 non-null  int64
3   Nationality               18207 non-null  object
4   Overall                   18207 non-null  int64
5   Potential                 18207 non-null  int64
6   Club                      17966 non-null  object
7   Value                     18207 non-null  object
8   Wage                      18207 non-null  object
9   Preferred Foot            18159 non-null  object
10  International Reputation  18159 non-null  float64
11  Weak Foot                 18159 non-null  float64
12  Skill Moves               18159 non-null  float64
13  Work Rate                 18159 non-null  object
14  Body Type                 18159 non-null  object
15  Position                  18147 non-null  object
16  Jersey Number             18147 non-null  float64
17  Joined                    16654 non-null  object
18  Loaned From               1264 non-null   object
19  Contract Valid Until      17918 non-null  object
20  Height                    18159 non-null  object
21  Weight                    18159 non-null  object
22  Crossing                  18159 non-null  float64
23  Finishing                 18159 non-null  float64
24  HeadingAccuracy           18159 non-null  float64
25  ShortPassing              18159 non-null  float64
26  Volleys                   18159 non-null  float64
27  Dribbling                 18159 non-null  float64
28  Curve                     18159 non-null  float64
29  FKAccuracy                18159 non-null  float64
30  LongPassing               18159 non-null  float64
31  BallControl               18159 non-null  float64
32  Acceleration              18159 non-null  float64
33  SprintSpeed               18159 non-null  float64
34  Agility                   18159 non-null  float64
35  Reactions                 18159 non-null  float64
36  Balance                   18159 non-null  float64
37  ShotPower                 18159 non-null  float64
38  Jumping                   18159 non-null  float64
39  Stamina                   18159 non-null  float64
40  Strength                  18159 non-null  float64
41  LongShots                 18159 non-null  float64
42  Aggression                18159 non-null  float64
43  Interceptions             18159 non-null  float64
44  Positioning               18159 non-null  float64
45  Vision                    18159 non-null  float64
46  Penalties                 18159 non-null  float64
47  Composure                 18159 non-null  float64
```

```
 48  Marking                   18159 non-null  float64
 49  StandingTackle            18159 non-null  float64
 50  SlidingTackle             18159 non-null  float64
 51  GKDiving                  18159 non-null  float64
 52  GKHandling                18159 non-null  float64
 53  GKKicking                 18159 non-null  float64
 54  GKPositioning             18159 non-null  float64
 55  GKReflexes                18159 non-null  float64
 56  Release Clause            16643 non-null  object
dtypes: float64(38), int64(4), object(15)
memory usage: 7.9+ MB

Summary Statistics:
                  ID           Age       Overall     Potential  \
count   18207.000000  18207.000000  18207.000000  18207.000000
mean   214298.338606     25.122206     66.238699     71.307299
std     29965.244204      4.669943      6.908930      6.136496
min        16.000000     16.000000     46.000000     48.000000
25%    200315.500000     21.000000     62.000000     67.000000
50%    221759.000000     25.000000     66.000000     71.000000
75%    236529.500000     28.000000     71.000000     75.000000
max    246620.000000     45.000000     94.000000     95.000000

       International Reputation     Weak Foot   Skill Moves  Jersey Number  \
count              18159.000000  18159.000000  18159.000000   18147.000000
mean                   1.113222      2.947299      2.361308      19.546096
std                    0.394031      0.660456      0.756164      15.947765
min                    1.000000      1.000000      1.000000       1.000000
25%                    1.000000      3.000000      2.000000       8.000000
50%                    1.000000      3.000000      2.000000      17.000000
75%                    1.000000      3.000000      3.000000      26.000000
max                    5.000000      5.000000      5.000000      99.000000

           Crossing     Finishing  …     Penalties     Composure  \
count  18159.000000  18159.000000  …  18159.000000  18159.000000
mean      49.734181     45.550911  …     48.548598     58.648274
std       18.364524     19.525820  …     15.704053     11.436133
min        5.000000      2.000000  …      5.000000      3.000000
25%       38.000000     30.000000  …     39.000000     51.000000
50%       54.000000     49.000000  …     49.000000     60.000000
75%       64.000000     62.000000  …     60.000000     67.000000
max       93.000000     95.000000  …     92.000000     96.000000

            Marking  StandingTackle  SlidingTackle      GKDiving  \
count  18159.000000    18159.000000   18159.000000  18159.000000
mean      47.281623       47.697836      45.661435     16.616223
std       19.904397       21.664004      21.289135     17.695349
min        3.000000        2.000000       3.000000      1.000000
```

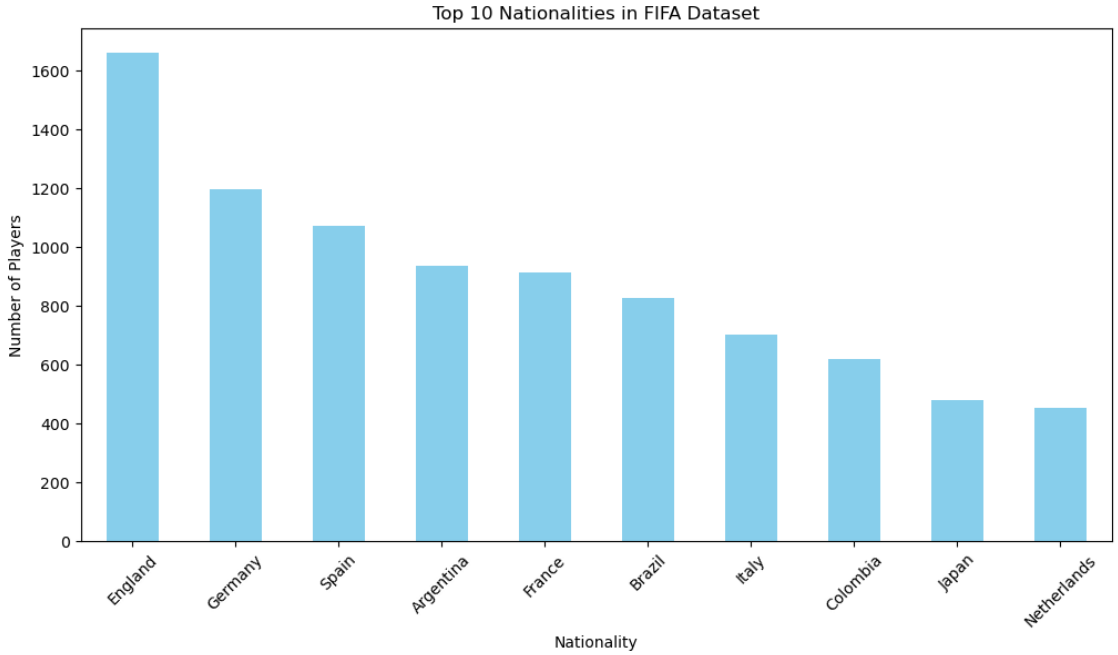
```
25%       30.000000       27.000000       24.000000       8.000000
50%       53.000000       55.000000       52.000000      11.000000
75%       64.000000       66.000000       64.000000      14.000000
max       94.000000       93.000000       91.000000      90.000000

         GKHandling     GKKicking  GKPositioning    GKReflexes
count  18159.000000  18159.000000   18159.000000  18159.000000
mean      16.391596     16.232061      16.388898     16.710887
std       16.906900     16.502864      17.034669     17.955119
min        1.000000      1.000000       1.000000      1.000000
25%        8.000000      8.000000       8.000000      8.000000
50%       11.000000     11.000000      11.000000     11.000000
75%       14.000000     14.000000      14.000000     14.000000
max       92.000000     91.000000      90.000000     94.000000

[8 rows x 42 columns]
```

```
<>:9: SyntaxWarning: invalid escape sequence '\D'
<>:9: SyntaxWarning: invalid escape sequence '\D'
C:\Users\Armin\AppData\Local\Temp\ipykernel_13544\2906780683.py:9:
SyntaxWarning: invalid escape sequence '\D'
  file_path = '\\Users\\armin\Documents\\Homework\\DSC630\\fifa_eda_stats.csv'
```

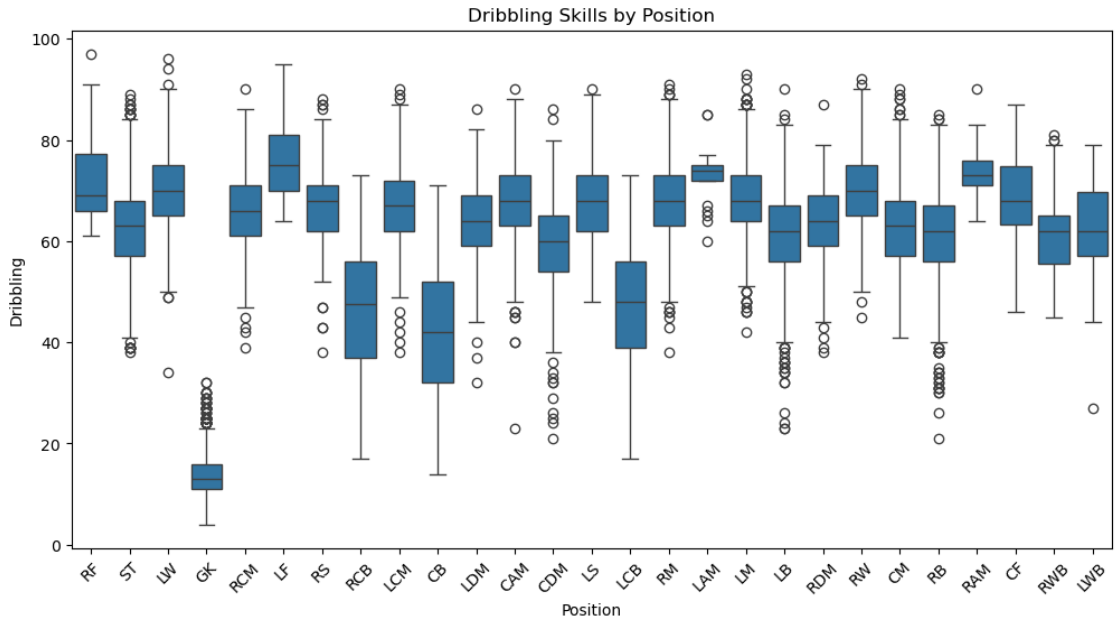
[33]:
```
# Histogram of Players' Ages
plt.figure(figsize=(10,6))
sns.histplot(df['Age'], bins=20, kde=True)
plt.title('Distribution of Players\' Ages')
plt.xlabel('Age')
plt.ylabel('Number of Players')
plt.show()
```



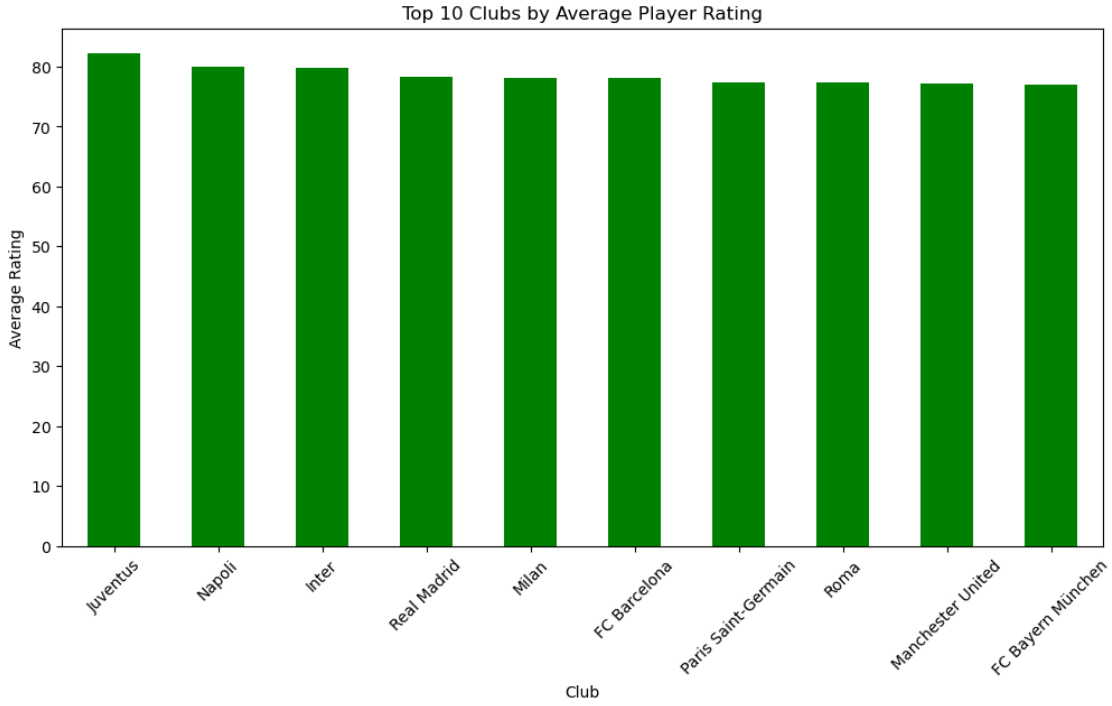
[35]:
```
# Bar Chart of Top 10 Nationalities
plt.figure(figsize=(12,6))
top_nationalities = df['Nationality'].value_counts().head(10)
top_nationalities.plot(kind='bar', color='skyblue')
plt.title('Top 10 Nationalities in FIFA Dataset')
plt.xlabel('Nationality')
plt.ylabel('Number of Players')
plt.xticks(rotation=45)
plt.show()
```



[37]:
```
# Boxplot of Player Skills by Position
skill_attributes = ['Dribbling']
df_filtered = df[['Position'] + skill_attributes].dropna()
plt.figure(figsize=(12,6))
sns.boxplot(x='Position', y='Dribbling', data=df_filtered)
plt.title('Dribbling Skills by Position')
plt.xticks(rotation=45)
plt.show()
```



[39]:
```
# Bar Chart of Average Player Ratings by Club
plt.figure(figsize=(12,6))
top_clubs = df.groupby('Club')['Overall'].mean().sort_values(ascending=False).
↪head(10)
top_clubs.plot(kind='bar', color='green')
plt.title('Top 10 Clubs by Average Player Rating')
plt.xlabel('Club')
plt.ylabel('Average Rating')
plt.xticks(rotation=45)
plt.show()
```



[42]:
```
# The key results from my analysis is that the histogram of players ages shows␣
 ↪that most of the pro soccer players are between 18 and 30 years old.
# The bar chart shows that the top 3 nationalities represented in professional␣
 ↪soccer seems to be England, Spain, and Germany.
# The boxplot that shows the dribbling skills among the different positions in␣
 ↪soccer shows that forwards have higher dribbling ratings, which is expected.
# The average player rating by club shows that the top European clubs like Real␣
 ↪Madrid, Barcelona, and Bayern Munich have the best players.
```

[ ]:
```

```