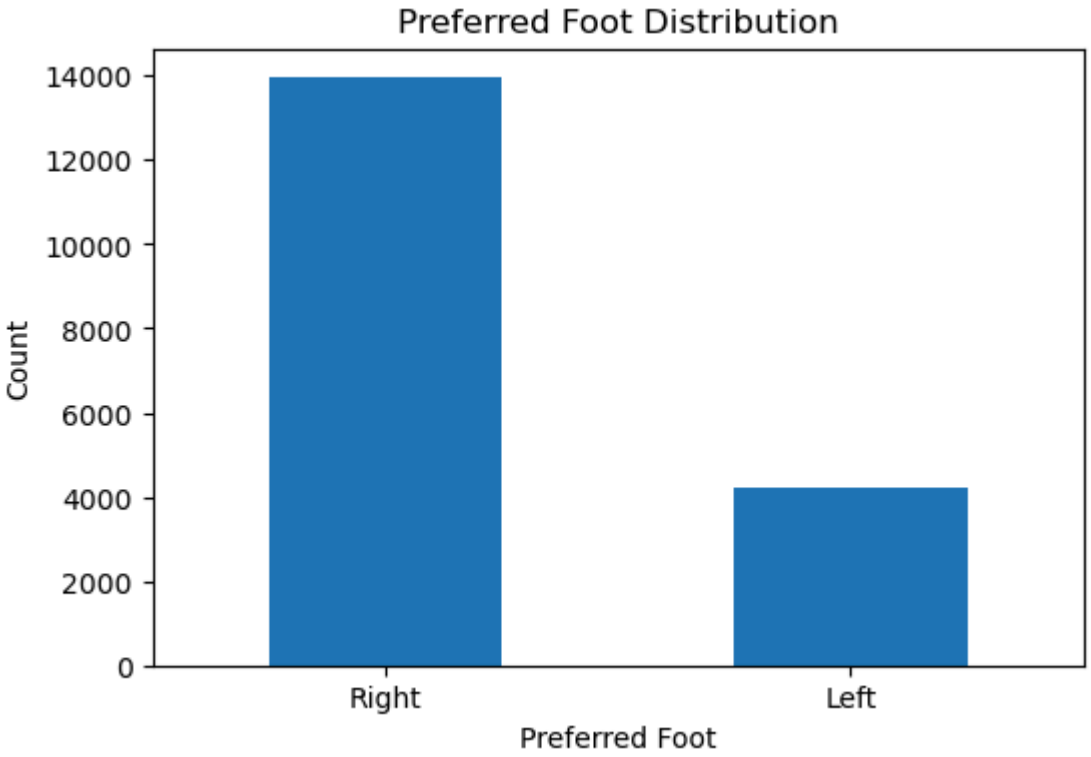
What is the distribution of preferred foot?

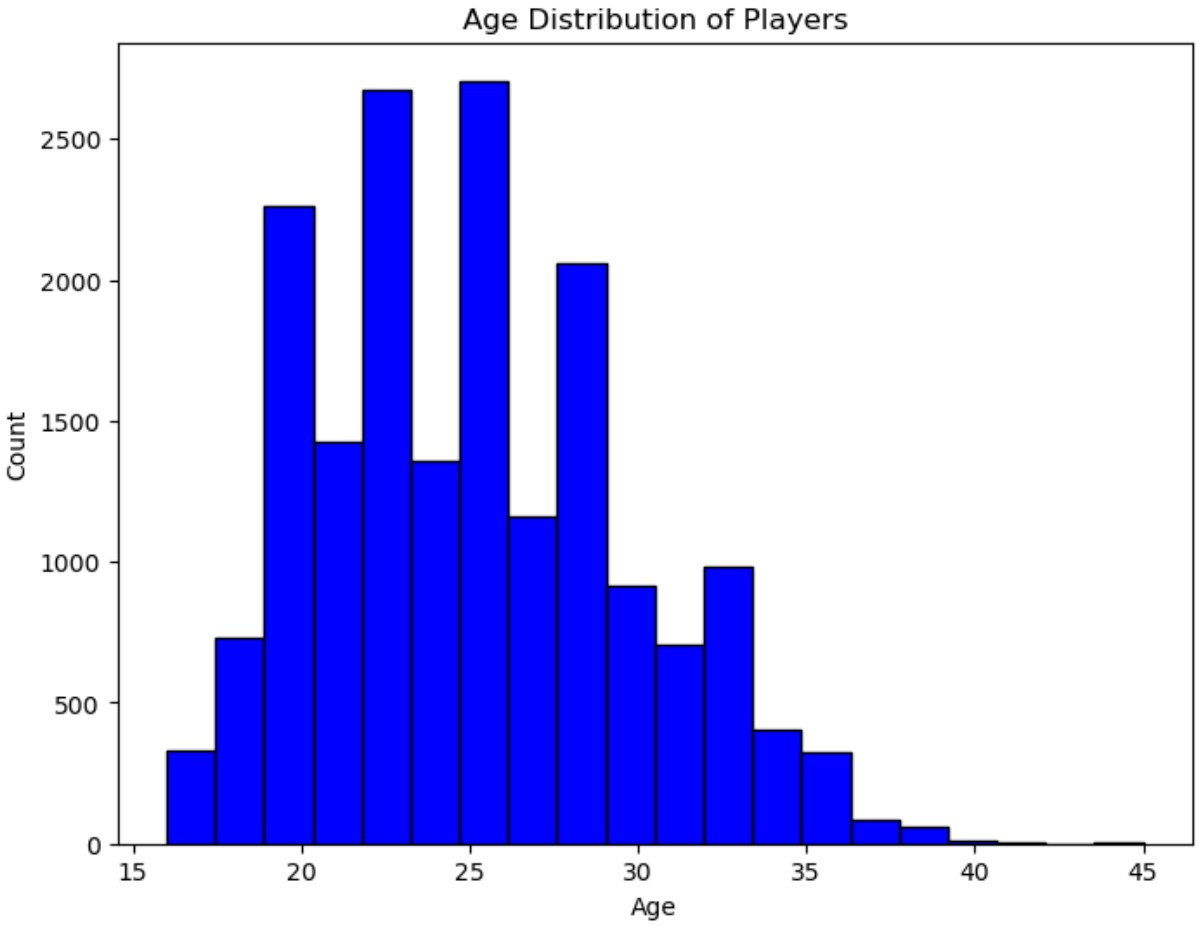
In [3]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
data = pd.read_csv("Footballer.csv")

# Analyze the distribution of the 'preferred foot' column
preferred_foot_counts = data['Preferred Foot'].value_counts()

# Plot the distribution using a bar chart
plt.figure(figsize=(6, 4))
preferred_foot_counts.plot(kind='bar')
plt.title("Preferred Foot Distribution")
plt.xlabel("Preferred Foot")
plt.ylabel("Count")
plt.xticks(rotation=0)
plt.show()
```



How can we visualize the age distribution of players?

In [5]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
data = pd.read_csv("Footballer.csv")

'age' column
```

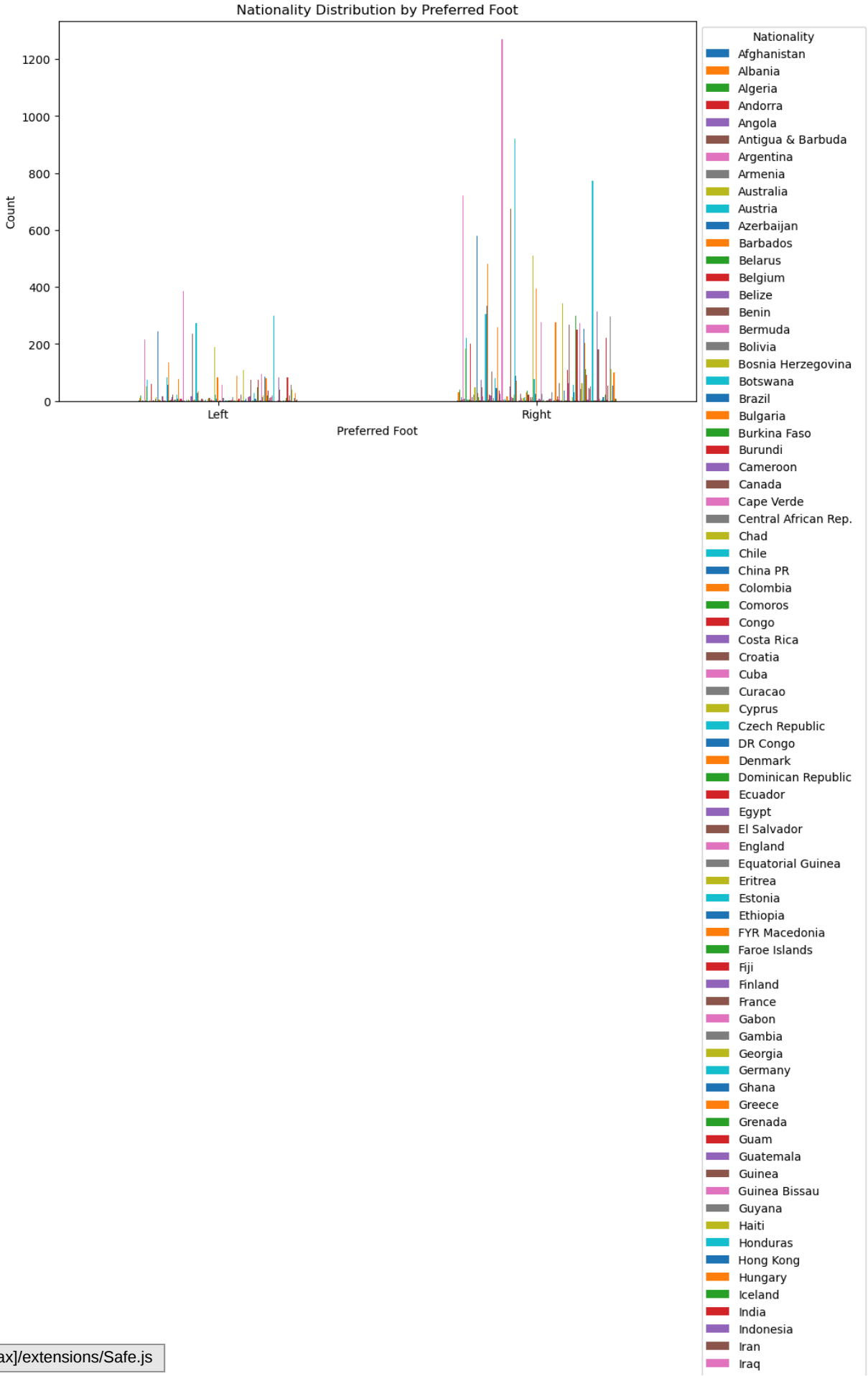
```
ages = data['Age']

# Plot the age distribution using a histogram
plt.figure(figsize=(8, 6))
plt.hist(ages, bins=20, color='blue', edgecolor='black')
plt.title("Age Distribution of Players")
plt.xlabel("Age")
plt.ylabel("Count")
plt.show()
```

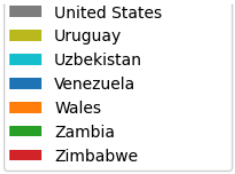


What is the international reputation of left and right foot?

In [13]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
data = pd.read_csv("Footballer.csv")

# Group data by preferred foot and nationality, then count occurrences
nationality_by_foot = data.groupby(['Preferred Foot', 'Nationality']).size()

# Plot the results using a bar chart for each foot
nationality_by_foot.plot(kind='bar', figsize=(10, 6))
plt.title("Nationality Distribution by Preferred Foot")
plt.xlabel("Preferred Foot")
plt.ylabel("Count")
```

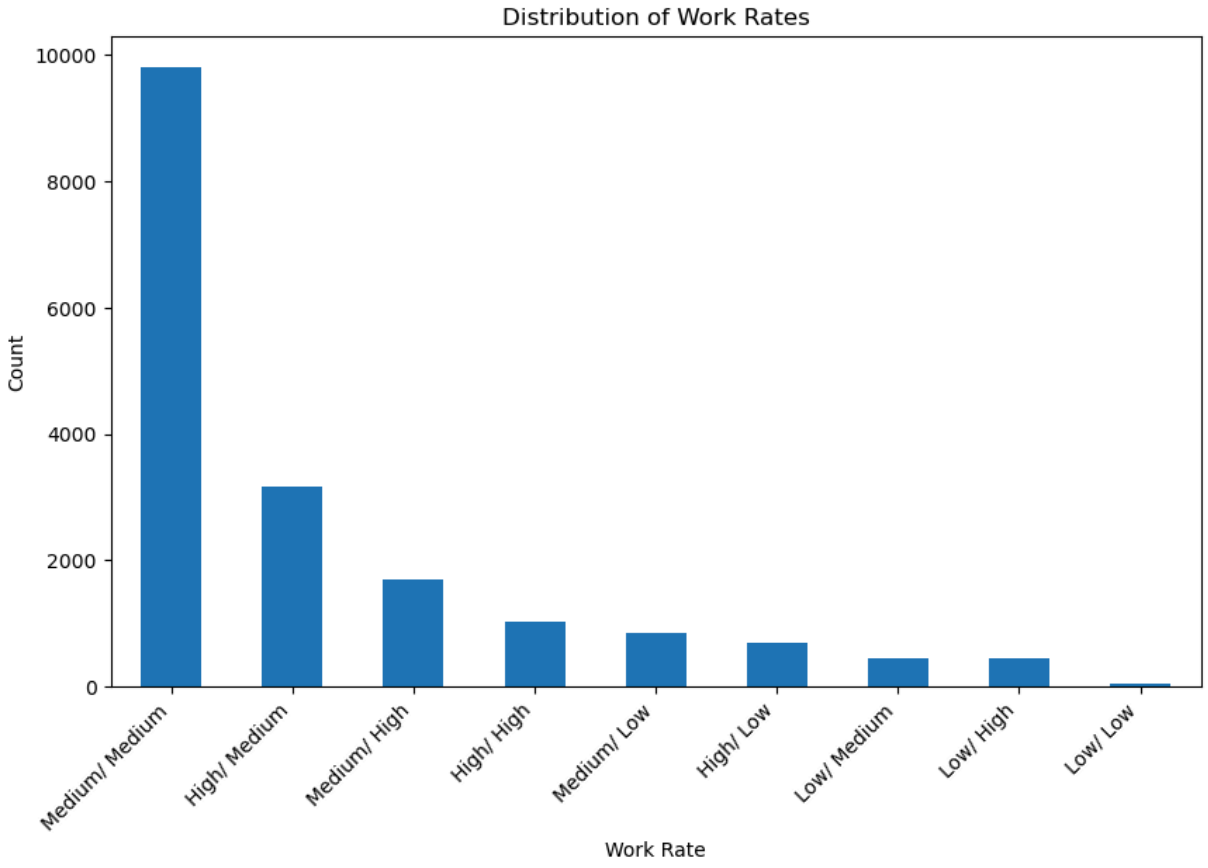
```
plt.xticks(rotation=0)
plt.legend(title="Nationality", bbox_to_anchor=(1, 1))
plt.show()
```





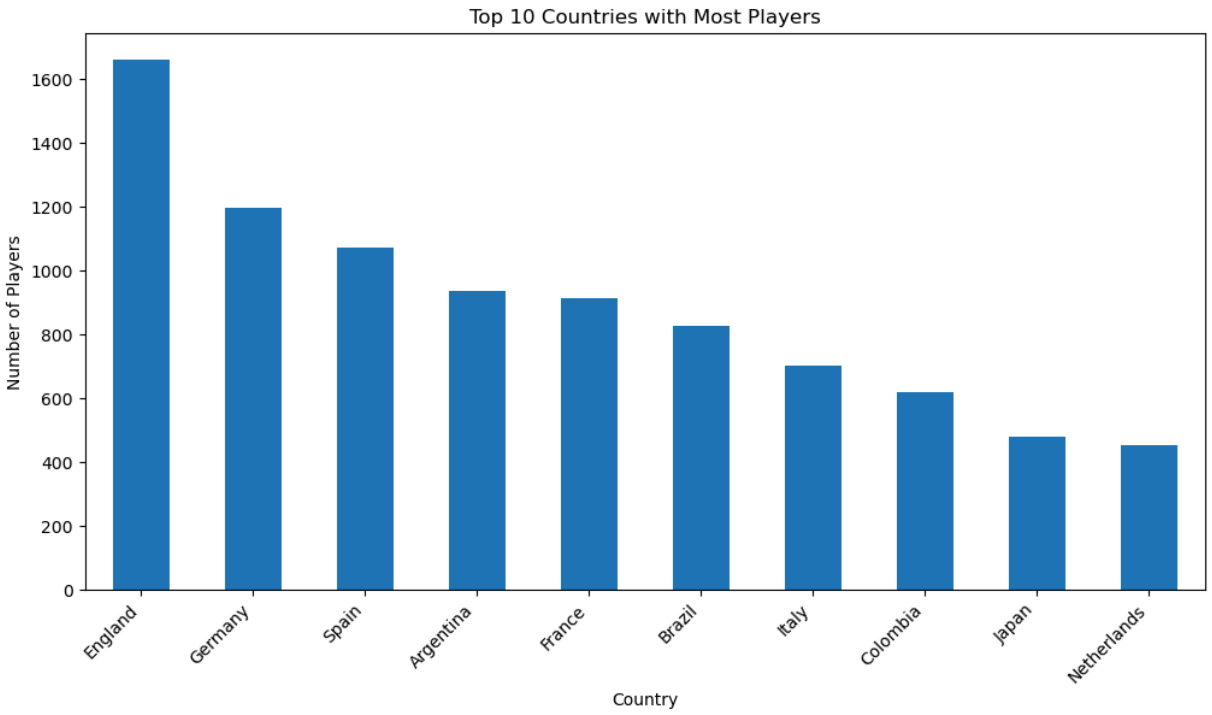
What is the different work rate of the Players?

In [15]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
data = pd.read_csv("Footballer.csv")

# Analyze the distribution of the 'work rate' column
work_rate_counts = data['Work Rate'].value_counts()

# Plot the distribution using a bar chart
plt.figure(figsize=(10, 6))
work_rate_counts.plot(kind='bar')
plt.title("Distribution of Work Rates")
plt.xlabel("Work Rate")
plt.ylabel("Count")
plt.xticks(rotation=45, ha="right")
plt.show()
```



Top 10 countries with most number of players

In [16]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
data = pd.read_csv("Footballer.csv")

# Group data by nationality and count players
country_counts = data['Nationality'].value_counts()

# Select the top 10 countries
top_countries = country_counts.head(10)

# Plot the results using a bar chart
plt.figure(figsize=(12, 6))
top_countries.plot(kind='bar')
plt.title("Top 10 Countries with Most Players")
plt.xlabel("Country")
plt.ylabel("Number of Players")
plt.xticks(rotation=45, ha="right")
plt.show()
```

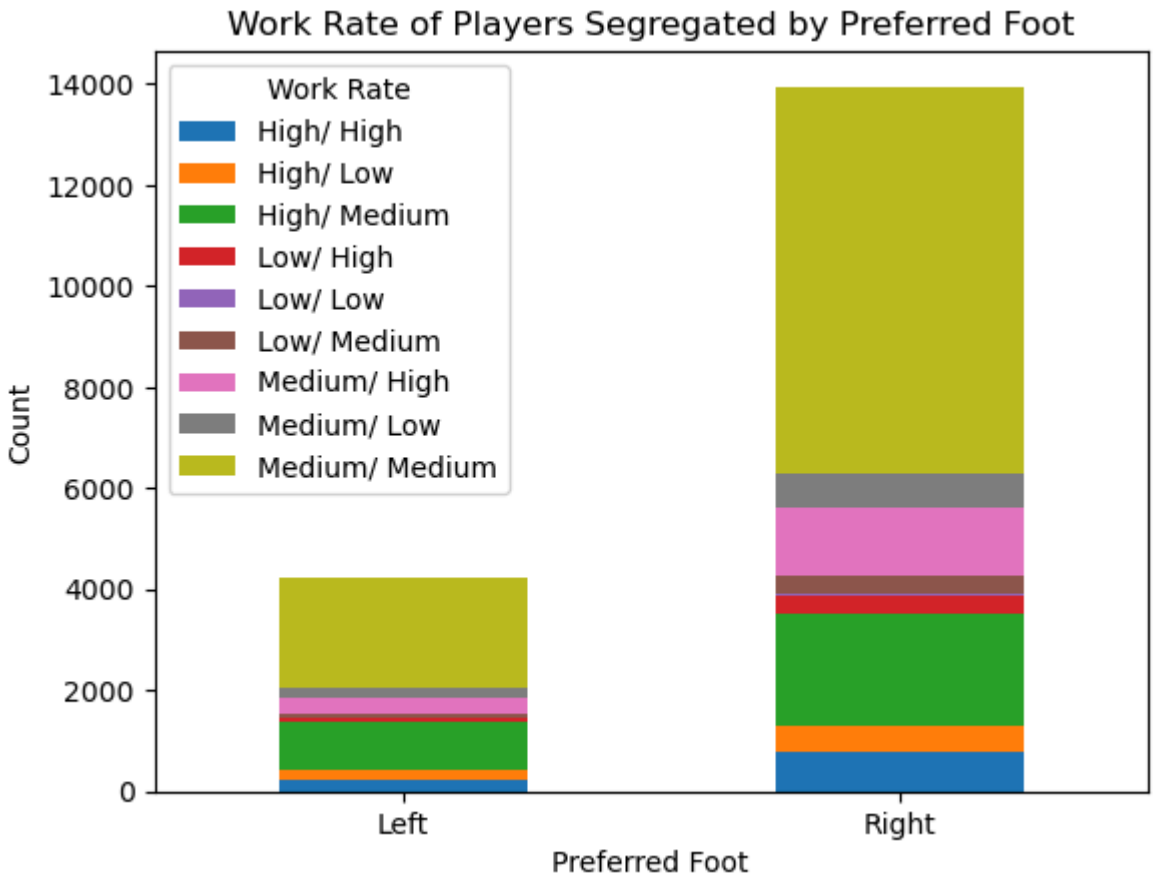


What is the work rate of players segregated by preferred foot?

In [17]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
data = pd.read_csv("Footballer.csv")

# Group data by 'preferred foot' and 'work rate', and count occurrences
work_rate_by_foot = data.groupby(['Preferred Foot', 'Work Rate']).size().uns
```

```
# Plot the results using a stacked bar chart
plt.figure(figsize=(12, 6))
work_rate_by_foot.plot(kind='bar', stacked=True)
plt.title("Work Rate of Players Segregated by Preferred Foot")
plt.xlabel("Preferred Foot")
plt.ylabel("Count")
plt.xticks(rotation=0)
plt.legend(title="Work Rate")
plt.show()
```

```
<Figure size 1200x600 with 0 Axes>
```

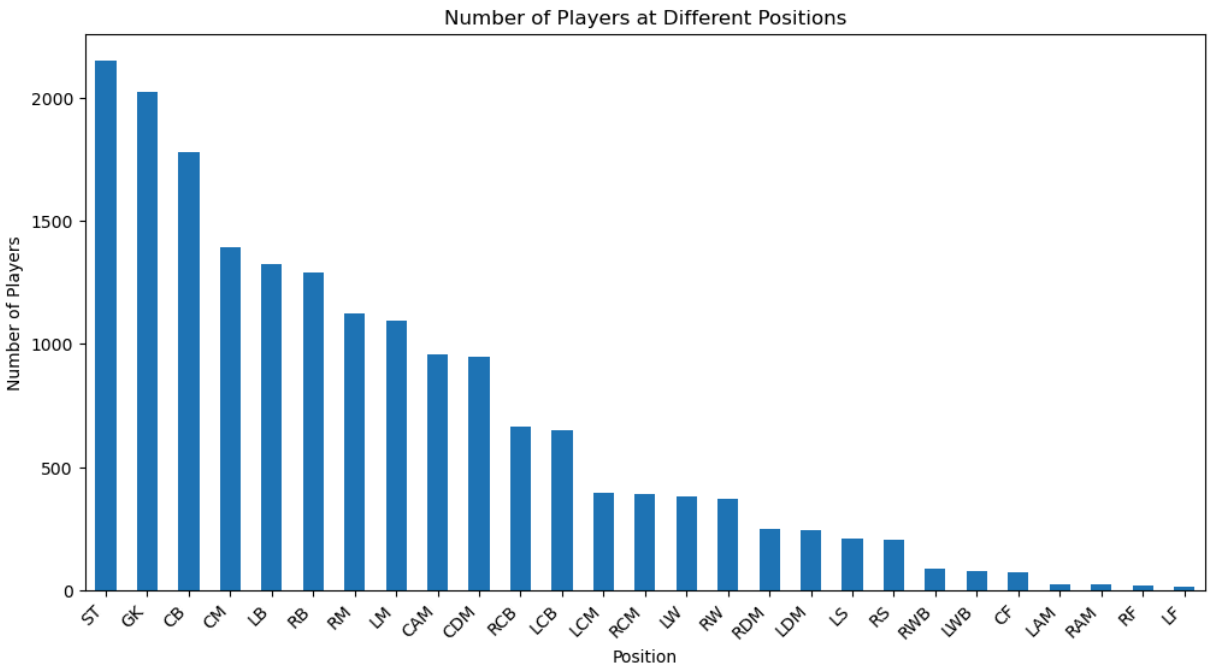


Number of players at different positions?

In [18]:

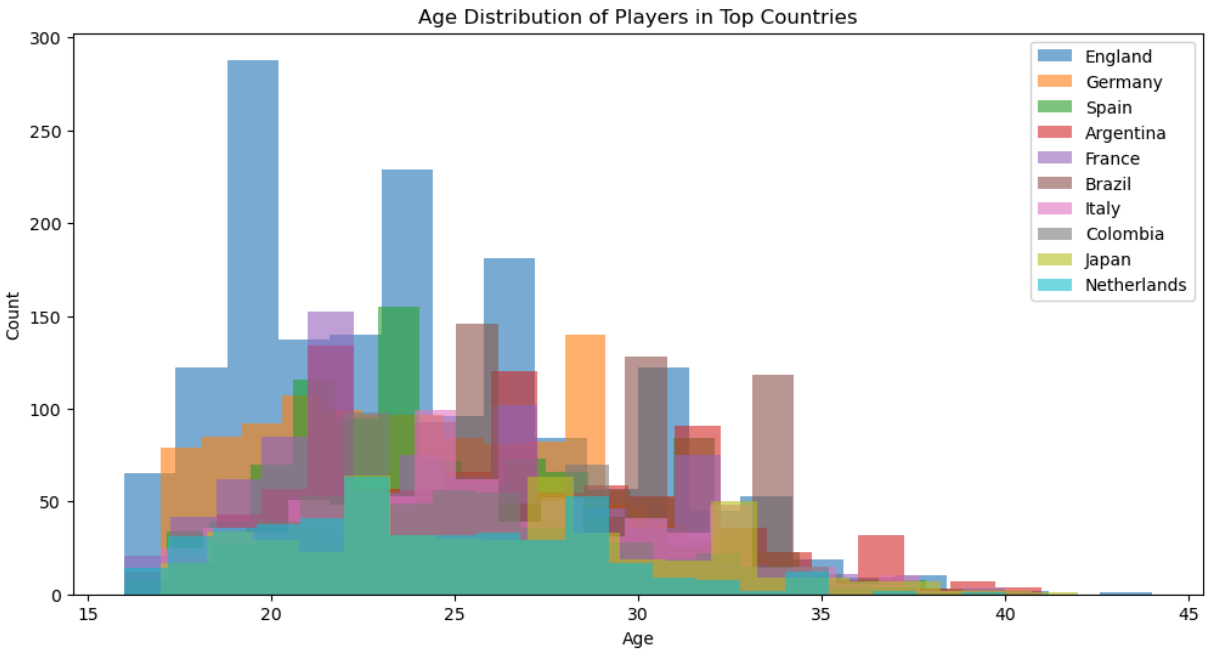
```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
data = pd.read_csv("Footballer.csv")

# Count the occurrences of each player position
position_counts = data['Position'].value_counts()

# Plot the distribution using a bar chart
plt.figure(figsize=(12, 6))
position_counts.plot(kind='bar')
plt.title("Number of Players at Different Positions")
plt.xlabel("Position")
plt.ylabel("Number of Players")
```

```
plt.xticks(rotation=45, ha="right")
plt.show()
```



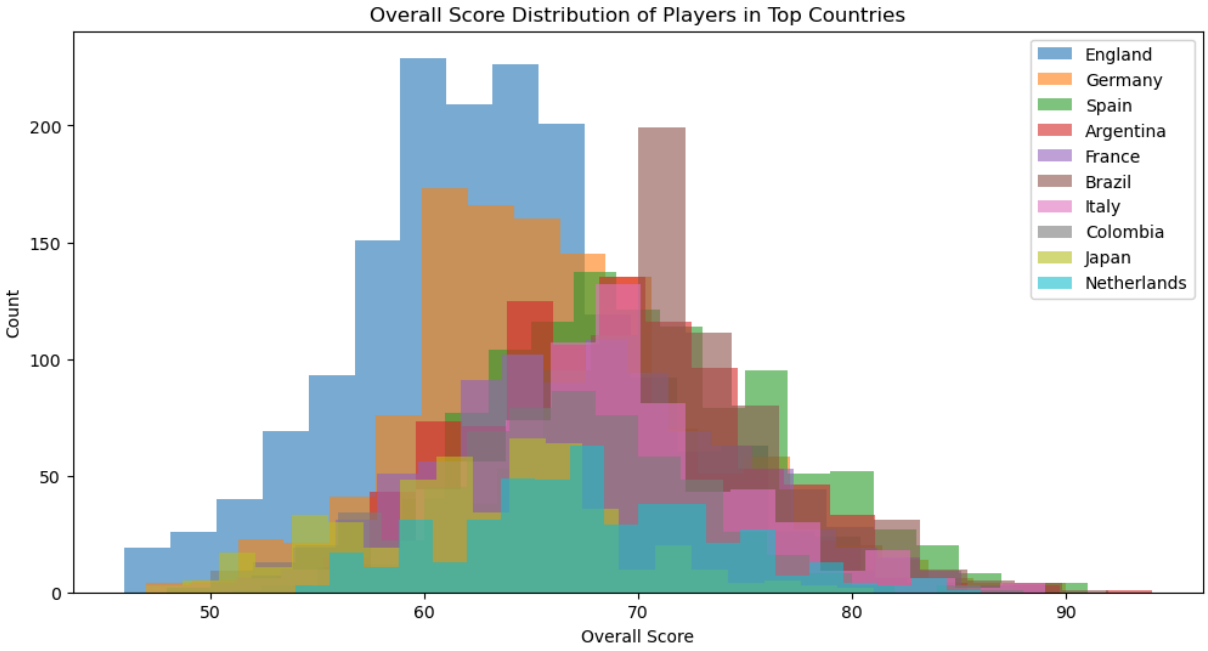
Age distribution in top countries

In [21]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
data = pd.read_csv("Footballer.csv")

# Determine the top countries based on the number of players
top_countries = data['Nationality'].value_counts().head(10).index.tolist()

# Filter the dataset to include only players from the top countries
top_countries_data = data[data['Nationality'].isin(top_countries)]

# Plot the age distribution for players from the top countries
plt.figure(figsize=(12, 6))
for country in top_countries:
    country_data = top_countries_data[top_countries_data['Nationality'] == c
    plt.hist(country_data['Age'], bins=20, alpha=0.6, label=country)

plt.title("Age Distribution of Players in Top Countries")
plt.xlabel("Age")
plt.ylabel("Count")
plt.legend()
plt.show()
```



What is the distribution of overall score of players from top countries?

In [24]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
data = pd.read_csv("Footballer.csv")

# Determine the top countries based on the number of players
top_countries = data['Nationality'].value_counts().head(10).index.tolist()

# Filter the dataset to include only players from the top countries
top_countries_data = data[data['Nationality'].isin(top_countries)]

# Plot the distribution of overall scores for players from the top countries
plt.figure(figsize=(12, 6))
for country in top_countries:
    country_data = top_countries_data[top_countries_data['Nationality'] == c
    plt.hist(country_data['Overall'], bins=20, alpha=0.6, label=country)

plt.title("Overall Score Distribution of Players in Top Countries")
plt.xlabel("Overall Score")
plt.ylabel("Count")
plt.legend()
plt.show()
```



In [ ]: