In [58]:
```
# Import necessary libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from warnings import filterwarnings
# Ignore warnings
filterwarnings(action='ignore')
```

In [10]:
```
# Load the dataset from a CSV file
data = pd.read_csv(r"C:\Users\kunal\Documents\PRODIGY Intership Material\PRODIGY_DS
```

In [11]:
```
# Display the first few rows of the dataset
data.head()
```

Out[11]:

	tour_id	tour_name	id	jersey_no	name	short_name	position_id	position	position_sho
0	148	ISL6	1514	3	Asamoah Gyan	Asamoah Gyan	2	Forward	FV
1	148	ISL6	2475	26	Deshorn Dwayne Brown	Deshorn Brown	2	Forward	FV
2	148	ISL6	2900	9	Amine Chermiti	Amine Chermiti	2	Forward	FV
3	148	ISL6	3753	7	Francisco Medina Luna	Piti	3	Midfielder	M
4	148	ISL6	3809	4	Rafael Lopez Gomez	Rafael Lopez Gomez	1	Defender	D

5 rows × 93 columns

In [12]:
```
# Display the last few rows of the dataset
data.tail()
```

Out[12]:

	tour_id	tour_name	id	jersey_no	name	short_name	position_id	position	posi
558	202	ISL7	144517	6	Hernan Daniel Santana Trujillo	Hernan Santana	3	Midfielder	
559	202	ISL7	144593	2	Scott Neville	Scott Neville	1	Defender	
560	202	ISL7	144594	4	Daniel Fox	Daniel Fox	1	Defender	
561	202	ISL7	144804	30	Ningombam Engson Singh	N Engson Singh	3	Midfielder	
562	202	ISL7	145060	44	Haobam Tomba Singh	Haobam Singh	3	Midfielder	

5 rows × 93 columns

In [16]:
```
# Drop the 'race/ethnicity' column if it exists, handle errors if the column is not
data = data.drop(columns = ['race/ethnicity'], axis=1, errors='ignore') # Corrected
```

In [18]:
```
# Fill missing values with 0
data.fillna(0, inplace=True)
```

In [19]:
```
# Display the First few rows of the dataset
data.head()
```

Out[19]:

	tour_id	tour_name	id	jersey_no	name	short_name	position_id	position	position_sho
0	148	ISL6	1514	3	Asamoah Gyan	Asamoah Gyan	2	Forward	FV
1	148	ISL6	2475	26	Deshorn Dwayne Brown	Deshorn Brown	2	Forward	FV
2	148	ISL6	2900	9	Amine Chermiti	Amine Chermiti	2	Forward	FV
3	148	ISL6	3753	7	Francisco Medina Luna	Piti	3	Midfielder	M
4	148	ISL6	3809	4	Rafael Lopez Gomez	Rafael Lopez Gomez	1	Defender	D

5 rows × 93 columns

In [26]:
```
# Check for any remaining missing values in the dataset
data.isnull().sum()
```

Out[26]:
```
tour_id         0
tour_name       0
id              0
jersey_no       0
name            0
               ..
country_id      0
country_name    0
dob             0
player_foot     0
height          0
Length: 93, dtype: int64
```

In [28]:
```
# Generate descriptive statistics of the dataset, including categorical data
data.describe(include='all')
```

Out[28]:

	tour_id	tour_name	id	jersey_no	name	short_name	position_id	p
count	563.000000	563	563.000000	563.000000	563	563	563.000000	
unique	NaN	2	NaN	NaN	399	400	NaN	
top	NaN	ISL7	NaN	NaN	Abhishek Halder	Shubham Sarangi	NaN	Mic
freq	NaN	306	NaN	NaN	2	2	NaN	
mean	177.349911	NaN	36417.937833	18.966252	NaN	NaN	2.287744	
std	26.921464	NaN	28532.321528	13.855916	NaN	NaN	1.024074	
min	148.000000	NaN	459.000000	1.000000	NaN	NaN	1.000000	
25%	148.000000	NaN	19083.500000	9.000000	NaN	NaN	1.000000	
50%	202.000000	NaN	31848.000000	17.000000	NaN	NaN	2.000000	
75%	202.000000	NaN	38918.000000	26.000000	NaN	NaN	3.000000	
max	202.000000	NaN	145060.000000	99.000000	NaN	NaN	4.000000	

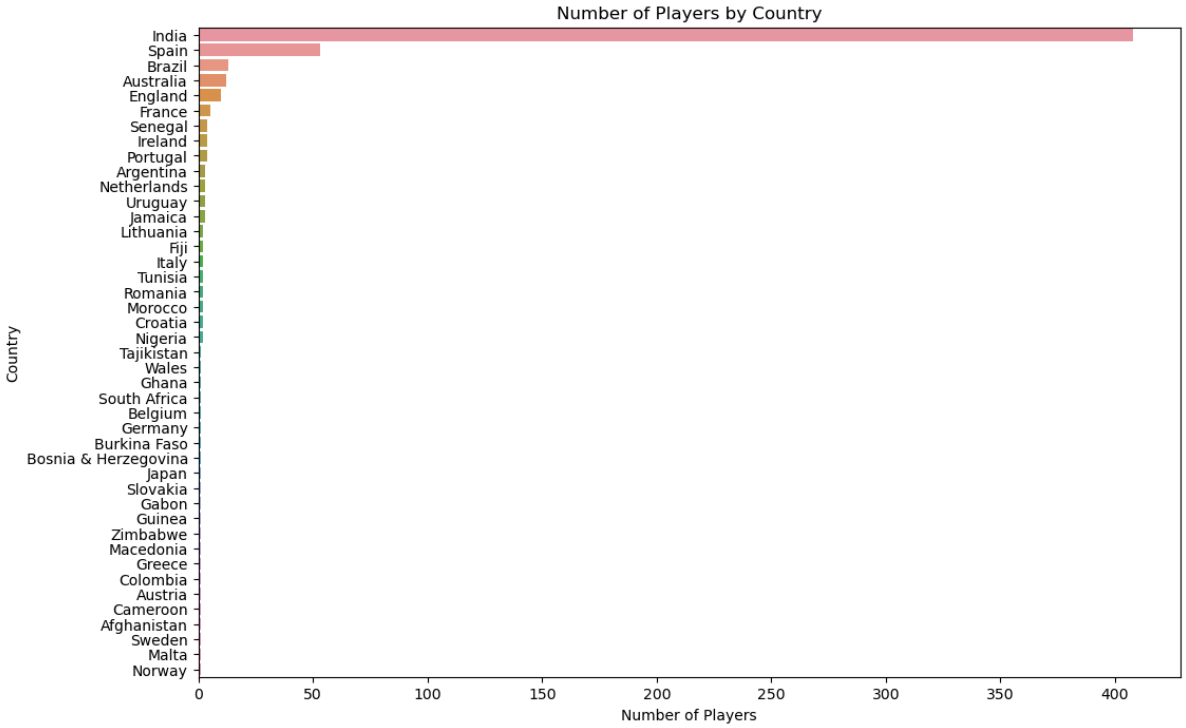
11 rows × 93 columns

In [29]:
```
# Display the data types of each column
data.dtypes
```

Out[29]:
```
tour_id          int64
tour_name       object
id               int64
jersey_no        int64
name            object
                 ...
country_id       int64
country_name    object
dob             object
player_foot     object
height         float64
Length: 93, dtype: object
```

In [48]:
```
# Visualize the number of players by country
plt.figure(figsize=(12, 8))
sns.countplot(y='country_name', data=data, order=data['country_name'].value_counts(
plt.title('Number of Players by Country')
plt.xlabel('Number of Players')
```

```
plt.ylabel('Country')
plt.show()
```



In [54]:

```
# Analyze and plot the trend of goals scored over seasons
season_goals_trend = data.groupby('name')['tour_id'].sum()
plt.figure(figsize=(20, 5))
season_goals_trend.plot(kind='line', marker='o')
plt.title('Number tours by Players Name')
plt.xlabel('Players Name')
plt.ylabel('Tour ID')
plt.grid(True)
plt.show()
```