nba_predictions

September 13, 2024

[11]:
```
# Load the dataset
import pandas as pd
df = pd.read_csv('nba_scores.csv')

df.shape
```

[11]:
```
(1316, 8)
```

[71]:
```
# Print out some summary statistics
df.describe()
```

[71]:

	season	date	away.team	away.score	home.team	\
count	1316.0	1316.000000	1316.000000	1316.000000	1316.000000	
mean	2015.0	87.086626	14.412614	100.971884	14.436170	
std	0.0	51.545912	8.662886	11.758921	8.663794	
min	2015.0	0.000000	0.000000	68.000000	0.000000	
25%	2015.0	43.000000	7.000000	93.000000	7.000000	
50%	2015.0	86.000000	14.000000	101.000000	14.000000	
75%	2015.0	130.250000	22.000000	109.000000	22.000000	
max	2015.0	209.000000	29.000000	147.000000	29.000000	

	home.score	line	over_under
count	1316.000000	1316.000000	1316.000000
mean	104.007599	-2.710486	204.595745
std	11.680406	6.947834	9.107258
min	68.000000	-21.500000	104.000000
25%	96.000000	-7.500000	198.500000
50%	104.000000	-3.500000	204.500000
75%	112.000000	2.500000	210.000000
max	144.000000	17.000000	233.500000

[73]:
```
# Encode the categorical variables in the dataset
from sklearn.preprocessing import LabelEncoder

label_encoder = LabelEncoder()

df['date'] = label_encoder.fit_transform(df['date'])
```

```
df['away.team'] = label_encoder.fit_transform(df['away.team'])
df['home.team'] = label_encoder.fit_transform(df['home.team'])
```

[43]:
```
# Choose which variables will be the Input and Output
X = df.drop(columns=['away.score', 'home.score'])
y = df.loc[:, ['away.score', 'home.score']]
```

```
Index(['season', 'date', 'away.team', 'away.score', 'home.team', 'home.score',
       'line', 'over_under'],
      dtype='object')
```

[67]:
```
# Create mock data to test on the model
# Create a new df with this data and encode it
data = [
    [2024, "2024-10-22", "New York Knicks", "Boston Celtics", -5.5, 222.5],
    [2024, "2024-10-22", "Minnesota Timberwolves", "Los Angeles Lakers", 1.5,␣
 ↪225.5],
    [2024, "2024-10-23", "Golden State Warriors", "Portland Trail Blazers", 3.
 ↪5, 226.5]
]
test_df = pd.DataFrame(data, columns=X.columns)

test_df['date'] = label_encoder.fit_transform(test_df['date'])
test_df['away.team'] = label_encoder.fit_transform(test_df['away.team'])
test_df['home.team'] = label_encoder.fit_transform(test_df['home.team'])
```

[77]:
```
# Create the model and make predictions
from sklearn.tree import DecisionTreeClassifier

model = DecisionTreeClassifier()
model.fit(X, y)
predictions = model.predict(test_df)
```

[75]:
```
# Output the predictions
for game in predictions:
    print(f"Away Score: {game[0]} : Home Score: {game[1]}")
```

```
Away Score: 108 : Home Score: 112
Away Score: 112 : Home Score: 92
Away Score: 122 : Home Score: 124
```