basketball

March 12, 2024

1 Team Project (Title tbd)

1.1 Ideas:

- Does regular season win record impact playoff performance?
  - Is a winning record enough to do well in the postseason?
- Can a player actually "carry" a team?
  -
-

[1]:
```
import pandas as pd
import plotly as px
import numpy as np
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
import plotly.express as pxo
```

[2]:
```
per = pd.read_csv("data/per.csv", index_col="Player")
ppg = pd.read_csv("data/ppg.csv")
ppg.drop(columns=["Player-additional", "Rk"], inplace=True)
per = per.rename(columns={"PER ":"PER"})
per.drop(columns="Player-additional", inplace=True)
per = per.merge(ppg, left_on="Player", right_on="Player", how="inner")
per = per[per["Tm"] != "TOT"]
```

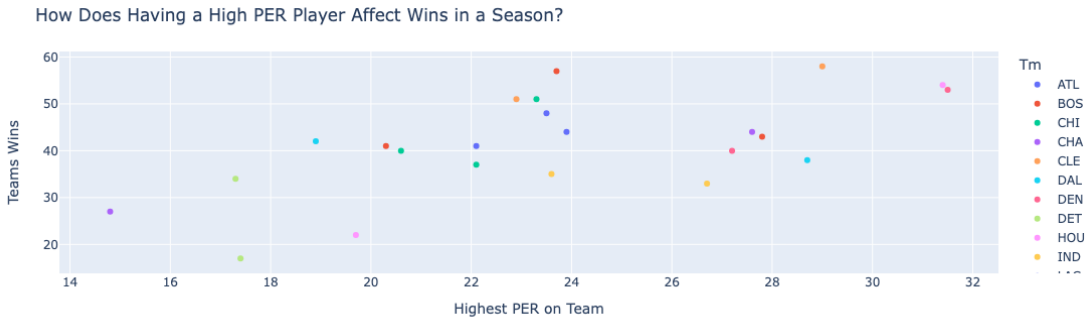
[3]:
```
three_pt_atmpts = pd.read_csv("data/3pa.csv", index_col="Team")
three_pt_atmpts.rename(index=lambda s: s[:-1] if s.endswith("*") else s,␣
↪inplace=True)
```

[4]:
```
abbreviations = pd.read_csv("data/abbreviations.csv")
teams = pd.read_csv("data/teams.csv")

# Rename index and columns to preferred names
abbreviations.rename(columns={"prefix_1": "abbrev"}, inplace=True)

# Regularize abbreviations
abbreviations["abbrev"] = abbreviations["abbrev"].apply(lambda s: s.upper())
```

```
# Drop unneeded column
abbreviations.drop(columns="prefix_2", inplace=True)
abbreviations.loc[abbreviations["abbrev"] == "NO", "name"] = "New Orleans␣
 ↪Pelicans"
abbreviations.loc[abbreviations["abbrev"] == "CHA", "name"] = "Charlotte␣
 ↪Hornets"

teams = teams.merge(abbreviations, left_on="Team", right_on="name", how="outer")
teams.drop(columns=["name", "Rk"], inplace=True)
```

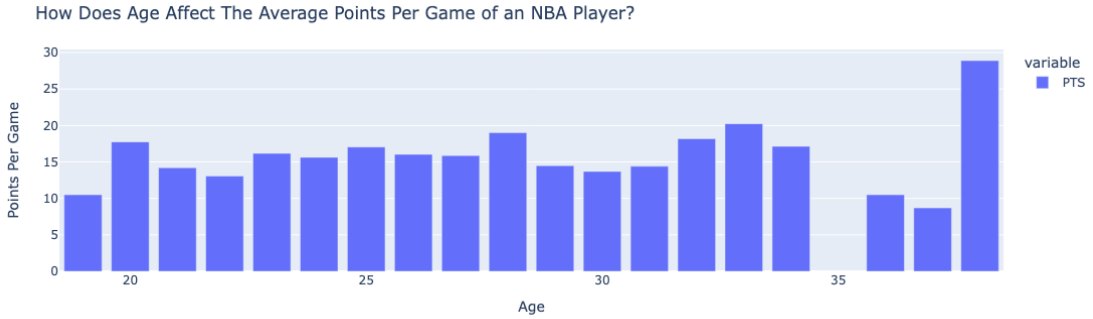
1.2 Question 1: How does player efficiency rating (PER) correlate with a team's win-loss record?

[5]:
```
teams[["wins", "losses"]] = teams["Overall"].str.split("-", expand=True).
 ↪astype(int)
# teams["W/L Ratio"] = round((teams["wins"] / (teams["losses"] +␣
 ↪teams["wins"])), 2)
# WL = teams["W/L Ratio"]
#display(teams)

top_players = per.groupby("Tm")["PER"].max().reset_index()
top_players.loc[top_players["Tm"] == "CHO", "Tm"] = "CHA"
top_players = pd.merge(top_players, teams[["abbrev", "wins"]], left_on="Tm",␣
 ↪right_on="abbrev")
top_players.drop(columns="abbrev", inplace=True)

required_columns = ["Tm", "PER", "wins"]
assert all(column in top_players.columns for column in required_columns),␣
 ↪"Missing columns"

pxo.scatter(top_players, x = "PER", y = "wins", color="Tm").
 ↪update_layout(xaxis_title = "Highest PER on Team", yaxis_title = "Teams␣
 ↪Wins", title="How Does Having a High PER Player Affect Wins in a Season?")
```



1.3 Question 2: What impact does age have on PPG?

[6]:
```
required_columns = ["Player", " Rk", "Pos", "Age", "Tm", "PER", "3PAr", "PTS"]
assert all(column in per.columns for column in required_columns), "Missing␣
 ↪columns"

age_avgs = per.groupby("Age")["PTS"].mean()
age_avgs
pxo.bar(age_avgs).update_layout( yaxis_title = "Points Per Game", title="How␣
 ↪Does Age Affect The Average Points Per Game of an NBA Player?")
```



1.3.1 Follow up questions:

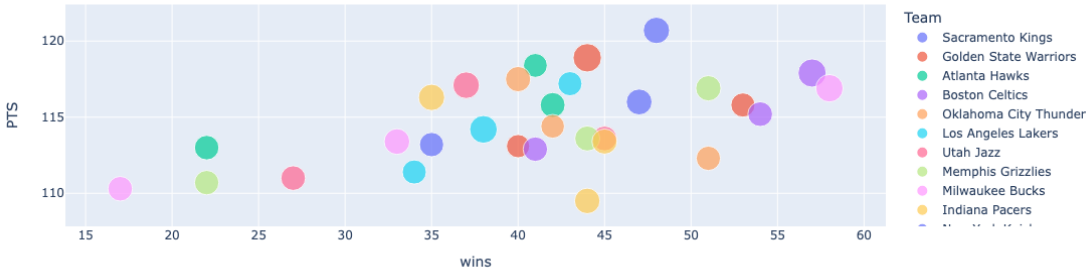
- Could this be more indicative of the impact that ppg has on the longevity of a career?

1.4 Question 3: Does the three-point shot frequency correlate with higher team scoring averages and how does this impact Win/Loss record?

[7]:
```
ThreePointShotFreq = per["3PAr"]
# display(teams)
teamPPG = three_pt_atmpts["PTS"]
wins = teams["wins"]

team_stats = three_pt_atmpts.merge(teams, left_on="Team", right_on="Team")
print(team_stats.columns)
team_stats["3PA"] = team_stats["3PA"]

required_columns = ["Team", "Rk", "G", "FG", "FGA", "FG%", "3P", "3PA", "3P%",␣
 ↪"2P",
                    "2PA", "2P%", "FT", "FTA", "FT%", "AST", "TOV", "PTS",
                    "Overall", "abbrev", "wins", "losses"]
assert all(column in team_stats.columns for column in required_columns),␣
 ↪"Missing columns"
```

```
pxo.scatter(team_stats, x = "wins", y = "PTS", size = "3PA", color = "Team")
```

```
Index(['Team', 'Rk', 'G', 'FG', 'FGA', 'FG%', '3P', '3PA', '3P%', '2P', '2PA',
       '2P%', 'FT', 'FTA', 'FT%', 'AST', 'TOV', 'PTS', 'Overall', 'abbrev',
       'wins', 'losses'],
      dtype='object')
```



1.4.1 Follow up questions:

- Is this more telling of the state of the league and the emphasis there is on the three-point shot?

[ ]: