Python Sport Analytics

SPRING 2025

A COMPARISON OF THE EFFECTS OF WIN SHARES IN HE SEC AND BIG 12

Data used for this project

- Big 12 and SEC
- 2022-23 basketball season (last time Arkansas was good)
- Conference game schedule data only (NOT non-conference)
- data from basketball-reference.com

PYTHON SPORT ANALYTICS

SYRACUSE UNIVERSITY

SPRING 2025

Input

One of the requirements was to allow the user to provide input (I just included a basic "Hello World" type input for now…more to come later).

[474]:
```
# Final Project

#follow output from every line
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
```

[475]:
```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns
```

[476]:
```
print('Enter your name:')
x = input()
print('Hello, ' + x)
```

```
Enter your name:
 Darren
Hello, Darren
```

[477]:
```
print('Nice to meet you,' + x)
```

```
Nice to meet you,Darren
```

Read_csv

- Secondly, I read in the data which I downloaded as csv files for all 24 teams from basketball-reference.com
- Each file contained player data. I also read in team Won loss record data as well (used to calc win pct)

```
BAMA_2023 = pd.read_csv("data/BAMA_2023.csv")
AUB_2023 = pd.read_csv("data/AUB_2023.csv")
ARK_2023 = pd.read_csv("data/ARK_2023.csv")
LSU_2023 = pd.read_csv("data/LSU_2023.csv")
TAM_2023 = pd.read_csv("data/TAM_2023.csv")
MIZZ_2023 = pd.read_csv("data/MIZZ_2023.csv")
MISS_2023 = pd.read_csv("data/MISS_2023.csv")
MSU_2023 = pd.read_csv("data/MSU_2023.csv")
UGA_2023 = pd.read_csv("data/UGA_2023.csv")
TENN_2023 = pd.read_csv("data/TENN_2023.csv")
KEN_2023 = pd.read_csv("data/KEN_2023.csv")
FLA_2023 = pd.read_csv("data/FLA_2023.csv")
USC_2023 = pd.read_csv("data/USC_2023.csv")
VAN_2023 = pd.read_csv("data/VAN_2023.csv")
#BAMA_2023
```

[481]:

```
# Reading in Big 12 Advanced data

TEX_2023 = pd.read_csv("data/TEX_2023.csv")
OKLA_2023 = pd.read_csv("data/OKLA_2023.csv")
OSU_2023 = pd.read_csv("data/OSU_2023.csv")
KAN_2023 = pd.read_csv("data/KAN_2023.csv")
KSU_2023 = pd.read_csv("data/KSU_2023.csv")
TTU_2023 = pd.read_csv("data/TTU_2023.csv")
TCU_2023 = pd.read_csv("data/TCU_2023.csv")
BAY_2023 = pd.read_csv("data/BAY_2023.csv")
ISU_2023 = pd.read_csv("data/ISU_2023.csv")
WVA_2023 = pd.read_csv("data/WVA_2023.csv")
```

Combining data

- To demonstrate the merging of data…
- A) I combined each team's individual cata using the concatenation of data frames into 2 large conference tables
- B) Under the "Awards" section, there wasn't an "NaN" award given in either conference, so we needed to get rid of the NaN (see bottom image)

[483]:
```
# Concatenate all Big 12 player data into one big table
print("Every player from the Big 12 conference.....\n")
BIG12_2023 = pd.concat([KAN_2023, TEX_2023, KSU_2023, BAY_2023, TCU_2023, TTU_2023, ISU_2023, OKLA_2023, OSU_2023, WVA_2023])
BIG12_2023
```

```
Every player from the Big 12 conference.....
```

[483]:

	Rk	Player	Pos	G	GS	MP	PER	TS%	3PAr	FTr	...	OWS	DWS	WS	WS/40	Awards	Player-additional	Team	Wins	Losses	pct
0	1	Dajuan Harris Jr.	G	18	18	648	16.8	0.523	0.311	0.119	...	1.4	0.4	1.7	0.106	NaN	dajuan-harris-1	Kansas	13	5	0.722222
1	2	Jalen Wilson	F	18	18	648	20.3	0.525	0.384	0.387	...	1.6	0.5	2.1	0.128	AA-1AP-AA-1NABC-AA-1SN-AA-1USBWA-AA-1W-AA-1	jalen-wilson-1	Kansas	13	5	0.722222
2	3	Gradey Dick	G	18	18	599	18.0	0.562	0.547	0.276	...	1.5	0.4	1.9	0.127	NaN	gradey-dick-1	Kansas	13	5	0.722222
3	4	Kevin McCullar Jr.	G	18	18	573	17.3	0.531	0.345	0.549	...	1.1	0.7	1.7	0.121	NaN	kevin-mccullar-1	Kansas	13	5	0.722222
4	5	KJ Adams Jr.	F	18	18	509	19.6	0.609	0.015	0.515	...	1.6	0.2	1.8	0.139	NaN	kj-adamsjr-1	Kansas	13	5	0.722222
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

[487]:
```
print("Every player from the S-E-C.....\n")
SEC_2023
```

```
Every player from the S-E-C.....
```

[487]:

	Rk	Player	Pos	G	GS	MP	PER	TS%	3PAr	FTr	...	OWS	DWS	WS	WS/40	Awards	Player-additional	Team	Wins	Losses	pct
0	1	Brandon Miller	F	18	18	598	28.1	0.635	0.552	0.316	...	2.9	1.2	4.1	0.271	AA-2AP-AA-1NABC-AA-2SN-AA-1USBWA-AA-2W-AA-2	brandon-miller-3	Alabama	16	2	0.888889
1	2	Mark Sears	G	18	18	537	19.8	0.597	0.588	0.45	...	1.7	0.9	2.7	0.198		mark-sears-1	Alabama	16	2	0.888889
2	3	Noah Clowney	F	18	18	487	20.2	0.585	0.443	0.45	...	1.5	1.0	2.5	0.204		noah-clowney-1	Alabama	16	2	0.888889
3	4	Jaden Bradley	G	18	18	371	13.1	0.477	0.13	0.772	...	0.6	0.5	1.1	0.123		jaden-bradley-1	Alabama	16	2	0.888889
4	5	Charles Bediako	C	18	18	367	17.7	0.585	0.063	0.241	...	0.9	0.8	1.6	0.179		charles-bediako-1	Alabama	16	2	0.888889
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
8	9	Justice Williams	G	13	6	221	6.7	0.427	0.234	0.255	...	0.0	-0.3	-0.2	-0.040		justice-williams-1	LSU	2	16	0.111111
		Shawn															shawn-				

So about the WS data used…..

- The WS_x category represents the average WS data for each team.
- The WS_y data represents the overall SUM of the WS (Win Shares) for each team.

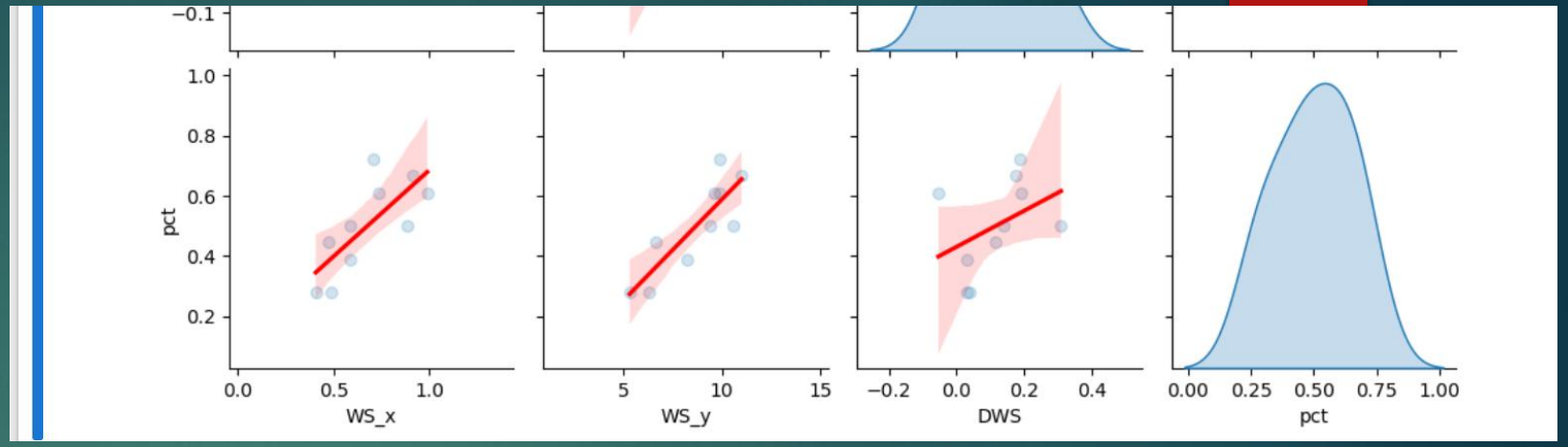
Team statistics for the 2022-23 season - Big 12

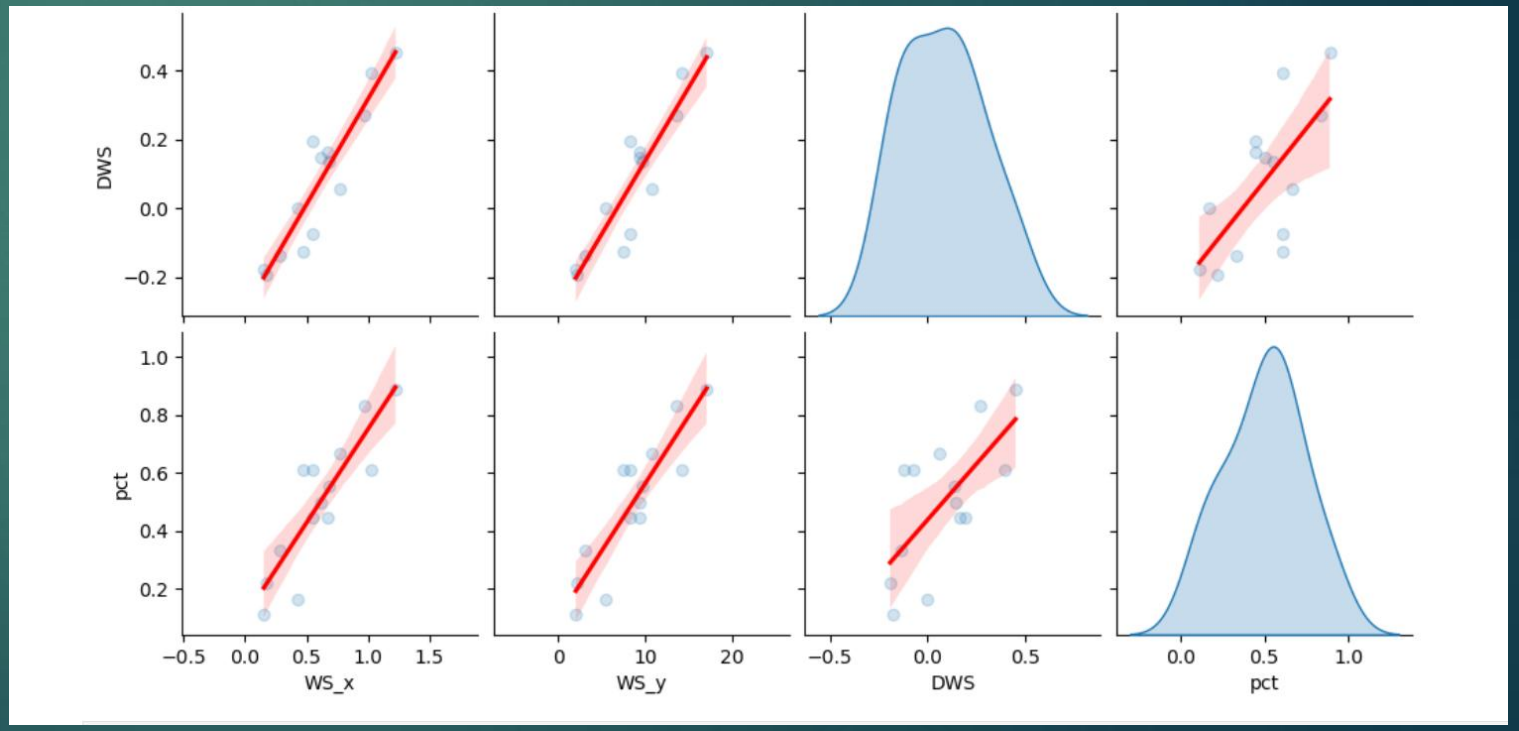
[499]:

Team	WS_x	WS_y	pct	OWS	DWS
Baylor	0.738462	9.6	0.611111	0.800000	-0.053846
Iowa State	0.883333	10.6	0.500000	0.566667	0.308333
Kansas	0.707143	9.9	0.722222	0.535714	0.185714
Kansas State	0.990000	9.9	0.611111	0.810000	0.190000
Oklahoma	0.407692	5.3	0.277778	0.392308	0.038462
Oklahoma State	0.471429	6.6	0.444444	0.364286	0.114286
TCU	0.587500	9.4	0.500000	0.450000	0.137500
Texas	0.916667	11.0	0.666667	0.725000	0.175000
Texas Tech	0.484615	6.3	0.277778	0.469231	0.030769
West Virginia	0.585714	8.2	0.388889	0.585714	0.028571

WS vs. Win %

- Definitively speaking, there is a correlaton between Win shares and win pct. Looking at the graph, there is a more defined graph for the Overall SUM (WS_y) of win shares (slightly) than the avg.
- Why is that? My theory: Most important factor is that this takes into consideration Games played, etc. so if a player has played 16 conference games, weight is heavier than a player that played in 8.





Interactive – Pt2 (Player data)

- Here the user is prompted to input which conference (Big 12 or SEC) for the next portion of the project.
- The idea is to determine of the minimum qualifiers (top 20 or 25 players by WS) do these players correlate to team success.

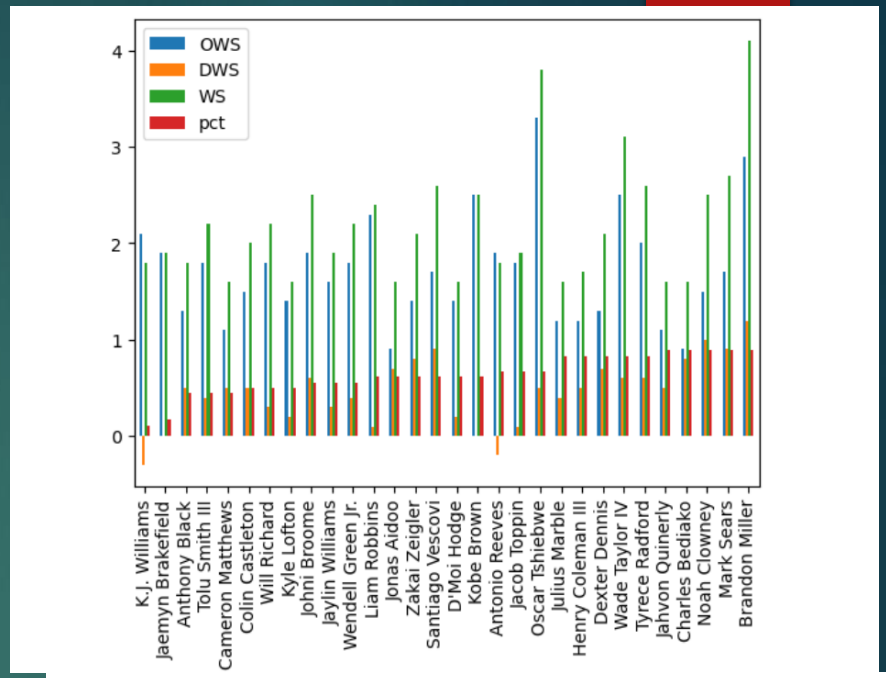
```
print('A part of my Final Project is to determine (visually...using the top 20 players)' + '\n' + 'the correlation between individual WS factors and w
print('Which Conference would you like to see final analysis?' + '\n' )
print('Type B for Big 12 Conference....Type S for Southeastern Conference...U for both\n')
y = input()

if y == 'B':
    print("\n\nYou chose the Big 12 Conference.\n\n")
    BIG12_2023_24[['OWS','DWS','WS','pct']].sort_values(by='pct').plot.bar()
    print("Here are the top 20 players by WS for 2023. \nThere appears to be no correlation AT ALL with any of the categories and win percentage.")
elif y == 'S':
    print("\n\nYou chose the S-E-C.\n\n")
    SEC_2023_24[['OWS','DWS','WS','pct']].sort_values(by='pct').plot.bar()
    print("Here are the top 20 players by WS for 2023. \nUnlike the Big 12, There appears to be a slight correlation \nbetween the DWS and win percent
elif y == 'U':
    print("\n\nYou chose both. Look and compare.\n\n")
    BIG12_2023_24[['OWS','DWS','WS','pct']].sort_values(by='pct').plot.bar()
    SEC_2023_24[['OWS','DWS','WS','pct']].sort_values(by='pct').plot.bar()
    print("The correlation between DWS and WS with win pct is a little stronger in the SEC than the Big 12")
else: print("Invalid Choice. ")
```

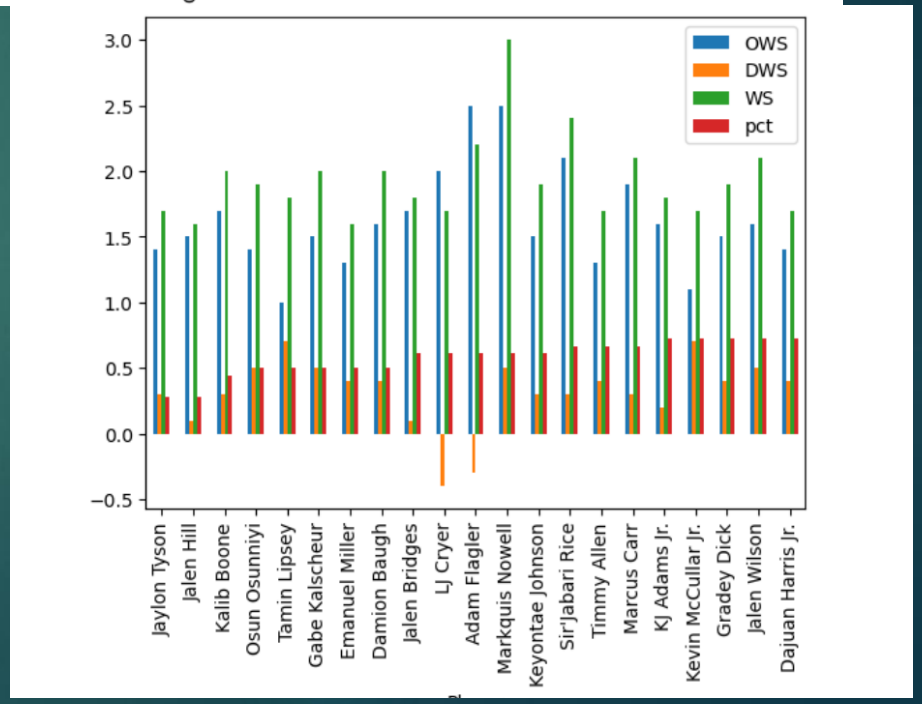
```
A part of my Final Project is to determine (visually...using the top 20 players)
the correlation between individual WS factors and win pct?

Which Conference would you like to see final analysis?

Type B for Big 12 Conference....Type S for Southeastern Conference...U for both

 S
```

Individual players:

Good players make up good teams…..
…..most of the time(?)

- In this study, It is important to note that only the 16-game conference schedule for each conference was used.
- There was a minimum (1.5) WS used where 21 Big 12 and 25 SEC players qualified
- There appears to be no connection AT ALL in the Big 12 between any of the Win share components used. (The data is sorted by win-pct – smallest to largest).
- However, in the SEC, the DWS has a moderate correlation, and the WS appears to slightly coincide with the win percentage.





Thanks for viewing