USC Football Defense Analysis

Data Visualization Project

Suraj Swarup

Motivation

The goal of this data visualization project was to compare the statistics of USC Football's Defense under Pete Carroll (starting in 2002 since that is the year that is considered the start of the dynastic era under him) to the defense under every head coach that coached at least a full season since Carroll left the program. Though the program has never reached the heights that it did under Carroll, there have been many seasons where the team's offense has been great, putting up a lot of points and even producing a Heisman Trophy Winner. The saying "defense wins championships" is often mentioned in sports, and for the most part is true. This analysis will look further into that and show the stark differences in defensive performance at the program under Carroll compared to the coaches that have lead the program since his departure.

Data

I manually pulled each game's stats using https://www.sports-reference.com/

```
data = pd.read_csv('USCDefense.csv')
data.head()
```

	Game	Year	Opponent	Conference	GameType	PA	Result	Coach	Pyd	Ryd	TotYd	Turnovers
0	1	2002	Auburn	SEC	OutConference	17	W	Carroll	140	141	281	2
1	2	2002	Colorado	Big12	OutConference	3	W	Carroll	24	37	61	2
2	3	2002	Kansas State	Big12	OutConference	27	L	Carroll	159	188	347	5
3	4	2002	Oregon State	Pac	Conference	0	W	Carroll	80	51	131	3
4	5	2002	Washington State	Pac	Conference	30	L	Carroll	315	201	516	2

Creating Data Frame For Each Coach

I decided to take each coach's performances and find out their wins and losses during their tenure as head coach, in conference games, out of conference games, and in bowl games. Note that for Clay Helton, I grouped the final 10 games of the 2021 season under his statistics even though he was fired mid-season. I also did not analyze Ed Orgeron's stats, as he was not even head coach for a full season. For Pete Carroll, I started the data collection from the 2002 season instead of the 2001 season (when Carroll was first hired at USC) because I wanted this analysis to focus on the statistics of the defenses under USC's dynastic era of college football, and more importantly to compare the defensive stats from then to how they have been under each coach that has succeeded Carroll.

Creating Data Frame For Each Coach (Code)

This example is for Pete Carroll, but similar code was ran for each other coach

```
PeteData = data.loc[data['Coach'] == 'Carroll']
#PeteData
```

```
RegCounts = PeteData.Result.value_counts().reset_index().rename(columns={"index": "Result", "Result": "Count"})
print(RegCounts)

PeteConfData = PeteData.loc[PeteData['Conference'] == 'Pac']
#PeteConfData
ConfCounts = PeteConfData.Result.value_counts().reset_index().rename(columns={"index": "Result", "Result": "Count"})
print(ConfCounts)

PeteOutConferenceData = PeteData.loc[PeteData['GameType'] == 'OutConference']
#PeteOutConferenceData
OutConferenceCounts = PeteOutConferenceData.Result.value_counts().reset_index().rename(columns={"index": "Result",
                                                                                                 "Result": "Count"})
print(OutConferenceCounts)

PeteBowlData = PeteData.loc[PeteData['GameType'] == 'Bowl']
#PeteBowlData
BowlCounts = PeteBowlData.Result.value_counts().reset_index().rename(columns={"index": "Result", "Result": "Count"})
print(BowlCounts)
```

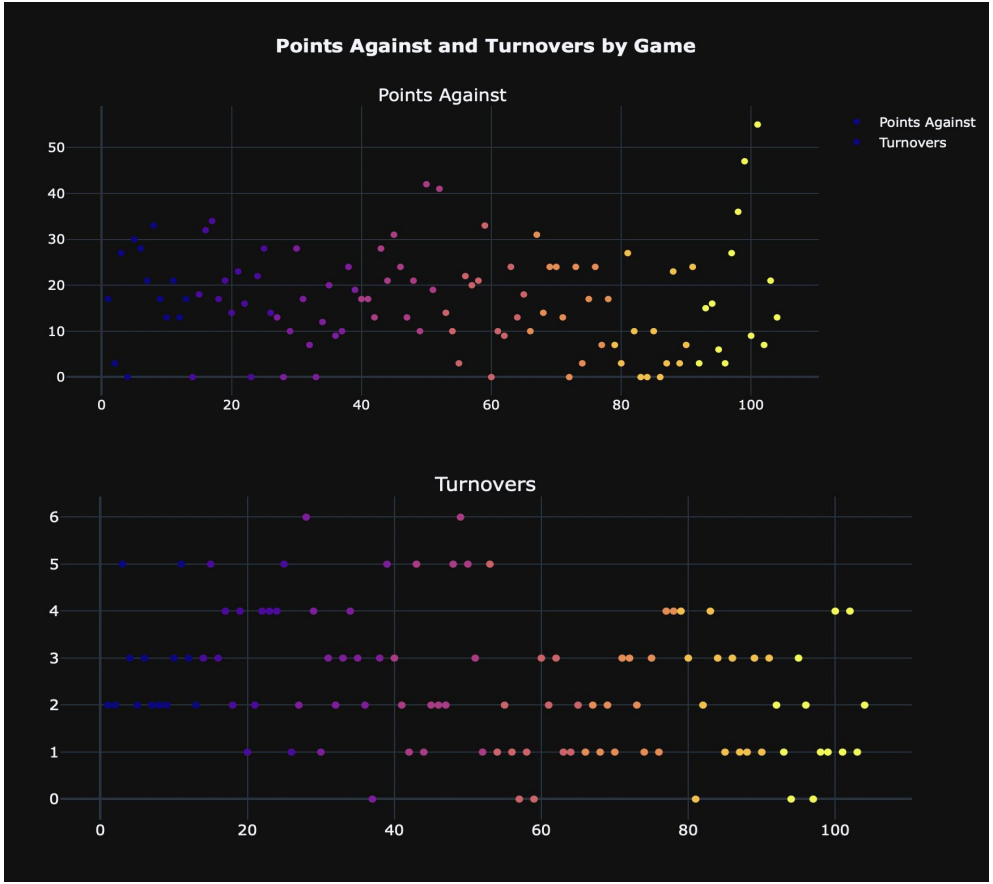
```
  Result  Count
0      W     91
1      L     13
  Result  Count
0      W     57
1      L     11
  Result  Count
0      W     27
1      L      1
  Result  Count
0      W      7
1      L      1
```

Carroll Season Analysis

These scatterplots show Points Against and Turnovers on the y axis, with each game number on the x axis. The colors of the points correspond to the year. When looking at the defense stats under Pete Carroll, we can see that his teams allowed the fewest points in the 2007 and 2008 seasons, but overall were stellar from 2002-2008. In 2009, we can see that the points against scatter plot has points that are higher on the y axis. In terms of turnovers, most games his team would force between 1-4 turnovers, but we can see the 2004 and 2005 teams had games where they forced 6 turnovers. The code for this is shown below.

```
fig=make_subplots(rows=2,cols=1,subplot_titles=('Points Against', 'Turnovers'))
fig.add_trace(go.Scatter(x = PeteData["Game"], y=PeteData["PA"],
                         mode = 'markers', marker_color = PeteData["Year"], text = PeteData["Year"],
                         name = "Points Against"), row = 1, col = 1)
fig.add_trace(go.Scatter(x = PeteData["Game"], y=PeteData["Turnovers"],
                         mode = 'markers', marker_color = PeteData["Year"], text = PeteData["Year"],
                         name = "Turnovers"), row=2,col=1)

fig.update_layout(height=900,width=900,title_text='<b>Points Against and Turnovers by Game')
fig.update_layout(template='plotly_dark',title_x=0.5)
```



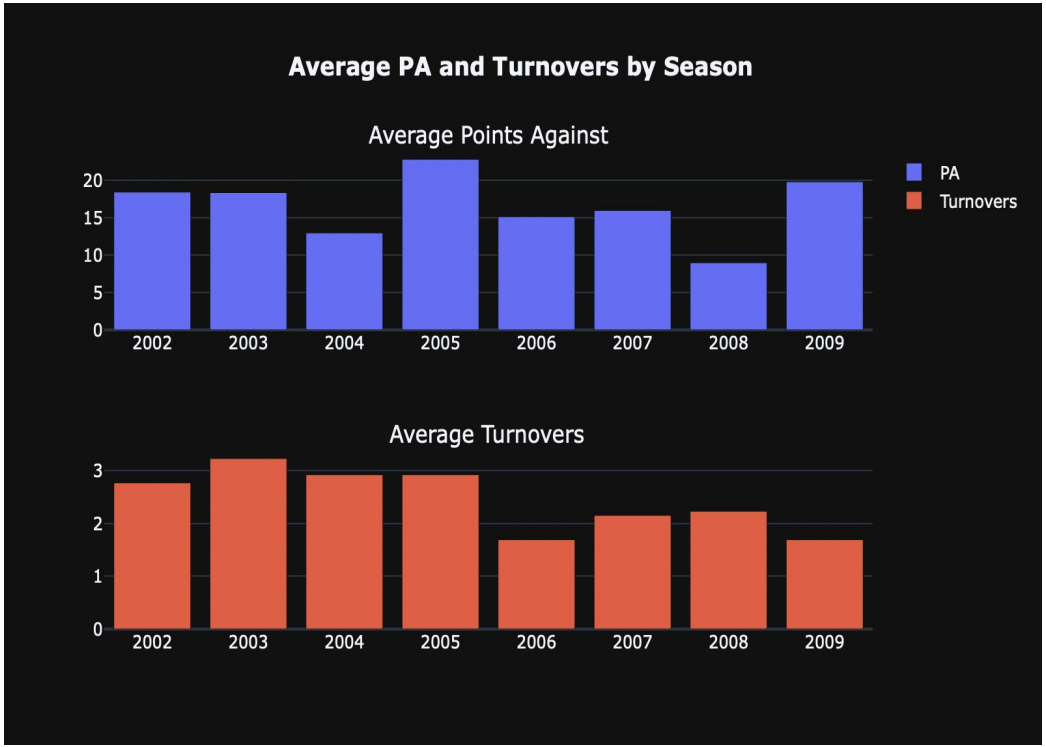
Carroll Season Analysis Cont.

This data frame shows the average of points against per game, pass yards given up per game, rushing yards given up per game, total yards given up per game, and turnovers forced per game for each season under Pete Carroll. We can see his first few seasons the defense forced more turnovers than later in his tenure, but also the team became a lot better at limiting yards towards the back end of his tenure as coach. The 2005 and 2009 teams were the only ones that allowed an average of over 19 points per game and were 2 of the 3 seasons where the defense allowed over 340 yards of total offense per game as well.

	Year	Game	PA	Pyd	Ryd	TotYd	Turnovers
0	2002	7.0	18.461538	201.692308	84.000000	285.692308	2.769231
1	2003	20.0	18.384615	276.230769	64.000000	340.230769	3.230769
2	2004	33.0	13.000000	207.615385	80.769231	288.384615	2.923077
3	2005	46.0	22.846154	230.384615	130.538462	360.923077	2.923077
4	2006	59.0	15.153846	204.769231	93.538462	298.307692	1.692308
5	2007	72.0	16.000000	189.000000	84.153846	273.153846	2.153846
6	2008	85.0	9.000000	134.384615	87.384615	221.769231	2.230769
7	2009	98.0	19.846154	212.307692	128.153846	340.461538	1.692308

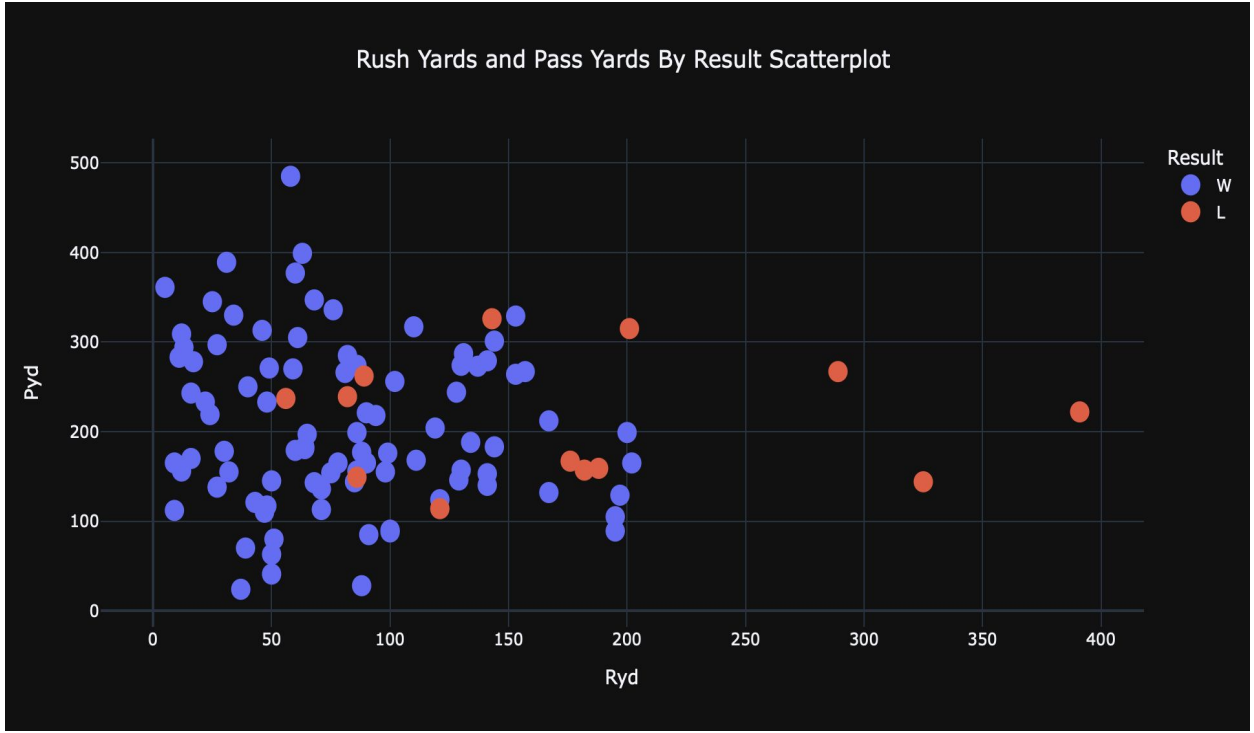
Carroll Season Analysis Cont.

```
fig=make_subplots(rows=2,cols=1, #shared_yaxes = 'all',
                  subplot_titles=('Average Points Against', 'Average Turnovers'))
fig.add_trace(go.Bar(x = y1["Year"], y=y1["PA"], name = "PA"), row=1,col=1)
fig.add_trace(go.Bar(x = y1["Year"], y=y1["Turnovers"], name = "Turnovers"),row=2,col=1)
fig.update_layout(height=500,width=800,title_text='<b>Average PA and Turnovers by Season')
fig.update_layout(template='plotly_dark',title_x=0.5)
fig.show()
```



Carroll Season Analysis Cont.

This scatterplot shows the pass yards and rush yards given up by game, and the colors correspond to whether USC won that game or not. As we can see, USC only won one game under Pete Carroll when they gave up over 200 yards rushing. Additionally, USC only lost 6 games under Pete Carroll when they gave up less than 200 yards rushing. This checks out, as dominating teams at the line of scrimmage is usually a recipe for winning games. The code for the plot is shown on the next slide.

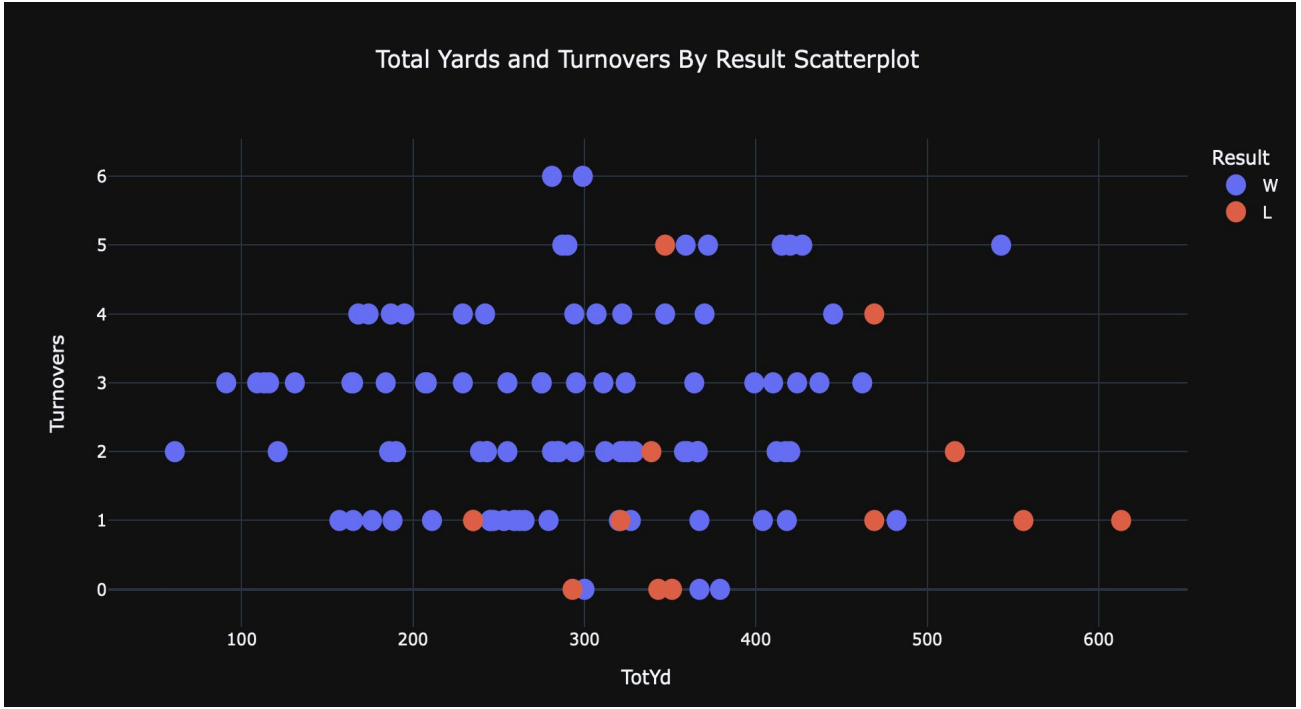


Carroll Season Analysis Cont. (Code)

```
# Scatter Plot
fig = px.scatter(PeteData, x="Ryd", y="Pyd", color="Result", hover_name = "Opponent",
                 title="Rush Yards and Pass Yards By Result Scatterplot")
fig.update_traces(marker_size=15)
fig.update_layout(template='plotly_dark',title_x=0.5)
fig.show()
```

Carroll Season Analysis Cont.

This scatterplot shows us the total yards and turnovers given up each game, and whether USC won that game. As we can see, USC only lost twice under Pete Carroll when forcing 3 or more turnovers and this included games in which they allowed nearly 500 yards of total offense. This highlights the importance of forcing turnovers and how that is a key to victory. The code for this is shown on the next slide.



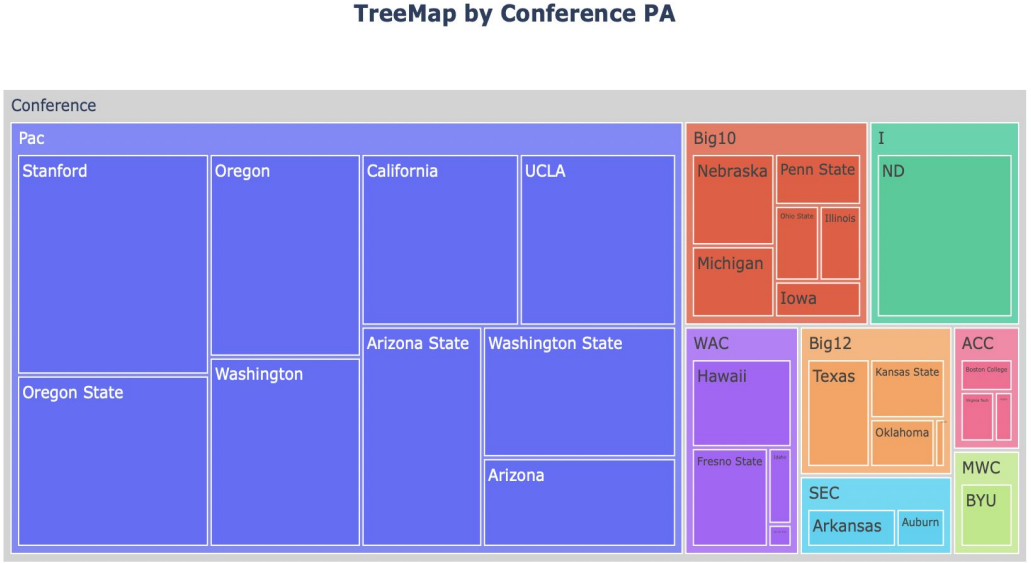
Carroll Season Analysis Cont. (Code)

```
# Scatter Plot
fig = px.scatter(PeteData, x="TotYd", y="Turnovers", color="Result",
                 hover_name = "Opponent", title="Total Yards and Turnovers By Result Scatterplot")
fig.update_traces(marker_size=15)
fig.update_layout(template='plotly_dark',title_x=0.5)
fig.show()
```

Carroll Season Analysis Cont. Tree Map

This tree map was just for fun—it shows the total points against by each opponent under Pete Carroll divided by conference. It was interesting to see that USC has given up the largest amount of total points to Stanford and Oregon State out of all opponents under Pete Carroll. Note that all the Pac-12 teams have such high values because USC is a member of that conference and plays those teams more often than they play teams from other conferences.

```
# TreeMap
fig=px.treemap(PeteData,path=[px.Constant('Conference'),'Conference','Opponent'],
               values='PA',title='<b>TreeMap by Conference PA')
fig.update_traces(root_color='lightgray')
fig.update_layout(title_x=0.5)
```



Coach Comparison

The next section of the analysis will include comparisons of the team defenses under each of the head coaches at USC since (and including) Pete Carroll.

```
confavg = data.groupby(['Conference', 'Coach']).mean().round(2).reset_index()
```

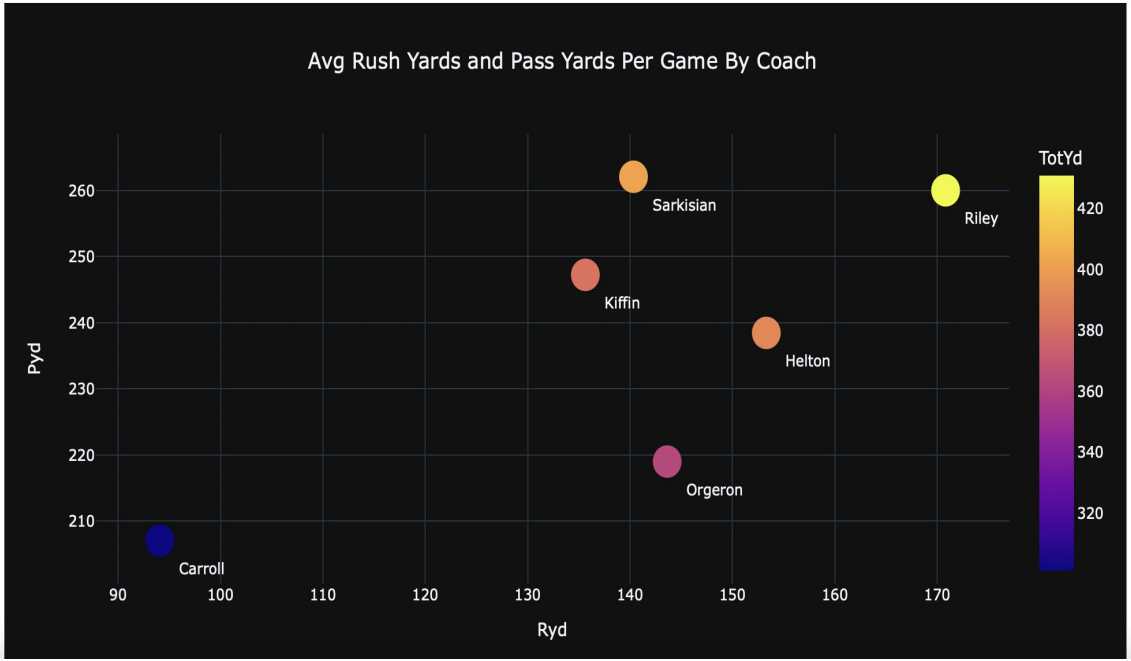
```
coachavg = data.groupby(['Coach']).mean().round(2).reset_index()
coachavg
```

	Coach	Game	PA	Pyd	Ryd	TotYd	Turnovers
0	Carroll	52.50	16.59	207.05	94.07	301.12	2.45
1	Helton	213.28	27.71	238.45	153.30	391.75	1.52
2	Kiffin	126.00	24.47	247.23	135.63	382.86	1.93
3	Orgeron	151.50	21.38	219.00	143.62	362.62	1.62
4	Riley	266.50	31.46	260.00	170.81	430.81	1.65
5	Sarkisian	165.50	23.00	262.06	140.33	402.39	1.78

Coach Comparison Cont.

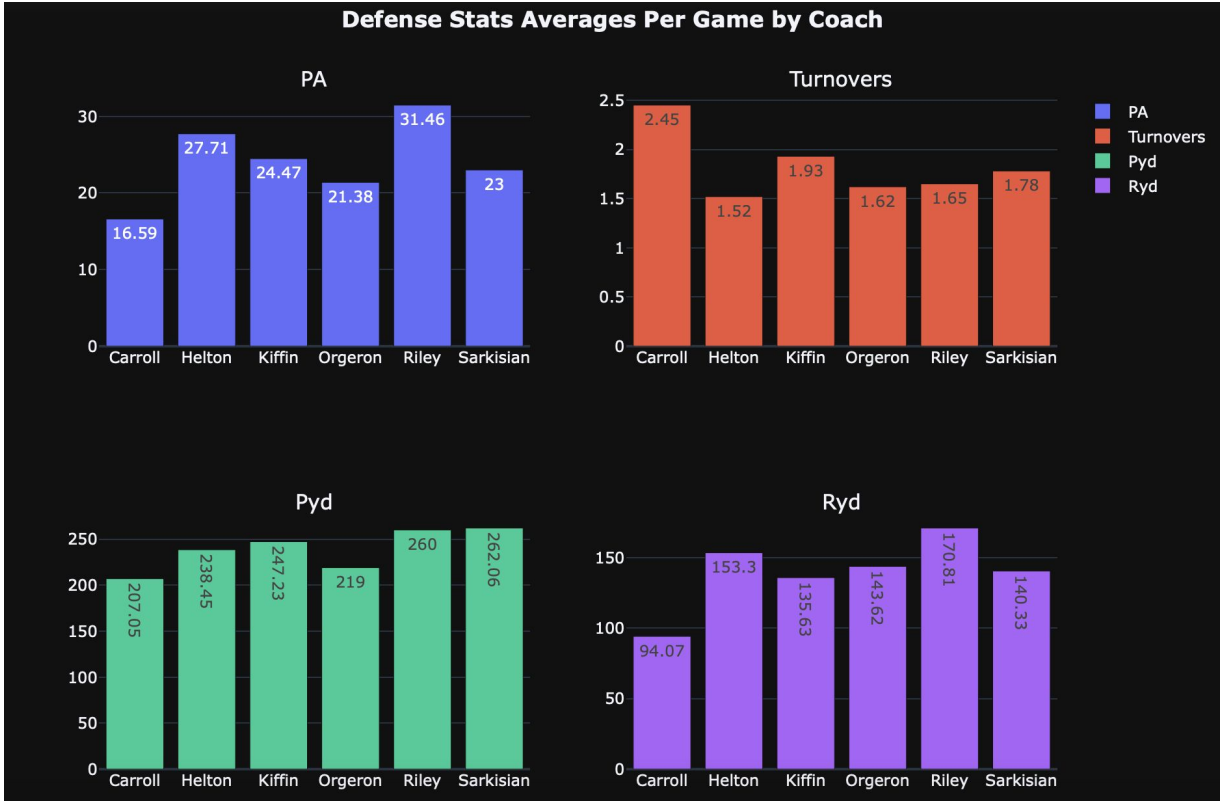
This scatterplot was eye opening. At the very bottom left giving up the least amount of passing and rushing yards per game is Pete Carroll. For coaches that have coached at least a full season at USC since him, Carroll's teams were averaging a minimum 30 fewer passing yards and a minimum of 40 fewer rushing yards per game compared to the other coaches. The worst point on the graph belongs to Lincoln Riley whose teams give up by far the most amount of rushing yards per game and roughly the same high amount of passing yards per game as Steve Sarkisian's teams. This should come as no surprise to USC Football fans, as Lincoln Riley parted ways with his defensive coordinator due to the team's poor performance on that side of the ball.

```
fig=px.scatter(coachavg,x="Ryd",y="Pyd",color="TotYd",text="Coach",title="Avg Rush and Pass Yards Per Game By Coach")
fig.update_traces(textposition = "bottom right")
fig.update_traces(marker_size = 25)
fig.update_layout(template='plotly_dark',title_x=0.5)
fig.show()
```



Coach Comparison Cont.

These bar plots show each of the following stats for each coach: average points against, average turnovers, average passing yards given up, and average rushing yards given up all per game. Pete Carroll teams were far superior at preventing teams from scoring, averaging nearly 7 fewer points per game than the next best defense a coach has led (Orgeron did not qualify as he did not coach a full season). We can also see how successful Carroll lead teams were at forcing turnovers compared to the other coaches. All coaches that have followed Carroll have struggled leading successful defenses—most of their stats are similar except for Riley's teams giving up over an astonishing 30 points per game. All in all, the gap between the success of USC's defense at their peak compared to now is quite severe. This is backed by both the data in the charts and the "eye test" of what is seen during games. The code for this plot is shown on the next slide.



Coach Comparison Cont. (Code)

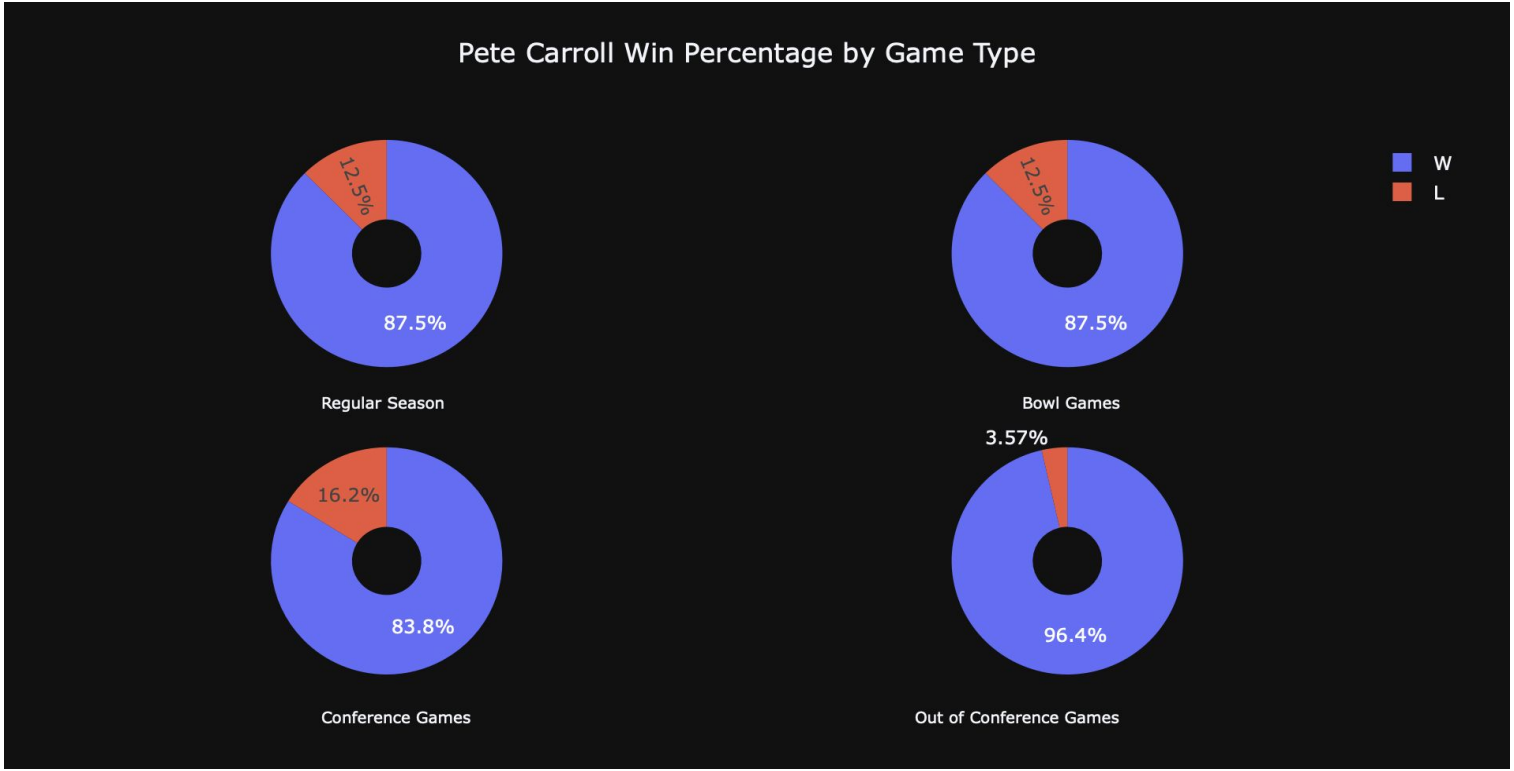
```
fig=make_subplots(rows=2,cols=2,subplot_titles=('PA', 'Turnovers', 'Pyd','Ryd'))
fig.add_trace(go.Bar(x = coachavg["Coach"], y=coachavg["PA"], name = "PA", text = coachavg["PA"]), row=1,col=1)
fig.add_trace(go.Bar(x = coachavg["Coach"], y=coachavg["Turnovers"], name = "Turnovers",
                     text = coachavg["Turnovers"]),row=1,col=2)
fig.add_trace(go.Bar(x = coachavg["Coach"], y=coachavg["Pyd"], name = "Pyd", text = coachavg["Pyd"]), row=2,col=1)
fig.add_trace(go.Bar(x = coachavg["Coach"], y=coachavg["Ryd"], name = "Ryd", text = coachavg["Ryd"]), row=2,col=2)

fig.update_layout(height=700,width=950,title_text='<b> Defense Stats Averages Per Game by Coach')
fig.update_layout(template='plotly_dark',title_x=0.5)
fig.show()
```

Coach Win Percentage by Game Type

The last visualizations for this analysis show each coach and their winning percentages overall, in conference games, out of conference games, and in bowl games. We can see that Pete Carroll had a winning percentage of nearly 90% which is remarkable. The next best coach—Lincoln Riley—is not even winning 70% of his games. Another stark difference is in the performance in conference games. Pete Carroll won over 80% of conference games, whereas of the coaches that followed him only Clay Helton and Lincoln Riley won over 60% of conference games. Lastly, the bowl game performances are slightly misleading, as Steve Sarkisian, Lane Kiffin, and at the time Lincoln Riley, only coached 1 bowl game so their 100%, 0%, and 0% winning percentages respectively in bowl games should be taken with a grain of salt.

Pete Carroll Win Percentage by Game Type



Pete Carroll Win Percentage by Game Type (Code)

Pete Carroll

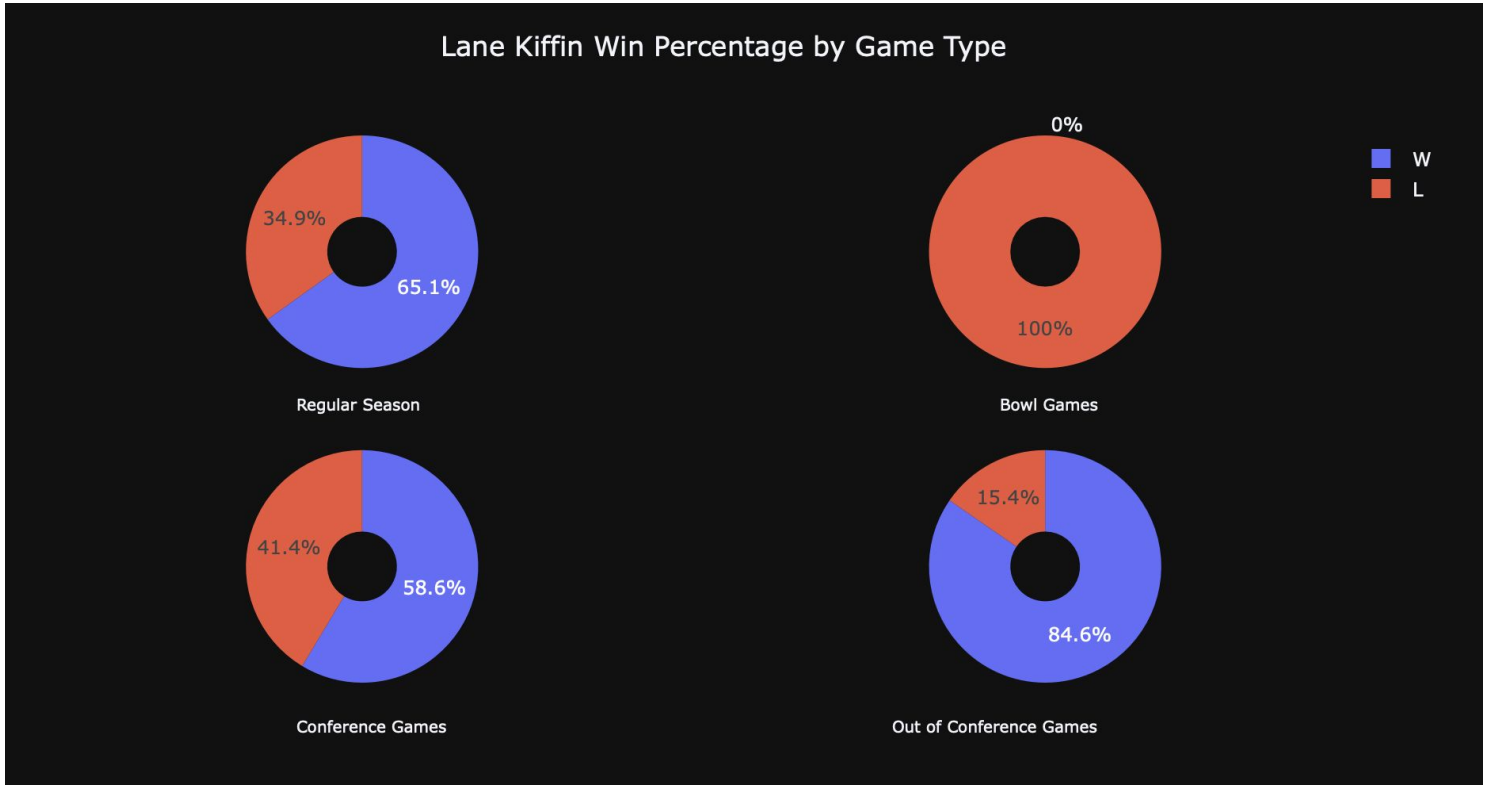
```
specs = [[{'type':'domain'}, {'type':'domain'}], [{'type':'domain'}, {'type':'domain'}]]
fig = make_subplots(rows=2, cols=2, specs=specs)

fig.add_trace(go.Pie(labels=RegCounts.Result, values=RegCounts.Count, name="Regular Season"),
              row = 1, col= 1)
fig.add_trace(go.Pie(labels=BowlCounts.Result, values=BowlCounts.Count, name="Bowl Games"),
              1, 2)
fig.add_trace(go.Pie(labels=ConfCounts.Result, values=ConfCounts.Count, name="Conference Games"),
              2, 1)
fig.add_trace(go.Pie(labels=OutConferenceCounts.Result, values=OutConferenceCounts.Count,
                     name="Out of Conference Games"), 2, 2)

fig.update_traces(hole=.3, hoverinfo="label+percent+name")

fig.update_layout(
    title_text="Pete Carroll Win Percentage by Game Type",
    annotations=[dict(text='Regular Season', x=0.17, y=0.51, font_size=10, showarrow=False),
                 dict(text='Bowl Games', x=0.82, y=0.51, font_size=10, showarrow=False),
                dict(text='Conference Games', x=.17, y=-.1, font_size=10, showarrow=False),
                dict(text='Out of Conference Games', x=0.82, y=-.1, font_size=10, showarrow=False)])
fig.update_layout(template='plotly_dark',title_x=0.5)
fig.show()
```

Lane Kiffin Win Percentage by Game Type



Lane Kiffin Win Percentage by Game Type (Code)

Lane Kiffin

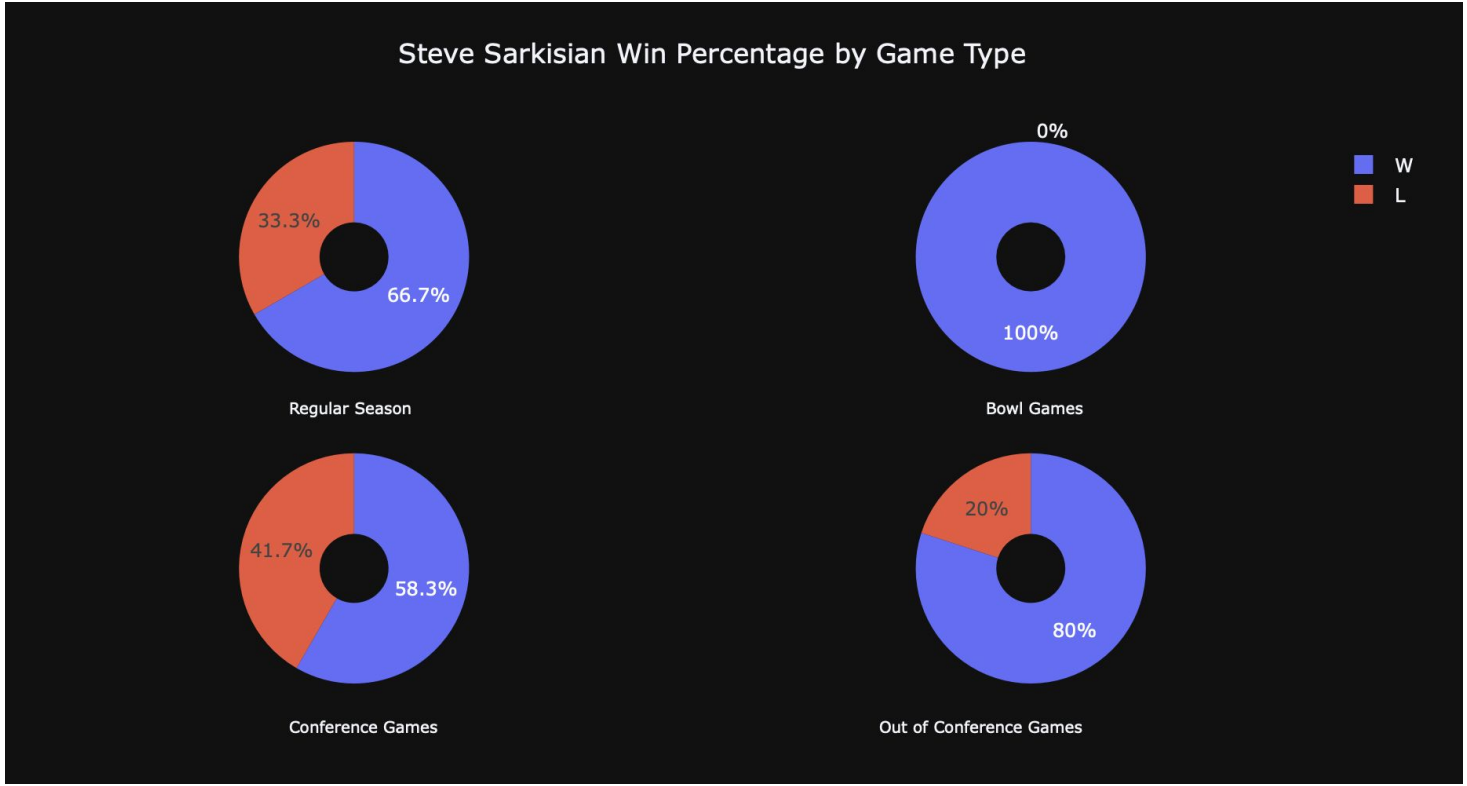
```
specs = [[{'type':'domain'}, {'type':'domain'}], [{'type':'domain'}, {'type':'domain'}]]
fig = make_subplots(rows=2, cols=2, specs=specs)

fig.add_trace(go.Pie(labels=KiffinRegCounts.Result, values=KiffinRegCounts.Count, name="Regular Season"),
              row = 1, col= 1)
fig.add_trace(go.Pie(labels=KiffinBowlCounts.Result, values=KiffinBowlCounts.Count, name="Bowl Games"),
              1, 2)
fig.add_trace(go.Pie(labels=KiffinConfCounts.Result, values=KiffinConfCounts.Count, name="Conference Games"),
              2, 1)
fig.add_trace(go.Pie(labels=KiffinOutConferenceCounts.Result, values=KiffinOutConferenceCounts.Count,
                     name="Out of Conference Games"), 2, 2)

fig.update_traces(hole=.3, hoverinfo="label+percent+name")

fig.update_layout(
    title_text="Lane Kiffin Win Percentage by Game Type",
    annotations=[dict(text='Regular Season', x=0.17, y=0.51, font_size=10, showarrow=False),
                 dict(text='Bowl Games', x=0.82, y=0.51, font_size=10, showarrow=False),
                dict(text='Conference Games', x=.17, y=-.1, font_size=10, showarrow=False),
                dict(text='Out of Conference Games', x=0.82, y=-.1, font_size=10, showarrow=False)])
fig.update_layout(template='plotly_dark',title_x=0.5)
fig.show()
```

Steve Sarkisian Win Percentage by Game Type



Steve Sarkisian Win Percentage by Game Type (Code)

Steve Sarkisian

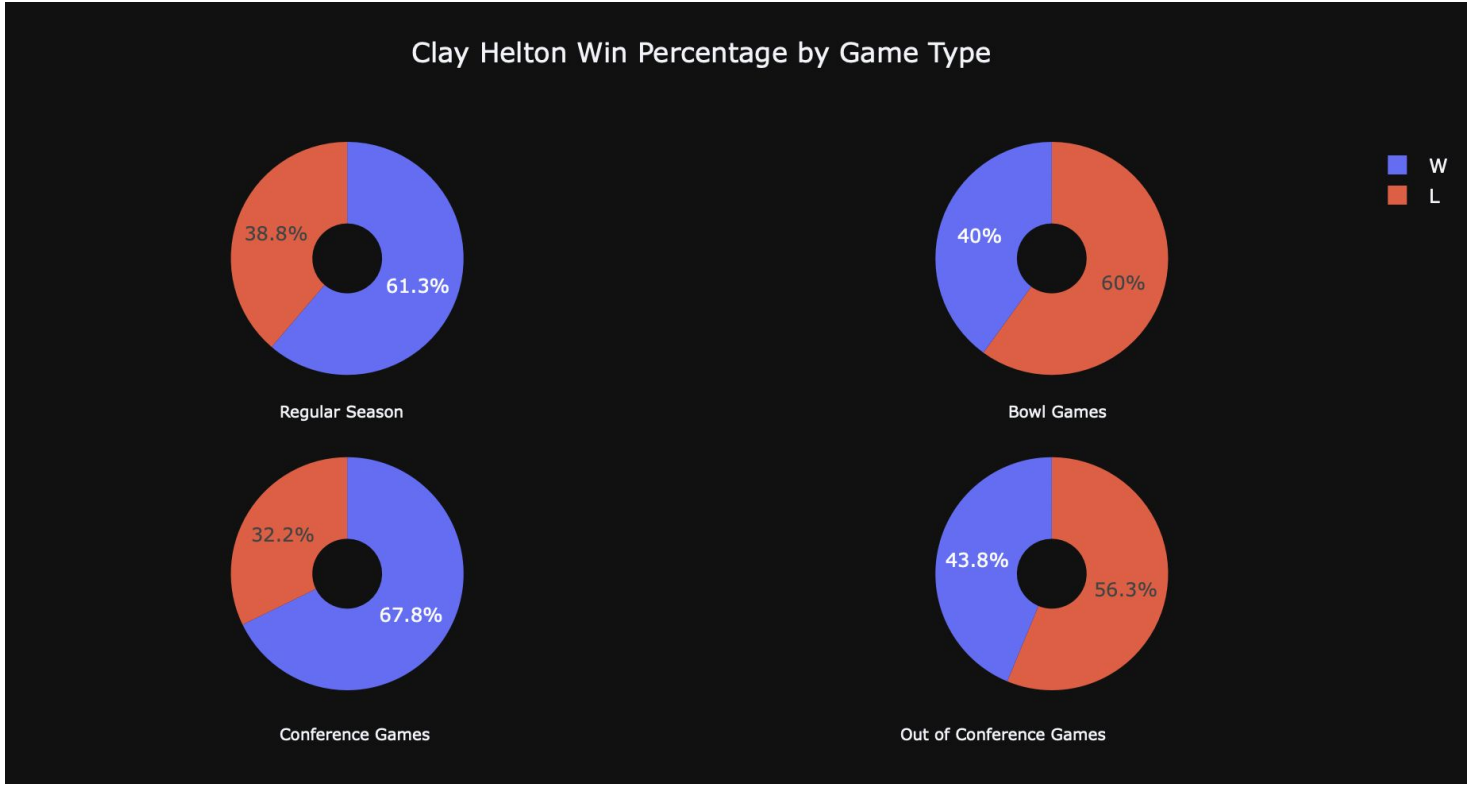
```
specs = [[{'type':'domain'}, {'type':'domain'}], [{'type':'domain'}, {'type':'domain'}]]
fig = make_subplots(rows=2, cols=2, specs=specs)

fig.add_trace(go.Pie(labels=SarkRegCounts.Result, values=SarkRegCounts.Count, name="Regular Season"),
              row = 1, col= 1)
fig.add_trace(go.Pie(labels=SarkBowlCounts.Result, values=SarkBowlCounts.Count, name="Bowl Games"),
              1, 2)
fig.add_trace(go.Pie(labels=SarkConfCounts.Result, values=SarkConfCounts.Count, name="Conference Games"),
              2, 1)
fig.add_trace(go.Pie(labels=SarkOutConferenceCounts.Result, values=SarkOutConferenceCounts.Count,
                     name="Out of Conference Games"), 2, 2)

fig.update_traces(hole=.3, hoverinfo="label+percent+name")

fig.update_layout(
    title_text="Steve Sarkisian Win Percentage by Game Type",
    annotations=[dict(text='Regular Season', x=0.17, y=0.51, font_size=10, showarrow=False),
                 dict(text='Bowl Games', x=0.82, y=0.51, font_size=10, showarrow=False),
                dict(text='Conference Games', x=.17, y=-.1, font_size=10, showarrow=False),
                dict(text='Out of Conference Games', x=0.82, y=-.1, font_size=10, showarrow=False)])
fig.update_layout(template='plotly_dark',title_x=0.5)
fig.show()
```

Clay Helton Win Percentage by Game Type



Clay Helton Win Percentage by Game Type (Code)

Clay Helton

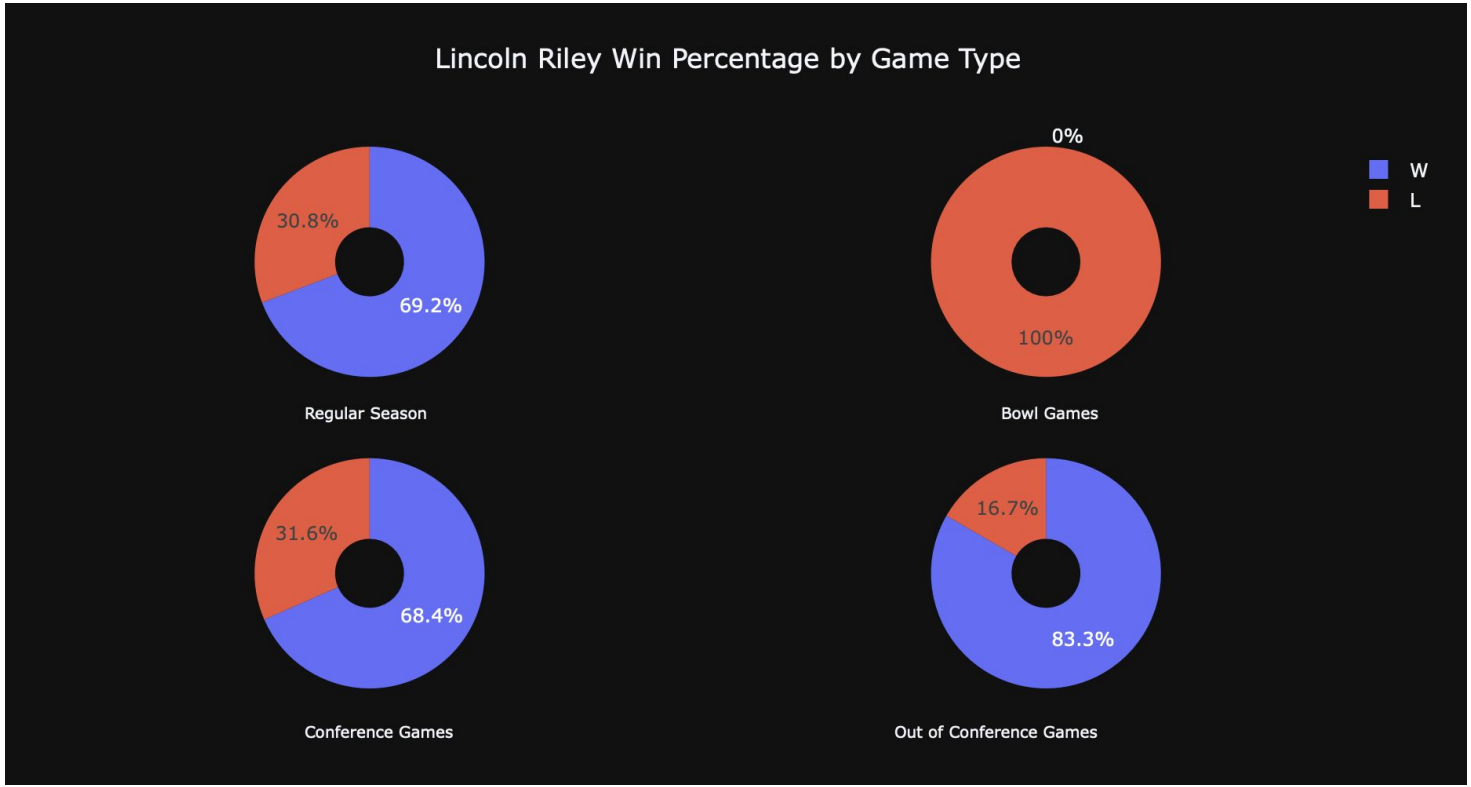
```
specs = [[{'type':'domain'}, {'type':'domain'}], [{'type':'domain'}, {'type':'domain'}]]
fig = make_subplots(rows=2, cols=2, specs=specs)

fig.add_trace(go.Pie(labels=HeltonRegCounts.Result, values=HeltonRegCounts.Count, name="Regular Season"),
              row = 1, col= 1)
fig.add_trace(go.Pie(labels=HeltonBowlCounts.Result, values=HeltonBowlCounts.Count, name="Bowl Games"),
              1, 2)
fig.add_trace(go.Pie(labels=HeltonConfCounts.Result, values=HeltonConfCounts.Count, name="Conference Games"),
              2, 1)
fig.add_trace(go.Pie(labels=HeltonOutConferenceCounts.Result, values=HeltonOutConferenceCounts.Count,
                     name="Out of Conference Games"), 2, 2)

fig.update_traces(hole=.3, hoverinfo="label+percent+name")

fig.update_layout(
    title_text="Clay Helton Win Percentage by Game Type",
    annotations=[dict(text='Regular Season', x=0.17, y=0.51, font_size=10, showarrow=False),
                 dict(text='Bowl Games', x=0.82, y=0.51, font_size=10, showarrow=False),
                dict(text='Conference Games', x=.17, y=-.1, font_size=10, showarrow=False),
                dict(text='Out of Conference Games', x=0.82, y=-.1, font_size=10, showarrow=False)])
fig.update_layout(template='plotly_dark',title_x=0.5)
fig.show()
```

Lincoln Riley Win Percentage by Game Type



Lincoln Riley Win Percentage by Game Type (Code)

Lincoln Riley

```
specs = [[{'type':'domain'}, {'type':'domain'}], [{'type':'domain'}, {'type':'domain'}]]
fig = make_subplots(rows=2, cols=2, specs=specs)

fig.add_trace(go.Pie(labels=RileyRegCounts.Result, values=RileyRegCounts.Count, name="Regular Season"),
              row = 1, col= 1)
fig.add_trace(go.Pie(labels=RileyBowlCounts.Result, values=RileyBowlCounts.Count, name="Bowl Games"),
              1, 2)
fig.add_trace(go.Pie(labels=RileyConfCounts.Result, values=RileyConfCounts.Count, name="Conference Games"),
              2, 1)
fig.add_trace(go.Pie(labels=RileyOutConferenceCounts.Result, values=RileyOutConferenceCounts.Count,
                     name="Out of Conference Games"), 2, 2)

fig.update_traces(hole=.3, hoverinfo="label+percent+name")

fig.update_layout(
    title_text="Lincoln Riley Win Percentage by Game Type",
    annotations=[dict(text='Regular Season', x=0.17, y=0.51, font_size=10, showarrow=False),
                 dict(text='Bowl Games', x=0.82, y=0.51, font_size=10, showarrow=False),
                dict(text='Conference Games', x=.17, y=-.1, font_size=10, showarrow=False),
                dict(text='Out of Conference Games', x=0.82, y=-.1, font_size=10, showarrow=False)])
fig.update_layout(template='plotly_dark',title_x=0.5)
fig.show()
```