2. DataExploration

March 20, 2019

1   Exploring the Last 10 Years of the NFL

My family and I have always had a huge passion for the NFL. Every Sunday we used to watch football all day long, with no interuption. Since coming to college, I have been able to maintain this tradition some how. I have built up this immense knowledge about the NFL, the sport, teams, players, etc. Now, I want to know how some season and game statistics have changed over the years, especially with a number of new rules and a decreasing television ratings. Also, as I delve further into Statistics and Data Science, I am curious to know what makes a team a SuperBowl Winner and what makes a team have the 1st overall draft pick for the upcoming season.

```
In [1]: import pandas as pd
        %matplotlib inline
        import matplotlib.pyplot as plt
        import numpy as np
        import matplotlib.patches as mpatches

In [2]: nfl = pd.read_csv("nfl.csv")
        nfl = nfl.set_index("Name")
        nfl;
```

## 2   Visualizing Season Statistics for All Teams

In recent years, there have been an increased number of fines for hitting a player near the head or neck area. This in turn has led to defensive players having to play more cautiously. As a result, it seems that NFL teams are throwing for more yards and TDs than before.

How has the amount of Passing and Rushing TDs changed in the last 10 years? How has the amount of Interceptions thrown and Fumbles changed in the last 10 years?

```
In [3]: for i in ["Passing TDs", "Rushing TDs", "Interceptions", "Fumbles", "Points per Game"]
            nfl.groupby(["Year"])[i].mean().plot.line(legend=True)

        fig = plt.gcf()
        fig.set_size_inches(12,10)
        plt.title("How Have TDs, INTs, and FUMs changed over the Past Decade", size = 25)
        ax = plt.gca()
        ax.set_ylabel("Count", size = 20)
        ax.set_xlabel("Year", size = 20);
```

How Have TDs, INTs, and FUMs changed over the Past Decade
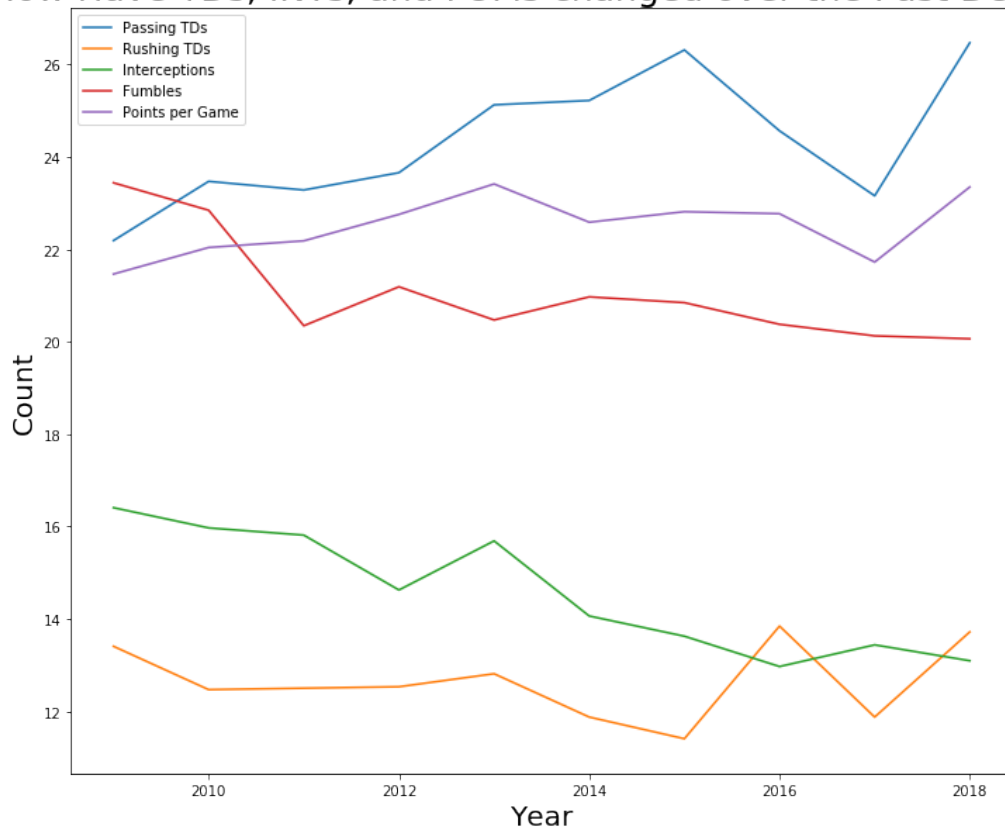
From this graph, we can see there was a clear upward trend in the number of passing TDs and a decreasing trend in the amount of Rushing TDs from 2009 to 2015. However, from 2015 - 2017, the opposite trend occurred - this can be accredited to defensive players adapting to the new rules and changing their style of play. Then, we see a surge from 2017 to 2018 - this is likely due to 2 things: 1) prolific passers are achieving new milestones, new gun-slinger QBs like Patrick Mahomes have taken the league by storm, and a more aggressive mindset by teams and coaches 2) a new concept that teams are incorporating into their offensive scheme called RPO. The Run Pass Option is a designated play where the QB can choose to give the ball to the running back or throw the ball to a wide receiver based on the position of a defender. For defenders, this is very hard to defend against and the play almost always ends up being a positive gain for the offense.

```
In [4]: nfl[["Pass Yards Per Game", "Rush Yards per Game"]].groupby(
            ["Name"]).mean().sort_values(by="Pass Yards Per Game").plot.barh(stacked=Tr

        fig = plt.gcf()
        fig.set_size_inches(12,10)
        plt.title("Passing Yards vs Rushing Yards Per Game by Team", size = 25)
        ax = plt.gca()
        ax.set_ylabel("Team", size = 20)
        ax.set_xlabel("Yards", size = 20);
```

Passing Yards vs Rushing Yards Per Game by Team
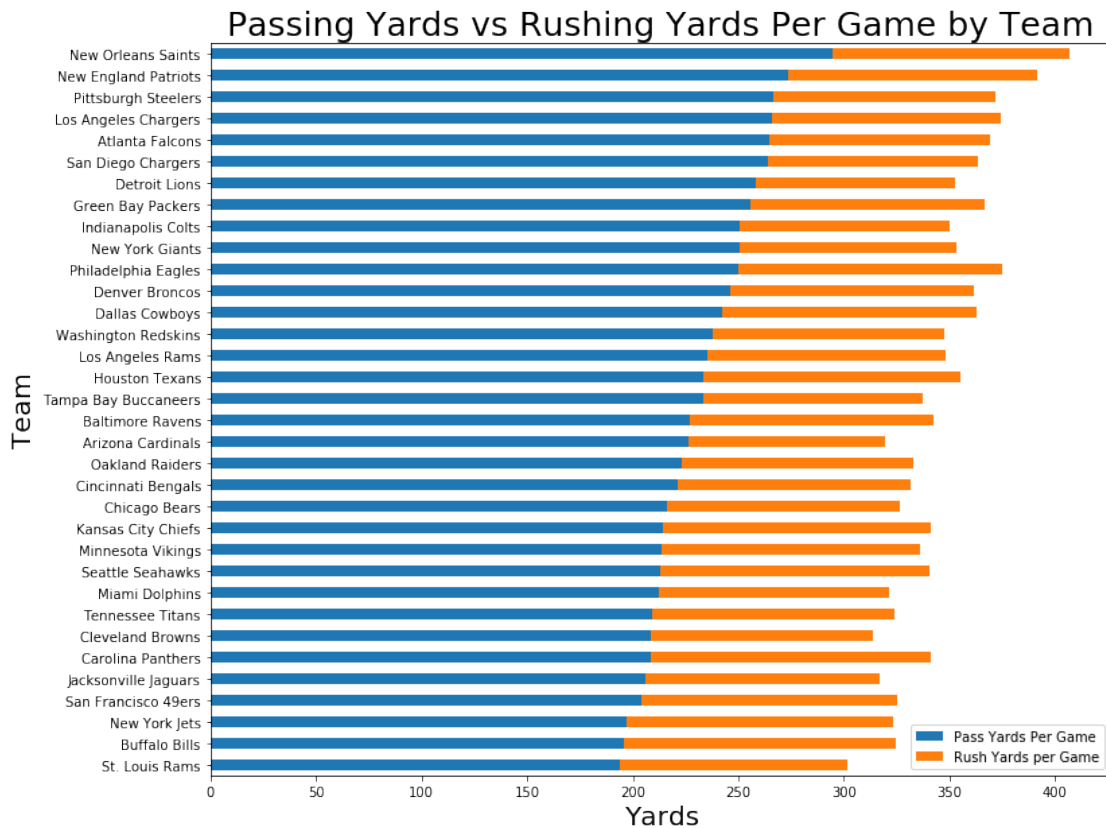
This graph has been made to compare the amount of passing yards to the amount of rushing yards for each team over the past 10 years (sorted by teams with the most passing yards per game to the teams with the least amount of passing yards per game). There is 1 common factor in all the teams at the top of the graph have - an elite QB. The teams at the top have all established there franchise QB, names like Drew Brees, Tom Brady, Peyton Manning, Aaron Rodgers. Note that all these QBs won a SuperBowl(s) during this 10 year period. This suggests that having a stable, elite QB increases your team's performance and likelihood of making the playoffs. The teams at the bottom of this graph like the 49ers, Jets, Bills, and Rams have been failing to find their long-term QB and hence missing out on the postseason nearly every year in comparison to the perennial playoff teams in the upper quarter.

Like I mentioned above, there are many more aggressive QBs and Head Coaches in the NFL than I have ever seen. For example, Philadelphia Eagles Head Coach, Doug Pederson, has been known to be a fearless head coach. In SuperBowl 52, on a fourth down, he called a trick play near the end of halftime to score a TD. This bold call helped lead Philadelphia to Super Bowl win against the New England Patriots. So, does this mean that in recent years teams are going for it more on 4th down? Are teams converting 4th downs more often than before?

```
In [5]: nfl.groupby(["Year"])["4th Downs Made"].sum().plot.line(label="4th Downs Converted", le
        nfl.groupby(["Year"])["4th Downs Attempted"].sum().plot.line(label="4th Downs Attempted
        fig = plt.gcf()
        fig.set_size_inches(12,10)
        plt.title("4th Down Over the Years", size = 25)
```

```
ax = plt.gca()
ax.set_ylabel("Count", size = 20)
ax.set_xlabel("Year", size = 20);
```
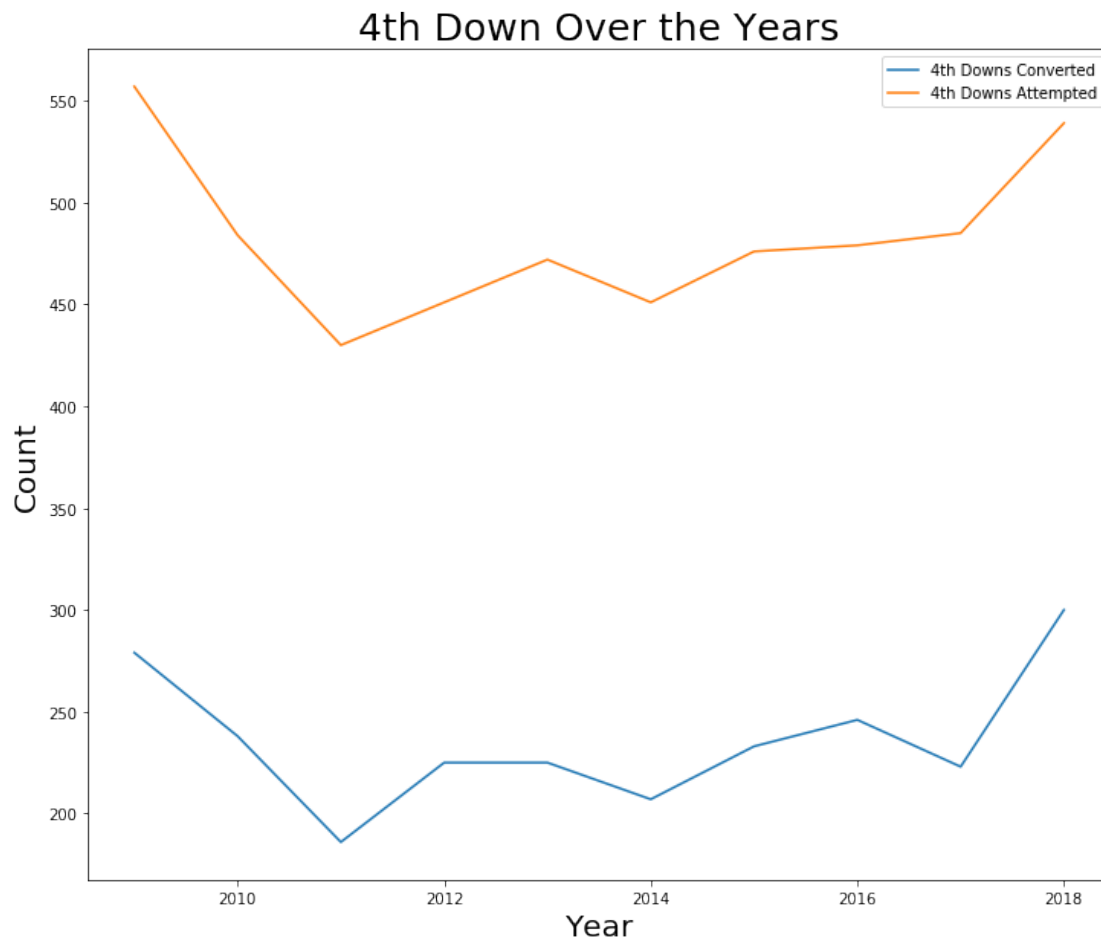

Well, it turns out that the number of 4th downs attempted peaked back in 2009 and then declined until 2011. Since 2011, both the number of 4th downs attempted and 4th downs converted have been increasing just like we suspected.

Now, lets isolate the dataframe to just one team. I will select the New York Giants for a couple of reasons: 1) This is my favorite team. 2) In this 10 year span, the Giants have been a mediocre team, a losing team, and a winning team, making the playoffs twice and winning the Super Bowl once.
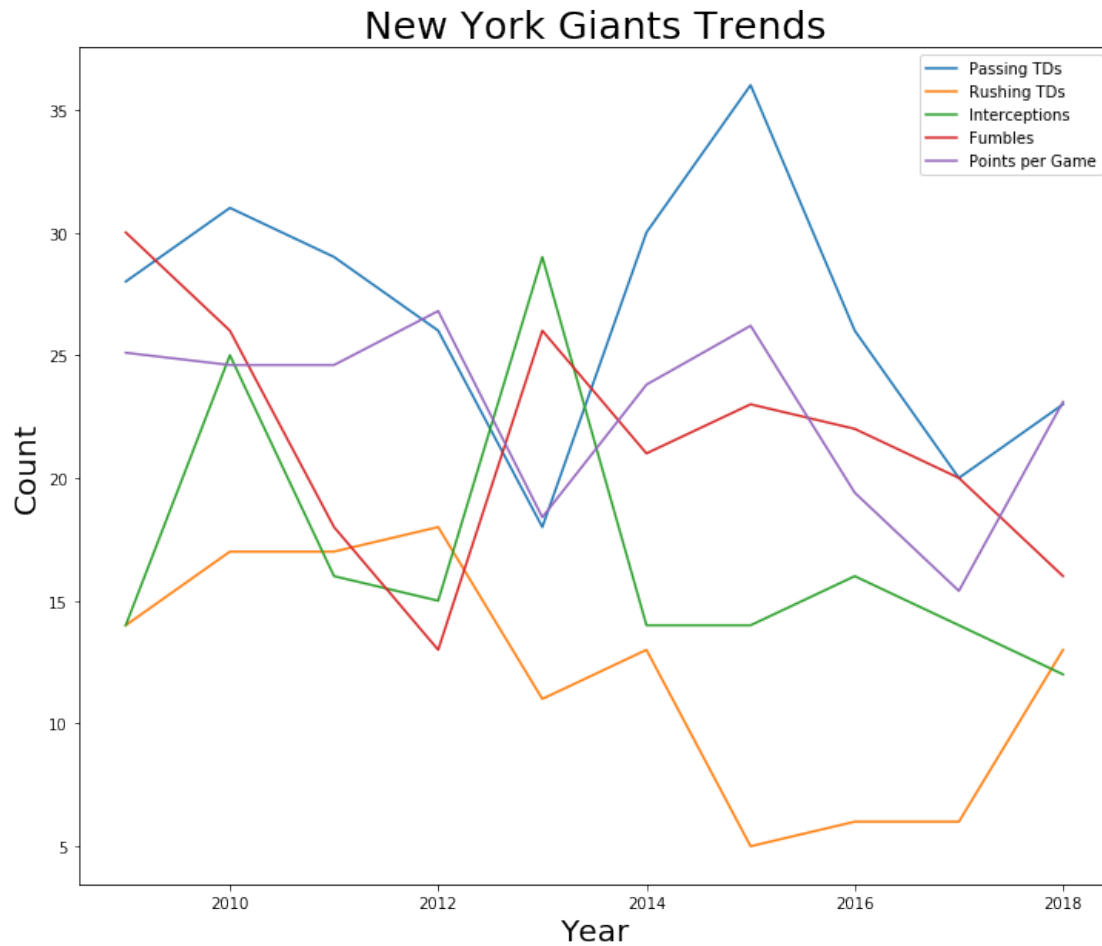
3   Season Stats for One Team

```
In [6]: nyg = nfl.loc["New York Giants"]
        nyg;

In [7]: for i in ["Passing TDs", "Rushing TDs", "Interceptions", "Fumbles", "Points per Game"]
            nyg.groupby(["Year"])[i].mean().plot.line(legend=True)
```

```
fig = plt.gcf()
fig.set_size_inches(12,10)
plt.title("New York Giants Trends", size = 25)
ax = plt.gca()
ax.set_ylabel("Count", size = 20)
ax.set_xlabel("Year", size = 20);
```
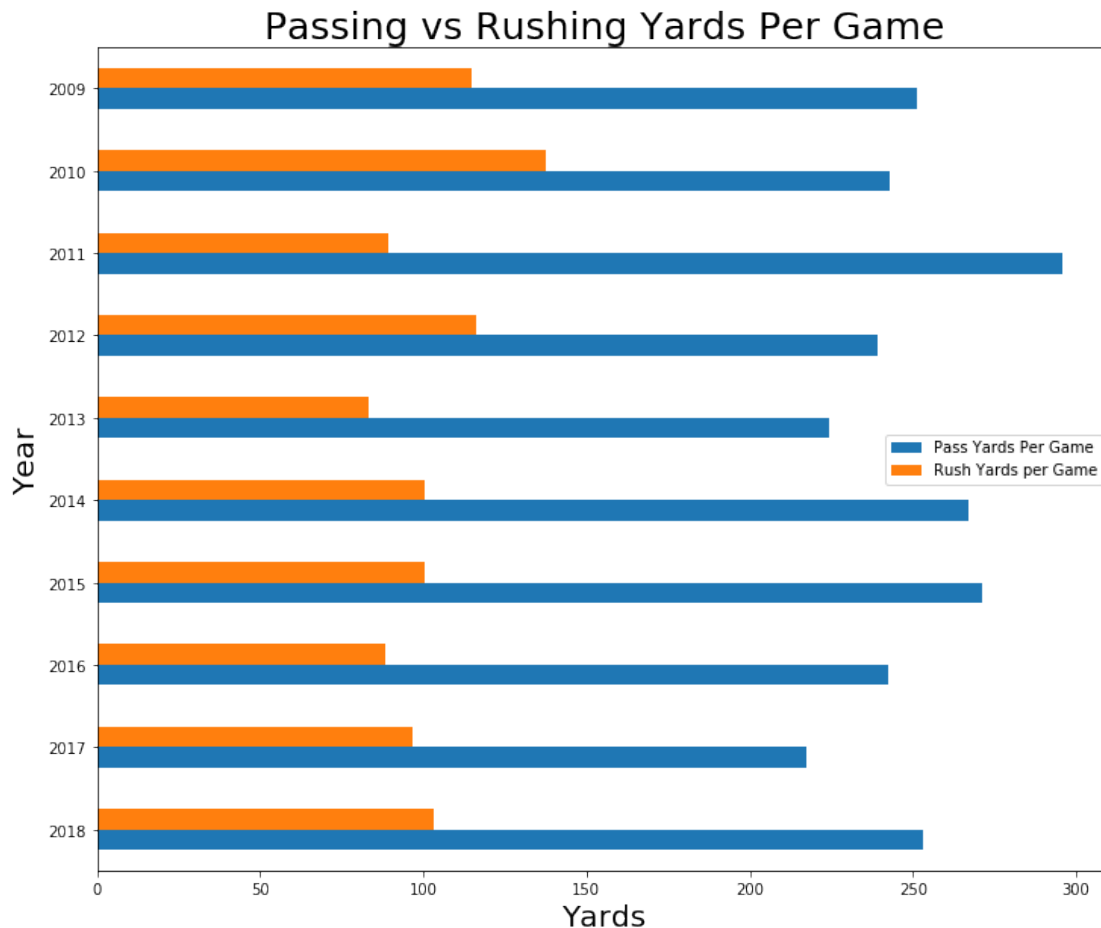
New York Giants Trends

```python
In [8]: import warnings
        warnings.filterwarnings('ignore')

        nyg["Proportion of Pass Yards Per Game"] = nyg["Pass Yards Per Game"]/nyg["Yards per Ga
        nyg["Proportion of Rush Yards Per Game"] = nyg["Rush Yards per Game"]/nyg["Yards per Ga
        nyg[["Pass Yards Per Game", "Rush Yards per Game","Year"]].groupby(
            ["Year"]).mean().sort_values(by="Year", ascending=False).plot.barh(stacked=False)

        fig = plt.gcf()
        fig.set_size_inches(12,10)
```

```
plt.title("Passing vs Rushing Yards Per Game", size = 25)
ax = plt.gca()
ax.set_ylabel("Year", size = 20)
ax.set_xlabel("Yards", size = 20);
```


Compared to the overall trends of the NFL, the New York Giants season statistics are much more scattered. Looking at the Pass Yards vs Rush Yards per Game, we see that the Giants best passing season came in 2011, but that year was also their worst rushing season. In 2018, they drafted a running back sensation, Saquon Barkley, and yet there does was not a dramatic positive increase in the amount of Rush Yards per Game leaving people wondering if he was worth the selection after finishing the season 5-11. In 2009, the Giants finished 8-8 which is surpising given that they had more Fumbles than Passing TDs and the same number of Interceptions as Rushing TDs. Ideally, you would want more Passing and Rushing TDs than Interceptions and Fumbles. Earlier, I mentioned that Giants made the playoffs twice and one the SuperBowl once. Do you think just by looking at the Trends you can tell which two years they made the playoffs? And of the 2 playoff appearances, can you guess which year they won the SuperBowl?

The Giants made the playoffs in 2011 and in 2016. The organization won its 4th SuperBowl, SuperBowl 46, in 2011. This is surprising given that I told you ideally as a team, you want to see

more TDs, less fumbles, and less interceptions. Out of these 10 years it would have seemed like 2012 would have been the most appropriate guess.

This now will lead me to investigate what separates a SuperBowl winning team from that year's worst NFL team.

4 What are the Biggest Differences Between a SuperBowl Winning Teanm and that Year's Worst NFL Team?

There are a lot that goes into a making a team a winning team and having an unsuccessful team, but bottomline is that your team has to execute - they need to score more points to win, get more yards, have more "splash / explosive" plays, and convert on 3rd down. (Note that splash/explosive plays are considered plays gaining more than 20 yards)

```
In [9]: superbowl_df = pd.read_csv("superbowl_df.csv")
        superbowl_df["Identifier"] = "SuperBowl Winner"

        worst_team_df = pd.read_csv("worst_team_df.csv")
        worst_team_df["Identifier"] = "Worst Team"

        df = pd.concat([superbowl_df, worst_team_df])

In [10]: superbowl_df.groupby("Year")["Yards per Game"].mean().plot.line(label="SuperBowl Winne
         worst_team_df.groupby("Year")["Yards per Game"].mean().plot.line(label="Worst Team", 
         fig = plt.gcf()
         fig.set_size_inches(12,10)
         plt.title("Yards Per Game for SuperBowl Winning Teams vs that Year's Worst Team", size
         ax = plt.gca()
         ax.set_ylabel("Yards Per Game", size = 20)
         ax.set_xlabel("Year", size = 20);
```

Yards Per Game for SuperBowl Winning Teams vs that Year's Worst Team
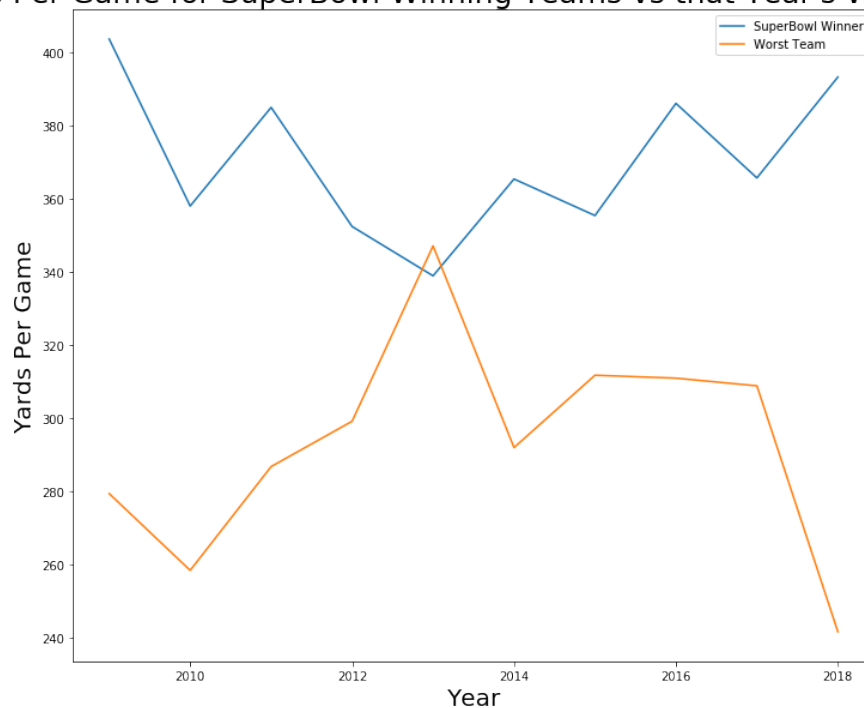
Here, we see that in all years except 2013, the SuperBowl winning team had drastically more Yards Per Game than the worst team. The largest difference came in 2018 - nearly a 150 yard per game difference in favor of the SuperBowl Champions. In 2013, the Houston Texans were the league's worst team, but not because they couldn't move the ball up and down the field as seen by the graph. Rather, they simply couldn't win games. This then led the team to not draft a QB and instead, they drafted college football's best defensive player, Jadeveon Clowney.

Above, I mentioned that explosive plays also define how well a team does. So, we will next plot at a grpah that plots the proportion of splash/explosive plays for each of the SuperBowl Winning Teams and Worst Teams. The graph will be organized so that as we look from top to bottom, the year is ascending and each barplot alternates between the SuperBowl Winning Team and that year's worst team.

```
In [11]: denominator = (df["Completed Passes Greater than 20 Yards"]+df["Completed Passes Grea
                        df["Rushes Greater than 20 Yards"]+df["Rushes Greater than 40 Yards"])

         for i in ["Completed Passes Greater than 20 Yards",
                 "Completed Passes Greater than 40 Yards",
                 "Rushes Greater than 20 Yards",
                 "Rushes Greater than 40 Yards"] :
             df["Prop. of " + i] = df[i] / denominator

         df[["Prop. of Completed Passes Greater than 20 Yards",
             "Prop. of Completed Passes Greater than 40 Yards",
             "Prop. of Rushes Greater than 20 Yards",
```

```
                "Prop. of Rushes Greater than 40 Yards" ,
                "Year", "Name",
                "Identifier"]].groupby(["Year", "Identifier", "Name"]).mean().sort_values(
                            by=["Year"], ascending=False).plot.barh(stacked=True)

        fig = plt.gcf()
        fig.set_size_inches(12,10)
        plt.title("Proportion of Explosive Plays for SuperBowl Winning Teams vs the NFL's Wors
        ax = plt.gca()
        ax.set_ylabel("Year, Team Type, Team", size = 20)
        ax.set_xlabel("Proportion of Big Plays", size = 20);
```


Proportion of Explosive Plays for SuperBowl Winning Teams vs the NFL's Worst Teams
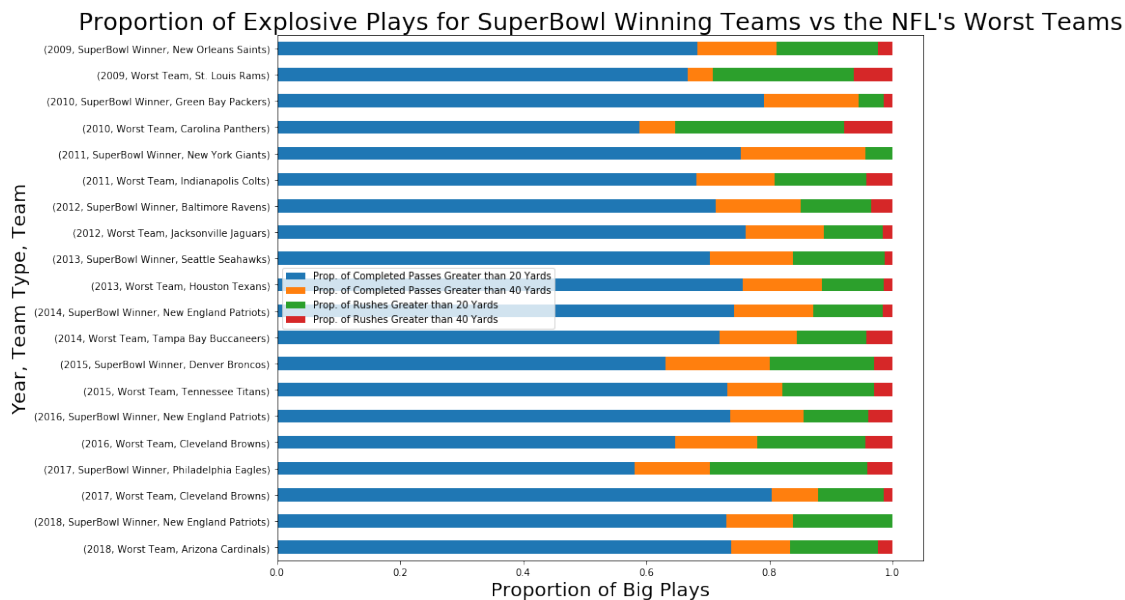
It appears there isn't an obvious trend in every team's explosive rushing plays, despite the team being either a SuperBowl winner or that year's worst team. However when it comnes to explosive passing plays, there is a difference. For the most part, we see that SuperBowl teams have a slightly greater proportion of Completed Passes Greater than 40 yards in comparison to the worst teams.

```
In [12]: colors = df["Identifier"].map({
            "SuperBowl Winner" : "b",
            "Worst Team" : "r"
        })

        df.plot.scatter(x="% 3rd Downs Converted", y="3rd Downs Attempted", c=colors)
        fig = plt.gcf()
        fig.set_size_inches(12,10)

        plt.title("3rd Down Conversion Rate for SuperBowl Teams vs the Worst Teams", size = 2
```
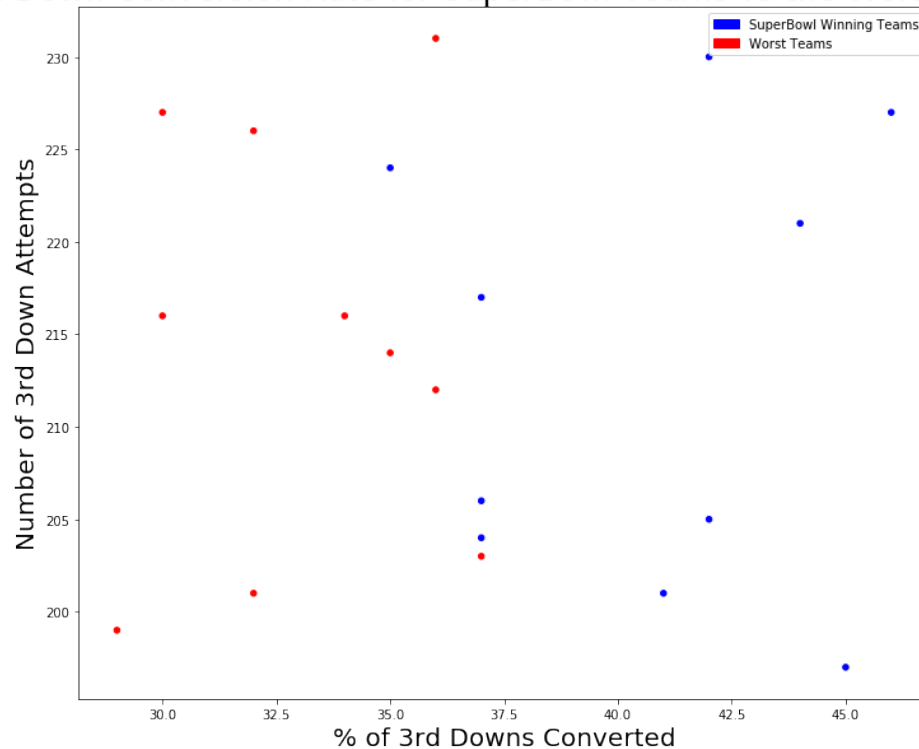
9

```
blue_patch = mpatches.Patch(color='blue', label='SuperBowl Winning Teams')
red_patch = mpatches.Patch(color='red', label='Worst Teams')
plt.legend(handles=[blue_patch, red_patch])

ax = plt.gca()
ax.set_ylabel("Number of 3rd Down Attempts", size = 20)
ax.set_xlabel("% of 3rd Downs Converted", size = 20);
```

3rd Down Conversion Rate for SuperBowl Teams vs the Worst Teams

From the scatter plot of 3rd down conversion rates, we see that SuperBowl Winning Teams are much better at converting on third down than the NFL's worst teams. This leads to longer drives down the field, and thus you get more opportunities to score points. Therefore, for a team looking to turn their franchise around, I recommend starting with your 3rd down conversion rate because this from this graph it appears that the NFL's best teams know how to stay on the field.