**Intro:**

Football is one of the most popular and most violent sports in the United States. On every level of competition be it high school, collegiate, or professional, football brings in more money and more attention in the United States than any other sport. The NFL or the National Football League is the biggest football organization in the world. With each team in the organization averaging around 2.86 billion dollars in net worth and each game pulling in over 158 million viewers with the Superbowl reaching around 100 million; the NFL is a monster of an association.

Football is played on a grass field that spans 120 yards with ten of those yards at each end representing the endzone for each team. At any given time in a game, there are eleven people from each team taking turns playing offence or defense with the goal of moving the ball to the other teams endzone or stopping the opposition from doing the same thing. There are many different positions on the offensive and defensive sides but the most important position by far is the quarterback. The quarterback is the leader of the offence and touches the ball every single play, because of this the quarterback has the most influence on the outcome of a game compared to all other positions which may only touch the ball a small percentage of the snaps. Generally, a quarterback can choose to either throw the ball to a receiver, run himself, or hand it off to a running back for a run. This means that the easiest way to compare quarterbacks is with stats such as passing yards or touch down to interception ratios because his main job is to throw the ball and when he hands it off someone else is doing the work. Throughout the course of a season sports analysists will look at these basic stats each game to determine how good a quarterback is playing, yet in the end, because a quarterback has so much to do with the outcome of a game, if a team does not make it to the playoffs and do well the brunt of the blame is places on him. Therefore, there are sometimes discrepancies between stats and championships, some QBs can put up numbers but their teams are not good enough to win while other QBs may put up worse stats but still win.

Sports in general tend to be a hard topic for data analytics because there are so many things that could affect the outcome of a game or how well a player performs. There are some players that do not put up good stats but have intangibles such as leadership or the ability to stay calm under pressure that may lead to them winning more games and are very hard to record and quantify. This leads to some players making you just believe in them or give you the feeling that they are good even if there is no real reason to think so.

The Pro Football Hall of Fame represents a list of all the best players determined by a committee of qualified individuals. When a player retires the commute reviews that players career and votes on whether they belong in the Hall of Fame. There are no specific rules or clear equation that results in a player getting the thumbs up or thumbs down from the committee; votes are cast based on whatever each member thinks are important. The goals of this project are to see if there are common trends that place a quarterback in the Hall of Fame and to be able to accurately determine if a player will eventually make it into the Hall of Fame.

Data was collected and merged from Pro Football Reference and a Kaggle data set called NFL-Statistics.

All the data graphed below is filtered to only plot the retired players. This is done because the graphs are meant to try to find differences between the hall of fame players and the rest of the group. If active players are included, players that might be in the hall of fame later will show up as normal players and potentially make the graphs misleading.

This graph plots career touchdown passes against career interceptions both in terms of their respective standard deviations. In football a quarterback is graded on how well they keep the ball out of the other teams’ hands and how well they put it in in the endzone. From the graph we can see that all the quarterbacks in the hall of fame are at least 1.5 standard deviations away from the pack in touchdown passes with most of them around 2 standard deviations away. However, most of the hall of fame players are also 1 standard deviation above the mean in interceptions. This leads to the conclusion that hall of famers probably played much longer than your average player, that is why they accumulated more touchdown passes and interceptions. There seems to be an almost linear correlation as we go from left to right on the plot.

Because the plot before this plot was potentially skewed by hall of fame players playing more total games, this plot shows per game averages. Touchdown passes per game is plotted on the x-axis while passing yards per game is plotted in the y-axis. This graph makes it pretty clear that hall of fame players tend to put up better numbers for both of these stats with very few players averaging more than 175 yards and 1.25 touchdowns a game and not making it in.

In football, the playoffs are as high stakes as it gets. If you lose you are out and have no hope of winning a championship. When a quarterback can bring his team back in the playoffs in front of millions of people it is almost always remembered. Some players can win big games despite not having good numbers on a stat sheet and because people vote hall of famers in and there is no formula, it is important to isolate quarterback performance in big games. It is clear that if a quarterback has even one game winning drive their odds of getting into the hall of fame skyrockets with anyone with more than three always getting in.

Getting sacked can change a game. If a quarterback is able to throw the ball away before being tackled for a big loss of yards the odds a drive stay alive are much higher. This graph plots sacks against passes incomplete. There is not much correlation here and the only thing that can be drawn from the plot is again that hall of fame players must have had longer careers in order to pile up more overall stats be it good or bad. Points seem to almost be random distributed.

In the same way that game winning drives in the playoffs can make a player famous, the Superbowl is the ultimate test. The Superbowl is the most viewed sporting event in the world every year and unlike playoff and regular season games there is only one a year. Most of the blame falls on a quarterback for the wins and losses throughout the year and if a team is able to win a Superbowl a lot of the credit is given to him. This bar graph shows number of Superbowl victories on the x-axis and number of players on the y-axis. Very few quarterbacks have won a Superbowl and if a quarterback wins even one their odds of getting in increase by a lot just how they did with playoff game winning drives.