Introduction

I will first preference this this project by saying that I am not at all a gambler. I went to the casino on my 21st birthday, sat at a slot machine, and proceeded to play it for six minutes. In those six minutes, I lost the twenty dollars that my older brother had given me to start out with and I decided that I was done. I then got an alcoholic beverage and watched everyone else in the building strive to win some money.

I decided to step away from the casinos because I didn't understand it. I saw it as giving my money to an institution that understands the game better than I do. Take it as paying Steph Curry to embarrass me in a 3pt contest. I can't shoot so how will I shoot better than the greatest shooter ever. One thing I do understand is football. I understand NFL teams, how they are structured, and who's who in the league.

Recently, as I have explored careers and jobs in data analytics, I have found myself trying to merge the knowledge of that I know and that I don't know. This comes in the form on sports betting. I have been to Las Vegas and seen these huge screens with sports books in hotel casino floors. I walked by and decided not to partake in the past, but now I would like to be able to have the knowledge to see and understand what is occurring so I can use football wits to make some betting money.

This recent NFL season had me paying more and more attention to which teams are being betted on, how many points are being scored to sustain certain bets, how the home team dictate betting lines. In looking at this project, I wanted to understand what exactly went on during the season to understand the big picture moving forward into the next season.

There are a couple terms that one must know to comprehend how to bet or what to bet on. Definitions provided by Athlon Sports are as follows:

Point Spread: "Commonly referred to as just the "spread." Betting line or odds used to determine favorite and underdog."

Over/Under: "The betting line on the total number of points or goals scored in a sporting event, with action taken on whether there will be more (the over) or fewer (the under) points or goals scored."

Cover: "To correctly pick the winning side of a point-spread bet. (e.g. "Alabama was favored by 10 points and won by 11, I covered.")"

Favorite: "The side expected to win a sporting event, with odds reflecting perceived confidence in favored team/person."

Underdog: "The side expected to lose a sporting event, with odds reflecting perceived lack of confidence in team/person.'

Line: "Betting line or odds used to determine the gambling margin between the favorite and underdog."

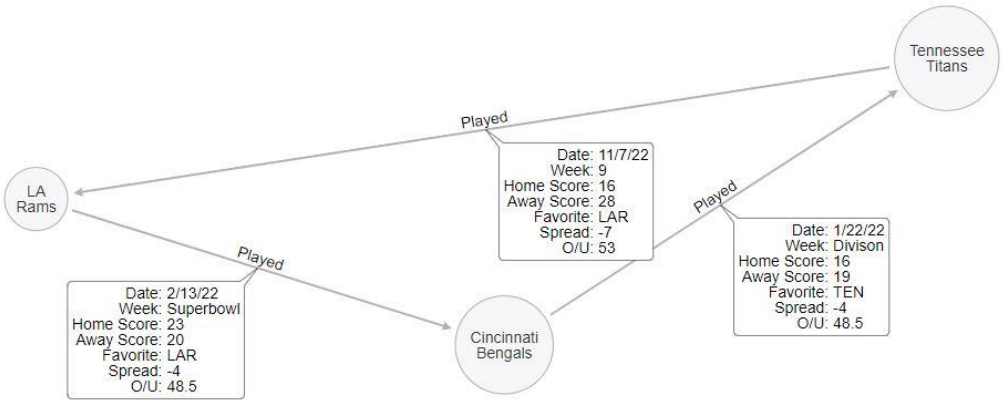
Dataset

To accomplish the purpose of examining betting data, I needed a dataset that had information that was specific enough that applied enough detail that the total picture could be

realized. I spent a lot of Google searches and time looking at NFL data. I realized that I couldn't look at NFL data because this was a different animal. I needed to turn to the casino side instead of the sports side because that's where the money lies. I found a spreadsheet by Toby Crabtree that contained NFL game results from 1966. The betting odds came later in 1979. The data was mixed together with scores from ESPN, NFL.com, and Pro Football Reference. Pro Football Reference also provided betting data which was cross referenced from sportsline.com. This was published on Kaggle and the link is provided below.

https://www.kaggle.com/datasets/tobycrabtree/nfl-scores-and-betting data?select=spreadspoke_scores.csv

Graph Data Model




```
CREATE

  (`1` :`LA Rams` ) ,
  (`2` :`Cincinnati Bengals` ) ,
  (`3` :`Tennessee Titans` ) ,
  (`1`)-[:`Played` {Date:'2/13/22',Week:'Superbowl',`Home Score`:'23',`Away
Score`:'20',Favorite:'LAR',Spread:'-4',`O/U`:'48.5'}]->(`2`),
  (`2`)-[:`Played` {Date:'1/22/22',Week:'Divison',`Home Score`:'16',`Away
Score`:'19',Favorite:'TEN',Spread:'-4',`O/U`:'48.5'}]->(`3`),
```

(`3`)-[:`Played` {Date:'11/7/22',Week:'9',`Home Score`:'16',`Away Score`:'28',Favorite:'LAR',Spread:'-7',`O/U`:'53'}]->(`1`)

Loading the Data

```
LOAD CSV WITH HEADERS FROM 'file:///2021_NFLSpread.csv' AS row
MERGE (h:Team {name: row.HomeTeam})
MERGE (a:Team {name: row.AwayTeam})
MERGE (h)-[r:INTERACTS]→(a)
ON CREATE SET r.Spread = toInteger(row.Spread)
ON CREATE SET r.OverUnder = toInteger(row.OverUnder)
ON CREATE SET r.Stadium = toString(row.Stadium)
ON CREATE SET r.HomeScore = toInteger(row.HomeScore)
ON CREATE SET r.AwayScore = toInteger(row.AwayScore)
ON CREATE SET r.FavoriteID = toString(row.FavoriteID)
ON CREATE SET r.Week = toString(row.Week)
ON CREATE SET r.Date = toString(row.Date)
```

Table

Code

Added 32 labels, created 32 nodes, set 1990 properties, created 280 relationships, completed after 185 ms.

Business Questions

For this project I selecting a couple questions that should be answered using this graph data.

The first one is: **Who should I bet on the over?**

Over/Under is one of the most popular bets when betting on the NFL. You pick a matchup and basically have to determine the number of points you believe will be scored. Vegas will set a certain number and you have two choices: the combined points scored will be more than this number or less than this number. This is very difficult to pinpoint because every game is different. The 2007 New England Patriots went 16-0 in the regular season and broke many scoring records on the way. They averaged 36.8 points a game (Pro Football Reference)

and their over/under for their Super Bowl matchup against the New York Giants was 55 points. The Patriots ended up only scoring 14 points in this contest and the under was the ultimate option by 24 points.

Some people take one team the whole season and ride them out on one of the options for 17 weeks. Frank Dimartin in our BDAT635 course stated that he took two teams for the over and two teams for under the season and did pretty well. He also took the Los Angeles teams a couple times (Rams and Chargers), which two of the highest scoring teams in the league and had different results.

In my studies to answer this question, I needed to examine which teams were hitting the over and which opponents they were doing so against. I ran three main queries to dissect who would consistently make me money by betting them and the over and essentially why. I have included these queries below.

```
MATCH (n:Team), (:Team)-[p:INTERACTS]-
(:Team) WHERE (p.AwayScore + p.HomeScore) > p.OverUnder RETURN p, n
```

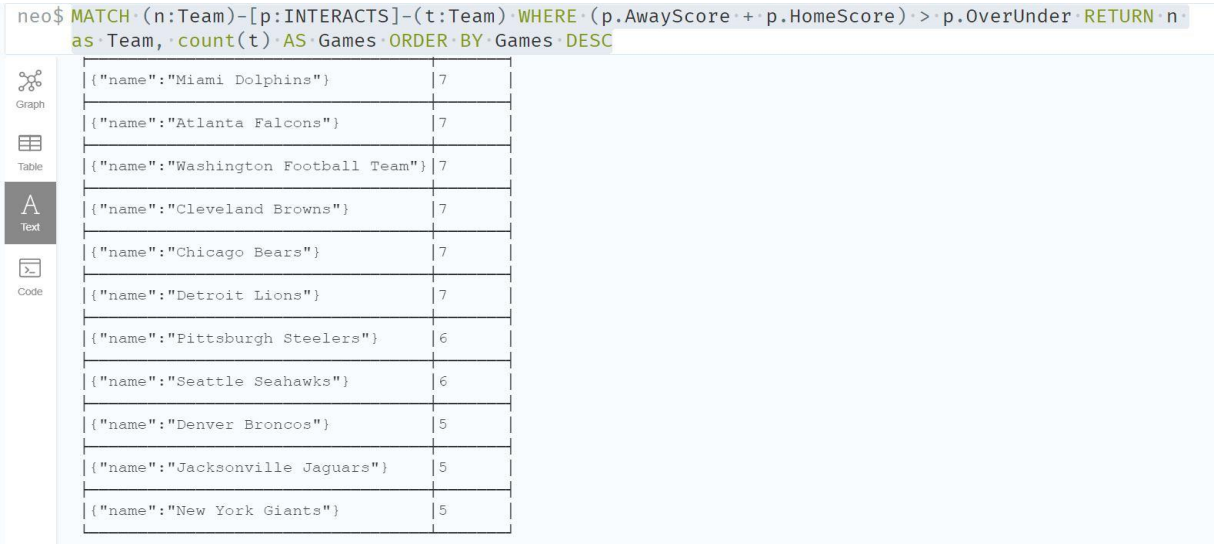
```
MATCH (n:Team)-[p:INTERACTS]-
(t:Team) WHERE (p.AwayScore + p.HomeScore) > p.OverUnder RETURN n as Team, count(t)
 AS Games ORDER BY Games DESC
```

```
MATCH (n:Team {name: 'Los Angeles Rams'})-(t:Team) WHERE (p.AwayScore +
p.HomeScore) > p.OverUnder RETURN p, n, t
```

What I discovered was somewhat shocking. The Minnesota Vikings lead the league in games in which the over should have been bet, followed by the New York Jets. These are two teams that did not make the playoffs and essentially underachieved this past season. But in

further investigation, we can see that these are two teams that rank in the bottom of the league in team defense (Pro Football Reference). This means that a team's defense is also factored into this bet. The offense and defense have to both be considered in determining the Over/Under because a team like the Denver Broncos who are ranked third in team defense are last in games in which the over should have been bet. On the other hand, the Jaguars are also ranked near the very bottom in terms of Over/Under and they are ranked toward the bottom of team defense as well. It takes a good disparity to rank towards the top of Over/Under in this case. Take the Los Angeles Rams for example, Frank said he did not have much success with betting them and it may be for good reason. The Rams ranked in the top 5 in team offense with would lead one to believe to bet the over, but also ranked in the top half of the league in team defense so the other team may not have been allowed to contribute much scoring to the cause.

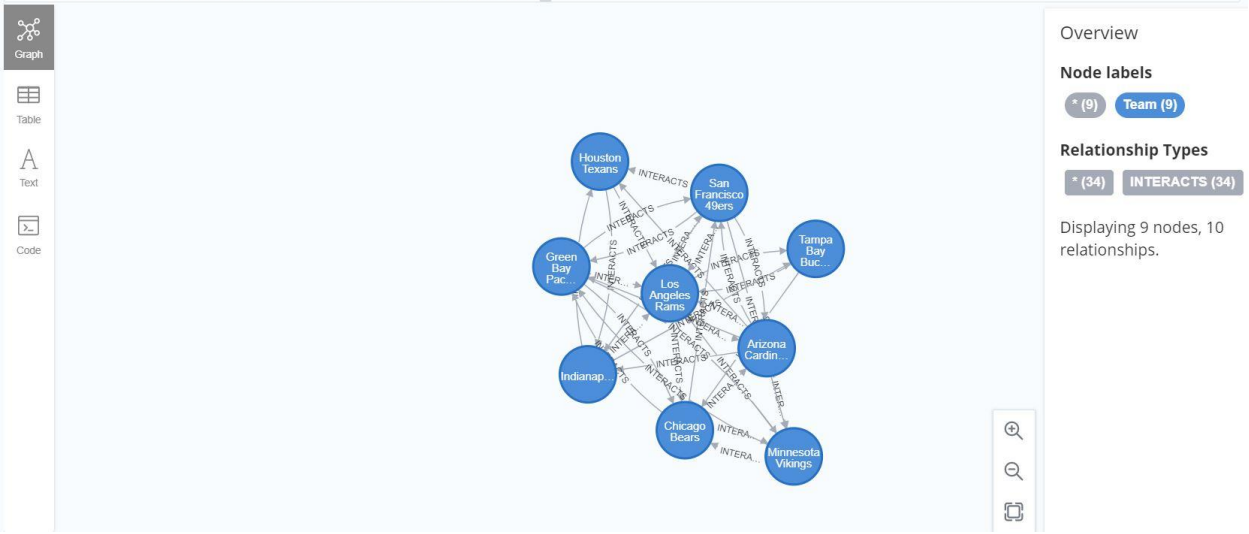
**The Rams rank high in the games over list but they played more games that everybody.*



```
neo$ MATCH (n:Team)-[p:INTERACTS]-(t:Team) WHERE (p.AwayScore + p.HomeScore) > p.OverUnder RETURN n
     as Team, t as Opponent, p.Week AS Week ORDER BY n ASC
```

Team	Opponent	Week
{"name":"Tampa Bay Buccaneers"}	{"name":"Atlanta Falcons"}	"2"
{"name":"Tampa Bay Buccaneers"}	{"name":"Dallas Cowboys"}	"1"
{"name":"Atlanta Falcons"}	{"name":"Miami Dolphins"}	"7"
{"name":"Atlanta Falcons"}	{"name":"Carolina Panthers"}	"14"
{"name":"Atlanta Falcons"}	{"name":"New York Jets"}	"5"
{"name":"Atlanta Falcons"}	{"name":"Washington Football Team"}	"4"
{"name":"Atlanta Falcons"}	{"name":"New Orleans Saints"}	"18"
{"name":"Atlanta Falcons"}	{"name":"New Orleans Saints"}	"9"
{"name":"Atlanta Falcons"}	{"name":"Tampa Bay Buccaneers"}	"2"
{"name":"Buffalo Bills"}	{"name":"Washington Football Team"}	"3"
{"name":"Buffalo Bills"}	{"name":"Carolina Panthers"}	"15"
{"name":"Buffalo Bills"}	{"name":"Tampa Bay Buccaneers"}	"14"

```
neo$ MATCH (n:Team {name:'Los Angeles Rams'})-[p:INTERACTS]-(t:Team) WHERE (p.AwayScore +
     p.HomeScore) > p.OverUnder RETURN p, n, t
```



The second one is: **Which teams should you stay from when betting the spread?**

One of my favorite sports talk television shows is First Things First on Fox Sports 1. Thursdays on the show, Nick Wright tells the audience his bets of the week. He tells which teams to bet on when betting on the Spread. Here, he likes to stay with a team based on the players playing. He also liked to pick players to stay away from. For example, The New York Jets quarterback once got hurt, and until he got back, Nick bet against the backup every week. This

got me to thinking, which teams can we bank on to stay away from or pick on every week. It may be easier to pick the loser rather than the winner.

To answer this I wanted to see which teams did not make the spread in their games. In looking at the list, it gives you an idea of the good teams as well as who just can't catch up to the spread. The teams to stay away from would have a high number of hits on the list, regardless of if they are the favorite.

```
MATCH (h:Team)-[p:INTERACTS]-
(a:Team) WHERE (p.AwayScore - p.HomeScore) < p.Spread RETURN h.name AS Team, p.Spr
ead AS SpreadSet, p.FavoriteID AS Favorite, p.Week AS Week
```

```
neo$ MATCH (h:Team)-[p:INTERACTS]-(a:Team) WHERE (p.AwayScore - p.HomeScore) < p.Spread RETURN h.name AS Team, p.Spread AS SpreadSet, p.FavoriteID AS Favorite,
    p.Week AS Week
```

"Cincinnati Bengals"	-3	"CIN"	"12"
"Cincinnati Bengals"	-7	"CIN"	"16"
"Cincinnati Bengals"	-6	"CIN"	"Wildcard"
"Detroit Lions"	-13	"ARI"	"15"
"Detroit Lions"	-4	"GB"	"18"
"Detroit Lions"	-8	"SEA"	"17"
"Detroit Lions"	-12	"DEN"	"14"
"Detroit Lions"	-3	"CHI"	"4"
"Detroit Lions"	-10	"GB"	"2"
"Houston Texans"	-20	"ARI"	"7"
"Houston Texans"	-11	"IND"	"6"
"Houston Texans"	-13	"SF"	"17"

As seen here, even bad teams such as the Detroit Lions and the Houston Texans have multiple hits even when you would think that they would not. Combing through this list, we can tell that there are certain teams that you should stay away from good, bad, or average.

The third one is: **Which teams carry a true home field advantage?**

Home is where the heart is. This is true both in life and in sports. As an athlete, didn't you feel better playing in front of your mom, sleeping in your own bed, or knowing the markers to look for in order to make split second decisions? This is especially true on the professional

level. More so because the crowd can play a big part in the gameplay on the field. There are a couple places in the NFL with a reputation of being a hostile playing environment for opponents. The Buffalo Bills have a rabid fan base that loves to jump through tables pregame and yell at the top of their lungs in game. It is said that it feels like there's an extra opponent on the field in Seattle with their crowd they call the 12th man. Weather also plays a part in the advantage with the frigid temperatures in Green Bay or the high altitude in Green Bay.

In order to determine the advantage in these home stadiums, I would presume that the offensive should play well because they are in a comfortable, controlled environment and the defense should play well because the crowd can make the opponents uncomfortable. If the defense keeps the opponents below the league average and the offense scores above the league average then an advantage has been realized. According to Pro Football Reference, the league average is 23 points so that is the mark that should be reached or prevented. The query to determine how accomplished this is as follows:

```
MATCH (h:Team)-[p:INTERACTS]-
(a:Team) WHERE (p.AwayScore < 23) AND (p.HomeScore > 23) RETURN h AS Team, p.Stad
ium AS Stadium, count(h) AS GreatGames ORDER BY GreatGames DESC
```

What was discovered was along the lines of what I thinking. A good team plus a home advantage means that home games will be won. The top five teams with games in which the offense and defense play well at home are the Buffalo Bills, Tampa Bay Buccaneers, Green Bay Packers, Dallas Cowboys, and New England Patriots. Three of these five teams play in a very cold climate so when the temperature turns at the later part of the season, their advantage increases. The shocker on this list is the Tampa Bay Buccaneers. When I think of Tampa, I don't

think of a great fanbase or a quirk that will lead to wins. In their case, they have the greatest quarterback of all time in Tom Brady so I guess he is the advantage. Dallas Cowboys owner, Jerry Jones also has the standard for a football stadium so that in itself can be an advantage as it's an experience to even walk in the place.

```
neo$ MATCH (h:Team)-[p:INTERACTS]-(a:Team) WHERE (p.AwayScore < 23) AND (p.HomeScore > 23) RETURN h
     AS Team, p.Stadium AS Stadium, count(h) AS GreatGames ORDER BY GreatGames DESC
```

Graph Table Text Code

"Team"	"Stadium"	"GreatGames"
{"name":"Buffalo Bills"}	"Highmark Stadium"	6
{"name":"Tampa Bay Buccaneers"}	"Raymond James Stadium"	5
{"name":"Green Bay Packers"}	"Lambeau Field"	5
{"name":"Dallas Cowboys"}	"AT&T Stadium"	4
{"name":"Cincinnati Bengals"}	"Paul Brown Stadium"	4
{"name":"New England Patriots"}	"Gillette Stadium"	4
{"name":"Los Angeles Rams"}	"SoFi Stadium"	3
{"name":"Tennessee Titans"}	"Nissan Stadium"	3
{"name":"Cleveland Browns"}	"FirstEnergy Stadium"	3
{"name":"Los Angeles Chargers"}	"SoFi Stadium"	3
{"name":"Denver Broncos"}	"Empower Field at Mile High"	3

Degree Centrality

Degree Centrality is used to find popular nodes within the graph model. In the social media universe, this can be used to determine the most important men and women in that space. In this graph, it can determine which teams contains the most relationships as the season's matchup are both incoming and outgoing.

```
1 MATCH (t:Team)
2 RETURN t.name AS team, size( (t)-[:INTERACTS]-() ) AS degree ORDER BY degree DESC
```

	team	degree
1	"Cincinnati Bengals"	21
2	"Tampa Bay Buccaneers"	19
3	"Los Angeles Rams"	19
4	"San Francisco 49ers"	19
5	"Kansas City Chiefs"	18
6	"Tennessee Titans"	18
7	"Las Vegas Raiders"	18
8	"Philadelphia Eagles"	18
9	"Green Bay Packers"	18
10	"Dallas Cowboys"	18
11	"Atlanta Falcons"	17
12	"Buffalo Bills"	17
13	"Carolina Panthers"	17

Started streaming 32 records in less than 1 ms and completed after 16 ms.

These teams have outlasted the other teams in the league. Most teams have a degree of centrality of 17. If the team has more than 17, then there is a reason of they are really good and get to last longer in the season. Or they connect with others in a way that they are popular and get shown a lot.

Weighted Centrality

The weighted degree centrality represents the nodes strength and relies on tie weights instead of number of ties. The weights in this case are the representations that the teams' have in terms of spread and over/under.

```
MATCH (t:Team)-[r:INTERACTS]-()
RETURN t.name AS team, avg(r.Spread) AS SpreadTotal, avg(r.OverUnder) AS OverUnderTotal ORDER BY SpreadTotal ASC
```

	team	SpreadTotal	OverUnderTotal
1	"Houston Texans"	-10.294117647058824	44.1764705882353
2	"Buffalo Bills"	-8.882352941176471	47.117647058823536
3	"Tampa Bay Buccaneers"	-8.052631578947368	49.10526315789474
4	"Detroit Lions"	-7.7058823529411775	45.1764705882353
5	"New York Jets"	-7.470588235294118	43.647058823529406
6	"Jacksonville Jaguars"	-7.352941176470587	45.411764705882355

```
neo4j$ MATCH (t:Team)-[r:INTERACTS]-() RETURN t.name AS team, sum(r.OverUnder) AS OverUnderTotal, sum…
```

	team	OverUnderTotal	SpreadTotal
1	"Cincinnati Bengals"	978	-87
2	"Kansas City Chiefs"	938	-109
3	"Tampa Bay Buccaneers"	933	-153
4	"Los Angeles Rams"	931	-113
5	"Dallas Cowboys"	902	-106
6	"San Francisco 49ers"	889	-86
7	"Tennessee Titans"	854	-92

Started streaming 32 records in less than 1 ms and completed in less than 1 ms.

```
MATCH (t:Team)-[r:INTERACTS]-()
RETURN t.name AS team, sum(r.Spread) AS SpreadTotal, sum(r.OverUnder) AS OverUnderTotal ORDER BY
SpreadTotal DESC
```

	team	SpreadTotal	OverUnderTotal
1	"Las Vegas Raiders"	-70	854
2	"Los Angeles Chargers"	-72	835
3	"Seattle Seahawks"	-75	793
4	"Pittsburgh Steelers"	-76	738
5	"Cleveland Browns"	-77	767
6	"Minnesota Vikings"	-77	821
7	"Philadelphia Eagles"	-81	839

Started streaming 32 records in less than 1 ms and completed in less than 1 ms.

As you can see here, the correlation between the two is not every apparent, but they are linked nonetheless. The weighted centrality and degree centrality also don't produce the same results as the measure becomes different when these teams' are investigated more closely.

K1 Coloring

K1 coloring is an algorithm that gives colors to every node on the graph. We try to use few colors as needed, but every node should have a different color than its neighbor. In this case, we have thirty-two teams so different colors were needed to be use. There are a couple base colors then variations were used to broaden the spectrum.

```
neo$ CALL gds.graph.project('myGraph', 'Team', 'INTERACTS')
```

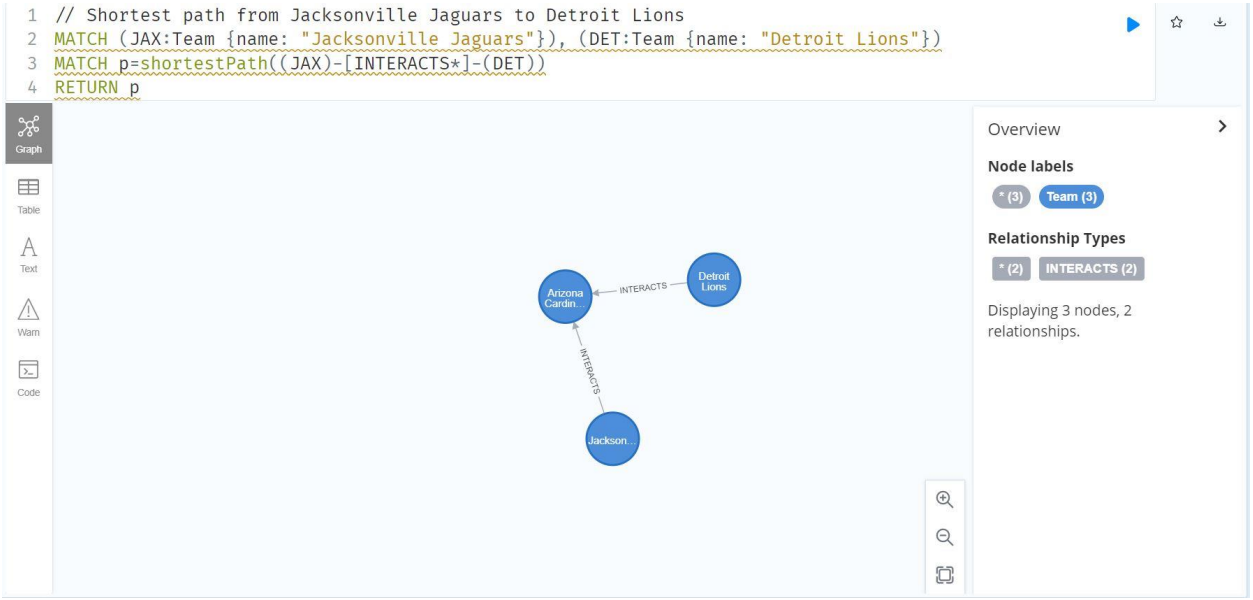
```
CALL gds.beta.k1coloring.stream(myGraph: String, configuration: Map)
YIELD nodeId, color
```

Pathfinding

In pathfinding, we explore the routes between nodes. The routes weave the interactions of the nodes until it arrives at the destination of the node wanted or the quickest path throughout.

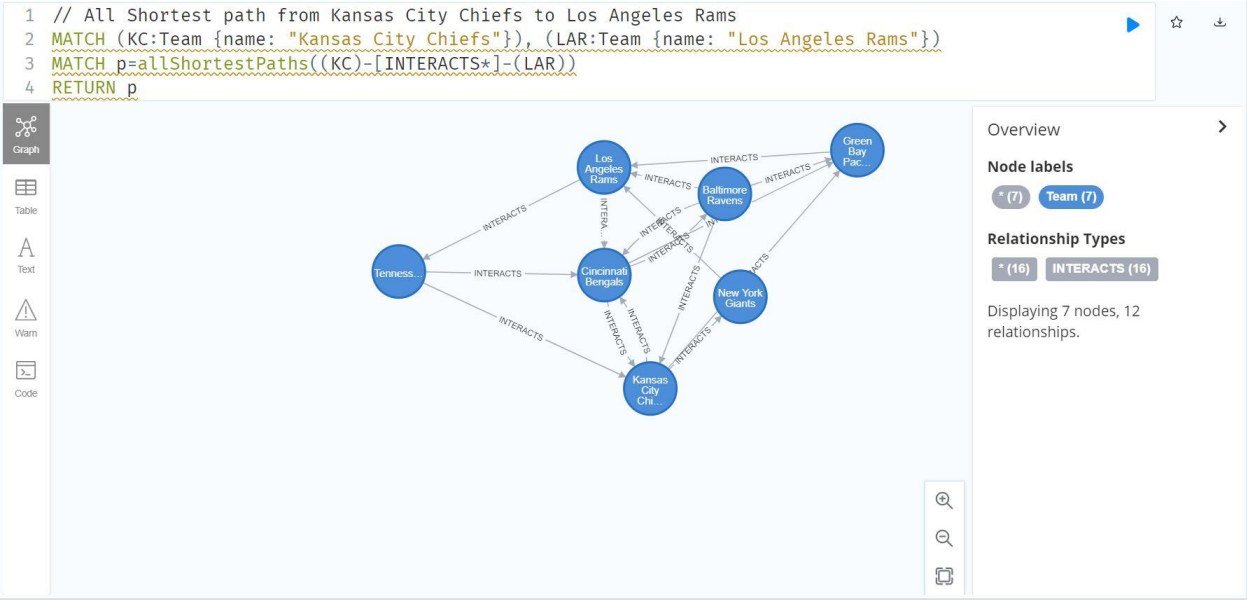
Shortest Path

The shortest path algorithm is the calculation of the shortest path between a pair of nodes. This is helpful because it works in real time and gives answers about degrees of separation. We usually use shortest path for social media. For example, on Facebook, it could tell how many people separate you to your mail main for instance. In this case, we can determine the shortest path to compare teams that don't play each other. For example, the Jacksonville Jaguars and the Detroit Lions were the worst teams in the league this year, but they did not play each other. To see who is the true worst team, we could compare the games both teams had with the Arizona Cardinals and compare the scores and statistics.



All Pairs Shortest Path

The all pairs shortest path measures the shortest path between all pairs of nodes. It keeps track of the distances calculated with all nodes running. All Pairs Shortest Path (APSP) is used most often when the shortest route is blocked. I look at this in terms of Google Maps. Apple Maps may show you the fastest way, but Google may show you the way if you are thinking about stopping at the gas station on the way. In this graph, we can determine how two teams would have fared if they played each other, but with more data to back it up. In the model below, the Rams and Chiefs did not play each other, but were two of the best teams in the league. This shows other good teams that they encountered and provides a full picture of what may have happened.



Conclusions

In examining this dataset and querying the data in Neo4j, there are a couple conclusions that I could reach. The point was to figure out exactly how my prior knowledge on NFL football could influence making some money on betting. First, on betting the over/under, one should

bet the over on teams what are average, but are built on the offensive side of the ball. The Minnesota Vikings are no one's favorite to win the Super Bowl, but they are built to score some points with players like Justin Jefferson. They are a great team for the over because they are going to score points, but are not necessarily going to stop the other team. Average teams could make wallets above average. Next, we learned that betting the spread is not an exact science. Good teams or bad teams, doesn't make a difference. A few teams may have been best last year, but this upcoming year, it may be a toss-up. Lastly, home field does matter, but if you're a good team then you're going to win games, whether home or away. It makes the games a lot better to experience both as players and in the crowd when the advantage is great, and it does correlate to the field.

Appendix

Constraint

```
CREATE CONSTRAINT ON (t:Team) ASSERT t.name IS UNIQUE;
```

Graph Model

```
MATCH n=(:Team)-[:INTERACTS]-(:Team) RETURN n
```

Shortest Path

```
// Shortest path from Jacksonville Jaguars to Detroit Lions
MATCH (JAX:Team {name: "Jacksonville Jaguars"}), (DET:Team {name: "Detroit Lions"})
MATCH p=shortestPath((JAX)-[INTERACTS*]-(DET))
RETURN p
```

All Shortest Path

```
// All Shortest path from Kansas City Chiefs to Los Angeles Rams
MATCH (KC:Team {name: "Kansas City Chiefs"}), (LAR:Team {name: "Los Angeles Rams"})
MATCH p=allShortestPaths((KC)-[INTERACTS*]-(LAR))
```

```
RETURN p
```

Degree Centrality

```
MATCH (t:Team)
RETURN t.name AS team, size( (t)-[:INTERACTS]-() ) AS degree ORDER BY degree DESC
```

Weighted Degree Centrality

```
MATCH (t:Team)-[r:INTERACTS]-()
RETURN t.name AS team, sum(r.Spread) AS SpreadTotal, sum(r.OverUnder) AS OverUnderTo
tal ORDER BY SpreadTotal ASC
```

```
MATCH (t:Team)-[r:INTERACTS]-()
RETURN t.name AS team, sum(r.OverUnder) AS OverUnderTotal, sum(r.Spread) AS SpreadTo
tal ORDER BY OverUnderTotal DESC
```

```
MATCH (t:Team)-[r:INTERACTS]-()
RETURN t.name AS team, avg(r.Spread) AS SpreadTotal, avg(r.OverUnder) AS OverUnderTot
al ORDER BY SpreadTotal ASC
```

Meet Over/Under

```
MATCH (n:Team), (:Team)-[p:INTERACTS]-
(:Team) WHERE (p.AwayScore + p.HomeScore) > p.OverUnder RETURN p, n
```

Times Exceed Over/Under

```
MATCH (n:Team)-[p:INTERACTS]-
(t:Team) WHERE (p.AwayScore + p.HomeScore) > p.OverUnder RETURN n as Team, count(t)
 AS Games ORDER BY Games DESC
```

```
MATCH (n:Team {name: 'Los Angeles Rams'})-(t:Team) WHERE (p.AwayScore +
p.HomeScore) > p.OverUnder RETURN p, n, t
```

Home Games Over OFF Avg, Under DEF Avg.

```
MATCH (h:Team)-[p:INTERACTS]-
(a:Team) WHERE (p.AwayScore < 23) AND (p.HomeScore > 23) RETURN h AS Team, p.Stad
ium AS Stadium, count(h) AS GreatGames ORDER BY GreatGames DESC
```

Spread Not Met

```
MATCH (h:Team)-[p:INTERACTS]-
(a:Team) WHERE (p.AwayScore - p.HomeScore) < p.Spread RETURN h.name AS Team, p.Spr
ead AS SpreadSet, p.FavoriteID AS Favorite, p.Week AS Week
```