

Wins Predictor Model



Project: NFL Team Wins Prediction Model

Selected topic and reasoning

We chose the topic of football win totals based on previous season historical data. We chose this to see if we could use multiple linear regression analysis that is close to the predicted win O/U win totals presented by Vegas casinos before the start of the regular NFL season.

Great	          
Good	     
Average	      
Meh	   
Bad	   

Sources



We will be collecting historical team data as well as betting data from the previous season and uploading into an excel file and creating SQL database tables. in these tables will be based on the previous season win totals, head coach return, total offensive ranking, total defensive ranking, Team QBR rating, OC remains, and DC remains

Questions we hope to answer

We hope to answer the question of what goes into the over/under totals that are given to teams. We want to know what the most important factor on if a team can be good or if a bad over under total can spark a team to a better record then was projected. and most importantly do these metrics lead to a team being over the projected total.



It's in the Data

We used data from ProFootball Reference, NFL.com, ESPN.com, and individual team websites.

We gathered offensive rating, defensive rating and QBR stats from ProFootball Reference and ESPN.com

NFL.com we found the divisions and their seasonal records for wins and losses.

Individual team websites provided if the head coaching returned or not information



Teams

Out[27]:

	team_name	div	team_id
0	Arizona	NFC	NFC001
1	Atlanta	NFC	NFC002
2	Baltimore	AFC	AFC001
3	Buffalo	AFC	AFC002
4	Carolina	NFC	NFC003
5	Chicago	NFC	NFC004
6	Cincinnati	AFC	AFC003
7	Cleveland	AFC	AFC004
8	Dallas	NFC	NFC005
9	Denver	AFC	AFC005
10	Detroit	NFC	NFC006
11	Green Bay	NFC	NFC007
12	Houston	AFC	AFC006
13	Indianapolis	AFC	AFC007
14	Jacksonville	AFC	AFC008
15	Kansas City	AFC	AFC009
16	Las Vegas	AFC	AFC010
17	Los Angeles Chargers	AFC	AFC016
18	Los Angeles Rams	NFC	NFC008
19	Miami	AFC	AFC011

15	Kansas City	AFC	AFC009
16	Las Vegas	AFC	AFC010
17	Los Angeles Chargers	AFC	AFC016
18	Los Angeles Rams	NFC	NFC008
19	Miami	AFC	AFC011
20	Minnesota	NFC	NFC009
21	New England	AFC	AFC012
22	New Orleans	NFC	NFC010
23	NY Giants	NFC	NFC011
24	NY Jets	AFC	AFC013
25	Philadelphia	NFC	NFC012
26	Pittsburgh	AFC	AFC014
27	San Francisco	NFC	NFC013
28	Seattle	NFC	NFC014
29	Tampa Bay	NFC	NFC015
30	Tennessee	AFC	AFC015
31	Washington	NFC	NFC016

Wins/Losses 2019-2021 Percentages

Out[38]:	team_name	div	team_id	wins_2021	losses_2021	win_pct_2021	wins_2020	losses_2020	win_pct_2020	wins_2019	...	qbr_2019	off_ranking_2021	off_ranking_2020	off_2019
0	Arizona	NFC	NFC001	11	6	0.647	8	8	0.5	5	...	87	8	6	
1	Atlanta	NFC	NFC002	7	10	0.412	4	12	0.25	7	...	93.8	29	18	
2	Baltimore	AFC	AFC001	8	9	0.471	11	5	0.688	14	...	109	6	19	
3	Buffalo	AFC	AFC002	11	6	0.647	13	3	0.813	10	...	82.9	5	2	
4	Carolina	NFC	NFC003	5	12	0.294	5	11	0.313	5	...	74.7	30	21	
5	Chicago	NFC	NFC004	6	11	0.353	8	8	0.5	8	...	83.9	24	26	
6	Cincinnati	AFC	AFC003	10	7	0.588	4	11	0.281	2	...	76.2	13	29	
7	Cleveland	AFC	AFC004	8	9	0.471	11	5	0.688	6	...	78.4	18	16	
8	Dallas	NFC	NFC005	12	5	0.706	6	10	0.375	8	...	99.5	1	14	
9	Denver	AFC	AFC005	7	10	0.412	5	11	0.313	7	...	84.1	19	23	
10	Detroit	NFC	NFC006	3	13	0.206	5	11	0.313	3	...	88.2	22	20	
11	Green Bay	NFC	NFC007	13	4	0.765	13	3	0.813	13	...	95.3	10	5	
12	Houston	AFC	AFC006	4	13	0.235	4	12	0.25	10	...	95.3	32	13	
13	Indianapolis	AFC	AFC007	9	8	0.529	11	5	0.688	7	...	85	16	10	
14	Jacksonville	AFC	AFC008	3	14	0.176	1	15	0.063	6	...	90	27	28	
15	Kansas City	AFC	AFC009	12	5	0.706	14	2	0.875	12	...	104.4	3	1	
16	Las Vegas	AFC	AFC010	10	7	0.588	8	8	0.5	5	...	85.6	9	11	
17	Los Angeles Chargers	AFC	AFC016	9	8	0.529	7	9	0.438	5	...	101	11	8	
18	Los Angeles Rams	NFC	NFC008	12	5	0.706	10	6	0.625	9	...	89	4	9	
19	Miami	AFC	AFC011	9	8	0.529	10	6	0.625	12	...	80	25	22	
20	Minnesota	NFC	NFC009	8	9	0.471	7	9	0.438	10	...	103.9	12	4	

Win/Loss 2019-2021 continued...

21	New England	AFC	AFC012	10	7	0.588	7	9	0.438	7 ...	88.2	15	27
22	New Orleans	NFC	NFC010	9	8	0.529	12	4	0.75	13 ...	110.2	28	12
23	NY Giants	NFC	NFC011	4	13	0.235	6	10	0.375	4 ...	80.6	26	32
24	NY Jets	AFC	AFC013	4	13	0.235	2	14	0.125	7 ...	86.4	31	31
25	Philadelphia	NFC	NFC012	9	8	0.529	4	11	0.281	9 ...	92.1	14	24
26	Pittsburgh	AFC	AFC014	9	7	0.559	12	4	0.75	8 ...	76.1	23	25
27	San Francisco	NFC	NFC013	10	7	0.588	6	10	0.375	13 ...	103.1	7	15
28	Seattle	NFC	NFC014	7	10	0.412	12	4	0.75	11 ...	105.3	20	17
29	Tampa Bay	NFC	NFC015	13	4	0.765	11	5	0.688	7 ...	84.1	2	7
30	Tennessee	AFC	AFC015	12	5	0.706	11	5	0.688	9 ...	108.3	17	3
31	Washington	NFC	NFC016	7	10	0.412	7	9	0.438	3 ...	83	21	30

32 rows × 24 columns

Extract, Transform, & Load Phase

We used the machine learning model linear regression

We collected seasonal data for each NFL Team

Each season win total we placed into a column as our Y value of our prediction

We created the tables with Postgres SQL using joins to make on large dataset

Using Python and Jupyter Notebook we input all the important columns into dataframes

We merged each data frame on Team name to make sure all the data matched up

Teams team_name div team_id	Sales team_id team_name sell_out_rate	QBR_table team_id qbr_average
Head_Coach team_id team_name coach_2020 coach_2021	NFC_Standings team_id team_name wins losses ties w-l%	Team_rating team_id team_name offensive_rate defensive_rate

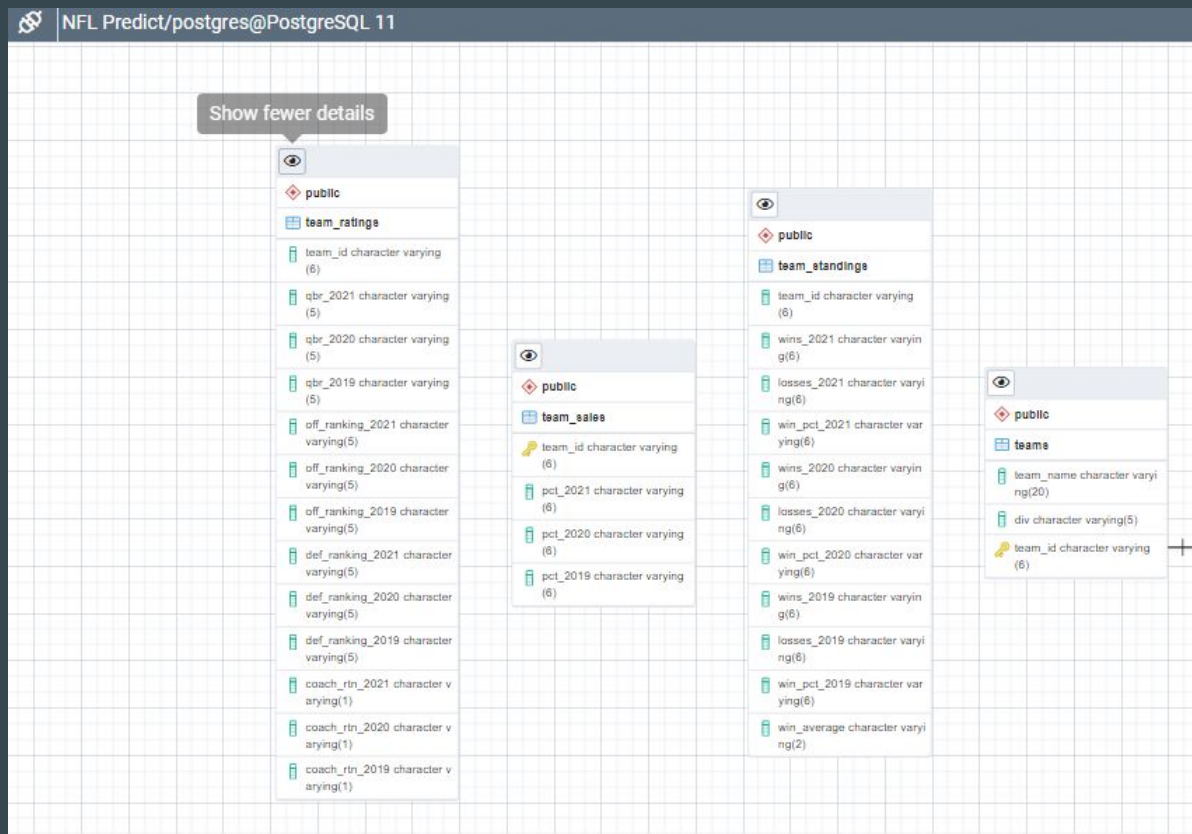
Extract, Transform, & Load Phase

After completing the initial phase of extracting and collecting NFL data we realized the need to increase the size of our dataset.

We increased the dataset by increasing the number of teams from 16 to all 32 teams of the NFL.

The dataset was also increased by expanding the data to include 3 years (2021, 2020, & 2019) instead of 1 year.

Our original 6 tables were joined in SQL and connected to our Jupyter notebook.



Extract, Transform, & Load Phase

The final code we used to create our tables in SQL is displayed here. As we expanded our dataset we needed to drop some of our original tables from the SQL database and rebuild them with new columns for the additional football data being included in the dataset.

```
1 -- Creating tables for NFL Predictor
2 CREATE TABLE teams (
3     team_name VARCHAR(20) NOT NULL,
4     div VARCHAR(3) NOT NULL,
5     team_id VARCHAR(6) NOT NULL,
6     PRIMARY KEY (team_id),
7     UNIQUE (team_name,team_id)
8 );
9
10 CREATE TABLE standings(
11     team_id VARCHAR(6) NOT NULL,
12     wins_2021 VARCHAR(6) NOT NULL,
13     losses_2021 VARCHAR(6) NOT NULL,
14     win_pct_2021 VARCHAR(6) NOT NULL,
15     wins_2020 VARCHAR(6) NOT NULL,
16     losses_2020 VARCHAR(6) NOT NULL,
17     win_pct_2020 VARCHAR(6) NOT NULL,
18     wins_2019 VARCHAR(6) NOT NULL,
19     losses_2019 VARCHAR(6) NOT NULL,
20     win_pct_2019 VARCHAR(6) NOT NULL,
21     FOREIGN KEY (team_id) REFERENCES teams(team_id),
22     PRIMARY KEY (team_id)
23 );
24
25 CREATE TABLE team_sales(
26     team_id VARCHAR(6) NOT NULL,
27     pct_2021 VARCHAR(6) NOT NULL,
28     pct_2020 VARCHAR(6) NOT NULL,
29     pct_2019 VARCHAR(6) NOT NULL,
30     FOREIGN KEY (team_id) REFERENCES teams(team_id),
31     PRIMARY KEY (team_id)
```

```
CREATE TABLE team_ratings (
    team_id VARCHAR(6) NOT NULL,
    qbr_2021 VARCHAR(5) NOT NULL,
    qbr_2020 VARCHAR(5) NOT NULL,
    qbr_2019 VARCHAR(5) NOT NULL,
    off_ranking_2021 VARCHAR(5) NOT NULL,
    off_ranking_2020 VARCHAR(5) NOT NULL,
    off_ranking_2019 VARCHAR(5) NOT NULL,
    def_ranking_2021 VARCHAR(5) NOT NULL,
    def_ranking_2020 VARCHAR(5) NOT NULL,
    def_ranking_2019 VARCHAR(5) NOT NULL,
    coach_rtn_2021 VARCHAR(1) NOT NULL,
    coach_rtn_2020 VARCHAR(1) NOT NULL,
    coach_rtn_2019 VARCHAR(1) NOT NULL
);

-- Join 'standings' table with 'teams' table
SELECT teams.team_id,standings.wins_2021, standings.losses_2021,standings.win_pct_2021,
standings.wins_2020, standings.losses_2020,standings.win_pct_2020,
standings.wins_2019, standings.losses_2019,standings.win_pct_2019
INTO team_standings
FROM teams
INNER JOIN standings
ON teams.team_id = standings.team_id;
```

Machine Learning Model - Linear Regression

Why linear regression?

“Linear-regression models have become a proven way to scientifically and reliably predict the future. Because linear regression is a long-established statistical procedure, the properties of linear-regression models are well understood and can be trained very quickly”.

(IBM website)

```
# Create our features
X = X_drop

# Create our target
y = NFL_DF["win_average"]

model = LinearRegression()
model.fit(X, y)

LinearRegression()

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=1, test_size =0.70)

model = LinearRegression()

model.fit(X, y)

LinearRegression()

y_pred = model.predict(X_test)
results = pd.DataFrame({"Prediction": y_pred, "Actual": y_test})

results
```

Machine Learning Model - Linear Regression

We divided our dataset into a 30/70 split within our linear regression model.

Features:

- Team Defensive Ranking (2019 -2021)
- Team Offensive Ranking (2019 -2021)
- Quarterback Rating (2019 -2021)
- Head Coach Return (2019 -2021)

Based on our linear regression model output we were able to very accurately predict the upcoming seasonal win totals using the selected features.

	Prediction	Actual
27	9.696243	10
3	10.904316	11
22	11.311308	11
18	10.451697	10
23	4.819526	5
17	7.304016	7
21	7.895677	8
28	9.986550	10
19	10.045337	10
14	3.228298	3
24	4.255365	4
10	3.939531	4
20	8.177007	8
26	10.060469	10
4	4.871696	5
2	11.016987	11
25	6.967059	7
6	5.032941	5
13	8.901431	9
7	8.144317	8
29	10.164000	10
1	5.969867	6
16	7.772368	8

For Example: NFC South

Carolina(4) had a projected win total of 4.87. Vegas O/U is 5.5. Very little changes to team.

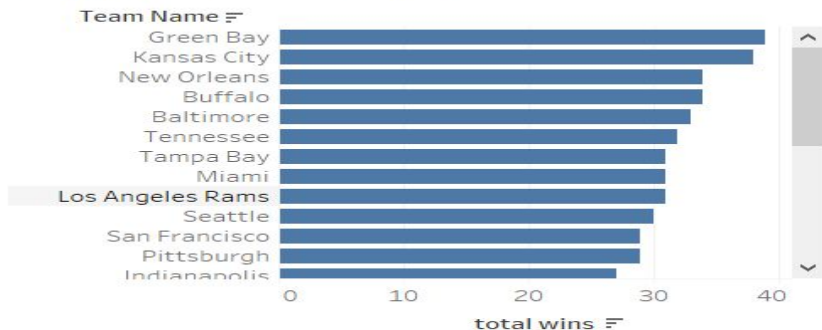
Atlanta (1) predicted is 5.97 and Vegas O/U is 5 only changing QB

New Orleans(22) was 11.3, over under in Vegas is 7.5 but they had a change in coach and QB.

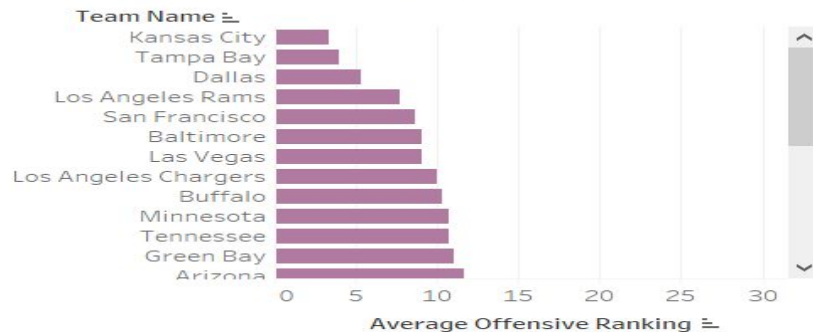
Tampa Bay(29) predicted 10.16, 11.5 from Vegas with a Head Coach change.

Metrics Equaling Wins?

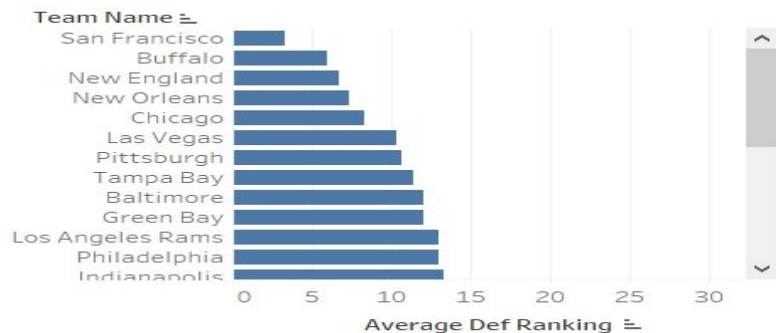
Team Wins from 2019-2020



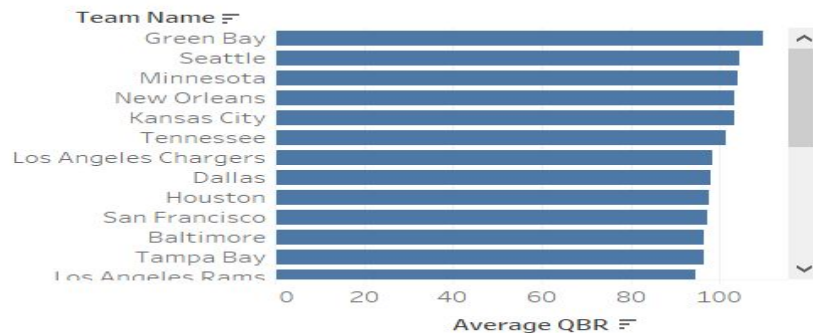
AVG Offensive Ranking 2019-2021



Avg Defensive Ranking

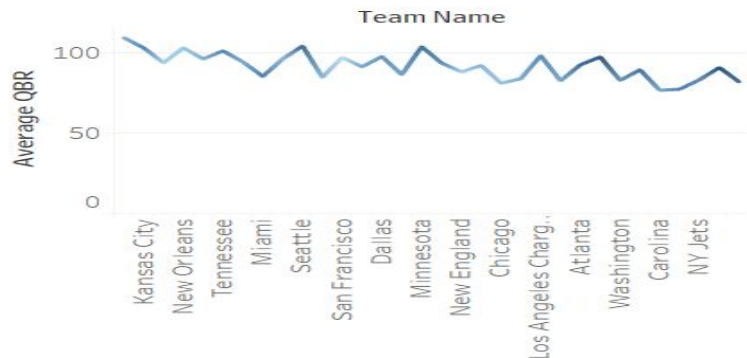


Average QBR from 2019-2022

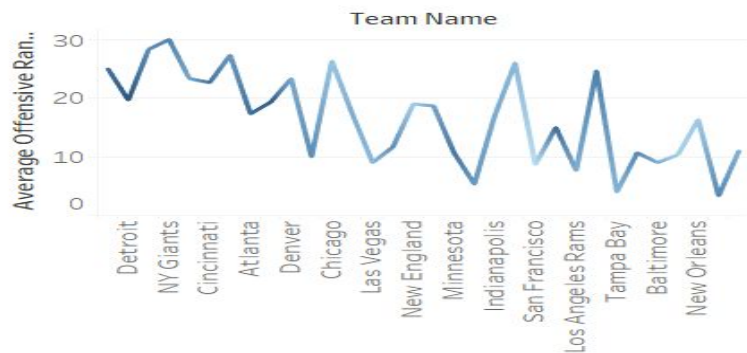


Metrics Equaling Wins?

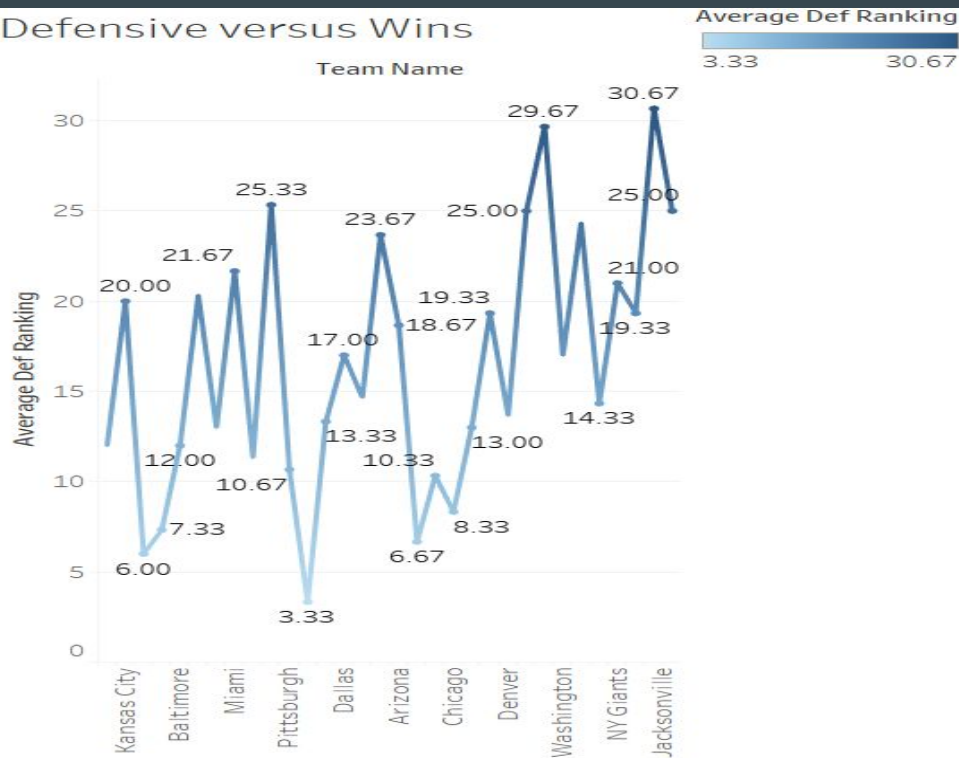
QBR Trend vs Wins



Wins versus Offensive Ranking



Defensive versus Wins



Technologies

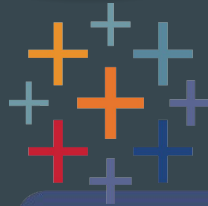
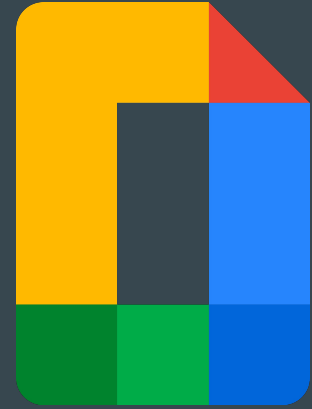
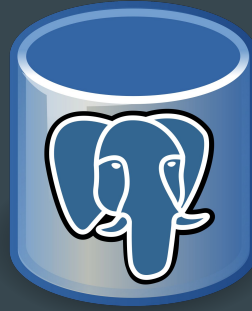
Python

Jupyter Notebook

Postgres SQL

Google Slides

Tableau



+ a b l e a u[®]



plotly