Exploratory Analysis on the Effect of Home-Team Advantage in the National Football League (NFL) using Spread and Scoring Statistics

David Stehr (202163162)

Introduction:

In every sport, playing in front of your own fans is considered an advantage. Home field advantage is a term that has been coined now for many generations throughout all team sports. But does playing in your own stadium, in front of crowds of fans cheering you on really make a difference? Throughout this report, we will be exploring the impact of home-field advantage in the National Football League (NFL) throughout 30-years of matchups using scoring and gambling statistics .

The National Football League (NFL) is the one of the most viewed sports leagues globally and generates billions of dollars in revenue annually. The legalization of online sports-betting throughout the US offers large benefits for the NFL, as the legalization is projected to not only increase the amount gambled on the league, but also lead to a projected 1.75 billion USD in new revenue through increased media-rights, sponsorship deals, increased fan-participation, and advertising[1].

The most popular method to bet on the NFL is spread betting, which is a method devised by mathematician Charles McNeil in 1940s. Spread betting aims to 'even the odds' between matchups, with the aim of creating an even-split of money bet between both teams[2]. In many cases, not just in the NFL, a team with superior talent will be favored over an 'underdog'. The betting returns of picking the favorite will be limited, therefore, to increase the returns for the favored team and to create equal number of bettors between both teams, the bookmaker will assign a 'spread' that the favor-team must exceed, which will offer more favorable returns for the bettor, usually resulting in even to near-even returns for betting either side of the point-spread[3].

For example, heading into Week 15 of the 2007 NFL Season. The New England Patriots had amassed an undefeated record up until this point, going 14-0. They faced the New York Jets, who to this point had managed a record of 3-10. The Vegas bookmakers had installed the Patriots as clear favorites for this matchup, with the Patriots playing on their home-field, Gilette Stadium, with a far superior roster in terms of talent at their disposal. The spread for the Patriots was set at –20.5, which meant the Patriots would have to score 21 points more than the Jets to 'cover' the spread. On the opposite spectrum, the New York Jets have a 20.5 point 'handicap'. The game ended with the score of 20-10 for the New England Patriots, with the Patriots extending their unbeaten streak but failing to 'cover' the spread[4].

---

[1] Betting Revenue for NFL – Source: (https://www.playusa.com/nfl-legal-sports-betting-revenue/)

[2] Spread Betting on Sports Explained – Source: (https://w88au.net/sports-spread-betting/)

[3] Spread Betting on Sports Explained – Source: (https://w88au.net/sports-spread-betting/)

[4] Recap of 2007 Matchup between Patriots vs. Jets – Source: (https://www.nfl.com/games/jets-at-patriots-2007-reg-15)

Preliminary Analysis

Data Collection and Cleaning

All data was collected via Kaggle[5], consisting of an 18 x 13187 table initially. After cleaning the dataset, we were left with 7688 observations per column. Initial relevant variables available being;

- Schedule-date – the date of the matchup
- Schedule-season – the season for the matchup
- Schedule-week – NFL Season consists of 17 weeks; weeks go from 1-17
- Schedule-playoff – whether the matchup was taking place in the regular season or playoffs(T/F)
- Team-home – the name of home team in the matchup
- Score-home – the score of the home team
- Team-away – the name of the away team
- Score-away – score of the away team
- Team-favorite-id – the short-form of the spread favorite team
- Spread-favorite – the value of spread for the favored team
- Over-under-line – the projected total score for the matchup

Additionally, several variables were added to further the analysis, these consisted of the absolute value of the spread, as spread for favorited teams is expressed with a negative value to signify the additional points the team must exceed in order to 'cover' the spread, variable 'actual-result' which represents the actual score difference, score-home and score-away, from the home-team's perspective, and several binary classifiers such as:

- Home-team-fav – if the team-id for favored team matches the home-team name
- Covered-spread – if the favored team covered the spread (variable actual-result must be greater than spread in absolute terms)

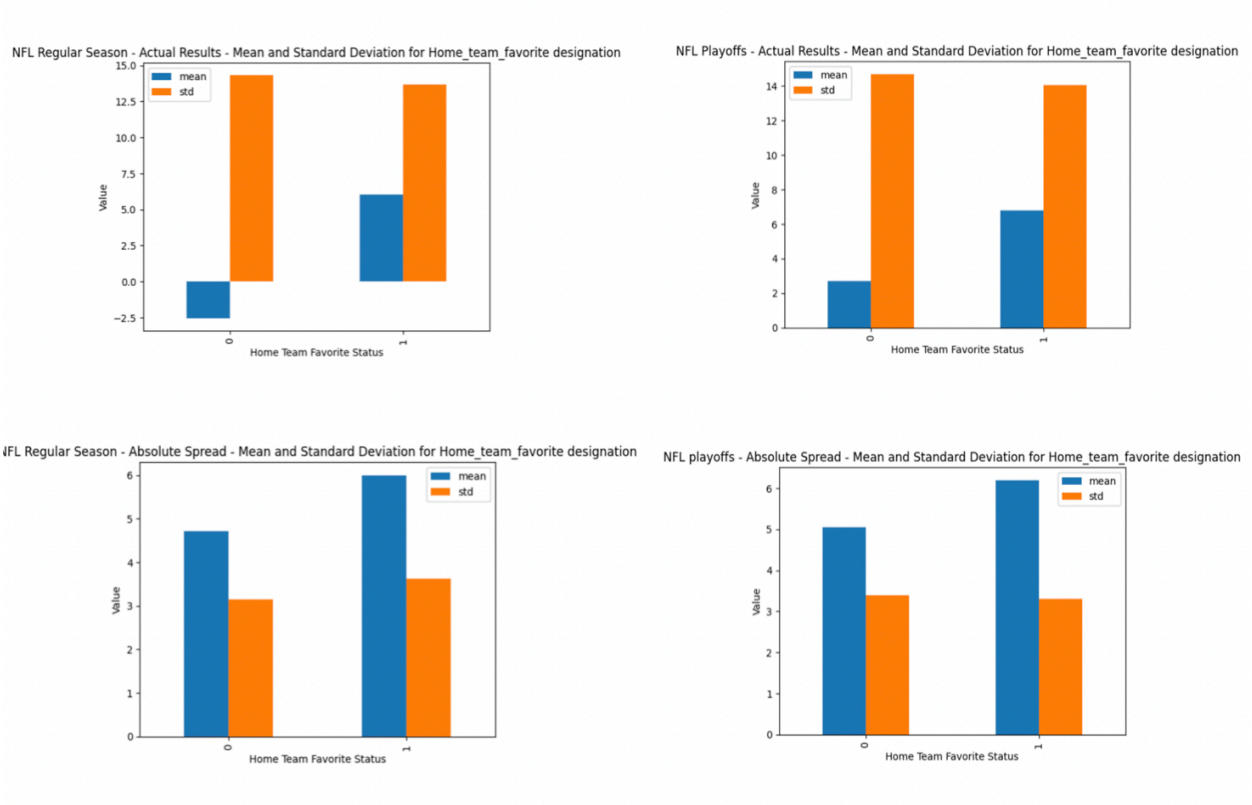
Following the addition of other variables, the dataset was split into two, one consisting of matchups covering the NFL Regular Season, and the other consisting of NFL Playoff only matchups. The purpose behind this is further capture the impact of home-team advantage. As mentioned before, the NFL Regular Season consists of 17 weeks, where 32-teams, which are split into two conferences, and then further divided into 4 divisions with 4 teams per division, face-off in pre-determined matchups to determine playoff qualification. To qualify, a team must finish with the best record in their respective division, where each division winner is seeded 1-4 based on number of wins or can qualify for one of 6

[5] Dataset Source – Source: (https://www.kaggle.com/tobycrabtree/nfl-scores-and-betting-data?select=nfl_teams.csv)

wild-card slots by being the team with the most wins from the remaining field in their conference, which are seeded 5-7 based on number of wins[6]. Throughout the playoffs, higher seeded teams retain home-team advantage

By filtering the NFL regular season-matchups in which home teams are favored to win, home teams were favored to win or cover the spread in 59.9% of the regular season matchups. Looking at summary statistics, we can see that the average absolute value of the spread issued by bookmakers is larger when the home team is favored, and also shows more variability in absolute spread. This would support the concept that home-field advantage does play a role in establishing spread favorites. Looking at the actual results variable from the matchups, the average actual scoring results would support the notion that home-field advantages play a significant factor in scoring, yet we can see that there is a higher variation when the home-team is not favored, which would indicate that scoring is greatly varied, regardless of matchups.

Figure 1 - 1.1 : NFL Regular Season vs. NFL Playoffs – Actual Result vs. Absolute Mean - Mean and Standard Deviation for Home-Team Favorite Designation



[6] NFL Playoff And Qualifications – Source:(https://blog.betway.com/nfl/nfl-explained-how-the-regular-season-and-playoffs-work/)

Table 1.1 : Grouped by Home-team favorite – NFL Regular Season Summary Statistics for Results and Spread

	actual_result								spread_abs							
	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%	50%	75%	max
home_team_fav																
0	3210.0	-2.553271	14.330008	-49.0	-11.0	-3.0	6.0	52.0	3210.0	4.712305	3.140402	0.0	2.5	3.5	6.5	18.5
1	4478.0	6.051586	13.669273	-45.0	-3.0	6.0	14.0	59.0	4478.0	5.995199	3.627004	1.0	3.0	5.5	8.0	26.5

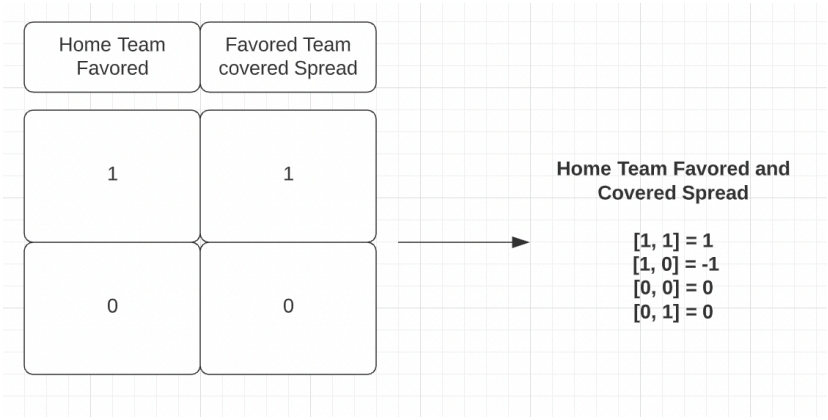
For playoff matchups, home teams were favored in 79.3% of the matchups, with both mean scoring and mean spread values being significantly higher for home teams favored, which supports the notion that home-field advantage is a bigger factor in playoffs. The standard deviation of scoring and spread were slightly higher when the home team was not favored, indicating larger variation in scoring and spread values for these matchups, but were relatively similar to the NFL regular statistics.

Table 1.2: Grouped by Home-team Favourite – NFL Playoff Summary Statistics for Results and Spread

	actual_result								spread_abs							
	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%	50%	75%	max
home_team_fav																
0	71.0	2.690141	14.674167	-30.0	-7.0	2.0	12.5	41.0	71.0	5.049296	3.389221	0.0	3.0	3.5	7.0	14.0
1	272.0	6.768382	14.034802	-35.0	-3.0	6.0	17.0	55.0	272.0	6.198529	3.310098	1.0	3.5	6.0	8.0	18.5

To further capture the impact of playing at home, we establish a two-by-two matrix that takes two previously defined binary criteria, whether the home team in the matchup was favored and if the team favored managed to cover the spread:

Figure 1.1- Matrix and Resulting Index that corresponds to home-team favourite status and covering spread.



The matrix yields three results; 1 indicating that the home-team was favored and managed to cover the spread, -1 indicating that the home-team was favored but failed to cover the spread, and 0 indicating that the home-team was not favored in the matchup. This criterion suits the analysis as it adds a further indicator to better capture the impact of home-field advantage, as it now includes a measure to see if the predicted status as favorite can be considered legitimate using the criteria of 'covering spread' or actual-result score exceeding the absolute value of spread.

Using these results, we look deeper into the 30 years of matchups. From the 7474 observations gathered, 28.4% result in the home-team being favored and covering the spread, 29.7% result in the home-team being favored and failing to cover the spread. We can also see the average value of absolute spread is slightly higher when the home team is favored but fails to cover, and as well having a higher standard deviation. Additionally, the maximum value of absolute spread associated with home-teams favored and failing to cover the spread is 5 points higher than when the home-team does manage to cover[7]. A similar trend can be seen with the results column, with the score difference associated with home teams being favored and covering being the sole positive average from the table, which can be explained by the home-team needing to score more points when they manage to cover the spread.

[7] Please refer to Table 1.3

Figure 1.3

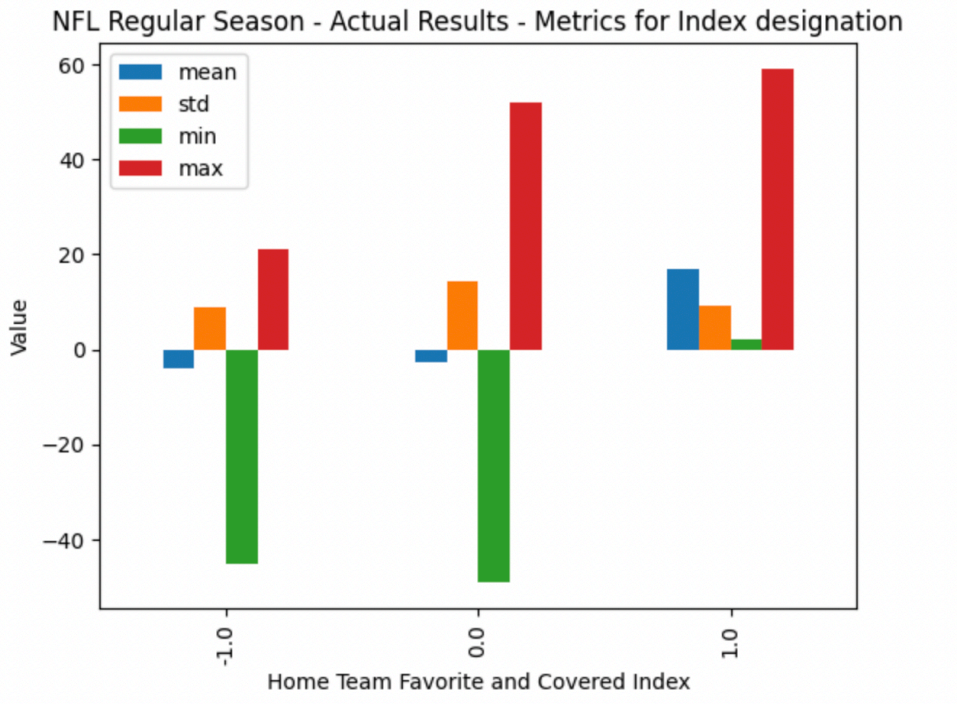
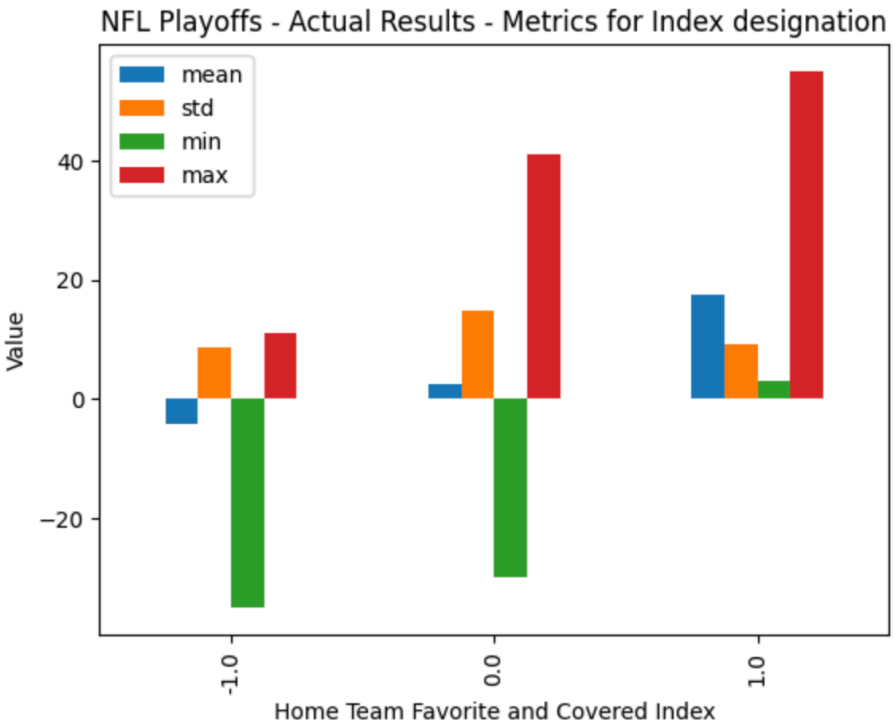


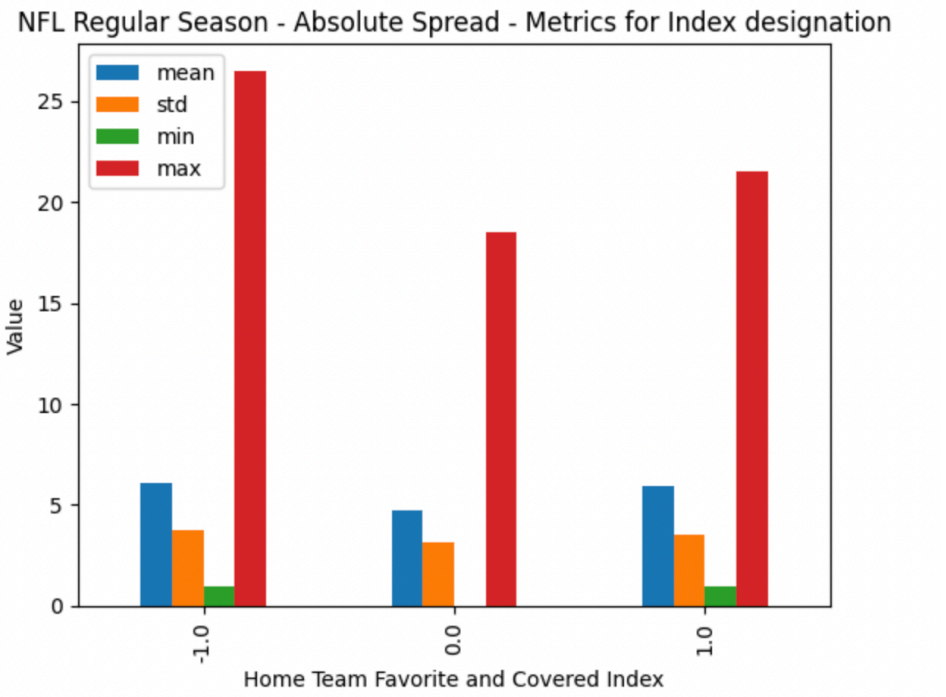
Figure 1.4 - Table 1.3

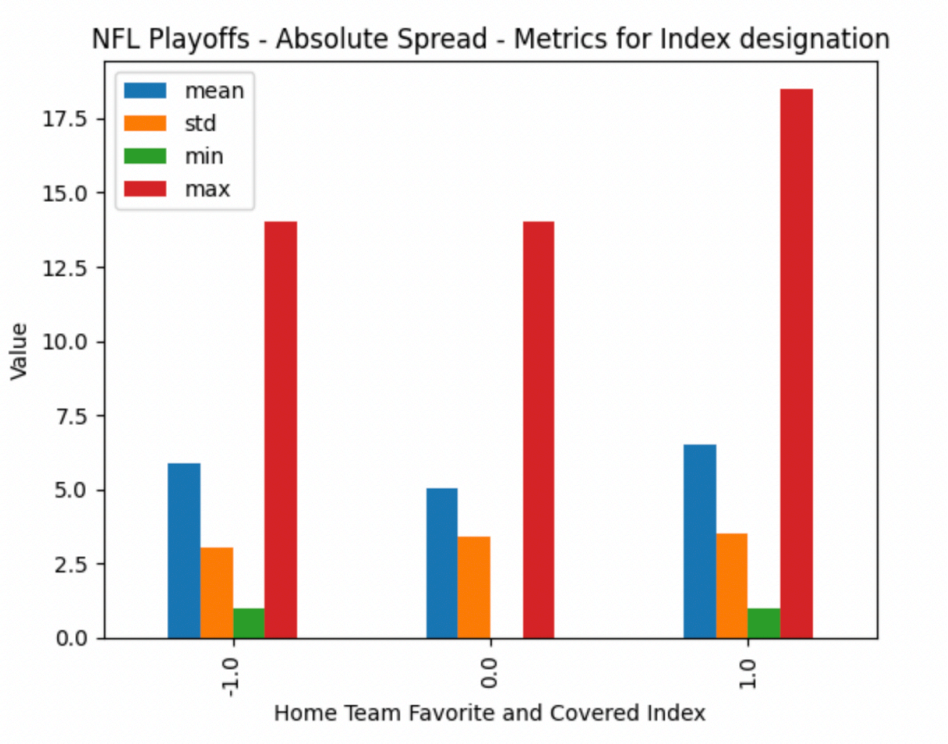


	actual_result								spread_abs							
	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%	50%	75%	max
home_fav_and_covered																
-1.0	2220.0	-4.114414	8.987522	-45.0	-8.0	-3.0	3.0	21.0	2220.0	6.111261	3.725242	1.0	3.0	5.5	8.0	26.5
0.0	3134.0	-2.704531	14.465434	-49.0	-12.0	-3.0	6.0	52.0	3134.0	4.737237	3.155956	0.0	2.5	3.5	6.5	18.5
1.0	2120.0	16.751415	9.312771	2.0	10.0	15.0	23.0	59.0	2120.0	5.924292	3.550359	1.0	3.0	5.5	7.5	21.5

For playoff matchups, 39.65% resulted in the home team being favored and covering, relative to 38.8% of the cases where the home team failed to cover the spread[8]. In the first case, the mean score difference was significantly higher than for the other two criteria, which was expected. Similarly, the mean values and standard deviation for absolute spread were higher than for the case of the home-team being favored and covering, which can be attributed to the fact that higher-placed teams in the regular season receive home-field advantage in the playoffs and are expected to win the matchups.

Figure 1.5 - Figure 1.6 - Table 1.4





	actual_result								spread_abs							
	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%	50%	75%	max
home_fav_and_covered																
-1.0	133.0	-4.278195	8.653562	-35.0	-10.0	-3.0	3.00	11.0	133.0	5.879699	3.020856	1.0	3.0	6.0	8.0	14.0
0.0	70.0	2.628571	14.770880	-30.0	-7.0	1.5	12.75	41.0	70.0	5.021429	3.405490	0.0	3.0	3.5	7.0	14.0
1.0	136.0	17.573529	9.151913	3.0	11.0	17.0	22.00	55.0	136.0	6.500000	3.503966	1.0	3.5	6.5	8.0	18.5

Figure 1.7 – NFL Regular Season: Scatterplot Comparing Home and Covered Spread vs. Not Covering the spread

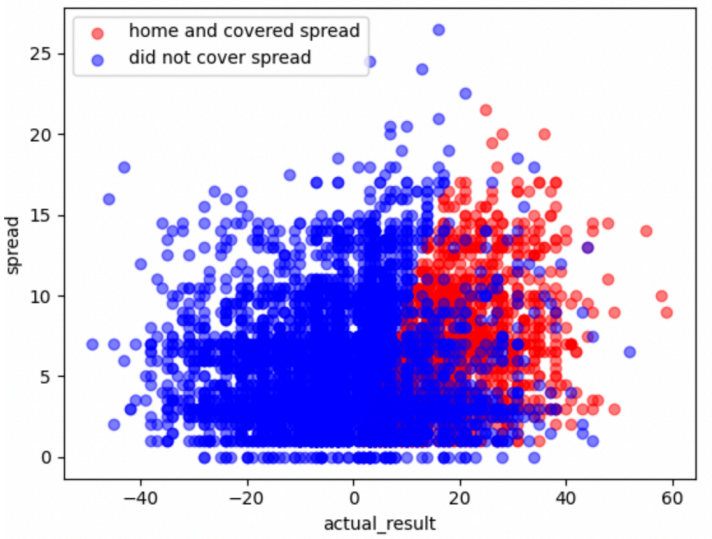
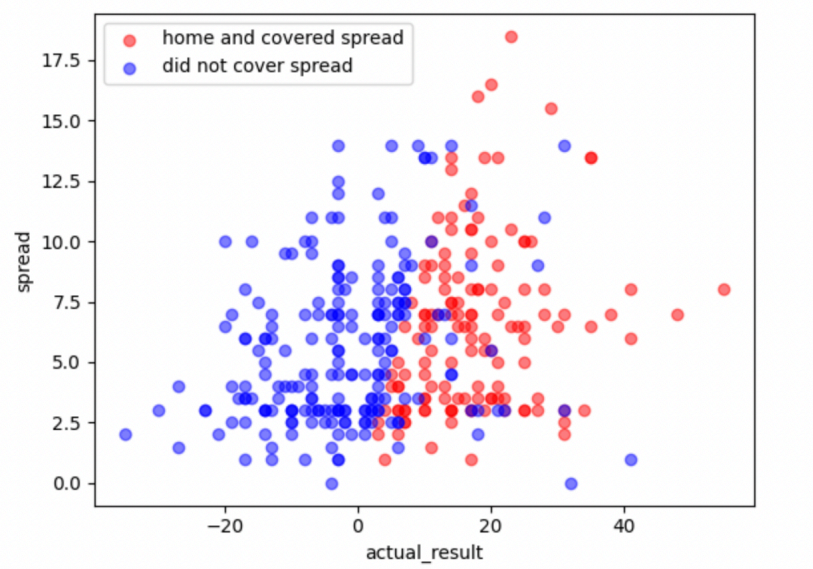


Figure 1.8 - NFL Playoff Matchups – Scatterplot comparing Scores and Spread for home-covered spread vs. Not covering spread



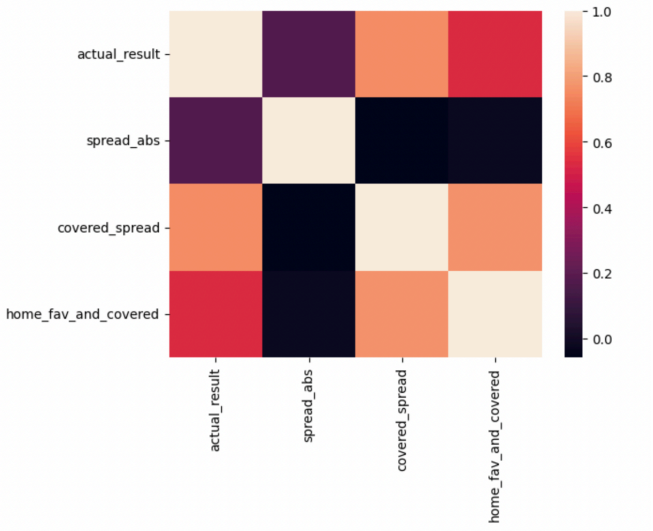
Looking at the graphs above and the scatter plot[9], we can see a similar trend highlighted in the table, as higher scores are associated with the home team being favored and covering the spread. There

appears to be no clear linear relationship in the scatter plot associated with scores and spread for the regular-season and playoffs matchups[10], but further investigation will be necessary.

Analysis

Performing initial correlation tests to gage the relationship between variables for the regular season matchups, there is a clear positive relationship (77.1%) between the home-team being favored and covering the spread and the binary indicator for the favored team covering the spread, which was to be expected. Additionally, there is a positive linear relationship (75.6%) between actual result scores and the binary indicator for spread, which indicates that the spread will be covered more often when scoring difference is largely positive[11]. Surprisingly, the home-team favored and covering the spread indicators show a weak positive correlation with actual scoring results and a very weak negative correlation to absolute values of spread.

Figure 2 / Table 2: Correlation Matrix with Pearson Correlations– NFL Regular Season

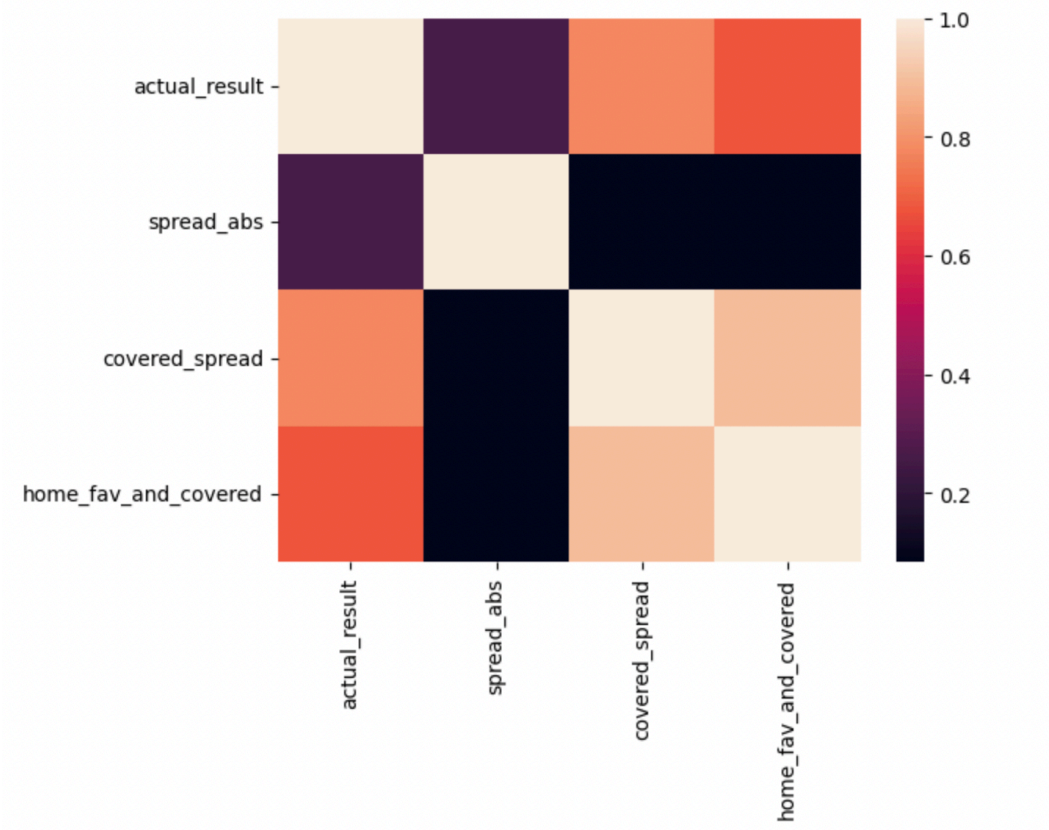


	actual_result	spread_abs	covered_spread	home_fav_and_covered
actual_result	1.000000	0.174804	0.756049	0.533577
spread_abs	0.174804	1.000000	-0.057953	-0.023021
covered_spread	0.756049	-0.057953	1.000000	0.771162
home_fav_and_covered	0.533577	-0.023021	0.771162	1.000000

For the playoff matchups, the correlation trends are similar to the regular season matchups but with slightly higher correlations for the home-favored and covered index, binary indicator for covering spread and the actual-result variable[12]. These patterns can be explained by the way the index is set up, as the actual-result variable would increase and show positive values associated with the home-team winning and covering the spread, but negative values when the index is negative, home-team favored but fails to cover, in which case the home-team result would be negative or below the value of absolute spread.

Figure 2.1 / Table 2.1: Correlation Matrix with Pearson Correlations – NFL Playoffs

	actual_result	spread_abs	covered_spread	home_fav_and_covered
actual_result	1.000000	0.261903	0.780675	0.681226
spread_abs	0.261903	1.000000	0.089423	0.083623
covered_spread	0.780675	0.089423	1.000000	0.891628
home_fav_and_covered	0.681226	0.083623	0.891628	1.000000



Unsupervised Learning Models

Hierarchical Clustering

Using Scikit-Learn hierarchical clustering tools, we can look to cluster the 3-value index for home-teams covering the spread for the regular season. In Agglomerative Clustering, a hierarchical clustering method, clusters are formed by measuring the 'distance' between points that are considered similar. The measure of distance is dependent on the method chosen, for the dataset in use, the best results appeared when using Euclidean-Distance over alternatives such as Manhattan or Cosine. The linkage that was chosen was complete, which looks to minimize the greatest distance between points in relative clusters, this was chosen over other methods such as average distance, which minimizes the average distance between points in clusters, or ward, which aims to minimize the variation, or variance, by minimizing the sum of squared distances between points.

Using the variables of the actual-result, absolute value of spread, the over-under-line, and the binary indicator for a favored team covering spread, with the aim of forming three clusters, the results show that clustering does not necessarily provide the best method of classifying the impact of home-field advantage throughout the regular season.

Figure 2.2 - **Dendrogram – Agglomerative Clustering with results limited to last 20 clusters and metrics for NFL Regular Season**

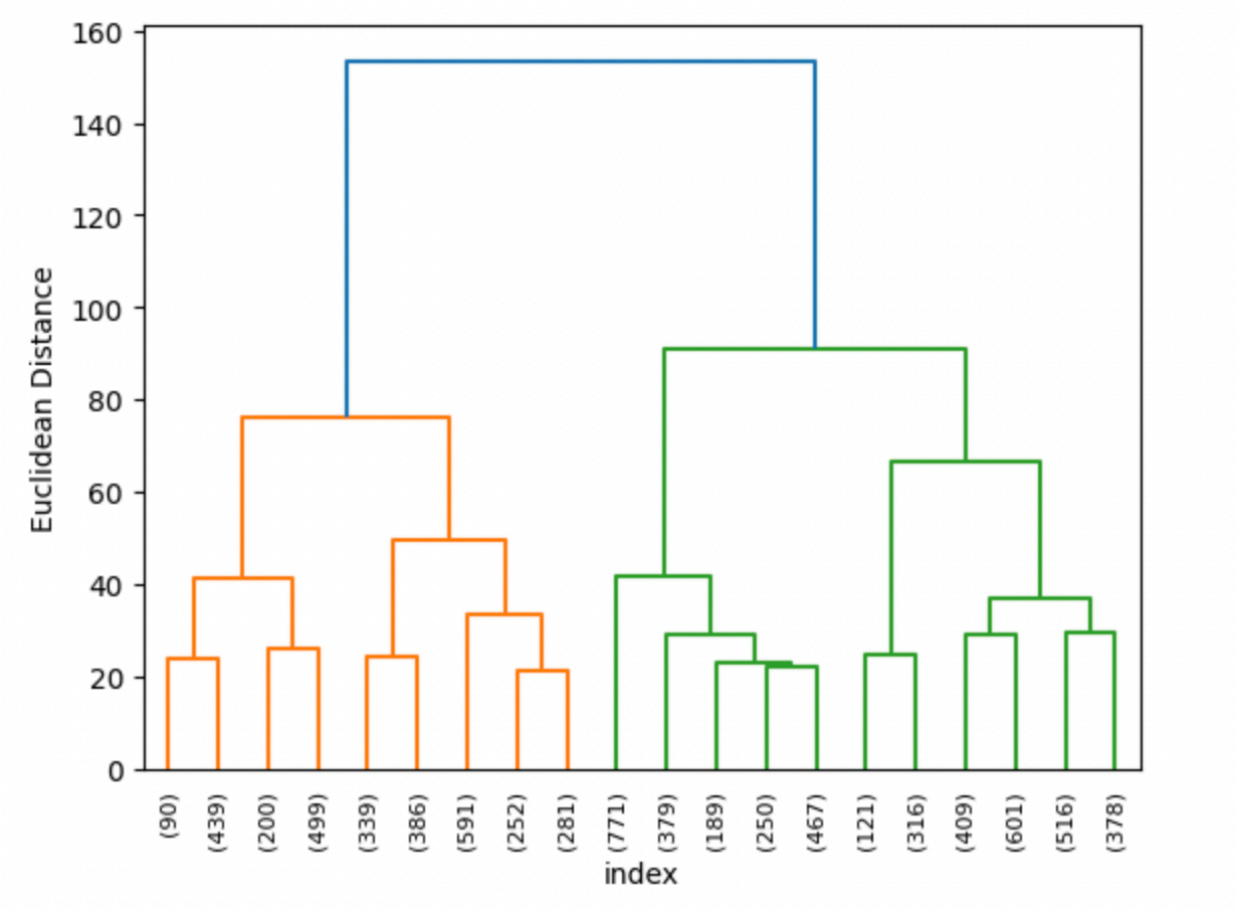


Table 2.2 - NFL Regular Season - Accuracy Metrics associated with Clustering

```
completeness 0.5494005038173192
homogeneity 0.38805756038305167
silhouette 0.1769757010026298
```

In comparison with the playoff dataset, the resulting metrics show clustering was far more effective for this dataset, with all three metrics (homogeneity, completeness and the silhouette score) showing higher values. The silhouette score especially shows much improvement, with the regular season result being 0.3281 points lower than the playoffs score. This highlights the significance of home-

field advantage in the playoffs setting vs. Regular season matchups, where the best performing teams exploit this advantage gained from finishing well in the regular season, and can be classified more appropriately, in comparison to the regular season results.

Figure 2.3 - Dendrogram – Agglomerative Clustering with results limited to last 20 clusters and metrics for NFL Playoffs

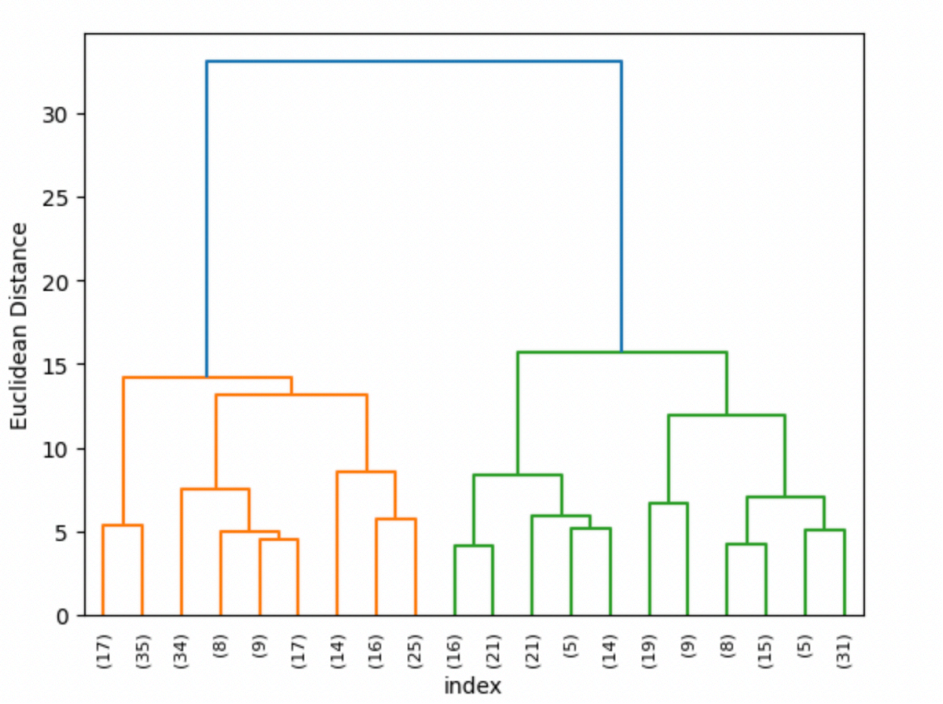


Table 2.3 - NFL Playoffs - Accuracy Metrics associated with Clustering

```
completeness 0.5578774970969246
homogeneity 0.48234518882599847
silhouette 0.5054696825626952
```

Overall, the resulting metrics show that clustering was not an effective way to distinguish home-team advantage as the resulting scores for homogeneity, completeness and the silhouette score are all relatively low for the regular season but showed more promising results for the playoff dataset.

Supervised Learning Models

Decision Trees

The decision tree method should yield more promising results for classification as we have classified our results into three categories already. The decision tree method attempts to classify a categorical-target variable into classes via the independent variables. The tree consists of three parts; nodes, which test the value of certain attributes, branches, which correspond to outcomes of the test connected to the node, and leaf nodes, which predict the outcome[13]. For the decision trees, the dataset received a 30% train-test split, in order to predict the outcome of home-field favored and covering the spread.

The resulting metrics show that accuracy between playoffs and regular season results are relatively similar, with the regular season having an accurate classification rate of 0.566, whereas in the playoffs, the accuracy rate increases slightly to 0.572. One factor to note is in both cases, the metrics for classifying home-favored and covering the spread, the index characterized by value 1, shows a significantly higher F1 value, which is defined as the harmonic mean of the precision and recall scores[14], for classification than the other two values. The model appears to find it easier to classify home-team and covered index value, in comparison to the other values.

[13] Decision Trees - Source:(https://www.springboard.com/blog/data-science/decision-tree-implementation-in-python/)

[14] F-Scores - Source(https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html)

Figure 2.4 – Decision Trees – Limited to Depth 5 - NFL Regular Season

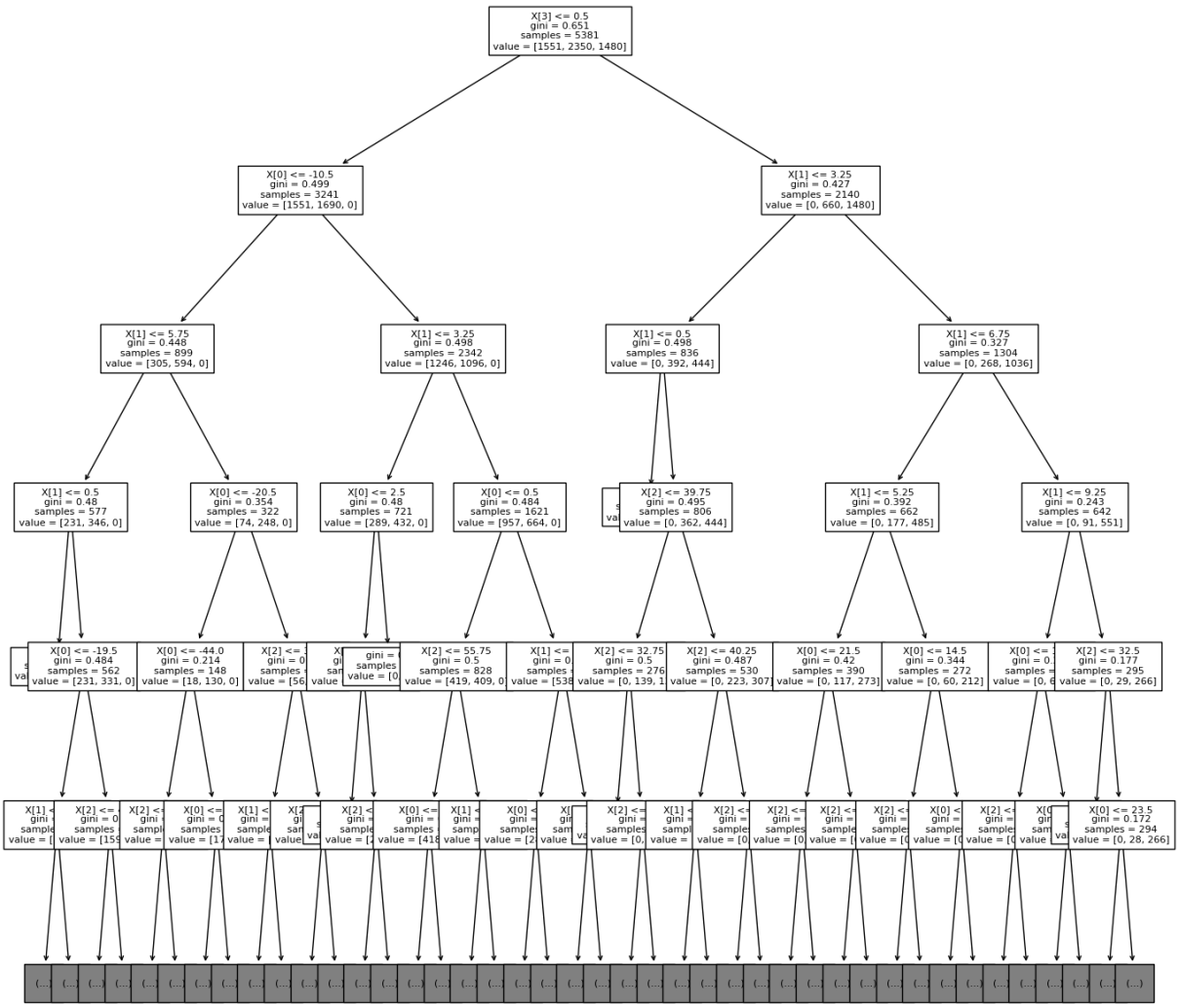


Table 2.4 - Metrics associated with Decision Tree NFL Regular Season

NFL Regular Season				
INDEX	Precision	Recall	F1	
-1	0.53	0.55	0.54	
0	0.52	0.51	0.52	
1	0.72	0.73	0.73	
Accuracy Score	0.58084			

Table 2.5 - Metrics associated with Decision Tree NFL Playoffs

NFL Playoffs				
INDEX	Precision	Recall	F1	
-1	0.56	0.55	0.56	
0	0.52	0.54	0.53	
1	0.73	0.69	0.71	
Accuracy Score	0.5852			

Logistic Regression

Using a logistic regression model, we can attempt to predict the classification of home team being favored and covering the spread using the same independent variables as for the clustering and decision trees. Logistic regression aims to predict the probability that the target variable belongs to a certain class or category. Like the Decision Trees, a 70/30 train-test split was used to feed target and predictive variables for prediction.

Table 2.6 - Confusion Matrix and Metrics associated with NFL Regular Season

```
0.6397919375812744
              precision    recall  f1-score   support

        -1.0       0.62      0.47      0.54       673
         0.0       0.58      0.55      0.57       983
         1.0       0.71      0.95      0.82       651

    accuracy                           0.64      2307
   macro avg       0.64      0.66      0.64      2307
weighted avg       0.63      0.64      0.63      2307

[[317 356   0]
 [195 541 247]
 [  0  33 618]]
```

The results for the regular season yielded a prediction score of 0.6385 for the regular season matchups and 0.656 score for the playoffs matchups. Similarly, to the decision trees, the F1 scores for predicting

the value 1, home team favored and covered, yielded far better scores in comparison to the two other values.

Figure 2.5 - Confusion Matrix for NFL Playoff Matchups

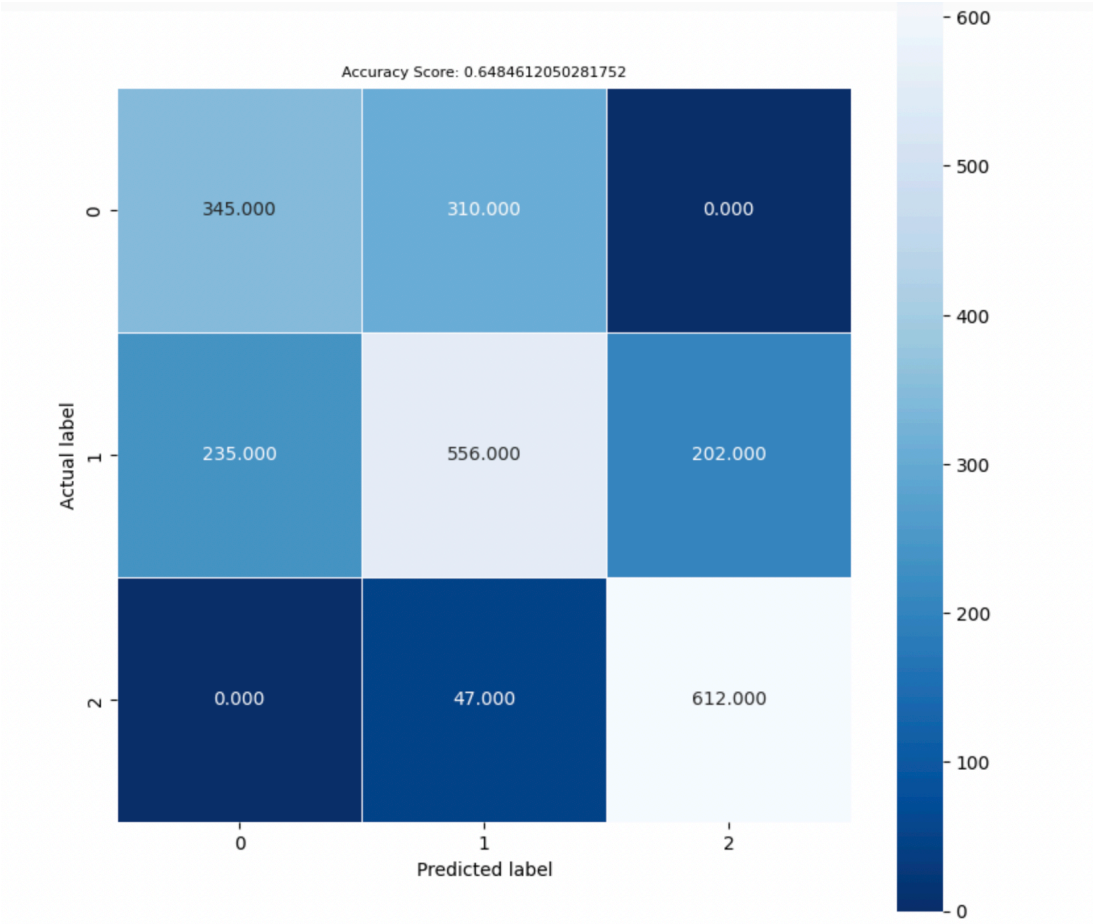


Table 2.7- Accuracy and Metrics associated with NFL Playoffs

```
0.6558300823580407
              precision    recall  f1-score   support

        -1.0       0.59      0.53      0.56       655
         0.0       0.61      0.56      0.58       993
         1.0       0.75      0.93      0.83       659

    accuracy                           0.66      2307
   macro avg       0.65      0.67      0.66      2307
weighted avg       0.65      0.66      0.65      2307

[[345 310   0]
 [235 556 202]
 [  0  47 612]]
```

Discussion and Conclusion

After attempting three different classification methods, one unsupervised learning and two supervised learning, it is evident that classification of the three-valued index used to measure the impact of home-team advantage does not fully capture the weight of home-field advantage throughout the NFL Regular Season and NFL Playoffs.

The supervised learning methods performed better in terms of accurately classifying the variables matchups in which the team with home-team advantage managed to cover the spread. This can be explained by the fact that unsupervised methods take unlabeled data, with the goal of gaining an understanding of the inherent structure of the input data via classification. In comparison, supervised learning methods take inputs in which data is labelled and look to predict the target output via the input variables. Through training the supervised learning models, the models were able to show better results in terms of predictive accuracy.

One main factor that impacted all models was the layout of the initial index. Through creating an index with three results (1, 0, -1), it limited the ability for the models to fully classify the associated variables. Had a binary index been used, the resulting metrics for prediction would probably have been higher, as both of the supervised learning models function best with binary classification. This theory was tested by switching our target indicator to 'covered-spread', the binary indicator for whether the favored team managed to cover spread, and used the same variables, in addition to the index for home-team being favored and covering spread from the NFL regular season dataset. The clustering results were slightly surprising, as all the metrics for classification (homogeneity, completeness and silhouette score) all fell below 0.1, indicating the variables are far too dissimilar for clustering. When attempting to

run the same binary target variables and independent variables using a Decision Tree model, the accuracy rose to 0.9979.

Table 3.1 - Metrics from Hierarchical Clustering – Binary Classification of 'Covered Spread'

```
completeness 0.0003189671746294725
homogeneity 0.00013777471612261443
silhouette 0.013251651623802093
```

Table 3.2 - Classification Report for Decision Tree – Metrics for Binary Classification of 'Covered Spread'

```
              precision    recall  f1-score   support

         0.0       1.00      1.00      1.00      1389
         1.0       1.00      0.99      1.00       918

    accuracy                           1.00      2307
   macro avg       1.00      1.00      1.00      2307
weighted avg       1.00      1.00      1.00      2307

[[1389    0]
 [   5  913]]
Accuracy: 0.9978326831382748
```

Attempting to measure the impact of home-advantage, let alone classify, remains a difficult task. The overall picture showed some surprising conclusions, one being that the absolute value of spread, which should rise depending on how 'favorable' the matchup is for the favored team, did not correlate with the home-favored and covered index. In the case of the regular season, the correlation value was negative, while for the playoffs, it increased ever so slightly to mildly positive[15].

When looking at summary statistics regarding the home-team favorite indicator, absolute spread and then home-team and favorite index on a per team basis[16], both in the regular-season and playoff setting, there appears no direct link between a team receiving home favorite status, signified by the mean value being closer to 1(home-favorite status) vs. The average of the index value. When sorting for teams by mean or count of home-favorite status, we see that that 'status' does not necessarily translate into the home-team managing to 'cover' the spread. One factor that is evident, is that the standard deviation of the generated index shows much larger variation for teams with more counts or higher means of favored status at home, which could signify that they are more likely to cover the spread with this favored status[17].

---

[15] Please refer to Figure 2/2.1 - Table 2 / Table 2.1

[16] Please refer to Table 3.3 and Table 3.4 in Annex

[17] Please refer to Table 3.3 and Table 3.4 in Annex

This leads me to conclude that spread might not necessarily been the best indicator for capturing the weight of home-field advantage. This was also evident during the testing, especially during the clustering of NFL-Regular season matchups, where the silhouette score showed very low results for clustering, but improved largely when using the NFL Playoff matchups, this can be attributed to more occasions where the home-team was favored and supported by the correlation results which indicated a stronger positive relationship for the indicator 'covering-spread' and the target variable.

Another factor that played a role was lack of supporting statistics regarding team-performance. As this dataset solely showed scoring metrics and spreads associated with these matchups, my ability to further capture relevant statistical variables that play a role in determining 'favorite statuses were limited. If additional variables associated with team-performance for these matchups, such as turnovers, the number of occasions where a team loses possession of the football, or advanced metrics about performance of both offences and defences, such as Yards per Play, Rushing Yards per play, Passing Yards per play[18], which would provide me with additional statistics to evaluate team-performance at home vs. Playing away from home.

[18] Defensive and Offensive Statistics NFL - (https://www.footballdb.com/stats/index.html)

Annex

Table 3.3 - Summary Statistics per Team - NFL Regular Season

	home_team_fav			spread_abs			home_fav_and_covered		
	count	mean	std	count	mean	std	count	mean	std
team_home									
Green Bay Packers	248	0.842742	0.364780	248	6.459677	3.496439	240	0.116667	0.911877
Pittsburgh Steelers	248	0.842742	0.364780	248	5.897177	3.564971	242	0.095041	0.912827
New England Patriots	248	0.802419	0.398979	248	7.088710	4.649474	239	0.037657	0.895040
Denver Broncos	248	0.794355	0.404990	248	6.322581	3.913898	237	-0.067511	0.889962
Los Angeles Chargers	32	0.781250	0.420013	32	5.828125	3.925946	31	-0.258065	0.855092
Baltimore Ravens	200	0.775000	0.418630	200	5.900000	3.829927	193	0.072539	0.874954
Dallas Cowboys	248	0.754032	0.431530	248	6.504032	4.087849	243	-0.028807	0.869117
Minnesota Vikings	248	0.750000	0.433888	248	5.429435	3.435357	241	0.000000	0.870823
Philadelphia Eagles	248	0.745968	0.436196	248	5.687500	3.437429	243	0.020576	0.869352
Kansas City Chiefs	248	0.745968	0.436196	248	5.633065	3.253219	242	-0.012397	0.866535
New Orleans Saints	247	0.736842	0.441241	247	5.390688	3.346244	242	-0.049587	0.857974
San Francisco 49ers	248	0.705645	0.456674	248	7.106855	4.247847	239	-0.012552	0.837569
Seattle Seahawks	248	0.689516	0.463627	248	5.840726	3.660689	240	0.008333	0.828325
Indianapolis Colts	248	0.689516	0.463627	248	5.891129	3.691307	239	-0.075314	0.831735
New York Giants	249	0.678715	0.467911	249	5.048193	3.112748	242	-0.020661	0.822143
Buffalo Bills	248	0.677419	0.468409	248	5.254032	3.637688	243	0.049383	0.826742
Atlanta Falcons	248	0.677419	0.468409	248	4.961694	3.030804	241	-0.049793	0.825133
Miami Dolphins	248	0.673387	0.469923	248	5.006048	3.144780	240	-0.079167	0.816913
Tennessee Titans	176	0.642045	0.480766	176	4.826705	2.906116	170	-0.041176	0.802038
Carolina Panthers	208	0.629808	0.484021	208	5.036058	2.872895	203	-0.009852	0.795968
Houston Texans	152	0.598684	0.491785	152	5.243421	3.109474	149	-0.026846	0.770629
Chicago Bears	248	0.596774	0.491537	248	4.518145	2.910033	238	-0.037815	0.775848

Table 3.4 - Summary Statistics per Team - NFL Playoffs

	home_team_fav			spread_abs			home_fav_and_covered		
	count	mean	std	count	mean	std	count	mean	std
team_home									
New England Patriots	33	1.000000	0.000000	33	6.984848	3.951829	32	0.125000	1.008032
Pittsburgh Steelers	23	0.913043	0.288104	23	5.717391	2.839759	23	-0.043478	0.975997
Green Bay Packers	22	0.954545	0.213201	22	7.159091	3.335903	21	0.095238	0.995227
San Francisco 49ers	20	0.900000	0.307794	20	7.275000	4.488216	19	0.157895	0.958190
Philadelphia Eagles	17	0.705882	0.469668	17	4.264706	1.969436	17	-0.235294	0.831370
Denver Broncos	17	0.823529	0.392953	17	6.735294	3.340648	17	-0.117647	0.927520
Dallas Cowboys	16	1.000000	0.000000	16	7.968750	3.635330	16	0.125000	1.024695
Kansas City Chiefs	15	0.866667	0.351866	15	4.933333	2.999206	15	-0.200000	0.941124
Indianapolis Colts	14	0.928571	0.267261	14	5.892857	2.890160	14	0.214286	0.974961
New Orleans Saints	13	0.923077	0.277350	13	6.115385	2.754949	13	-0.461538	0.877058
Seattle Seahawks	13	0.923077	0.277350	13	6.384615	3.434833	13	0.000000	1.000000
Buffalo Bills	12	1.000000	0.000000	12	6.416667	3.146667	12	0.166667	1.029857
Minnesota Vikings	10	0.800000	0.421637	10	6.800000	4.230839	10	0.000000	0.942809
Chicago Bears	9	0.888889	0.333333	9	5.222222	2.728451	9	-0.222222	0.971825
New York Giants	8	0.875000	0.353553	8	4.187500	1.944544	8	0.125000	0.991031
Baltimore Ravens	8	0.875000	0.353553	8	4.937500	3.052370	8	-0.125000	0.991031
San Diego Chargers	8	0.000000	0.000000	8	5.562500	2.808628	8	0.000000	0.000000
Miami Dolphins	8	0.625000	0.517549	8	2.500000	0.597614	8	0.125000	0.834523
Carolina Panthers	7	0.714286	0.487950	7	4.071429	2.935821	7	0.428571	0.786796

Environment

Python 3.7.10

Packages used: SciKit Learn, Pandas, Seaborn, Pyplot, XLRD

References

NFL Logo - Source[online]: (https://1.bp.blogspot.com/-Skddoeh3zoE/VCCqnD7Q2EI/AAAAAAAAEwg/ExbHZkWTwzY/s1600/Logo%2BNFL.png)

Betting Revenue for NFL [online]– Source: (https://www.playusa.com/nfl-legal-sports-betting-revenue/)

NFL Playoff and Qualifications[online] – Source: (https://blog.betway.com/nfl/nfl-explained-how-the-regular-season-and-playoffs-work/)

Recap of Patriots vs. Jets [online]– Source: (https://www.nfl.com/games/jets-at-patriots-2007-reg-15)

Dataset Source [online]– Source: (https://www.kaggle.com/tobycrabtree/nfl-scores-and-betting-data?select=nfl_teams.csv)

Hierarchical Clustering [online]– Source: (https://www.askpython.com/python/examples/hierarchical-clustering)

Decision Trees [online]– Source: (https://www.springboard.com/blog/data-science/decision-tree-implementation-in-python/)

F-Scores [online]– Source: (https://scikit learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html)

NFL Football Statistics [online]– Source: (https://www.footballdb.com/stats/index.html)

Spread Betting on Sports Explained [online] – Source: (https://w88au.net/sports-spread-betting/)