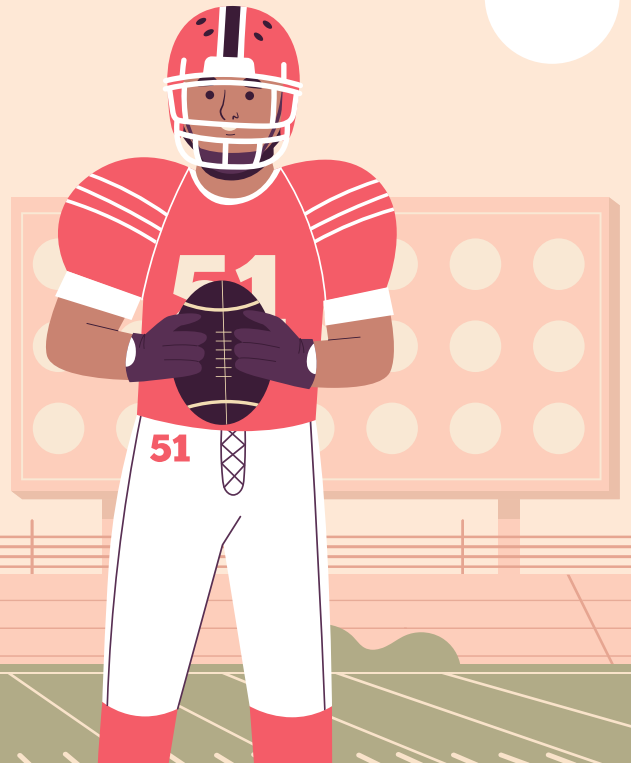
BETTING LINES IN THE NFL

Sahil Gupta, Fraser Kerr, Shlok Shetty



INTRODUCTION

Betting on various sports has been a tradition that has been going on for many centuries at this point. Usually, people would visit their local casinos, where they could place bets on certain games, and have the opportunity to make money. However, with the rise of mobile technology and the internet, many bookmakers have went online, exposing the average person to the world of sports betting. Nowadays, a person can place a bet in a snap of a finger.

Our goal is to analyze the behavior of bookmakers to see how accurate the lines they make are, and come up with an answer to how these lines are placed based on the outcomes of the corresponding NFL games. We data analysis is mainly targeted towards NFL enthusiasts and people who are interested in sports betting in general. From our analysis they will be able to make more informed decisions the next time they decide to place bets on NFL games.

The questions we will be answering:

- How often to favorites win based on the spread?
- How accurate are the over/under totals compared with the total score of the game?
- Effect of indoor/outdoor stadiums on over/under totals?
- Over/under lines through the decades. Have odds makers matched evolving offenses?
- Spreads favoring home team or away team?
- Underdogs win more at home/away?
- How often do the lines set by bookmakers hit? How have they changed over time, with the advent of technology?

THE DATASETS

KAGGLE, BETTING DATA

This dataset contains data on the betting lines set for NFL games starting from the 1960s. The data also contains relevant weather information. There are no ethical concerns using the data, and the dataset did not have any quality issues.

team_favorite_id <chr>	spread_favorite <dbl>	over_under_line <chr>
PHI	-7.0	31.5
PIT	-2.0	37
PICK	0.0	36
PICK	0.0	35.5
DEN	-3.0	32
LAR	-13.0	39
CLE	-7.0	42
PIT	-6.0	40
PIT	-4.0	36
DET	-3.0	39

PRO-FOOTBALL REFERENCE

This dataset reflects NFL game data of NFL games with the most common scores. This dataset contains information, about offensive yardage, turnovers, and winners of the game. There are also no quality issues present, due to the reliability of PFR, and there are no ethical concerns using this dataset.

Rk	Week	Day	Date	Winner/tie		Loser/tie		PtsW	PtsL	YdsW	TOW	YdsL	TOL
1	5	Sun	1944-10-15	W Cleveland Rams	@	Detroit Lions	boxscore	20	17	177	3	236	6
2	8	Sun	1947-11-09	W Chicago Bears		Green Bay Packers	boxscore	20	17	320	9	291	4
3	15	Sun	1947-12-07	W New York Yankees	@	Brooklyn Dodgers	boxscore	20	17	357	3	175	1
4	7	Sun	1951-11-11	W Cleveland Browns		Philadelphia Eagles	boxscore	20	17	202	3	310	4
5	11	Sun	1951-12-09	W Washington Redskins	@	Chicago Cardinals	boxscore	20	17	326	2	318	2
6	6	Sun	1952-11-02	W Chicago Bears	@	San Francisco 49ers	boxscore	20	17	206	1	317	2
7	10	Sun	1954-11-28	W Pittsburgh Steelers		Chicago Cardinals	boxscore	20	17	339	2	437	7
8	1	Sun	1955-09-25	W Green Bay Packers		Detroit Lions	boxscore	20	17	294	4	311	2
9	4	Sun	1956-10-21	W Detroit Lions		San Francisco 49ers	boxscore	20	17	393	1	257	2
10	9	Sun	1956-11-25	W Washington Redskins	@	Cleveland Browns	boxscore	20	17	237	2	184	2
11	10	Sun	1956-12-02	W San Francisco 49ers	@	Baltimore Colts	boxscore	20	17	360	2	302	4
12	5	Sun	1959-10-25	W San Francisco 49ers		Chicago Bears	boxscore	20	17	332	1	320	4
13	2	Sun	1961-09-24	W Cleveland Browns		St. Louis Cardinals	boxscore	20	17	245	4	185	3
14	11	Fri	1961-11-17	W Boston Patriots		Oakland Raiders	boxscore	20	17	225	2	203	2
15	12	Sun	1961-12-03	W Baltimore Colts		San Francisco 49ers	boxscore	20	17	346	3	309	3
16	12	Sun	1961-12-03	W Green Bay Packers		New York Giants	boxscore	20	17	404	3	262	3
17	5	Sun	1962-10-14	W Pittsburgh Steelers	@	New York Giants	boxscore	20	17	322	1	291	3
18	7	Sun	1962-10-21	W Dallas Texans		New York Titans	boxscore	20	17	342	2	276	3
19	14	Sun	1962-12-16	W Green Bay Packers	@	Los Angeles Rams	boxscore	20	17	427	3	303	1
20	Champ	Sun	1962-12-23	W Dallas Texans	@	Houston Oilers	boxscore	20	17	237	1	359	5
21	7	Sun	1964-10-25	W San Diego Chargers	@	Houston Oilers	boxscore	20	17	345	2	398	2
22	2	Sun	1965-09-26	W Green Bay Packers		Baltimore Colts	boxscore	20	17	184	3	309	6
23	14	Sat	1965-12-18	W Baltimore Colts	@	Los Angeles Rams	boxscore	20	17	264	0	267	2
24	7	Sun	1966-10-23	W Baltimore Colts		Minnesota Vikings	boxscore	20	17	347	2	317	1
25	9	Sun	1966-11-06	W Minnesota Vikings	@	Green Bay Packers	boxscore	20	17	290	0	292	1
26	9	Sun	1966-11-06	W St. Louis Cardinals	@	New York Giants	boxscore	20	17	264	2	210	2

DATA PROCESSING

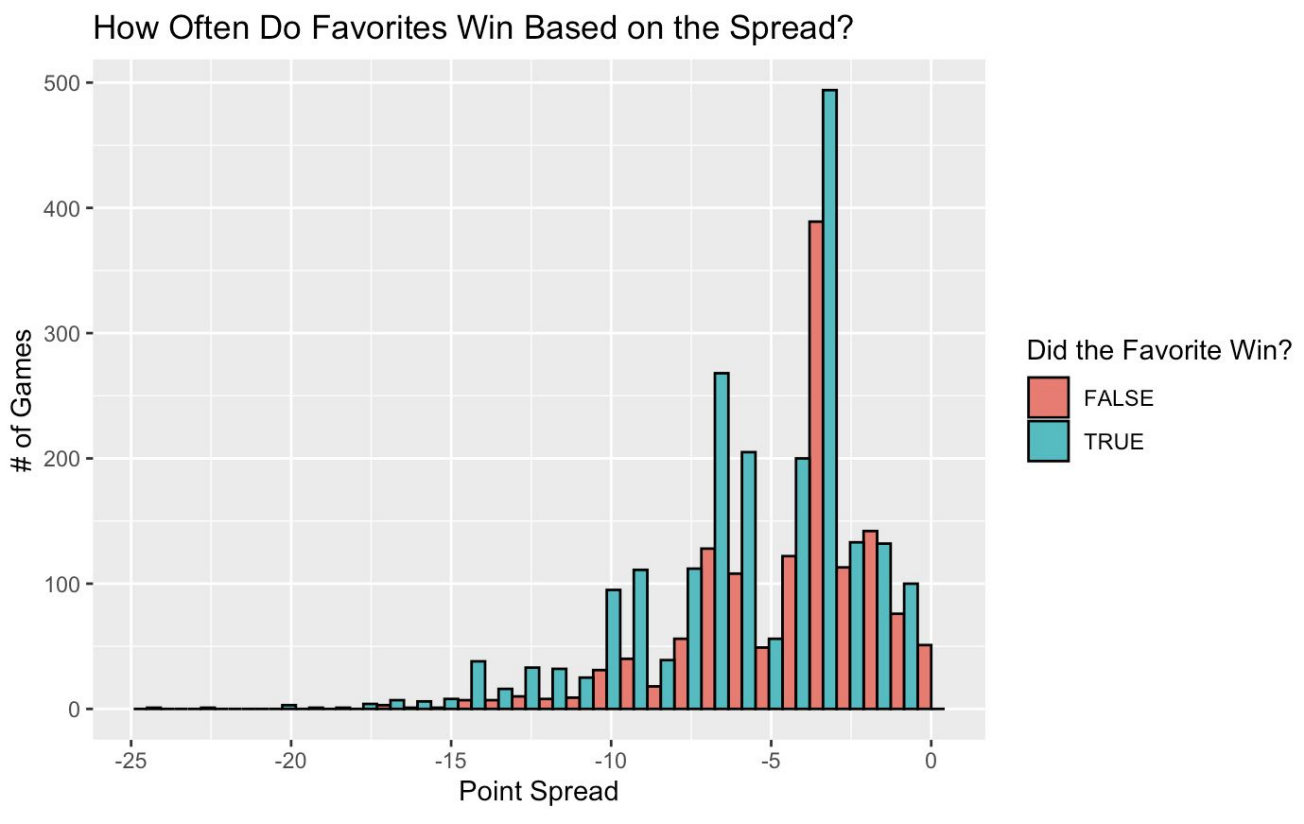
Before we could use the datasets, we had to process the datasets, so we could merge them together into one big dataset. In order to do this, we first dropped extraneous columns from both datasets, such as columns that were present in both datasets, or columns that were not needed for our purposes. Then we had to create a merge key, for both data sets. The issue we encountered was that the datasets did not have a distinct common key we could use, so we had to create one. We did this by taking the date of each NFL game, and then adding on the id(2-3 letter abbreviation of each team, “Seattle Seahawks” is “SEA”) of the home team, so we would end up with something like “202311300sea” as a game_id. Then we merged the datasets, using the game_id as a key, to get something that looked like this:

game_id <chr>	week <chr>	day <chr>	date <date>	winner <chr>	loser <chr>	PtsW <dbl>	PtsL <dbl>	YdsW <dbl>	TOW <dbl>
197909160ari	3	Sun	1979-09-16	PIT	ARI	24	21	398	4
197909160dal	3	Sun	1979-09-16	DAL	CHI	24	20	420	3
197909160sea	3	Sun	1979-09-16	SEA	LVR	27	10	419	0
197909160cin	3	Sun	1979-09-16	NE	CIN	20	14	332	0
197909230nyg	4	Sun	1979-09-23	PHI	NYG	17	13	326	3
197909230pit	4	Sun	1979-09-23	PIT	IND	17	13	392	4
197909230ari	4	Sun	1979-09-23	WAS	ARI	17	7	191	1
197909230ne	4	Sun	1979-09-23	NE	LAC	27	21	291	2
197909230min	4	Sun	1979-09-23	MIN	GB	27	21	376	2
197909230cin	4	Sun	1979-09-23	TEN	CIN	30	27	361	3

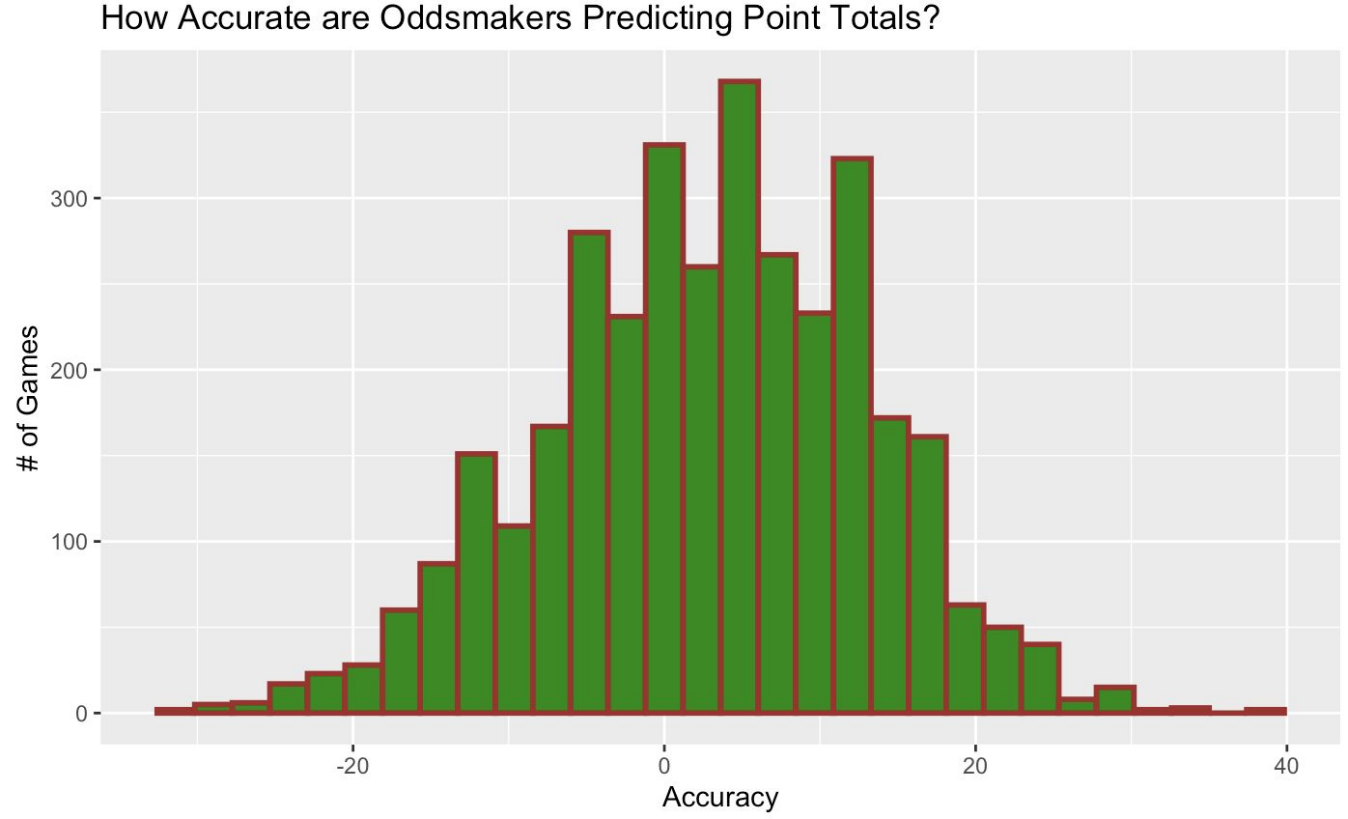
winner <chr>	loser <chr>	PtsW <dbl>	PtsL <dbl>	YdsW <dbl>	TOW <dbl>	YdsL <dbl>	TOL <dbl>	atHome <lgl>	TODiff <dbl>
PIT	ARI	24	21	398	4	212	1	FALSE	-3
DAL	CHI	24	20	420	3	318	2	TRUE	-1
SEA	LVR	27	10	419	0	431	4	TRUE	4
NE	CIN	20	14	332	0	198	1	FALSE	1
PHI	NYG	17	13	326	3	229	2	FALSE	-1
PIT	IND	17	13	392	4	285	2	TRUE	-2
WAS	ARI	17	7	191	1	376	5	FALSE	4
NE	LAC	27	21	291	2	232	2	TRUE	0
MIN	GB	27	21	376	2	249	1	TRUE	-1
TEN	CIN	30	27	361	3	265	2	FALSE	-1

schedule_playoff <lgl>	team_favorite_id <chr>	spread_favorite <dbl>	over_under_line <dbl>	stadium <chr>	stadium_neutral <lgl>	weather_temperature <dbl>	weather_wind_mph <dbl>	weather_humidity <dbl>	weather_detail <chr>
FALSE	PIT	-6.0	40.0	Busch Memorial Stadium	FALSE	63	5	56	NA
FALSE	DAL	-9.0	37.0	Texas Stadium	FALSE	68	8	41	NA
FALSE	SEA	-3.0	44.0	Seattle Kingdome	FALSE	72	0	NA	indoor
FALSE	NE	-6.0	41.0	Cinergy Field	FALSE	58	7	59	NA
FALSE	PHI	-5.0	37.0	Giants Stadium	FALSE	61	16	58	NA
FALSE	PIT	-14.0	38.0	Three Rivers Stadium	FALSE	54	10	65	NA
FALSE	ARI	-3.0	39.0	Busch Memorial Stadium	FALSE	61	8	53	NA
FALSE	NE	-3.0	44.0	Foxboro Stadium	FALSE	55	14	72	NA
FALSE	MIN	-1.0	34.0	Metropolitan Stadium	FALSE	59	14	53	NA
FALSE	TEN	-3.0	37.0	Cinergy Field	FALSE	56	11	64	NA

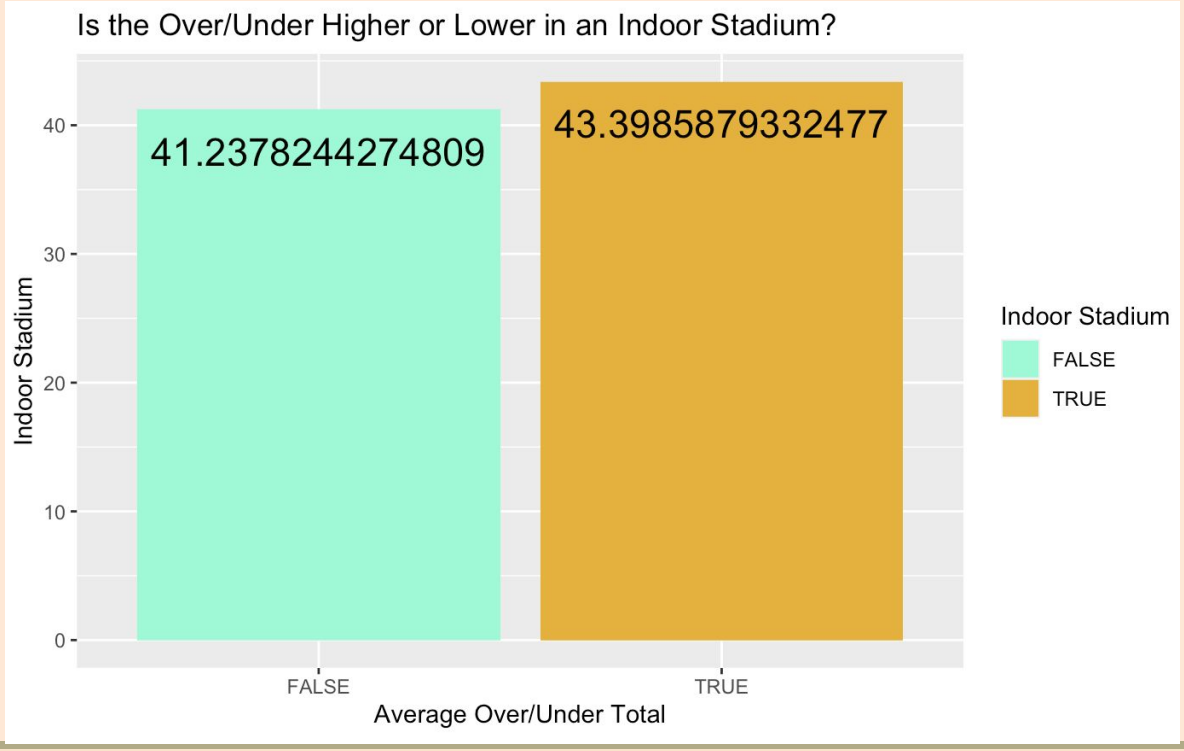
DO TEAMS WITH BIGGER SPREADS WIN MORE OFTEN?



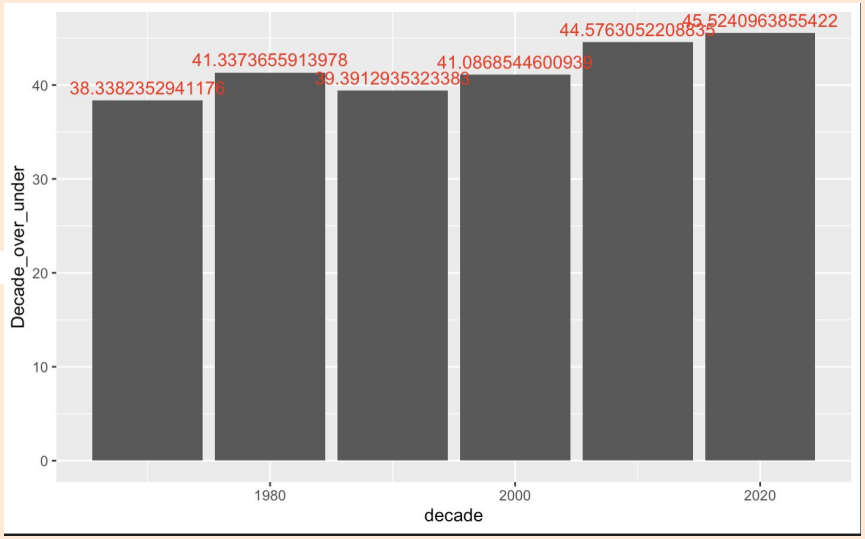
ARE THE OVER/UNDER TOTALS ACCURATE?

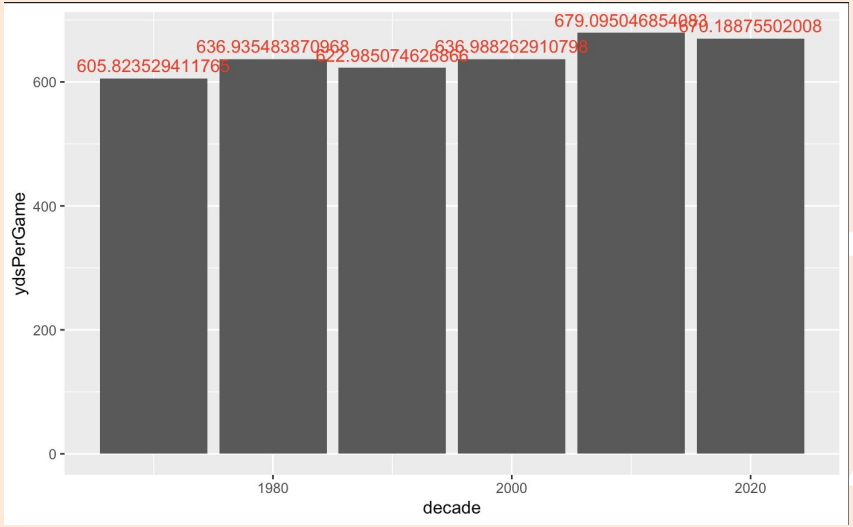


DO ODDSMAKERS THINK INDOOR STADIUMS LEADS TO MORE POINTS?

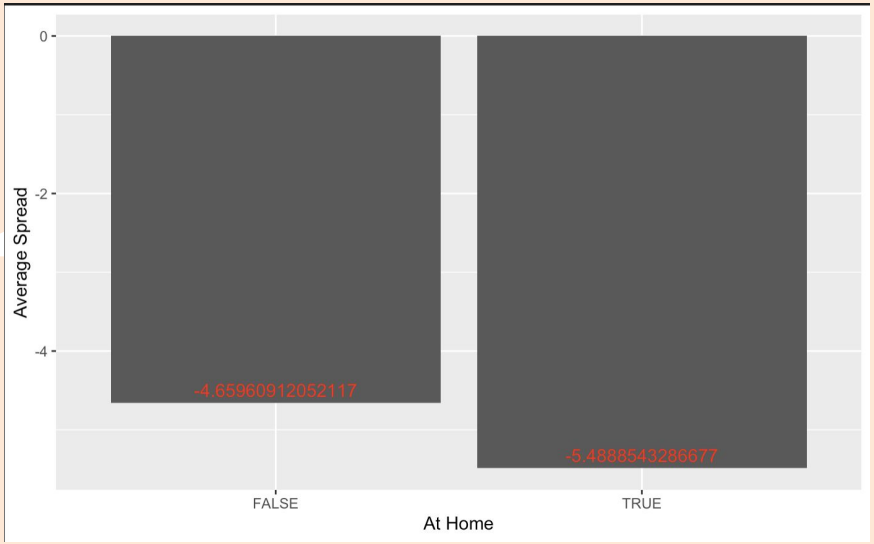


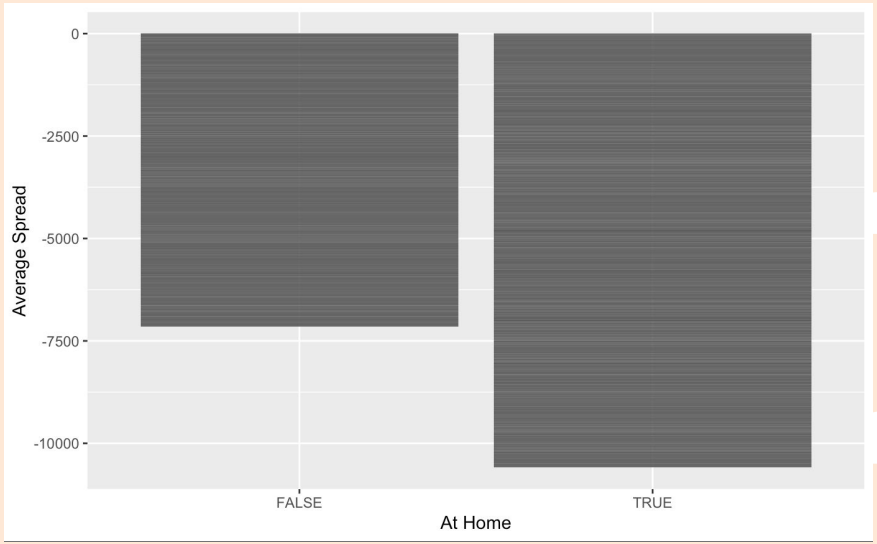
HOW HAS OVER/UNDERS SWITCHED THROUGH DECADES



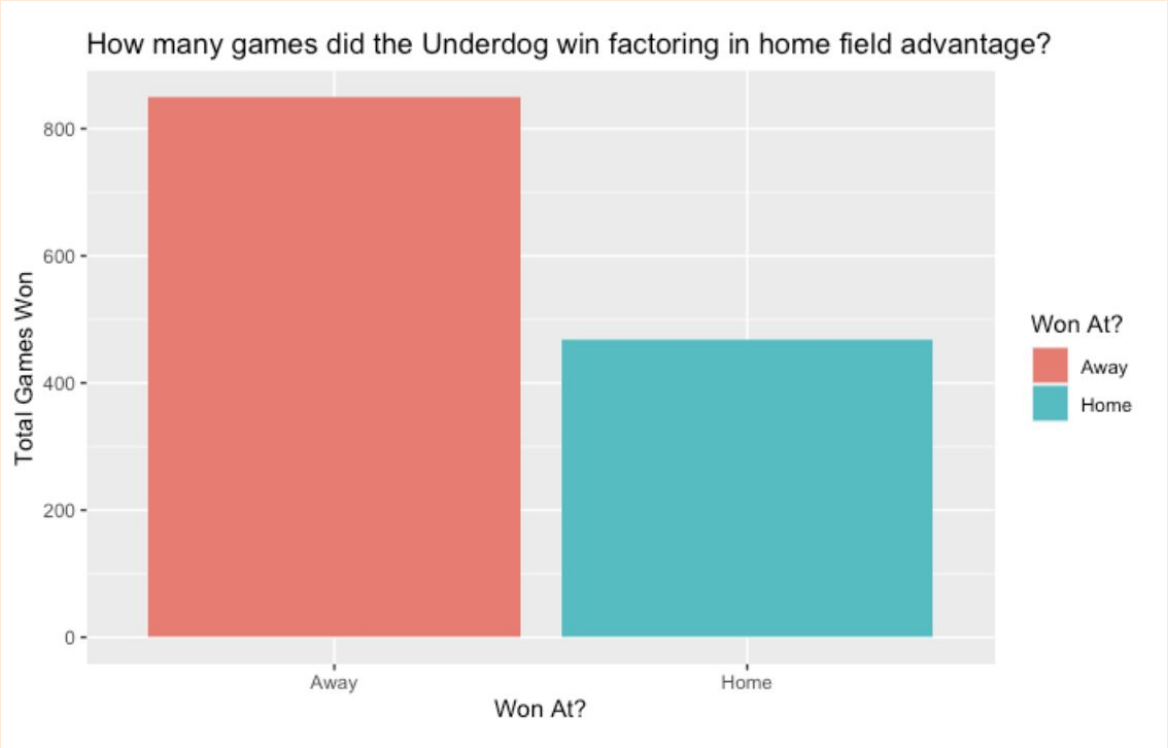


SPREADS AT HOME

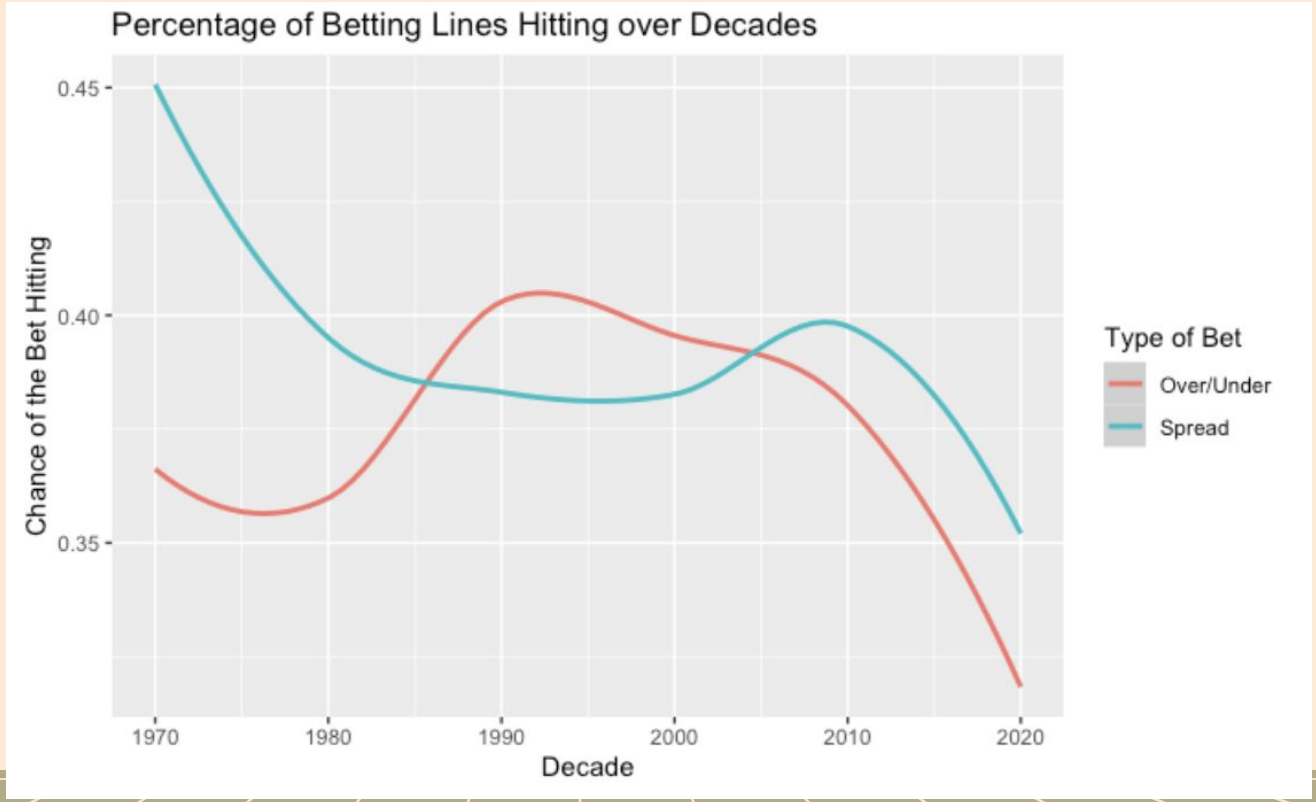




HOME FIELD ADVANTAGE AND UNDERDOGS



BETTING ACCURACY OVER TIME



MAIN TAKEAWAYS

- While Spreads placed on NFL games are quite close the to actual final score of the games, they rarely hit.
- Over/Under lines are usually lean towards the over, so it's better to lean towards the over rather than the under.
- Indoor stadiums have a higher over/under total placed on them, as indoor stadiums are not affected by weather conditions.
- Bookmakers seem to match the evolution of offense over decades which is reflected in their over/under marks.
- Spreads at home usually favor the home team, and is significantly larger for home teams in general.
- Contrary to popular belief, more upsets do not occur at home, and more upsets occur away from home.
- Betting lines set by bookmakers don't hit very often, so it is always a good idea to take the over, and the side of the spread the favors the underdog.

THANK YOU