

Fantasy Football Draft Assistant

The General Managers - Group 7

Data Science Capstone Project
Exploratory Data Analytics Report

Date:

03/09/2025

[GitHub Repository](#)

(two branches: EDA and web-scraping)

Team Members:

Name: Caleb Miller

Name: Hashim Afzal

Name: Thomas Kiefer

Name: David Blankenship

Analyzing the basic metrics of variables

The data for this project includes player statistics from the 2022 - 2024 NFL seasons, and can be expanded to include subsequent seasons by editing our web scraping script. This section presents an analysis of the dataset variables, their types, and key descriptive statistics. It covers various aspects of the game such as passing, rushing, receiving, defensive, and special teams statistics. Additionally, it includes calculated fields such as fantasy points and player average draft positions (ADP).

The features in the dataset can be categorized into categorical and numerical types. Categorical variables can be expanded into nominal and ordinal types, with numerical variables further classified as discrete or continuous. Below is a summary of the data types:

Categorical:

Nominal: Player Name, Position, Team

This dataset contains 970 non-null player names, and team information which can correspond with one of the 32 NFL teams. In certain instances, this team field may be null if a player did not play in a given season, or hold the value “2TM” or “3TM” if a player was traded or played for multiple teams in a given season. This dataset also contains player positions for the common positions associated with fantasy football. These being: QB, RB, WR, TE, DEF and K. This column can also be null if a player did not play in a given season.

Ordinal: ESPN ADP, Sleeper ADP, NFL ADP, RTSports ADP, Average ADP, Positional ADP

Ordinal positions include the average draft position for a given player from one of four different sources. These were then averaged into an average ADP, and a positional ADP which was also calculated to determine how players were drafted in their own positional subgroup.

	2022	ESPN ADP	2022	Sleeper ADP	2022	NFL ADP	2022		2023	ESPN ADP	2023	Sleeper ADP	2023	NFL ADP	2023	
count	225.000000		243.000000		291.000000		277.000000		count	402.000000		267.000000		289.000000		271.000000
mean	113.342222		122.440329		147.347679		140.252708		mean	221.472637		135.528909		146.892734		136.594046
std	65.551801		70.924820		85.952552		81.316529		std	140.942200		79.498062		85.795414		79.002239
min	1.000000		1.000000		1.000000		1.000000		min	1.000000		1.000000		1.000000		1.000000
25%	57.000000		61.500000		73.500000		70.000000		25%	101.250000		67.500000		73.000000		68.500000
50%	113.000000		122.000000		146.000000		140.000000		50%	202.500000		134.000000		146.000000		136.000000
75%	170.000000		182.500000		220.500000		211.000000		75%	345.500000		281.500000		219.000000		204.500000
max	227.000000		246.000000		300.000000		280.000000		max	499.000000		302.000000		300.000000		273.000000
2022 Average ADP		2022 Positional ADP		2023 Average ADP		2023 Positional ADP		2024 Average ADP		2024 Positional ADP		2025 Average ADP		2025 Positional ADP		
count	686.000000		686.000000		686.000000		686.000000		count	653.000000		653.000000		653.000000		653.000000
mean	232.564140		68.088921		232.564140		68.088921		mean	301.764472		78.336987		78.336987		78.336987
std	92.798386		35.985268		92.798386		35.985268		std	165.399939		57.413711		57.413711		57.413711
min	1.000000		1.000000		1.000000		1.000000		min	1.000000		1.000000		1.000000		1.000000
25%	174.050000		29.000000		174.050000		29.000000		25%	168.300000		26.000000		26.000000		26.000000
50%	301.000000		97.000000		301.000000		97.000000		50%	284.000000		62.000000		62.000000		62.000000
75%	301.000000		97.000000		301.000000		97.000000		75%	500.000000		154.000000		154.000000		154.000000
max	301.000000		97.000000		301.000000		97.000000		max	580.000000		154.000000		154.000000		154.000000
2024 Average ADP		2024 Positional ADP		2024 NFL ADP		2024 RTSports ADP		2024 Average ADP		2024 Positional ADP		2024 NFL ADP		2024 RTSports ADP		
count	383.000000		290.000000		601.000000		268.000000		count	345.792013		135.518657		345.792013		135.518657
mean	214.187990		149.665517		214.187990		149.665517		mean	89.449693		232.401632		89.449693		78.597655
std	139.002679		89.449693		139.002679		89.449693		std	151.000000		151.000000		151.000000		67.750000
min	1.000000		1.000000		1.000000		1.000000		min	1.000000		1.000000		1.000000		1.000000
25%	96.500000		73.250000		96.500000		73.250000		25%	195.000000		146.500000		195.000000		135.500000
50%	323.000000		223.750000		323.000000		223.750000		50%	509.000000		509.000000		509.000000		282.250000
75%	499.000000		326.000000		499.000000		326.000000		75%	899.000000		899.000000		899.000000		272.000000
max	663.000000		663.000000		663.000000		663.000000		max	79.601810		79.601810		79.601810		79.601810
2024 Average ADP		2024 Positional ADP		2024 NFL ADP		2024 RTSports ADP		2024 Average ADP		2024 Positional ADP		2024 NFL ADP		2024 RTSports ADP		
count	374.246456		79.601810		260.774983		64.238731		count	316.000000		63.000000		316.000000		63.000000
mean	169.800000		27.000000		169.800000		27.000000		mean	529.250000		115.000000		529.250000		115.000000
std	316.000000		63.000000		316.000000		63.000000		std	900.000000		211.000000		900.000000		211.000000
min	529.250000		115.000000		529.250000		115.000000		min	900.000000		211.000000		900.000000		211.000000
25%	529.250000		115.000000		529.250000		115.000000		25%	900.000000		211.000000		900.000000		211.000000
50%	529.250000		115.000000		529.250000		115.000000		50%	900.000000		211.000000		900.000000		211.000000
75%	529.250000		115.000000		529.250000		115.000000		75%	900.000000		211.000000		900.000000		211.000000
max	529.250000		115.000000		529.250000		115.000000		max	900.000000		211.000000		900.000000		211.000000

Numerical:

The numerical features of our dataset can be broken down into either discrete or continuous. The discrete variables mainly encapsulated counting statistics like Games Played, Passing Attempts and Passing Touchdowns. Continuous features were mainly “total” statistics like Passing Yards, Receiving Yards and Fantasy Points. A full breakdown of our numerical features with descriptive statistics can be found below.

Discrete: Games Played, Games Started, Passing Attempts, Passing Completions, Passing Touchdowns, Interceptions Thrown, Rushing Attempts, Rushing Touchdowns, Targets, Receptions, Receiving Touchdowns, Fumbles, Field Goals Attempted 0-39, Field Goals Made 0-39, Field Goals Attempted 40-49, Field Goals Made 40-49, Field Goals Made 50+, Field Goals Made 50+, Field Goals Attempted, Field Goals Made, Extra Points Attempted, Extra Points Made, Total Plays, Takeaways, First Downs Allowed, Passing Touchdowns Allowed, Rushing Touchdowns Allowed, Penalties Committed, First Downs by Penalty, ST_Sacks, ST_Interceptions, ST_Fumble Recoveries, ST_Forced Fumbles, ST_Safeties, ST_Special Teams Touchdowns, XP2

	2022 Games Played	2022 Games Started	2023 Games Played	2023 Games Started
count	686.000000	686.000000	653.000000	653.000000
mean	11.959184	5.476676	12.643185	5.799387
std	5.280574	6.136540	4.979384	6.260785
min	1.000000	0.000000	1.000000	0.000000
25%	8.000000	0.000000	9.000000	0.000000
50%	14.000000	2.000000	15.000000	3.000000
75%	17.000000	11.000000	17.000000	11.000000
max	17.000000	17.000000	17.000000	17.000000

	2024 Games Played	2024 Games Started
count	663.000000	663.000000
mean	12.220211	5.675716
std	5.222518	6.172896
min	1.000000	0.000000
25%	8.000000	0.000000
50%	14.000000	3.000000
75%	17.000000	11.000000
max	18.000000	17.000000

Targets represents the amount of times a player is thrown to by his quarterback, and XP2 is the amount of two point conversions that were scored by a player.

	2022 Targets	2022 Receptions	2022 Receiving Touchdowns	2022 XP2	2023 Targets	2023 Receptions	2023 Receiving Touchdowns	2023 XP2
count	586.000000	586.000000	586.000000	24.000000	561.000000	561.000000	561.000000	23.000000
mean	29.510239	19.790102	1.278157	1.166667	31.135472	21.023173	1.342246	1.391304
std	36.892323	24.834604	2.107236	0.380693	38.181982	25.937173	2.212904	0.656376
min	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	1.000000
25%	2.000000	2.000000	0.000000	1.000000	2.000000	1.000000	0.000000	1.000000
50%	13.500000	9.000000	0.000000	1.000000	17.000000	11.000000	0.000000	1.000000
75%	43.750000	30.000000	2.000000	1.000000	44.000000	31.000000	2.000000	2.000000
max	184.000000	128.000000	14.000000	2.000000	181.000000	135.000000	13.000000	3.000000

	2024 Targets	2024 Receptions	2024 Receiving Touchdowns	2024 XP2
count	566.000000	566.000000	566.000000	22.000000
mean	30.022968	20.517668	1.420495	1.409091
std	36.985113	25.335600	2.399086	0.590326
min	0.000000	0.000000	0.000000	1.000000
25%	2.000000	1.000000	0.000000	1.000000
50%	13.500000	9.000000	0.000000	1.000000
75%	49.000000	33.000000	2.000000	2.000000
max	175.000000	127.000000	17.000000	3.000000

	2022 Passing Attempts	2022 Passing Completions \
count	113.000000	113.000000
mean	159.902655	102.699115
std	204.670422	134.771161
min	0.000000	0.000000
25%	2.000000	1.000000
50%	61.000000	38.000000
75%	303.000000	192.000000
max	733.000000	490.000000

	2023 Passing Attempts	2023 Passing Completions \
count	114.000000	114.000000
mean	160.649123	103.570175
std	203.324188	133.354804
min	0.000000	0.000000
25%	1.000000	1.000000
50%	46.500000	26.500000
75%	300.250000	194.750000
max	612.000000	410.000000

	2022 Passing Touchdowns	2022 Interceptions Thrown
count	113.000000	113.000000
mean	6.637168	3.699115
std	9.724841	4.292612
min	0.000000	0.000000
25%	0.000000	0.000000
50%	1.000000	2.000000
75%	12.000000	6.000000
max	41.000000	15.000000

	2023 Passing Touchdowns	2023 Interceptions Thrown
count	114.000000	114.000000
mean	6.614035	3.771930
std	9.552379	4.713068
min	0.000000	0.000000
25%	0.000000	0.000000
50%	2.000000	1.000000
75%	10.000000	7.000000
max	36.000000	21.000000

	2024 Passing Attempts	2024 Passing Completions \
count	107.000000	107.000000
mean	166.448598	108.682243
std	203.405777	136.053305
min	0.000000	0.000000
25%	1.000000	1.000000
50%	42.000000	27.000000
75%	303.500000	189.500000
max	652.000000	460.000000

	2024 Passing Touchdowns	2024 Interceptions Thrown
count	107.000000	107.000000
mean	7.560748	3.616822
std	10.864671	4.630997
min	0.000000	0.000000
25%	0.000000	0.000000
50%	2.000000	1.000000
75%	12.500000	7.000000
max	43.000000	16.000000

	2022 Rushing Attempts	2022 Rushing Touchdowns	2022 Fumbles
count	586.000000	586.000000	586.000000
mean	25.179181	0.831058	1.039249
std	56.224188	2.118811	1.969033
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	2.000000	0.000000	0.000000
75%	17.000000	1.000000	1.000000
max	349.000000	17.000000	16.000000

	2023 Rushing Attempts	2023 Rushing Touchdowns	2023 Fumbles
count	561.000000	561.000000	561.000000
mean	26.012478	0.837790	1.062389
std	56.277950	2.253663	2.004823
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	2.000000	0.000000	0.000000
75%	17.000000	0.000000	1.000000
max	280.000000	18.000000	14.000000

	2024 Rushing Attempts	2024 Rushing Touchdowns	2024 Fumbles
count	566.000000	566.000000	566.000000
mean	25.934629	0.902827	1.024735
std	58.154396	2.457301	1.957809
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	1.000000	0.000000	0.000000
75%	18.000000	0.000000	1.000000
max	345.000000	16.000000	13.000000

	2022 Field Goals Attempted	2022 Field Goals Made \
count	64.000000	64.000000
mean	16.593750	14.109375
std	15.634434	13.397604
min	0.000000	0.000000
25%	0.000000	0.000000
50%	19.000000	16.000000
75%	32.000000	27.000000
max	43.000000	37.000000

	2023 Field Goals Attempted	2023 Field Goals Made \
count	58.000000	58.000000
mean	18.275862	15.706897
std	15.803390	13.930918
min	0.000000	0.000000
25%	0.000000	0.000000
50%	23.500000	16.500000
75%	32.750000	29.000000
max	42.000000	36.000000

	2023 Extra Points Attempted	2023 Extra Points Made
count	58.000000	58.000000
mean	20.155172	19.327586
std	18.796978	18.276039
min	0.000000	0.000000
25%	0.000000	0.000000
50%	21.000000	18.000000
75%	35.000000	33.750000
max	61.000000	60.000000

	2024 Field Goals Attempted	2024 Field Goals Made \
count	61.000000	61.000000
mean	18.278689	15.360656
std	15.368942	13.267797
min	0.000000	0.000000
25%	0.000000	0.000000
50%	20.000000	16.000000
75%	31.000000	26.000000
max	47.000000	41.000000
	2024 Extra Points Attempted	2024 Extra Points Made
count	61.000000	61.000000
mean	20.344262	19.491803
std	18.424517	17.691451
min	0.000000	0.000000
25%	0.000000	0.000000
50%	19.000000	18.000000
75%	33.000000	32.000000
max	67.000000	64.000000

Total Plays represented the amount of plays a defensive team was on the field for during the entire season. First Downs Allowed represented the total number of first downs a defense allowed during the entire season.

2022 Total Plays		2022 First Downs Allowed \	
count	32.000000	32.000000	32.000000
mean	1066.750000	333.312500	327.437500
std	39.615083	23.669005	25.637595
min	991.000000	286.000000	254.000000
25%	1042.500000	317.000000	311.000000
50%	1076.500000	330.000000	328.500000
75%	1093.000000	350.250000	344.500000
max	1134.000000	377.000000	380.000000
2022 Passing Touchdowns Allowed		2022 Rushing Touchdowns Allowed	
count	32.000000	32.000000	32.000000
mean	23.437500	15.218750	14.687500
std	4.287548	5.528427	4.617201
min	15.000000	7.000000	6.000000
25%	20.750000	11.000000	11.000000
50%	23.000000	14.500000	14.500000
75%	26.000000	18.500000	15.500000
max	33.000000	31.000000	25.000000
2023 Total Plays		2023 First Downs Allowed \	
count	32.000000	32.000000	32.000000
mean	1072.84375	327.437500	327.437500
std	35.44110	25.637595	25.637595
min	1084.00000	254.000000	254.000000
25%	1051.75000	311.000000	311.000000
50%	1078.50000	328.500000	328.500000
75%	1092.00000	344.500000	344.500000
max	1147.00000	380.000000	380.000000
2023 Passing Touchdowns Allowed		2023 Rushing Touchdowns Allowed	
count	32.000000	32.000000	32.000000
mean	23.56250	14.687500	14.687500
std	5.18022	4.617201	4.617201
min	17.00000	6.000000	6.000000
25%	20.00000	11.000000	11.000000
50%	22.50000	14.500000	14.500000
75%	26.25000	15.500000	15.500000
max	39.00000	25.000000	25.000000
2024 Total Plays		2024 First Downs Allowed \	
count	32.000000	32.000000	32.000000
mean	1056.625000	335.312500	335.312500
std	35.460792	23.750637	23.750637
min	999.000000	293.000000	293.000000
25%	1032.750000	323.750000	323.750000
50%	1051.500000	330.500000	330.500000
75%	1068.000000	344.750000	344.750000
max	1154.000000	417.000000	417.000000
2024 Passing Touchdowns Allowed		2024 Rushing Touchdowns Allowed	
count	32.000000	32.000000	32.000000
mean	25.281250	15.968750	15.968750
std	4.371715	4.835917	4.835917
min	17.000000	7.000000	7.000000
25%	22.750000	12.000000	12.000000
50%	25.500000	15.000000	15.000000
75%	28.000000	20.000000	20.000000
max	35.000000	25.000000	25.000000

Takeaways are how many times a defense forced a turnover, this could be either an interception or a fumble. ST_Sacks are how many sacks (when the defense tackles the quarterback before throwing) a defense had. ST_Special Teams Touchdowns represented the amount of touchdowns a team scored with their special teams unit.

2022 Takeaways	2022 ST_Sacks	2022 ST_Interceptions \\\n	2023 Takeaways	2023 ST_Sacks	2023 ST_Interceptions \\\n
count	32.000000	32.000000	count	32.000000	32.000000
mean	22.187500	40.500000	mean	22.937500	44.062500
std	4.855575	9.738517	std	5.453247	7.573032
min	13.000000	28.000000	min	11.000000	27.000000
25%	19.750000	35.750000	25%	18.000000	40.750000
50%	22.500000	40.000000	50%	24.000000	45.000000
75%	25.000000	44.250000	75%	27.000000	48.000000
max	33.000000	70.000000	max	31.000000	60.000000
2022 ST_Fumble Recoveries	2022 ST_Forced Fumbles	2022 ST_Safeties	2023 ST_Fumble Recoveries	2023 ST_Forced Fumbles	2023 ST_Safeties
count	32.000000	32.000000	count	32.000000	32.000000
mean	9.125000	12.000000	mean	9.406250	12.625000
std	2.756225	3.172462	std	2.637929	3.713532
min	3.000000	7.000000	min	3.000000	4.000000
25%	7.000000	10.000000	25%	8.000000	10.750000
50%	9.500000	11.000000	50%	9.500000	13.000000
75%	11.000000	14.000000	75%	11.000000	14.250000
max	17.000000	21.000000	max	15.000000	21.000000
2022 ST_Special Teams Touchdowns			2023 ST_Special Teams Touchdowns		
count	32.000000		count	32.000000	
mean	2.28125		mean	2.496250	
std	1.59099		std	1.521022	
min	0.000000		min	0.000000	
25%	1.000000		25%	1.000000	
50%	2.000000		50%	2.000000	
75%	3.000000		75%	3.000000	
max	8.000000		max	7.000000	
2024 Takeaways	2024 ST_Sacks	2024 ST_Interceptions \\\n	2024 Takeaways	2024 ST_Sacks	2024 ST_Interceptions \\\n
count	32.000000	32.000000	count	32.000000	32.000000
mean	20.562500	41.062500	mean	20.937500	42.093750
std	6.324237	7.266747	std	4.416579	4.000000
min	9.000000	28.000000	min	9.000000	28.000000
25%	17.000000	36.750000	25%	18.000000	40.000000
50%	25.000000	45.000000	50%	25.000000	45.000000
75%	33.000000	63.000000	75%	33.000000	63.000000
2024 ST_Fumble Recoveries	2024 ST_Forced Fumbles	2024 ST_Safeties	2024 ST_Fumble Recoveries	2024 ST_Forced Fumbles	2024 ST_Safeties
count	32.000000	32.000000	count	32.000000	32.000000
mean	8.375000	12.156250	mean	8.375000	12.156250
std	3.424206	3.418256	std	3.424206	3.418256
min	3.000000	5.000000	min	3.000000	5.000000
25%	6.000000	10.000000	25%	6.000000	10.000000
50%	8.000000	12.000000	50%	8.000000	12.000000
75%	10.000000	15.000000	75%	10.000000	15.000000
max	16.000000	18.000000	max	16.000000	18.000000
2024 ST_Special Teams Touchdowns			2024 ST_Special Teams Touchdowns		
count	32.000000		count	32.000000	
mean	2.000000		mean	2.000000	
std	1.367833		std	1.367833	
min	0.000000		min	0.000000	
25%	1.000000		25%	1.000000	
50%	2.000000		50%	2.000000	
75%	3.000000		75%	3.000000	
max	5.000000		max	5.000000	
2022 Passing Yards Allowed	2022 Rushing Yards Allowed \\\n		2023 Passing Yards Allowed	2023 Rushing Yards Allowed \\\n	
count	32.000000	32.000000	count	32.000000	32.000000
mean	3701.187500	2059.187500	mean	3721.718750	1915.218750
std	370.259929	361.000441	std	444.510621	258.107971
min	3057.000000	1307.000000	min	2800.000000	1468.000000
25%	3428.500000	1834.250000	25%	3498.500000	1733.500000
50%	3673.000000	2068.000000	50%	3851.000000	1917.000000
75%	3921.750000	2237.250000	75%	4000.250000	2105.750000
max	4671.000000	2894.000000	max	4457.000000	2434.000000
2022 Penalty Yards			2023 Penalty Yards		
count	32.000000		count	32.000000	
mean	780.500000		mean	800.625000	
std	75.881487		std	108.621791	
min	647.000000		min	530.000000	
25%	716.250000		25%	762.000000	
50%	785.500000		50%	803.000000	
75%	832.000000		75%	847.500000	
max	926.000000		max	986.000000	
2024 Passing Yards Allowed	2024 Rushing Yards Allowed \\\n		2024 Passing Yards Allowed	2024 Rushing Yards Allowed \\\n	
count	32.000000	32.000000	count	32.000000	32.000000
mean	3699.500000	2036.687500	mean	3699.500000	2036.687500
std	325.018114	330.044712	std	325.018114	330.044712
min	2961.000000	1361.000000	min	2961.000000	1361.000000
25%	3581.500000	1755.000000	25%	3581.500000	1755.000000
50%	3706.500000	2056.000000	50%	3706.500000	2056.000000
75%	3851.250000	2244.250000	75%	3851.250000	2244.250000
max	4375.000000	3057.000000	max	4375.000000	3057.000000
2024 Penalty Yards			2024 Penalty Yards		
count	32.000000		count	32.000000	
mean	887.906250		mean	887.906250	
std	116.456283		std	116.456283	
min	624.000000		min	624.000000	
25%	806.750000		25%	806.750000	
50%	869.500000		50%	869.500000	
75%	967.250000		75%	967.250000	
max	1185.000000		max	1185.000000	

Continuous: Passing Yards, Rushing Yards, Receiving Yards, Total Yards Allowed, Passing Yards Allowed, Rushing Yards Allowed, Penalty Yards, Percent Drives Scored On, Percent Drives Takeaways, Fantasy Points from Points, PPR Fantasy Points Scored, Standard Fantasy Points Scored

	2022 Passing Yards	2022 Rushing Yards	2022 Receiving Yards	
count	113.000000	586.000000	586.000000	
mean	1124.203540	112.313993	216.598976	
std	1473.904037	259.748426	302.625160	
min	0.000000	-18.000000	-10.000000	
25%	17.000000	0.000000	9.000000	
50%	390.000000	5.000000	90.000000	
75%	2163.000000	70.000000	308.750000	
max	5250.000000	1653.000000	1809.000000	

	2023 Passing Yards	2023 Rushing Yards	2023 Receiving Yards	
count	114.000000	561.000000	561.000000	
mean	1127.508772	109.167558	228.985740	
std	1480.065188	241.652666	319.657752	
min	-7.000000	-23.000000	-10.000000	
25%	11.250000	0.000000	10.000000	
50%	277.500000	4.000000	93.000000	
75%	2036.500000	74.000000	313.000000	
max	4624.000000	1459.000000	1799.000000	

	2024 Passing Yards	2024 Rushing Yards	2024 Receiving Yards	
count	107.000000	566.000000	566.000000	
mean	1187.140187	115.077739	224.226148	
std	1510.332185	271.626392	305.854668	
min	0.000000	-7.000000	-5.000000	
25%	17.000000	0.000000	7.000000	
50%	276.000000	2.000000	80.000000	
75%	2080.500000	70.000000	342.000000	
max	4918.000000	2005.000000	1708.000000	

	2022 Percent Drives Scored On	2022 Percent Drives Takeaway	
count	32.000000	32.000000	
mean	36.275000	11.381250	
std	3.958331	2.521832	
min	25.700000	7.100000	
25%	33.800000	10.150000	
50%	36.050000	11.550000	
75%	38.825000	13.300000	
max	43.500000	16.200000	

	2023 Percent Drives Scored On	2023 Percent Drives Takeaway	
count	32.000000	32.000000	
mean	35.593750	11.568750	
std	4.759231	2.721176	
min	24.800000	5.500000	
25%	32.075000	9.650000	
50%	34.950000	11.750000	
75%	38.550000	14.125000	
max	46.800000	16.000000	

	2024 Percent Drives Scored On	2024 Percent Drives Takeaway	
count	32.000000	32.000000	
mean	38.846875	10.696875	
std	4.856388	3.271429	
min	31.600000	4.400000	
25%	35.050000	8.775000	
50%	39.000000	9.700000	
75%	42.825000	12.850000	
max	51.400000	17.500000	

Fantasy from points represents the sum of how many fantasy points a defense earned from points allowed for each week, this was a calculated field. PPR Fantasy Points Scored represents the amount of fantasy points a player would have scored in a PPR (Points per Reception) league. Standard Fantasy Points Scored represents how many fantasy points a player would have scored in a fantasy league under standard scoring. These were both calculated fields.

	2022 Fantasy Points From Points	2022 PPR Fantasy Points Scored	
count	32.000000	686.000000	
mean	5.281250	75.548805	
std	8.666712	80.295509	
min	-12.000000	-5.000000	
25%	-1.250000	10.000000	
50%	5.500000	47.450000	
75%	12.000000	115.075000	
max	27.000000	403.400000	

	2023 Fantasy Points From Points	2023 PPR Fantasy Points Scored	
count	32.000000	653.000000	
mean	6.250000	79.612956	
std	9.466409	82.510807	
min	-20.000000	-4.200000	
25%	0.750000	11.960000	
50%	5.500000	50.300000	
75%	12.250000	126.800000	
max	26.000000	399.200000	

	2023 Standard Fantasy Points Scored	
count	653.000000	
mean	61.551700	
std	68.763344	
min	-4.200000	
25%	7.700000	
50%	33.900000	
75%	103.380000	
max	380.640000	

	2024 Fantasy Points From Points	2024 PPR Fantasy Points Scored	
count	32.000000	663.000000	
mean	2.312500	79.342413	
std	10.399888	84.473830	
min	-26.000000	-8.700000	
25%	-6.000000	10.500000	
50%	2.000000	51.800000	
75%	9.250000	123.000000	
max	21.000000	420.380000	

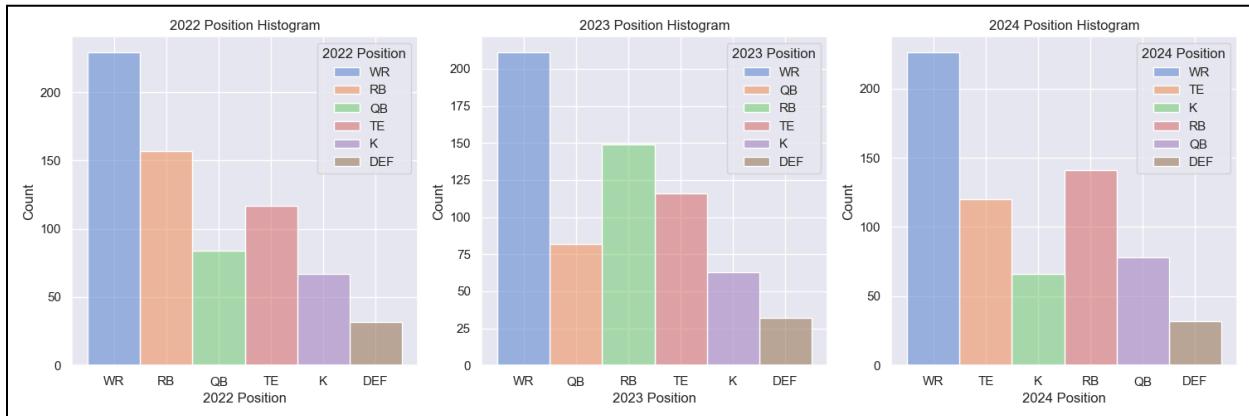
	2024 Standard Fantasy Points Scored	
count	663.000000	
mean	61.826576	
std	71.585064	
min	-9.700000	
25%	5.300000	
50%	35.800000	
75%	96.900000	
max	420.380000	

Non-graphical and graphical univariate analysis

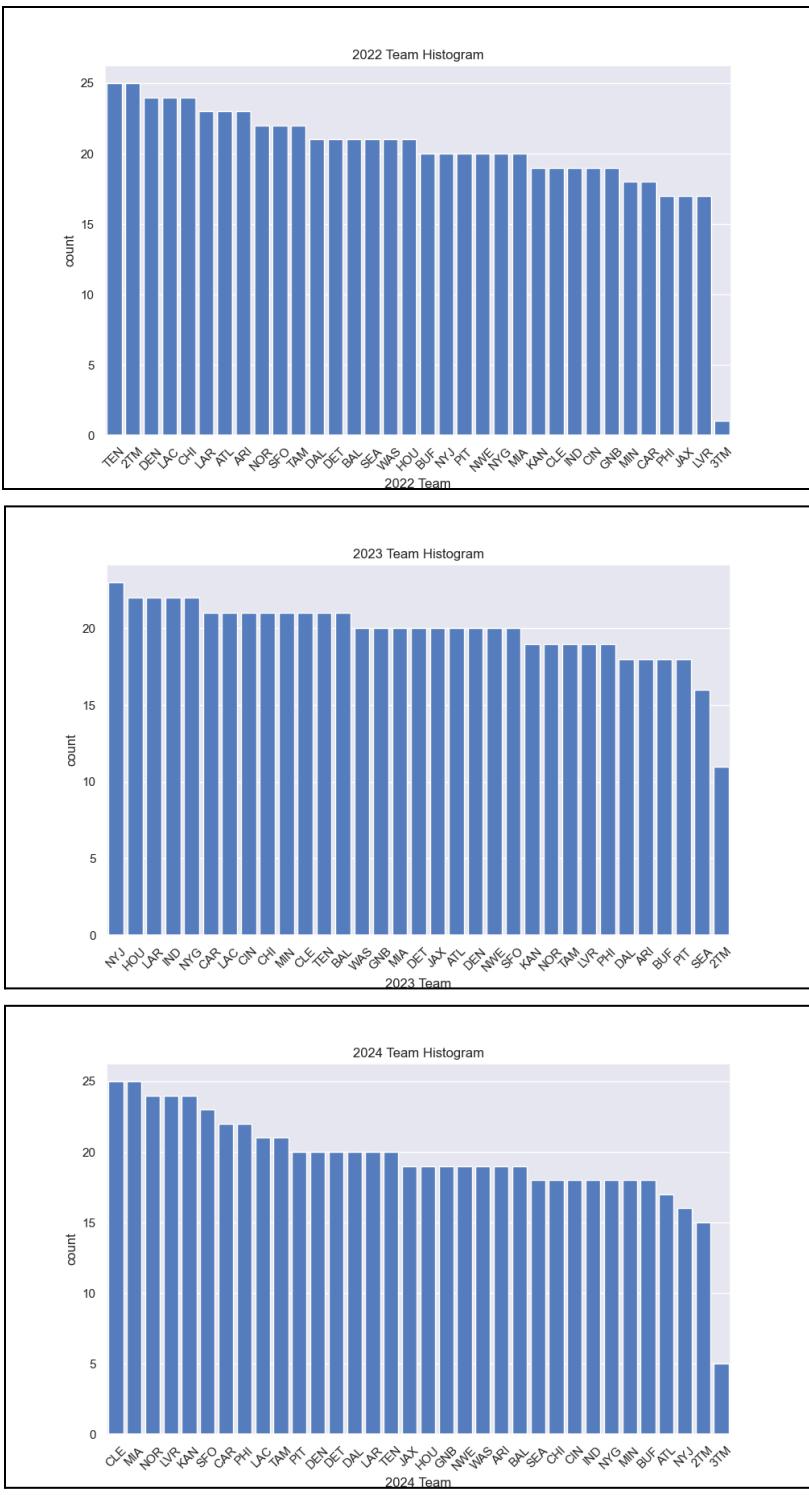
Categoricals

We have 6 categorical columns representing 2 different features across 3 years. These features are Position and Team for the years 2022, 2023, and 2024. For the positions we do see some variation amongst the exact numbers for each position, but the general order is maintained. From the most to the least we see Wide Receivers (WR), Running Backs (RB), Tight Ends (TE), Quarterbacks (QB), Kickers (K), and Defensive Teams (DEF). We do not anticipate an issue with data imbalance as this is representative of how many options there are in the league. See the table and histogram below.

	WR	RB	TE	QB	K	DEF
2022	229	157	117	84	67	32
2023	211	149	116	82	63	32
2024	226	141	120	78	66	32



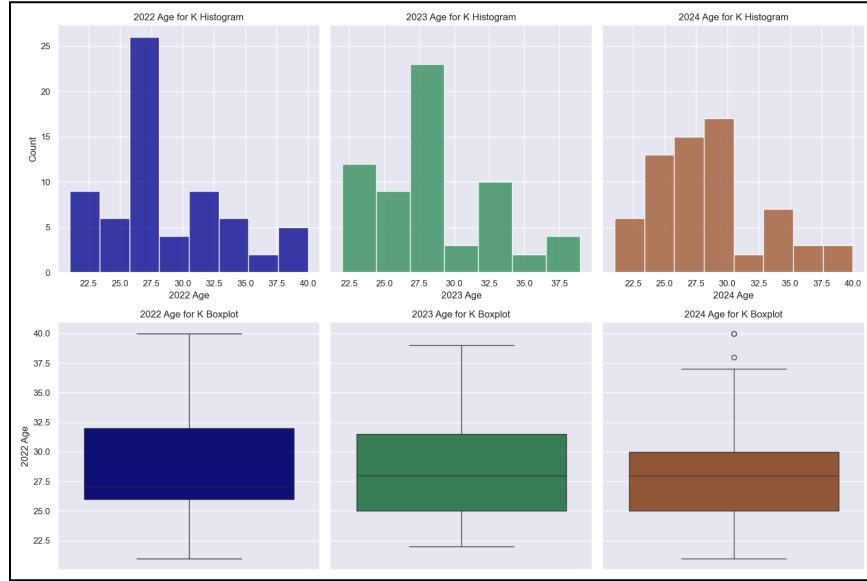
We also analyzed the distributions of players across teams for the 3 years we look at. Looking at the tables below we see changes from year to year with no clear pattern. For example the New York Jets (NYJ) are roughly middle of the pack in 2022, the top in 2023, and the third from the bottom in 2024. There's no clear and easy pattern to extract from these graphs. However, it is notable that the New York Jets had a particularly rough year with injuries in 2023 and this may explain their year to year shifts. For reference, Aaron Rodgers (their star quarterback) tore his achilles in 2023, which led the team to experiment with a plethora of other quarterback options, including Zach Wilson and Tim Boyle. This leads us to believe that this year to year changes in positional counts has to do with injuries, player performance, and other statistics that this dataset does not include.



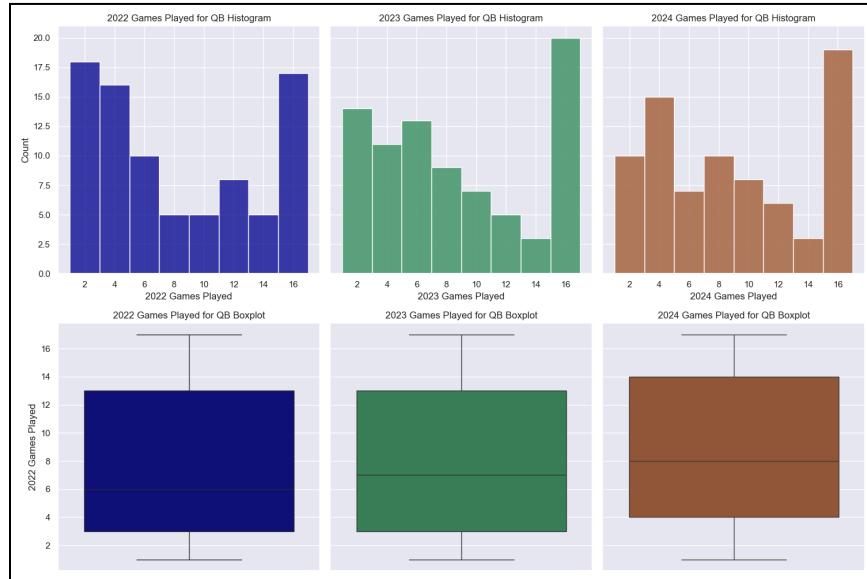
Numericals

For this analysis we compare the features by positions and highlight the most relevant patterns seen in the data. We display this information through histograms and box plots. Due to the sheer volume of features, we elect to show only relevant plots directly in the work and the reminder can be seen in our

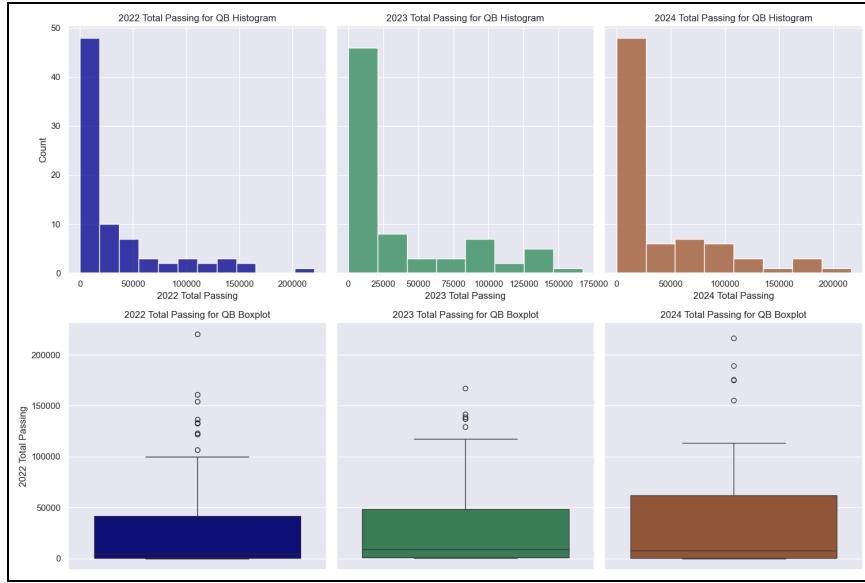
code repository in the Categorical and Univariate Analysis Jupyter Notebook in the interest of good formatting of the report.



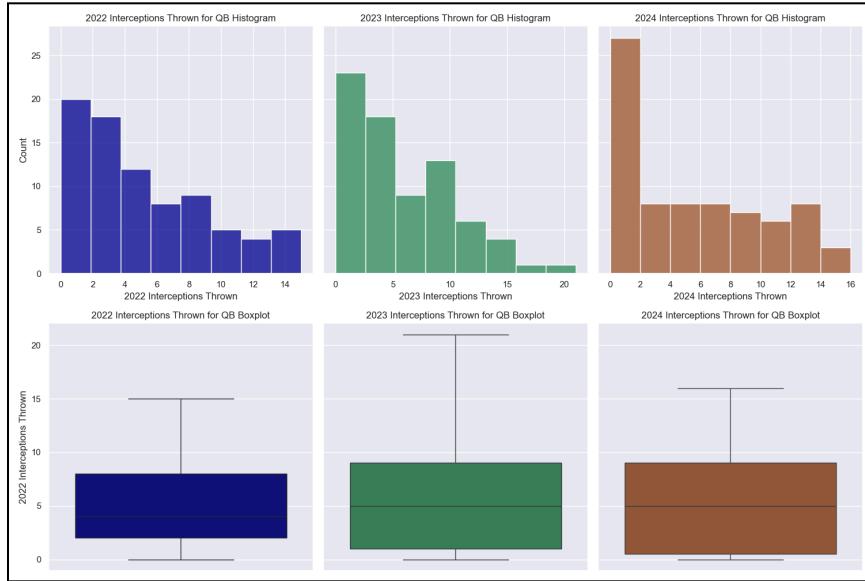
For most of the positions, the age histogram displays a right skewed distribution that contains more younger players. However we see that the Kicker age histogram has a relatively unique shape with a tendency for older players demonstrated by the highest mean age amongst positions. This may be a unique feature of the position that's perhaps less based on pure athleticism and more on technique and skill.



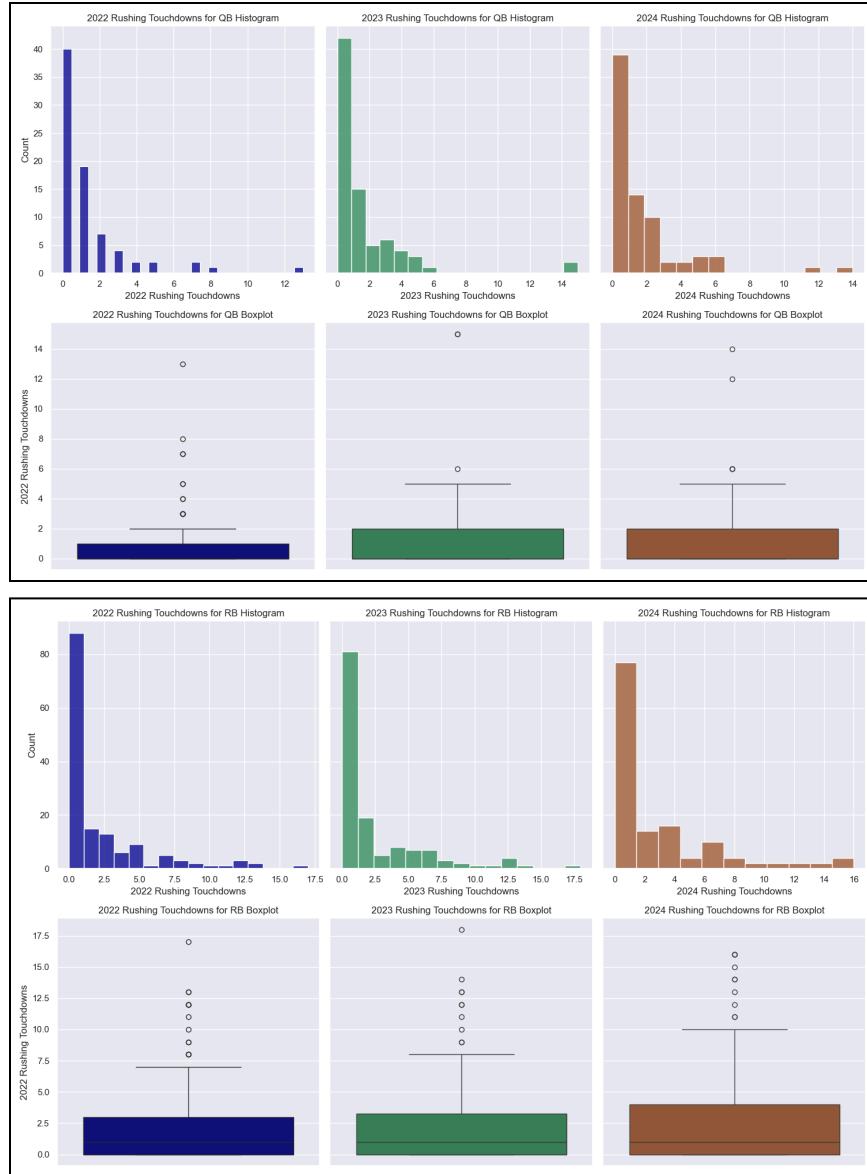
For all players the Games Played feature tends to be left skewed with most positions playing in most of the games. The notable exception is Quarterbacks which have more bimodal distribution with peaks at the high and low numbers of games played. Games Started is similar to Games Played but less extreme and right skewed for everyone instead.



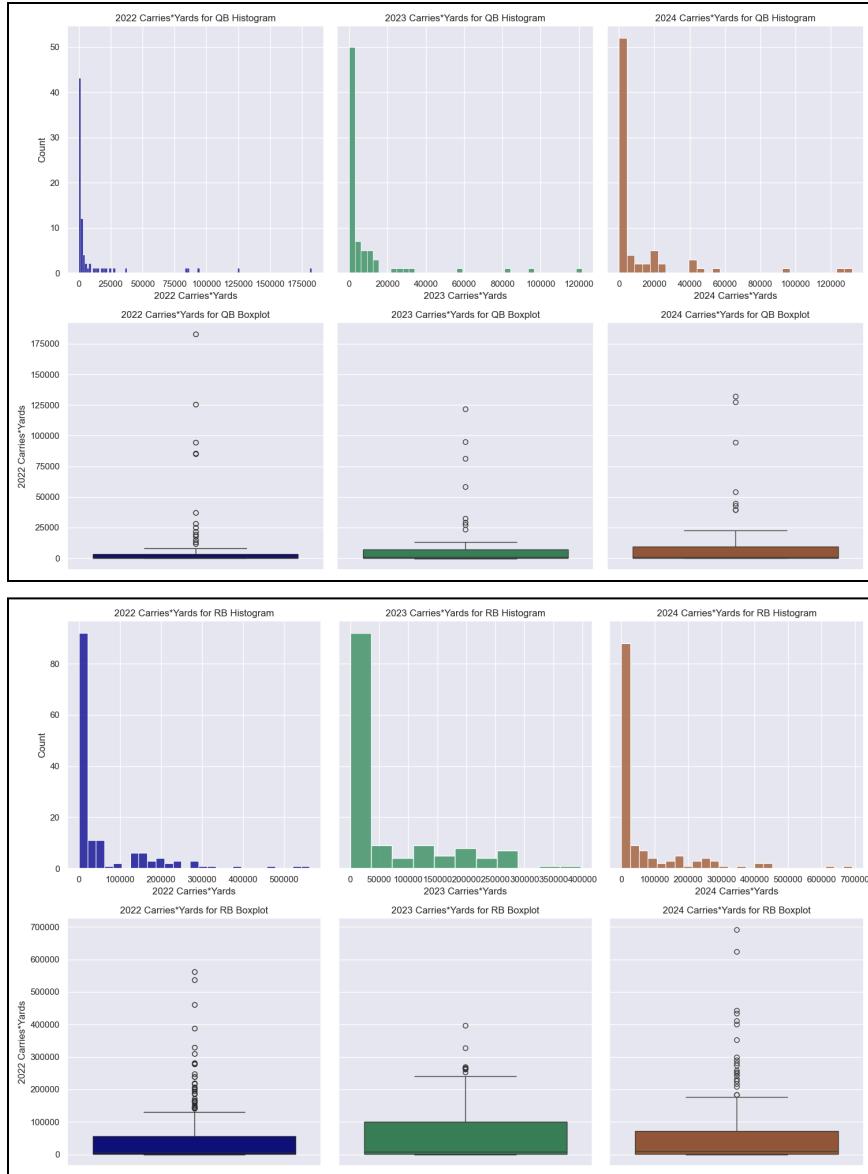
For Total Passing we see it is decidedly right skewed, indicating some players simply perform better. These are the quarterbacks we would wish to select.



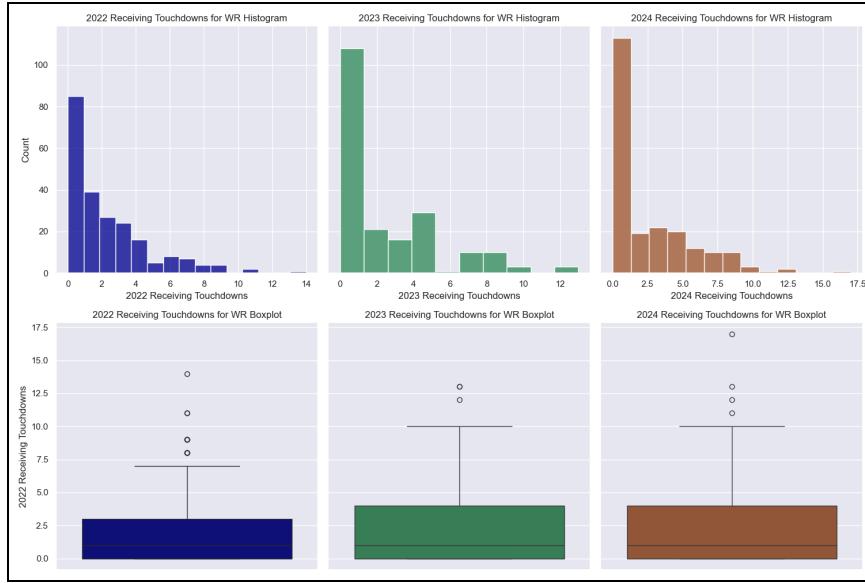
Interceptions Thrown is generally right skewed. This suggests high performance (low interceptions) is the most common for Quarterbacks. This makes sense as most quarterbacks are not going to get enough opportunity to throw more interceptions.



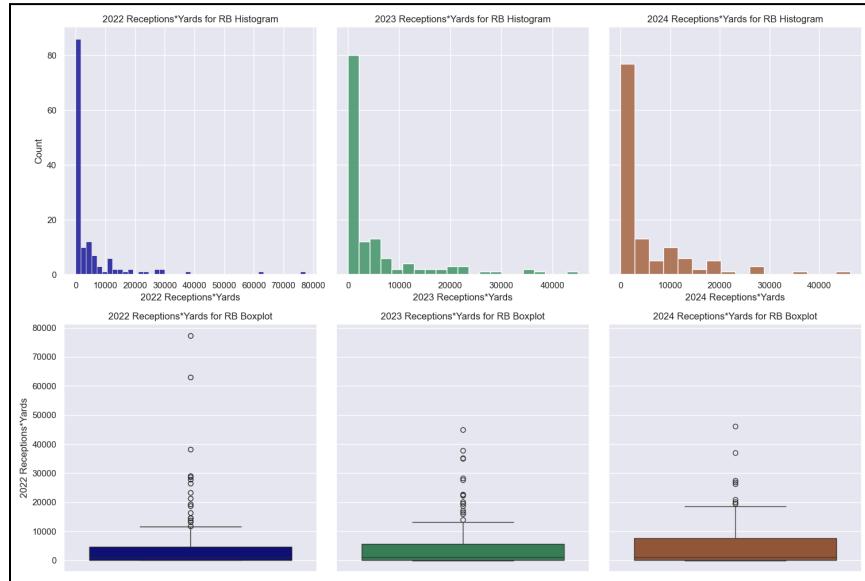
The Rushing Touchdowns variable tends to be right skewed with serious outliers. TEs and WRs have some outliers, but much less compared to QBs and RBs which also have a higher range of rushing touchdowns.

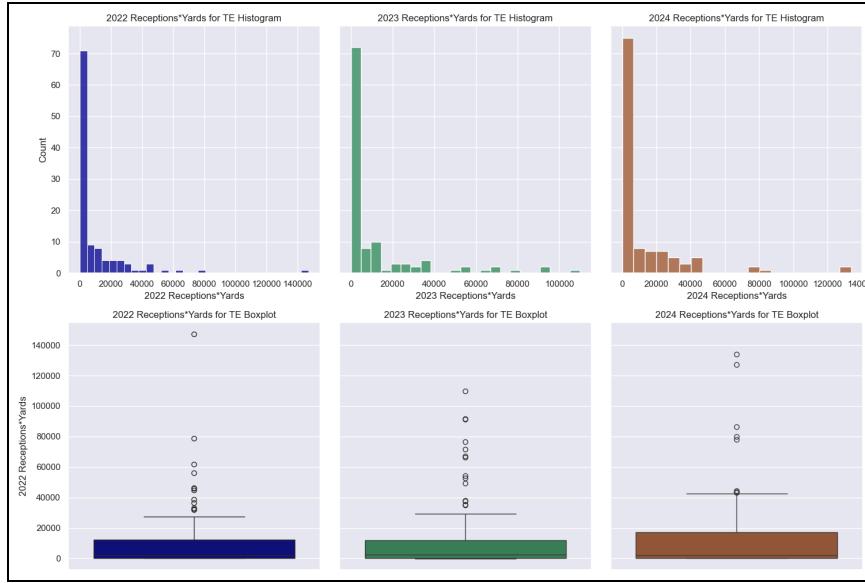


The Carries*Yards feature has an **extreme** right skew. The higher performers are in a totally different league. QBs and RBs have significantly higher values for this statistic. WRs and TEs have values, but they are extremely small compared with QBs and RBs.

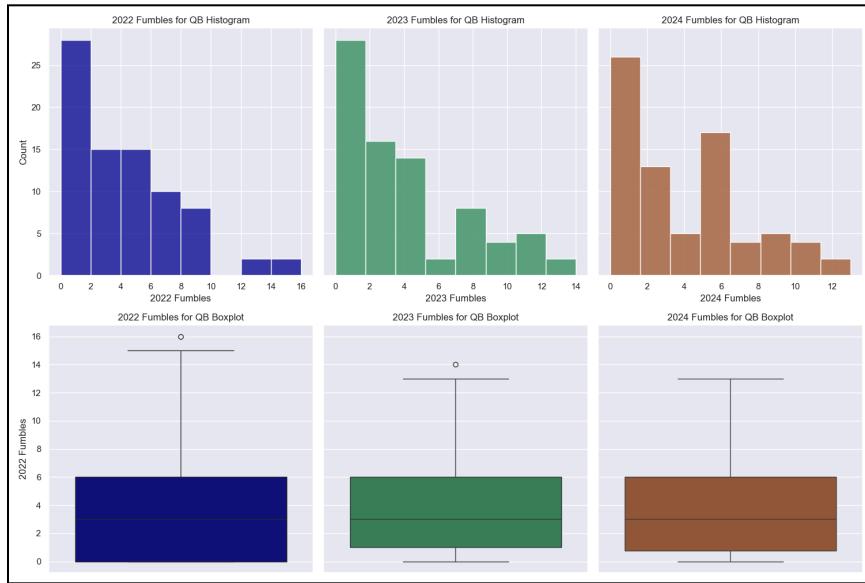


For Receiving Touchdowns we can see a tendency towards a right skew with WRs having the most touchdowns. Logically, this makes the most sense as receivers are targeted more frequently than other positions like RBs and TEs.

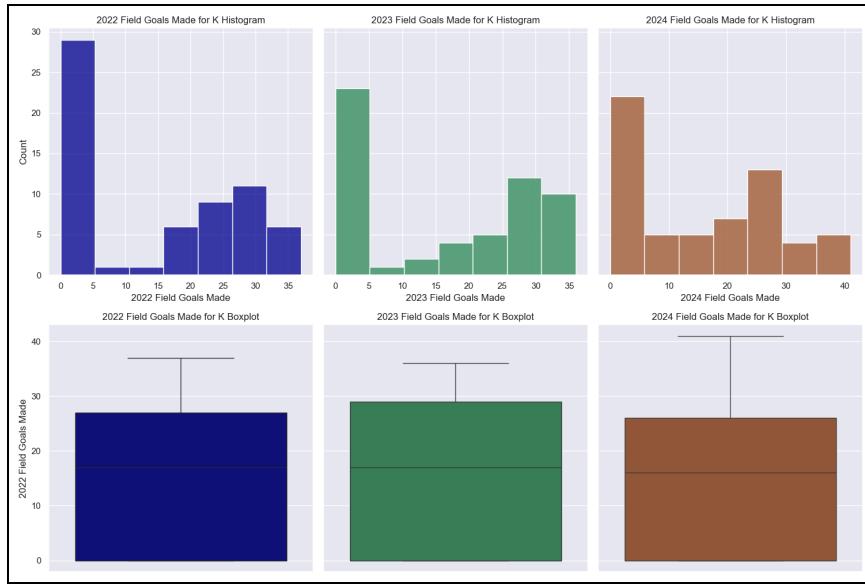




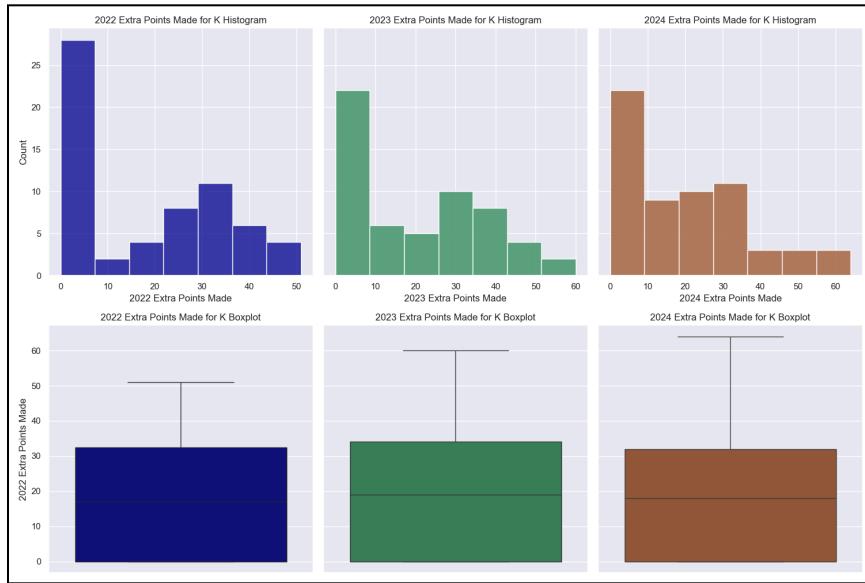
Reception*Yards has an **extreme** right skew showing some players are simply performing at another level. WRs, TEs, and RBs are the most relevant positions for this statistic.



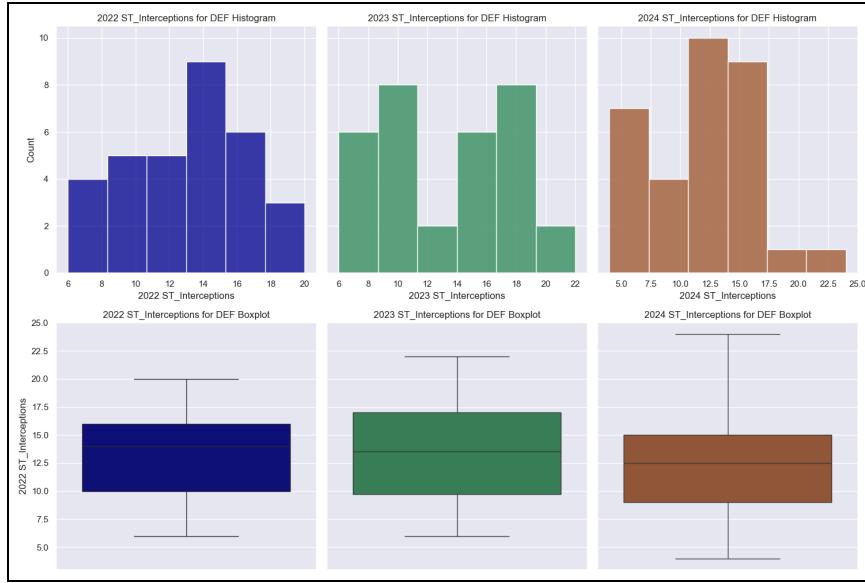
QBs have the widest range of Fumbles, likely because they have hands on the ball more than any other position. This makes fumbles a hard metric to approach. In theory higher values should be worse, but it is almost certainly a function of how much time you have handling the ball which will be correlated with other positive metrics.



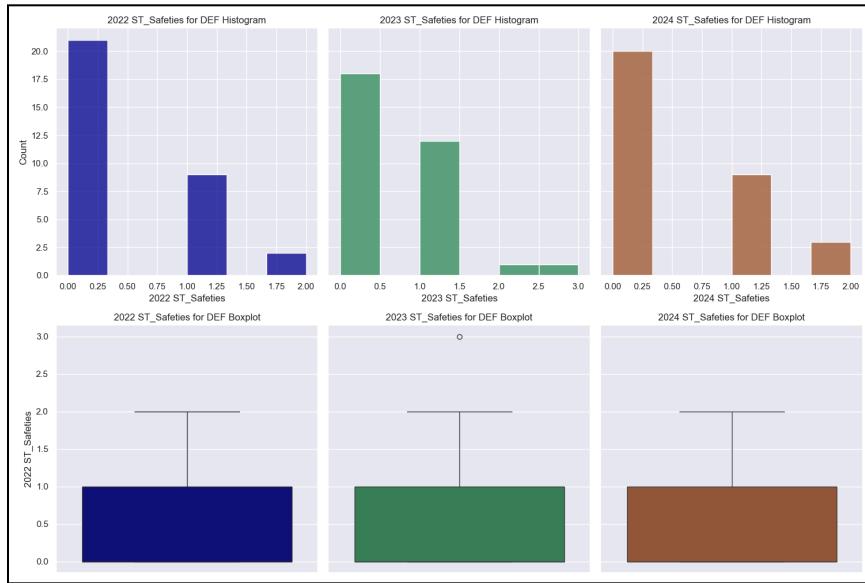
In Field Goals Made we see two clear groups are present, a normal distribution of kickers scoring field goals and a large group with little to no field goals.



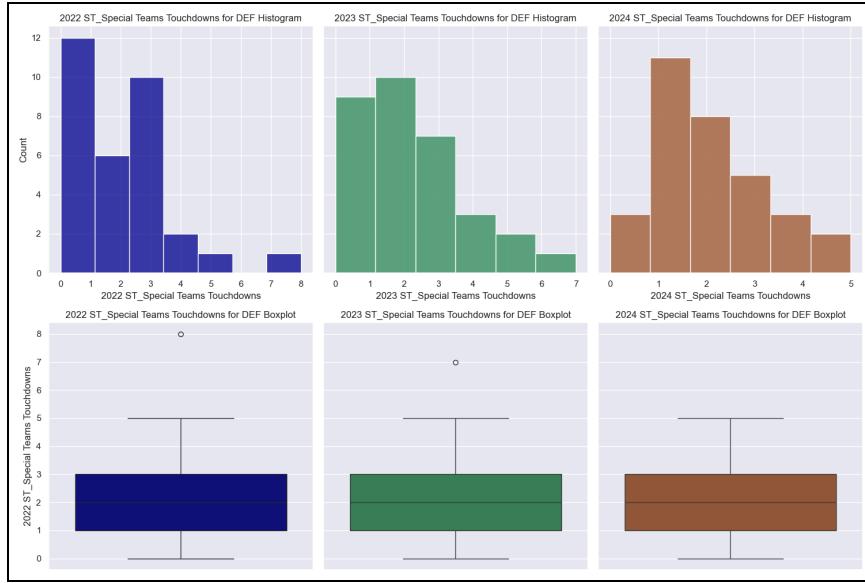
For Extra Points Made we see a similar distribution to field goals made with the exception of 2024 which seems to have shifted to fewer extra points made.



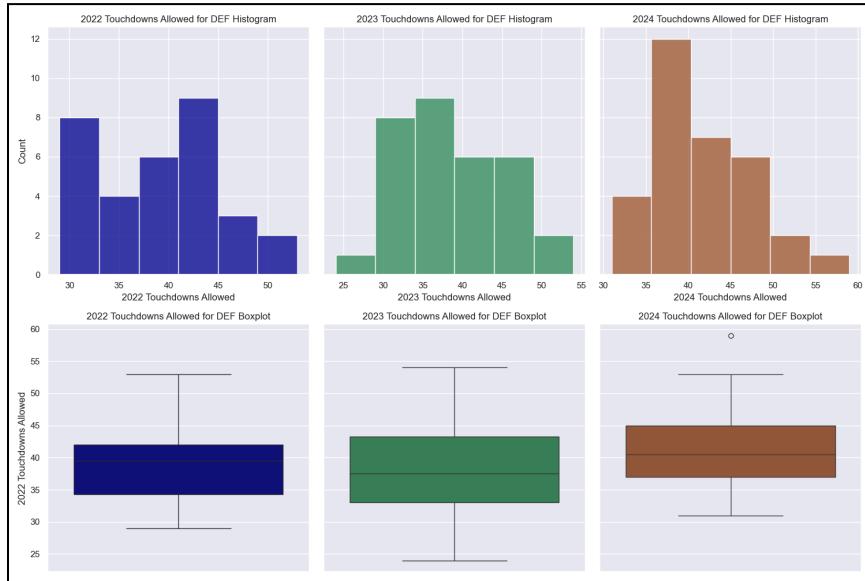
There are significant shifts to the distribution of Special Team Interceptions from year to year, with no clear pattern to extract.



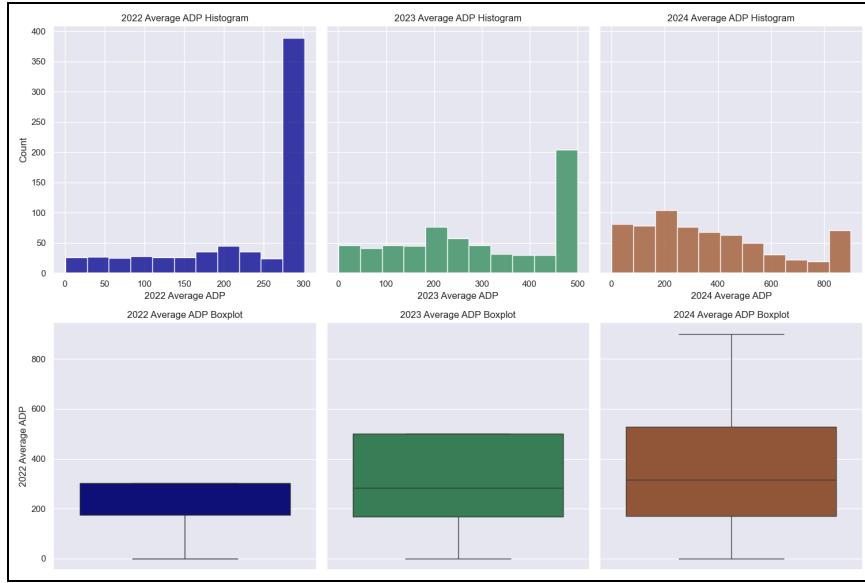
The most notable takeaway from our analysis of special teams statistics is the relative rarity of safeties. Safeties are relatively rare in the NFL, as in order to record one, there is a lot of situational luck that must take place (i.e. have the offense backed up in their own red zone, offensive line misses a block, etc.).



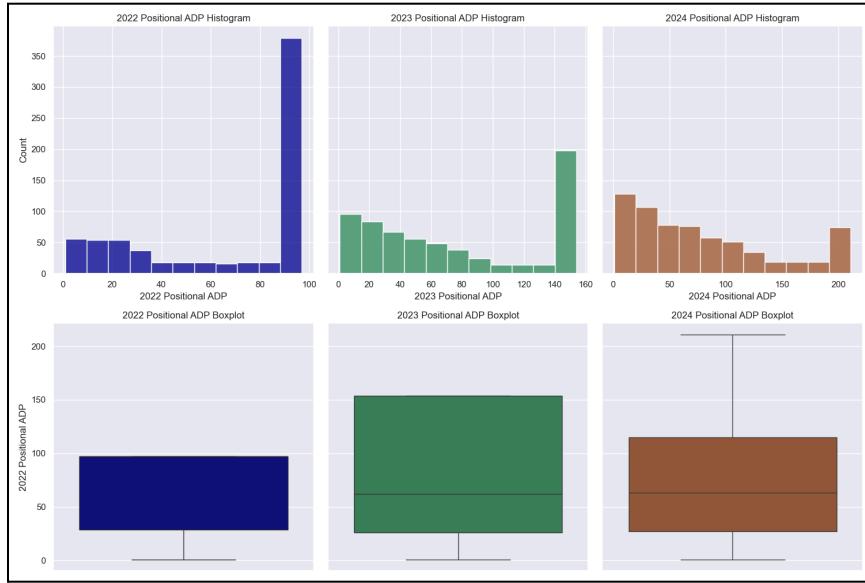
There is no clear pattern to the Special Team Touchdowns but one observation is that between 2022 and 2024, the maximum special teams touchdowns scored dropped each year. Whether or not this is a real pattern is yet to be confirmed.



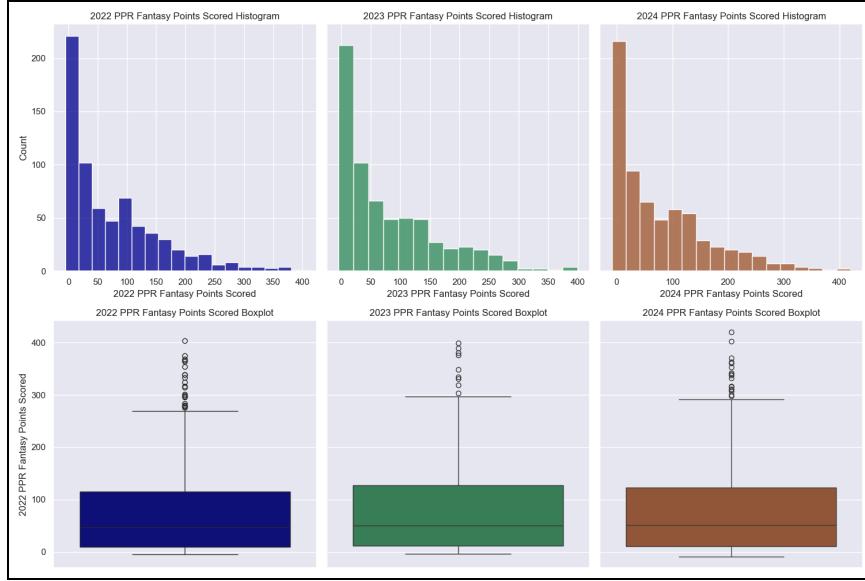
Defensive touchdowns allowed seems to have a somewhat normal distribution. It seems that more touchdowns are occurring in recent years, and would likely be an interesting metric to continue monitoring into the future.



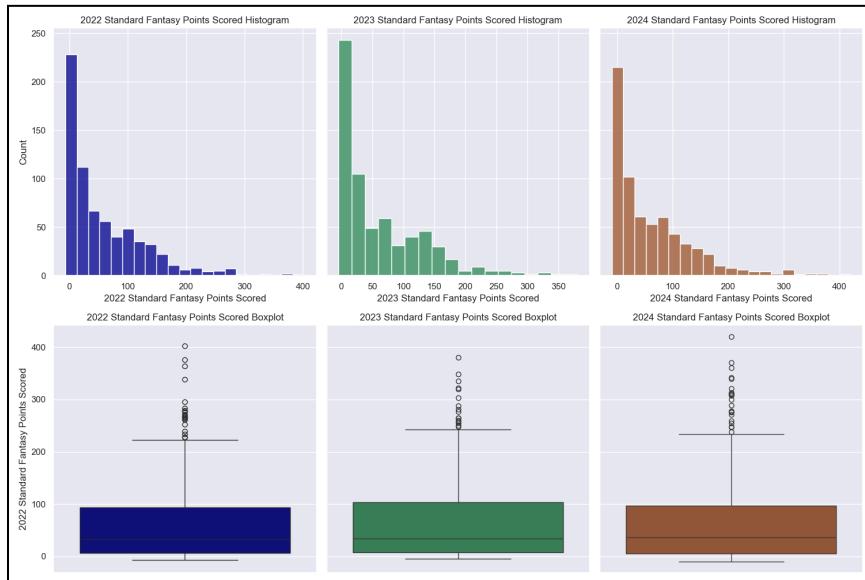
For the Average ADP, if we look at this metric holistically then there are not too many takeaways. However, if we break down by position, we see that the QBs, WRs, TEs, and RBs all have players across the full spread of ranges, while the Ks and the DEFs have values that start around 80-100 suggesting that these positions are selected much later in drafts. The graphs that break down by position can be found in the appendix.



There is a tendency to have a region of competitive positional ADP (low values) and a large amount of noncompetitive positional ADP (higher values). This suggests that certain players in each position are preferred due to a clearly higher performance.



For the PPR Fantasy points scored we see that there is a strong right skew indicating that the high performing players are both rarer and performing at a much higher level. Breaking them down by positions, the DEF fantasy points tend to start at a higher value which makes sense as all defenses will play unlike individual players and has a tendency towards a normal distribution. For the remaining positions, we see heavily right skewed data. We also see that the higher end of QBs is well above all other positions at around 400. WRs and RBs are roughly in the same ballpark around 350 followed by TEs around 300. Ks and DEFs tend to max around 160. This suggests a tendency for a rough order to draft that may emerge from our recommender system. The graphs broken down by position can be found in the appendix.



The general trends within Standard scoring are the same as PPR except the QBs are much more emphasized as WRs, TEs, and RBs have much lower values. Positional breakdowns can also be found in the appendix.

Missing Value Analysis and Outlier Analysis

Missing Values

Our final dataset was populated via our finely-tuned web scraper that ensured the accuracy and cleanliness of the data pulled from our “golden standard” reputation sources (Pro Football Reference, NFL.com, etc.). As such, we did not encounter typical “missing” values. In this scraped dataset, null values are present and expected where a player did not participate in a given season. For instance, if a player was injured or suspended, their data for that season would correctly be marked as null, reflecting their absence from games. Similarly, a player first drafted into the league in 2024 (example: QB Jayden Daniels) would appropriately have null data for columns representing the 2022 season.

	Player Name	2022 Position	2022 Total Passing	2022 Average ADP	\
461	Jayden Daniels	NaN	NaN	NaN	
<hr/>					
461	2022 PPR Fantasy Points Scored	2022 Games Played			
<hr/>					
	Player Name	2024 Position	2024 Total Passing	2024 Average ADP	\
461	Jayden Daniels	QB	89200.0	102.6	
<hr/>					
461	2024 PPR Fantasy Points Scored	2024 Games Played			
<hr/>					
461		339.82	17.0		

Certain player statistics are naturally null because they do not apply to the player’s role. For instance, it is entirely expected and appropriate for a wide receiver to have null values for all columns dealing with field goals. We also expect these null values to potentially fluctuate across seasons, especially when players are not available due to injuries or suspension. Given the context of our data and our eventual modeling strategy, it is entirely reasonable for a player to have null values for the 2023 season if they were inactive, while their 2022 and 2024 season columns are fully populated with appropriate data. This variability is to be expected and is consistent with the nature of player availability across different seasons. It is clear to see from our non-null calculation lists (by season) below that fluctuations in null values exist as we expected. Another key insight from these lists is that for most defensive/special teams stats, there should be the same amount of non-null values from season to season. This is expected, and a further sign of our dataset’s quality, as there will only ever be 32 NFL team defenses to choose from to draft.

1	2022 Age	654	non-null
2	2022 Position	686	non-null
3	2022 Team	686	non-null
4	2022 Games Played	686	non-null
5	2022 Games Started	686	non-null
6	2022 Total Passing	113	non-null
7	2022 Interceptions Thrown	113	non-null
8	2022 Rushing Touchdowns	586	non-null
9	2022 Carries*Yards	586	non-null
10	2022 Receiving Touchdowns	586	non-null
11	2022 Receptions*Yards	586	non-null
12	2022 Fumbles	586	non-null
13	2022 Field Goals Made	64	non-null
14	2022 Extra Points Made	64	non-null
15	2022 ST_Interceptions	32	non-null
16	2022 ST_Safeties	32	non-null
17	2022 ST_Special Teams Touchdowns	32	non-null
18	2022 XP2	24	non-null
19	2022 Touchdowns Allowed	32	non-null
20	2022 Average ADP	686	non-null
21	2022 Positional ADP	686	non-null
22	2022 PPR Fantasy Points Scored	686	non-null
23	2022 Standard Fantasy Points Scored	686	non-null
24	2023 Age	621	non-null
25	2023 Position	653	non-null
26	2023 Team	653	non-null
27	2023 Games Played	653	non-null
28	2023 Games Started	653	non-null
29	2023 Total Passing	114	non-null
30	2023 Interceptions Thrown	114	non-null
31	2023 Rushing Touchdowns	561	non-null
32	2023 Carries*Yards	561	non-null
33	2023 Receiving Touchdowns	561	non-null
34	2023 Receptions*Yards	561	non-null
35	2023 Fumbles	561	non-null
36	2023 Field Goals Made	58	non-null
37	2023 Extra Points Made	58	non-null
38	2023 ST_Interceptions	32	non-null
39	2023 ST_Safeties	32	non-null
40	2023 ST_Special Teams Touchdowns	32	non-null
41	2023 XP2	23	non-null
42	2023 Touchdowns Allowed	32	non-null
43	2023 Average ADP	653	non-null
44	2023 Positional ADP	653	non-null
45	2023 PPR Fantasy Points Scored	653	non-null
46	2023 Standard Fantasy Points Scored	653	non-null
47	2024 Age	631	non-null
48	2024 Position	663	non-null
49	2024 Team	663	non-null
50	2024 Games Played	663	non-null
51	2024 Games Started	663	non-null
52	2024 Total Passing	107	non-null
53	2024 Interceptions Thrown	107	non-null
54	2024 Rushing Touchdowns	566	non-null
55	2024 Carries*Yards	566	non-null
56	2024 Receiving Touchdowns	566	non-null
57	2024 Receptions*Yards	566	non-null
58	2024 Fumbles	566	non-null
59	2024 Field Goals Made	61	non-null
60	2024 Extra Points Made	61	non-null
61	2024 ST_Interceptions	32	non-null
62	2024 ST_Safeties	32	non-null
63	2024 ST_Special Teams Touchdowns	32	non-null
64	2024 XP2	22	non-null
65	2024 Touchdowns Allowed	32	non-null
66	2024 Average ADP	663	non-null
67	2024 Positional ADP	663	non-null
68	2024 PPR Fantasy Points Scored	663	non-null
69	2024 Standard Fantasy Points Scored	663	non-null

Additionally, no null values were found in the “Player Name” column, confirming that the dataset retains complete identification for every player who has participated at least partially across the 2022-2024 seasons. As seen below, there are 970 non-nulls in the “Player Name” column, perfectly matching the total length of 970 rows of our dataset. The df.info() function was used to identify these instances, but it is vital to again emphasize that due to the nature of this data, nulls are not errors. It should also be noted that the difference between null values and 0 values is important in this context. In many cases, a statistic with a value equal to 0 is entirely valid. For instance, a backup Quarterback who only participated in 1 offensive drive all season may have completed only 1 pass for 0 yards. In this case, 0 is appropriate because, while the player did not produce much in terms of statistical value, they did participate in gametime, meaning a null value would be inaccurate.

RangeIndex: 970 entries, 0 to 969
Data columns (total 70 columns):
Column

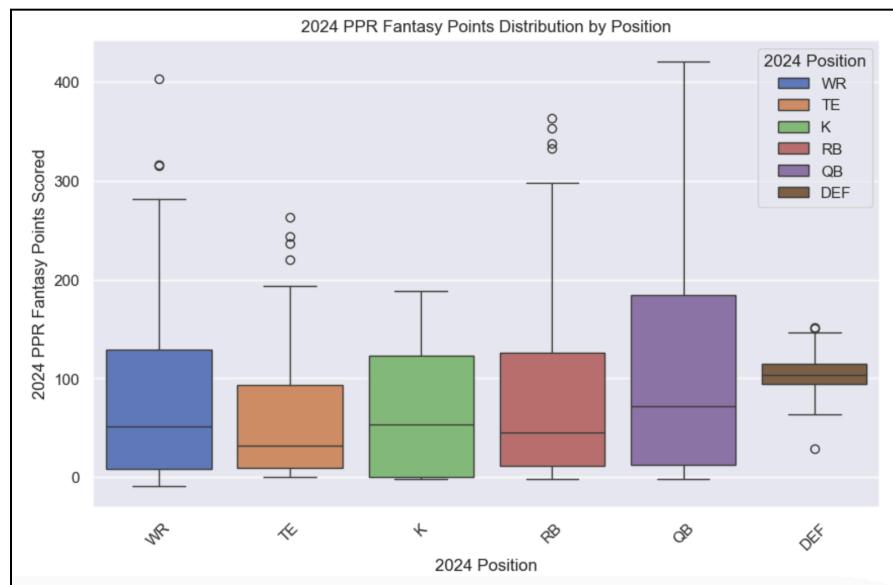
0 Player Name
Non-Null Count Dtype
970 non-null object

Outlier Analysis

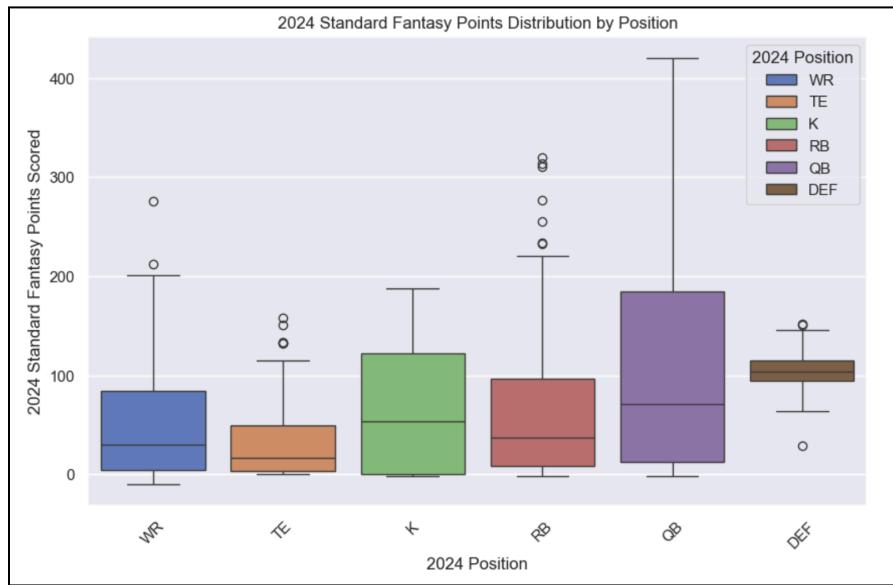
In most “traditional” machine learning situations, identifying and dealing with outliers is a very important part of both preprocessing and EDA. However, due to the context of our Fantasy Football Draft Assistant looking to aid managers in identifying high performing players, certain outliers in our dataset are actually a key consideration for what our recommendation system should aim to prioritize. In other words, our primary strategy is not to remove outliers, but instead keep them intact due to their potential to

represent high value players and potential draft picks. In this section, we will be discussing different boxplots and distributions of our data in the context of outlier detection by position (and potential ideas for transformation), as well as focusing on some specific outliers in our data that we find provide insight into the dimensions, properties, and purpose of our dataset in relation to our recommender system. While we will only be highlighting specifically insightful examples in this section, the entirety of our outlier analysis plot collection can be found in our project's GitHub repository. It should be noted that if our recommendation system's performance in the future indicates issues that may be caused by this lack of addressing outliers and/or heavily skewed distributions even with transformation, we will further explore how to handle these outliers through more advanced strategies such as winsorization.

To further emphasize the atypical relevance of outliers in the context of our project, we will first analyze some boxplots dealing with the primary feature which our recommendation system will ultimately revolve around; PPR (point per reception) and Standard fantasy scoring. Seen below is the fantasy point distributions by position for the 2024 season measured by the PPR scoring system. In general, we can see that several position groups have “positive” outliers in terms of high fantasy scoring. In many other machine learning projects, one may be inclined to remove these outliers as they are placed significantly outside of their respective position’s box and whiskers. However, in our case these outlying players are exactly the kind of data points we want our model to capture and prioritize, as they represent the highest performing players in terms of scoring that will potentially be available to draft. Another fascinating element of our data that we can see through this graph is the smaller variance in fantasy scoring amongst defensive units compared to other position groups. This makes sense and aligns with the context of the data for multiple reasons. Individual offensive players may have situationally dependent roles or playing time, such as when a team relies less on running from RBs due to poor in-game performance or time constraints when trailing late in a game. 32 defensive units, and only 32 units, will always play a full game, regardless of any game scripting or individual roster adjustments. Similarly, a defensive unit consists of 11 players on the field, often the highest performing on the team at each respective defensive position, at any given time. This means that defensive units are less susceptible to individual variance or game-by-game performance fluctuations that we see in individual players.



We see the exact same distribution trends across each position group in the standard fantasy points boxplot for 2024 below, including the aforementioned smaller variance amongst defensive units. However, the shift in y-axis “fantasy points scored” scaling, while expected, is statistically significant for our model. Standard fantasy scoring, as opposed to point-per-reception/PPR scoring, does not additionally reward players for successful receptions in passing situations. As a result, player groups that are affected by this change (WRs, TEs, RBs) have lower scoring.

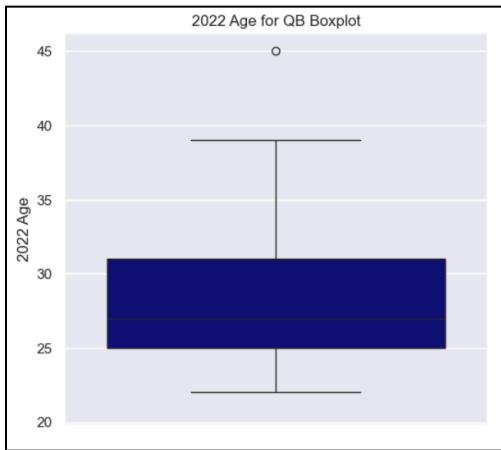


As seen in the previous non-graphical and graphical univariate analysis section, the majority of the distribution plots for non-defensive stats show a moderate to extreme rightward skew. In the following examples of Rushing Touchdowns, Carries*Yards, and Receiving Touchdowns for the 2023 season, we can see that many of these distributions follow a power-law like pattern. In the context of our model, a small number of elite players producing disproportionately high on-field output while the majority remain clustered at lower values is expected due to the varying talent levels, positional roles, and opportunities given to all players in the NFL. This skew is crucial to our model because it strongly influences and indicates outlier detection of elite players versus standard talents. Additionally, we applied a variety of transformations to our data in an attempt to further explore and validate this expectation. Strategies of Log-Scaled, Box-Cox, Yeo-Johnson, and Sqrt (square root) transformations were applied as part of our analysis. As seen below, they had varying levels of success in “normalizing” our distributions. Our strategy moving forward will be to discover and apply each individual feature’s most successful transformation technique prior to modeling.

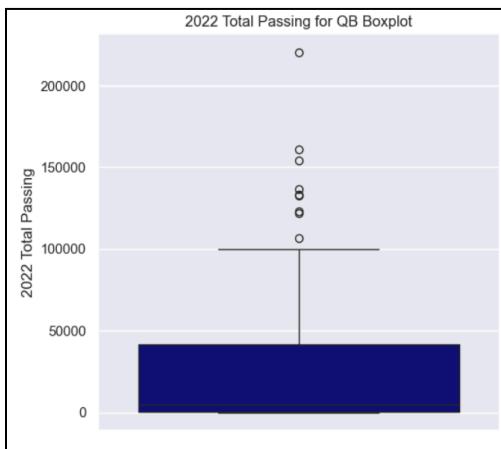


Quarterbacks (QBs)

One interesting display of an outlier we can clearly identify amongst quarterback stat distributions comes in the QB Age boxplot for the 2022 season, seen below. The plot shows a fairly standard box and whisker spread from ages 22-39, with the exception of one extreme outlier. This outlier represents the 2022 season of NFL legend Tom Brady, who was 45 years old at the time. This distribution is a great example of the typically extraordinary longevity and consistency of most older outliers in our Age plots. To be able to play a pivotal position like QB while still producing at an elite level is the exact sort of unique player one might be interested in drafting.

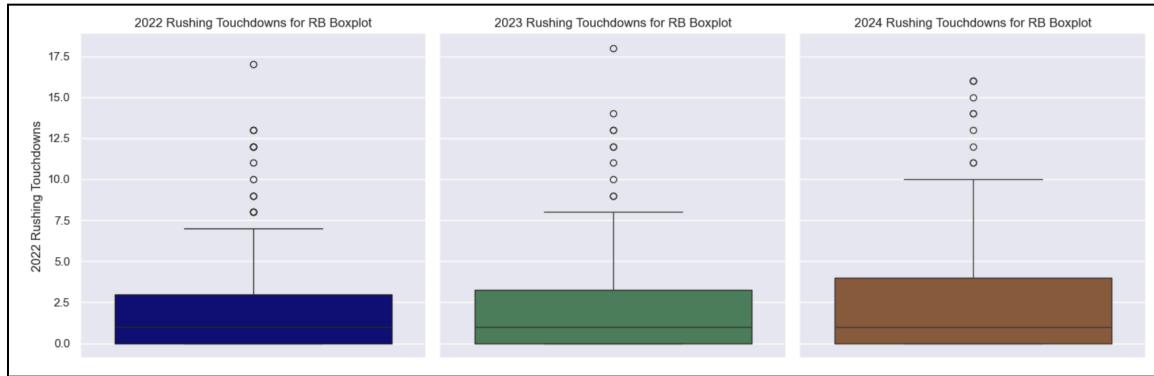


Another example of an elite outlier can be seen in the 2022 QB Total Passing boxplot. While there are several “elite” outliers to be seen in this plot, one extreme outlier can be seen with a value of well over 200000 Total Passing for the season. This outlier is Kansas City Chiefs quarterback Patrick Mahomes, who led his team to a 2022 Super Bowl victory and won the NFL MVP award for his efforts. His output this season represents the most recent NFL quarterback to throw for over 5000 yards in a season, and is another example of an outlier whose status should not be removed but prioritized for the model.



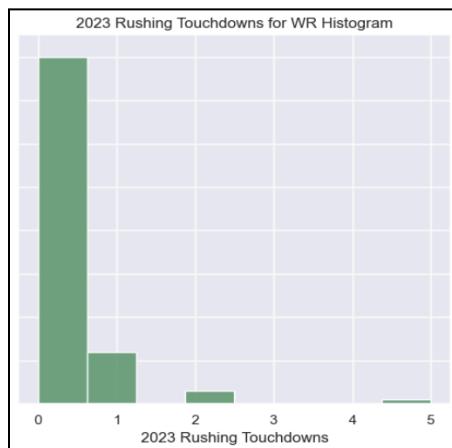
Running Backs (RBs)

The most interesting outliers for our RB positional stats can be seen in the 2022-2024 Rushing Touchdown boxplots below. Despite playing less games in their seasons than 2024 teams did, respectively, Detroit Lions RB Jamaal Williams's 17 rushing touchdowns in 2022 and Miami Dolphins RB Raheem Mostert's 18 rushing touchdowns in 2023 stand out as extreme outliers. Conventionally, elite running backs, especially ones who get to the endzone at a high level, are arguably the most sought after players in any fantasy football draft, so identifying them through analysis like this will be crucial for a successful recommendation system.



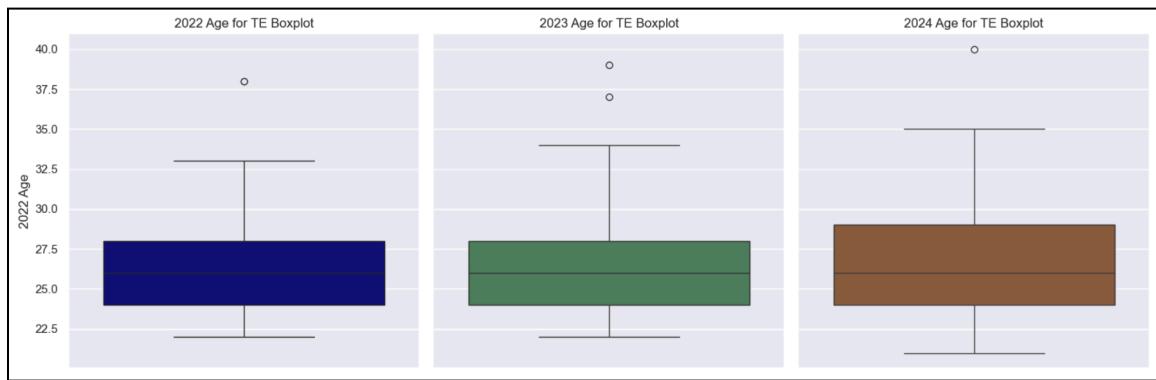
Wide Receivers (WRs)

One example of an outlier in stats among WRs that points to potentially increased fantasy value is Deebo Samuel on the San Francisco 49ers and his tendency to be used in special rushing plays near the endzone. Due to his agility and speed on the ball, a player like Deebo Samuel is trusted more in these described situations, as seen by his 5 rushing touchdowns in the 2023 season in the histogram below, a relative anomaly amongst his WR peers. This type of “flex” player, who is not only competent in more traditional receiving areas, but that also can be used in other unique game situations, represents an outlier that we want to prioritize due to their increased fantasy point output.



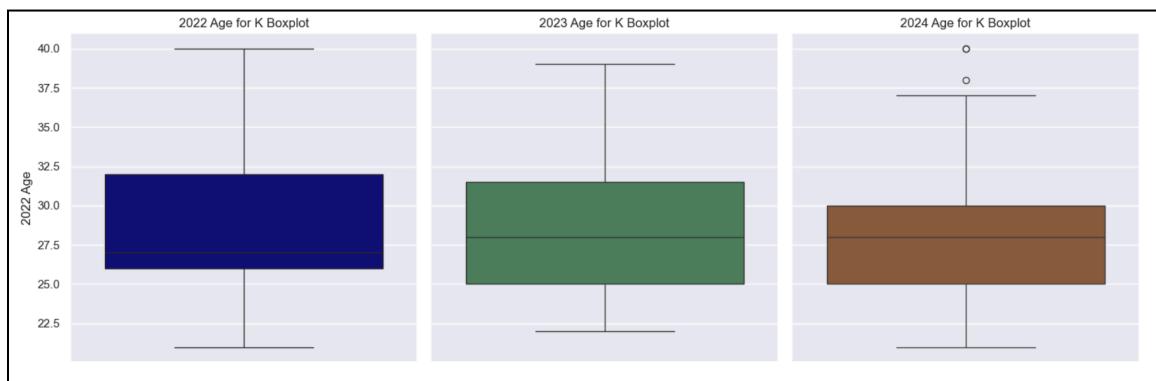
Tight Ends (TEs)

Looking again at our Age boxplots, this time with specifically tracking TEs, we can see two particular outliers across the 2022-2024 seasons. Tight end Marcedes Lewis, the player represented by the most extreme outlier point across each season, has played 19 seasons in the NFL. In 2022, Lewis played for the Green Bay Packers before signing a contract with the Chicago Bears for the 2023-2024 seasons. His participation in each season, as well as the fact that multiple teams were interested in having them on their roster well into his late 30s, is a remarkable testament to his consistency and durability. However, for the 2023 season, he was not the lone outlier, as TE Jimmy Graham signed a one-year contract with the New Orleans Saints at the age of 37 despite not playing since the 2021 season. While he did not continue to play into the 2024 season, it is fascinating to see this example of outliers who were sought out by multiple teams despite their relatively older ages. These types of players may also be valuable to a model as they exemplify players that teams are willing to get involved in their offensive schemes and scoring.



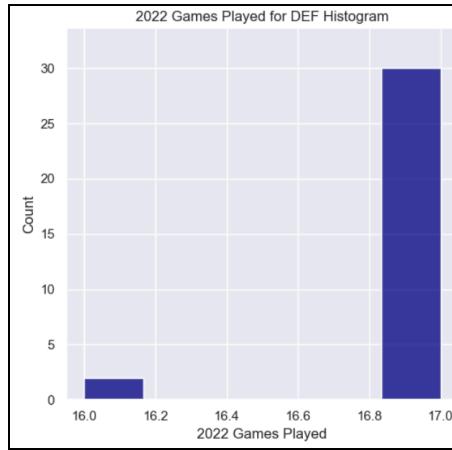
Kickers (Ks)

The boxplots below reinforce the previously explored insight that kickers tend to be able to potentially perform at a high level for longer periods of time. The kicker position group regularly have “normal” Age values in the mid to high 30s. This is expected due to the fact that kicking the ball enacts practically no physical toll on the body relative to other football position groups, leading to longer lasting careers. This unique trend in kickers is something we want the model to be able to pick up on, as Age values are not necessarily as indicative of success when compared to other draftable positions.



Defensive Units (DSTs)

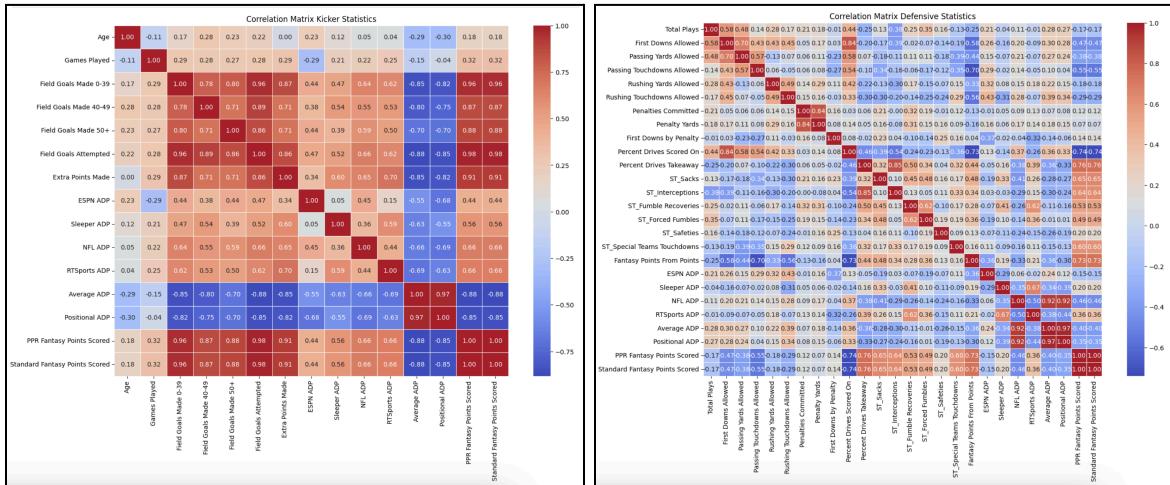
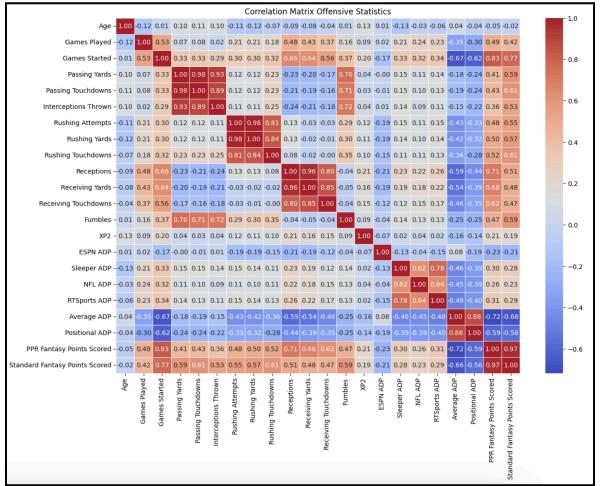
While not directly valuable to our model, one outlier we have detected for our analysis of stats amongst DSTs tells the story of an entirely unpredictable Games Played value instance. As seen in the histogram below, all but two teams played 17 games in the 2022 season. This outlier represents the abandonment of the January 3rd, 2023 game between the Buffalo Bills and Cincinnati Bengals due to the unexpected on-field cardiac arrest of Bills player Damar Hamlin. While our intended inclusion of outliers in our model is relatively unique, this outlier instance displays exactly why it is important to fully understand the context of the data you are working with. To some, this may have looked like a data error and would have led to potentially incorrectly altered data points. However, due to contextual knowledge of The General Managers team, we were able to identify and highlight this outlier for the incredibly unique scenario it represents.



Feature engineering and analysis

Our correlation matrix was created as a part of a broader effort to manage multicollinearity in our dataset. First, we ran Variance Inflation Factor (VIF) scores on three subsets – offensive stats, kicker stats, and defensive stats – to identify variables with excessive collinearity. To simplify this analysis, we focused exclusively on the 2023 season, as statistical relationships are expected to remain consistent across years. This approach provided a representative assessment of VIF and multicollinearity while reducing computational complexity.

After identifying problematic features, we dropped a plethora of variables that were severely affected, such as passing attempts, rushing attempts, field goals at different distances, and total yards allowed. Next, we examined the relationships of the remaining variables through a correlation matrix to better understand which variables were strongly interdependent.



From our initial correlation matrix analysis, we found that multiple statistics within each position group were highly correlated, leading to inflated VIF scores. This was especially apparent in:

- Quarterbacks (QBs): Strong correlations existed between passing yards, touchdowns, and interceptions thrown, which makes sense given that these stats naturally scale together. Players with high passing volume tend to accumulate more touchdowns and interceptions, leading to redundant predictive power.
- Running Backs (RBs): Rushing attempts, rushing yards, and rushing touchdowns were highly interdependent. Since rushing volume directly impacts total yardage and scoring opportunities, these variables showed excessive collinearity.
- Wide Receivers (WRs) & Tight Ends (TEs): Receptions, receiving yards, and receiving touchdowns were strongly correlated, reinforcing that opportunity (receptions) directly translates to production.
- Kickers (Ks): Field goals attempted and made (at different distances) showed near-perfect correlation, making individual breakdowns redundant.
- Defensive Units (DSTs): Penalties committed and penalty yards are strongly correlated and therefore a bit redundant.

Due to these strong relationships, many variables exceeded acceptable VIF thresholds. This prompted us to drop redundant features (e.g., specific ADP sources, certain breakdowns of field goals made/attempted) and create new combined variables where necessary.

To further reduce multicollinearity and reduce dimensionality (from an initial 172 features), we engineered combined features for each season (2022-2024):

- Passing Performance: We multiplied passing yards by passing touchdowns to create a single metric, emphasizing the importance of excelling in both areas
- Total Touchdowns Allowed: Passing touchdowns allowed and rushing touchdowns allowed were combined into a single defensive metric to reduce feature overlap
- Receiving Production: We multiplied receptions by receiving yards to consolidate receiving metrics while retaining the key information
- Rushing Performance: Rushing yards and rushing attempts were multiplied to create a unified rushing statistic

These transformations, along with the removal of highly collinear features, reduced the dataset's dimensions from 172x971 to 70x971.

After implementing these feature engineering steps, we re-evaluated the dataset using a final correlation matrix for the remaining numeric columns. This confirmed that our adjustments effectively mitigated excessive collinearity while preserving meaningful relationships between key variables. By refining the correlation structure, we ensured that the dataset was better suited for downstream modeling and analysis.

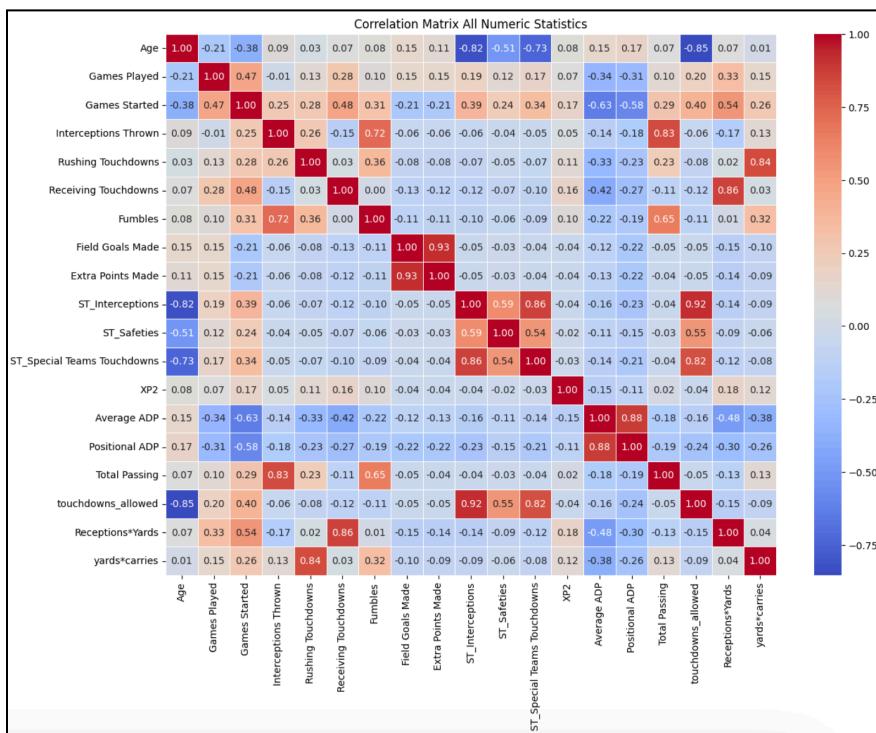


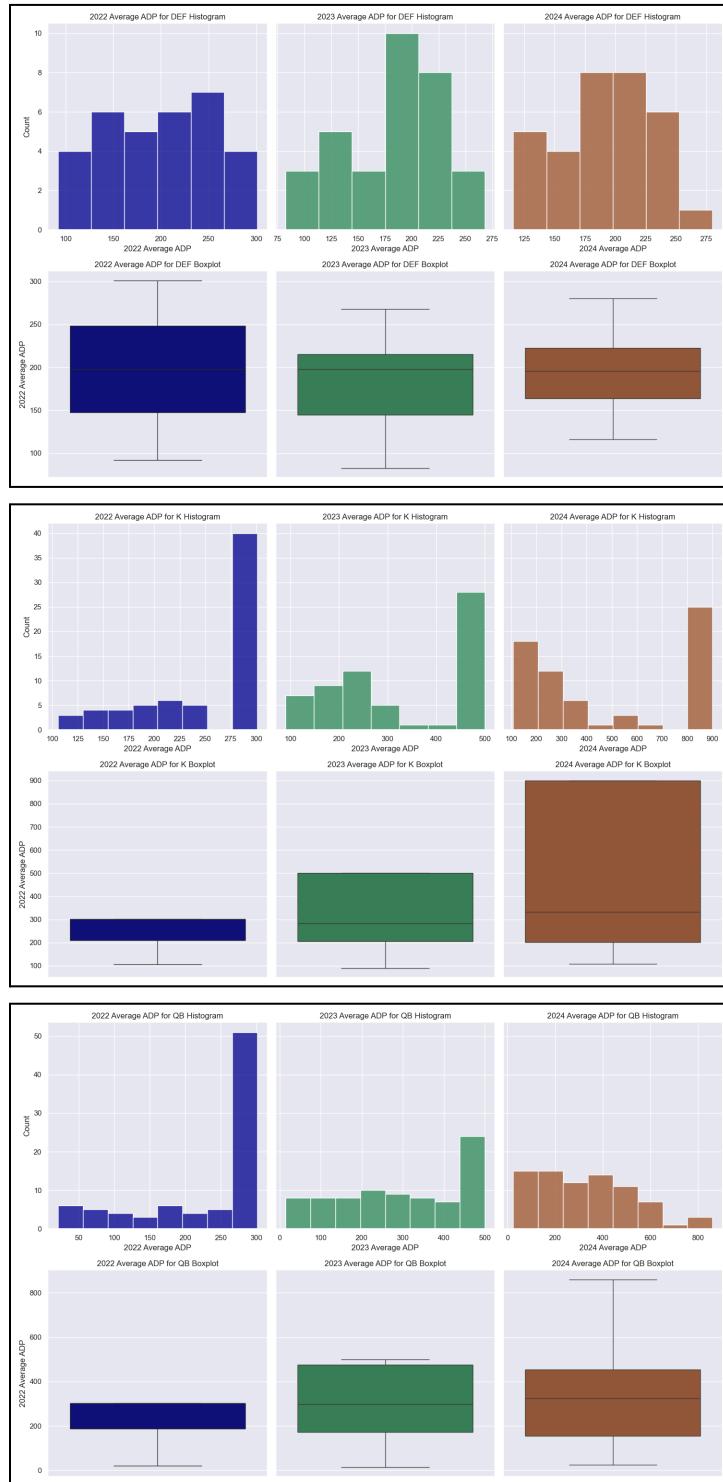
Table of Contributions

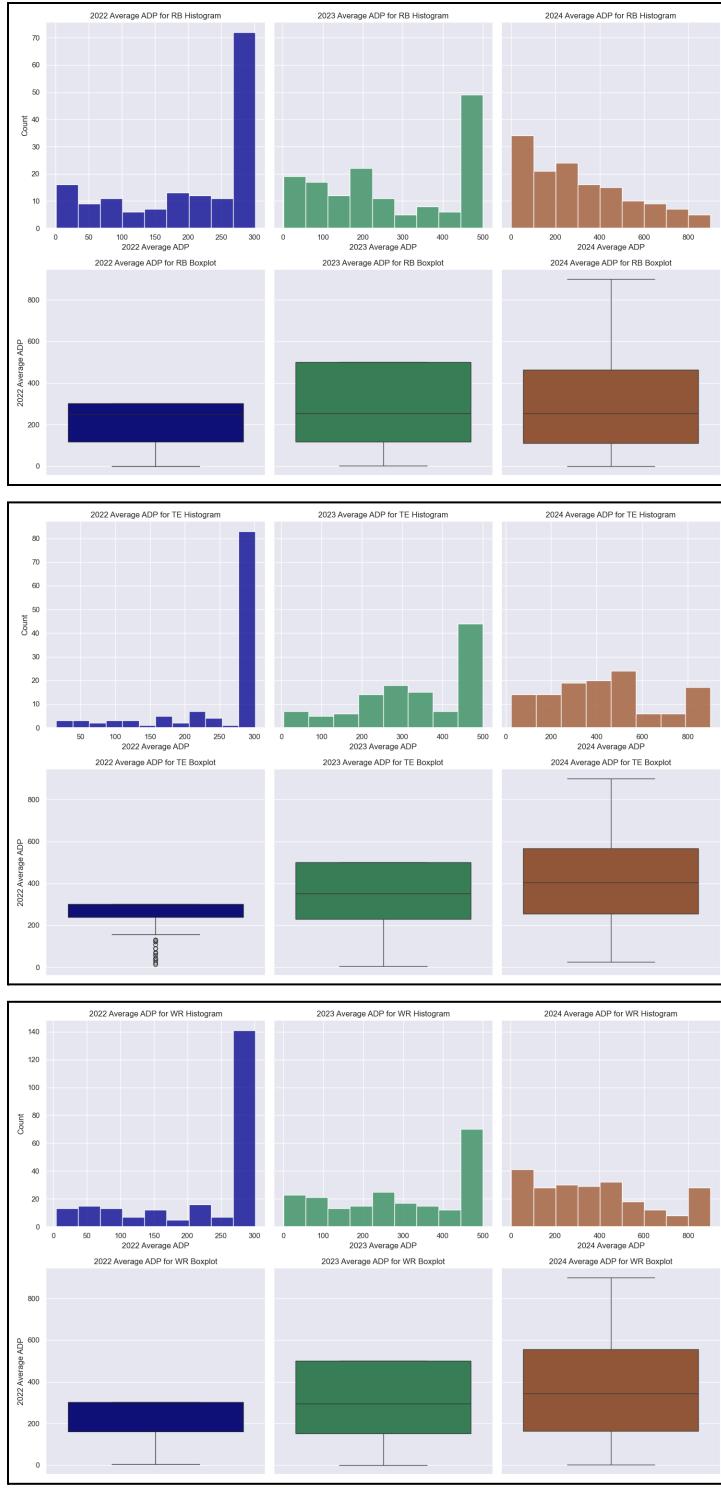
The table below identifies contributors to various sections of this document.

	Section	Writing	Editing
1	Analysis the basic metrics of variables	Hashim	Caleb, Hashim, David, Tommy
2	Non-graphical and graphical univariate analysis	David	Hashim, David, Tommy, Caleb
3	Missing value and outlier analysis	Thomas	David, Tommy, Caleb, Hashim
4	Feature engineering and analysis	Caleb	Tommy, Caleb, Hashim, David
5	Appendix	Hashim, David, Tommy, Caleb	Caleb, Tommy, David, Hashim

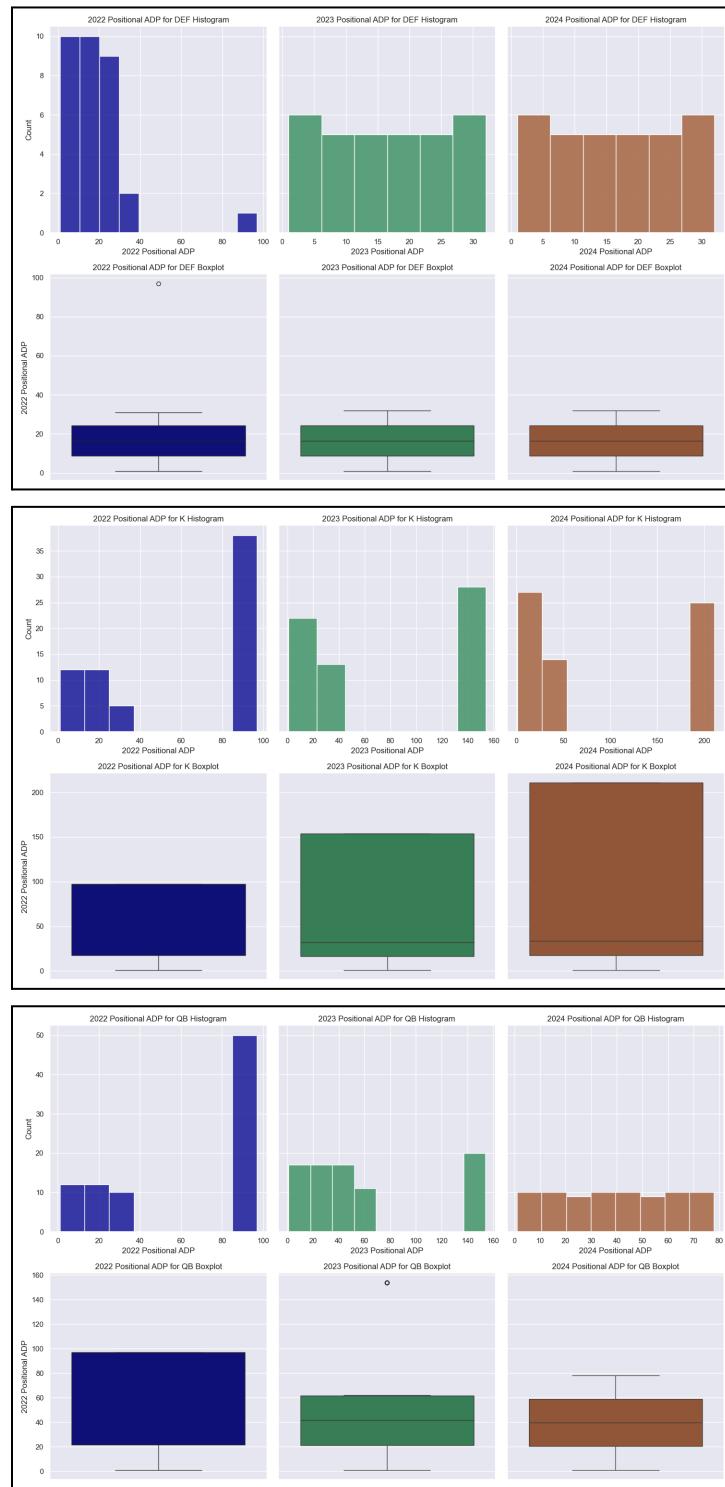
Appendix

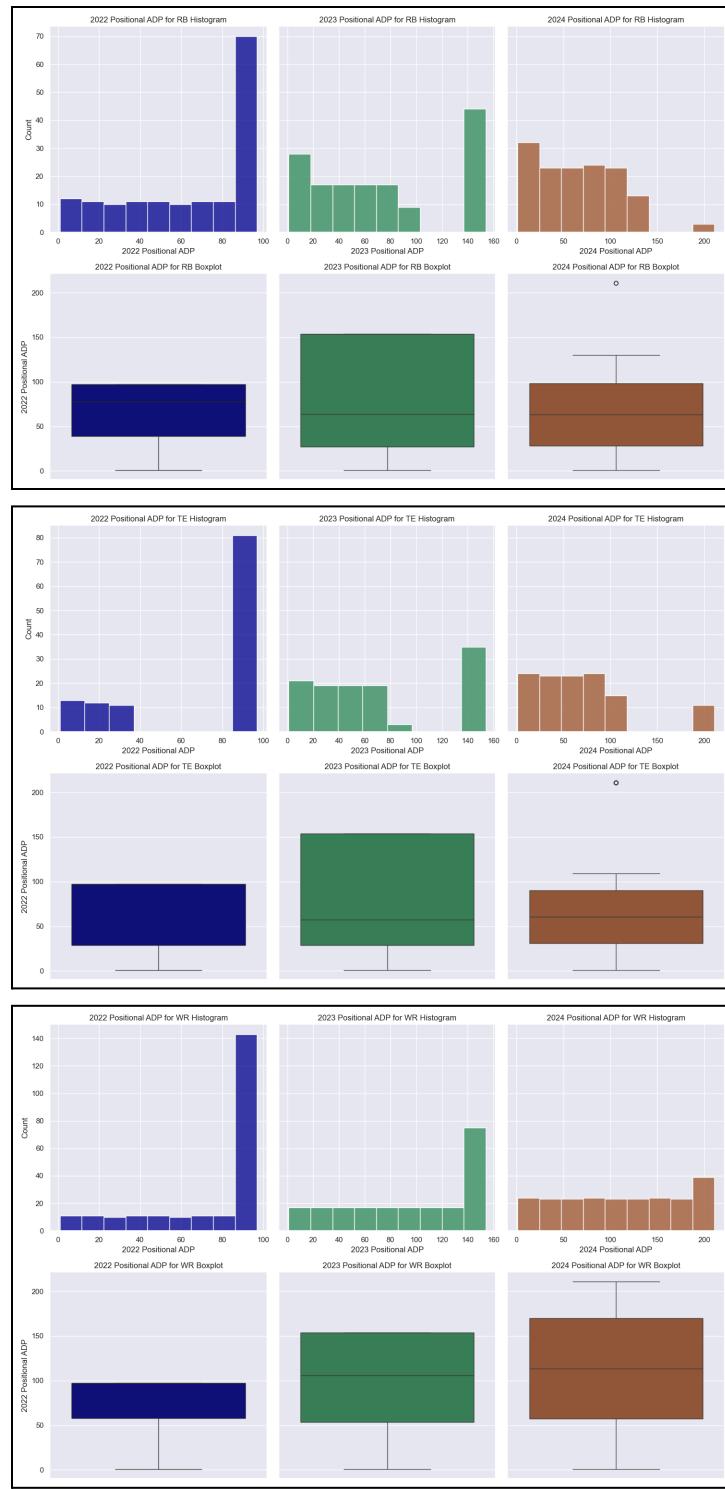
Average ADP by Position



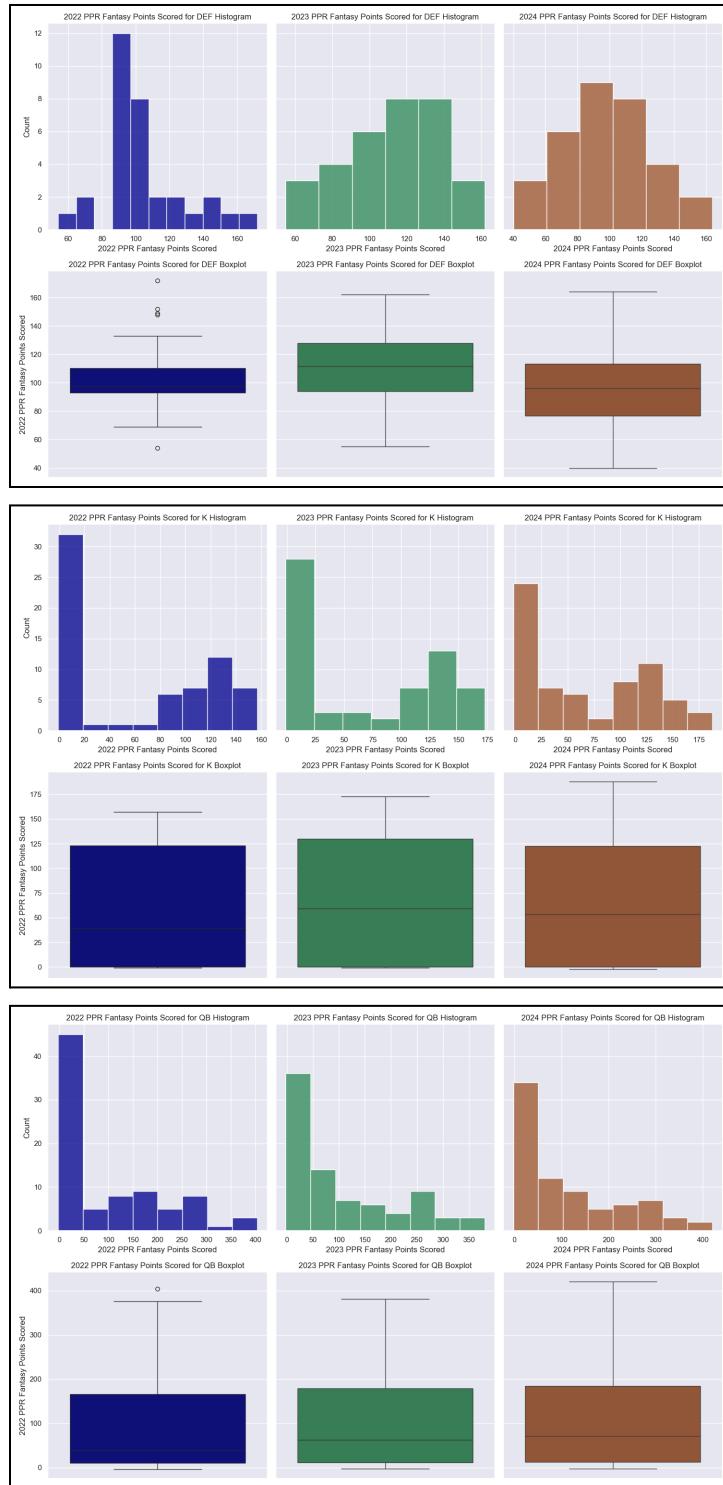


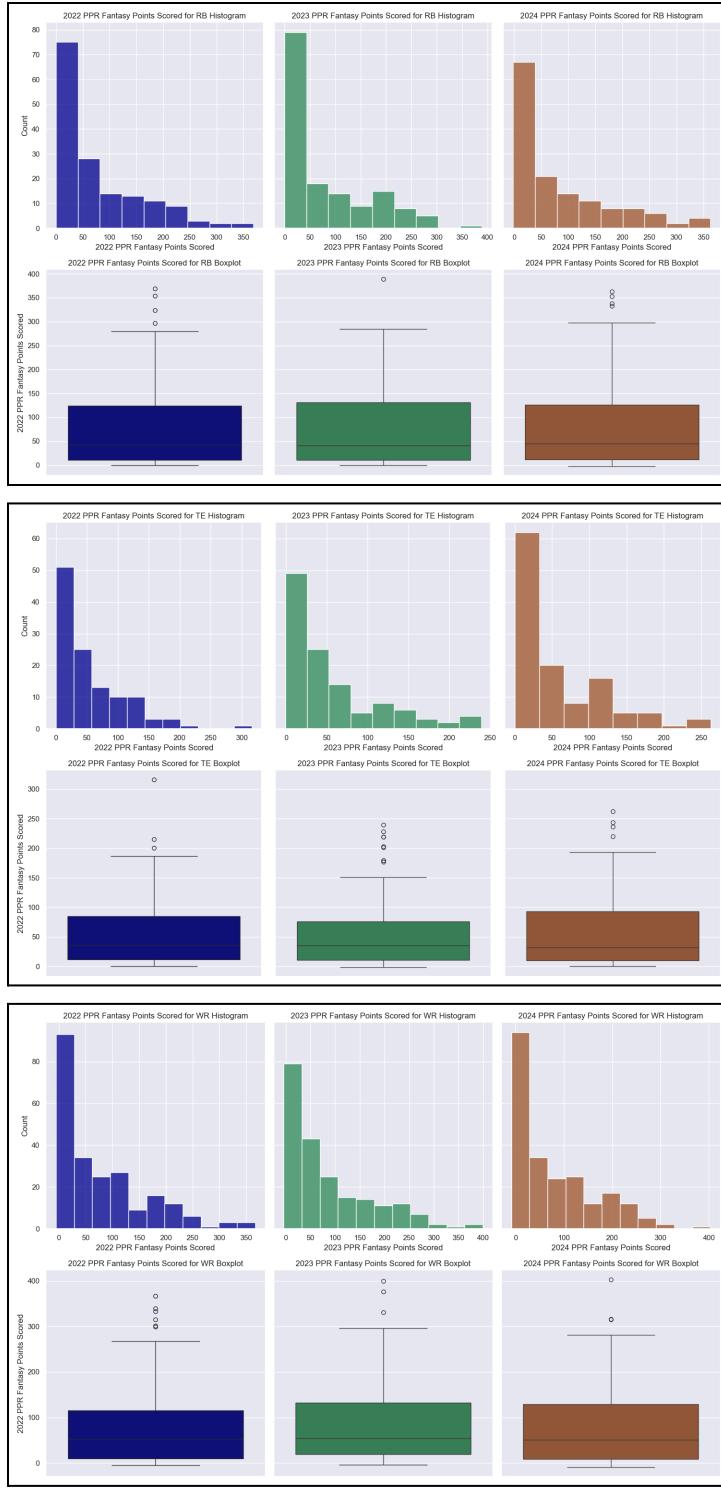
Positional ADP by Position





PPR Fantasy Points Scored by Position





Standard Fantasy Points Scored by Position

