NBA PLAYER STATISTICS: INSIGHTS ON GAME OUTCOMES

BY: DAVID ADEWOLE, KLAUDIA CZABANSKI, SAMUEL HANNAH, AND ANGELY LEE

AGENDA

01 OBJECTIVE

02 DATA COLLECTION

03 EDA

04 DATA EXPLORATION

05 FIRST ROUND OF MODELS

06 FIRST CONCLUSION

07 SECOND ROUND OF MODELS

08 FINAL CONCLUSION

OBJECTIVE

Create and evaluate predictive models to estimate the impact of player performance metrics on Wins

Share findings with key decision makers (owners/general managers) as they develop game strategies and select players

DATA COLLECTION

NBA Kaggle Set

Kaggle Set Link

Our targeted outcome variable

Table of Variables

Only 5 N/A observations so we filled it in by researching player position

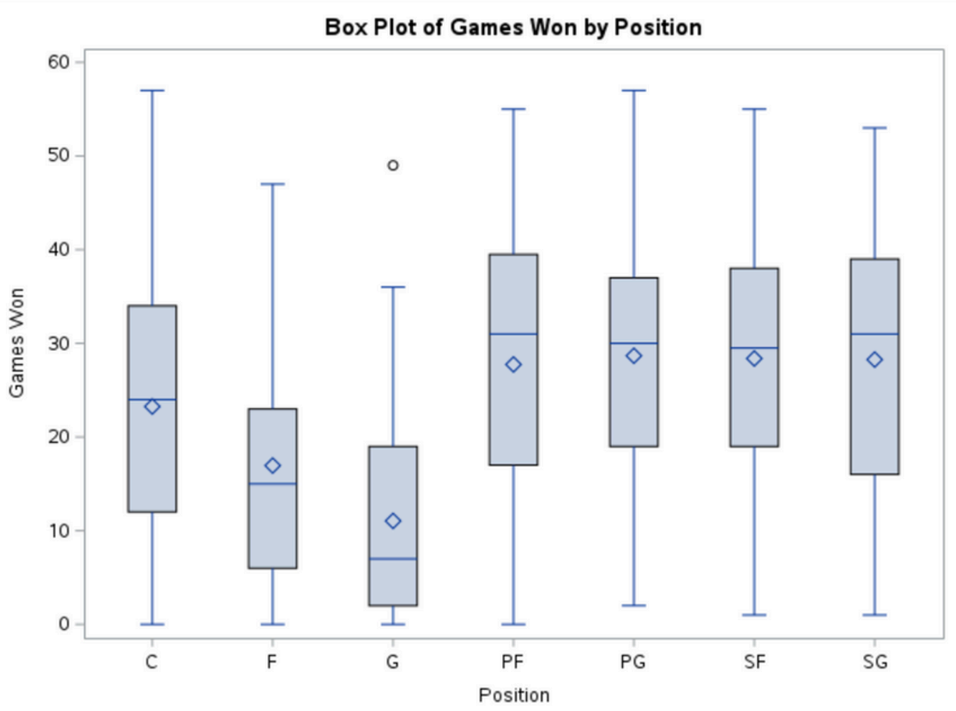
Description	Label	New Label
The name of the basketball player	PName	
The player's position in the game, including 'N/A'	POS	
The abbreviation of the team the player is currently playing for this season	Team	
The age of the player	Age	
The total number of games the player has played in this season	GP	
The total number of games won by the player	W	
The total number of games lost by the player	L	
The total minutes the player has played in this season	Min	
The total points made by the player	PTS	
The total number of field goals made by the player	FGM	
The total number of field goals attempted by the player	FGA	
The percentage of successful field goals made by the player	FG%	FGP
The total number of 3-point field goals made by the player	3PM	_3PM
The total number of 3-point field goals attempted by the player	3PA	_3PA
The percentage of successful 3-point field goals made by the player	3P%	_3PP
The total number of free throws made by the player	FTM	
The total number of free throws attempted by the player	FTA	
The percentage of successful free throws made by the player	FT%	FTP
The total number of offensive rebounds made by the player	OREB	
The total number of defensive rebounds made by the player	DREB	
The total number of rebounds (offensive + defensive) made by the player	REB	
The total number of assists made by the player	AST	
The total number of turnovers made by the player	TOV	
The total number of steals made by the player	STL	
The total number of blocks made by the player	BLK	
The total number of personal fouls made by the player	PF	
The total number of NBA fantasy points made by the player	FP	
The total number of double-doubles made by the player	DD2	
The total number of triple-doubles made by the player	TD3	
The total difference between the player's team scoring and the opponents' scoring while the player is in the game	+/-	Efficiency

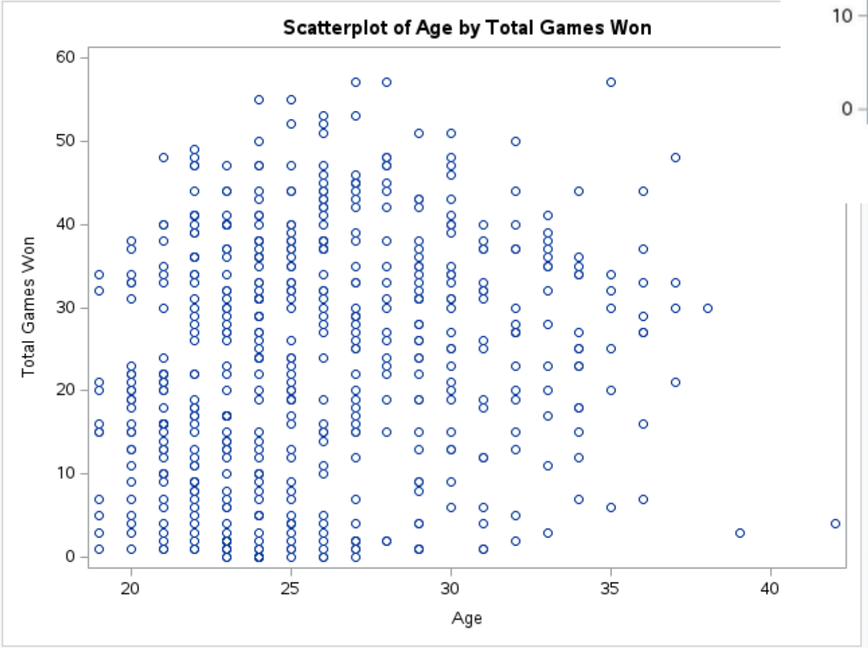
We modified certain variable names to enhance compatibility and efficiency when writing code in SAS.

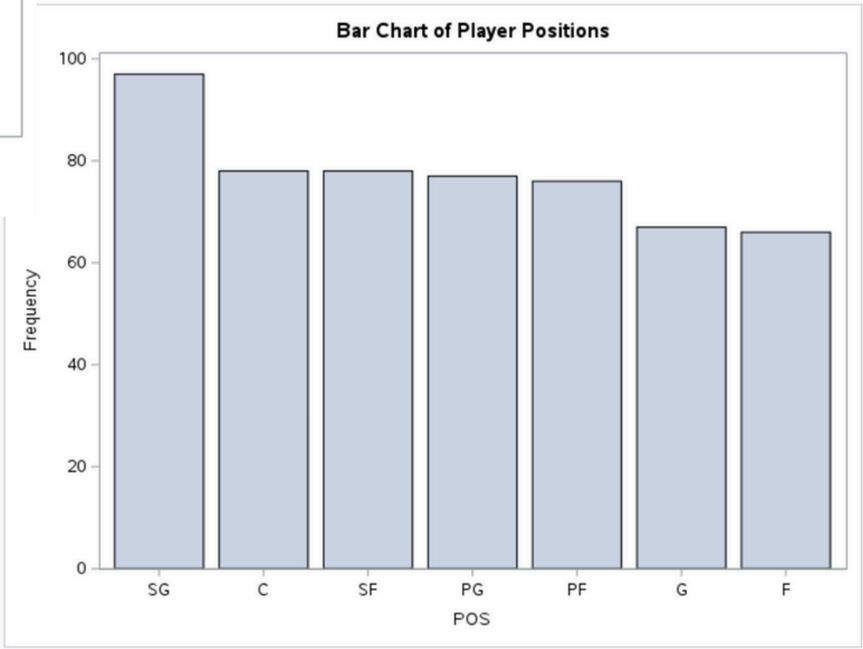
View of Dataset in Excel

PName	POS	Team	Age	GP	W	L	Min	PTS	FGM	FGA	FGP	_3PM	_3PA	_3PP	FTM	FTA	FTP	OREB	DREB	REB	AST	TOV	STL	BLK	PF	FP	DD2	TD3	Efficiency
Jayson Tatum	SF	BOS	25	74	52	22	2732.2	2225	727	1559	46.6	240	686	35	531	622	85.4	78	571	649	342	213	78	51	160	3691	31	1	470
Joel Embiid	C	PHI	29	66	43	23	2284.1	2183	728	1328	54.8	66	200	33	661	771	85.7	113	557	670	274	226	66	112	205	3706	39	1	424
Luka Doncic	PG	DAL	24	66	33	33	2390.5	2138	719	1449	49.6	185	541	34.2	515	694	74.2	54	515	569	529	236	90	33	166	3747	36	10	128
Shai Gilgeous-Alexander	PG	OKC	24	68	33	35	2416	2135	704	1381	51	58	168	34.5	669	739	90.5	59	270	329	371	192	112	65	192	3425	3	0	149
Giannis Antetokounmpo	PF	MIL	28	63	47	16	2023.6	1959	707	1278	55.3	47	171	27.5	498	772	64.5	137	605	742	359	246	52	51	197	3451	46	6	341
Anthony Edwards	SG	MIN	21	79	40	39	2841.5	1946	707	1541	45.9	213	578	36.9	319	422	75.6	47	411	458	350	259	125	58	186	3311	9	0	97
Julius Randle	PF	NYK	28	77	44	33	2737.3	1936	658	1432	45.9	218	636	34.3	402	531	75.7	141	626	767	316	216	49	21	233	3324	40	0	170
Donovan Mitchell	SG	CLE	26	68	44	24	2432.4	1922	679	1402	48.4	245	635	38.6	319	368	86.7	63	226	289	301	180	99	27	168	2918	5	0	338
Trae Young	PG	ATL	24	73	38	35	2540.7	1914	597	1390	42.9	154	460	33.5	566	639	88.6	56	161	217	741	300	80	9	104	3253	40	0	100
Zach LaVine	SG	CHI	28	77	38	39	2767.9	1913	673	1388	48.5	204	544	37.5	363	428	84.8	42	303	345	327	194	69	18	159	2885	2	0	18
Damian Lillard	PG	POR	32	58	27	31	2106.6	1866	556	1202	46.3	244	658	37.1	510	558	91.4	44	233	277	425	191	50	18	109	2849	16	2	105
De'Aaron Fox	PG	SAC	25	73	44	29	2435.2	1826	682	1331	51.2	119	367	32.4	343	440	78	40	266	306	447	181	83	23	172	3001	11	0	176
DeMar DeRozan	SF	CHI	33	74	37	37	2681.6	1816	657	1303	50.4	46	142	32.4	456	523	87.2	34	309	343	377	153	83	36	186	2997	6	0	67
Jaylen Brown	SG	BOS	26	67	46	21	2404.9	1784	679	1383	49.1	163	487	33.5	263	344	76.5	78	381	459	232	197	75	26	172	2789	13	0	256
Pascal Siakam	PF	TOR	29	71	35	36	2652	1720	630	1313	48	93	287	32.4	367	474	77.4	131	425	556	415	169	65	36	228	3144	24	2	123
Lauri Markkanen	PF	UTA	25	66	32	34	2272.5	1691	571	1144	49.9	200	510	39.2	349	399	87.5	130	440	570	123	127	42	38	137	2673	28	0	163
Nikola Jokic	C	DEN	28	69	48	21	2323	1690	646	1022	63.2	57	149	38.3	341	415	82.2	167	650	817	678	247	87	47	174	3842	58	29	640
Jalen Green	SG	HOU	21	76	20	56	2602.2	1683	566	1359	41.6	187	554	33.8	364	463	78.6	43	241	284	281	200	59	18	131	2476	0	0	-447
Jordan Poole	SG	GSW	23	82	44	38	2458.1	1675	550	1278	43	214	637	33.6	361	415	87	32	193	225	369	252	63	21	214	2499	1	0	-11
Mikal Bridges	SF	BKN	26	83	42	41	2963.2	1671	593	1267	46.8	169	442	38.2	316	353	89.5	79	285	364	273	127	91	61	159	2846	2	0	36
Stephen Curry	PG	GSW	35	56	30	26	1941.2	1648	559	1133	49.3	273	639	42.7	257	281	91.5	39	302	341	352	179	52	20	117	2622	12	1	268
Jalen Brunson	PG	NYK	26	68	40	28	2378.7	1633	587	1195	49.1	134	322	41.6	325	392	82.9	40	201	241	421	142	61	15	152	2640	5	0	152
Kyrie Irving	PG	DAL	31	60	32	28	2240.6	1623	594	1203	49.4	188	496	37.9	247	273	90.5	59	245	304	331	128	66	45	165	2689	5	0	222
Ja Morant	PG	MEM	23	61	40	21	1948.3	1596	566	1214	46.6	92	300	30.7	372	497	74.8	61	296	357	493	206	66	16	100	2804	20	7	302
LeBron James	SF	LAL	38	55	30	25	1953.9	1590	609	1219	50	121	377	32.1	251	327	76.8	65	392	457	375	178	50	32	88	2769	18	2	214
CJ McCollum	SG	NOP	31	75	38	37	2649.2	1568	587	1344	43.7	211	543	38.9	183	238	76.9	55	273	328	429	183	70	38	153	2746	6	0	59
Bam Adebayo	C	MIA	25	75	40	35	2598	1529	602	1114	54	1	12	8.3	324	402	80.6	184	504	688	240	187	88	61	208	2975	31	0	117
Dejounte Murray	SG	ATL	26	74	38	36	2693.4	1515	612	1319	46.4	133	387	34.4	158	190	83.2	53	336	389	448	160	112	19	106	2887	8	1	-62
Domantas Sabonis	PF	SAC	26	79	47	32	2735.6	1510	577	938	61.5	31	83	37.3	325	438	74.2	251	722	973	573	230	65	39	279	3619	65	14	212
Klay Thompson	SG	GSW	33	69	38	31	2278.9	1509	546	1252	43.6	301	731	41.2	116	132	87.9	39	247	286	163	123	49	29	130	2208	2	0	163
Kristaps Porzingis	C	WAS	27	65	27	38	2119.5	1505	507	1018	49.8	137	356	38.5	354	416	85.1	114	432	546	174	137	58	100	196	2758	20	0	80
Darius Garland	PG	CLE	23	69	40	29	2447.2	1490	522	1129	46.2	169	412	41	277	321	86.3	28	157	185	538	199	85	9	148	2602	18	0	438
Franz Wagner	SF	ORL	21	80	33	47	2608.7	1485	542	1118	48.5	130	360	36.1	271	322	84.2	70	259	329	283	167	77	17	183	2419	2	0	50
Devin Booker	SG	PHX	26	53	34	19	1835.4	1471	527	1067	49.4	111	316	35.1	306	358	85.5	46	194	240	293	145	51	18	159	2261	5	0	247
Jimmy Butler	SF	MIA	33	64	35	29	2138	1466	479	888	53.9	36	103	35	472	555	85	141	234	375	340	101	117	21	80	2739	11	0	47
Anthony Davis	PF	LAL	30	56	31	25	1904.5	1451	542	962	56.3	19	74	25.7	348	444	78.4	195	507	702	148	122	59	114	146	2912	40	0	205
Nikola Vucevic	C	CHI	32	82	40	42	2746.4	1447	597	1148	52	121	347	34.9	132	158	83.5	159	744	903	265	139	60	57	179	3140	51	1	-20
Paolo Banchero	PF	ORL	20	72	31	41	2429.7	1437	479	1122	42.7	85	285	29.8	394	534	73.8	84	413	497	269	200	60	39	160	2534	14	0	-215
RJ Barrett	SG	NYK	22	73	40	33	2475.5	1431	510	1176	43.4	121	390	31	290	392	74	60	308	368	201	164	31	15	179	2148	3	0	-62
Keldon Johnson	SF	SAS	23	63	17	46	2063.3	1385	503	1113	45.2	134	407	32.9	245	327	74.9	56	262	318	183	134	46	11	125	2078	5	0	-491
Spencer Dinwiddie	PG	BKN	30	79	41	38	2725.3	1369	460	1050	43.8	181	490	36.9	268	330	81.2	28	242	270	515	145	67	24	187	2594	9	0	-8
Kevin Durant	PF	PHX	34	47	34	13	1671.8	1366	483	862	56	93	230	40.4	307	334	91.9	17	296	313	235	156	34	67	99	2241	5	1	258
Kyle Kuzma	SF	WAS	27	64	29	35	2238.9	1357	512	1142	44.8	160	481	33.3	173	237	73	55	408	463	239	190	36	29	145	2276	14	1	10
Tyler Herro	PG	MIA	23	67	37	30	2335.4	1347	487	1109	43.9	203	537	37.8	170	182	93.4	27	333	360	280	158	51	16	101	2242	6	1	170
Buddy Hield	SG	IND	30	80	35	45	2481.9	1344	475	1038	45.8	288	677	42.5	106	129	82.2	67	333	400	225	139	92	26	156	2377	3	0	-84
Fred VanVleet	SG	TOR	29	69	34	35	2535	1335	437	1112	39.3	207	606	34.2	254	283	89.8	30	250	280	495	140	123	38	193	2757	10	1	195
Paul George	F	LAC	33	56	32	24	1938.9	1332	459	1004	45.7	158	426	37.1	256	294	87.1	45	297	342	288	176	83	20	159	2307	8	1	157
Terry Rozier	SG	CHA	29	63	19	44	2220.9	1329	493	1188	41.5	165	505	32.7	178	220	80.9	51	208	259	319	134	74	16	119	2254	1	0	-288

DATA EXPLORATION







Summary of Wins by Team

Analysis Variable : W W						
Team	N Obs	N	Mean	Median	Minimum	Maximum
ATL	18	18	24.722	29.500	1.000	41.000
BKN	20	20	24.750	27.500	0.000	42.000
BOS	18	18	33.778	37.500	3.000	57.000
CHA	17	17	14.294	16.000	0.000	24.000
CHI	17	17	24.765	32.000	3.000	40.000
CLE	17	17	29.471	31.000	3.000	48.000
DAL	21	21	19.857	20.000	1.000	37.000
DEN	16	16	36.125	37.000	13.000	53.000
DET	19	19	8.895	10.000	0.000	16.000
GSW	17	17	27.118	30.000	3.000	44.000
HOU	15	15	13.733	16.000	0.000	22.000
IND	18	18	21.889	27.500	1.000	36.000
LAC	18	18	26.278	29.500	2.000	44.000
LAL	18	18	26.722	30.500	1.000	45.000
MEM	18	18	32.000	37.500	0.000	51.000
MIA	17	17	24.353	28.000	4.000	42.000
MIL	18	18	33.056	33.000	4.000	57.000
MIN	18	18	25.222	27.000	7.000	41.000
NOP	16	16	26.438	27.000	2.000	40.000
NYK	16	16	31.000	36.500	3.000	47.000
OKC	16	16	24.875	26.000	3.000	37.000
ORL	17	17	18.765	20.000	1.000	33.000
PHI	18	18	32.278	38.500	1.000	52.000
PHX	16	16	30.813	33.000	15.000	44.000
POR	21	21	16.905	15.000	0.000	38.000
SAC	20	20	27.950	30.000	1.000	48.000
SAS	21	21	12.190	12.000	1.000	31.000
TOR	18	18	24.667	28.000	4.000	39.000
UTA	20	20	17.250	18.000	0.000	34.000
WAS	20	20	17.300	21.500	0.000	33.000

Summary of Games Played by Team

Analysis Variable : GP GP						
Team	N Obs	N	Mean	Median	Minimum	Maximum
ATL	18	18	52.167	64.000	2.000	80.000
BKN	20	20	47.400	50.500	1.000	83.000
BOS	18	18	49.778	62.000	4.000	82.000
CHA	17	17	42.294	46.000	4.000	73.000
CHI	17	17	50.941	67.000	5.000	82.000
CLE	17	17	46.353	48.000	3.000	79.000
DAL	21	21	41.952	45.000	1.000	78.000
DEN	16	16	57.125	61.000	17.000	80.000
DET	19	19	41.737	47.000	1.000	76.000
GSW	17	17	49.529	57.000	4.000	82.000
HOU	15	15	52.467	59.000	4.000	82.000
IND	18	18	49.444	62.000	3.000	80.000
LAC	18	18	52.167	56.000	4.000	81.000
LAL	18	18	49.389	56.000	4.000	81.000
MEM	18	18	51.111	58.500	1.000	80.000
MIA	17	17	45.588	54.000	7.000	80.000
MIL	18	18	47.500	52.000	7.000	81.000
MIN	18	18	50.167	53.000	15.000	79.000
NOP	16	16	52.500	60.000	5.000	79.000
NYK	16	16	56.250	65.500	3.000	82.000
OKC	16	16	52.125	55.000	6.000	76.000
ORL	17	17	44.765	51.000	2.000	80.000
PHI	18	18	51.167	59.000	1.000	80.000
PHX	16	16	56.375	59.500	25.000	79.000
POR	21	21	41.857	52.000	1.000	80.000
SAC	20	20	48.100	53.000	2.000	82.000
SAS	21	21	40.905	38.000	4.000	73.000
TOR	18	18	49.889	55.500	8.000	77.000
UTA	20	20	38.100	44.500	1.000	74.000
WAS	20	20	41.600	50.000	2.000	78.000

Summary of Wins by Player Position

The MEANS Procedure

Analysis Variable : W W						
POS	N Obs	N	Mean	Median	Minimum	Maximum
C	78	78	23.256	24.000	0.000	57.000
F	66	66	16.955	15.000	0.000	47.000
G	67	67	11.045	7.000	0.000	49.000
PF	76	76	27.750	31.000	0.000	55.000
PG	77	77	28.675	30.000	2.000	57.000
SF	78	78	28.385	29.500	1.000	55.000
SG	97	97	28.268	31.000	1.000	53.000

Summary of Games Played by Player Position

The MEANS Procedure

Analysis Variable : GP GP						
POS	N Obs	N	Mean	Median	Minimum	Maximum
C	78	78	47.462	51.000	1.000	82.000
F	66	66	35.273	34.500	1.000	82.000
G	67	67	25.015	16.000	1.000	76.000
PF	76	76	54.487	62.000	1.000	82.000
PG	77	77	55.338	60.000	4.000	82.000
SF	78	78	56.462	63.000	1.000	83.000
SG	97	97	55.485	61.000	1.000	82.000

Pearson Correlation Coefficients, N = 539
Prob > |r| under H0: Rho=0

	Min	FGM	FGA	_3PM	_3PA	FTM	FTA	OREB	DREB	AST	TOV	STL	BLK	PF	FP	DD2	TD3	Efficiency
Min Min	1.00000	0.91550 <.0001	0.91899 <.0001	0.76417 <.0001	0.78370 <.0001	0.74929 <.0001	0.75713 <.0001	0.58748 <.0001	0.83596 <.0001	0.77136 <.0001	0.85819 <.0001	0.86923 <.0001	0.54583 <.0001	0.89119 <.0001	0.94404 <.0001	0.50694 <.0001	0.18692 <.0001	0.22646 <.0001
FGM FGM	0.91550 <.0001	1.00000	0.98560 <.0001	0.73479 <.0001	0.75702 <.0001	0.88548 <.0001	0.88950 <.0001	0.53117 <.0001	0.82489 <.0001	0.79034 <.0001	0.91839 <.0001	0.77981 <.0001	0.51148 <.0001	0.77693 <.0001	0.97451 <.0001	0.61454 <.0001	0.27661 <.0001	0.27403 <.0001
FGA FGA	0.91899 <.0001	0.98560 <.0001	1.00000	0.80744 <.0001	0.83502 <.0001	0.86784 <.0001	0.86385 <.0001	0.44125 <.0001	0.76928 <.0001	0.80618 <.0001	0.91710 <.0001	0.79238 <.0001	0.43232 <.0001	0.75570 <.0001	0.95277 <.0001	0.53745 <.0001	0.23390 <.0001	0.23642 <.0001
_3PM _3PM	0.76417 <.0001	0.73479 <.0001	0.80744 <.0001	1.00000	0.99145 <.0001	0.56535 <.0001	0.53231 <.0001	0.09979 0.0205	0.47558 <.0001	0.62109 <.0001	0.65393 <.0001	0.65309 <.0001	0.16792 <.0001	0.55541 <.0001	0.69220 <.0001	0.18594 <.0001	0.07910 0.0665	0.23341 <.0001
_3PA _3PA	0.78370 <.0001	0.75702 <.0001	0.83502 <.0001	0.99145 <.0001	1.00000	0.59563 <.0001	0.56615 <.0001	0.11611 0.0070	0.49800 <.0001	0.64475 <.0001	0.68765 <.0001	0.67081 <.0001	0.17606 <.0001	0.57896 <.0001	0.71469 <.0001	0.20734 <.0001	0.08782 0.0415	0.20269 <.0001
FTM FTM	0.74929 <.0001	0.88548 <.0001	0.86784 <.0001	0.56535 <.0001	0.59563 <.0001	1.00000	0.99135 <.0001	0.42733 <.0001	0.69330 <.0001	0.72284 <.0001	0.85189 <.0001	0.64117 <.0001	0.41685 <.0001	0.62712 <.0001	0.86735 <.0001	0.59872 <.0001	0.29274 <.0001	0.27463 <.0001
FTA FTA	0.75713 <.0001	0.88950 <.0001	0.86385 <.0001	0.53231 <.0001	0.56615 <.0001	0.99135 <.0001	1.00000	0.48943 <.0001	0.73276 <.0001	0.71557 <.0001	0.86056 <.0001	0.64227 <.0001	0.45990 <.0001	0.65577 <.0001	0.87909 <.0001	0.63761 <.0001	0.30942 <.0001	0.26435 <.0001
OREB OREB	0.58748 <.0001	0.53117 <.0001	0.44125 <.0001	0.09979 0.0205	0.11611 0.0070	0.42733 <.0001	0.48943 <.0001	1.00000	0.80197 <.0001	0.29901 <.0001	0.48239 <.0001	0.46724 <.0001	0.73304 <.0001	0.71499 <.0001	0.63024 <.0001	0.67218 <.0001	0.18984 <.0001	0.13434 0.0018
DREB DREB	0.83596 <.0001	0.82489 <.0001	0.76928 <.0001	0.47558 <.0001	0.49800 <.0001	0.69330 <.0001	0.73276 <.0001	0.80197 <.0001	1.00000	0.60706 <.0001	0.77588 <.0001	0.66464 <.0001	0.69040 <.0001	0.83612 <.0001	0.88991 <.0001	0.78930 <.0001	0.33170 <.0001	0.25213 <.0001
AST AST	0.77136 <.0001	0.79034 <.0001	0.80618 <.0001	0.62109 <.0001	0.64475 <.0001	0.72284 <.0001	0.71557 <.0001	0.29901 <.0001	0.60706 <.0001	1.00000	0.88211 <.0001	0.77109 <.0001	0.25233 <.0001	0.60767 <.0001	0.83154 <.0001	0.54595 <.0001	0.41084 <.0001	0.27829 <.0001
TOV TOV	0.85819 <.0001	0.91839 <.0001	0.91710 <.0001	0.65393 <.0001	0.68765 <.0001	0.85189 <.0001	0.86056 <.0001	0.48239 <.0001	0.77588 <.0001	0.88211 <.0001	1.00000	0.76750 <.0001	0.43599 <.0001	0.76273 <.0001	0.92676 <.0001	0.62797 <.0001	0.34395 <.0001	0.20525 <.0001
STL STL	0.86923 <.0001	0.77981 <.0001	0.79238 <.0001	0.65309 <.0001	0.67081 <.0001	0.64117 <.0001	0.64227 <.0001	0.46724 <.0001	0.66464 <.0001	0.77109 <.0001	0.76750 <.0001	1.00000	0.43223 <.0001	0.77149 <.0001	0.83774 <.0001	0.38224 <.0001	0.19001 <.0001	0.24207 <.0001
BLK BLK	0.54583 <.0001	0.51148 <.0001	0.43232 <.0001	0.16792 <.0001	0.17606 <.0001	0.41685 <.0001	0.45990 <.0001	0.73304 <.0001	0.69040 <.0001	0.25233 <.0001	0.43599 <.0001	0.43223 <.0001	1.00000	0.66590 <.0001	0.59986 <.0001	0.50887 <.0001	0.08184 0.0576	0.20687 <.0001
PF PF	0.89119 <.0001	0.77693 <.0001	0.75570 <.0001	0.55541 <.0001	0.57896 <.0001	0.62712 <.0001	0.65577 <.0001	0.71499 <.0001	0.83612 <.0001	0.60767 <.0001	0.76273 <.0001	0.77149 <.0001	0.66590 <.0001	1.00000	0.84107 <.0001	0.50241 <.0001	0.16008 0.0002	0.15570 0.0003
FP FP	0.94404 <.0001	0.97451 <.0001	0.95277 <.0001	0.69220 <.0001	0.71469 <.0001	0.86735 <.0001	0.87909 <.0001	0.63024 <.0001	0.88991 <.0001	0.83154 <.0001	0.92676 <.0001	0.83774 <.0001	0.59986 <.0001	0.84107 <.0001	1.00000	0.68051 <.0001	0.31699 <.0001	0.30213 <.0001
DD2 DD2	0.50694 <.0001	0.61454 <.0001	0.53745 <.0001	0.18594 <.0001	0.20734 <.0001	0.59872 <.0001	0.63761 <.0001	0.67218 <.0001	0.78930 <.0001	0.54595 <.0001	0.62797 <.0001	0.38224 <.0001	0.50887 <.0001	0.50241 <.0001	0.68051 <.0001	1.00000	0.51498 <.0001	0.25847 <.0001
TD3 TD3	0.18692 <.0001	0.27661 <.0001	0.23390 <.0001	0.07910 0.0665	0.08782 0.0415	0.29274 <.0001	0.30942 <.0001	0.18984 <.0001	0.33170 <.0001	0.41084 <.0001	0.34395 <.0001	0.19001 <.0001	0.08184 0.0576	0.16008 0.0002	0.31699 <.0001	0.51498 <.0001	1.00000	0.23879 <.0001
Efficiency Efficiency	0.22646 <.0001	0.27403 <.0001	0.23642 <.0001	0.23341 <.0001	0.20269 <.0001	0.27463 <.0001	0.26435 <.0001	0.13434 0.0018	0.25213 <.0001	0.27829 <.0001	0.20525 <.0001	0.24207 <.0001	0.20687 <.0001	0.15570 0.0003	0.30213 <.0001	0.25847 <.0001	0.23879 <.0001	1.00000

VARIABLES

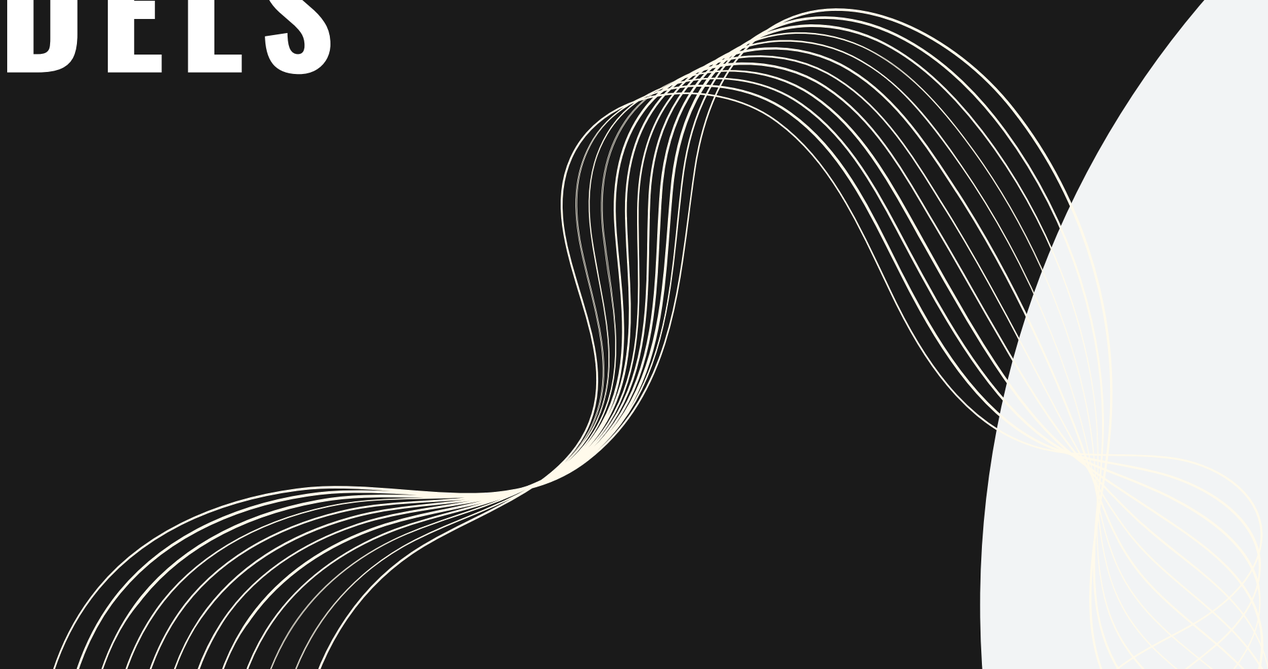
Variables in first model
W
POS
AGE
GP
MIN
PTS
FGM
FGA
FGP
_3PM
_3PA
_3PP
FTM
FTA
FTP
OREB
DREB
REB
AST
TOV
STL
BLK
PF
FP
DD2
TD3
EFFICIENCY

Excluded teams and losses:

- The Teams variable itself, we thought, does not directly provide predictive information for number of games won by the player.
- Losses is often inversely related to Wins. Including "Losses" in the model introduces redundancy and potentially multicollinearity

FIRST ROUND OF MODELS

Linear Regression
CART
Neural Network



FIRST MULTIPLE LINEAR REGRESSION

Multiple linear regression was used to predict player wins based on key statistics and performance metrics. The model demonstrated strong explanatory power, identified significant predictors, and highlighted multicollinearity issues, providing actionable insights for evaluation and improvement.

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	NBA
Random Number Seed	12345
Sampling Rate	0.8
Sample Size	432
Selection Probability	0.801484
Sampling Weight	0
Output Data Set	NBA_PART

The data was partitioned 80% allocated to the training dataset (432 observations) and 20% for validation. Simple random sampling ensured an unbiased split.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	81374	27125	1275.15	<.0001
Error	428	9104.23910	21.27159		
Corrected Total	431	90478			

Root MSE	4.61211
R-Square	0.89938
Adj R-Sq	0.89867
AIC	1758.76617
AICC	1758.90701
SBC	1341.03987
ASE (Train)	21.07463
ASE (Validate)	16.66570

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-0.541125	0.510504	-1.06	0.2898	0
GP	1	0.529854	0.010929	48.48	<.0001	1.41375
FTA	1	-0.008181	0.002037	-4.02	<.0001	1.50947
Efficiency	1	0.033663	0.001494	22.53	<.0001	1.08318

Selection Information

Selection Method	Stepwise

Approximately 89% of the variability in player wins is reflected in the model. Predictors such as GP, FTA and Eficiency are statistically significant, indicating that they could be important variables.

MULTIPLE LINEAR REGRESSION - DIFF METRICS

Stepwise:

Parameter Estimates						
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	8.953047	0.677662	13.21	<.0001	0
_3PM	1	0.065608	0.009337	7.03	<.0001	2.30466
OREB	1	0.032832	0.011864	2.77	0.0059	2.88244
PF	1	0.120617	0.011283	10.71	<.0001	3.96652
FTM	1	-0.017465	0.004697	-3.72	0.0002	1.87092
Efficiency	1	0.028046	0.002558	10.96	<.0001	1.11941
POS_SF	1	3.039488	1.084356	2.80	0.0053	1.02080

Backward:

Parameter Estimates						
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	9.012977	0.673159	13.39	<.0001	0
_3PA	1	0.014710	0.002915	5.05	<.0001	1.54662
DREB	1	0.033348	0.007572	4.40	<.0001	8.08738
PF	1	0.102712	0.011585	8.87	<.0001	4.17868
DD2	1	-0.377459	0.075778	-4.98	<.0001	3.49747
Efficiency	1	0.028833	0.002559	11.19	<.0001	1.11920

Stepwise:

The MEANS Procedure

Variable	N	Mean
mape_fit	423	73.830
mape_acc	104	107.247
mae_fit	432	6.231
mae_acc	107	6.875
mse_fit	432	59.365
mse_acc	107	68.723

Backward:

The MEANS Procedure

Variable	N	Mean
mape_fit	423	72.519
mape_acc	104	106.122
mae_fit	432	6.253
mae_acc	107	6.539
mse_fit	432	59.553
mse_acc	107	62.706

- 3PM _3PA OREB DREB AST TOV STL BLK PF FTM FTA DD2 TD3 EFFICIENCY POS_PG POS_SG POS_SF POS_PF POS_F POS_G

FIRST MULTIPLE LINEAR REGRESSION - EVALUATION

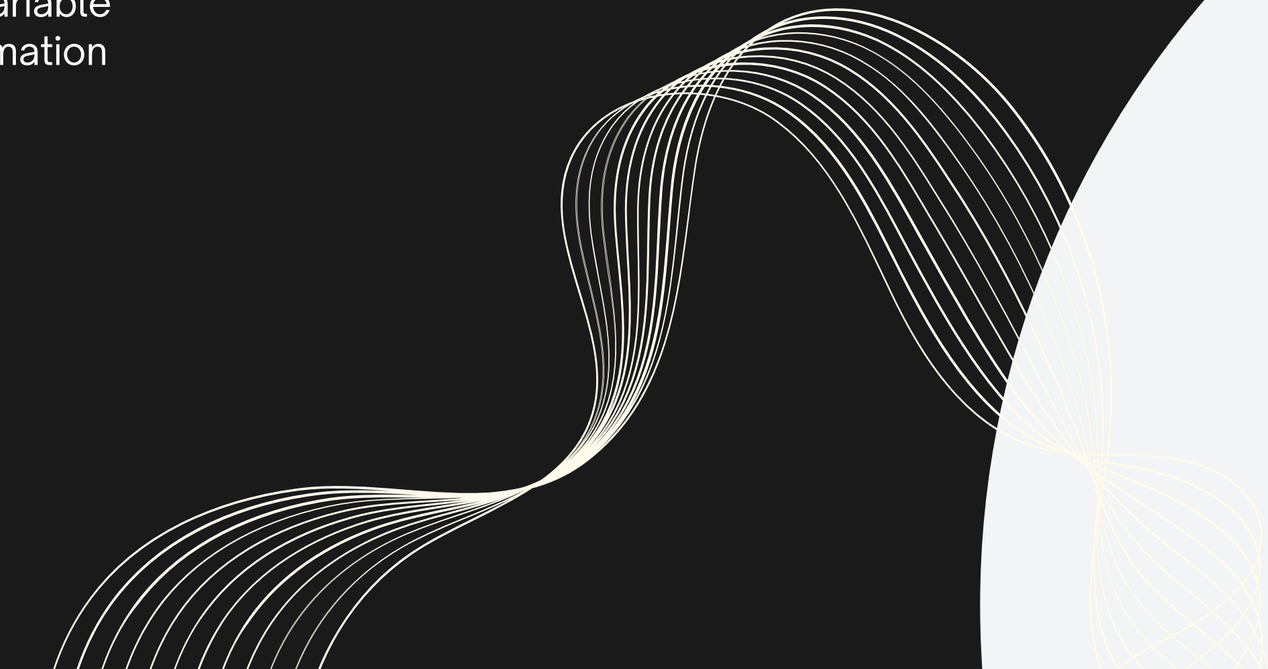
- No clear signs of overfitting were observed, as the performance between the training (MAPE = 23.45, MSE = 21.08) and validation (MAPE = 26.11, MSE = 16.67) datasets was within reasonable limits.
- Low error measures, including MAE values (3.43 training, 3.00 validation), indicate relatively high accuracy in predicting player wins.
- The results suggest that the model generalizes well to unseen data, making it a reliable tool for evaluating player performance and predicting outcomes.

The MEANS Procedure

Variable	N	Mean
mape_fit	423	23.451
mape_acc	104	26.105
mae_fit	432	3.432
mae_acc	107	2.998
mse_fit	432	21.075
mse_acc	107	16.666

CART

- Numerical Outcome Variable
- Minimal Data Transformation
- Easy Interpretation
- Large Dataset



FIRST CART MODEL - OUTPUT

Model Information	
Split Criterion Used	Variance
Pruning Method	Cost-Complexity
Subtree Evaluation Criterion	Cost-Complexity
Number of Branches	2
Maximum Tree Depth Requested	10
Maximum Tree Depth Achieved	10
Tree Depth	8
Number of Leaves Before Pruning	261
Number of Leaves After Pruning	25

Number of Observations Read	539
Number of Observations Used	539
Number of Training Observations Used	432
Number of Validation Observations Used	107

Variable Importance							
		Training		Validation			
Variable	Variable Label	Relative	Importance	Relative	Importance	Relative Ratio	Count
GP	GP	1.0000	263.6	1.0000	135.6	1.0000	9
Efficiency	Efficiency	0.4028	106.2	0.3293	44.6541	0.8175	7
_3PM	_3PM	0.0491	12.9437	0.0704	9.5493	1.4344	1
Min	Min	0.0433	11.4109	0.0379	5.1338	0.8747	1
Age	Age	0.0457	12.0420	0.0241	3.2691	0.5278	1
STL	STL	0.0754	19.8873	0.0000	0	0.0000	1
FGP	FGP	0.0734	19.3451	0.0000	0	0.0000	1
FTA	FTA	0.0544	14.3381	0.0000	0	0.0000	1
OREB	OREB	0.0507	13.3698	0.0000	0	0.0000	1
AST	AST	0.0482	12.7109	0.0000	0	0.0000	1

CART MODEL - INTERPRETING NODE

Node Information					
		Training Data		Validation Data	
ID	Path	Count	Pred	Count	Pred
0	Root Node	432	24.6528	107	21.4579
1	Root Node	432	24.6528	107	21.4579
	GP < 47.74	172	10.7674	57	10.9474
2	Root Node	432	24.6528	107	21.4579
	GP >= 47.74 or Missing	260	33.8385	50	33.4400
3	Root Node	432	24.6528	107	21.4579
	GP < 47.74	172	10.7674	57	10.9474
	GP < 20.68	77	3.8182	28	3.7857
4	Root Node	432	24.6528	107	21.4579
	GP < 47.74	172	10.7674	57	10.9474
	GP >= 20.68 or Missing	95	16.4000	29	17.8621
5	Root Node	432	24.6528	107	21.4579
	GP >= 47.74 or Missing	260	33.8385	50	33.4400
	Efficiency < -1 or Missing	134	28.8433	19	27.8421

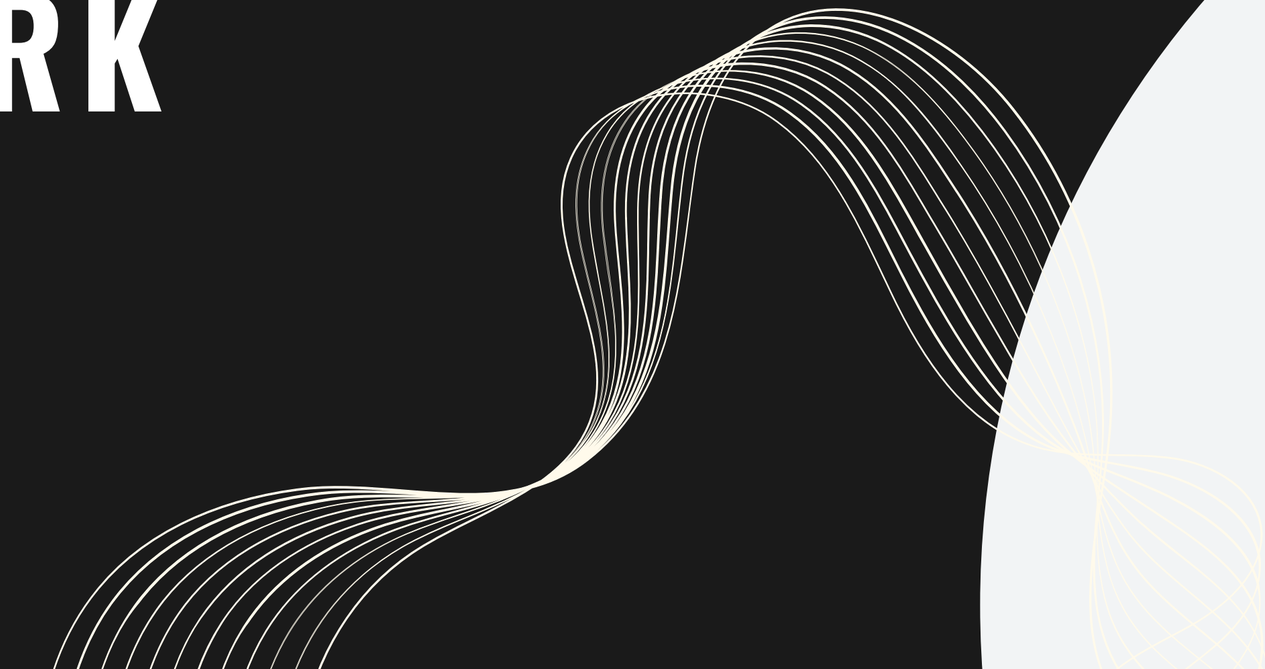
FIRST CART MODEL - EVALUATION

The MEANS Procedure

Variable	N	Mean
mape_fit	423	21.8001631
mape_acc	104	29.3792970
mae_fit	432	3.3295582
mae_acc	107	3.7757293
mse_fit	432	18.3806581
mse_acc	107	23.9465153

- No clear signs of overfitting as there is no drastic drop-off in performance from training to validation.
- Low error measures indicate relatively high model accuracy in terms of predicting players wins

NEURAL NETWORK



FIRST NEURAL NETWORK OUTPUT

Model Information	
Data Source	WORK.NBA_PART
Architecture	MLP
Number of Input Variables	26
Number of Hidden Layers	1
Number of Hidden Neurons	26
Number of Target Variables	1
Number of Weights	885
Optimization Technique	Limited Memory BFGS

Number of Observations Read	539
Number of Observations Used	539
Number Used for Training	432
Number Used for Validation	107

Train: Average Absolute Error	Valid: Average Absolute Error
3.362792	2.954259

The MEANS Procedure

Variable	N	Mean
mape_fit	423	22.998
mape_acc	104	26.507
mae_fit	432	3.363
mae_acc	107	2.954
mse_fit	432	20.111
mse_acc	107	15.865

- After building the model, AAE doesn't indicate any sign of overfitting
- Based on MAE and MSE, the model performs slightly better on the test set than the training set.
- Overall, no drastic difference between the test and training sets. Error measures are low suggesting that neural network is also a reliable tool for evaluating player performance and predicting outcomes.

CONCLUSION FROM FIRST ROUND OF MODELS

The MEANS Procedure

Variable	N	Mean
mape_fit	423	23.451
mape_acc	104	26.105
mae_fit	432	3.432
mae_acc	107	2.998
mse_fit	432	21.075
mse_acc	107	16.666

LINEAR REG

The MEANS Procedure

Variable	N	Mean
mape_fit	423	21.8001631
mape_acc	104	29.3792970
mae_fit	432	3.3295582
mae_acc	107	3.7757293
mse_fit	432	18.3806581
mse_acc	107	23.9465153

CART

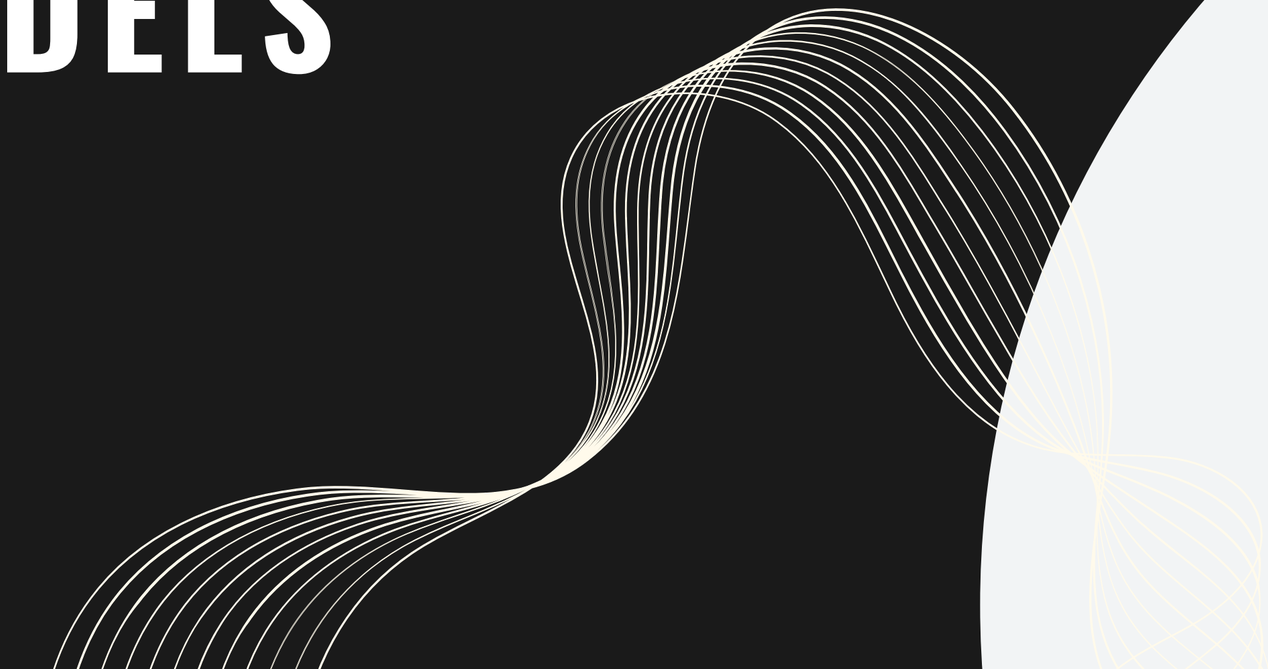
The MEANS Procedure

Variable	N	Mean
mape_fit	423	22.998
mape_acc	104	26.507
mae_fit	432	3.363
mae_acc	107	2.954
mse_fit	432	20.111
mse_acc	107	15.865

NEURAL NETWORK

SECOND ROUND OF MODELS

Linear Regression
CART
Neural Network



VARIABLES FOR SECOND MODEL

Predictors in second model
AGE
GP
FGM
_3PM
FTM
OREB
DREB
AST
STL
BLK
EFFICIENCY

- Streamlined model based on the variable importance from the first round of linear regression and CART models.
- Also included a few other performance metrics that we thought would be significant

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-1.120727	0.733929	-1.53	0.1275	0
GP	1	0.516311	0.017056	30.27	<.0001	1.66862
_3PM	1	0.183204	0.039574	4.63	<.0001	56.92545
_3PA	1	-0.064249	0.015399	-4.17	<.0001	59.36984

Linear Model

Variable Importance

Variable	Variable Label	Training		Validation		Relative Ratio	Count
		Relative	Importance	Relative	Importance		
GP	GP	1.0000	263.6	1.0000	135.6	1.0000	9
Efficiency	Efficiency	0.4028	106.2	0.3293	44.6541	0.8175	7
_3PM	_3PM	0.0491	12.9437	0.0704	9.5493	1.4344	1
Min	Min	0.0433	11.4109	0.0379	5.1338	0.8747	1
Age	Age	0.0457	12.0420	0.0241	3.2691	0.5278	1
STL	STL	0.0754	19.8873	0.0000	0	0.0000	1
FGP	FGP	0.0734	19.3451	0.0000	0	0.0000	1
FTA	FTA	0.0544	14.3381	0.0000	0	0.0000	1
OREB	OREB	0.0507	13.3698	0.0000	0	0.0000	1
AST	AST	0.0482	12.7109	0.0000	0	0.0000	1

CART Model

SECOND MULTIPLE LINEAR REGRESSION

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	NBA
Random Number Seed	12345
Sampling Rate	0.8
Sample Size	432
Selection Probability	0.801484
Sampling Weight	0
Output Data Set	NBA_PART

The data was partitioned 80% allocated to the training dataset (432 observations) and 20% for validation. Simple random sampling ensured an unbiased split.

Selection Information	
Selection Method	Stepwise

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	81362	27121	1273.29	<.0001
Error	428	9116.20884	21.29955		
Corrected Total	431	90478			

Root MSE	4.61514
R-Square	0.89924
Adj R-Sq	0.89854
AIC	1759.33376
AICC	1759.47461
SBC	1341.60747
ASE (Train)	21.10234
ASE (Validate)	16.70372

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-0.527627	0.510579	-1.03	0.3020	0
GP	1	0.528320	0.010787	48.98	<.0001	1.37549
Efficiency	1	0.033700	0.001499	22.48	<.0001	1.08932
FTM	1	-0.009776	0.002479	-3.94	<.0001	1.47703

89.9% of the variability in player wins is explained by the model, as indicated by the R-Square (0.89924) value. GP, Efficiency, and FTM are statistically significant ($p < 0.05$), with low VIF values (1.37549, 1.08932, and 1.47703) indicating minimal multicollinearity

SECOND MULTIPLE LINEAR REGRESSION - EVALUATION

- No clear signs of overfitting were observed, as the performance between the training dataset (MAPE = 23.466, MSE = 21.102) and the validation dataset (MAPE = 26.155, MSE = 16.704) was consistent.
- Low error measures, including MAE values (3.440 training, 3.002 validation), indicate high accuracy in predicting player wins.
- The results suggest that the model generalizes well to unseen data, making it a reliable tool for evaluating player performance and predicting outcomes.

The MEANS Procedure

Variable	N	Mean
mape_fit	423	23.466
mape_acc	104	26.155
mae_fit	432	3.440
mae_acc	107	3.002
mse_fit	432	21.102
mse_acc	107	16.704

SECOND CART MODEL

Model Information	
Split Criterion Used	Variance
Pruning Method	Cost-Complexity
Subtree Evaluation Criterion	Cost-Complexity
Number of Branches	2
Maximum Tree Depth Requested	10
Maximum Tree Depth Achieved	10
Tree Depth	8
Number of Leaves Before Pruning	271
Number of Leaves After Pruning	29

Number of Observations Read	539
Number of Observations Used	539
Number of Training Observations Used	432
Number of Validation Observations Used	107

Variable Importance							
		Training		Validation			
Variable	Variable Label	Relative	Importance	Relative	Importance	Relative Ratio	Count
GP	GP	1.0000	264.4	1.0000	136.0	1.0000	11
Efficiency	Efficiency	0.4050	107.1	0.3327	45.2633	0.8216	8
_3PM	_3PM	0.0799	21.1233	0.0736	10.0178	0.9218	3
FGM	FGM	0.0677	17.8939	0.0224	3.0503	0.3313	1
AST	AST	0.0758	20.0540	0.0000	0	0.0000	2
STL	STL	0.0752	19.8873	0.0000	0	0.0000	1
OREB	OREB	0.0658	17.4083	0.0000	0	0.0000	2

SECOND CART MODEL - EVALUATION

The MEANS Procedure

Variable	N	Mean
mape_fit	423	21.1854773
mape_acc	104	29.3405080
mae_fit	432	3.1390508
mae_acc	107	3.7729383
mse_fit	432	16.7273192
mse_acc	107	23.2801754

- Narrowing down variables resulted in better model performance.
- No clear signs of overfitting from the MAPE and MAE, as deterioration from training set to validation set is within reasonable limits.
- MSE shows that it may be picking up some noise.

SECOND NEURAL NETWORK EVALUATION

Data Access Information			
Data	Engine	Role	Path
WORK.NBA_PART	V9	Input	On Client
WORK.NBAOUTNEURAL1	V9	Output	On Client

Model Information	
Data Source	WORK.NBA_PART
Architecture	MLP
Number of Input Variables	11
Number of Hidden Layers	1
Number of Hidden Neurons	11
Number of Target Variables	1
Number of Weights	144
Optimization Technique	Limited Memory BFGS

Number of Observations Read	539
Number of Observations Used	539
Number Used for Training	432
Number Used for Validation	107

Train: Average Absolute Error	Valid: Average Absolute Error
3.375923	2.869599

The MEANS Procedure

Variable	N	Mean
mape_fit	423	23.056
mape_acc	104	25.094
mae_fit	432	3.376
mae_acc	107	2.870
mse_fit	432	20.225
mse_acc	107	15.466

- After following the other models with narrowing down the predictors, there aren't any drastic changes and these results also don't indicate overfitting.
- Error measures are still low suggesting that neural network continues to be a reliable tool for evaluating player performance and predicting outcomes.

FINAL CONCLUSION

The MEANS Procedure

Variable	N	Mean
mape_fit	423	23.466
mape_acc	104	26.155
mae_fit	432	3.440
mae_acc	107	3.002
mse_fit	432	21.102
mse_acc	107	16.704

LINEAR REG

The MEANS Procedure

Variable	N	Mean
mape_fit	423	21.1854773
mape_acc	104	29.3405080
mae_fit	432	3.1390508
mae_acc	107	3.7729383
mse_fit	432	16.7273192
mse_acc	107	23.2801754

CART

The MEANS Procedure

Variable	N	Mean
mape_fit	423	23.056
mape_acc	104	25.094
mae_fit	432	3.376
mae_acc	107	2.870
mse_fit	432	20.225
mse_acc	107	15.466

NEURAL NETWORK

CONCLUSION FROM FIRST ROUND OF MODELS

The MEANS Procedure

Variable	N	Mean
mape_fit	423	23.451
mape_acc	104	26.105
mae_fit	432	3.432
mae_acc	107	2.998
mse_fit	432	21.075
mse_acc	107	16.666

LINEAR REG

The MEANS Procedure

Variable	N	Mean
mape_fit	423	21.8001631
mape_acc	104	29.3792970
mae_fit	432	3.3295582
mae_acc	107	3.7757293
mse_fit	432	18.3806581
mse_acc	107	23.9465153

CART

The MEANS Procedure

Variable	N	Mean
mape_fit	423	22.998
mape_acc	104	26.507
mae_fit	432	3.363
mae_acc	107	2.954
mse_fit	432	20.111
mse_acc	107	15.865

NEURAL NETWORK

APPLICATION

- Front Office/Ownership
  - Drafting players
  - Player contract extensions
  - Staff hires
- Coaching Staff
  - Player minute distribution
  - Player development
  - Personnel decisions
- Players
  - Improvement plans/focused training.

IDEAS?

- Collect data over multiple seasons to identify possible trends.
- Collect data across different levels of basketball
- Transform data to run a logistic regression model
  - Create binary outcome for teams with roughly 45 wins (historical winning % in an 82-game season).

THANK YOU!