

פרויקט גמר

שאלת מחקר: הפרויקט שלי ינסה לענות על השאלה האם קיים מגמה של ירידה בכמות הסנטרים בליגת ה - NBA?

רקע - בעבר הלא רחוק העמדה הכי חשובה או בין השתיים הכי חשובות (תלוי את מי שואלים..) היה עמדת הסנטר, "הלב" של הקבוצה. כל קבוצה ידעה שבשביל לבנות קבוצה חזקה, צריך קודם כל סנטר דומיננטי. הסגנון משחק לרוב היה מבוסס סביב הסנטר שיקבל את הכדור מתחת לסל וייצר נקודות, ישלוט בריבאונד ויחסום כל שחקן יריב שחודר לסל. שחקן שמשחק בעמדת הסנטר היה באופן טבעי מאוד מבוקש בגלל שלמצוא סנטר גם גבוה (יחסית לשאר השחקנים), גם אתלטי (יחסית לשחקן מעל 2.05) וגם עם כישורי כדורסל ברמה גבוהה זה שילוב שיש למעט מאוד שחקנים לאורך ההיסטוריה, ולכן יש תחרות בין הקבוצות על הסנטרים הטובים וגם הפחות.

בשנים האחרונות חל שינוי משמעותי בענף הכדורסל כלל וב - NBA בפרט, בכך שיותר ויותר קבוצות (בהובלת גולדן סטייט) התחילו פחות לחדור לסל ויותר לזרוק ממרחק, בעיקר משלוש, במילים אחרות עידן ה - small ball. בעקבות כך, קבוצות חיפשו לצרף אליהם קודם כל שחקנים שיש להם יכולת קליעה, יכולת לשמור על קו השלוש ושחקנים מהירים או לפחות לא איטיים. שלושת הדרישות האלו הם לרוב החולשות של הסנטרים, מה שהוביל לשינוי משסנטר הוא הגורם אחד החשובים בקבוצה לכך שהוא נטל בהגנה ובהתקפה ויש קבוצות שאפילו החליטו לוותר בכלל על העמדה הזאת.

קווים מנחים:

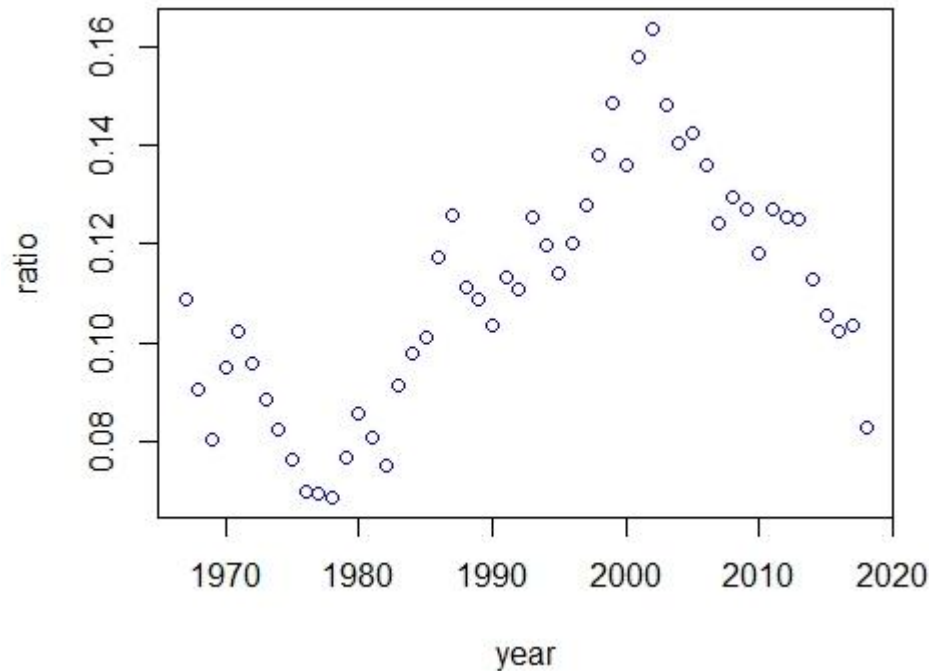
- סנטר יוגדר עבור שחקנים שהם רק סנטרים ולא שחקנים שהם גם סנטרים וגם פורוורדים או כל שילוב אחר.
- השנים עליהם מתבסס הפרויקט הם שנים שבהן יש לפחות 200 שחקנים לאותה שנה.

שיטות מחקר:

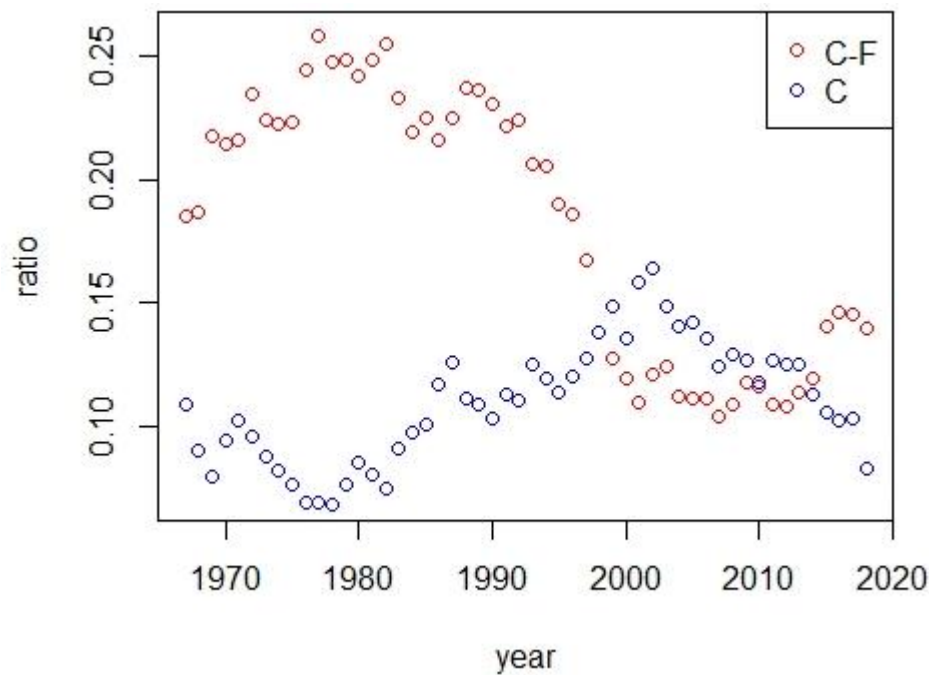
- Exploratory data analysis
- T – test
- PCA
- Prediction

Exploratory data analysis

מספר הסנטרים ביחס לשאר השחקנים:



מספר הסנטרים ביחס לשאר השחקנים משולב עם מספר הסנטרים-פורוורדים ביחס לשאר השחקנים:



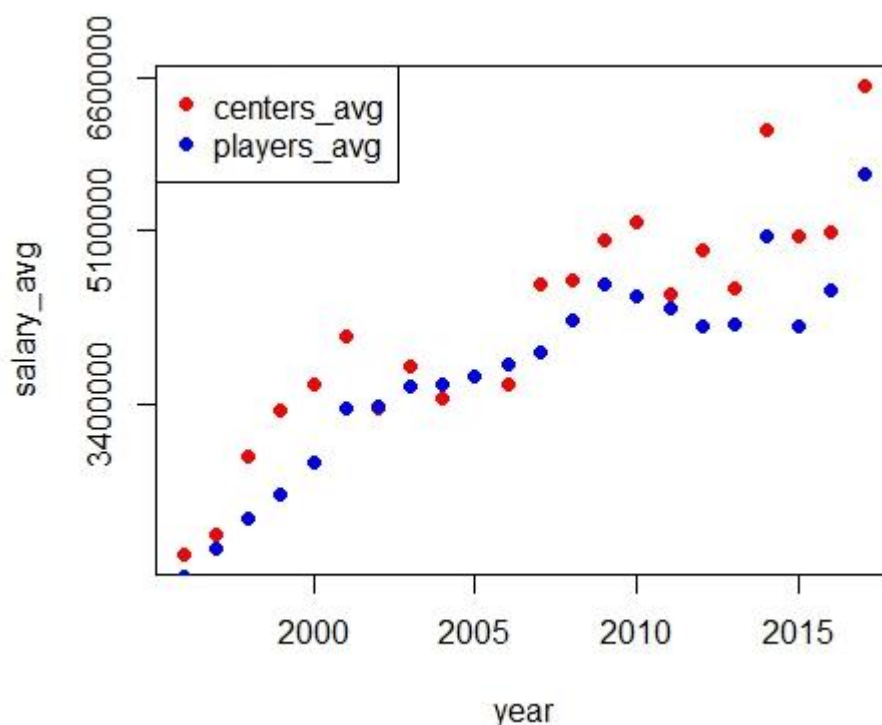
בתרשים הראשון ניתן לראות את מגמת העלייה חדה החל מסוף שנות ה-70 ועד לתחילת שנות ה-2000 במספר הסנטרים ב-NBA, ולאחר מכן חלה ירידה חדה עד תחילת העשור הקודם (2010) שנעצרת מעט וכמה שנים אח"כ מתחילה שוב ירידה חדה.

התרשים השני בא לנסות להסביר את מספר המועט של סנטרים בין השנים 1970-2000 בכך שיכול להיות שהגדירו את תפקיד הסנטר-פורוורד ואת תפקיד הסנטר באופן שונה או שלא מעט סנטרים נדרשו להיות גם פורוורדים או כל סיבה מקצועית אחרת שסנטרים נדרשו להוות גם את תפקיד הפורוורד.

בסיכום שני התרשימים ניתן לראות שמשנות ה-2000 יש ירידה בכמות הסנטרים וגם בסנטרים – פורוורדים. השינוי היותר מעניין הוא בתחילת – אמצע העשור הקודם שמספר הסנטרים ממשיך לרדת ובנוסף המספר

הסנטרים-פורוורדים חזר לעלות לאחר שנים של ירידה. אולי כתוצאה מדרישות להפוך את תפקיד הגבוה ליותר ורסטילי או כל סיבה מקצועית אחרת.

ממוצע המשכורות של סנטרים וממוצע המשכורות של כלל השחקנים (כולל סנטרים):



בתרשים הנ"ל מתאר את המשכורות הממוצעות של הסנטרים (באדום) ושל כלל השחקנים (כולל סנטרים בכחול).

לפי התרשים, מלבד שלוש עונות, בין השנים 2004 – 2006, שבהן ממוצע שכר של כלל השחקנים גבוה או שווה לממוצע השכר של סנטרים, בכל שאר העונות הממוצע שכר של הסנטרים גבוה ולפעמים משמעותית משכר ממוצע של כלל השחקנים למרות שיש ירידה במספר הסנטרים הכללי בליגה. אני חושב שכנראה תפקיד הסנטר עדיין חשוב בכדורסל ואין הרבה שחקנים מתאימים לתפקיד הזה ולכן הממוצע שכר של הסנטרים נשאר גבוה ואף גדל.

T – test

YEAR_1	YEAR_2	AVG_MIN_1	AVG_MIN_2	DIFF	P.VAL	SIG
1977	2007	22.17361	18.44054	3.73306981	0.020570646	*
1978	2008	22.44262	17.41818	5.02444259	0.001384237	**
1979	2009	21.70298	17.61307	4.08991718	0.014034023	*
1980	2010	23.24406	18.51358	4.73047862	0.001621663	**
1981	2011	21.49833	16.98430	4.51402593	0.004156759	**
1982	2012	20.94164	17.67838	3.26326319	0.024044768	*
1983	2013	19.56300	17.15524	2.40776170	0.096203614	
1984	2014	20.62093	17.41762	3.20330199	0.043191317	*
1985	2015	19.07912	19.16925	-0.09012724	0.949881737	
1986	2016	19.14139	17.52570	1.61569440	0.321157925	
1987	2017	19.36124	18.13896	1.22227903	0.412853364	

Ho – אין הבדל בממוצע הדק'

H1 – יש הבדל בממוצע הדק'

אחרי שעשיתי data exploratory analysis, הבנתי שהירידה בכמות הסנטרים ב NBA התחילה כנראה לפני שנת 2000 ולכן החלטתי ללכת 30 שנה לאחור ולעשות t - test בין התקופות 1977-1987 ולתקופה 2007-2017. לכל שנה חישבתי עבור כל עמדה ואת מספר הדק' הממוצע לאותה עמדה ובטבלה כלולים הנתונים הללו עבור הסנטרים. למשל, ב – 1977 ממוצע הדק' הכולל לשחקנים בעמדת הסנטר הוא 22.17361.

לפי הטבלה, במבחני t שבין השנים 1977-1984 לבין השנים 2007-2014 קיים הבדל משמעותי, ברמת מובהקות שלפחות 95%, בדקות המשחק הממוצע עבור עמדת הסנטר, למרות שביחס כמות הסנטרים לכמות כלל השחקנים (לפי הסעיף הקודם) לא רק שאין ירידה אלא להיפך ויש יותר סנטרים בשנים 2007-2014 לעומת השנים 1977-1984. אחת הסיבות לפי דעתי היא הוספת קשת ה – 3 ל NBA בשנת 1977 וכמו לא מעט דברים בחיים לקח זמן הסתגלות, שבסופו של דבר ההשפעה שלו ניכרת בעיקר אחרי 1981 שהממוצע דק' ירד באופן משמעותי.

לשאר מבחני t אין הבדל משמעותי ולא קיים רמת מובהקות שלפחות 95% ולכן לא ניתן לדחות את Ho.

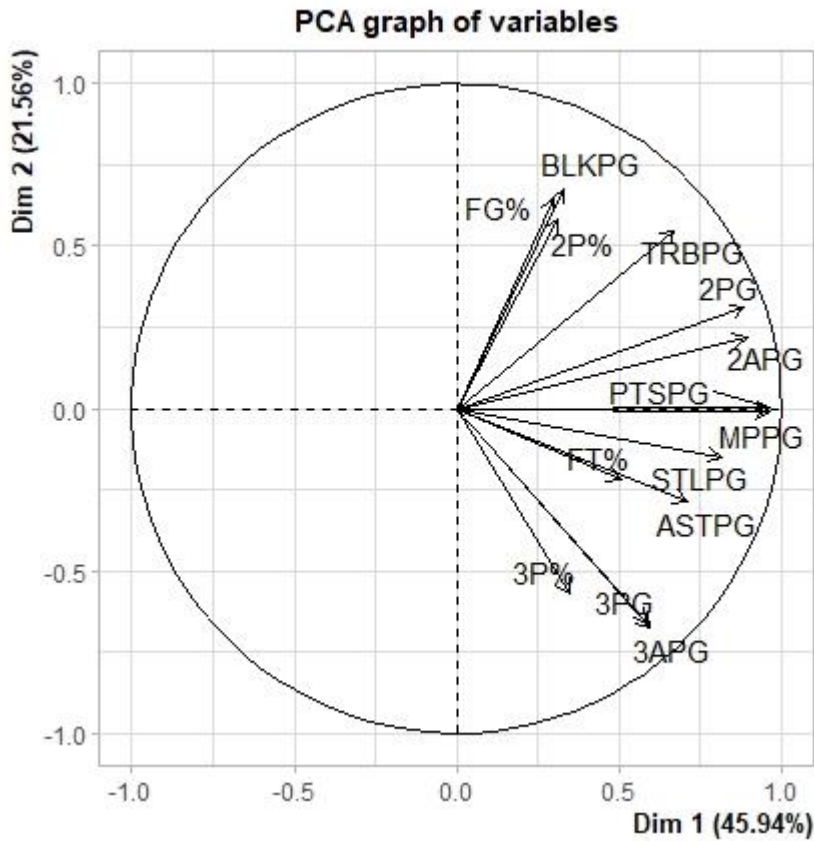
YEAR_1	YEAR_2	AVG_MIN_1	AVG_MIN_2	DIFF_AVG	P.VAL	SIG
1996	2007	17.47720	18.44054	-0.9633412	0.5798512	
1997	2008	16.72153	17.41818	-0.6966489	0.7375410	
1998	2009	18.80632	17.61307	1.1932534	0.4167228	
1999	2010	16.83844	18.51358	-1.6751382	0.6282819	
2000	2011	17.59057	16.98430	0.6062668	0.6712581	
2001	2012	17.86593	17.67838	0.1875468	0.7108002	
2002	2013	17.47005	17.15524	0.3148073	0.9059231	
2003	2014	17.59924	17.41762	0.1816208	0.8652149	
2004	2015	16.48184	19.16925	-2.6874111	0.0879438	
2005	2016	16.39738	17.52570	-1.1283197	0.2849461	
2006	2017	16.46465	18.13896	-1.6743109	0.1340344	

בטבלה הנ"ל אין באף אחד ממבחני t רמת מובהקות מעל 95% לכן לא ניתן לקבל את H1 ונקבל את H0. למרות שיש יותר סנטרים בשנים 1996-2006, הכמות דק' לא השתנתה. אני מעריך שזה נובע שאמנם יש פחות סנטרים ב NBA, אבל הסנטרים שכן נמצאים בליגה הם עדיין חשובים לקבוצות שלהם כמו לפני עשור ולכן הממוצע דק' שלהם לא השתנה.

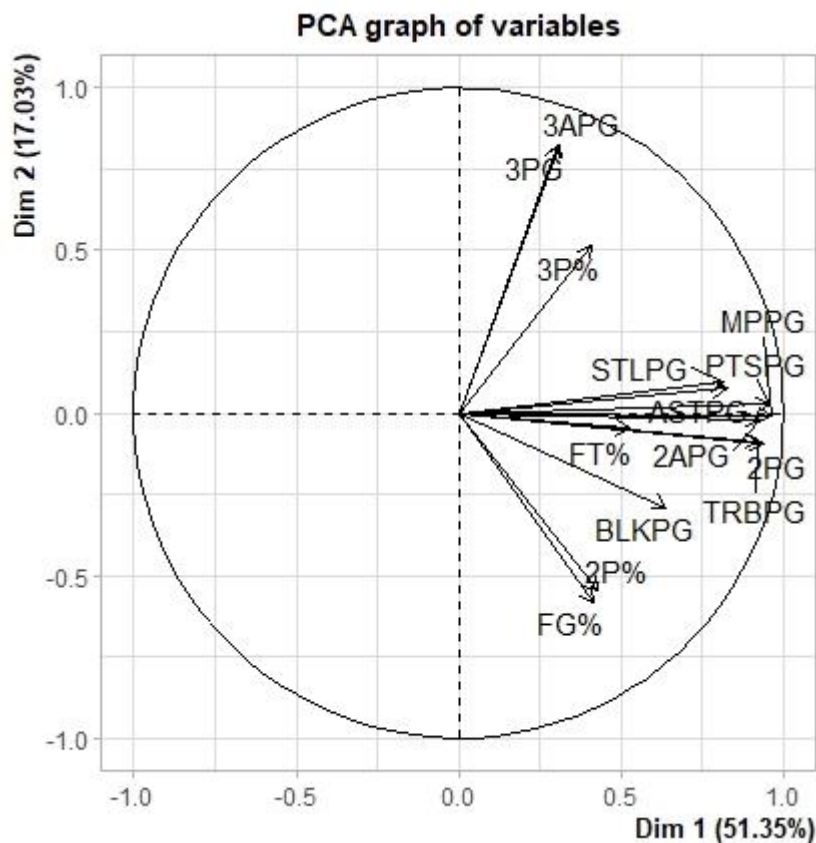
PCA

בסעיף הזה אני הולך לבדוק את קשר בין כמות הדק' של כלל השחקנים (כולל סנטרים) ושל הסנטרים בנפרד. הנתונים כולם הם לפי ממוצע פר משחק. למשל, 2APG מציין כמות הזריקות למשחק בממוצע

PCA של כלל השחקנים ושל הסנטרים בעונת 2009-10

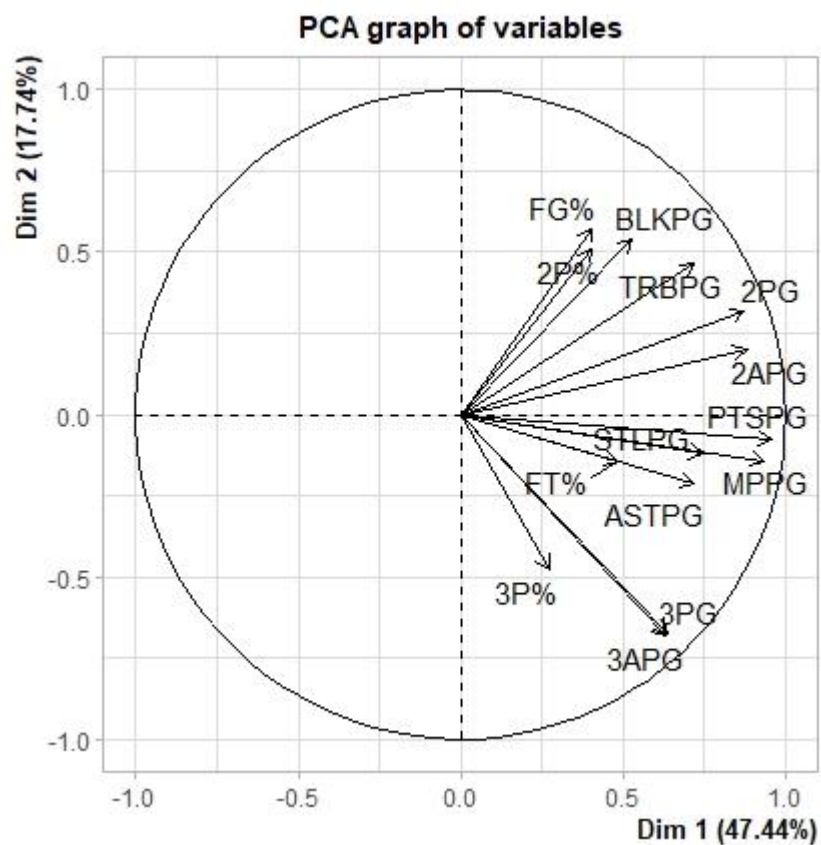


← כלל השחקנים

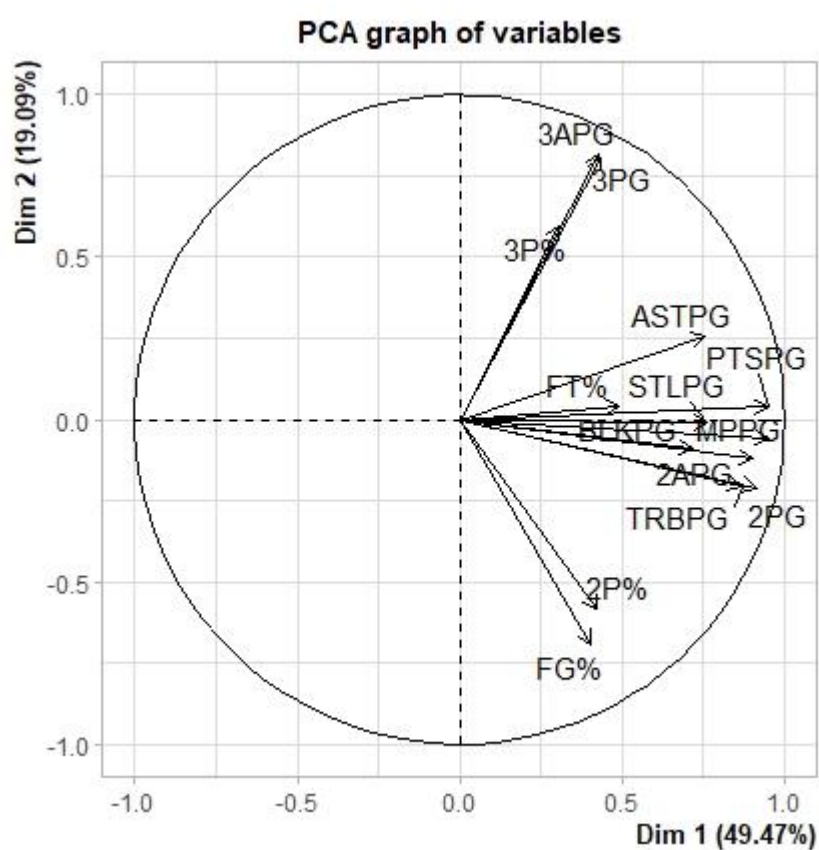


← סנטרים

PCA של כלל השחקנים ושל הסנטרים בעונת 2018-19



← כלל השחקנים



← סנטרים

טבלת קורלציות בין מספר הדק' לשאר הנתונים הסטטיסטיים

	STAT	p_2009_10	c_2009_10	p_2018_19	c_2018_19
1	FG%	0.2169459	0.3271199	0.2509859	0.3631286
2	3P%	0.2964650	0.3440686	0.2634563	0.1835714
3	2P%	0.2260201	0.3333182	0.2815011	0.3579157
4	FT%	0.4707158	0.4402315	0.4542605	0.4095076
5	MPPG	1.0000000	1.0000000	1.0000000	1.0000000
6	TRBPG	0.6869302	0.9356830	0.6431428	0.8962831
7	ASTPG	0.6842789	0.8123086	0.6759064	0.6906167
8	STLPG	0.7909215	0.8100761	0.7386616	0.7518710
9	BLKPG	0.3408317	0.6427116	0.4420470	0.6861746
10	PTSPG	0.9126598	0.9170505	0.8744446	0.8828363
11	2PG	0.8379032	0.9096891	0.7421373	0.8724561
12	2APG	0.8539714	0.8956407	0.7715823	0.8622966
13	3PG	0.5645797	0.2539983	0.6865204	0.3492314
14	3APG	0.5714693	0.2588689	0.6940797	0.3455162

אני אתייחס לחלק מהטבלה ואחתוך את החלקים הפחות מעניינים ולכן אתייחס לטבלה הבאה וסטטיסטיקה של ממוצע אסימטרים למשחק:

	STAT	p_2009_10	c_2009_10	p_2018_19	c_2018_19
1	FG%	0.2169459	0.3271199	0.2509859	0.3631286
2	3P%	0.2964650	0.3440686	0.2634563	0.1835714
3	2P%	0.2260201	0.3333182	0.2815011	0.3579157
11	2PG	0.8379032	0.9096891	0.7421373	0.8724561
12	2APG	0.8539714	0.8956407	0.7715823	0.8622966
13	3PG	0.5645797	0.2539983	0.6865204	0.3492314
14	3APG	0.5714693	0.2588689	0.6940797	0.3455162

לפי הקורלציה וגרף ה- PCA, ניתן לראות שקיים עלייה משמעותי בקורלציה בין כמות הזריקות לשלוש גם של כלל השחקנים וגם של הסנטרים לבין כמות הדק' משחק של שחקן, ובאותה נשימה יש ירידה בקורלציה של זריקות לשתיים של כלל השחקנים וירידה קטנה של הסנטרים. אצל הסנטרים, למרות שהם לוקחים יותר זריקות לשלוש, חל ירידה משמעותית בקורלציה שבין האחוזים לשלוש בעונת 2009-10 לעומת עונת 2018-2019.

לפי דעתי הסיבה לכך וכמו שראינו בסעיף הקודם שאין ירידה בכמות הדק' של סנטרים בממוצע למשחק, הסנטרים מסתגלים לשינויים וכלל הנראה נדרשים יותר לזרוק לשלוש מבעבר גם במחיר של אחוזים נמוכים לשלוש. בנוסף, עוד נתון מעניין הוא הירידה בקורלציה שבין ממוצע אסימטרים של סנטרים למשחק בעונת

2009-10 לעומת עונת 2018-19. הנתון הזה נובע לפי דעתי מכך שבעבר סנטרים היו מרכז ההתקפה וכל התקפה הייתה עוברת דרכם, ובשנים האחרונות ההתקפה פחות עוברת בהם והם פחות מנהלים את המשחק בעבר.

Prediction

הסעיף אבנה מודל שינסה לנבא את השכר של סנטרים ולבדוק איזה תכונות או נתונים סטטיסטים משפיעים על השכר.

בהתחלה לקחתי את כל הנתונים וחישבתי קורולציה בין השכר לבין שאר הנתונים

Cor	name
0.29902845	Age
0.24633469	G
0.50169428	GS
0.56048381	MP
0.53176546	FG
0.53395384	FGA
0.04661500	FG.
0.20735583	X3P
0.20900455	X3PA
0.11731927	X3P.
0.52041255	X2P
0.53839900	X2PA
-0.04831837	X2P.
0.04235963	eFG.

Cor	name
0.4437761	FT
0.4724653	FTA
0.0448831	FT.
0.4838225	ORB
0.5386943	DRB
0.5561702	TRB
0.4338165	AST
0.3962917	STL
0.3263858	BLK
0.4716328	TOV
0.5217394	PTS

בטבלאות הללו אין שום תכונה שנמצא בקורולציה חזקה מאוד, חיובית או שלילית. לאחר מספר ניסיונות של לקיחת מספר מינימלי של תכונות, החלטתי להשתמש רק בתכונות שהקורולציה שלהם גדולה יותר מ- 0.52 כי ככה קיבלתי מספר מינימלי של תכונות עם דיוק גבוה ובנוסף הסרתי תכונות דומות כמו אחוזים לשתיים, ריבאונד הגנה ואחוזים מהשדה והשארתי את זריקות לשתיים, ריבאונד כולל וזריקות מהשדה כי הקורולציה שלהם טיפה יותר חזקה.

Cor	name
0.5604838	MP
0.5339538	FGA
0.5383990	X2PA
0.5561702	TRB
0.5217394	PTS

← התכונות שעליהן בנית את המודל

האלגוריתם בו השתמשתי לבנות את המודל הוא random forest כי מכל המודלים הוא ניבא באחוזי הדיוק הגבוהים ביותר.

טבלאות של השכר, השכר שהמודל ניבא ואחוז דיוק:

Salary	predicSalary	Accuracy
22116750	22295503.1432666	99.191774816523
1551659	1519719.55726667	97.9415939498734
5318313	5291555.08346667	99.4968720996051
15756438	16248893.6037	96.8745753088357
2730000	2828333.9031	96.3980255274725
2202240	2193422.563	99.5996150737431
17145838	17128474.2447667	99.8987290371382
1034865	1073289.7497	96.2869794900784
5e+06	5021120.7935	99.57758413
1.7e+07	16956168.7258	99.7421689752941
4826160	5022375.59436667	95.9343330024975
14382023	14561421.2923	98.7526212946538
1921320	1956609.59766667	98.163262878299
5960160	6504836.97156667	90.8613699704929

Salary	predicSalary	Accuracy
3094013	3101179.3775	99.7683792052587
1015696	1051300.8812	96.4945336793686
2380440	2498727.8769	95.0308398069264
15500000	15713197.0103333	98.6245354172043
1024000	1123123.94516667	90.3199272298177
3140517	3194742.21016667	98.2733667683803
2568600	2561994.79113333	99.7428478989851
8250000	8267125.49773333	99.792418209293
15330435	15446490.7124	99.2429718243481
2898000	2917422.35256667	99.3298014987348
9769821	9780751.55690001	99.8881191692252
22642350	22193967.6722	98.0197182368437
7843500	7881224.78473333	99.5190312394552
2328652	2345657.45106667	99.2697298236634

Salary	predicSalary	Accuracy
6900000	6989389.33716667	98.7045023599033
3200000	3182990.12303333	99.4684413447916
17884176	18056792.3966	99.0348093387138
14112360	14135961.3123	99.8327614070219
4995120	4993681.9526	99.971210953891
5225000	5202262.15663333	99.5648259642743
2116955	2074106.5765	97.9759407498034
593780	677008.3065	85.9833092222709
6100266	6522024.69606667	93.0862245012485
1524305	1539446.9009	99.0066357520313
11422536	11511979.2153333	99.2169583415335
2947304	3194405.5779	91.6160132141103
1471382	1467715.76116667	99.750830251197
2947304	2899374.96946667	98.373800919982

Salary	predicSalary	Accuracy
2483040	2564321.2694	96.7265420855081
1.6e+07	16142999.1987	99.106255008125
13788500	13808062.1499	99.8581270631323
1471382	1464048.9897	99.5016243028663
22471911	22538703.7709333	99.7027721810873
3086784	3083358.42933333	99.8890246072719
2328652	2213396.47116667	95.0505473195079
1471382	1465742.34833333	99.6167105709689
3477600	3524226.10946667	98.6592446093091
28928710	26116582.0451333	90.2791104239813
2393887	2407712.60993333	99.422461881729
25434263	24994177.4651	98.2697138309059
9490740	9503607.29113334	99.8644226779646
3382000	3524757.0486	95.7789163630988

Salary	predicSalary	Accuracy
3057240	3054347.78823333	99.9053979482583
28928710	25304043.2223667	87.4703477008365
4350000	4366624.0536	99.6178378482759
2955840	2976014.5278	99.3174688819422
13528090	13545096.2061667	99.8742896730679
8175840	8411353.04853333	97.1194024279666
16900000	16888319.7494333	99.9308860913215
2659800	2734323.4166	97.1981571321153
4449000	4512459.46716667	98.5736240241253
3627842	3652398.44403333	99.3231115348096
24434262	23640005.9511333	96.7494166639178
2393887	2449371.5078	97.682241985524
10577466	10559672.3374667	99.8317776437822
14242782	14258045.0833667	99.8928363618381

Salary	predicSalary	Accuracy
14242782	14258045.0833667	99.8928363618381
12917808	12937353.9775333	99.8486896729435
6e+06	6023423.20333333	99.6096132777778
1764240	1738089.15783333	98.5177276239816
2098196	2081902.97763333	99.2234747198704
2799720	2819875.20903333	99.280099115864
2393887	2398730.28956667	99.7976809445614
1496500	1514676.5757	98.7853942064818
3628920	3644683.59916667	99.5656118303334
2393887	2397533.69923333	99.8476661917069

סה"כ אחוז דיוק: 98.10411

סיכום

ניתן לומר שכמות הסנטרים ב – NBA נמצאת בירידה מתחילת שנות ה – 2000 ולמרות שיש ירידה חדה יותר בתחילת העשור הקודם, הנסיגה בכמות הסנטרים התחילה עוד לפני העשור הקודם.

מבחינת מספר הדקות בממוצע על הפרקט הירידה המשמעותית היא החל מתחילת שנות ה - 80 ונעצרה באמצע העשור הזה והתייצבה לאורך השנים על סביב אותם ממוצעים ולא נפגע גם בעשור האחרון. בנוסף, המשכורות הממוצעת לסנטר היא גבוהה יותר ברוב המקרים מממוצע של שכר כלל השחקנים מה שמעיד על כך שעדיין גם היום תפקיד הסנטר הוא חשוב בכדורסל.

בשנים האחרונות תפקיד הסנטר השתנה מעט בכך שהם נדרשים ככל הנראה לזרוק יותר לשלוש מהשאר בעבר הלא רחוק לפי מה שרואים בסעיף של PCA. אני חושב שהסיבה היא שבעידן הנוכחי בשביל להיות שחקן NBA צריך את היכולת של זריקה לשלוש וגם לאט לאט הסנטרים מתחילים לזרוק יותר משלוש מאשר בעבר בשביל להתאים לכדורסל החדש של ה – NBA. למרות זאת, בינתיים אין קשר ישיר בין השכר לכמות זריקות לשלוש ועדיין אחת התכונות החשובות של סנטר הוא ריבאונד והוא גורם משמעותי בהשפעה על השכר.

לסיכום, תפקיד הסנטר יישאר גם בשנים הקרובות ב – NBA גם אם תהיה כמות פחותה ואני חושב שהסנטר הנוכחי והעתידי לא יהיו דומים לסנטרים מעשורים קודמים.

Resource

כל השחקנים בין השנים 2009/10 – 2019/20

<https://data.world/dwpeterson/nba-data-analysis/workspace/file?filename=NBAPlayerTotals.csv>

כל השחקנים בין השנים 1950 – 2017

<https://www.kaggle.com/drgilermo/nba-players-stats?select=Seasons Stats.csv>

"players" | "salaries_1985to2018"

<https://data.world/datadavis/nba-salaries>

משכורות השחקנים בין 2016 – 2019

<https://www.kaggle.com/davra98/nba-players-20162019>