

LeBron James karrier elemzés

Farkas Adél

Az elemzésem célja a világ egyik legjobb kosárlabdázójának tartott LeBron James karrierjének felderítése, és annak vizsgálata, hogy mennyire befolyásolja az egyéni teljesítménye a csapata sikerességét. A használt adathalmazom forrása a [Kaggle](#). Minden lejátszott alapszakaszbeli mérkőzése szerepel az adathalmazban a 2003-04-es szezontól kezdődően a 2019-20-as szezonig. Minden rekord egy mérkőzést és a hozzá tartozó mutatókat jelenti, összesen 1265 mérkőzést tartalmaz. A következő információk állnak rendelkezésre:

- game – a mérkőzés száma (szeznonként)
- date – dátum
- age – LeBron életkora az adott mérkőzés napján
- team – LeBron csapata
- opp – az ellenfél csapat
- result – eredmény (zárójelben a különbség)
- mp – minutes played, vagyis hogy mennyiz volt pályán
- fg – field goals (sikeres 2 pontos dobások)
- fga – field goals attempted (2 pontos próbálkozások)
- fgp – field goal percentage (2 pontos dobások sikerességének százalékos mutatója)
- three - sikeres 3 pontos dobások
- threeat - 3 pontos próbálkozások
- threep - 3 pontos dobások sikerességének százalékos mutatója
- ft – sikeres büntető dobások
- fta – büntető próbálkozások
- ftp – büntető dobások sikerességének százalékos mutatója
- orb – offensive rebounds (támadói labdaszerzés)
- drb – defensive rebounds (védekezői labdszerzés)
- trb – összes labdaszerzés
- ast – assists (olyan passz ami kosárral végződött)
- stl – steals (labdalopás)
- blk – blocks (ellenfél dobásának blokkolása)
- tov – turnovers (labda elvesztése)
- pts – points made (összes pontja a mérkőzés során)

- `game_score` – a későbbiekben részletesen magyarázom
- `plus_minus`

Adattisztítás

Szerencsére az adathalmazban kevés a hiányzó érték, csak olyan helyen kellett kitöltenem, ahol 0 lett volna a mutató (pl. „threep” hiányzott néhány helyen, mert a „three” és „threeat” értékek 0-k voltak és a „threep” ezeknek a hányadosa), így egy egyszerű `fillna(0)`-val oldottam meg. Ezen kívül csak néhány kisebb változtatást kellett végezni. Először is az „age” oszlopot alakítottam át, mivel „év-nap” formátumban volt és nekem csak az évre volt szükségem. Ezt úgy oldottam meg, hogy létrehoztam egy „age_” nevű oszlopot, ahová az „age” oszlop első két számjegyét tettem, majd int-té alakítottam. Ugyanezt tettem a „result” oszloppal, ahol zárójelben benne volt az is, hogy hány ponttal nyert vagy veszített az adott mérkőzésen. Itt létrehoztam a „result_” oszlopot, ahová csak az L (veszteség) vagy W (győzelem) betűk kerültek. Ezután töröltem az eredeti oszlopokat és a „game”-t, mert ezekre nem volt szükségem az elemzés során.

Elemzés

Csapatok szerint

Arra voltam kíváncsi, hogy melyik csapatnál játszott a legtöbb meccset és hol volt a legeredményesebb. Először a „game_score” oszlopot használtam az eredményesség mérésére, amit most el is magyaráznék. A `game_score` mérőszámot John Holliger találta ki azzal a céllal, hogy egyetlen mértékkel lehessen egy játékos teljesítményét számszerűsíteni mérkőzésenként. A képlete:

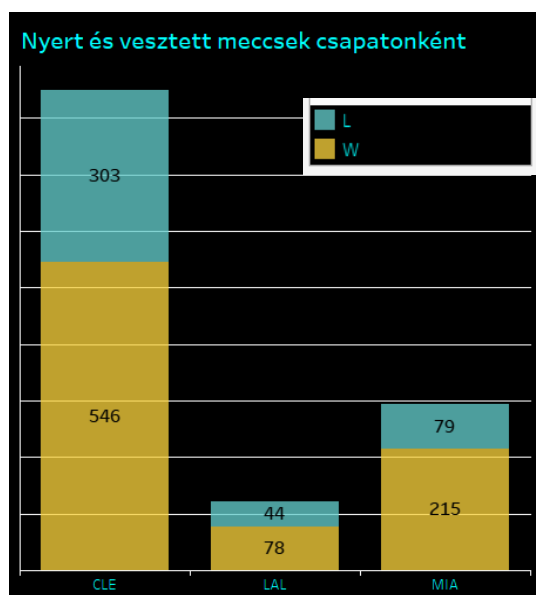
$$\text{game_score} = \text{PTS} + 0.4 * \text{FG} - 0.7 * \text{FGA} - 0.4 * (\text{FTA} - \text{FT}) + 0.7 * \text{ORB} + 0.3 * \text{DRB} + \text{STL} + 0.7 * \text{AST} + 0.7 * \text{BLK} - 0.4 * \text{PF} - \text{TOV}$$

Először megsámoltam, hogy egyes csapatoknál hány mérkőzést játszott. A Cleveland Cavaliers-nél játszott a legtöbb meccset, 849-et, majd a Miami Heat-nél 294-et, végül pedig a jelenlegi

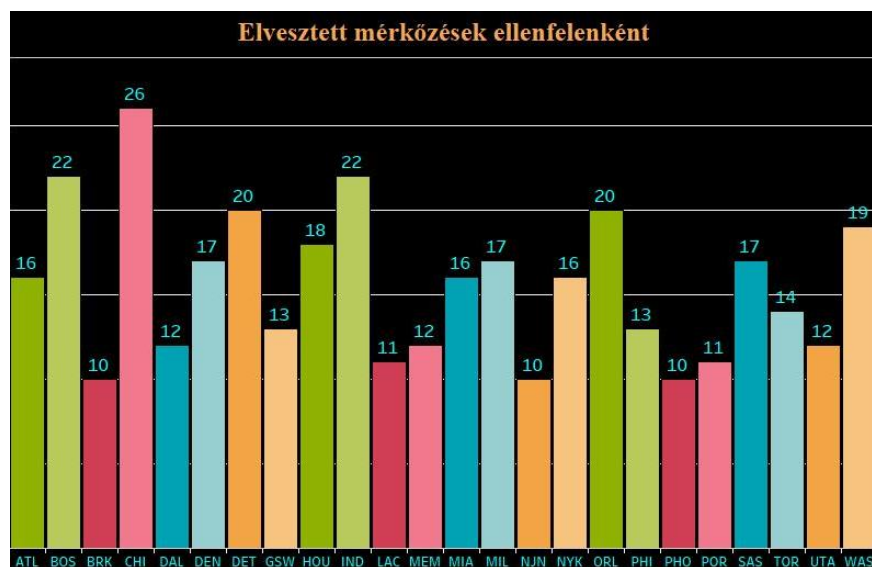
csapatánál, a Los Angeles Lakers-nél 122-t. Ezután a game_score alapján megnéztem, hogy átlagosan melyik csapatnál teljesített a legjobban. A Heat-nél teljesített a legjobban, majd a Lakers-nél, végül pedig a Cavaliers-nél volt a leggyengébb.

```
Out[14]: team
CLE      21.998704
LAL      22.133607
MIA      22.914626
Name: game_score, dtype: float64
```

Ezután megnéztem, hogy melyik csapatával nyerte legtöbb mérkőzést.



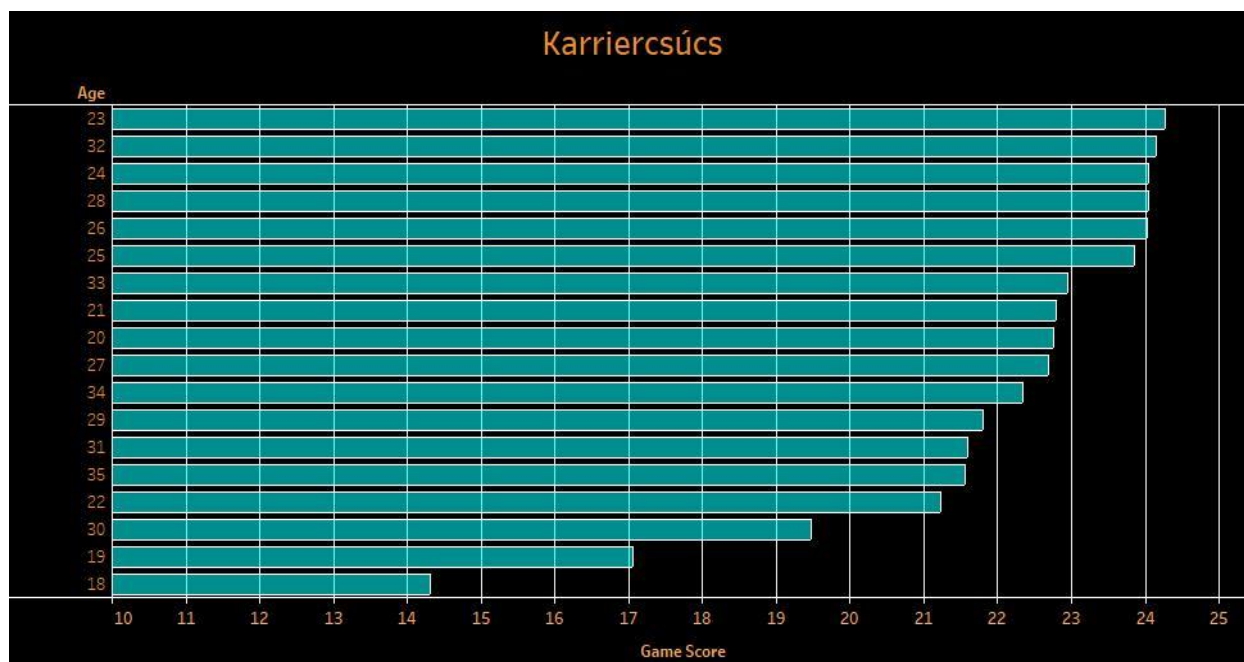
Azt is megvizsgáltam, hogy melyik ellenfél csapat ellen vesztette el a legtöbb meccset a karrierje során. Eredményül a Chicago Bulls csapatát kaptam, akik ellen 26 mérkőzést veszített.



LeBron játékának elemzése

Először arra voltam kíváncsi, hogy átlagosan hány percet tölt a pályán egy mérkőzés alatt (egy NBA mérkőzés 4x12 perces). Eredményül 38,40-et kaptam, ami ezt jelenti, hogy átlagosan minden mérkőzés majdnem 80%-át a pályán tölti. Ez komoly fizikai kondícióról árulkodik, kevés játékos képes ennyi időt végigjátszani.

Azt is megvizsgáltam, hogy game_score alapján hány évesen volt a karrierje csúcsán. Eredményül azt kaptam, hogy 23 éves korában játszott a legjobban. Az utóbbi 3 évben (33, 34, 35 éves korában) egyre romlott a teljesítménye, de ez teljesen érthető, legtöbb játékos már visszavonult az ő korában.



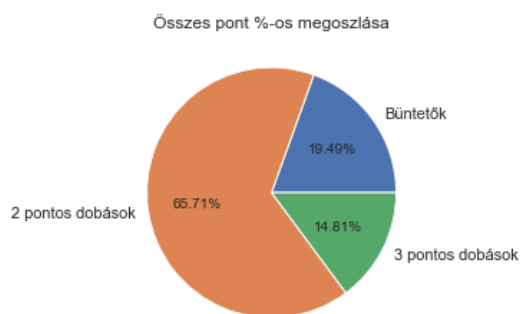
Ezután azt vizsgáltam, hogy melyik dobásai a legsikeresebbek átlagosan. Az “ftp” oszlop átlagát számoltam ki a büntető dobásokhoz, az “fgp” átlagát a két pontos dobásokhoz és a “threep”-ét a három pontosokhoz. A büntetők 72,03%-a sikeres, a 2 pontos dobásainak 50,49%-a megy be, míg a 3 pontos dobásai 30,07%-ban sikeresek.

Majd arra voltam kíváncsi, hogy LeBron összesen hány pontot dobott karrierje során. Ehhez egyszerűen csak összegeztem a „pts” oszlopot, és az eredmény 34241 lett. Mivel az adataim még nem tartalmazza az idei szezon adatait, illetve a nem alapszakaszokban szerzett pontokat, ezért valójában több pontot is szerzett, pontosan 35318 pontnál jár jelenleg. Ezzel egyébként ő

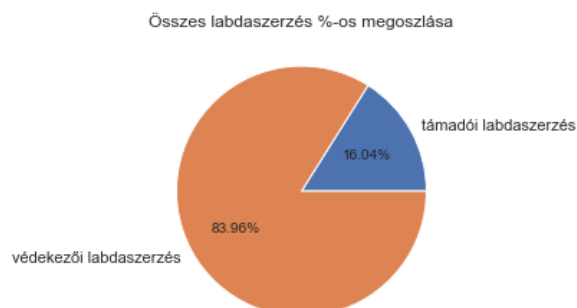
mindenidők harmadik legeredményesebb pontszerzője, megelőzve Kobe Bryant-et és Michael Jordan-t.

Points Leaders		
RK	PLAYER	PTS
1	Kareem Abdul-Jabbar	38,387
2	Karl Malone	36,928
3	LeBRON JAMES	35,318
4	Kobe Bryant	33,643
5	Michael Jordan	32,292

Azt is megnéztem, hogy milyen százalékban oszlanak meg a különböző fajta dobások az összes szerzett pontjában. Itt összegeztem az összes sikeres büntetőt, 2 pontost és 3 pontost, majd a 2 pontosokat 2-vel, a 3 pontosokat pedig 3-mal kellett szoroznom.



Ezután a játékának a többi aspektusát vizsgáltam. Először is azt számoltam ki, hogy átlagosan hány védekezői illetve támadói labdszerzése van egy meccs alatt. Támadói esetén 1,1928 lett az eredmény, védekezői esetén pedig 6,2418. Majd a pontokhoz hasonlóan megnéztem az összes labdaszerzés eloszlását.



Kiszámoltam, hogy átlagosan hány asszisztja, labdalopása, blokkolása, illetve labda vesztese van egy mérkőzés alatt. Az eredmények sorrendben: 7,39 ; 1,6 ; 0,76 és 3,5.

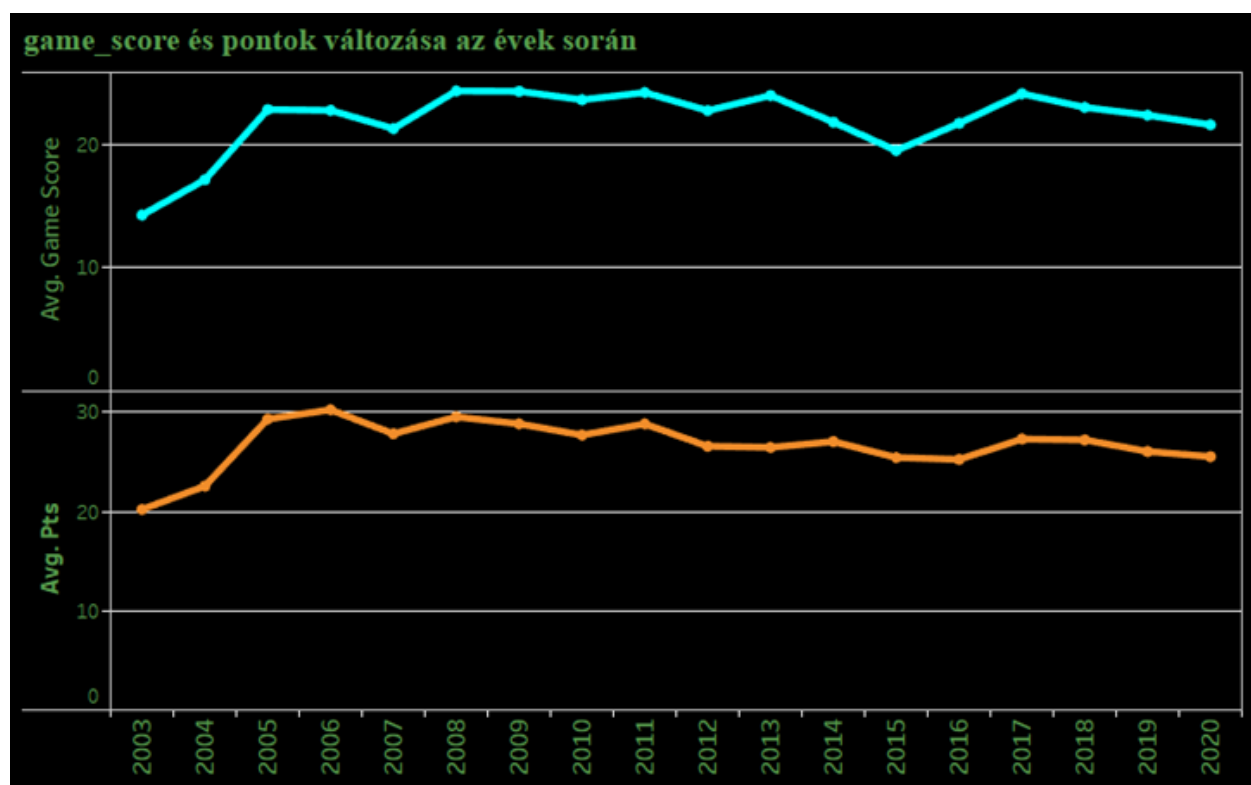
Ebből látszik, hogy nem csak pontokat tud dobni, jól tud passzolni és labdát szerezni is. Blokkolás szempontjából is kimagaslóan magas a mutatója.

Ezután megkerestem azt a mérkőzést, ahol a legeredményesebb volt karrierje során.

Egy 2017.11.03.-án játszott mérkőzésen volt a legmagasabb a game_score mutatója (53,2).

date	team	opp	mp	fg	fga	fgp	three	threeatt	...	drb	trb	ast	stl	blk	tov	pts	game_score	result	age
2017-11-03	CLE	WAS	43	23	34	0.676	2	4	...	6	11	7	3	2	3	57	53.2	W	32

Tableau segítségével megvizsgáltam, hogyan változott a game_score mutatójának és a dobott pontjainak átlaga évenként.



Láthatjuk, hogy az átlagosan dobott pontjainak száma és game_score-ja 2005 óta alig változik, 25 körül mozog a game_score, az átlagosan szerzett pontjai pedig 25 és 30 között. Egy ilyen átlag fenntartása 15 éven keresztül magáért beszél.

Hogyan hat az egyéni teljesítménye a mérkőzés kimenetére?

Arra voltam kíváncsi, hogy mennyire függ a csapat győzelme LeBron egyéni teljesítményétől. Ezt úgy vizsgáltam meg, hogy készítettem egy logisztikus regressziót, ahol a meccsek végeredményét (W vagy L) prediktáltam az egyéni mutatói alapján. Minél pontosabb a predikció, annál nagyobb befolyása van LeBron-nak a meccsek kimenetére. Első lépésként a 'result_' oszlop értékeit kicseréltem 0-ra, ahol L volt és 1-re ahol W szerepelt.

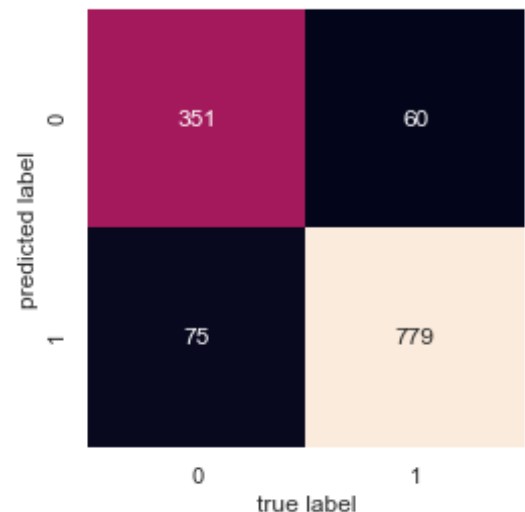
A szükséges csomagok beimportálása után definiáltam a training_features-t és a célváltozót, vagyis outcome_label-t ('result_'). Ezután a numerikus változókat normális eloszlásúra hoztam StandardScaler() segítségével, majd létrehoztam a regresszort és ráillesztettem az adatokra.

Ezután kiértékeltem a regressziót és készítettem egy konfúziós mátrixot:

Accuracy: 89.32806324110672 %

Classification Stats:

	precision	recall	f1-score	support
0.0	0.85	0.82	0.84	426
1.0	0.91	0.93	0.92	839
accuracy			0.89	1265
macro avg	0.88	0.88	0.88	1265
weighted avg	0.89	0.89	0.89	1265



Láthatjuk, hogy a predikció 89,33%-ban volt pontos.

A konfúziós mátrixról leolvasható, hogy az összes mérkőzésből 779 esetben prediktált helyesen győzelmet, 351-szer prediktálta helyesen a veszteséget, 75 meccs esetén hamis győzelmet, 60 mérkőzésnél pedig hamis veszteséget prediktált.

Elkészítettem a logisztikus regressziót Azure Machine Learning Studio-ban is, tesztelésre az adatok 35%-át használtam. Eredmény:

True Positive	False Negative	Accuracy	Precision
284	17	0.883	0.890
False Positive	True Negative	Recall	F1 Score
35	107	0.944	0.916
Positive Label	Negative Label		
1	0		

Itt rosszabb eredményt kaptam, 88,3%-ban volt pontos a predikció.

<https://gallery.cortanaintelligence.com/Experiment/LeBron-log-reg-3>

Végül még Naive Bayes-el is elvégeztem a predikciót Azure-ban, szintén az adatok 35%-át használva tesztelésre.

True Positive	False Negative	Accuracy	Precision
283	18	0.901	0.916
False Positive	True Negative	Recall	F1 Score
26	116	0.940	0.928
Positive Label	Negative Label		
1	0		

Itt lett a legpontosabb a predikció, 90,1%-os.

<https://gallery.cortanaintelligence.com/Experiment/LeBron-bayes-3>

Következtetés

A bemutatott adatok nem tükrözik meglepő eredményt, LeBron James köztudottan minden idők egyik legeredményesebb játékosa statisztikailag, számos rekordot döntött és dönt meg. Ami mégis megdöbbentő az az, hogy mennyire pontosak lettek a predikciók csak az ő teljesítménye alapján. Már több ízben igazolt olyan csapathoz, akik előtte play-off-ig sem jutottak el és LeBron érkezése után NBA-bajnokok lettek (Miami Heat 2012-ben és 13-ban, Cleveland Cavaliers 2016-ban és a Lakers 2020-ban). Ezt szokták úgy emlegetni, hogy a “LeBron effect” és a bemutatott eredmények alapján kijelenthetjük, hogy ez nem csak a véletlen műve.