

Budapesti Corvinus Egyetem  
Tudományos Diákköri Konferencia  
2024

**A középpályások értékelésének vizsgálata a  
labdarúgás világában**

**Analysing the valuation process of  
midfielders in the world of football**

Gazdálkodástudományi Kar  
Gazdálkodási és menedzsment alapképzési szak

Hajdu Bálint – VG8M5M  
Konzulens: Madari Zoltán

2024

## Absztrakt

A labdarúgásban a Transfermarkt német weboldallal és az általuk biztosított piaci értékekkel már mindenki találkozhatott, aki egy minimálisan is követi a sportágat. A weboldal hitelességével és megbízhatóságával már rengeteg tanulmány foglalkozott, de azt még kevésbé vizsgálták, hogy az általuk 2021-ben publikált legfontosabb befolyásoló tényezők valóban mennyire is magyarázzák a becsült piaci értékeket. Továbbá a sportban és ezáltal a futball világában is egyre elfogadottabbak és elérhetőbbek lettek a modern adatelemzési mutatók, köszönhetően az erre szakosodott úttörő cégeknek, mint például az OPTA Sports-nak. A kutatásom tehát azt fogja vizsgálni, hogy a Transfermarkt piaci értéket mennyire magyarázzák jól az általuk javasolt 16 különböző befolyásoló tényezők, illetve, hogy a modern adatelemzési mutatók (xG, xA, progresszív passzok) milyen hatással vannak az így készített modellre. A dolgozatom központi témája a 2022/23-as szezon 682 középpályásának adatelemzési leírása lesz az 5 legerősebbnek megítélt európai ligában, melyek a következők: angol, spanyol, olasz, francia, illetve német elsőosztályú bajnokságok. A kutatáshoz a használt adatbázist a Transfermarkt, FBref oldalak és a FIFA videójáték által nyújtott statisztikák mentén készítettem el. A dolgozat során a modellépítést és OLS-regressziót követően láthatóvá válik, hogy egy nagyon erős magyarázóerejű modell jön létre. Ez igazolja a publikált változók hitelességét és azt, hogy valóban ezek mentén és figyelembevételével tesz javaslatot a tömeg és ezek szerint dönt az önkéntes „bíró”. A kutatásban továbbá fény derül arra is, hogy a modern adatelemzési mutatók mind szignifikánsan pozitívan magyarázzák a piaci értéket, de a hagyományos statisztikai mutatópárjukra épített modellek még mindig pontosabb visszaigazolást adnak a Transfermarkt értékek vizsgálatakor.

**Kulcsszavak:** Transfermarkt, piaci érték, xG, xA, progresszív passz, OLS

## Tartalomjegyzék

1. Bevezetés.....	1
2. Szakirodalom.....	5
2.1: A Transfermarkt piaci értékek fontossága .....	5
2.2: A Transfermarkt és a tömegértékelés hitelességének az alapjai .....	6
2.3: A Transfermarkt oldal bemutatása és az értékelések működésének alapjai.....	8
2.4: Elért eredmények .....	9
3. Kutatási módszertan .....	16
4. Adatok bemutatása .....	17
4.1: A Transfermarkt ajánlott változói .....	17
4.2: A használt változók bemutatása .....	18
4.3.: Leíró statisztika .....	22
4.4.: Szükséges változók transzformálása .....	24
5. Eredmények.....	27
5.1.: Első kutatási eredményei.....	27
5.2.: Második kutatási kérdés eredményei .....	30
6. Konklúziók .....	34
6.1.: Összefoglalás.....	34
6.2.: Limitációk, kitekintés.....	35
Irodalomjegyzék.....	36
Mellékletek.....	I

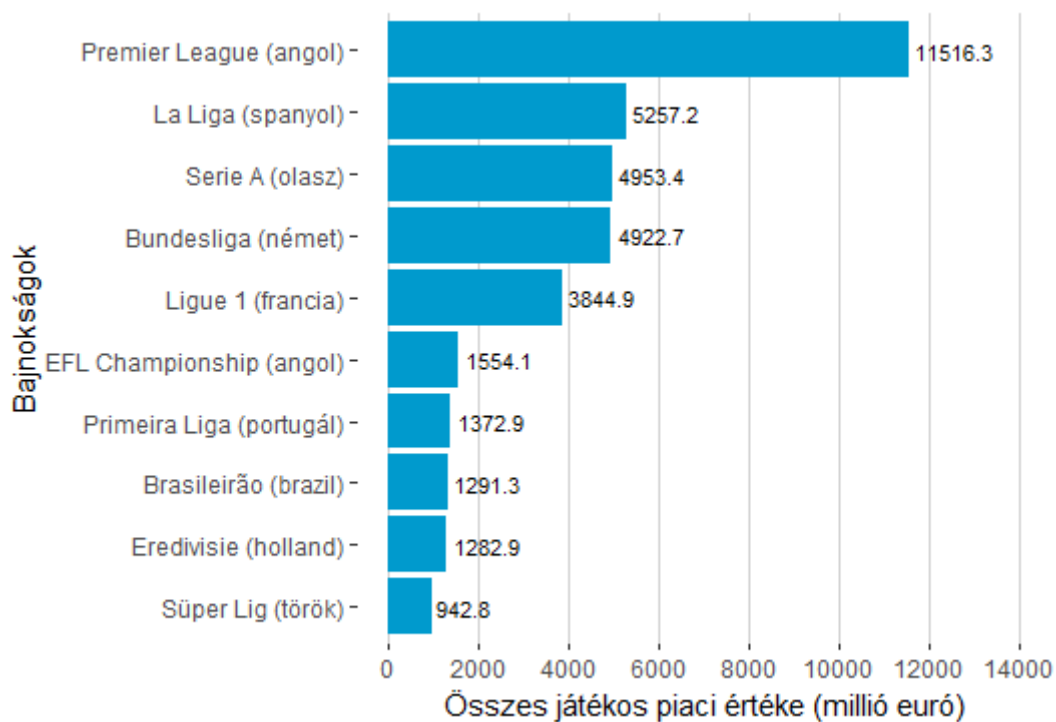
## Táblázat- és ábrajegyzék

1. táblázat: A Transfermarkt által javasolt tényezők és a hozzájuk kapcsolt változók .....	17
2. táblázat: A használt változók bemutatása.....	18
3. táblázat: leíró statisztika a numerikus változókra. ....	22
4. táblázat: Transfermarkt által javasolt változók és a piaci érték OLS regressziós eredményei .....	27
5. táblázat: A modellek összehasonlítása .....	31
1. ábra: A 10 legértékesebb és legerősebb bajnokság a játékosok piaci értéke alapján 2024. februárjában.....	2
2. ábra: Németországban a hírportálok online és mobil látogatásainak száma 2023. októberében .....	6
3. ábra: A Transfermarkt piaci értékelések működése .....	9
4. ábra: A bajnokságok játékosértékeinek dobozábrája .....	23
5. ábra: A sérülékenység változó dobozábrája .....	24
6. ábra: A 2023-as piaci érték eloszlása az eredeti és a logaritmált adatokkal.....	25
7. ábra: A 2022-es piaci érték eloszlása az eredeti és a logaritmált adatokkal.....	25
8. ábra: A Wikipedia keresések száma az eredeti és a logaritmált adatokkal .....	26
9. ábra: Az életkor kvadratikus tagjának kapcsolata az piaci értékkel .....	26

## 1. Bevezetés

A kutatási témám kiválasztása során kifejezetten fontos szempont volt számomra, hogy olyan területet vizsgáljak, amely valóban érdekel és van hozzá valami kapcsolódási pontom. Ezen tényező miatt döntöttem úgy, hogy a labdarúgás világával kapcsolatos kutatást és azon belül is kifejezetten a középpályások vizsgálatát szeretném a középpontba helyezni.

A kutatásom első kérdésköre a népszerű Transfermarkt által becsült piaci értékek vizsgálatát helyezi előtérbe és járja körül. A német weboldal a honlapján 2021-ben meghatározta azokat a legfontosabb tényezőket, amelyek figyelembevételét javasolják a felhasználóknak, mielőtt meg hozzák és kinyilvánítják a véleményüket egy játékos értékét illetően az ahhoz rendelt fórumon (Transfermarkt, 2021). A kutatásomban ezen legfontosabb tényezőkhöz változókat fogok kapcsolni és a 2022/23-as szezon alapján az 5 legerősebbnek ítélt európai bajnokságokban (angol, spanyol, német, olasz, francia) OLS regresszióval megvizsgálom, hogy milyen kapcsolat található a javasolt tényezők és a felhasználók által becsült piaci értékek között. Az említett bajnokságokban szereplő összes játékos piaci értékén alapuló rangsor az 1. ábrán látható. Ez a vizualizáció megmutatja, hogy a játékosok piaci értéke alapján mely bajnokságok a legértékesebbek, legerősebbek. A modell erősségét ezek után korrelációs index mentén értékelem. Ezek alapján tehát az első kutatási kérdésem azt vizsgálja, hogy a Transfermarkt által legfontosabbként javasolt tényezők mennyire magyarázzák a felhasználók által becsült piaci értékeket. Ennek alapján pedig megfogalmazom az első hipotézisem melynek nullhipotézise az, hogy a javasolt tényezőket a felhasználók figyelembe veszik az értékelés során, tehát van kapcsolat a becsült piaci értékek és a javasolt tényezők között. Ezzel ellentétben az alternatív hipotézisem az, hogy a felhasználók nem veszik figyelembe a javasolt tényezőket az értékelés során, tehát nincsen kapcsolat a javasolt tényezők és a becsült piaci értékek között.



**1. ábra: A 10 legértékesebb és legerősebb bajnokság a játékosok piaci értéke alapján  
2024. februárjában**

*Forrás: Football Benchmark (2024). Saját szerkesztés.*

A piaci értékek és az azokra ható változók vizsgálatával már jelentek meg szakirodalmak, de az egy kevésbé kutatott terület, hogy a Transfermarkt által 2021-ben javasolt tényezők valóban mennyire is magyarázzák jól a piaci értékeket. Ebben a kutatásban tehát én nem azt fogom vizsgálni, hogy mely változók magyarázzák a legjobban a Transfermarkt piaci értékeket, hanem csak a meghatározott tényezőkhez kapcsolt változók és az eredményváltozó kapcsolatát fogom vizsgálni ebben a kutatási kérdésben. A témában megjelent általam vélt legfontosabb kutatásokat és eredményeket, illetve a Transfermarkt működését, népszerűségét és hitelességét részletesen a szakirodalmi fejezetben fogom bemutatni.

A kutatásom második kérdésköre az újszerű, modern statisztikai mutatók kérdéskörét járja körbe, amely tartalmazza az xG, xA, 90 percre vetített progresszív passzok számát, illetve a progresszív passzok arányát. Az xG rövidítés a várható gólok számát mutatja egy adott pozícióból a historikus adatok elemzése alapján (Whitmore, 2023). Az xA rövidítés a várható gólpasszok számát mutatja, amely szintén hasonló metódus alapján működik, mint előző modern mutatótársa (Whitmore, 2021). A progresszív passzok egy játékos esetében a mélységi passzokat igyekszik számszerűsíteni, amelynek meghatározása eltérő lehet egyes adatszolgáltató fórumokon. A kutatásomban az FBref által biztosított adatokat fogom használni, amely cég leírása alapján a progresszív passzmutató 2 tényezőből adódik össze.

Progresszív passznak számít az ellenfél kapuja felé haladó legalább 10 yardos passz vagy szintén az ellenfél kapuja felé haladó passz a tizenhatoson belülre. Viszont ezen passzok csak akkor kerülnek számolásra, ha az átadás az egész pályának az ellenfél kapujához eső 60%-án történik (Sports Reference, n. d.). A modern statisztikai mutatókról egy átfogóbb és mélyebb megértést segítő meghatározást az OPTA Sport vagy az FBref oldalain lehet olvasni. A progresszív passzok aránya mutató pedig az összes passzkísérleten belül a progresszív passzok kísérleteit számszerűsíti. Ebbe egyaránt beletartoznak a sikeres és sikertelen kísérletek is.

Ehhez a kutatáshoz az előző kutatási kérdésben felépített modellt fogom használni, amelyet alapmodellnek fogok majd tekinteni. Ebben az alapmodellben mind a 4 modern statisztikai mutatónak azonosítok egy hagyományos mutatópárt, amelyet kicserélve megvizsgálom az újonnan kapott modelleket. Ezek alapján az xG-nek a gólok száma, az xA-nak a gólpasszok száma, a progresszív passzoknak a passzok száma és a progresszív passzok arányának a passzpontosság a mutatópárja a kutatásomban. A dolgozatban a progresszív passzok vizsgálatát egyszerre folytatom és ebben az esetben 2 mutatót helyettesítek az alapmodellhez képest. Ezek alapján 3 új OLS modellt fogok vizsgálni és AIC és BIC információs kritériumok, illetve a korrelációs index mutatók mentén összehasonlítom, hogy a hagyományos vagy a modern statisztikai mutatókon alapuló modell bír-e nagyobb magyarázóerővel. Fontos megemlíteni, hogy ezekkel az új változókkal nem bővítem a modellt, hanem csak az adott modern mutatót cserélem ki a hagyományosra, minden más változatlansága mellett.

Ezt a kutatást azért gondolom fontosnak, mivel egyre népszerűbbek a modern és újszerű mutatók a sport és a labdarúgás világában és egy mérkőzés megnézése közben is már találkozhatunk ezekkel a mutatókkal a televízióban. Ezen kívül amint már említettem az első kutatási kérdésemnél, az adatbázisomban a középpályásokat fogom vizsgálni, akiknél a passz minősége is fontos. Az xA esetében a csapattárstól való kiszolgáltatottságot szeretném kiszűrni, mivel egy játékos akkor ad gólpasszt, ha a társa értékesíti az általa adott passzból adódó helyzetet. Az xA viszont kiszűri ezt a hatást és csak azt vizsgálja, hogy a középpályás milyen minőségi passzt biztosított a társának és nem veszi már figyelembe azt, hogy abból gól születik-e vagy nem. A progresszív passzok szintén a passzok minőségét szemléltetik, mivel a passzok száma változó tartalmaz olyan átadásokat is, amelyek az ellenfél kapujától távolabb helyezik a csapatot és veszélytelenebbé tesznek egy támadási kísérletet. Ez a modern mutató ezzel ellentétben pedig azokat a mélységi passzokat vizsgálja, amelyek az ellenfél kapujához közelebb juttatják a labdát és veszélyeztetik az ellenfél hálóját. A passzokkal és azok minőségi értékelésével Bransen és szerzőtársai (2019) hosszasan foglalkozott kutatásában és betekintést

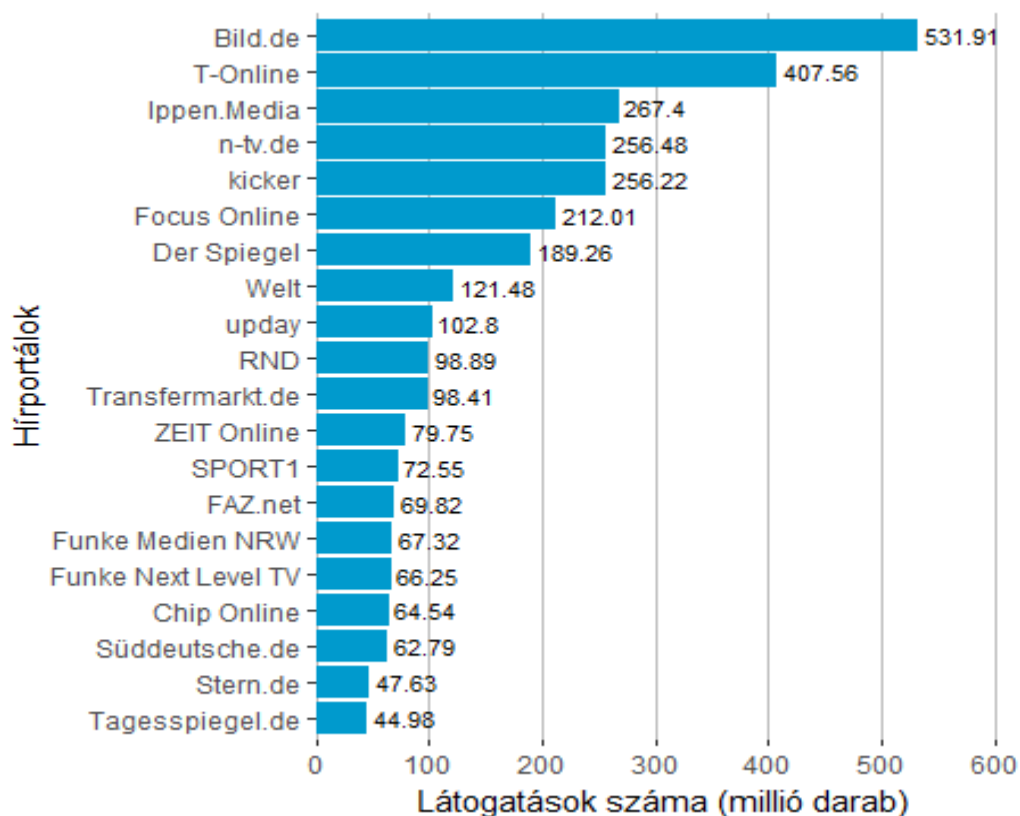
nyújt ezen érdekes terület témájába. Ezen információk alapján megfogalmazom a második kutatási kérdésemhez kapcsolódó nullhipotézisem, mely szerint a modern statisztikai mutatókon alapuló modellek egyenként pontosabb becslést és nagyobb magyarázóerőt fognak biztosítani, mint a hagyományos változópárjukon alapuló modellé. Ezzel ellentétben az alternatív hipotézisem az, hogy a modern statisztikai mutatókon alapuló modellek egyenként nem fognak pontosabb becslést és nagyobb magyarázóerőt biztosítani, mint a hagyományos változópárjukra épített modell.



## 2. Szakirodalom

### *2.1: A Transfermarkt piaci értékek fontossága*

A futballban a Transfermarkt piaci értékek könnyű elérhetőségük miatt nagyon népszerűek, így rengeteg portál és a labdarúgás világában dolgozó személy használja aktívan a becsült értékeket. Az edzők is ezekkel az értékekkel kalkulálnak, amikor egy játékos átigazolását vagy cseréjét fontolgatják. Az újságírók a mérkőzések előtt sokszor foglalkoznak a keretek összértékeivel vagy éppen az átigazolási időszakokban a potenciális transferekkel és akkor is ezek a becslések az iránymutatók számukra. A szakértők a televíziós adások vagy éppen podcastek során szintén ezekre a számadatokra hivatkoznak játékosok értékelése vagy összehasonlítása során. A klubok számára is elengedhetetlenek ezek az adatok, mivel a játékoskeretük alapján lesznek értékelve a többi csapathoz képest. Egy klub továbbá a pénzügyi helyzetének javításán is tud segíteni azzal, ha egy játékosuk magasra van értékelve a piacon és őt értékesítik, hogy bevételi forráshoz jussanak. Bár azt fontos megemlíteni, hogy a sportban a játékosokat számviteli értelemben eszközként kezelni egy kétséges és kevésbé vizsgált téma (Adiwiyana és szerzőtársai, 2021). Az átigazolások során a játékosügynökök erre az értékre hivatkoznak, hogy az érdeklődő kluboknál jobb feltételeket érjenek el a szóban forgó képviseltjeinek (Peeters, 2017; Herm és szerzőtársai, 2014; Coates és Parshakov, 2021). Emellett érdemes megemlíteni, hogy a hétköznapi emberek is előszeretettel használják a Transfermarkt által biztosított adatokat, hogy megnézhessék, hogy az általuk kedvelt vagy éppen érdekesnek talált játékos hogyan teljesít a piacon a jelen helyzetben. A 2. ábrán, az IVW (2023) adatai szerint látható, hogy Németországban 2023 októberében az 98,41 millió látogatást kapott a weboldal, amellyel a hónap 11-dik és a Kicker után a sportportálok második leglátogatottabb forrása volt. Szóval látható, hogy nagyon fontos szerepet játszik a német weboldal a labdarúgás életében és elengedhetetlenek az általuk biztosított adatok mind a szurkolók, mind pedig az ezzel foglalkozó személyek számára.



**2. ábra: Németországban a hírportálok online és mobil látogatásainak száma 2023. októberében**

*Forrás: IVW (2023). Saját szerkesztés.*

## **2.2: A Transfermarkt és a tömegértékelés hitelességének az alapjai**

Amint már az előző bekezdésben olvasható volt, a Transfermarkt német weboldal adatai nagyon népszerűek, viszont ehhez az is kell, hogy a tömeg által becsült értékek valóban hitelesek legyenek és ne legyenek részrehajlások az értékelések során. A tömeg által hozott döntések témában az alapokat Surowiecki (2005) tette le a könyvében, aki meghatározta a szükséges fogalmakat és definiálta a kollektív tudást és annak értékeit. A tömeges, csoportos értékelések nem csak a futball, hanem a gazdasági, társadalmi élet minden területén megjelentek, ahol lehetőség volt adatok becslésére vagy a kimenetek megjósolására. Az egyének és a csoportok döntéshozatalát vizsgálva azok az eredmények születtek, hogy utóbbiak racionálisabb és pontosabb döntést tudnak hozni, mint az egyének (Charness és Sutter, 2012). A tömeg hitelességét igazoló egyik ilyen kutatás a gazdasági életben a seekingalpha.com oldalhoz kötődik, amely egy népszerű véleményfórumként funkcionált és funkcionál napjainkban is. Itt kimutatták a szerzők a felhasználók szöveges üzeneteinek az elemzése alapján, hogy a jövőbeli részvényhozamok szempontjából előrejelző erővel bírnak a tömeg által létrehozott

bejegyzések. Ebből tehát az a konklúzió, hogy a pénzügyi piacokon segítő szerepük van a felhasználóknak (Chen és szerzőtársai, 2014). Egy további példa lehet a tömeg hitelességét illetően a kickstarter.com oldal esete. Ezen az oldalon a regisztrált felhasználók projekteket, ötleteket tudnak támogatni pénzügyileg részesedésért, termékért vagy egyéb választható juttatásért cserébe. A kutatás célpontjában a színházi előadások és azok sikerességének az előrejelzése volt. Ezt az oldalt vizsgálva arra jutottak a szerzők, hogy az oldal használói jobb előrejelzői egy színházi előadás sikerének, mint egy szűkebb létszámú szakértői csoport (Mollick és Nanda, 2015).

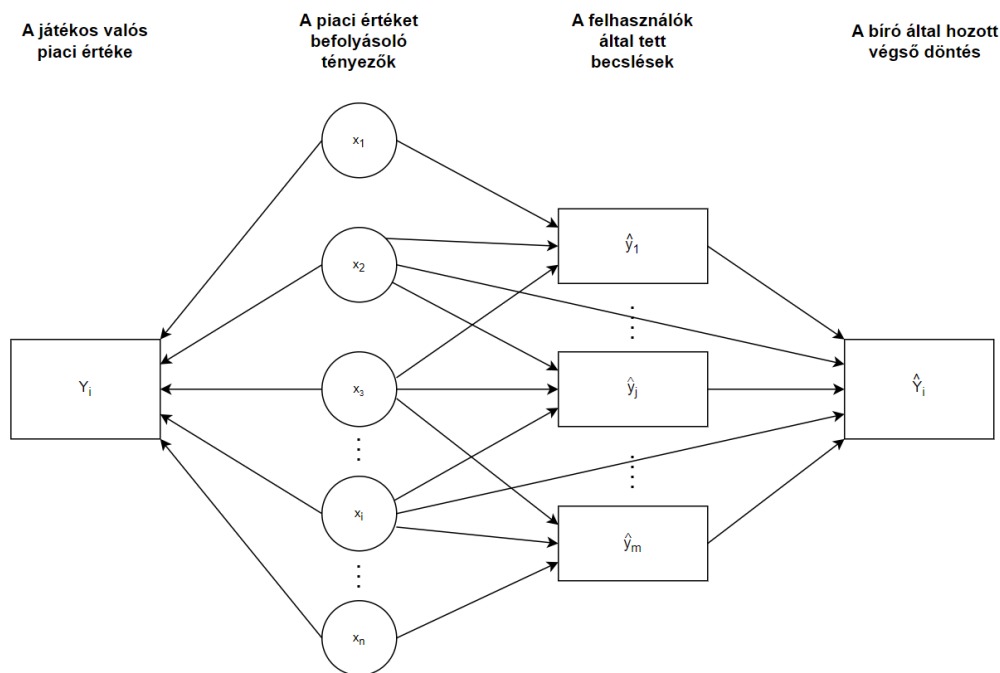
Végső soron pedig napjainkra már rengeteg kutatás alátámasztja a Transfermarkt piaci értékelések hitelességét és azt, hogy ebben a témában valóban használhatóak az általuk biztosított adatok. Herm szerzőtársai. (2014) kutatásában a becsült piaci értékek 90%-ban magyarázzák a játékosok átigazolási díjait. Ez az eredmény azért fontos, mert a piaci értékek mindig elérhetőek, az átigazolási díjakkal szemben, amelyek csak akkor merülnek fel, amikor egy játékost meg akar venni egy másik klub és tárgyalásokat folytatnak a játékos ügynökével vagy csapatával. Az átigazolási díjak nem elérhetőek minden játékosnál, sokkal ritkábban szolgáltat adatot, az esetleges részrehajlás veszélye is fennáll vagy éppen a szerződés lejártával hamis értéket biztosít. Továbbá fontos megjegyezni, hogy az előbbi értékeket eredményváltozóként használva pontosabb modelleket tudtak létrehozni a szerzők, mintha az átigazolási díjakat vették volna alapul (Franceschi, és szerzőtársai, 2023). A piaci értékek elérhetősége és bizonyított hitelessége miatt a témával foglalkozó szakirodalmak a Transfermarkt adatait tekintik egy játékos valós értékének a kutatásaik során (Müller és szerzőtársai, 2017; Herm és szerzőtársai, 2014; Peeters, 2017; He és szerzőtársai, 2015; Stanojevic és Gyarmati, 2016). A Transfermarkt oldal felhasználóinak a független döntéshozatalával és részrehajlásával is foglalkozott kutatás, amely azt vizsgálta, hogy a felhasználók magasabb árat tanácsolnak-e akkor, ha a kedvenc csapatuk játékosáról van szó. A szerző megállapította, hogy a használt adatok alapján nincsen részrehajlás a kedvenc csapat játékosait illetően és kizárható a részrehajlás az értékelés során (Peeters, 2017). Ezen források és eredmények figyelembevételével a kutatásomban a futballisták valós piaci értékét a Transfermarkt által biztosított adatokkal teszem egyenlővé és ez a változó lesz majd a kutatásom eredményváltozója.

### ***2.3: A Transfermarkt oldal bemutatása és az értékelések működésének alapjai***

A weboldal 2001-ben lett létrehozva Németországban és 2009 óta elérhető az angol nyelvű verziója is. Az oldal átfogóan tartalmazza az elmúlt és a jelenlegi bajnokságok mérkőzésadatait, klubértékeit, átigazolási híreket, játékosok adatait és természetesen a játékosok piaci értékeit. Az oldal azért jött létre, mivel a szakértők véleménye helyett a felhasználók gondolatait szerették volna előtérbe helyezni (Müller és szerzőtársai, 2017).

A játékosok piaci értékeinek az értékelését a felhasználók végzik, amelyeket az oldalon kijelölt fórumokon tehetnek meg. Bárki felhasználóvá tud válni, ehhez csak annyit kell tennie, hogy regisztrál a megfelelő adatok megadásával, amelyet ingyenesen tehet meg. A regisztrációhoz név, e-mail cím és az oldal használati feltételeinek az elfogadása szükséges. A regisztrációt követően a felhasználók elkezdhetik a fórumokon megosztani a gondolataikat a játékosok piaci értékeivel kapcsolatban vagy becsatlakozhatnak már folyamatban lévő megbeszélésekbe, gondolatkifejtésekbe. Ezek a vélemények a weboldalon aggregálódnak, összegyűjtésre kerülnek és évente kétszer meghatározásra, módosításra kerülnek a játékosok piaci értékei. Mindkét értékelés az átigazolási időszakokban történik, az egyik a téliben, a másik a nyáriban. Nyáron fejeződnek be a bajnokságok, így ekkor a szezon végi teljesítmények alapján, télen pedig a szezon felénél az addigi teljesítmények alapján tudnak módosítani vagy meghatározni egy új értéket a játékosok számára a felhasználók. Egyes esetekben, például egy új fiatal játékos esetén ezeken az időszakon kívüli értékelések is megtörténhetnek. A publikált értékek nem egy közgazdasági modell, átlag vagy medián alapján kerülnek meghatározásra, hanem kijelölt döntéshozók, úgynevezett bírok alapján. Ez a döntéshozatali forma tehát nem demokratikus alapon, hanem hierarchikus módon valósul meg. A Transfermarkt meghatároz egy változólistát, amelyet ajánlott, de nem kötelező figyelembe venni a felhasználóknak a javaslatuk megtételekor. A bírok ezeket a javaslatokat aggregálják és mindenki mástól függetlenül, de a vélemények figyelembevételével meghatároznak egy becsült piaci értéket (Herm és szerzőtársai, 2014).

A bírók szerepe kisebb, amikor egy közepes, népszerűbb vagy top bajnokság játékosának az értékét kell meghatározni, mivel rengeteg a felhasználói vélemény és ezek alapján egy átfogóbb képet kap a döntéséhez. Ezzel ellenben egy kevésbé népszerű bajnokság játékosainál, például a magyar NB1 esetében nagyobb a szerepük, mivel kevesebb a felhasználói vélemény és ezeknek a javaslatoknak a megfelelő súlyozásával és kiszűrésével kell meghozniuk a döntésüket a piaci érték meghatározásakor (Müller és szerzőtársai, 2017). A Transfermarkt játékosértékelésének könnyebb megértését az 3. ábra biztosítja.



### 3. ábra: A Transfermarkt piaci értékelések működése

Forrás: Herm és szerzőtársai (2014), Müller és szerzőtársai (2017). Saját szerkesztés.

#### 2.4: Elért eredmények

Ebben a bekezdésben a labdarúgás világában a játékosok Transfermarkt piaci értékeivel foglalkozó szakirodalmakat és az általam legfontosabbnak ítélt eredményeket fogom bemutatni. A kutatásomra a legnagyobb hatással Müller és szerzőtársai (2017) cikke volt. Ebben az elemzésben azt vizsgálták, hogy a játékosok által nyújtott statisztikák milyen hatással voltak a Transfermarkt piaci értékekre a top 5 bajnokságokban a 2009/10-2014/15-ös szezonig. A vizsgált magyarázóváltozókat 3 kategóriára osztották, amelyek rendre a következők: játékosok adottságai, játékosok teljesítménye és játékosok népszerűsége. Az első kategóriába tartozott az életkor, magasság, kétlábasság (azonosan tudja-e használni mindkét lábát) változók, a második csoportba a szezonokban elért teljesítményekkel, statisztikákkal kapcsolatos változók és a végső kategóriába pedig a Wikipedia oldal keresések, a Reddit cikkek és a YouTube videók száma. A modelleket OLS regresszióval hozták létre és kezdetben csak az első kategóriát tartalmazták, majd mindig a következő modellnél a soron következő csoporttal bővítették azt. Így születtek azok az eredmények, hogy a végső modell bírt a legkedvezőbb AIC, BIC értékekkel. Fontos megemlíteni, hogy a piaci értékeket és a népszerűségi mutatókat az erős jobbra elnyúlás miatt logaritmálták, illetve az életkor változónak a kvadrátikus tagját használták a modellben, mivel az nem korrelált lineárisan az eredményváltozóval. A végső

modellben a népszerűségi változók rendre szignifikánsak voltak, az első csoportból az életkor, a másodiktól pedig csak a gól, gólpasz, passz, labdavezetés, szerelés, sárga lap változók bizonyultak szignifikánsnak. A szakirodalom azonosította a Transfermarkt limitációit is, miszerint ezek az értékek csak 6-12 havonta elérhetőek, a fiatal, kevésbé ismert játékosoknál hiányosságot véltek felfedezni és mivel ezek az adatok bárki számára elérhetőek, így nem jelentenek versenyelőnyt a klubok számára (Müller és szerzőtársai, 2017).

A témában készült egy szisztematikus összefoglaló elemzés is, amely arra a kérdésre kereste a választ, hogy melyek a leghasználtabb változók és ezek közül melyek szignifikánsak, amikor a játékosok piaci értéke vagy az átigazolási díjuk volt az eredményváltozó. A vizsgált cikkek 3-4 évtizedes időszakot ölelnek fel és nagyjából 3000 cikket olvastak át, amelyből kiválasztották a számukra megfelelő 29 szakirodalmat. A leggyakrabban használt változókat 3 csoportba sorolták, amelyek a következők voltak: teljesítményalapú változók, játékosok adottságai, klub karakterisztikái. A többi változók ritkábbak, így azokat nem klaszterezték. Ezek alapján a 10 leggyakoribb változók között azonosították az életkor, életkor négyzete, gól, gólpasz, sárga lap, piros lap, játékkercek, mérkőzések száma, felnőtt nemzetiség dummy és kétlábás dummy változókat. A vizsgált 29 cikkben 111 modellspecifikációt azonosítottak, melyből 94-ben OLS regressziót alkalmaztak. A szakirodalmak eredményeinek a koefficienseit és p-értékeit hasonlították össze és azt kapták eredményül, hogy a sárga-piros lap és kétlábás dummy változók adatai ugyan könnyen elérhetőek, de nem szignifikánsak, így nem érdemes ezeket alkalmazni a modellekben. A maradék 7 változónak a használata ajánlott, mivel rendre szignifikánsak és pozitívan hatnak a piaci értékre, az életkor és az életkor négyzete változók kivételével. A szerzők megemlítik, hogy a modern világban az adatok egyre könnyebben elérhetőek és érdemes lehet olyan változók vizsgálata, mint például a népszerűségi változók vagy a szerződés hossza, hátralévő ideje változók (Francheschi és szerzőtársai, 2023).

A Top 5 bajnokságokon kívül a kisebb bajnokságokat is vizsgálták szerzők a kutatásaikban. Az egyik szakirodalom az MLS amerikai futballbajnokságot vizsgálta a 2006-2015-ös szezonig. A kutatási kérdés középpontjában az állt, hogy mennyire pontosak a Transfermarkt piaci értékek becslései és ezek az adatok milyen összhangban vannak a játékosok fizetéseivel. Az eredményváltozó a piaci érték logaritmus volt, a magyarázóváltozók pedig nem tartalmaztak semmilyen teljesítményalapú változókat, csak a csapatokat, a szezonokat, az életkort, annak négyzetét, köbét és az erősebbik láb változót. A kutatás eredményeként megállapították azt, hogy a piaci érték egy jó proxyja a fizetéseknek, legyen az jelenlegi vagy jövőbeli. Szerintük egyre jobban növekvő szerepe van a német oldal adatainak a futball gazdasági világában és

előszeretettel ajánlják annak használatát (Prockl és Frick, 2018). Egy másik kutatás a török bajnokságot vizsgálta a 2018-19-es szezonban. A kutatás témája azt vizsgálta, hogy az Artificial Neural Network (ANN) modellel hogyan lehet megbecsülni a játékosok piaci értékét. A modell output rétege a Transfermarkt piaci értéket tartalmazta, az input rétege pedig csak teljesítménymutatókat, amelyek rendre a következők voltak: játszott percek, gól, gólpassz, xG, xA számadatai és passz, kapura lövés, szerelés, fejpárbaj százalékos változók. A cikk a középpályásokat vizsgálta az említett időszakban és a teljesítménymutatókat az OPTA Sports oldalról használták fel. A szerzők csak azokat a játékosokat vizsgálták, akik megfeleltek a pozíció kikötésnek, illetve legalább 90 percet játszottak a bajnokságban a szezon alatt. A kutatás eredményeként azt kapták az írók, hogy pusztán a teljesítménymutatók és a piaci érték között gyenge a korreláció, így az OPTA adatokból nem lehet felépíteni ANN modellt, amely kapcsolatban állna a piaci értékkel. Ezek után készítettek egy szimulált adatbázist is, amely a minimum és a maximum között random értékeket generált a játékosok számára, majd ezek közül kiválasztottak véletlenszerűen 60 megfigyelést. Az így kapott eredmények sikeresek voltak, mivel a szimulált adatbázis már erősen korrelált a piaci értékkel és így fel tudtak építeni egy ANN modellt, amely már kapcsolatban állt a Transfermarkt adataival. A szerzők megemlítik a sérülékenység változó figyelembevételét a jövőben és az OPTA Sports adatainak a megbízhatóságát is dokumentálják. A kutatás a végső bekezdésben egy szélsőséges kijelentést tesz, miszerint bármilyen modellről legyen szó, a piaci értékeket csak a teljesítményváltozókkal kellene vizsgálni, mert a nézők ezeket látják a pályán, nem pedig a követők számát, az életkort vagy egyéb változók adatait (Inan és Cavas, 2021). Egy további kutatás az Indonéz ligában azt vizsgálta, hogy melyik változók magyarázzák legjobban a bajnokság játékosainak a piaci értékeit. A vizsgált szezon a 2017-es volt, amelyben 207 játékost gyűjtöttek össze, akik a 15 legjobb csapatba tartoztak és támadók, szélsők vagy középpályások voltak. A magyarázóváltozók közé tartoztak az életkor, gól, gólpassz, sárga, piros lapok száma, játszott percek, kezdett mérkőzések száma, külföldi nemzetiség dummy és top 4 csapat dummy változók. A kutatásban OLS regressziót használtak és Stepwise analízis módszerrel határozták meg a legvégén a legjobb modellt. Az így kapott eredmények eltérőek voltak a top 5 bajnokságot vizsgált szakcikkek eredményeitől. Az életkor, gólpassz, sárga lap, top 4 csapat és külföldi dummy változók mind szignifikánsak, a többi magyarázóváltozó pedig nem voltak azok. A szakirodalmak eredményeitől való különbözőség okát a szerzők az Indonéz liga eltérő erősségében látták, mivel ez a bajnokság jóval gyengébb az 5 legerősebbnek ítélt európai bajnokságtól (Adiwayana és Harymawan, 2021).

A témában voltak olyan szakirodalmak is, amelyek a teljesítménymutatók helyett kifejezetten a népszerűségi mutatókat és azok hatását vizsgálták a piaci értékre vetítve. Ebben a bekezdésben ezeket az általam relevánsnak vélt kutatásokat szeretném bemutatni. Az első ilyen cikk, amelyet megemlítenék, azt vizsgálta, hogy a közösségi média aktivitás milyen hatással van a sportolók értékére, pénzügyi helyzetére. A szerzők a 2013/14-es szezonban a 95 legértékesebb piaci értékű játékost vizsgálták. A magyarázóváltozók az életkor, gól, gólpassz, Instagram és Twitter követők száma volt. Az alacsony mintaszám miatt Qualitative Comparative Analysis-t (QCA) végeztek a szerzők és azt az eredményt kapták, hogy míg az idősebb játékosoknak minden változó mentén jól kell teljesíteniük, addig a fiatalok a gyengébb teljesítményt kompenzálni tudják a közösségi hálókön elért aktivitással, hogy szinten tartsák a magas piaci értéküket (Korzynski és Paniagua, 2016). Egy következő cikk a Top 5 bajnokságot vizsgálta, a 2018/19-es szezonban, 1428 megfigyeléssel. 43 változót használtak, amelyeket 4 csoportba osztottak: játékos adottságai, játékos teljesítménye, munka és klubfeltételek, illetve népszerűségi mutatók. A kutatás azt vizsgálta, hogy a Google trendek, keresések száma hogyan befolyásolja egy játékos piaci értékét. A szerzők nem OLS regressziót használtak, hanem Multiple Linear Regression-t (MLR), Random Forest-et (RF) és Gradient Boost Method-ot (GBM). Az eredmények részeként megemlítik, hogy a népszerűségi mutatók jelentős változók és hatással vannak a játékosok piaci értékére, így használni lehet ezeket a későbbi kutatásokban (Malagón-Selma és szerzőtársai, 2023). Egy harmadik szakirodalom, amely a közösségi médiával foglalkozott, az a Bundesliga 2017/18-as szezonját vizsgálta 600 megfigyelésen keresztül. Két hipotézist fogalmazott meg, az első nullhipotézise az volt, hogy a kor hatással van a közösségi média aktivitásra, a második szerint pedig a követők száma hatással van a piaci értékre. A változók nem tartalmaztak teljesítménymutatókat, csak a népszerűségi mutatókat vette figyelembe, amelyek a Twitter, Instagram és Facebook követők száma volt. A kutatás eredményeként az első nullhipotézist elutasította, a másodikat pedig elfogadta. Ezek alapján tehát nincs kapcsolat az életkor és a közösségi média aktivitás között, viszont a platformokon elért követők száma szignifikánsan hatással voltak a piaci értékre. Ezek a változók gyenge magyarázóerővel bírtak, de ennek az volt az oka, hogy további változó kategóriákat nem alkalmaztak. A szerzők dokumentálják, hogy a népszerűségi változók vizsgálata lényeges és érdemes használni őket, de megemlítik azt is, hogy a jövőbeli kutatásoknál figyelembe lehetne venni a posztok tartalmát és a követők visszajelzéseit, reakcióit (Frenger és szerzőtársai, 2019). Egy másik számomra érdekes kutatás azt vizsgálta, hogy az EA Sports által készített FIFA 18-as játék adatai, a Wikipedia keresések száma és az előző évi teljesítménymutatók hogyan befolyásolják a játékosok értékét. A vizsgált időszak 2017 augusztusától 2018 januárjáig tartott,



ahol a szerzők különböző modellek mentén tesztelték a változók magyarázóerejét. A kutatás eredményeként az OLS és Ridge regressziót modelljeit állapították meg a legjobbnak R-négyzet és átlagos négyzetes hiba mentén. Mindkét modell mentén megállapították, hogy a Wikipedia keresések száma és a FIFA játék adatai szignifikáns változók és jelentős hatással vannak a piaci értékre tekintve (Singh és Lamba, 2019).

A kutatásomra Herm és szerzőtársai (2014) irodalma is nagy hatással volt. Ők a Bundesliga 2011/12-es szezonban azt vizsgálták, hogy mennyire tudják pontosan előrejelezni a piaci értékek a valós átigazolási díjakat, illetve, hogy melyek a legfontosabb változók, amelyek a piaci értéket magyarázzák. A változókat 2 kategóriára osztották, amelyek a játékos tehetsége és a külső tényező csoportok voltak. Előbbibe tartoztak az életkor, passzpontosság, gólok, gólpasszok száma, párharcok, illetve a kétlábás dummy változók. Utóbbi változók közé tartoztak az átlagos média osztályzatok, átlagos játékosügynök portfólió összege, klub játékoskeretének értéke, játszott meccsek száma és a Google keresések száma. Fontos megemlíteni, hogy az eredményváltozót logaritmálták. A piaci értékek előrejelzésénél 67 megfigyelést vizsgáltak, mivel ennyi átigazolás volt az évben, a változók magyarázóerejének a vizsgálatánál pedig 338 játékost figyeltek meg. Mindkét esetben OLS regressziót alkalmaztak és utóbbinál a heteroszkedaszticitás miatt White-féle korrigált standard hibákkal számoltak. A kutatás eredményeként azt kapták, hogy a Transfermarkt adatok 90%-ban magyarázzák az átigazolási díjakat és a vizsgált magyarázóváltozók pedig 70%-ban a piaci értékeket. A szerzők megemlítik, hogy ezzel a 2 kategóriával nagyon jól lehet becsülni az eredményváltozót és érdemes használni ezeket a jövőben. Kitérnek arra is, hogy a kategóriák száma ugyan jó, de a változókat bővíteni kellene, különösen a külső tényezők kategóriánál, mivel azok nagy magyarázóerővel bírnak. Az első cikként hivatkozik magára, aki a játékosügynök változót használja a kutatásban ilyen célra (Herm és szerzőtársai, 2014).

A piaci értékekkel kapcsolatban készültek olyan kutatások is a témában, akik azt vizsgálták, hogy a játékosértékek mennyire képesek előrejelezni a meccsek végkimeneteleit. Az első ilyen szakirodalom, amelyet bemutatnék azt vizsgálta, hogy a Transfermarkt értékekkel mennyire lehet megjósolni egy nemzetközi meccs kimenetelét, illetve pontosabb becsléseket biztosít-e, mint a FIFA vagy ELO sorrend. A vizsgált időszak 2008-2014-ig terjedt és 1020 hivatalos nemzetközi mérkőzést vizsgált. A használt változók között volt a piaci érték, a csapat játékosainak száma, győzelem, döntetlen, vereség oddsai, külön-külön, az ELO pontok, a FIFA rangsor helyezés és a hazai dummy változó, amely azt mutatta, hogy melyik csapat rendezte a mérkőzést. A kutatásban OLS és LOGIT modelleket hasonlított össze a szerző és azt az

eredményt kapta, hogy egy szimpla modell, amely a játékosok értékeit, a játékosok számát és a hazai dummy változót, az pontosabb becslést ad a mérkőzés kimenetelére, mintha az ELO vagy FIFA rangsort venné számításba. A szerző tehát ajánlja a piaci értékek használatát, mivel megbízhatóbb és pontosabb becslést tudnak biztosítani egy mérkőzés kimenetelére, mint más változó társai (Peeters, 2017).

Egy további szakirodalom szintén a játékosok teljesítményét vizsgálta a piaci értékre vetítve. A vizsgált időszak a 2014/15-ös La Liga spanyol bajnoksága volt és a támadók voltak a kutatás középpontjában. Ennek megfelelően a vizsgált változók is ehhez a pozícióhoz mérten lettek kiválasztva. A használt változók tartalmazták a faultok, tizenhatoson belüli és kívüli gólok, kaput találó lövések, sikeres labdavezetések és tizenhatoson belüli lövések számát. A szerzők megállapították, hogy minél jobb egy játékos teljesítménye, annál nagyobb a piaci értéke. Észrevették, hogy van egy plafon piaci érték, a játékosoknál, amelyet, ha elér egy játékos, akkor hiába javul a teljesítménye, az érték ezt nem követi le ennek a hatását. Ezt az érdekességet a top játékosoknál állapították meg, ahol nagyon hasonlóak voltak a piaci értékek, de a teljesítménymutatók eltérőek voltak (He és szerzőtársai, 2015)

Az előző szakirodalmaktól elrugaszkodva született olyan kutatás is, ami nem a teljesítménymutatókat vagy a népszerűségi mutatókat, hanem a relatív életkort helyezte a középpontba és ennek a változónak a kapcsolatát vizsgálták a piaci értékkel. A relatív életkor azt feltételezi, hogy két ember születhetett ugyanabban az évben, viszont. A cikk azt vizsgálta, hogy valóban magasabbak-e azoknak a játékosoknak a piaci értékei, akik az első negyedévben születtek. A relatív életkor hatás tehát azt vizsgálja, hogy az azonos évben születettek között van-e különbség a szerint, hogy ki melyik negyedévben született, jobbak-e a kilátásaik a jövőre nézve, csak ezért, mert ők a naptári évben korábban születtek. A kutatás a top 5 bajnokságokat vizsgálta 2008-2017-ig. A használt változók tartalmazták a piaci értéket, a születési dátumot, a születési negyedévet, életkort, szezon, ligát, illetve pozíciót. A kutatás eredményeként azt állapították meg a szerzők, hogy egy egyenesen arányos relatív életkor hatás figyelhető meg, tehát aki a korábbi negyedévben született, annak magasabb a piaci értéke. Ez igazolja tehát, hogy számít az, hogy ki melyik negyedévben született és jobbak a kilátásai azoknak a játékosoknak, akik például januárban születtek, azokkal szemben, akik például augusztusban. Megállapították azt is, hogy ennek következtében a relatív fiatalabb játékosok alulértékeltnek lehetnek és érdemes őket leigazolni fiatal korban. Az életkor növekedésével szűkül a negyedévek közötti különbség. A cikk megemlíti, hogy ezeket a különbségeket úgy lehetne

csökkenteni, hogy az utánpótlásokban nem a születési évek alapján lennének a korosztályok kialakítva, hanem negyedévek összevonása alapján (Gyimesi és Kehl, 2021).

Egy további szakirodalom azzal foglalkozott, hogy a top 5 bajnokságok és az európai kupasorozatok hogyan befolyásolják a játékosok piaci értékét. A kutatás a 2017/18-as szezon bajnokságait vizsgálta, 2259 játékos figyelembevételével és tartalmazták a piaci értéket, a születési negyedéveket, az életkor, a liga és az európai kupasorozatok kategoriális változókat. A szerzők OLS regressziót alkalmaztak és azokat az eredményeket kapták, hogy a liga és az európai kupasorozatok szignifikánsan befolyásolják egy játékos piaci értékét. Ezek közül megállapították, hogy a Premier League-ben szereplő játékosok szignifikánsan magasabb értékkel rendelkeztek, mint a spanyol, német, olasz és francia elsőosztályú bajnokságokban szereplő játékosok. Ugyanez a hatás volt megfigyelhető a Bajnokok Ligája versenysorozat esetében, azok a játékosok, akik szerepeltek ebben a kupasorozatban, azok szignifikánsan magasabb értékkel rendelkeztek, mint azok, akik az Európa Ligában szerepeltek vagy nem voltak részesei egyetlen európai kupasorozatnak sem. A szerző megemlíti, hogy előbbinek az oka az is lehet, hogy az angol bajnokság csapatai rendelkeznek a legmagasabb televíziós jogdíjakkal és ennek köszönhetően a csapatok átlagosan magasabb értékű játékosokat tudnak megvásárolni, mint a többi ország bajnokságának a klubjai (Felipe és szerzőtársai, 2017)

### 3. Kutatási módszertan

A kutatásban a modellépítések során rendre OLS regressziót fogok alkalmazni. Ez a módszer megmutatja, hogy a magyarázóváltozók milyen hatással vannak az eredményváltozóra és mekkora szignifikancia tartozik ezekhez, tehát hogyan állnak meg a helyüket a mintán kívüli világban, azaz a sokaságban (Hunyadi és Vita, 2008).

A modell magyarázóerejét a transzformált eredményváltozó miatt nem az R-négyzetek figyelembevételével, hanem a korrelációs index (I) értékének a vizsgálatával fogom megtenni. Ez a mutató 0-1 közötti értékeket vehet fel és minél közelebb található az 1-hez, a változók annál jobban magyarázzák a vizsgált eredményváltozót (Kerékgyártó és szerzőtársai, 2009).

A második kutatási kérdésemhez a modellek összehasonlításához az említett korrelációs index mutató mellett az AIC és BIC információs kritériumok vizsgálatát is figyelembe fogom venni. A mutatók önmagukban nem értelmezhetők, csak modellek összehasonlításakor. Mindkét mutató esetében akkor tekint egy modellt jobbnak, pontosabbnak, ha az információs kritérium értéke alacsonyabb, mint a másik mutatónak (Jánosa, 2015).

A konzisztencia és a torzítatlanság érdekében multikollinearitás és heteroszkedaszticitás vizsgálatokat fogok elvégezni a modelleken. Ha a kettő közül bármelyik fennáll, akkor azokat a megfelelő módokon fogom kezelni a szakirodalomnak megfelelően (White, 1980; Belsley és szerzőtársai, 2005).

## 4. Adatok bemutatása

### 4.1: A Transfermarkt ajánlott változói

Az első kutatási kérdésemhez kapcsolódóan láthatóak az 1. táblázat első oszlopában az a 16 legfontosabb tényező, amelyeket a Transfermarkt a felhasználóinak kiindulási alapként, ajánlasként meghatározott. Ezen kívül a német weboldal 2 további csoportban is meghatároz változókat, de azokat ezen kutatásban nem vizsgálom, mivel azok egyéni, speciális esetekre vonatkozó változókat tartalmaznak (Transfermarkt, 2021). A második oszlopban láthatóak azok a változók, amelyeket képes voltam számszerűsíteni vagy kategorizálni és ennek eredményeképpen felhasználni a kutatásomban. Ezek azok a változók, amelyekkel vizsgálni fogom a játékosok piaci értékét. A változók kifejtése és bemutatása a következő bekezdésben lesz olvasható.

**1. táblázat: A Transfermarkt által javasolt tényezők és a hozzájuk kapcsolt változók**

<b>Transfermarkt ajánlások legfontosabb tényezői</b>	<b>Általam számszerűsített változó</b>
1. Jövőbeli tehetség	FIFA 23: 23 év alatti dummy
2. Életkor	Életkor
3. Teljesítmény a klubban és a válogatottban	Bajnokságban elért teljesítménymutatók
4. A liga szintje és státusza piaci és sportszempontról	Top 5 bajnokság kategória
5. Hírnév/elismertség	EA FC 24: nemzetközi népszerűség
6. Fejlődési potenciál	23 év alatti dummy
7. Bajnokság specifikus sajátosság, szabály	<i>Nincsen ilyen vizsgált változó</i>
8. Marketing érték	Wikipedia keresések száma
9. Érdeklődő klubok száma és hírneve	<i>Nincsen ilyen vizsgált változó</i>
10. Potenciális teljesítmény	<i>Nincsen ilyen vizsgált változó</i>

11. Tapasztalat	Életkor
12. Sérülékenység	FIFA 23: Sérülékenység dummy
13. A bajnokságok és klubok eltérő pénzügyi helyzete	<i>Nincsen ilyen vizsgált változó</i>
14. Általános kereslet, piaci trend a játékosokat illetően	<i>Nincsen ilyen vizsgált változó</i>
15. Átigazolási díjak fejlődése	<i>Nincsen ilyen vizsgált változó</i>
16. Egyéb külső tényezők: COVID-19	<i>Nincsen ilyen vizsgált változó</i>

---

*Forrás: Transfermarkt (2021). Saját szerkesztés. Megjegyzés: Dőlttel jelölve, amely tényezőkhöz nem tudtam változót párosítani*

#### 4.2: A használt változók bemutatása

A kutatáshoz használt adatbázis a 2022/23-as szezon top 5 bajnoksága alapján tartalmaz adatokat a középpályásokról. A használt adatok és a források, ahonnan azokat kigyűjtöttem a 2. táblázat összefoglalásában látható. A táblázat továbbá tartalmazza azokat a szakirodalmakat, amelyek mentén döntöttem az adott változó beveteléről.

**2. táblázat: A használt változók bemutatása**

Változó	Változó neve R-ben	Rövid leírás	Forrás	Hivatkozás
Piaci érték (2023)	market_value_in_eur_y	A játékos piaci értéke a 2022/23-as szezon befejezte után	Transfermarkt	(1), (3)-(17)
Életkor	age	A játékos életkora 2023. június 30-án.	FBref	(1), (3)-(5), (7), (9), (12), (13), (15), (17)
Piaci érték (2022)	market_value_2022	A játékos piaci értéke a 2022/23-as szezon megkezdése előtt.	Transfermarkt	(13), (16)
Bajnokság	comp_level	Az öt legerősebbnek ítélt európai elsőosztályú bajnokságok: angol, spanyol, német, olasz, francia.	FBref	(4), (7)
Játszott percek	minutes_pct	Játszott percek aránya az elméletileg lejátszható percekből.	FBref	(1), (5), (9), (10), (13), (17)
Gólok	goals_per90	Gólok száma 90 percre vetítve.	FBref	(1), (3), (5), (8)-(13), (17)
Gólpasszok	assists_per90	Gólpasszok száma 90 percre vetítve	FBref	(1), (3), (5), (8)-(13), (17)

Várható gólok (xG)	xg_per90	A játékos a helyzete alapján mennyi gólt szerezhetett volna. 90 percre vetítve.	FBref	(10)
Várható gólpasszok (xA)	xg_a_per90	A játékosnak mennyi gólpasszal kellene rendelkeznie lövést megelőző átadásai alapján. 90 percre vetítve.	FBref	(1), (10)
Összes passzok	passes_per90	Sikeres és sikertelen passzok száma összesen. 90 percre vetítve.	FBref	(13), (17)
Sikeres passzok	passes_pct	Sikeres passzok pontosságának aránya.	FBref	(9), (13), (17)
Győztes fejpárbaj	aerials_won_pct	Győztes fejpárbajok aránya.	FBref	(10), (12), (17)
Progresszív passzok száma	progressive_passes_per90	Progresszív passzok száma. 90 percre vetítve.	FBref	(1), (10), (12)
Progresszív passzok aránya	prog_pass_pct	Progresszív passzok aránya, az összes passzt tekintve. Sikeres, sikertelenek egyaránt. 90 percre vetítve.	FBref	(1), (10), (12)
Cselezési kísérletek	take_ons_per90	Labdavezetéssel való elhaladási kísérlet egy másik játékos mellett. 90 percre vetítve.	FBref	(8), (12), (13), (17)
Szerelési kísérletek	challenge_tackles_pct	Egy másik játékos labdavezetését megakadályozó sikeres szerelési kísérletek aránya.	FBref	(10), (13), (17)
Nemzetközi hírnév	international_reputation	Az EA FC 24 által meghatározott nemzetközi népszerűség. 1-5-ös skálán vehet fel értéket.	EA FC 24	(16)
Sérülékenység	injury_prone	A játékos sérülékenyebb típus a többi társához képest.	FIFA 23	(10)
Jövőbeli tehetség	future_prospect	A 23 év alatti játékosok átlaghoz mért fejlődési potenciáját mutatja. Igen és nem értéket vehet fel.	FIFA 23	(5)
Wikipedia keresések	total_pageviews	Wikipédia keresések száma a játékosra 2022. május 1. és 2023. július 31 között.	Pageview Analysis	(5), (6), (9), (11)-(13), (16)

*Saját szerkesztés. Hivatkozások: (1) Adiwiyana és Harymawan (2021); (2) Bransen, Van Haaren, és Van de Velden (2019); (3) Coates és Parshakov (2021); (4) Felipe és szerzőtársai (2020); (5) Franceschi és szerzőtársai (2023); (6) Frenger és szerzőtársai (2019); (7) Gyimesi és Kehl (2021); (8) He és szerzőtársai (2015); (9) Herm és szerzőtársai (2014); (10) Inan és Cavas (2021); (11) Korzynski és Paniagua (2016); (12) Malagón-Selma és szerzőtársai (2023); (13) Müller és szerzőtársai (2017); (14) Peeters (2017); (15) Prockl és Frick (2018); (16) Singh és Lamba (2019); (17) Stanojevic és Gyarmati (2016)*

A vizsgált bajnokságok középpályásainak a játékosneveit, akkori életkorokat és a hozzájuk tartozó ligabeli teljesítménymutatókat az FBRef oldalról gyűjtöttem be. Itt néhány kikötés szerint szereztem be az adatokat. Az első és legfontosabb az, hogy amint már említettem, a vizsgált játékosnak a Top 5 bajnokság valamelyikében kellett játszani, mivel ezek a ligák könnyen összehasonlíthatóak, nagy számban elérhetőek és megbízhatóak a rendelkezésre álló adatok. A következő kikötés az volt, hogy az adott szezont egy csapatnál kellett eltölteni és végig játszani a középpályásnak, szóval az adatbázis nem tartalmaz olyan játékost, aki télen átigazolt egyik vizsgált bajnokságból a másik 4 vizsgált bajnokság valamelyikébe. A harmadik és egyben utolsó kritérium a kiválasztásnál pedig az volt, hogy az adott labdarúgónak legalább 900 percet kellett eltöltenie a vizsgált szezonban a pályán. Ez a mennyiség nagyjából 10 egész meccs játékidejének felel meg és ez már elegendő adatot biztosít egy játékosról, anélkül, hogy lefelé vagy felfelé kilógó outlier értéket generáljon az adatbázisban. Ennek köszönhetően 683 középpályást sikerült leszűrni, viszont ebből további 1-et el kellett távolítanom, mivel hiányzott egy megfigyelt teljesítménymutatója, így 682 megfigyeléssel folytattam tovább a munkát.

Az adatbázis eredményváltozója a játékosok piaci értéke lett, amely adatokat a Transfermarkt oldalról gyűjtöttem össze. A 2022/23 szezon végi piaci értékek a 2023. április 1. – augusztus 10. közötti adatok közül a legkorábbiakat tartalmazza. Az adatbázis 95%-nál ez az intervallum május-június közé esik, mert a 2022-2023-es szezon után ilyenkor értékelik újra a játékosokat az oldalon, de vannak olyan fiatal vagy kevésbé ismert játékosok, akiknél szükség volt egy tágabb intervallumban nézni a megfigyelt adatokat.

A teljesítménymutatóknál a játékosperceket a bajnokságok különböző meccsszáma miatt korrigálnom kellett. Ez azt jelenti, hogy a Bundesliga kivételével, minden bajnokságban 20 csapatos rendszert alkalmaznak a mai napig (a Ligue 1-ben ez volt az utolsó ilyen menetrend) és a németeknél pedig csak 18 csapat verseng a bajnokságban. Ezért én az összes futballpályán töltött percet elosztottam az összes mérkőzés perceinek a számával és egy százalékos értéket kaptam. Minden mérkőzésnél csak a rendes játékidővel, azaz 90 perccel számoltam, amely az angol, spanyol, olasz, francia bajnokságoknál 38 meccset a németeknél pedig 34 mérkőzést jelent. Utóbbiakban 3420 perc, előbbiben pedig 3060 perc volt a pályán tölthető maximális elméleti mennyiség. Az így említett százalékos változót az oldal által biztosított egyéni mutatók felhasználásával számoltam ki. A hosszabbításokat nem vettem figyelembe, mivel a rendelkezésre álló adatok is csak a 90 percből a pályán töltött adatokat tartalmazza.



A játékosokhoz tartozó előző szezon végi értékeket szintén a Transfermarkt oldalról gyűjtöttem össze. Itt az összes játékosnak maradéktalanul megtalálható volt minden vizsgált megfigyelése. A tavalyi piaci értékek 2022. április 1. – augusztus 10. közötti adatok közül a legkorábbit tartalmazza. Itt szintén elmondható az, mint az eredményváltozónál, hogy egyes játékosoknál szükség volt tágabb intervallumot megvizsgálni.

A nemzetközi hírnév, sérülékenység és jövőbeli tehetség változókhoz tartozó adatokat az EA Sports FIFA 23, illetve EA FC 24 videójátékok alapján társítottam a játékosokhoz. Ezek a videójátékok mindig szeptember végén jelennek meg és az új szezon kezdete előtti adatok alapján értékelik újra a játékosokat. Tehát a 2021-22-es szezon teljesítményei alapján korrigálják a játékosok adatait, amelyek a FIFA 23 videójátékban láthatóak. Természetesen az EA FC 24 is hasonló módszer alapján készült el, ott a 2022-23-as szezon teljesítményei alapján korrigálták a játékosok jellemzőit. A nemzetközi hírnév változó az utóbbi említett videójáték adatbázisából lett kigyűjtve, mivel az tartalmazza a vizsgált szezon alatt elért nemzetközi ismertséget. A FIFA 23-as játék tartalmazta utoljára a játékosokról a sérülékenység jellemzőt, ezért itt ezeket az adatokat használtam fel. Természetesen 1-es kódolást kapott az, aki a játék szerint sérülékeny típus és 0-s kódolást az, aki nem tartozik ebbe a kategóriába. A jövőbeli tehetség változónál a cél az volt, hogy minél objektívebben közelítsem meg és számszerűsítsem az eredményeket. Ennél a változónál két kritériumnak kellett megfelelnie egy középpályásnak, hogy jövőbeli tehetségnek számítsa. Először is én minden olyan labdarúgónál vizsgáltam ezt a jellemzőt, aki a 24 éves kort nem töltötte be a szezon kezdetéig, tehát maximum 23 éves volt a bajnokság indulása pillanatakor. Ennek csupán az az oka, hogy az utolsó életkor szerinti korosztály az U23 és ezután már csak a felnőtt korcsoport a következő állomás és ez még egy utolsó lépcsőfok a teljes érettség eléréséhez. A következő lépés az volt, hogy a FIFA 23 által biztosított alapértékelések és potenciális jövőbeli értékelések közötti különbségeket kiszámoltam. Ezeket az eredményeket átlagoltam és akik ezen számérték felett helyezkedtek el, azok 1-es kódolást, akik pedig ez alatt, azok 0-s kódolást kaptak. Az 1-es kódolású játékosok az átlagosnál nagyobb jövő előtt állnak, tehát fiatal tehetségek, a 0-s kódolásúak pedig az átlagosnál kisebb jövő előtt állnak, tehát nem tekinthetőek a kritériumok alapján jövőbeli tehetségeknek. Azok a játékosok, akik nem teljesítik az első, életkorhoz kapcsolódó kritériumot, ők is 0-s kódolást kaptak, mert őket már kész, érett játékosnak tekintem a kutatásomban.

Az utolsó változó, amelyről említést kell még tennem, az a Wikipédia keresések száma. Ezen számadatokat a Wikipedia Pageview oldal által biztosított adatok mentén gyűjtöttem ki. Az

időintervallum, amelyen az összes keresések számát vizsgáltam, az a 2022-23-as szezon kezdetének első hónap első napjától az utolsó hónap utolsó napjáig tartott. Ez 2022. augusztus 1. és 2023. május 31. közötti időtartamot jelenti. Ez a számadat tartalmazza az összes Platformról érkező keresések számát, amelyet a Wikipedia felajánl (Asztali, Mobilalkalmazás, Mobil web), illetve csak a valós felhasználók által indított műveleteket vizsgáltam, tehát az eredmények nem tartalmazzák a keresőmotorok vagy egyéb más robotizált eszközök kereséseit.

#### 4.3.: Leíró statisztika

A használt változókhoz tartozó leíró statisztika a 3. táblázatban látható. A táblázat adataiból látható, hogy a top 5 bajnokságban egy középpályás átlagos piaci értéke nagyjából 16,2 millió euró és ehhez egy igen magas, átlagot meghaladó 19,2 millió eurós szórás tartozik. Ez és a magas ferdeségi mutató már arra utal, hogy az eredményváltozó balra ferde, jobbra elnyúló. Ez nem meglepő, hiszen a pénzügyi adatoknál ez egy gyakran előforduló jelenség. Ugyanez vehető észre a 2022-es piaci értékeknél is, ott egy körülbelül 14,6 millió eurós átlaghoz 17,5 millió eurós szórás és szintén magas ferdeségi mutató társul. Egy szintén hasonló jelenség figyelhető meg a Wikipedia keresések számánál, amelynél egy játékosra átlagosan körülbelül 266 ezerszer kerestek rá, amelyhez 1,04 milliós szórás és magas a ferdeségi mutató. A többi változónál nem figyelhető meg említésre méltó érdekesség, nincsen nem normális eloszlásra utaló jel.

**3. táblázat: leíró statisztika a numerikus változókra.**

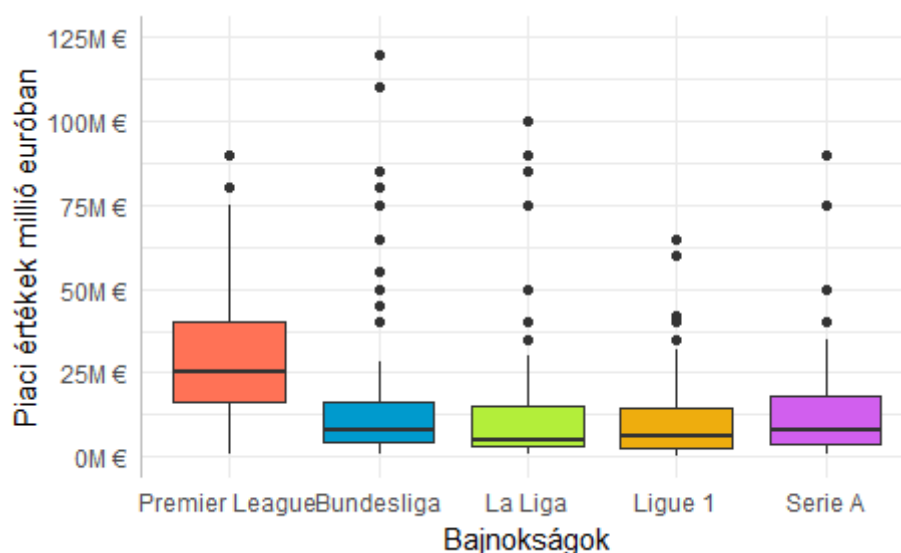
Változó neve	Átlag	Szórás	Medián	Minimum	Maximum	Terjedelem	Ferdeség	Csúcsosság
Piaci érték (2023)	16133284	19211411	9000000	200000	120000000	119800000	2,23	5,68
Életkor	25,89	4,00	25,50	16,00	37,00	21,00	0,26	-0,34
Játszott percek	0,57	0,19	0,56	0,26	0,99	0,73	0,21	-1,06
Gólok	0,15	0,14	0,12	0,00	0,82	0,82	1,30	2,08
Gólpasszok	0,12	0,11	0,10	0,00	0,64	0,64	1,29	2,22
xG	0,15	0,12	0,12	0,00	0,68	0,68	1,20	1,38
xA	0,13	0,08	0,11	0,00	0,51	0,51	1,20	1,87
Összes passzok	10,54	5,39	9,27	2,22	33,19	30,97	0,98	0,69
Passzpontosság	0,78	0,08	0,79	0,50	0,94	0,44	-0,52	0,03
Győztes fejpárbajok	0,42	0,14	0,42	0,00	0,79	0,79	-0,20	-0,29

Progresszív passzok	1,06	0,63	0,90	0,09	3,57	3,48	1,20	1,44
Progresszív passzok aránya	0,29	0,05	0,29	0,16	0,51	0,36	0,23	0,37
Piaci érték (2022)	14555938	17471256	8000000	50000	85000000	84950000	1,98	3,73
Cselezési kísérletek	0,53	0,38	0,43	0,01	2,04	2,03	1,24	1,51
Szerelési kísérletek	0,42	0,13	0,42	0,00	0,79	0,79	-0,12	0,19
Wikipedia keresések	265955	1039208	47467	220	23624981	23624761	17,43	375,33

Megfigyelések száma: 682

*Saját szerkesztés.*

A numerikus adatokon kívül megvizsgáltam még a bajnokságokon belüli piaci értékek hogyan alakulnak, melynek eredményei a 4. ábrán láthatóak. Ezen az ábrán látható, hogy az angol elsőosztály, a Premier League játékosai szemmel láthatóan értékesebbek, mint a másik 4 vizsgált bajnokságé. A másik 4 liga játékosértékeinek az eloszlása ránézésre azonosnak mondható. Ezek alapján az lesz a feltételezésem, hogy a Premier League az átlagnál szignifikánsan magasabb értéket fognak képviselni. A modellépítés során majd kiderül, hogy valóban jogosak lesznek-e a feltételezéseim.

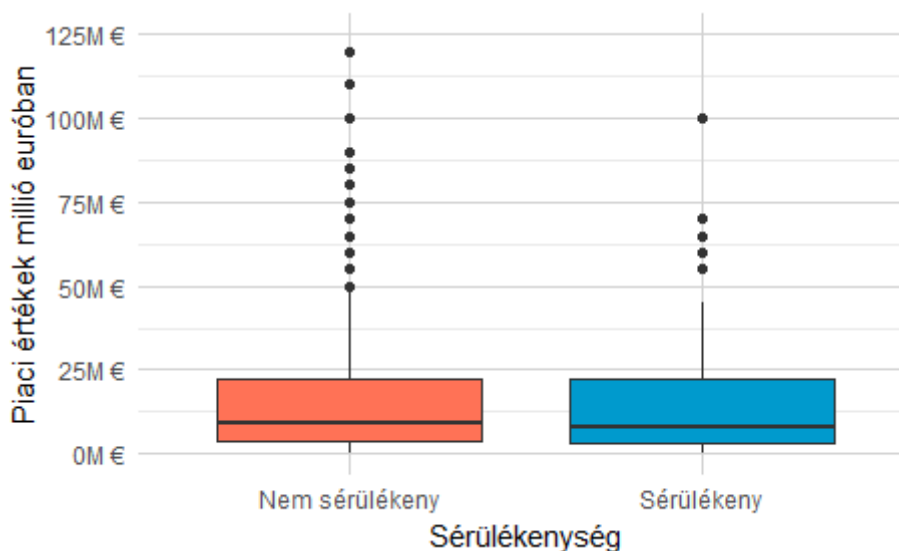


**4. ábra: A bajnokságok játékosértékeinek dobozábrája**

*Saját szerkesztés.*

Ezen kívül továbbá megvizsgáltam a sérülékenység dummy változónál, hogy a modell felépítése előtt előzetesen látható-e valamilyen kapcsolat a változó és a piaci érték között.

Ennek eredményei az 5. ábrán láthatóak. Itt meg lehet figyelni, hogy szinte semmilyen különbség nem lelhető fel a sérülékeny és a nem sérülékeny játékosok dobozábrái között. Ezek alapján az a feltételezésem, hogy nem lesz szignifikáns kapcsolat a sérülékenységi változó és a piaci érték között.

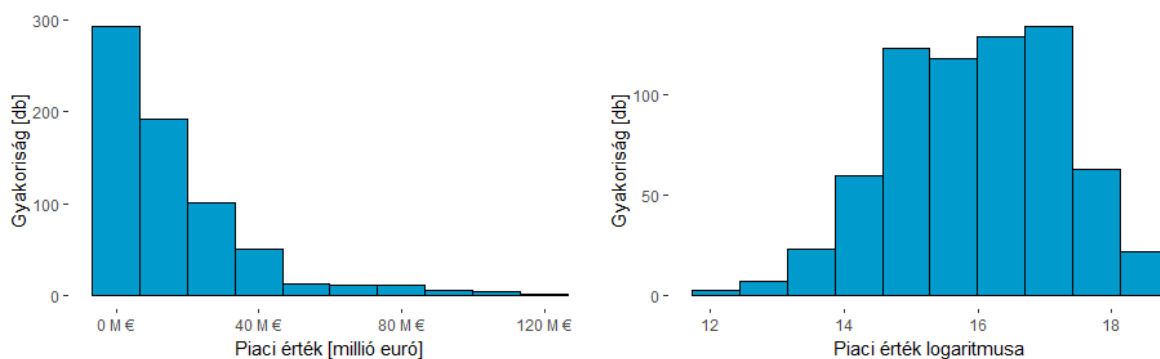


**5. ábra: A sérülékenységi változó dobozábrája**

Saját szerkesztés.

#### 4.4.: Szükséges változók transzformálása

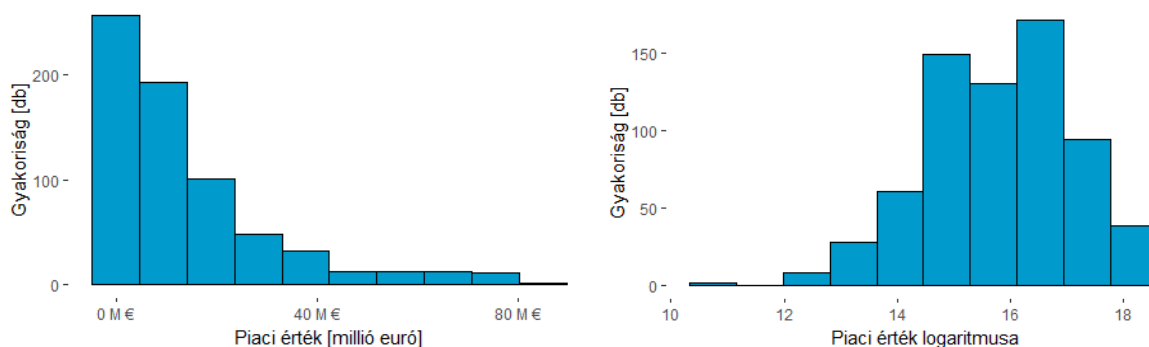
Az előző 3 bekezdésben említett változó eloszlását hisztogramon is megvizsgálom, hogy vizuálisan is leellenőrizsem a transzformációk szükségességét. Az eredményváltozó normális eloszlása egy modellfeltevés, ezt ugyan kiváltja a nagy mintaelemszám, de a modell pontosságát így is javítja, ha ez a feltétel is teljesül mellette. A 6. ábrán látható, hogy a piaci értékek valóban durván jobbra elnyúlóak és a logaritmálással ezt kezelni lehet és egy normális eloszláshoz közelítő változót kaptam. Továbbá fontos megemlíteni, hogy számos szakcikk is észrevette ezt a problémát és a piaci érték logaritmált adatait használták (például Müller és szerzőtársai, 2017; Gyimesi és Kehl, 2021; Frenger és szerzőtársai, 2019). Ezen információk figyelembevételével én is logaritmálom az eredményváltozót és ezt fogom használni a modell építése során.



**6.ábra: A 2023-as piaci érték eloszlása az eredeti és a logaritmált adatokkal**

*Saját szerkesztés.*

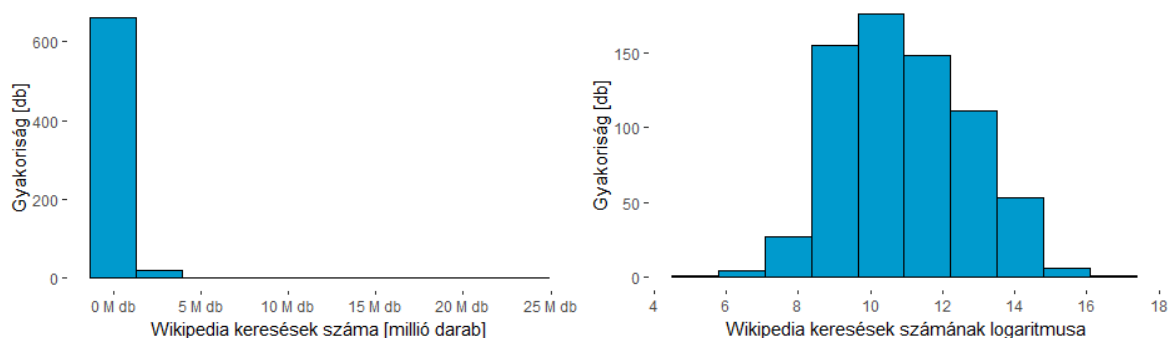
A 2022-es piaci értékeknél is megvizsgáltam a transzformált és az az előtti hisztogramokat és a 7. ábra alapján itt is látható, hogy a logaritmált változó eloszlása normális eloszlást követ. Szintén megvizsgáltam egyes szakcikkeket, akik alkalmazták ezen változó használatát és akik az eredményváltozót logaritmálták, azok hasonlóan jártak el ezen változó mentén is (Müller és szerzőtársai, 2017; Singh és Lamba). Így ezen információkat megvizsgálva én azt a döntést hoztam, hogy transzformálom a tavalyi piaci értékek adatait és ezt fogom használni a modellépítés során.



**7.ábra: A 2022-as piaci érték eloszlása az eredeti és a logaritmált adatokkal**

*Saját szerkesztés.*

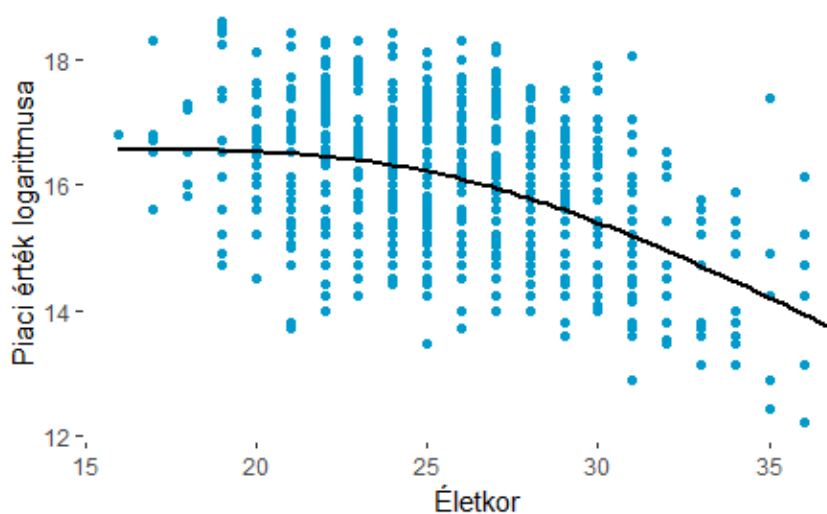
Végző soron, pedig megvizsgáltam a Wikipedia keresések számát is a hisztogramokon és itt is beigazolódott a 8. ábra alapján, hogy a transzformált adatok normális eloszlást követnek. Szintén ellenőriztem a szakirodalmakat és a népszerűségi mutatókat vizsgáló cikkek vegyesen kezelték ezt a problémát. Az egyik csoport nem logaritmálta az adatokat (Herm és szerzőtársai, 2014), a másik viszont igen (Müller és szerzőtársai, 2017). A kutatásomban én a logaritmált adatokat fogom alkalmazni, mivel ha már két másik magyarázóváltozót is logaritmáltam, akkor itt is ezt az elvet fogom követni.



**8.ábra: A Wikipedia keresések száma az eredeti és a logaritmált adatokkal**

*Saját szerkesztés.*

A leíró statisztikából ugyan nem látszódik, de az előző 3 változón kívül egy negyedik vizsgálata is szükséges, ez pedig az életkor. Számos cikk felfedezte, hogy az életkor és a piaci érték között nem található lineáris kapcsolat (például: Franchesci és szerzőtársai, 2023; Gyimesi és Kehl, 2021) és ennek érdekében a kvadratikus tag használatára is szükség lesz a modellépítést követően. A 9. ábrán látható, hogy szükség van az életkor transzformálására, mivel fordított parabolára hasonlít, így én ezt megteszem és használni fogom ezen transzformált változót a modellépítést követően.



**9.ábra: Az életkor kvadratikus tagjának kapcsolata az piaci értékkel**

*Saját szerkesztés.*

## 5. Eredmények

### 5.1.: Első kutatási eredményei

Az adatok bemutatása és transzformálása után OLS regresszióval megvizsgáltam a felépített változókból készült modelleket. A modell heteroszkedasztikus (1. melléklet és 2. melléklet), így a standard hibákat White-féle korrigált módszerrel számoltam újra és a változók szerinti standard hibák, t-teszték és p-értékei ezen számítás követően láthatóak a 4. táblázatban. A modellem 86,63%-os korrigált R-négyzettel rendelkezik, de fontos megemlíteni, hogy mivel az eredményváltozó nem-lineáris, így sok következtetést nem lehet ebből levonni. Ennek következtében a modellre kiszámoltam a korrelációs index mutatószámot, amely egy nagyon erős 0,933-as értéket mutatott. Ezek után már bátran kijelenthetem, hogy a Transfermarkt által javasolt legfontosabb változók erős magyarázóerővel bírnak a piaci érték meghatározásakor, a felhasználók valóban ezek a változók alapján teszik meg a javaslataikat a bíró, döntéshozó felé a kijelölt fórumokon. A modellben nem található multikollinearitás, a változók 5 alatti VIF értékeket vettek fel, amely azt jelenti, hogy zavaró multikollinearitás sem áll fent (3. melléklet). Ez alól egyetlen kivételt képez az életkor és az életkor-négyzete változó, amely 192-es és 182-es VIF értéket vettek fel, viszont ez strukturált multikollinearitás, amelyet én hoztam létre így itt nincs szükség ennek a kezelésére.

**4. táblázat: Transfermarkt által javasolt változók és a piaci érték OLS regressziós eredményei**

<b>Magyarázóváltozók</b>	<b>Modell eredményei</b>
age	0.045 (0.085)
age_square	-0.003** (0.002)
comp_level1	0.321*** (0.049)
comp_level2	-0.056 (0.036)
comp_level3	-0.006 (0.040)
comp_level4	-0.221*** (0.040)

minutes_pct	0.313** (0.160)
goals_per90	0.939*** (0.165)
assists_per90	0.923*** (0.197)
passes_per90	0.040*** (0.007)
passes_pct	1.090*** (0.318)
aerials_won_pct	0.133 (0.171)
log_market_value_2022	0.385*** (0.050)
take_ons_per90	0.065 (0.062)
challenge_tackles_pct	-0.159 (0.146)
international_reputation1	-0.117* (0.071)
international_reputation2	0.026 (0.055)
international_reputation3	0.062 (0.061)
international_reputation4	0.171** (0.075)
injury_prone1	0.005 (0.054)
future_prospect1	0.251** (0.102)
log_total_pageviews	0.113*** (0.026)
Constant	7.913*** (1.267)
Használt változók száma	16
R-négyzet	0.8706
Korrigált R-négyzet	0.8663
Korrelációs index	0,9331



### Megfigyelések száma: 682

*Saját szerkesztés. Megjegyzés: \*  $p < 0,1$ , \*\*  $p < 0,05$ , \*\*\*  $p < 0,01$ , a standard hibák zárójelben szerepelnek.*

Az életkor kvadratikus tagja, a játszott percek százaléka, a 90 percre vetített gólok, gólpasszok és passzok száma, a passzpontosság, a logaritmált tavalyi piaci értékek, a fiatal tehetség és a logaritmált Wikipedia keresések mind szignifikáns változóknak tekinthetők a modellem eredményei alapján. Ezek közül az életkor kivételével minden változó pozitív hatással van a játékosok piaci értékére. A bajnokságok és nemzetközi hírnév kategorikus változóknál találhatóak szignifikáns kategóriák. Mindkét esetben kontraszt kódolást alkalmaztam és az átlaghoz hasonlítottam. Az angol elsőosztályú bajnokság középpályásainak értéke magasabb, a francia bajnokság játékosainak az értéke pedig alacsonyabb az átlaghoz képest 1%-os szignifikanciaszinten. Akik az EA FC 24 videójáték alapján 1-es nemzetközi hírnévvel rendelkeznek, ők 10%-os szignifikanciaszinten az átlagnál alacsonyabb, akik pedig 4-es nemzetközi hírnévvel rendelkeznek, azok 5%-os szignifikanciaszinten az átlagnál magasabb piaci értékkel rendelkeznek. A győztes fejpárbajok százaléka, a cselezések, a szerelési kísérletek és a sérülékenység változók egyik szokásos vizsgált alfán sem tekinthetők szignifikánsnak.

A leíró statisztikánál megfigyelt elvárások is beigazolódtak, mely szerint a Premier League játékosainak a piaci értéke szignifikánsan magasabb, mint a többi bajnokságé, illetve a sérülékenység változó nem biztosít szignifikáns hatást a piaci értékre nézve.

A modellhez tartozó általános regressziós egyenletet is elkészítettem, a koefficiensekhez tartozó pontos értékek 4. táblázatban láthatóak. A Transfermarkt által javasolt tényezőkre épített modell regressziós egyenlete a következőképpen alakul:

$$\begin{aligned} Y_{\text{Log Piaci Érték}} = & B_0 + B_1 * X_{\text{Életkor}} + B_2 * X_{\text{Életkor}^2} + B_3 \\ & * X_{\text{Bajnokság}} + B_4 * X_{\text{Játszott percek}} + B_5 * X_{\text{Gólok}} + B_6 \\ & * X_{\text{Gólpasszok}} + B_7 * X_{\text{Összes passzok}} + B_8 \\ & * X_{\text{Sikeres passzok}} + B_9 * X_{\text{Győztes fejpárbajok}} + B_{10} \\ & * \ln(X_{\text{Piaci érték (2022)}}) + B_{11} * X_{\text{Cselezési kísérletek}} + B_{12} \\ & * X_{\text{Szerelési kísérletek}} + B_{13} * X_{\text{Nemzetközi hírnév}} + B_{14} * X_{\text{Sérülékenység}} + B_{15} \\ & * X_{\text{Jövőbeli tehetség}} + B_{16} * \ln(X_{\text{Wikipedia keresések}}) \end{aligned}$$

Más kutatási eredményekkel is összevetettem az eredményeimet és azt figyeltem meg, hogy az életkor kvadratikus tagjának, a játszott percek, gólok, gólpasszok, passzok, passzpontosság, előző szezon értékei, a Wikipedia keresések számai és a győztes fejpárbaj változók eredményei

azonosak, viszont a szerelési változó eredményei nem egyeznek Müller és szerzőtársai (2017) kutatásaival. Az említett szakirodalom számára ez a változó is szignifikáns, viszont nálam ez nem mondható el. Egy további kutatásban kapott eredményekkel szintén egyeznek az általam készített modell eredményei, mely szerint az életkor kvadratikus tagja szignifikánsan negatív, viszont a gólok, gólpasszok és játszott percek szignifikánsan pozitív hatással vannak a piaci értékre tekintve (Franchesci és szerzőtársai, 2023). A Gyimesi és Kehl (2021) és Prockl és Frick (2018) által vizsgált kutatásokban az életkor és az életkor kvadratikus tagjai is egyaránt szignifikánsak. Az én modellemben csak az utóbbi változó hatása bizonyult szignifikánsnak a piaci értéket vizsgálva. Adiwiyana és Harymawan (2021) kutatásában a gólok száma és játszott percek száma változók nem lettek szignifikánsak, amelyek az eredményeimmel nem egyeznek, viszont a gólpasszok száma változó a cikkben is szignifikánsan pozitív. Herm és szerzőtársai (2014) kutatásaival is összevettem az eredményeimet és a Wikipedia keresések száma, a passzpontosság és a gólok száma változók eredményei is megegyeznek.

Egyes kivételektől eltekintve tehát elmondható, hogy a szakirodalmakkal azonos eredményeket kaptam a változók szignifikanciáját és hatását illetően. Az eredményeket megvizsgálva egy megbízható és erős kapcsolat jött létre a Transfermarkt által ajánlott változó és a becsült piaci értékek között. Az első kutatási kérdésemhez kapcsolódó nullhipotézisemet tehát elfogadom és kijelenthetem, hogy van kapcsolat az ajánlott Transfermarkt változók és a felhasználók által becsült piaci érték között. A kutatásom további részében ezen változók mentén felépített modellemet tekintem az alapmodellemnek.

### ***5.2.: Második kutatási kérdés eredményei***

A kutatásom második vizsgált kérdésköre a modern adatelemzési mutatók vizsgálatához kapcsolódik. Ezek a mutatók az xG, xA, progresszív passzok 90 percre vetítve és a progresszív passzok aránya. Mind a 4 változó hatását és szignifikanciáját úgy vizsgálom, hogy kicserélem az alapmodellben található hasonló tartalmú változópárjára. Ennek következtében 3 új modellt fogok vizsgálni. A modelleket szintén OLS regresszióval vizsgálom és az AIC, BIC információs kritériumok és az I, korrelációs index mentén hasonlítom őket össze az alapmodell eredményeivel. Mindhárom új modellnél az alapmodellhez hasonlóan fennáll a heteroszkedaszticitás jelensége, így itt is White-féle korrigált módszerrel fogom számítani a standard hibákat és az ebből adódó t-teszteket és p-értékeket. Először az xG hatását vizsgáltam az alapmodellben úgy, hogy kicseréltem a változópárjára, ami az én esetemben a 90 percre vetített gólok száma. A következő modellben az xA, azaz a várható gólpasszok kicserélése történt a változópárjára, ami a 90 percre vetített gólpasszok száma. Majd legvégül kicseréltem

a 90 percre vetített passzok számát a 90 percre vetített progresszív passzok számával és a passzpontoság változót a progresszív passzok arányával. Ezen modellek eredményei az 5. táblázatban láthatóak. Az xG és progresszív passz mutatók bevételeével a multikollinearitás továbbra is az 5-ös VIF érték alatt voltak és még a zavaró hatást sem érték el, úgy, mint az alapmodellnél. Az xA mutató bevételeével viszont sérül ezen állítás, mivel a 90 percre vetített passzok száma minimálisan átlépi az 5-ös VIF érték határát (4. melléklet). Ez az érték 0,002-es átlépést jelent és már zavarónak tekinthető a multikollinearitás, viszont mivel ilyen közel van ehhez a határhoz és ennyire minimálisan haladja csak meg azt, így én nem tekintem ezt problémának és nem látom szükségesnek a kezelését. Mindhárom modell esetében az alapmodellhez hasonlóan szintén figyelmen kívül hagyom az általam készített strukturált multikollinearitást az életkor változókat tekintve.

**5. táblázat: A modellek összehasonlítása**

Magyarázóváltozók	Modellek eredményei			
	Alapmodell	Modell 2	Modell 3	Modell 4
age	0.045 (0.085)	0.047 (0.083)	0.038 (0.086)	0.048 (0.084)
age_square	-0.003** (0.002)	-0.003** (0.002)	-0.003* (0.002)	-0.003** (0.002)
comp_level1	0.321*** (0.049)	0.331*** (0.050)	0.310*** (0.049)	0.287*** (0.051)
comp_level2	-0.056 (0.036)	-0.036 (0.035)	-0.040 (0.036)	-0.075** (0.038)
comp_level3	-0.006 (0.040)	-0.018 (0.040)	-0.002 (0.041)	-0.005 (0.041)
comp_level4	-0.221*** (0.040)	-0.241*** (0.042)	-0.224*** (0.041)	-0.178*** (0.041)
minutes_pct	0.313** (0.160)	0.301* (0.162)	0.323* (0.172)	0.590*** (0.140)
goals_per90	0.939*** (0.165)		0.956*** (0.165)	0.649*** (0.168)
assists_per90	0.923*** (0.197)	0.956*** (0.199)		0.888*** (0.200)
xg_per90		1.105*** (0.212)		
xg_a_per90			0.880***	

			(0.302)	
passes_per90	0.040*** (0.007)	0.042*** (0.007)	0.040*** (0.007)	
passes_pct	1.090*** (0.318)	1.239*** (0.315)	1.093*** (0.354)	
aerials_won_pct	0.133 (0.171)	0.097 (0.172)	0.161 (0.172)	0.201 (0.175)
progressive_passes_per90				0.302*** (0.047)
prog_pass_pct				-1.306*** (0.419)
log_market_value_2022	0.385*** (0.050)	0.382*** (0.050)	0.385*** (0.051)	0.393*** (0.052)
take_ons_per90	0.065 (0.062)	0.069 (0.062)	0.061 (0.065)	-0.044 (0.060)
challenge_tackles_pct	-0.159 (0.146)	-0.163 (0.155)	-0.176 (0.150)	0.033 (0.154)
international_reputation1	-0.117* (0.071)	-0.115 (0.071)	-0.127* (0.076)	-0.122* (0.069)
international_reputation2	0.026 (0.055)	0.033 (0.055)	0.013 (0.061)	0.033 (0.051)
international_reputation3	0.062 (0.061)	0.060 (0.060)	0.059 (0.065)	0.105* (0.057)
international_reputation4	0.171** (0.075)	0.164** (0.075)	0.143* (0.078)	0.187** (0.076)
injury_prone1	0.005 (0.054)	0.034 (0.054)	0.008 (0.053)	-0.012 (0.056)
future_prospect1	0.251** (0.102)	0.260** (0.103)	0.258** (0.105)	0.252** (0.105)
log_total_pageviews	0.113*** (0.026)	0.108*** (0.027)	0.114*** (0.026)	0.127*** (0.027)
Constant	7.913*** (1.267)	7.846*** (1.259)	8.004*** (1.308)	8.744*** (1.186)
Használt változók száma	16	16	16	16
AIC	846.24	855.62	859.73	877.00
BIC	954.85	964.21	968.33	985.60
Korrelációs index	0.9331	0.9321	0.9317	0.9299

A modellek eredményei alapján látható, hogy mind a 4 modern adatelemzési mutató minden szokásos alfán szignifikáns a piaci érték becslésekor. Itt pontosan erre számítottam, mert ezek az újszerű statisztikák mélyebb betekintést nyújtanak a játékosok teljesítményét illetően. Ezek a változók a progresszív passzok arányának a kivételével mind pozitív hatást nyújtanak a piaci értékre nézve. Ez meglepő volt számomra, arra számítottam, hogyha valaki nagyobb arányban ad mélységi passzokat, az magasabb piaci értékkel rendelkezik. Ennek a negatív hatásnak talán az lehet az oka, hogy ez a mutató tartalmazza a sikertelen progresszív passzokat is. Ennek következtében minél nagyobb az aránya a mélységi passzoknak, annál több hibát tartalmazhat és feltehetőleg ez eredményezi a negatív hatást.

Az új modelleket összehasonlítottam az információs kritériumok és a korrelációs indexek értékei alapján. Ezen eredmények voltak számomra a legmeglepőbbek, mivel mindhárom esetben a várakozásaimmal ellentétes eredmények születtek. A legnagyobb magyarázóerővel az alapmodell bír, amely mind az AIC, mind a BIC információs kritériumok mentén a legalacsonyabb értékekkel rendelkezik és a korrelációs indexet tekintve pedig az 1-hez legközelebbi értékkel rendelkezik. Ezek alapján kijelenthető, hogy a piaci értékek meghatározása során a hagyományos mutatók, mint a gólok, gólpaszok és progresszív passzok pontosabb becslést biztosítanak, mint a modern statisztikai mutatókon alapuló párjaik. Ennek következtében tehát elvetem a második kutatási kérdésem nullhipotézisét, miszerint a modern statisztikai mutatók által becsült modellek nagyobb magyarázóerővel bírnak, mint a hagyományos mutatószámokon alapulók.

A várakozásaimmal ugyan ellentétes eredmények születtek, viszont ez nem jelenti azt, hogy nem születtek értékes eredmények. A hagyományos mutatóknál ugyan nem nyújtanak pontosabb becslést az újszerű statisztikai mutatók, viszont minden szokásos alfán szignifikánsak így kijelenthető az, hogy ezek a változók a későbbi kutatásokban is használhatóak és ha nem is azonos, de nagyon hasonló eredményeket lehet elérni ezekkel a változókkal egy OLS modell felépítése során.

## 6. Konklúziók

### 6.1.: Összefoglalás

A labdarúgás világában a Transfermarkt által publikált játékosértékek nagyon népszerűek és könnyen elérhetőek a hétköznapi emberek és szakemberek számára. A kutatásomban bemutattam ennek a német weboldalnak a működését, amely, ha nem is demokratikus módon működik, mégis képes egy hierarchikus struktúrával biztosítani azt, hogy a publikált piaci értékek megbízhatóak és használhatóak legyenek későbbi kutatások, mint például ezen dolgozat alapjának.

A kutatásom első kérdéskörként megvizsgáltam a Transfermarkt által javasolt, de nem kötelezően felhasználandó 16 legfontosabb tényezőt, amely kiindulási alapot biztosít az oldal felhasználóinak a játékosok piaci értékeinek meghatározásához. Ezek után kiválasztottam azokat a tényezőket, amelyekhez változókat tudtam kapcsolni és megvizsgáltam, hogy a Transfermarkt által javasolt tényezők mennyire magyarázzák pontosan a becsült piaci értékeket. A 16 tényezőből 9-hez tudtam változókat párosítani. A változók tartalmaztak adatokat a Transfermarkt és az FBRef.com oldalakról és az EA Sports által készített FIFA 23 és FC 24 videójátékaiból. A változók transzformálása után a 9 tényezőhöz 15 magyarázóváltozót tudtam kapcsolni. Ezek után lefuttattam az OLS regressziót és azokat az eredményeket kaptam, hogy a Transfermarkt által javasolt tényezők és a becsült piaci értékek között egy nagyon erős kapcsolat található, amelyet a korrelációs index 0,933-as értéke erősít meg.

A kutatásom második kérdésköre a modern statisztikai mutatókat és azok hatását vizsgálta a piaci értékre vetítve. Ezek az újszerű mutatók egy mélyebb elemzést adnak a játékosok teljesítményéről. Ezek alapján az volt a feltételezésem, hogy az xG, xA és progresszív passzal kapcsolatos modern statisztikai mutatók szignifikánsan pozitív hatással lesznek a piaci értékre és egyenként kicserélve a hagyományos mutatópárjukra, egy pontosabb modellt biztosítanak számomra. Ezek a várakozások csak részben teljesültek, mivel ugyan mindegyik változó szignifikáns lett a szokásos alfákon vizsgálva, viszont a progresszív passzok aránya negatív hatással volt a piaci értékre a modellemben. Ezen kívül a modern statisztikai mutatók alul szerepeltek a hagyományos mutatópárjukkal szemben, mind a 4 modellt figyelembe véve az alapmodell bír a legalacsonyabb AIC és BIC információs kritériumokkal és a legmagasabb korrelációs index mutatóval.

## **6.2.: Limitációk, kitekintés**

A kutatásomban vizsgált modellnek és a Transfermarkt adatoknak vannak limitációi és vannak olyan területek, amelyeket fejleszteni lehetne. Először is kezdeném a Transfermarkt adatokkal. Számos szakirodalom (Stanojevic és Gyarmati, 2016; Inan és Cavas, 2021; Müller és szerzőtársai, 2017) vizsgálja és megemlíti, hogy az adatalapú becselőmodellek pontosabbak és hasznosabbak lehetnek a Transfermarkt piaci becsléseknél, mivel ezek a közgazdasági modellek gyorsabban képesek reagálni a játékosok változó teljesítményére és 1-1 meccs után is képesek egy játékos értékét kalkulálni a német weboldallal szemben, ahol évente csak 2-szer történik értékelés. Ezen gondolatokkal én is egyetértek és a jövőben úgy gondolom szükség lenne egy ilyen publikus közgazdasági modell létrehozására, amely mindenki számára publikus és könnyen elérhető lenne.

Az első kutatási kérdésemnél a Transfermarkt által meghatározott legfontosabb tényezőket nem tudtam mind számszerűsíteni és úgy gondolom, hogy ezt a kutatást még lehetni javítani azzal, hogy a kihagyott 7 tényezőhöz változókat kapcsolok és ezzel is javítani lehetne már így is erős magyarázóerővel bíró modellemet. Továbbá a kutatási kérdést ki lehetne bővíteni a Transfermarkt által meghatározott másik 2 tényezőcsoportra is, amely az egyéni és szituációs csoportok ajánlott tényezőit tartalmazza. Úgy gondolom, hogy ezeknek a tényezőknek a változóvá alakítása egy még pontosabb és még nagyobb magyarázóerővel bíró modellt tudna létrehozni.

A második kutatási kérdésemnél az adatbázisom alapján azokat az eredményeket kaptam, hogy a modern statisztikai mutatók ugyan mind szignifikánsak, de alul szerepelnek, a hagyományos mutatókkal szemben. Itt fontos kiemelni, hogy én csak a középpályásokat vizsgáltam a top 5 legerősebbnek ítélt európai bajnokságokból. Ezek alapján további kutatást ki lehetne terjeszteni tágabb vizsgált időszakra, a támadókra, akiket hasonló változók alapján értékelnek vagy éppen másik bajnokságokra és megvizsgálni, hogy ott milyen eredmények születnének ugyanezek változók figyelembevételével.

Amint már említettem a modern statisztikai mutatók eredményeinél, attól mert még a hagyományos mutatópárjuknál gyengébb magyarázóerővel bírnak, az nem jelenti azt, hogy ezek a változók nem használhatóak. Ezek a mutatók is mind szignifikánsak és a jövőbeli kutatásokban is egy alapot biztosíthat ez a tanulmány ezeknek a változóknak a használatára.

## Irodalomjegyzék

### Szakirodalmak

- Adiwiyan, H. I., & Harymawan, I. (2021). Factors that determine the market value of professional football players in Indonesia. *Jurnal Dinamika Akuntansi*, 13(1), 51–61.
- Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). Regression diagnostics: Identifying influential data and sources of collinearity. Wiley.
- Bransen, L., Van Haaren, J., & Van de Velden, M. (2019). Measuring soccer players' contributions to chance creation by valuing their passes. *Journal of Quantitative Analysis in Sports*, 15(2).
- Chen, H., De, P., Hu, Y., & Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, 27(5), 1367–1403.
- Coates, D., & Parshakov, P. (2021). The wisdom of crowds and transfer market values. *European Journal of Operational Research*, 301(2), 523–534.
- Felipe, J. L., Fernandez-Luna, A., Burillo, P., de la Riva, L. E., Sanchez-Sanchez, J., & Garcia-Unanue, J. (2020). Money Talks: Team Variables and Player Positions that Most Influence the Market Value of Professional Male Footballers in Europe. *Sustainability*, 12(3709), 3709.
- Franceschi, M., Brocard, J.-F., Follert, F., & Gouguet, J.-J. (2023). Determinants of football players' valuation: a systematic review. *Journal of Economic Surveys*, 1–24.
- Frenger, M., Follert, F., Richau, L., & Emrich, E. (2019). Follow me ... on the relationship between social media activities and market values in the German Bundesliga. *Working Papers of the European Institute for Socioeconomics*, 32.
- Gyimesi, A., & Kehl, D. (2021). Relative age effect on the market value of elite European football players: a balanced sample approach. *European Sport Management Quarterly*, 23(4), 1-17.
- He, M., Cachucho, R., & Knobbe, A. (2015). Football player's performance and market value. Proceedings of the 2nd Workshop on Machine Learning and Data Mining for Sports Analytics, 87–95.
- Herm, S., Callsen-Bracker, H.-M., & Kreis, H. (2014). When the crowd evaluates soccer players' market values:



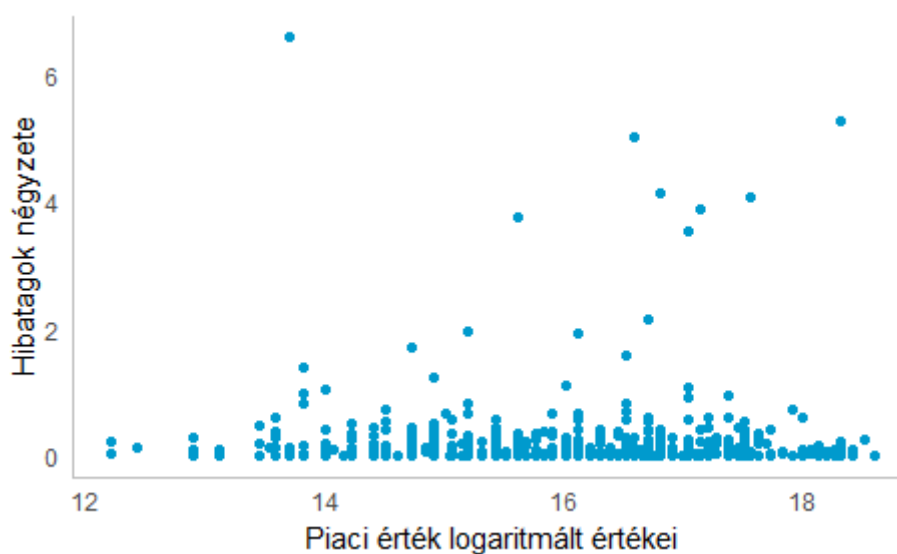
- Accuracy and evaluation attributes of an online community. *Sport Management Review* 17(4):484-492
- Hunyadi, L., & Vita, L. (2008). Statisztika II. Akadémiai Kiadó.
- Inan, T., & Cavas, L. (2021). Estimation of Market Values of Football Players through Artificial Neural Network: A Model Study from the Turkish Super League. *Applied Artificial Intelligence*, 35(13).
- Jánosa, A. (2015). Adatelemzés IBM SPSS Statistics megoldások alkalmazásával. Magyar Könyvvizsgálói Kamara Oktatási Központ Kft.
- Kerékgyártó Gy., L. Balogh I., Sugár A., & Szarvas B. (2009). Statisztikai módszerek és alkalmazásuk a gazdasági és társadalmi elemzésekben. Aula Kiadó.
- Korzynski, P., & Paniagua, J. (2016). Social media recipes for sports stars. *Business Horizons* 59(2):185-192
- Malagón-Selma, P., Debón, A., & Domenech, J. (2023). Measuring the popularity of football players with Google Trends. *PLoS ONE*, 18(8).
- Mollick, E., & Nanda, R. (2015). Wisdom or madness? Comparing crowds with expert evaluation in funding the arts. *Management Science*, 62(6), 1533–1553.
- Müller, O., Simons, A., & Weinmann, M. (2017). Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research*, 263(2), 611-624.
- Peeters, T. (2017) Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results
- Prockl, F., & Frick, B. (2018). Information precision in online communities: Player valuations on www.transfermarkt.de. *International Journal of Sport Finance*, 13(4), 319–335.
- Singh, P., & Lamba, P. S. (2019). Influence of crowdsourcing, popularity and previous year statistics in market value estimation of football players. *Journal of Discrete Mathematical Sciences and Cryptography*, 22(2), 113-126.
- Stanojevic, R., & Gyarmati, L. (2016). Towards data-driven football player assessment.
- Surowiecki, J. (2005). The wisdom of crowds. New York, NY, USA: Random House.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817–838.

## Internetes források

- Football Benchmark. (2024). Most valuable soccer leagues worldwide as of February 2024, by combined player value (in million euros). *Statista*. Letöltés dátuma: 2024.03.18. Letöltés helye: <https://www.statista.com/statistics/1454070/soccer-leagues-aggregate-player-value/>
- IVW. (2023). Number of online and mobile visits on news portals in Germany in October 2023 (in millions). *Statista*. Letöltés dátuma: 2024.03.18. Letöltés helye: <https://www.statista.com/statistics/442988/most-visited-news-websites-germany/>
- Sports Reference. (n.d). All About Fbref.com. FBref. Letöltés dátuma: 2024.02.11. Letöltés helye: <https://fbref.com/en/about/>
- Transfermarkt. (2021). Transfermarkt Market Value explained—How is it determined?. Transfermarkt.Co.In. Letöltés dátuma: 2024.03.11. Letöltés helye: <https://www.transfermarkt.co.in/transfermarkt-market-value-explained-how-is-it-determined-/view/news/385100>
- Whitmore, J. (2021). What is Expected Assists (xA)?. Opta Analyst. Letöltés dátuma: 2024.03.20. Letöltés helye: <https://theanalyst.com/eu/2021/03/what-are-expected-assists-xa/>
- Whitmore, J. (2023). What is Expected Goals (xG)?. Opta Analyst. Letöltés dátuma: 2024.03.20. Letöltés helye: <https://theanalyst.com/eu/2023/08/what-is-expected-goals-xg>

## Mellékletek

### 1. melléklet: az alapmodell heteroszkedaszticitása ábrán



*Saját szerkesztés.*

### 2. melléklet: az alapmodell heteroszkedaszticitásának tesztelése

Teszt neve	Teszt p-értéke	Teszt értelmezése
Kolmogorov-Smirnov	2.2e-16	A hibatagok nem normális eloszlásúak. Koenker-korrekciózott BP-teszt kell.
White (interakciók nélkül)	6.55e-19	A modell heteroszkedasztikus.
White (interakciókkal)	9.18e-13	A modell heteroszkedasztikus.
Breusch-Pagan (Koenker-korrekciózott)	2.2e-16	A modell heteroszkedasztikus.
Megfigyelések száma: 682		

*Saját szerkesztés.*

### 3. melléklet: Az alapmodell multikollinearitás vizsgálata

Magyarázóváltozó	VIF	Szabadságfok	Korrigált VIF
age	192.345	1	13.869
age_square	182.823	1	13.521
comp_level	2.154	4	1.101
minutes_pct	3.633	1	1.906
goals_per90	1.717	1	1.310
assists_per90	1.493	1	1.222

passes_per90	4.668	1	2.160
passes_pct	2.057	1	1.434
aerials_won_pct	1.297	1	1.139
log_market_value_2022	2.571	1	1.603
take_ons_per90	1.896	1	1.377
challenge_tackles_pct	1.270	1	1.127
international_reputation	1.981	4	1.089
injury_prone	1.104	1	1.051
future_prospect	2.203	1	1.484
log_total_pageviews	2.485	1	1.576
Megfigyelések száma: 682			
<i>Saját szerkesztés</i>			

#### 4. melléklet: A 3. modell (xA modell) multikollinearitás vizsgálata

Magyarázóváltozó	VIF	Szabadságfok	Korrigált VIF
age	192.345	1	13.869
age_square	182.823	1	13.521
comp_level	2.154	4	1.101
minutes_pct	3.633	1	1.906
goals_per90	1.717	1	1.310
assists_per90	1.493	1	1.222
passes_per90	4.668	1	2.160
passes_pct	2.057	1	1.434
aerials_won_pct	1.297	1	1.139
log_market_value_2022	2.571	1	1.603
take_ons_per90	1.896	1	1.377
challenge_tackles_pct	1.270	1	1.127
international_reputation	1.981	4	1.089
injury_prone	1.104	1	1.051
future_prospect	2.203	1	1.484
log_total_pageviews	2.485	1	1.576
Megfigyelések száma: 682			
<i>Saját szerkesztés.</i>			