

OTDK-dolgozat

2023

Budapesti Corvinus Egyetem
Gazdálkodástudományi kar



**Mi történt a Dortmundtal? Dárdai ismét összeszedi a Herthát? A
Schalkénak befellegzett?**

**What happened to Dortmund? Dárdai rebuilds Hertha again? Schalke's
dreams are over?**

szerző: Bodnár Gergely

Konzulens: Havran Zsolt

Kézirat lezárásának dátuma: 2021.04.12.

Tartalomjegyzék:

Ábra- és táblázatjegyzék:	5
1. Absztrakt, bevezetés:	6
2. A sport, illetve a labdarúgás, mint üzlet:	7
3. A német labdarúgóbajnokság és bemutatása:	11
3.1. A német labdarúgó bajnokság rendszere:	11
3.2. A német (első, illetve másodosztályú) labdarúgóklubok működése, nézőszámok alakulása:	11
4. Labdarúgás és adatelemzés kapcsolata:	14
4.1. Az adatelemzés megjelenése, kezdeti elemzési módszerek:	14
4.2. Labdarúgás és Big Data elemzés kapcsolata, az OPTA megalakulása:	16
4.3. Adatgyűjtési eszközök és módszertanuk a XXI. század labdarúgásában:	17
5. Teljesítményre ható tényezők a labdarúgásban, releváns mutatószámok a szakirodalomból:	20
5.1. Labdabirtoklás, illetve passzpontosság és teljesítmény kapcsolata:	20
5.2. Kapura lövések, illetve támadások hatása a teljesítményre:	21
5.3. Védekezést, illetve mentalitását jellemző mutatószámok:	22
5.4. Transzferpiaci tevékenység, valamint a játékosok értékének hatása a teljesítményre:	23
5.5. Játékoskeretet jellemző mutatók és a teljesítmény összefüggése:	24
5.6. Nézőszámok, illetve a szurkolói aktivitás és a teljesítmény kapcsolata:	26
5.7. Egyéni teljesítmény hatása a csapat szintű teljesítményre:	28
6. Kutatás, a német Bundesliga elmúlt négy szezonjának vizsgálata az ökonometria eszköztárával:	29
6.1. Adatbázis összeállítása, változók kiválasztása:	29
6.2. Modell szűkítése lineáris korrelációk segítségével:	31
6.3. Modell szűkítése backward módszerrel:	33
6.4. Végző modell változói, illetve regressziós eredmények összegzése:	34
6.5. Multikollinearitás vizsgálata:	36
6.6. Heteroszkedaszticitás tesztelése:	36
6.7. Regressziós modell robusztus standard hibákkal:	39
7. Előrejelzés a modellel, célszámok megfogalmazása, illetve elérésükhöz szükséges teljesítmény:	41
7.1. Célpontszámok meghatározása, az „átlagos csapat”:	41
7.2. Teljesítmény becslése a szignifikáns mutatószámok segítségével:	44
7.3. Elaszticitási együtthatók vizsgálata a szignifikáns változók esetén:	47

7.4. A végső lineáris regressziós modell alkalmazása a 2020/2021-es szezon előrejelzésére:	48
7.5. Regressziós modellem becslésének összevetése a tényleges 2020/2021-es teljesítménnyel:	50
8. A kutatás korlátatai, további kutatási irányok bemutatása:	56
8.1. A kutatás korlátai:	56
8.2. További kutatási irányok:	56
Forrás- és irodalomjegyzék:	58
Mellékletek:	68

Ábra- és táblázatjegyzék:

1. ábra: Nézőszámok alakulása.	13
2. ábra: Pontok megoszlása a Bundesligában.....	27
1. táblázat: Pontszámok megoszlása a négy vizsgált szezonban.....	27
2. táblázat: A modell változói és csoportosításuk, illetve a korreláció iránya az előzetes várakozások alapján.	31
3. táblázat: A magyarázó változók lineáris korrelációja, illetve szignifikanciája a pontszám alakulásával.	32
4. táblázat: Végző modell regressziós eredményei.....	35
5. táblázat: Az ANOVA-teszt eredményei.	35
3. ábra: Hibatagok szóródása a regresszió standard értékeinek függvényében.....	37
4. ábra: A White-teszt eredményei.	38
6. táblázat: A Breusch-Pagan, illetve Koenker-próbák eredményei.	38
7. táblázat: A regresszió eredményei robosztus standard hibákkal kalkulálva.	39
8. táblázat: A pontszámok összefoglalása.	42
9. táblázat: Célpontszámok, illetve győzelmei összefoglalása.....	43
10. táblázat: Teljesítmény becslése a változók egységnyi változtatásai alapján.....	45
11. táblázat: Az elaszticitási együtthatók az egyes célpontszámok szintjén.	47
12. táblázat: Szezonvégi teljesítmény előrejelzésének eredményei.	49
13. táblázat: Valós eredmények összevetése modellem becsléseivel a 2020/2021-es Bundesliga-szezon első 26 fordulójára alapján.	51
14. táblázat: CSGR mutatók a 2020/2021-es szezon alapján.....	53
15. táblázat: A Dortmund győzelmi aránya az elmúlt öt év Bundesliga szezonjaiban.	54

1. Absztrakt, bevezetés:

A modern labdarúgás a klubvezetők, befektetők, tulajdonosok, valamint a szurkolók számára is rengeteg statisztikát bocsát rendelkezésre. Ezek értelmezése és megfelelő módon való kezelése segítséget nyújt kulcsfontosságú döntések meghozatalában is a professzionális labdarúgóklubok vezetői számára. Az adatokon alapuló döntéshozatal a labdarúgásban nem volt mindig jelen, mely a digitalizáció, illetve a rendelkezésre töké hiányának volt betudható a XX. század első és középső harmadában. Dolgozatomban a modern labdarúgás újszerű adatelemzési, adatgyűjtési eszközeiről, illetve az elterjedt statisztikai mutatók bemutatásáról, használatáról írok. A kutatás során arra keresek választ, hogy a modern labdarúgásban vizsgált statisztikai mutatószámok közül melyek azok, amelyek szoros lineáris korrelációs kapcsolatban állhatnak egy csapat teljesítményével hosszabb távon (teljes labdarúgó szezonokat vizsgálva). Kutatómunkám során Európában az egyik legjobban ismert labdarúgó bajnokság, a német Bundesliga csapatainak teljesítményét vizsgálom majd négy teljes szezonon keresztül (2016/17-től 2019/20-ig). A kutatást kvantitatív módszertannal végzem, amelyhez a felhasznált adatbázis sajátkezű, manuális adatgyűjtés eredménye volt. Az adatbázisból az SPSS szoftver segítségével ökonometriai (lineáris regressziós) modellt építék, amelyből különféle modellszelekciós eljárásokkal kiszűröm az egyes csapatok teljesítményét nem szignifikánsan befolyásoló magyarázó változóként használt mutatószámokat. A végső modell segítségével az egyes, statisztikailag szignifikáns változókhoz definiálok célszámokat, amelyek elérésével a Bundesligában egy csapat elérheti a nevezetes helyezésekhez (pl.: bentmaradás, Bajnokok Ligája indulás, Európa Liga indulás) szükséges pontszámot egy szezon során. A professzionális labdarúgás köszönhetően az évről évre növekvő piacméretnek, illetve a verseny élesedésének napjainkra teljes mértékben eredménycentrikussá vált, minden klub számára a cél, hogy minél több mérkőzést megnyerjenek (Kuper, Szymanski, 2014). Dolgozatom eredményei alapján definiálni fogok tudni olyan mutatószámokat, amelyek szignifikánsan befolyásolják a csapatok teljesítményét az általam készített mintavétel alapján. Végezetül készítek egy előrejelzést a 2020/2021-es Bundesliga szezon 26. fordulójának utáni statisztikák segítségével a bajnokság végeredményére az általam alkotott regressziós modell alapján, valamint a modellel készített becslést összevetem a tényleges eredményeivel a 2020/2021-es szezonnak.

2. A sport, illetve a labdarúgás, mint üzlet:

Napjainkban hatalmas iparaggá vált világszerte a sport. A globális sportpiac teljes bevétele csaknem elérte 2017-ben a 91 milliárd dollárt (Scrapehero, 2017). Ez részben annak köszönhető, hogy a XXI. században hatalmas mértékben megnőtt a sportközvetítések száma és egyre több sportág válik globális sporttá. Globális sportágak azon sportágak nevezzük, melyeket világszerte közvetítenek az interneten, valamint televízión keresztül, illetve azon sportágak, amelyek során a megrendezett versenyek helyszínei a Föld valamennyi kontinensét lefedik (András et al. 2012). A sportban, akár csak a vállalati szférában is erős a versenyhelyzet, az egyes országok hazai bajnokai (akár csapat, akár egyéni szinten) mind törekednek arra, hogy nemzetközi szinten, világversenyeken (világbajnokságok, Olimpia, kontinensviadalok) megmérettessék magukat (András, Havran, Jandó 2012) sikeresen. A sportpiac méretét tekintve is hatalmas és évről évre növekszik globálisan a mérete, 2011 és 2018 között csaknem másfélszeresére nőtt (Statista, 2020d). 2018-ra a sportpiac mérete körülbelül 471 milliárd dollárra volt tehető, sőt Wood (2019) üzleti előrejelzése szerint 2022-re akár a 614 milliárd dollárt is meghaladhatja. A sport töretlen és globális népszerűsége a piacon köszönhető a nagyfokú média érdeklődésnek a sportesemények iránt (Kajos, 2019), illetve a sportmárkák, valamint a sportolók személyes marketingértékének egyaránt. A sportok közül is üzletileg, illetve pénzügyileg egyaránt kiemelkedik világszerte a labdarúgás. A labdarúgás, illetve a labdarúgószövetségek a teljes globális sportpiac méretének körülbelül 43%-át tették ki 2020 februárjában (Torrens, 2020), ezzel toronymagasan a legjelentősebb sportnak számított a futball piacméret tekintetében.

Az 1990-es évektől Európa legrangosabb ligáinak labdarúgó klubjai egyre jellemzőbben váltottak működési formát, és kezdtek el vállalként, gazdasági társaságként működni (András, 2003). Míg szűk 30 évvel később, 2019-re globális sportpiac bevételeinek csaknem a harmadát, megközelítőleg 29 milliárd dollárt termelt meg bevételként az európai labdarúgópiac (Statista, 2019a). Labdarúgásban, mint az üzlet megannyi területén, szintén jelen van a globalizáció, azaz a potenciális világméretűség (Chikán, 2017). A labdarúgás a rendkívül magasság nézettség mellett gyakran szociológiai, illetve politikai (Giulianotti, Robertson, 2012) kontextusban is megjelenik, főként a XXI. században. Kuper és Szymanski (2014) szerint azonban hosszabb távon már nem kizárólag Európa lesz a labdarúgás középpontja, ugyanis az utóbbi néhány évtizedben már több tengerentúli (Egyesült Államok), illetve távol-keleti országban (Japán, Kína, Arab Emírségek, Katar) is tért hódított magának ez a sportág, amely további növekedési lehetőségeket jelent a jövőre nézve. Kínában például 2018-ban a Manchester United

legnépszerűbb futballklubként az országban már 107 millió követővel rendelkezett (Wilson, 2018), mellyel a „Vörös Ördögök” csapatának egyik legjelentősebb piaca a kínai rajongótáboruk. A rajongás az európai futball iránt ezeken a kontinenseken tehát már jó néhány évtizede jelen van, azonban kifejezetten új jelenségnek számít, hogy a labdarúgók idősebb korukban egy új kontinensen népszerűsítik tovább a sportot, illetve személyüket, valamint emelik ezzel az adott ország bajnokságának színvonalát is. Az egyik legelső ilyen transzfer még David Beckham átigazolása volt, aki 2007-ben a Real Madridtól az amerikai bajnokságban szereplő Los Angeles Galaxy csapatához szerződött. Beckham a mai napra már ebben a bajnokságban (a Major League Soccer-ben) egy 2020-ban alapított klub, az Inter Miami CF tulajdonosa, illetve működési igazgatója (Intermiamicf, 2021). A dél-amerikai kontinensen pedig már a kezdetektől fogva, az első világbajnokság (1930) idején is nagy hagyománya volt a labdarúgásnak, mint sportnak. (A legelső világbajnokságon indult 13 csapatból 7 ország csapata dél-amerikai volt (Fifa, 2021)).

A labdarúgásban tevékenykedő legnagyobb kluboknak nem elsődleges célja a profit termelése (Kuper, Szymanski, 2014) a vállalati szektorban tevékenykedő cégekkel ellentétben (Chikán, 2017). Ezen klubok tulajdonosai, illetve tulajdonosi csoportjai hatalmas mennyiségű pénzt fektetnek a működésbe (játékos transzferek, sztáredzők szerződtetése, edzőközpontok működtetése, alkalmazottak fizetése), mellyel a legfőbb céljuk, hogy a pályán sikeres legyen a csapatuk és minél több trófeát sikerüljön elhódítaniuk (Kuper, Szymanski, 2014). A labdarúgásban azonban az elért sikereknek, illetve a pénzügyi sikereknek szoros kapcsolata van egymással. A klubok szurkolói bázisa növekedik világszerte, valamint a környékbeli szurkolótáborok mérete is nagy mértékben meg tud nőni abban az esetben, amennyiben egy csapat jól teljesít huzamosabb ideig. A legelhivatottabb rajongók, a német bajnokság fanatikus szurkolói egyes kutatások szerint saját pénzüket is hajlandóak lennének feláldozni, hogy csapatuk a pályán sikereket érjen el (Wicker, Whitehead, Johnson, Mason, 2016). A jó eredmények mindezek mellett gyakran befolyásolják pozitív irányban a városok lakosainak a boldogságát, valamint büszkeségét, illetve a városok külső megítélését is egyaránt (Wicker, Whitehead, Johnson, Mason, 2016). Kuper és Szymanski (2014) szerint eredményektől függetlenül már a nemzetközi futballrendezvények (világbajnokságok, illetve kontinensviadalok) rendezése is növelheti az általános elégedettséget, illetve a boldogság mértékét. Az ő kutatásuk eredményei szerint, különösen a férfiak körében csökkent az öngyilkosságok száma abban a néhány hétben, amikor egy adott ország valamely jeles nemzetközi futballeseménynek szolgált helyszínül.

Ettől függetlenül a labdarúgóklubok bizonyos aspektusok szerint üzleti alapon működnek, bevételi forrásait különböző piacok jelentik. András (2003) szerint ezen piacok a fogyasztói piac, a közvetítési jogok piaca, a szponzori piac, a merchandising piac, illetve a játékospiac, amely tranzakcióiról az átigazolási időszakok során rengeteg információ áll rendelkezésünkre. Ezen piacok közül a leginkább különböző működését tekintve a játékospiac, egyrészt ezen piacon a tranzakciók csupán szezonálisan vannak engedélyezve, másrészt a tranzakciók tárgyát emberek, profi labdarúgójátékosok képezik. A labdarúgó szezonok során egy év alatt kettő átigazolási időszak van (Transfermarkt, 2020a), amikor a klubok játékosaikkal szabadon kereskedhetnek (természetesen a játékosok, illetve ügynökeik beleegyezésével). A téli időszak jellemzően a januári hónapot szokta jelenteni (egyes országokban, bajnokságokban néhány nap eltérés adódhat a nyitó, illetve záró időpont között), míg a nyári átigazolási időszak, mely jellemzően jóval nagyobb pénz-, illetve játékosmozgással jár már a májusi hónap végén nyílik meg, és általában augusztus végéig nyitva marad a legtöbb, topligás bajnokságot üzemeltető országban. A nyári átigazolási időszak során a bajnokságok túlnyomó része szünetel, a csapatok nem játszanak tétmeccseket, illetve ekkor szokták megrendezni a nemzeti válogatottak közötti sporteseményeket is (Világ bajnokság, Európa-Bajnokság, Konföderációs Kupa). Egy nyári átigazolási szezon alatt a pénzmozgás tehát valóban jelentős, a The Guardian (2020) által készített kimutatás szerint az öt európai topligában (angol, spanyol, német, olasz, illetve francia első osztály) 2020 nyarán összesen 1422 üzlet kötött meg és ezek értéke körülbelül 4,13 milliárd dolláros volt.

A labdarúgás során a televíziós, illetve egyéb (internetes, régebben rádiós) közvetítési jogok piacán is hatalmas pénztömegeket mozgat meg. A Forbes (2020) szerint a 2020-as évben például a német Bundesliga (első, illetve másodosztályú) csapatok 4,95 milliárd dollárt kaptak azért, hogy mérkőzéseiket televízión keresztül is közvetítsék. Ez a 2020-as pénzösszeg ráadásul az első az elmúlt tizenöt évben, mely (enyhe) recessziót mutatott az előző évhez képest, melynek az oka vélhetően az, hogy sikerült a piac nagy részét lefedniük, rengeteg országban nézhetőek már televízión keresztül is a német labdarúgás legfelsőbb osztályainak mérkőzései. Napjainkban minden versenysorozatnak vannak főszponzorai, legyen az klub vagy nemzetközi szintű verseny. A stadionokban, illetve a televíziós közvetítések előtt közvetlenül megjelenő reklámok mellett a játékosok mezein is multinacionális vállalatok szponzorációjának végtermékeit láthatjuk (Hanke, Kirchler, 2013). A világ leggazdagabb bajnokságának tartott angol első osztályú sorozatban, a Premier League-ben már 2019-2020-as szezon elején minden egyes csapat egymillió dollár feletti főszponzori szerződéssel rendelkezett (Financefootball,

2019). A legmagasabb értékű főszponzori szerződés ebben a szezonban a Chevrolet és a Manchester United együttműködése, mely több, mint 84 millió dollár értékű megállapodás volt.

Egy ekkora éves bevételeket termelő piacon minden egyes játékostranszfert, beruházást, befektetést, döntést nagyon komplex módon hoznak meg a labdarúgó klubok vezetői, adatokra, számításokra, illetve előrejelzésekre támaszkodva. A sportban az adatelemzés, illetve a statisztikák szerepe a döntéstámogatás mellett több másik iparág számára nyújt támogató funkciót. Például az úgynevezett Sports Fantasy játékok alapjául is szolgálnak (Scrapehero, 2017) a statisztikák. Ezen játékok segítségével a hétköznapi emberek gyakorlatilag úgy hozhatnak döntéseket, ahogyan egy edző vagy egy sportklub menedzsere a való életben, melyhez legtöbbször a rendelkezésre álló statisztikákat, illetve a játékosok formáját is figyelembe kell venniük, hogy a lehető legjobb eredményt ériék el. A fantasy játékokra való regisztráció leggyakrabban ingyenes, és gyakorlatilag bármilyen általunk kedvelt sportot (legjellemzőbben csapatsportokat) követhetünk segítségükkel, ezáltal is izgalmasabbá téve az adott sportágat, illetve az eredmények alakulását számunkra.

Egy ennél jelentősebb pénztömegeket megmozgató felhasználása a különféle sportstatisztikáknak a szerencsejátékok, illetve a sportfogadás esetén van. Az Amerikai Egyesült Államokban éves szinten 2019-ben több, mint 900 milliárd dollár (Statista, 2020a) volt a sportfogadásokon realizált bevétele a fogadóirodáknak, ez tehát azt jelenti, hogy a sportfogadás piaca jóval nagyobb méreteket ölt globális szinten, mint a professzionális sportoké. Talán ez a magyarázata a sportban (a labdarúgásban kifejezetten gyakran) rendszeresen előforduló különféle bundabotrányoknak, illetve fogadási csalásoknak. Rengeteg korábbi kísérlet volt mérkőzések eredményeinek befolyásolására a sportolók, játékvezetők, de akár az edzők bevonásával is Európa és világszerte egyaránt.

3. A német labdarúgóbajnokság és bemutatása:

Dolgozatom kutatási részében a Bundesliga, azaz a német első osztályú labdarúgó bajnokság csapatainak az elmúlt négy évben elért teljesítményére ható tényezőket vizsgálom majd meg regressziós elemzés segítségével. Ebben a fejezetben a német bajnokság rendszerét, illetve a klubok működési modelljét mutatom be.

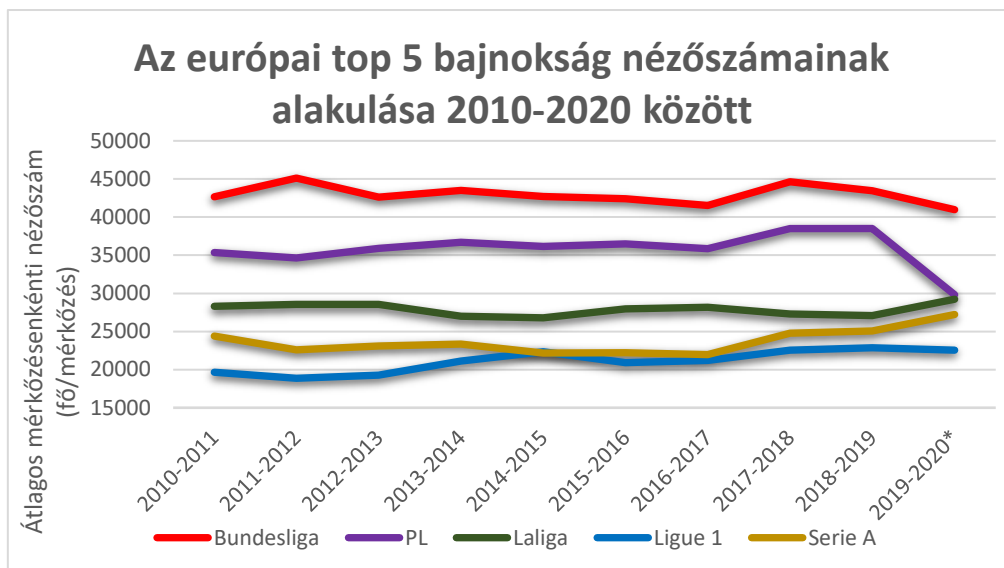
3.1. A német labdarúgó bajnokság rendszere:

A Bundesliga egy szezonjában 18 csapat versenyez, minden csapat kétszer (hazai pályán, illetve idegenben) játszik egymás ellen, egy szezonban tehát 34 forduló van (Bundesliga, 2020a). A 2019/20-as szezon szabályzata alapján a német Bundesliga első négy helyezett csapata jogosult elindulni a következő szezon során az UEFA Bajnokok Ligája sorozatban. Az ötödik helyezett automatikusan elindulhat Európa második legrangosabb kupasorozatának az UEFA Európa Ligának a csoportkörében, míg a hatodik helyezett az Európa Liga selejtezőkörében indulhat el a következő szezon során. A Bundesliga utolsó két helyezettje kiesik a német másodosztályba, a Bundesliga2-be, illetve a 16. helyezett (hátról rangsorolva a harmadik), pedig osztályozó mérkőzést játszik a Bundesliga2-ben a dobogó legalsó fokán végzett csapat ellen (Bundesliga, 2020a). A másodosztályban pontosan ugyanez a rendszer működik. A harmadosztályban, a Bundesliga3-ban már 20 csapat versenyez, ahol az első két helyezett jut fel, a 3. helyezett pedig játszik a másodosztály 16. helyezettje ellen osztályozó mérkőzést. A Bundesliga3 utolsó négy helyezettje esik ki (Bundesliga, 2020a). A harmadosztály alatt már nem országos, hanem regionális szinten működnek a bajnokságok Németországban.

3.2. A német (első, illetve másodosztályú) labdarúgóklubok működése, nézőszámok alakulása:

A német labdarúgó első -, illetve másodosztályban a klubok irányítására, működésére vonatkozóan létezik egy egész Európában egyedi szabályrendszer, az úgynevezett 50+1%-os szabály. Az 50+1%-os szabály értelmében a német labdarúgóklubokat érintő döntések meghozatala több, mint 50%-ban az anyaegyesületek kezében kell, hogy maradjon, hogy elindulhassanak a Bundesliga első, illetve második osztályában. A különféle befektetői csoportok tehát legfeljebb 49%-ban képviseltethetik érdekeiket a klubok irányítása során

(Pethő, 2014a). A szabály 1999 január elsejétől lépett életbe és azon klubok mentesültek ez alól a működési rendszer alól, amelyek már a szabály meghozatalakor nagy mértékben függtek külső tulajdonban álló vállalatoktól (Pethő, 2014b). Ezen rendelkezés a jelenlegi első osztály mezőnyéből két klubot érintett: A Bayer Leverkusent (tulajdonos: Bayer gyógyszer), illetve a VFL Wolfsburg (tulajdonos: Volkswagen autógyár). A német labdarúgó szövetség ezen a szabályon 2011-ben módosított, mely értelmében bármely külső befektető rendelkezhet többségi döntési joggal abban az esetben, ha az adott klubbal legalább 20 éve együttműködik (Pethő, 2014b). Ez a rendelkezés elősegíti azt a tényt is, hogy a klubok és a szurkolók közötti kapcsolat Németországban egész Európában egyedülálló, illetve az 50+1%-os szabálynak nagyon fontos szerepe van abban, hogy a Bundesliga jegyárai kivételesen alacsonynak mondhatóak a többi európai top ligához képest (Bundesliga, 2018). Az alacsony jegyárak, illetve az erős kötődés kialakulása a szurkolók és klubjaik között együttesen járultak ahhoz, hogy a Bundesliga mérkőzései a Covid-19 vírus megjelenése előtt a legmagasabb átlagos nézőszámmal rendelkezzenek a világ összes labdarúgó ligáját figyelembe véve. Ez a nézőszám már több, mint 10 éve 40 ezer néző feletti számot jelent mérkőzésenként, a szurkolói aktivitás (stadionok kihasználtsága) pedig átlagosan 91,76%-os volt a 2016/17-es, illetve a 2019/20-as labdarúgó szezonok közötti négy labdarúgó szezonban. A 2016/2017-es szezon során ráadásul mindössze két csapat volt, amelyek stadionjának kihasználtsága 90% alatt volt, a Hertha Berlin, illetve az FSV Mainz 05. A legtöbb csapat esetén hatalmas számban találkozhatunk bérlettel rendelkező szurkolóval, például a Borussia Dortmund esetén már a 2010/11-es szezonban is több, mint 51 ezer szezonbérlettel (ez a 17 hazai meccsre való kilátogatásra jogosítja fel a szurkolót) rendelkező szurkolójuk volt (Whitney, 2010).



1. ábra: Nézőszámok alakulása.

(források: Transfermarkt (2021), Bundesliga (2020c), Statista (2020b))

Saját készítés.

*: A 2019-2020-as labdarúgó szezon során a mérkőzések márciustól zárt kapuk mögött zajlottak, az angol bajnokság („PL”) esetén a zártkapuval megrendezett mérkőzéseket is beleszámolták a statista.com adatgyűjtése során.

A szurkolói aktivitás Németországban az alacsonyabb osztályokban is rendkívül magas, a Bundesliga2-ben 2016 és 2019 között a teljes bajnokság átlagos nézőszáma sosem volt 17 ezer néző alatt, sőt a 2016/17-es szezonban az átlagos nézőszám mérkőzésenként a 21 ezret is meghaladta (Worldfootball, 2020). Ezzel Európában egyetlen alsóbb osztályú liga van, amely képes a versenyt tartani, az angol Championship (másodosztály), mely átlagos nézőszámai a Covid-19 megjelenése előtt szintén 17 és 21 ezer néző között szóródtak.

4. Labdarúgás és adatelemzés kapcsolata:

4.1. Az adatelemzés megjelenése, kezdeti elemzési módszerek:

A futballban a statisztikák eleinte a szurkolók tájékoztatására készültek el, azonban napjainkra a leghatalmasabb futballklubok már adatelemző csapatokat foglalkoztatnak. Ezen adatelemzők olyan adatokhoz és információkhoz férnek hozzá a labdarúgókról, melyeket egy hétköznapi rajongó sem értelmezni, sem hozzáférni nem tudna. A hagyományos, kezdetleges mutatók leíró statisztikai mutatók voltak, mind játékosok (például: egy játékos által szerzett gólok száma, adott gólpasszok száma, később megnyert fejpárbajok, szerelések száma, passzok száma, futómennyiség), mind a csapatok (például büntetőlapok száma, gólkülönbség, szerzett pontok hazai pályán, illetve idegenben) szintjén. A probléma ezekkel a leíró jellegű statisztikai mutatókkal Kocsis (2020) szerint, hogy ezen mutatók, (elsősorban a játékos szintűek) kontextus nélkül nem rendelkeznek elegendő magyarázóerővel. Például egy védekező mentalitású csapat vélhetően kevesebbet birtokolja a labdát, a védőik több szereléssel is rendelkezhetnek. Illetve egy játékos passzpontossága is lehet 100%-os, hogyha folyamatosan néhány méteres hátrafelé irányuló passzokkal próbálkozik, ettől nem nyújt valaki kiemelkedő teljesítményt.

A labdarúgás történetének első adatelemzői közé tartozott az angol Charles Reep, aki először próbált egy átfogó elemzési rendszert alkotni a labdarúgásra (Pollard, 2010). Reep, aki fiatalkorában az angol katonai légierőben szolgált éveken keresztül, lett az első hivatásos adatelemző a labdarúgásban. Az ő megfigyelései még teljesen manuálisan készültek el. Reep elsősorban a hatékony labdarúgásra próbált egyfajta metódust alkotni, arra kereste a választ, hogy hány passz szükséges egy-egy gól eléréséhez. Ezen megfigyelések arra adtak következtetést, hogy a gólok nagy részéhez néhány passz volt szükséges. Ezen elemzési keretrendszernek az eredménye tehát egy olyan stratégia volt, melyben a labdát a lehető leggyorsabban az ellenfél térfelére kellett juttatni. Reep ötlete a kevés passzos futballal azonban kifejezetten mély nyomott hagyott a brit, illetve az északi labdarúgás egészére. Larsson (2010) szerint a Skandináv országokra nagy hatással volt a 60-as évek végétől a Reep ötlete nyomán Nagy-Britanniában megjelenő „újfajta” játéktípus. Norvégiában Charles Hughes és Allan Wade is tett látogatásokat a 60-as években, közülük Hughes dolgozott korábban együtt Charles Reep-pel. Svédországban pedig két, a 70-es években fiatal angol edző fektette le a brit játéktípus alapjait, ők Bob Houghton (a Malmö FK csapatát vette át), illetőleg Roy Hodgson (a Halmstad BK edzője lett) voltak (Larsson, 2010).

Később egy híres ukrán edző, Valeriy Lobanovskij volt az, aki beemelte a statisztikát, illetve az adatelemzést taktikai elemként a csapatai játékába. Lobanovskij 32 éves edzői pályafutása során 33 hivatalos trófeát nyert csapataival, az ukrán (korábban szovjet) Dinamo Kijev csapatával három alkalommal is európai szinten is kiemelkedő csapatot tudott felépíteni (Dimiean, 2019). Lobanovskij sikerének egyik magyarázata az a tény, hogy elméleti fizikusokkal, illetve matematikusokkal dolgoztatott ki statisztikákat, elemzéseket a játékosai teljesítményéről.

Az egyik leghíresebb elméleti fizikus, akivel együtt dolgozott a kijevi egyetem professzora Anatolij Zelentsov volt (Dimiean, 2019). Lobanovskij és az elemzői csapata egy nagyon hasonló játéktípust alkottak meg a hollandoktól (Rinus Michels és az 1970-es évek Ajaxa volt erről a stílusról rendkívül híres) eredeztetett „totális futball” nevű játéktípushoz. (A totális futball lényegében azt jelentette, hogy a csapat felállása fix az egész meccs során, azonban a játékosok a pozíciójukat folyamatosan változtatták a mérkőzés folyamán. (Siregar, 2018) Lobanovskij ezt a játéktípust egy extra elemmel, az intenzív letámadással egészítette ki, ehhez azonban a játékosainak átlagon felüli fizikai állapotra volt szükségük, melyet intenzív edzőmunkával értek el (Dimiean, 2019).

Az ukrán edző a statisztikákat, melyeket vezettek az elemzői egy idő után a játékosai számára úgy fordítottak le, hogy bizonyos poszton játszó játékosok számára egy-egy statisztikai mutatóra egy célszámot rendelt (Ronay, 2012, Kele, 2020). Ez azt jelentette, hogy az edző akkor volt elégedett a játékosok teljesítményével, hogyha elérték ezeket a kitűzött célszámokat. Ez a játékosok felé támasztott elvárás legtöbbször olyan mutató volt, mely az erőállapottal volt összeegyeztethető. Leggyakrabban a játékosok által megtett sprintek darabszámára volt egy célkitűzése Lobanovskijnek, például az a (általában szélső) játékos, aki megtett 50 sprintet egy mérkőzésen, az vélhetően mindent megtett a győzelemért, jól teljesített. Az adatelemzések és a statisztikák helyes felhasználásával és értelmezésével tehát a XX. században versenyelőnyre lehetett szert tenni. Lobanovskij nyomdokaiba lépett a XX század végén, illetve a XXI. században egyre több és több sikeredző is: Pep Guardiola, José Mourinho, Arséne Wenger, valamint Sam Allardyce mind hittek a statisztika, illetve adatelemzés alapú játékos-, illetve futballklubmenedzsmentben (Ronay, 2012, Kuper, Szymanski, 2014).

4.2. Labdarúgás és Big Data elemzés kapcsolata, az OPTA megalakulása:

1996-ban megalakult az OPTA (McGuinness, 2012), egy brit szervezet, amely kizárólag sportstatisztikával, illetve sportstatisztikai adatok gyűjtésével, majd ezen adatok elemzésével kezdett el foglalkozni. Az OPTA megalakulása mellett elkezdtek a ProZone rendszerek beépítését (Arastey, 2018, Kele, 2020) a footballstadionokba. Kameratek eddig is voltak, azonban a '90-es évek legvégétől a ProZone segítségével sokkal több szögből, látványosabban elemezhetők a sportesemények történései. A ProZone rendszereinek felszerelésével a játékosok gyakorlatilag minden mozdulata elemezhetővé vált a futballpályán. A különbség az OPTA, illetve a ProZone között, hogy noha mindkét szervezet alapvetően az adatelemzéssel foglalkozik, a ProZone elsősorban a futballklubok számára nyújt szolgáltatásokat, míg az OPTA ügyfelei között sportfogadással foglalkozó nagyvállalatokkal is találkozhatunk (Medeiros, 2014). A ProZone egyik korábbi munkatársa, Simon Wilson 2006 óta a Manchester City F.C. munkatársaként dolgozik stratégiai teljesítményelemzőként. Wilson és a ProZone korai alkalmazása, illetve az a tény, hogy 2008-tól a Manchester City futballklub az Abu Dhabi United Group tulajdonába került együttesen hozták el a sikereket a klub számára. 2008 óta négy angol bajnoki címmel, két FA-kupa győzelemmel, öt Angol Ligakupa győzelemmel gazdagodott a klub (Cricfooty, 2017).

A mai napra az OPTA elemzői a sporteseményeket, mérkőzéseket élőben, illetve a mérkőzések után is figyelik, analizálják a történő eseményeket (Optasports, 2021). Az OPTA elemzőinek akkor van munkájuk, amikor a mérkőzések zajlanak, így például a Premier League-gel, az angol labdarúgó bajnokság elemzésével foglalkozó szakemberek karácsonykor is dolgoznak (McGuinness, 2012) az Angliában egyedinek számító játéknapon, December 26-án (Angliában ezt a napot „Boxing day”-nek nevezik). Kuper és Szymanski (2014) szerint az adatelemzők gyakran többször is megnézik ugyanazt a mérkőzést az adatok megbízhatóságát ezzel is tovább növelve. A mérkőzések alatti, illetve utáni adatszolgáltatás között azonban statisztikai hibahatáron belül vannak az eltérések. Az OPTA napjainkra 350 sportcsapatot támogat a legprofibb adatelemzési eszközökkel, módszerekkel (Optasportspro, 2020). A labdarúgás egyik legkorszerűbb teljesítménymutatója, az X/G (expected goal) mutató (bővebben később) is az OPTA adatbázisaiban található mérkőzéseken, illetve ezeken leadott kapura lövési kísérleteken alapszik (Footballxg, 2021). Rengeteg nagyvállalati partnerrel működnek együtt, többek között labdarúgóklubokkal, szövetségekkel, sportfogadással

foglalkozó szervezetekkel, illetve különböző sportstatisztikák szolgáltatásával foglalkozó weboldalak (pl.: whoscored.com) egyaránt támogatnak.

4.3. Adatgyűjtési eszközök és módszertanuk a XXI. század labdarúgásában:

A modern labdarúgópiac éves szinten már 2017-ben közel harminc milliárd dolláros bevétellel (Scrapehero, 2017) dolgozott, a profi labdarúgóklubok szinte tökéletes körülményeket teremtenek meg a játékosaik számára a fejlődéshez, illetve ahhoz, hogy minél tökéletesebben végezzék a munkájukat. Az XXI. század egyik legnagyobb angol futballklubja, a Manchester City edzőközpontja, az úgynevezett Etihad Campus például több, mint 30 hektáron terül el (Cityfootball-performance, 2020), tele edzőpályákkal, konditermekkel, uszodával, biztosítva a legjobb körülményeket a felnőtt csapatok, illetve az akadémisták számára egyaránt. A játékosoktól a megfelelő körülmények biztosítása mellett el van várva, hogy ők maguk is mindent megtegyenek saját fejlődésükért, illetve a csapat sikereiért.

A játékosok teljesítményét mérkőzéseken és edzéseken már a legapróbb részletekig képes a klubok vezetősége megfigyelni, amely részben a 2006-ban Ausztráliában (Catapultsports, 2020) megjelenő úgynevezett „katapult” segítségével valósul meg. A „katapult” egy mérőberendezés, amelyet a játékosok a mellkasukon rögzítve viselnek terhelés (edzések, vagy akár mérkőzések) során. A berendezést már több sportágban is használják a labdarúgáson kívül, igen elterjedt a használata az amerikai sportágakban (baseball, kosárlabda, jégkorong, amerikai futball). Ez a rendszer olyan célt szolgál, mint az egyéb gazdasági szereplők (különböző üzleti vagy állami szervezetek) esetén a munkavállalók teljesítményét mérő eszközök, felmérések. A labdarúgók (illetve egyéb sportolók) esetén azonban folyamatosan megfigyelhető a munkavégzésük minősége, intenzitása, ezáltal is ösztönözve vannak arra, hogy a saját fejlődésük, illetve a csapatuk teljesítményének érdekében is maximumot nyújtsanak. A „katapult” méri többek között a játékosok szívverését, pulzusát, futási sebességét, mozgását, tehát rengeteg adatot bocsát a klubok rendelkezésére. A rendszer segítségével optimalizálható a játékosok terhelése, illetve az egyes feladatok időtartama az edzések során. Hosszú távon pedig, a korábbi adatok összehasonlításával monitorozható a játékosok fizikai állapotának fejlődése is (Channelstv, 2017). A „katapult” segítségével az edzések során összegyűjtött adatokhoz általában kizárólag a klubok férhetnek hozzá. Ezen adatok a külső szemlélők, de még az ellenfelek számára sem bírnának komoly jelentőséggel,

ismeretük nem okozna semmiféle versenyelőnyt. Ilyen adat lehet például a játékosok pulzusa az edzések bizonyos gyakorlatai során vagy a futómennyiség, esetleg a legmagasabb futási sebessége a játékosoknak.

Mérkőzéseken azonban a játékosok teljesítményének titkosítása lehetetlenné válik, hiszen a televíziós közvetítéseknek, illetve a stadionokat behálózó kamerarendszereknek köszönhetően bármely játékos mozgása, illetve teljesítménye megfigyelhető a pályán. Az adatok összegyűjtése a mérkőzések során pedig nem kizárólag az egymás ellen játszó kluboknak, illetve stábjaiknak hordozhatnak értékes információt. A mérkőzések során külső szemlélők is végeznek adatgyűjtést, hogy az érdeklődő szurkolók, illetve akár a sportfogadók számára is minél több és minél részletesebb adat álljon rendelkezésre a játékosok teljesítményéről. Ez az adatgyűjtés az úgynevezett „playertracking” rendszerek segítségével és támogatásával jön létre. Ez egy egészen új aspektusát jeleníti meg a modern labdarúgásnak, ez pedig a videóelemzése a futballpályán történő eseményeknek (Baysal, Duygulu, 2015), melyhez többek között a korábban említett Prozone rendszerek nyújtanak támogatást. Ezen rendszerek által sokkal több információhoz juthatunk a játékosok mozgásáról, erőállapotáról, de akár arról is, hogy mely játékos teljesít egy mérkőzésen önmagához képest alul, esetleg felül. Az alulteljesítésnek természetesen számos oka lehet, például az, hogy a játékosok apróbb sérülésekkel kell, hogy a pályára lépjenek. Amin et al. (2018) szerint a legnagyobb kihívás a labdarúgók mozgásának elemzése során, az a tény, hogy a lépéseik szinte megjósolhatatlanok, pillanatok alatt váltanak pozíciót, valamint sebességet főleg azokban az esetekben, amikor a labda közelébe kerülnek. A labdarúgás úgynevezett aciklikus mozgásformát megkövetelő sportág, mely lényegében azt jelenti, hogy a játékosok mozgásában nem figyelhető meg semmilyen szabály szerinti ismétlődés (ciklikusság) ezzel tovább nehezítve a videóelemzők feladatát. Ezek a „tracking” rendszerek a stadionokban elhelyezett kamerák megfelelő beállításával képesek a játékosok mozgásának folyamatos megfigyelésére. A Baysal és Duygulu (2015) által megalkotott rendszer egy fiktív (modellezett) futballpályát apró, pontszerű részekre oszt fel (mintha egy koordinátarendszer egy pontjai lennének ezek a részek). Egy adott pillanatban egy játékos a fiktív pálya valamely pontszerű részén tartózkodik. A játékosok minden egyes lehetséges tartózkodási helyéhez három koordinátát rendelnek hozzá, mellyel a pályán egy adott pillanatban felvett pozíciójuk jól definiálhatóvá válik.

A videóelemzés nem csupán a játékosok erőállapotának felmérésével, hanem a mérkőzések során előforduló taktikai elemekkel, illetve a játékosok helyezkedésével is egyaránt foglalkozik. Elsősorban a taktikai elemzések összetettségének köszönhetően a videóelemző,

mint külön szakma is megjelent a labdarúgásban. A világ és Európa legnagyobb klubjai mind rendelkeznek videóelemzőkkel, sőt a modern futball során már az utánpótlásról gondoskodó akadémiákon is dolgoznak videóelemző szakértők. László (2021), az UTE labdarúgó akadémiáján videóelemző-edzőként dolgozó szakember világít rá arra, hogy nagy eltérés van azonban a felnőtt és az utánpótlás korú labdarúgók videóelemzési módszertana között. Az utánpótlásban elsősorban az egyéni mozgásokon és játékosok saját helyezkedésén van a hangsúly, illetve a mérkőzéseken elért eredmény kevésbé fontos, mint az egyéni teljesítmény és fejlődés. A felnőttkorú, hivatásos labdarúgóknál azonban a videóelemzések során az ellenfelek tanulmányozásán van a hangsúly (László, 2021), mivel a klubok célja az éppen soron következő ellenfél legyőzése. A professzionális futballban is jelen van tehát a versenytársak megfigyelése azzal a különbséggel, hogy hétről hétre változnak ezek a versenytársak, ez pedig azt eredményezheti, hogy a klubok stratégiája (taktikája) dinamikusabban változhat a labdarúgásban, mint például a vállalati szektorban. A játékosokon és az edzőkön, illetve a klubokat segítő szakembereken hatalmas a nyomás, ugyanis néhány hónap rossz teljesítmény következtében (főleg az edzők esetén) nagyon gyakori, hogy leváltják őket, szerződést bontanak velük, ennek következtében munkanélkülivé válnak.

5. Teljesítményre ható tényezők a labdarúgásban, releváns mutatószámok a szakirodalomból:

Dolgozatom ezen fejezetében arra keresem a választ, hogy mely változók, illetve mutatószámok alakulása korrelálhat a csapatok által nyújtott teljesítménnyel. Ezen mutatószámokat (amelyeket lehetőségem nyílt) beépítettem a kutatásomhoz felhasznált adatbázisba is.

5.1. Labdabirtoklás, illetve passzpontosság és teljesítmény kapcsolata:

Itt többek között olyan alapvető mutatók, mint a *labdabirtoklás*, vagy a *passzpontosság* is helyet kaptak. A *labdabirtoklás* fontossága azon összefüggésen alapul, hogy az a csapat vezet támadást, amelyik a labdát birtokolja, tehát az a csapat támad többet egy mérkőzésen, amelyik többet birtokolt labdát. Jones (2004) angol első osztályú csapatok teljesítménye esetén vizsgálták a *labdabirtoklás* jelentőségét. Az ő modelljük kizárólag a három másodpercnél tovább tartó *labdabirtoklásokat* vette figyelembe, (a labdarúgásban ennél rövidebb ideig tartó támadásból szinte lehetetlen gólhelyzetet kialakítani) és azt az eredményt szolgálta, hogy az élcsapatok jellemzően hosszabb támadásokat építenek, ezek során tovább birtokolják a labdát. A várakozások szerint tehát pozitív korreláció mutatkozhat egy csapat labdabirtoklása, illetve elért pontszáma (teljesítménye) között.

A *passzolás pontosságát* is ugyanezen az elven fogom kalkulálni. A *passzpontosság* és a teljesítmény között azonban a pozitív korreláció fennállása kérdéses, ugyanis a passzpontosság, mint változó minden passzát tartalmazza a mérkőzéseknek. Egy labdarúgó mérkőzésen egy adott passz pontosságát meghatározza például a passzoló játékos elfoglalt helyzete is a pályán, illetve az is, hogy milyen irányba (előre, hátra, oldalra) megy egy adott passz. Saját térfélen jelentősen több a pontos passz, mint az ellenfél térfelén (Whoscored, 2020a), mivel saját térfélen jellemzőbben kisebb az ellenfél által kifejtett nyomás a passzoló játékosra, könnyebb csapattársat megjátszani a labdával. Goes et al. (2018) szerint a passzokat nem kizárólag az irányuk határozza meg, hanem azok elfoglalt szerepe egy taktikai rendszerben, ez tehát azt jelentheti, hogy a hátrafelé irányuló passzok is lehetnek hatékonyságot növelő passzok, amelyek egy-egy figura/taktikai elem megvalósításához hozzájárulhatnak. Mao et al. (2017) a kínai első osztályban (ebben az évben már jó néhány Európa-szerte ismert

játékos játszott Kínában, így bajnokság csapatai szakmai felkészültségben és költségvetésben is fel tudták venni a versenyt az európai bajnokságok csapataival) vizsgálták a jó teljesítményt eredményező kulcs mutatószámokat. Ezen kutatás eredményei alapján pozitív korreláció mutatkozott a *passzpontosság* (emellett az úgynevezett „through ball” (Through ball (magyarul indítás) típusú passz Mao et al. (2017) definíciója alapján olyan passz, amely az ellenfél utolsó védelmi vonalát töri át, és egy csapattársat gólhelyzetbe hoz.) típusú passzok száma is hasonlóan pozitívan korrelált), illetve a mérkőzések megnyerésének valószínűsége között. A *passzpontosságról*, mint változóról azonban előzetes várakozások alapján nem feltételezném biztosan az erős pozitív korrelációs kapcsolatot a csapatok teljesítményével, hiszen ezen mutatószám minden passzt magában foglal minőségtől, valamint hatékonyságtól függetlenül.

5.2. Kapura lövések, illetve támadások hatása a teljesítményre:

A következő mutatószám, a szakirodalomban *SGR* néven ismert, azaz a Shot/Goal Ratio lesz, amelyet érdemes megvizsgálni. Az *SGR* a csapatok által a szezon során szerzett összes góljának, illetve az összes kapura lövésének hányadosa. Ezen mutató százalékosan értelmezhető, azonban előfordulhat egy vizsgált sokaságban gyakrabban, hogy 0 értéket fog felvenni abban az esetben, hogyha egy csapat egyetlen gólt sem tudott szerezni egy adott mérkőzésen. Az *SGR* lényegében a lövések hatékonyságát minősítő mutatószám, így szinte biztosan feltételezhető, hogy minél hatékonyabban értékesíti lövéseit egy csapat, annál jobb eredményeket ér el az egyes szezonok során. A továbbiakban a *lövés* (amely az *SGR* mutató nevezője volt), illetve a *kaput eltaláló lövések* számának vizsgálata is mutathat korrelációt a teljesítménnyel („Lövésből lesz a gól” szállóige alapján legalábbis mindenképpen). Itt fontos megjegyezni, hogy definíció szerint kaput eltaláló lövésnek csak azon lövések számítanak, melyek a hálót érnék abban az esetben, ha nem állna kapus a kapuban (Bigsoccer, 2020). A *kaput eltaláló lövések* közé tehát például a kapufáról kipattanó lövések nem (befelé pattanóak például viszont már igen) számítanak bele. Az összes, illetve a *kaput eltaláló lövések* esetén sem kijelenthető biztosan a pozitív korreláció a teljesítménnyel, ugyanis ezek a mutatószámok a *passzpontossághoz* hasonlóan nem veszik figyelembe a minőséget. Pappalardo és Cintia (2017) ugyan pozitív korrelációt találtak a *lövészsám* és a teljesítmény között a futballpályán, azonban Yue, Broich és Mester (2014) 2011-es mérkőzéseket vizsgálva arra a következésre jutottak, hogy a futballban a *lövés* esetén a minőség jóval fontosabb a mennyiségnél.

Ugyanezen a logika mentén jutottak a modern labdarúgás elemzői, illetve statisztikusai az xG mutató megalkotására (ld: mellékletben az xG-ről szóló fejezet).

A modellbe beépítésre került még a mérkőzésenként *elszenvedett szabálytalanságok* száma is. A szabálytalanságok elszenvedője a labdarúgásban az esetek túlnyomó részében az a csapat, amely éppen támadást indít, tehát ez a mutató is jól jellemezheti egy csapat támadójátékának hatékonyságát, illetve a mérkőzések során támadással töltött időt. Pappalardo és Cintia (2017) a szabálytalanságok alakulásáról az olasz első osztályú labdarúgó bajnokság 2013-14-es szezonját vizsgálva, azt találta, hogy a gyengébben teljesítő (a tabella hátsó régiójában található) csapatok jellemzően több szabálytalanságot követnek el, mint a jobban teljesítő, élbolyba tartozó klubok. Ebből én azt a következtetést vontam le, hogy nagy valószínűséggel azon csapatok, akik ellen az ellenfelek több szabálytalanságot követnek el jobban teljesíthetnek a pályán. Várakozásaim szerint tehát az *elszenvedett szabálytalanságok*, illetve a szerzett pontszám között (erős) pozitív korreláció mutatkozhat.

5.3. Védekezést, illetve mentalitását jellemző mutatószámok:

A szakirodalomban gyakran vizsgálják a védekezés mentalitását jellemző mutatókat, mint például a szerzett *sárga*, illetve *piros lapok* számát, illetve a mérkőzésenként *elkövetett szabálytalanságok* számát is, mint esetlegesen a teljesítményre (többnyire negatívan) ható mutatószámokat. Mindhárom védekezési mentalitást (*szabálytalanságok*, *sárga*, illetve *piros lapok* száma) jellemző változót vizsgálta spanyol labdarúgó mérkőzések kapcsán Lago-Peñas et al. (2010). Az általuk vizsgált mérkőzések során kizárólag a *piros lapok* száma, illetve a teljesítmény között találtak összefüggést, mégpedig erős negatív irányú korrelációt, amely nem meglepő, hiszen egy emberhátrány gyakran hat negatívan egy csapat teljesítményére. Vannak továbbá olyan változók, melyek egy csapat védekezésének hatékonyságát próbálják majd mérni. Ilyen például a *sikeres szerelési kísérletek* száma. Pappalardo és Cintia (2017) a *szerelések* száma és a győzelmek megszerzése között negatív korrelációt vélt felfedezni, ennek a magyarázata az lehet, hogy egy rosszabbul teljesítő csapatnak vélhetően többet kell védekeznie jobban teljesítő ellenfeleikkel szemben, így több szerelést is szükséges bemutatniuk a védőknek a mérkőzések során hétről hétre. Az utolsó vizsgált változó, amely a védekezés hatékonyságát jellemezheti az az *elszenvedett kapura lövések* száma. A *kapura lövések* esetén a korábban is említettem, hogy Pappalardo és Cintia (2017) regressziós elemzésük során pozitív

korrelációt talált a jó teljesítmény és a lövések száma között. Ebből arra a következtetésre jutottam fordított logikával, hogy az *elszenvedett kapura lövések* alakulása és a szerzett pontszám között a negatív lineáris kapcsolat fennállása valószínűsíthető.

5.4. Transzferpiaci tevékenység, valamint a játékosok értékének hatása a teljesítményre:

A további változók, mutatószámok már kevésbé foglalkoznak a mérkőzéseken ténylegesen történő eseményekkel, sokkal inkább a klubcsapatok felkészültségét, futballfilozófiáját, illetve anyagi háttérét hivatottak megvizsgálni. Ezen változók többek között a transzferpiaci tevékenységét jellemzik a klubcsapatoknak. Matesanz et al. (2018) átfogó, több, mint kétezer európai futballklubot vizsgáló elemzésének eredményei azt mutatták, hogy a *transzferpiaci költségek* mértéke és a pályán elért teljesítmény (kontinentális szinten (Bajnokok Ligája, Európa Liga), illetve a saját, hazai bajnokságban is egyaránt) között pozitív kapcsolat áll fenn. Ennek a magyarázata a kutatás (Matesanz et al., 2018) szerint azonban elsősorban az egyenlőtlen pénzügyi feltételekben rejlik az európai futball élmezőnyében. A legtöbb topbajnokság élmezőnyében tartósan megragadó klubok jelentősen nagyobb átigazolási kiadásokkal működnek, mint a hektikusabb teljesítményt nyújtó, szerényebb költségvetéssel rendelkező ellenfelek.

Pénzben kifejezett, azonban az átigazolásokhoz közvetettebben kapcsolódó változó még egy adott klub *játékosainak átlagos értéke* is. A *játékosok értékét* rengeteg tényező határozza meg. Tunaru és Viney (2010) szerint olyan, nem kizárólag statisztikai mutatók is meghatározzák a labdarúgók értékét, mint a sérüléseik száma, a pályán tanúsított magatartás (fegyelmezettség), illetve a pályán nyújtott fizikális teljesítmény (futómennyiség, futási sebesség). Mindezek mellett olyan, elsőre kevésbé jelentősnek tűnő tényezők is alakíthatják a labdarúgók értékét, mint a marketing értékük, illetve akár a személyes (családi) háttérük. Kuper és Szymanski szerint (2014) a játékosok posztjai is befolyásoló tényezők lehetnek, szerintük a legdrágább játékosok leggyakrabban a csatárok, hiszen az ő teljesítményük outputja (a gólszerzés) az egyik leglátványosabb, illetve legjelentősebb elem a labdarúgásban. A *játékosok átlagos értéke* egy csapat játékoskeretében tehát várhatóan pozitívan fog korrelálni az elért teljesítménnyel, azonban fordult már elő a labdarúgásban nagyon sokszor, hogy egy-egy szerényebb bevétellel, illetve alacsonyabb (becsült) értékű játékosokkal rendelkező (angolul ezeket a csapatokat úgynevezett „underdog” néven szokták emlegetni) klub nagyon jó

eredményt ért el egy szezon során. Az ilyen „meglepetés” eredmények vélhetően kiszűrhetők abban az esetben, hogyha több szezonon keresztül vizsgáljuk egy bajnokság eredményeinek alakulását. Kuper és Szymanski (2014) a labdarúgók fizetése, illetve a jó teljesítmény között találtak pozitív korrelációs kapcsolatot még, azonban ezen pénzügyi adatokhoz nehéz hozzáférni, illetve a játékosok pontos fizetése a rendkívül sok szerződésükben foglalt bónusz miatt nagyon nehezen becsülhető meg pontosan.

5.5. Játékoskeretet jellemző mutatók és a teljesítmény összefüggése:

A szakirodalomban találkoztam a klubok játékoskeretének összetételét jellemző különféle mutatókkal. Ilyen volt például a játékosok *átlagéletkora* a keretben. Ezen változó korrelációja a teljesítménnyel kérdéses lehet, hiszen Dendir (2016) kutatása alapján a profi labdarúgók 25 és 27 éves koruk között van karrierjük csúcsán, ez a labdarúgók karrierjének nagyjából a közepére tehető. Ez a csúcs azonban játékosoknál eltérő lehet, hiszen egy karriert a sportban számos külső tényező befolyásolhat, mint például a súlyosabb sérülések. Torgler és Schmidt (2013) szerint az egyéni teljesítmény és a csapattársak életkora között van statisztikailag szignifikáns korrelációs kapcsolat, így nem kizárt, hogy az átlagéletkor befolyásolja a csapatteljesítményt is. A kapcsolat irányából megállapítható lehet az is, hogy a fiatalabb, vagy esetleg az idősebb játékoskeretű klubok a sikeresebbek.

A futballban a saját nevelésű, illetve a fiatal játékosok kérdésköre emberi erőforrás menedzsment szempontból is érdekes lehet (Maqueira et al., 2019). Hiszen akár a vállalati szektorban is vezérelhetik a cégeket eltérő motivációk, jelen esetben a fiatal munkaerő megtartása, fejlesztése, „kinevelése”, vagy esetleg tapasztaltabb/tehetségesebb munkavállalók alkalmazása magasabb fizetésért cserébe. A labdarúgásban is érvényes ez a kérdés, hogy melyik működési modell a kifizetődőbb, illetve melyik a sikeresebb, a fiatal labdarúgókat kinevelni, vagy esetleg megvásárolni őket és idővel, fokozatosan a csapatba beépíteni. A sportokban különös jelentősége van a tehetségmenedzsment kérdéskörének. Hatum (2010) négy fő fázisát különböztette meg a tehetségmenedzsmentnek, melyek a tehetségek bevonása; a felismerésük/feltérképezésük; a tehetségek fejlődéséhez a megfelelő körülmények megteremtése; végül a tehetségek megtartása. Kuper és Szymanski (2014) világítanak rá arra, hogy milyen nehézségekkel kell szembenézzen egy (fiatal) labdarúgó amikor egy új klubhoz csatlakozik. Napjainkban a klubok az új játékosok integrációjának segítésére alkalmaznak munkatársakat,

akik segítenek a játékosoknak a kezdeti nehézségek leküzdésében, illetve sok terhet vesznek le a játékosokról. Az ilyen háttérmunkálatoknak (mint például a nyelvtanulás elkezdése vagy akár a lakás; autókeresésben való segítségnyújtás) komoly szerepe van a fiatal játékosok fejlődésében, hiszen tudnak maximálisan a munkájukra, a labdarúgásra koncentrálni, illetve a beilleszkedési folyamat is nagy mértékben felgyorsulhat ezáltal. Rengeteg kutatás, illetve filozófia szerint a legtöbb sportban a legalkalmasabb életkor fiatal játékosok leszerződtetésére a főiskolai kor elején van, azaz a 20-as éveik legelején (Lewis, 2011, Kuper, Szymanski, 2014). Ebben az életkorban a sportolók ugyanis (amennyiben igazán tehetségesek) már minimális tapasztalattal (legalább 2-3 szezon során már kaptak játéklehetőséget nevelőklubjuk felnőtt keretében) rendelkeznek a felnőtt korúak közötti mérkőzésekről.

Azon változók, melyekkel a fiatal játékosok (tehetségek) bevonásának mértékét mérni lehet, azok a *saját nevelésű játékosok aránya* az egyes csapatok játékoskereteiben az adott szezonban, illetve ezen saját nevelésű játékosok esetén az egyesek szezonok során a *pályán töltött játékidő* (percben). A *játékidő* által árnyaltabb képet kaphatunk arról, hogy mely csapatok adnak több szerepet a szezon során saját nevelésű játékosainak. A korreláció erőssége ezen változók esetén is kérdéses lehet, azonban az irányuk alapján (pozitív/negatív korreláció) ezen változók vizsgálata is hordozhat érdekességeket.

Érdeemes lehet vizsgálni továbbá a *fiatal játékosok játékidőjét* is az egyes szezonokban. Ezzel a változóval kiszűrhetők azon csapatok, amelyek kevesebb lehetőséget adtak fiataljaiknak, ezáltal kissé háttérbe szorították fejlődésüket (véltetően a minél jobb eredmény elérése érdekében). Ezen játékoskeretet jellemző változó esetén is a korreláció irányára érdemes a hangsúlyt helyezni, mivel nehéz megítélni a korreláció erősségét előzetesen.

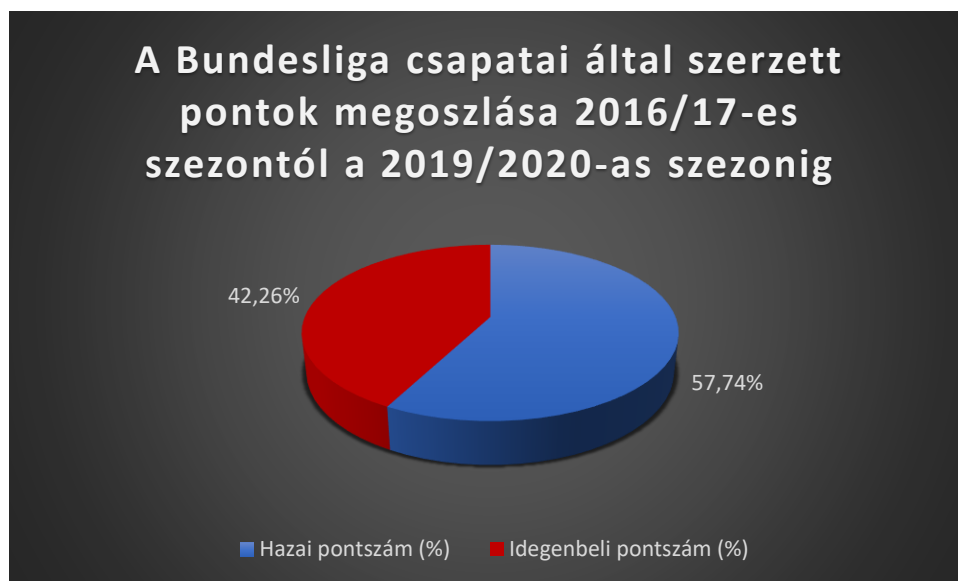
Végül a *külföldi játékosok aránya* a csapatok játékoskeretében is lehet hatással egy klub teljesítményére. Royuela és Gásquez (2018) szerint az *idegen nyelvű játékosok* magas *aránya* elsősorban a nemzetközi porondon való szereplést befolyásolja, még hozzá pozitívan korrelál a több külföldi játékos a jobb szerepléssel. Azonban a saját, hazai bajnokságokban ez a korreláció már nincsen feltétlenül jelen, hiszen a csapatokra ugyanazok a szabályok vonatkoznak az idegenlégiósok foglalkoztatására (különbféle megkötések vannak egyes bajnokságokban, például maximalizálják az egy időben pályán tartózkodó külföldi játékosok számát). Az idegen nyelvű játékosok mennyiségének szabályozása alapvetően a hazai játékosok védelmét és fejlődését célozza meg, ezzel a hazai labdarúgás fejlődését szem előtt tartva. A CIES (International Centre for Sports Studies) labdarúgással foglalkozó részlege 2017-ben végzett egy kutatást (ESPN, 2017), mely során azt vizsgálták, hogy az európai klubok a külföldi játékosokat az

elérhetőségük (elérhetőség alatt egy adott meccsen a játékoskeretbe való bekerülést értjük, tehát azt az időszakot, amikor egy játékos futballra bevethető, nem sérült, illetve nem is eltiltott) hány százalékában tartották a pályán. Ha egy klub esetén ez a százalékos mutató nagyon magas, abban az esetben az a klub nagy mértékben támaszkodik a külföldi játékosokra. A CIES kutatásának eredménye az lett, hogy Európában kiemelkedően a ciprusi klubok támaszkodnak a külföldi játékosokra, mintegy 80.2%-ban játszatják az elérhető külföldi labdarúgókat. A kutatásom során vizsgált bajnokság, a német Bundesliga esetén ez az arányszám 50.5%-ra volt tehető, ez a mutatószám Európa legrangosabb bajnokságai esetén átlagosnak volt mondható. A Bundesligában nincsenek megkötések a külföldi játékosok számának korlátozását illetően, így a Royuela és Gásquez (2018) által vélt pozitív korreláció a *külföldi játékosok* magasabb *aránya*, illetve a jó teljesítmény között megmutatkozhat.

5.6. Nézőszámok, illetve a szurkolói aktivitás és a teljesítmény kapcsolata:

A továbbiakban a szurkolói aktivitást jellemző mutatókat is érdemes vizsgálni. Ezen paramétert az *átlagos mérkőzésenkénti nézőszám*, illetve a *stadionok kihasználtsága* egyaránt jól jellemezhetik. A *kihasználtság* az *átlagos nézőszám*, illetve a saját stadion befogadóképességének hányadosa, százalékosan értelmezhető mutatószám, minél nagyobb az értéke, annál jobban megtöltik a szurkolók egy csapat stadionját egy átlagos bajnoki mérkőzés során. Korábban említettem, hogy az általam vizsgált német labdarúgó első osztályban nagy jelentőséggel bír a szurkolói réteg, illetve 2014-ben a német labdarúgó bajnokságban Pudasaini (2014) szerint a világ összes sportligája közül második legmagasabb volt az átlagos nézőszám a német bajnoki mérkőzéseken több, mint 42 ezer fővel. Ezt a rendkívül magas átlagos nézőszámot részben a klubok tulajdonosi szerkezetének, részben a (európai toptligák viszonylatában) kedvező jegyáraknak, részben pedig a német emberek futballszeretetének köszönhetik. Fontos kihangsúlyozni, hogy az adatbázisom összeállításához a 2019/2020-as labdarúgó szezonban az átlagos nézőszám a COVID-19 vírus előtti időszakból származik, hiszen a vírus megjelenése és a bajnokság újrakezdése (2020. 03.11 és 2020.05.16-a között nem rendeztek meg egyetlen mérkőzést sem a bajnokságban) után már kizárólag zártkapus mérkőzéseket rendeztek a német Bundesligában. A nézőszámok, illetve a stadionok kapacitásának kihasználása esetén kérdéses a korreláció fennállása a teljesítménnyel, azonban Littkemann és Kleist (2002) is megfigyelték, hogy a hazai csapatok gyakrabban kerülnek ki győztesként a német labdarúgó bajnokságban.

Az alábbi kördiagram a német Bundesliga elmúlt négy befejezett szezonjában a hazai pályán, illetve idegenben szerzett pontok megoszlását láthatjuk:



2. ábra: Pontok megoszlása a Bundesligában.
Saját készítés.

Az 2. ábrán jól látható, hogy a pontok csaknem 58%-át szerezték négy vizsgált szezonon keresztül a hazai csapatok, tehát a saját szurkolók előtt játszott mérkőzéseken többnyire sikerebben szerepeltek a klubok, mint idegenben.

Azonban szezononként vizsgálva külön-külön az alábbi módon alakult a hazai, illetve idegenbeli pontszám megoszlása:

Szezon	Hazai pontszám (%)	Idegenbeli pontszám (%)
2016-2017	62,09%	37,91%
2017-2018	59,88%	40,12%
2018-2019	57,63%	42,37%
2019-2020	51,41%	48,59%

1. táblázat: Pontszámok megoszlása a négy vizsgált szezonban.
Saját készítés.

Az 1. táblázatban egyértelműen látszik, hogy a trend, mely szerint a hazai csapatok szereznek több pontot a Bundesligában mérséklődik. Az is látható, hogy a 2019/2020-as szezonban már szinte ugyanannyi pontot szereztek a csapatok hazai pályán, mint idegenben. Ebben a szezonban azonban 2020 március 16-ától a szurkolók nem látogathattak mérkőzéseket a Bundesligában. Valószínűleg ez is hozzájárult ahhoz, hogy a hazai pályán szerzett pontok

aránya erősen visszaesett a korábbi évekhez képest. A *szurkolói aktivitás*, illetve a jó hangulat a hazai mérkőzéseken, tehát várakozásaim szerint pozitív hatással lesz az elért teljesítményre.

5.7. Egyéni teljesítmény hatása a csapat szintű teljesítményre:

Torgler és Schmidt (2013) a kutatásuk során azt vizsgálták, hogy a csapatok teljesítménye, illetve működése mennyiben befolyásolja az egyéni teljesítményt a német labdarúgásban. Én ezen kutatási kérdésen kísérleti jelleggel fordítottam és olyan mutatószámokat kerestem, amelyek segítségével esetleg található korrelációs kapcsolat az egyéni, illetve a csapatteljesítmény között. Littkemann és Kleist (2002) a Kicker magazin *játékosértékeléseit* is felhasználták kutatásuk során. A játékosok értékelése mérkőzésenként zajlik 10-es skálán. A Whoscored (2020b) értékelései alapján azt vizsgálom meg majd, hogy az egyes csapatok *legjobb játékosának* szezonvégi (minden meccsnek az átlagát vett értékelés) *értékelése* és a csapatok teljesítménye között mutatkozik-e korreláció. További mutatószámok lehetnek, melyek a kiemelkedő egyéni teljesítmények mérésére fognak szolgálni az adott csapatokban a szezon során *legtöbb gólt szerző játékosok* találatainak száma, illetve a *legtöbb gólpasszt jegyző játékosok* gólpasszainak száma. Ezen egyéni teljesítményt mérni szándékozó változók problémája lehet, hogy egy klubon belüli holtversenyt nem képesek kezelni. Ezen változókkal a célom az elemzésem során, hogy megvizsgáljam, hogy egyetlen kiemelkedő teljesítmény okozhat-e szignifikáns teljesítményjavulást a futballpályán csapat szinten egy teljes labdarúgó szezon során. Amennyiben okozhat, úgy arra is választ kaphatunk, hogy milyen stílusú játékos kiemelkedő teljesítményének van nagyobb ráhatása egy csapat teljesítményére (bármely poszton kiemelkedő játékos (a *rating* alapján), kiemelkedő *góllövő* játékos, illetve kiemelkedően sok *gólpasszt szerző* játékos).

6. Kutatás, a német Bundesliga elmúlt négy szezonjának vizsgálata az ökonometria eszköztárával:

A kutatás célja, hogy korrelációs kapcsolatot találjak (többségében leíró) sportstatistikai mutatók, illetve a csapatok elért eredményei, a labdarúgóklubok teljesítménye között. A kutatás eredményei akár prediktív jelleggel is szolgálhatnak azonban a regressziós modellek hibatagja (egyes szakirodalmak errornak, vagy véletlen tagnak nevezik) miatt a regressziós modellekkel készített előrejelzések, prediktív statisztikák készítése során hordozhat pontatlanságokat. Kutatásom során lineáris regressziós elemzéssel fogom az általam, manuálisan összeállított adatbázist elemezni. A regressziós modellem eredményváltozója az aktuális szezonban *szerzett pontok száma* lesz egy adott szezon végén, a labdarúgóklubokat nominális változóként fogom elhelyezni a modellben, így a kvantitatív elemzések során nem lesz szerepe ezen változónak. Attól, hogy mely német első osztályban szereplő labdarúgó klub érte el az adott pontszámot, teljes mértékben el fogok tekinteni. A csapatok nevei, illetve helyezésük az adott szezonban szintén fel van tüntetve a modellben, azonban ezen változókat sem használtam fel semmilyen kalkulációra. Adatbázisomban az 5. fejezetben ismertetett változók foglalnak helyet. Az átláthatóság kedvéért a következő táblázat foglalja össze az általam elemzésre kiválasztott huszonöt darab változót. A táblázat tartalmaz egy hipotézist is arra vonatkozóan, hogy az előzetes várakozások, illetve szakirodalmi kutatások szerint milyen irányú korreláció mutatkozhat az egyes magyarázó változók, illetve az eredményváltozó között. A változókat további csoportokra osztottam melyek a következők lettek: támadást jellemző mutatószámok, védekezést jellemző mutatószámok, transzferpiaci, illetve pénzben kifejezett változók, játékoskeret-specifikus változók, valamint egyéb változók (ezek bekategorizálása nem volt egyértelmű).

6.1. Adatbázis összeállítása, változók kiválasztása:

Változó csoportja	Változó neve (mértékegysége)	Várakozás
Eredményváltozó	Egy teljes szezon során szerzett pontszám	
Támadást jellemző	Labdabirtoklás (%)	pozitív
	Passzpontosság (%)	kérdéses

	SGR (%)	pozitív
	Lövések (db/mérkőzés)	kérdéses
	Kaput eltaláló lövések (db/mérkőzés)	kérdéses
	Elszenvedett szabálytalanságok (db/mérkőzés)	pozitív
Védekezést jellemző	Sárga lapok (db)	kérdéses
	Piros lapok (db)	negatív
	Elkövetett szabálytalanságok (db/mérkőzés)	kérdéses
	Sikeres szerelések (db/mérkőzés)	negatív
	Elszenvedett kapura lövések (db/mérkőzés)	negatív
Transzferpiaci tevékenységet jellemző	Nettó átigazolási bevétel (millió euró)	kérdéses
	Nettó átigazolási költség (millió euró)	pozitív
	Nettó átigazolási egyenleg (millió euró)	kérdéses
	Átlagos játékosérték (millió euró)	pozitív
Játékoskeret-specifikus	Átlagéletkor (év)	kérdéses
	Saját nevelésű játékosok aránya (%)	kérdéses
	Saját nevelésű játékosok játékideje (perc)	kérdéses
	20 év alatti játékosok játékideje (perc)	kérdéses
	Külföldi játékosok aránya (%)	pozitív
	Stadion kihasználtsága (%)	pozitív

Egyéb, kísérleti jellegű változók	Átlagos nézőszám (néző)	pozitív
	Legjobb játékosértékelés	kérdéses
	Legtöbb gól/játékos	kérdéses
	Legtöbb gólpasz/játékos	kérdéses

2. táblázat: A modell változói és csoportosításuk, illetve a korreláció iránya az előzetes várakozások alapján.

Saját készítés.

A modern labdarúgás nagy előnye, hogy rengeteg statisztikával, eredményekkel, elemzésekkel foglalkozó adatbázist segítségül hívhatunk egy elemzés, illetve adatbázis elkészítéséhez. Az általam összeállított adatbázis alapjául a világ egyik legnépszerűbb footballstatisztikai portálját a Whoscored-ot (2020b) használtam, amely szintén az OPTA adatbázisai alapján gyűjt adatokat. Az adatgyűjtést manuálisan végeztem, mely során a fókusz arra helyeztem, hogy minél több mutatószám helyet kapjon a végső adatbázisban. A kész adatbázis huszonöt darab mutatószámot (a regressziós modell magyarázó változóit) tartalmazott, egy darab eredmény változót (*szerzett pontszám*), illetve két darab nominális változót (szezón/helyezés, valamint a csapatok neve). A nominális változóknak nem volt szerepük a kalkuláció során. Az adatbázis elkészítésének nagyfokú manualitása eredményezte azt, hogy a vizsgált sokaság nagysága végül 72 elemből állt (ez négy teljes Bundesliga szezón adatait jelenti 2016/17-es szezontól kezdve a 2019/2020-as szezónig). Ekkora sokaságon egy sokváltozós modellt nem érdemes lefuttatni, ugyanis az eredmények kevésbé lesznek pontosak, így egy a hagyományostól eltérő módon végeztem a modellszelekciót első körben.

6.2. Modell szűkítése lineáris korrelációk segítségével:

Először az SPSS program segítségével a változókra külön-külön futtattam egyváltozós lineáris regressziót. Minden egyes változó determinációs együtthatóját (R mutató), mely a lineáris korreláció mértékét és irányát mutatja meg a magyarázó változóval, illetve empirikus szignifikancia-szintjét is figyelembe véve fogom a modellem változóinak számát leszűkíteni a valóban szignifikáns változókra (Claro, 2019).

Magyarázó változó	Korreláció	Szignifikancia
Labdabirtoklás	0,8	0,000***
Paszspontosság	0,685	0,000***

SGR	0,806	0,000***
Lövések	0,684	0,000***
Kaput eltaláló lövések	0,831	0,000***
Elszenvedett szabálytalanságok	0,146	0,222
Sárga lapok	-0,481	0,000***
Piros lapok	-0,342	0,003***
Elkövetett szabálytalanságok	-0,422	0,000***
Sikeres szerelések	-0,205	0,084
Elszenvedett kapura lövések	-0,751	0,000***
Átigazolási bevétel	0,4	0,000***
Átigazolási költség	0,573	0,000***
Átigazolási egyenleg	0,071	0,551
Játékosérték	0,833	0,000***
Átlagéletkor	-0,236	0,046**
Saját nevelésű játékosok aránya	0,243	0,040**
Saját nevelésű játékosok játékidéje	0,007	0,952
20 év alatti játékosok játékidéje	0,191	0,108
Külföldi játékosok aránya	0,036	0,766
Stadion kihasználtsága	0,289	0,014**
Átlagos nézőszám	0,51	0,000***
Legjobb játékosértékelés	0,812	0,000***
Legtöbb gól/játékos	0,759	0,000***
Legtöbb gólpassz/játékos	0,789	0,000***

3. táblázat: A magyarázó változók lineáris korrelációja, illetve szignifikanciája a pontszám alakulásával.

Saját készítés.

Jelmagyarázat: ***: 1%-on; **: 5%-on; *: 10%-on szignifikáns a változó

A lineáris korrelációk segítségével a 3. táblázat első oszlopában piros betűszínnel jelzett változók elhagyásával szűkült a modellemben helyet kapó változók száma. Az eliminált

mutatószámok azok, amelyek 5%-os szignifikancia-szinten nem bizonyultak szignifikáns változónak (Ramanathan, 1993). Látható, hogy egy szignifikáns változót is megjelöltem az elhagyandó változók között, ez pedig a kapura lövések száma volt. Ennek az oka, hogy a lövések, illetve kaput eltaláló lövések közötti magyarázóváltozók közül további kalkulációra a szignifikánsabb, illetve magasabb determinációs együtthatóval rendelkező kaput eltaláló lövések változót fogom felhasználni a modellszűkítés után. A két fent említett változó között erős a multikollinearitás, amely a magyarázó változók közötti lineáris kapcsolatot jelenti (Hunyadi, Vita, 2018). Ezt az úgynevezett VIF-mutatóval is (Variance Inflation Factor) mérhetjük, mely megmutatja, hogy valamely változó becsült együtthatójának tényleges varianciája hányszorosa annak, mint ami a multikollinearitás teljes hiányában lenne (Polt, 2018). Értéke 5 (de a GRETL ökonometria szoftver szerint például csupán 10 fölött) fölött mondható károsnak. Egy háromváltozós (pontszám, lövések, kaput eltaláló lövések) lineáris regresszió lefuttatása után egyértelműen megmutatkozott a lineáris kapcsolat mértéke a magyarázó változók között, a VIF-mutató értéke ezen modellben 5,864 volt, így az egyik változó eliminálása további kalkulációk végzéséhez mindenképpen indokolt. Ezen fejezetben tehát hét változót (ld.: 3. táblázat) elimináltam a teljes modellből.

6.3. Modell szűkítése backward módszerrel:

Az SPSS statisztikai szoftverben a modellben bent maradt tizenhét változót tovább szűkítettem a programban található funkcióval, úgynevezett backward módszerrel. A backward módszer lépésként távolít el változókat a lineáris regressziós modellből, oly módon, hogy egy adott magyarázó változó elhagyásával kapott lineáris regressziós modell magyarázóereje és szignifikanciája jelentősen ne csökkenjen. A változók eliminálásának befejezését két módszerrel lehet meghatározni (Choueiry, 2019): Az egyik, a szignifikanciák vizsgálata a modellben minden egyes eliminálás után. Itt egy tetszőleges szignifikancia-szint (jellemzően 1%, 5%, 10% vagy ritkább esetekben 20%) meghatározása után az utolsó eliminálás a modellből akkor történik, amikor az egyetlen, a kiválasztott szignifikancia-szinten nem szignifikáns változót is eltávolítottuk. A következő modellben ezek után még előfordulhat, hogy lesznek (a kiválasztott szignifikancia-szinten) nem szignifikáns változók, azonban elhagyásukkal már jelentősebben csökkenne a modell magyarázóereje. A másik módszer az információs kritériumok vizsgálata. Az információs kritériumok segítenek, hogy eldöntsük, a bővebb, esetleg a szűkebb modell hasznosabb számunkra. Az egyik leggyakrabban használt

információs kritérium az Akaike-féle, mely a modell devianciájának (hibájának), illetve a paraméterek kétszeresének összege. Az Akaike-kritérium hiba alapú mutató (Ferenci, 2018a), tehát minél nagyobb értéket vesz fel, annál rosszabb egy vizsgált regressziós modell.

Én a modellszűkítés leállítását az első módszer, azaz egy meghatározott szignifikancia-szint alapján végeztem el így tizenkettő további változót hagyhattam el a modellből egyesével. Az elimináció sorrendje a következő volt:

1. *Átlagos nézőszám*
2. *Passzpontosság*
3. *Átigazolási költségek*
4. *Egy játékos által adott legtöbb gólpassz*
5. *Stadion kihasználtsága*
6. *Kaput eltaláló lövések száma*
7. *Átigazolási bevételek*
8. *Átlagéletkor*
9. *Egy játékos által szerzett legtöbb gól*
10. *Elkövetett szabálytalanságok száma*
11. *Játékosok átlagos értéke*
12. *Saját nevelésű játékosok aránya*

6.4. Végző modell változói, illetve regressziós eredmények összegzése:

A végző modell már csupán összesen hat magyarázó változót tartalmazott, ezen modell korrelációs együtthatója 0,939 volt, míg a magyarázó változók az eredményváltozó (pontszám) szóródását 88,2%-ban magyarázzák ($R^2=0,882$). A modell magyarázóereje tehát rendkívül magas, azonban, eredményeinek értelmezése előtt szükséges a modellben a multikollinearitás, illetve a heteroszkedaszticitás jelenlétét is letesztelni, ezek a jelenségek ugyanis torzíthatják egy regressziós elemzés eredményeit (Ferenci, 2018a, 2018b).

A backward-módszer után kapott regressziós eredmények:

Változó	Együttható	Standard hiba	t érték	p-érték	VIF-mutató
konstans	-53,795	30,169	-1,783	0,079	-
Labdabirtoklás	0,470	0,243	1,931	0,058*	3,600
SGR	2,054	0,362	5,669	0,000***	2,048
Sárga lapok	-0,155	0,075	-2,072	0,042**	1,533
Piros lapok	-0,850	0,386	-2,202	0,031**	1,201
Elsz. kapura lövések	-1,955	0,502	-3,897	0,000***	2,606
Egyéni rating	12,522	3,792	3,302	0,002***	2,731

4. táblázat: Végző modell regressziós eredményei.
Saját készítés.

Ezen modellre elvégeztem egy ANOVA-tesztet is, melynek segítségével tesztelhető a modell magyarázó változóinak szignifikanciája. Az R-négyzet magas értéke, illetve a 4. táblázat szignifikanciái alapján várhatóan az ANOVA-teszt eredményei is statisztikailag szignifikánsnak fogják a modellt nyilvánítni.

$H_0: \beta_{\text{labdabirt}} = \beta_{\text{SGR}} = \beta_{\text{sarga}} = \beta_{\text{piros}} = \beta_{\text{elsz_lovesek}} = \beta_{\text{rating}} = 0$, tehát a modell egyik változója sem szignifikáns

H_1 : valamely $\beta_j \neq 0$, tehát a modell tartalmaz szignifikáns változót

Variancia eredete	Eltérés négyzetösszeg	Szabadságfok	Átlagos négyzetösszeg	F próbafüggvény	p-érték
Regresszió	14098,817	6	2349,803	80,822	0,000
Maradék	1889,794	65	29,074		
Σ	15988,611	71	2378,877		

5. táblázat: Az ANOVA-teszt eredményei.
Saját készítés.

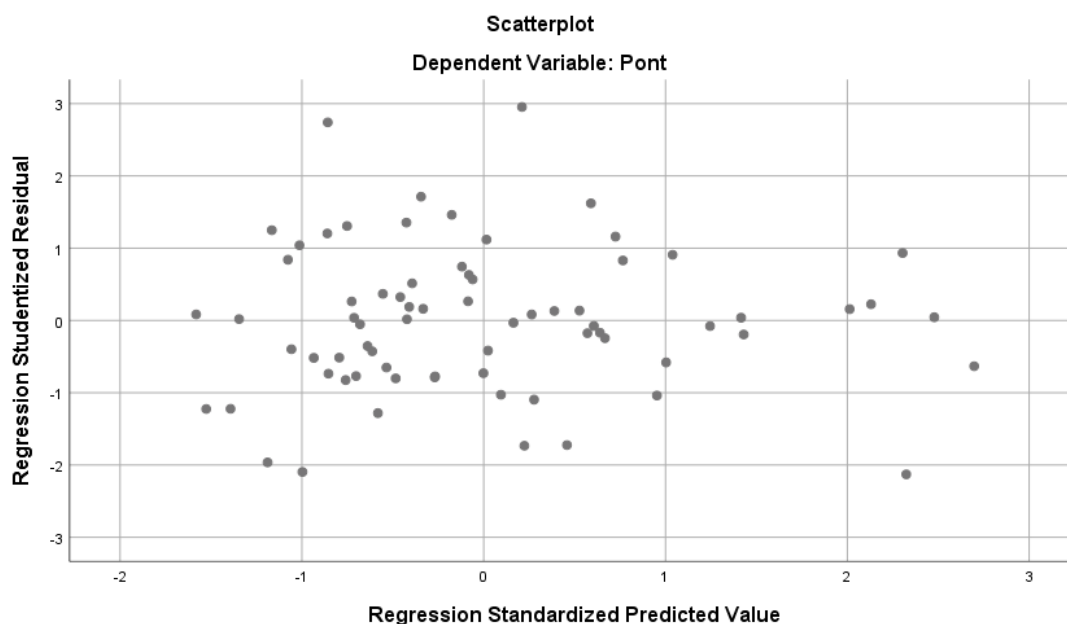
Döntés p-érték alapján: Az 5. táblázatban az ANOVA-teszt p-értékének értéke 0,000, amely kisebb az összes szokásos szignifikancia-szint értékénél. 1% (5%, illetve 10%) -os szignifikancia-szinten is a H_1 hipotézis kerül elfogadásra, tehát a modell tartalmaz nullától szignifikánsan eltérő regressziós együtthatójú változót. Tehát az ANOVA-teszt alapján is kijelenthető, hogy a végző regressziós modell statisztikailag szignifikáns.

6.5. Multikollinearitás vizsgálata:

Multikollinearitás alatt a magyarázó változók közötti lineáris kapcsolatot értjük, melynek jelenléte torzíthatja egy regressziós elemzés eredményeit (2018a), illetve a multikollinearitás fennállása semmissé teszi a ceteris paribus elvet a regressziós együtthatók értelmezésekor (hiszen a változók nem függetlenek egymástól). A multikollinearitás jelenlétét modellemben a VIF-mutatók alapján vizsgáltam meg (ld.: 4. táblázat utolsó oszlopa). Polt (2018) szerint a VIF-mutató értéke 5 fölötti érték esetén káros, az én modellemben a legnagyobb felvett érték 3,6 (a labdabirtoklás esetén), ezen változó esetén már kis mértékben zavaró a multikollinearitás, azonban a változó elhagyása a modellből további kalkulációk elvégzéséhez nem szükséges.

6.6. Heteroszkedaszticitás tesztelése:

Heteroszkedaszticitás alatt azt értjük, amikor a regressziós modell hibatagjainak szóródásának varianciái nem állandók (Ramanathan, 1993, Hunyadi, Vita, 2008). Heteroszkedaszticitás fennállása mellett a legkisebb négyzetek módszere (Ordinary Least Squares – OLS), amellyel én is dolgozni fogok a regressziós elemzésem során, elveszíti a hatásosságát (Ferenci, 2018b). (A hatásosságot a becslések BLUE, azaz Best Linear Unbiased Estimator tulajdonságának is nevezi a szakirodalom) A BLUE tulajdonság elvesztése azt eredményezné, hogy a nagyobb varianciájú megfigyelések kevesebb információtöbbletet nyújtanak a kisebb varianciájú megfigyeléseknél (Claro, 2019).



3. ábra: Hibatagok szóródása a regresszió standard értékeinek függvényében.

Saját készítés SPSS segítségével.

A 3. ábra alapján nem dönthető el egyértelműen, hogy a modell heteroszkedasztikus. Az értékek nagy része a diagramon -2 és 2 közé esik, tehát egy meghatározott tartományon belül vannak azonban találkozhatunk néhány kiugró értékkel is. Továbbá az sem kijelenthető, hogy a nagyobb regressziós értékekhez nagyobb hibatagok tartoznának. Ezáltal a további tesztek eredményei alapján fogok döntést hozni a modell heteroszkedaszticitásáról, illetve annak kezeléséről. A heteroszkedaszticitás további vizsgálatához először a White-tesztet fogom alkalmazni, amely az SPSS programcsomagnak nem része, így a kalkulációt a GRETL ökonometriai programmal végeztem el. A White-teszt a vizsgált regressziós modellhez egy segédregressziót készít, melynek eredményváltozója a hibatagok négyzete. Ezen segédregresszió tagja lesz az összes változó négyzete, illetve az összes magyarázó változó interakciója (interakció alatt két változó értékeinek keresztszorzatait értjük) páronként. A White teszt Khí-négyzet eloszlást követő próba, melynek hipotézispárja a következő:

H0: A vizsgált regressziós modell homoszkedasztikus (a hibatagok varianciájának szóródása közel azonos)

H1: A vizsgált regressziós modell heteroszkedasztikus

Ezen módszerrel tehát összesen huszonkét új változó került a modellbe, mely az alábbi eredményeket hozta:

```
White's test for heteroskedasticity -
Null hypothesis: heteroskedasticity not present
Test statistic: LM = 38,8555
with p-value = P(Chi-square(27) > 38,8555) = 0,0652993
```

4. ábra: A White-teszt eredményei.
Saját készítés GRETL szoftver segítségével.

Döntés p-érték alapján: A p-érték 0,0652993, mely nagyobb a meghatározott 0,05-ös szignifikancia-szintnél, így a White-teszt alapján 5%-os szignifikancia-szinten a nullhipotézis került elfogadásra, mely szerint a modell homoszkedasztikus.

A White-teszt megbízhatósága megkérdőjelezhető kisebb mintán végzett mintavétel esetén (Ferenci 2018b), mivel az általam vizsgált sokaság csupán 72 elemet tartalmaz, így újabb tesztek szükségesek a modellen elvégezni, hogy egyértelműen döntést tudjak hozni a hibatagok varianciájának szóródásáról.

A következő tesztek, amelyeket a modellen végrehajtottam azok úgynevezett LM-próbák voltak (Ferenci, 2018b). Az LM próbák nagy előnye, hogy minden rendelkezésre álló információt felhasználnak, azonban a hibanormalitásra érzékenyek lehetnek. A két teszt az úgynevezett Breusch-Pagan-próba, illetve ezen próba hibanormalitásra robosztusabb változata, a Koenker-próba lesznek. Ezen tesztek hipotézispárjukat tekintve ekvivalensek a White-teszttel. Ezen két teszt lefuttatásának eredményeit a lineáris regressziós modellel az alábbi, 6. táblázat mutatja be:

Heteroszkedaszticitás-teszt	Érték
Breusch-Pagan ($K\chi^2$)	17,5973
Breusch-Pagan (p-érték)	0,0073
Koenker ($K\chi^2$)	14,6347
Koenker (p-érték)	0,0233

6. táblázat: A Breusch-Pagan, illetve Koenker-próbák eredményei.
Saját készítés.

A hipotézisvizsgálat során döntésemet ismételten az empirikus szignifikancia-szintek (p-értékek) alapján hoztam meg. 5%-os szignifikancia-szintnél mindkét próba esetén kisebb értékeket 0,0073-et, illetve 0,0233-et kaptam a p-értékekre, melyek következtében a próbák

alternatív hipotézisét kell elfogadjam, mely szerint a modellem heteroszkedasztikus. A heteroszkedaszticitás bizonyos szinten orvosolható lineáris regressziós modellben robusztus standard hibák használatával (Claro, 2019), így a regressziós modellt ismételten lefuttattam, ezúttal a robusztus standard hibák használatával.

6.7. Regressziós modell robusztus standard hibákkal:

A lineáris regresszió eredményei robusztus standard hibákkal lefuttatva tehát a következők lettek (a standard hibák, a t-értékek, illetve a p-értékek is változtak kis mértékben; a regressziós együtthatókat, illetve a VIF-mutatókat nem érintették a változások):

Változó	Együttható	Standard hiba	t érték	p-érték	VIF-mutató
konstans	-53,795	26,673	-2,017	0,048**	-
Labdabirtoklás	0,470	0,273	1,719	0,090*	3,600
SGR	2,054	0,385	5,335	0,000***	2,048
Sárga lapok	-0,155	0,088	-1,764	0,083*	1,533
Piros lapok	-0,850	0,315	-2,699	0,009***	1,201
Elsz. kapura lövések	-1,955	0,629	-3,109	0,003***	2,606
Egyéni rating	12,522	3,072	4,076	0,000***	2,731

7. táblázat: A regresszió eredményei robusztus standard hibákkal kalkulálva.
Saját készítés.

Jól látható, hogy a modell változóinak p-értékei kis mértékben változtak a korábbihoz képest, a Piros lapok száma már 5% helyett 1%-on is szignifikáns változó lett, míg a Sárga lapok száma 5% helyett már csupán 10%-os szignifikancia-szinten szignifikáns. A korrelációs együttható (r) értéke 0,939 lett, míg a determinációs együttható értéke 0,882. A modell magyarázóerejéből nem veszített, a magyarázó változók tehát 88,2%-ban magyarázzák a csapatok által szezon végén szerzett pontszám szóródását.

A végső modell regressziós egyenlete pedig a következő lett:

$$\hat{Y}_{\text{pont}} = -53,795 + 0,470 \cdot X_{\text{labdabirt}} + 2,054 \cdot X_{\text{SGR}} - 0,155 \cdot X_{\text{sarga}} - 0,850 \cdot X_{\text{piros}} - 1,955 \cdot X_{\text{elsz_lov}} + 12,522 \cdot X_{\text{egyeni_rating}}$$

Ezen egyenlet együtthatóinak értelmezése:

A lineáris regressziós modell eredményei alapján amennyiben egy klub 1%-kal magasabb labdabirtoklást ér el egy szezon során, akkor várhatóan ceteris paribus 0,47-dal emelkedik a szezon végén szerzett pontjainak száma. Abban az esetben, hogyha egy klub által elért SGR növekszik 1%-kal, akkor azonban már várhatóan c.p. 2,054 ponttal emelkedik ezen klub szezon végi pontszáma. A lövési hatékonyság javítása tehát rendkívül fontos szempont, amelyet érdemes a csapatoknak szem előtt tartani. A lövési hatékonyság javítása elsősorban a leadott lövések minőségének javításával érhető el. A lövések minőségét jellemző mutatószám az úgynevezett xG-mutató. Ez a mutató megmutatja, hogy mekkora valószínűséggel születik gól egy lövésből historikus adatokat felhasználva, illetve több külső tényezőt is figyelembe véve. (Az xG-mutató használatáról, illetve kiszámításáról bővebben a mellékletben olvashat.)

A teljes szezonban szerzett sárga lapok száma amennyiben eggyel nő, abban az esetben a szerzett pontszám c.p. 0,155-del fog csökkenni. Míg a szezon során szerzett piros lapok számának egységnyi (egy darabbal való) növekedése már c.p. 0,85-dal kevesebb várható pontszámot jelent a modell alapján. A továbbiakban az elszenvedett kapura lövések számának minimalizálása is fontos lehet a minél jobb teljesítmény elérésének érdekében, hiszen amennyiben mérkőzésenként eggyel több lövést szenved el egy csapat a szezon során, akkor c.p. 1,955-al csökken a várható szerzett pontszáma a német Bundesliga-ban. Végezetül amennyiben az egyes csapatok legjobb játékosának mérkőzésenkénti átlagos értékelése a teljes szezon során egy egész ponttal nőne, abban az esetben c.p. 12,522 ponttal növekedne a csapatok teljesítménye. Ennek az a magyarázata, hogy ezen változó rendkívül alacsony szórással rendelkezik, illetve egy tízes skálán felvett számértéket jelöl.

7. Előrejelzés a modellel, célszámok megfogalmazása, illetve elérésükhöz

szükséges teljesítmény:

7.1. Célpontszámok meghatározása, az „átlagos csapat”:

Az alábbi, 8. táblázatban foglaltam össze, hogy az általam vizsgált négy Bundesliga-szezon helyezettjei hány pontot gyűjtöttek átlagosan, illetve mennyi pont volt az adott helyezéseken zárt csapatok pontszámának maximuma az elmúlt négy szezont vizsgálva. Azért tartottam fontosnak a maximumok megjelenítését is, mivel így kijelenthető, hogy a négy vizsgált szezon legjobb elért teljesítményével vetem össze modellem előrejelzéseit. Ezzel nagyobb megbízhatósággal kijelenthető, hogy egy adott pontszám elegendő valamely „nevezetes” helyezés eléréséhez. A táblázatba bekerült még egy olyan oszlop, amely azt a pontszámot hivatott kiszámolni, amivel az adott helyezés közvetlenül megelőzi a mögötte levő helyezés maximumát (képlet: előző helyezés maximum pontszáma +1 pont). A táblázatban kék háttérszínnel jelöltem a Bajnokok-Ligája helyet érő pozíciókat, sárga színnel az Európa-Ligát érő helyezéseket, narancssárgával az osztályozós helyezést, illetve piros színnel a kiesést érő helyeket:

Helyezés	Átlagos pontszám	Maximális pontszám	Előző helyezés maximumát közvetlenül megelőző pontszám
1.	81,5	84	77
2.	68,75	76	67
3.	62,75	66	66
4.	60	65	64
5.	55,5	63	56
6.	52,25	55	55
7.	50,5	54	54
8.	48,75	53	52
9.	47	51	46*
10.	42,75	44	45*

11.	42	43	44
12.	41	43	41*
13.	37,5	39	40*
14.	35,75	38	39*
15.	35,25	37	38
16.	32,25	37	33
17.	28,5	32	26
18.	21,5	25	-

8. táblázat: A pontszámok összefoglalása.
Saját készítés.

*: 2 pontot adtam hozzá az adott helyezésekhez, hogy ne alakuljon ki holtverseny.

Legelőször egy olyan fiktív csapat pontszámát számoltam ki, amely minden egyes, a végső modellben helyett kapott változó átlagával teljesített, az egyes változók átlagát a regressziós egyenletbe behelyettesítve 46,887 pontot kaptam eredményül. Ezzel a pontszámmal a 8. táblázat alapján egy csapat várhatóan a 9. vagy legrosszabb esetben 10. helyen végezne a szezon végén. Amennyiben a tényleges szezonok végeredményét vizsgálom, úgy a négy vizsgált szezon közül három esetén szintén ezen helyek valamelyikén végezne ez a fiktív csapat, amely kerekítve 47 ponttal zárna a szezont. A 2016/2017-es szezon esetén a 47 pontos teljesítmény azonban a 8. helyre is elegendő lett volna. Tehát egy olyan klub, amely az általam kiválasztott, az elért pontszámot szignifikánsan magyarázó mutatószámok közül a ligaátlagon teljesített, az szezon végén a modell alapján várhatóan a 8. és 10. helyek között fog végezni a német Bundesliga-ban. A 47 pontos teljesítmény eléréséhez egy klubnak a szezon során esedékes 34 mérkőzéséből legalább hét, legfeljebb pedig tizenhat mérkőzését kell megnyernie. Ezek a szélsőértékek lennének (hét győzelem kizárólag akkor lenne elegendő, hogyha egy csapat utána minden mérkőzésén döntetlent játszana), azonban a realitás talaján maradva egy középmezőnyben (a tabella középső hatodán végzett csapat, azaz a 9., 10., 11. hely valamelyikét elérő klub) végző Bundesliga-klubnak a valóságban minden esetben legalább tíz mérkőzésen (átlagosan ezen helyezettek xy győzelemmel zártak) győztesen kellett elhagynia a pályát (Whoscored, 2020c) hogy a 9.-11. helyen végezzen a szezon végén. Abban az esetben, hogyha az általam kalkulált 47 pontos teljesítmény elérését vettem figyelembe és nem a helyezést a győzelmek száma azonban már minden esetben legalább tizenhárom mérkőzés volt, amely a teljes szezon mérkőzéseinek több, mint harmadát jelenti.

A továbbiakban a nevezetes helyezésekhez elért pontszámokat fogom vizsgálni, mely helyezésekhez a 8. táblázat alapján egy célpontszámot is kitűztem. A célpontszámok minden nevezetes helyezés eléréséhez a 8. táblázat utolsó oszlopában található pontszámok lettek. Ezen pontszámokkal mind a négy általam vizsgált Bundesliga-szezonban egy csapat elérette valamely nevezetes helyezést. Ezekhez a célpontszámokhoz azonban már a szezon kezdetekor is hozzá lehet rendelni győztes mérkőzéseket. Ezzel a módszerrel a klubok vezetői is tisztábban tudnak definiálni elvárásokat a játékosok, illetve a velük dolgozó edzői stáb felé. Az alábbi táblázat tartalmazza a célpontszámokat, illetve az ezek eléréséhez szükséges győzelmek számát, illetve arányát a szezon során lejátszandó összes (34) mérkőzéshez viszonyítva:

Helyezés (jelentése)	Célpontszám	Győzelmek minimuma	Győzelmek maximuma	Győzelmek ténylegesen	Győzelmi %
1.(bajnoki cím)	77	21	26	26	76,47%
4.(BL indulás)	64	15	22	19	55,88%
6.(EL-indulás)	55	11	19	17	50%
15.(biztos bentmaradás)	38	2	13	11	32,35%
16.(osztályozó bentmaradásért)	33	0	11	9	26,47%

9. táblázat: Célpontszámok, illetve győzelmek összefoglalása.
Saját készítés.

A 9. táblázat célpontszámai tehát minden helyezés esetében azok a pontszámok lettek, melyekkel a vizsgált helyezés a közvetlenül mögötte lévő helyezést minden vizsgált szezonban megelőzte volna, tehát jobban teljesít a mögötte levő helyezés maximumánál. A győzelmek minimuma, illetve maximuma az a legtöbb, illetve legkevesebb győzelmek számát jelenti, melyekkel matematikailag elérhető, esetleg meghaladható az adott pontszám. A tényleges győzelmek száma esetén mindegyik általam vizsgált szezon során azon csapat győzelmeinek számát vettem alapul mind a négy szezonban, amely csapat teljesítménye (helyezéstől függetlenül) elérte, esetleg meghaladta a célpontszámot. Ezen négy győzelmi számból végül számtani átlagot kalkuláltam, majd az eredményt egész győzelmekre kerekítettem felfelé. A felfelé kerekítés oka, hogy amennyiben a kapott számtani átlag 12,1 győzelem lenne, ez azt jelentené, hogy 12 darab nyertes mérkőzés még nem volt elegendő biztosan az adott pontszám eléréséhez, ennél tehát több győzelemre volna szükség az adott pontszám eléréséhez. Végül a

győzelmi % alatt a tényleges győzelmek, illetve az összes forduló hányadosa került kiszámításra, ezen ráta kalkulációja során tehát a döntetleneket egyáltalán nem vettem figyelembe. A 9. táblázat alapján tehát kijelenthető, hogy egy csapat, amely megnyeri mérkőzéseinek a felét a német Bundesligában, az a következő szezonban nagyon jó eséllyel elindulhat legalább a következő évi Európa-Liga selejtezőiben. A biztos bentmaradás eléréséhez pedig elengedhetetlen legalább minden harmadik találkozó megnyerése a szezon során.

7.2. Teljesítmény becslése a szignifikáns mutatószámok segítségével:

A továbbiakban tehát a 9. táblázatban látható célszámokhoz az egyes szignifikáns mutatószámokban megmutatkozott teljesítményt is figyelembe kellett vennem. Nem szabad azt sem elfelejteni, hogy a szignifikáns változók közül három változó pozitív, illetve szintén három változó negatív korrelációt mutatott a pontszám alakulásával, így a változók közül a labdabirtoklás, a SGR, illetve a legjobb játékos értékelése maximalizálandó, míg a sárga, illetve piros lapok száma, valamint az elszenvedett kapura lövések száma minimalizálandó változók a minél jobb teljesítmény elérésének érdekében. A regressziós együtthatókból (ld.: 7. táblázat) kiolvasható, hogy az egyetlen játékos értékelése változó növelése, illetve csökkentése reagál legérzékenyebben a pontszámra, így ezen változó során többször is elő fog fordulni, hogy nem egységnyi (egy egész értékelési pont) értékekkel, hanem századokkal, illetve tizedekkel fogom változtatni, hiszen már ebben az esetben is lesz jelentősnek mondható hatása az eredményváltozó alakulására. A továbbiakban különféle módszerekkel fogom bemutatni, hogy a regressziós modell együtthatóinak változtatásával hogyan érhetőek el a nevezetes helyezések a német Bundesligában az elmúlt négy szezon adatai alapján. Ezen kalkulációk eredményeit az alábbi táblázat foglalja össze:

Helyezés	Átlag	1.(bajnok)	4.(BL)	6.(EL)	15.(bentmaradó)	16.(osztályozó)
Célpontszám	47	77	64	55	38	33
Labdabirtoklás	50%	54%	52%	51%	49%	48%
SGR	11,45%	16,45%	14,45%	13,45%	9,45%	9,45%
Sárga	60,04	56,04	58,04	59,04	61,04	63,04
Piros	2,74	0,74	0,74	1,74	3,74	4,74
Elszenvedett lövések	12,975	7,975	9,975	11,975	13,975	14,975
Egyéni rating	7,24	7,74	7,44	7,34	7,14	7,04

10. táblázat: Teljesítmény becslése a változók egységnyi változtatásai alapján.
Saját készítés.

A 10. táblázatban olyan módszerrel dolgoztam, hogy a változók együttthatóit egymás után változtattam egységnyivel az átlagos szintről, mindaddig, amíg el nem érik valamely nevezetes helyezéshez szükséges pontszámot. Ez a gyakorlatban tehát úgy néz ki, hogy amennyiben a modell mind a hat változóját az átlagos szintről egységnyivel (a *rating* esetén egységnyinek 0,1-et vettem, mely lényegében azt jelenti, hogy együttthatójának értéke a regressziós egyenletben mintha tizedére csökkent volna, tehát 1,255 lett) változtatjuk olyan irányban, hogy a regressziós egyenlet eredménye növekedjen úgy ez a változás az eredményváltozón pontosan 6,74 pontos, tehát csaknem 7 pontos növekedést jelentene. A változtatásokat lépésről lépésre végeztem (így a *ceteris paribus* elv sem sérült egyszerre egyetlen változó változtatásával), mely azt eredményezte, hogy a változókat rangsorolnom kellett. A változókat regressziós együttthatójuk abszolút értéke alapján tettem csökkenő sorrendbe (sorrend: *SGR*, *Elszenvedett kapura lövések*, *Legjobb játékos értékelése*, *Piros lapok*, *Labdabirtoklás*, *Sárga lapok*). Itt meg kell jegyezni, hogy például a *piros lapok* számát az átlagos szintről például legfeljebb csak 2 egységet csökkenthettem, hiszen utána a mutató negatív szám lett volna, amely a futballban nem értelmezhető.

A 10. táblázat eredményeit összefoglalva a lineáris regressziós modellem alapján a német bajnoki cím elhódításához (77 pont eléréséhez) elegendő az 54%-os *labdabirtoklás* (4% javulás az átlagoshoz képest), azonban az *SGR-mutatón* szignifikáns javulásra van szükség, az

átlagos szinthez képest, ezen változót 5%-ponttal növeltem, amely a ligaátlag csaknem 144%-a lett. A bajnoki cím eléréséhez egy csapatnak tehát száz lövésből az átlagosnál ötten többet (összesen 16,45-öt) gólra kell váltania ezen módszer alapján. A *sárga lapok* számát négy, illetve a *piros lapok* számát két (ez a maximális egységnyi csökkentés, amit a piros lapokon eszközölhettem) darabbal (egységgel) csökkentettem az átlagos szintről, ezzel is tovább növelve a pontszámot, tehát a Bundesliga-ban az elmúlt négy szezon alapján a szerzett büntetőlapok számának csökkentésével javítható a teljesítmény is a modell alapján. Továbbá kijelenthető, hogyha egy csapatnak 8 (átlag: 12,975) alá sikerül csökkentenie a mérkőzésenként *elszenvedett lövések* számát, akkor jó eséllyel célozza meg a bajnoki címhez szükséges 77 pontos határt. Ez a csökkentés jelentős mértékű, mintegy közel 40%-kal kellene egy átlagos szinten teljesítő csapatnak az ellenfelek lövéseinek számát lecsökkentenie. Végül a modell alapján arra a következtetésre jutottam, hogy egy csapat, ahol egy játékos kiemelkedően, a szezon összes mérkőzésén átlagosan 7,74-es *értékelést* tud szerezni, az a csapat jó eséllyel pályázik a bajnoki címre. Ez, illetve ennél jobb teljesítmény a vizsgált sokaságban háromszor fordult elő azonban összesen, a 2016/2017-es bajnok (Thiago Alcantara: 8,23), illetve a 2019/2020-as bajnok Bayern München (Robert Lewandowski: 8,13) csapatában, illetve a 2016/2017-es második helyezett Leipzig csapatában (Emil Forsberg: 7,75) (Whoscored, 2020c).

A 10. táblázat alapján a változók egységnyi változtatásának módszerével a Bundesliga-ban biztos bent maradáshoz szükséges 38 pontos (15. hely) határ eléréséhez elegendő az átlagos szintnél 1%-kal alacsonyabb 49%-os *labdabirtoklási* mutató. Ami a lövések hatékonyságát illeti, ez akár 2%-kal is romolhat az átlagoshoz képest, azonban a 38 pont eléréséhez a 9,45%-os szintet érdemes elérnie egy csapatnak. Ennél alacsonyabb *SGR*-rel a vizsgált sokaságban összesen hat csapat tudott bent maradást érő helyen végezni. A *sárga*, illetve *piros lapok* számának elegendő volt egyaránt egységnyivel növelni az átlagos szintjét 61,04, illetve 3,74 (~4) darab szerzett lapra. Az *elszenvedett kapura lövések* esetén is elegendő volt egységnyivel csökkenteni az átlagos szintet, mely mérkőzésenként 13,975 lövést jelent a szezon során, tehát átlagosan. A biztos bent maradás eléréséhez 14 lövés alatt célszerű tartania egy csapatnak az ellenfeleit átlagosan. A csapatok *legjobb játékosának értékelését* is elegendő volt egységnyivel, azaz 0,1 ponttal csökkenteni 7,14 pontosra. A vizsgált sokaságban egyetlen egyszer fordult elő, hogy az utolsó három helyen végzett valamely csapat legjobb játékosa ennél jobb *átlagos mérkőzésenkénti értékeléssel* zárjon egy szezont. Ez a 2016/2017-es szezon 17. helyezett Ingolstadt csapatának játékosa, az osztrák Markus Suttner volt, aki 7,28-os *átlagos értékeléssel*

zárta a szezont. Ezzel Suttner az értékelések alapján a 20. helyen végzett a Bundesliga ezen szezonjában, amely figyelemre méltó annak ellenére, hogy csapata kiesett a szezon végén.

7.3. Elaszticitási együtthatók vizsgálata a szignifikáns változók esetén:

A továbbiakban megvizsgáltam, hogy az elért célpontszámokhoz, illetve a 10. táblázatban becsült mutatószámok további 1%-os változtatása hány százalékos változást eredményez a pontszám alakulásán (Ferenci, 2018, Ramanathan, 1993). Ezen statisztikai mutatószámot nevezzük rugalmassági, vagy elaszticitási együtthatónak. A rugalmassági együttható sajátossága, hogy kizárólag egy rögzített helyen (a változók értékét így a 10. táblázatban becsült értékeken rögzítettem) tudja megvizsgálni a magyarázó változók elaszticitását. (Képlet: $El(\hat{y}, x) = \frac{\beta_j \cdot x}{\hat{y}}$ (Hunyadi, Vita, 2008))

Az egyes változók szintjeihez tartozó rugalmassági együtthatókat az alábbi táblázat tartalmazza:

Helyezés	Átlag	1.(bajnok)	4.(BL)	6.(EL)	15.(bentmaradó)	16.(osztályozó)
Labdabirtoklás	0,5%	0,33%	0,38%	0,44%	0,61%	0,68%
SGR	0,5%	0,44%	0,46%	0,5%	0,51%	0,59%
Sárga	-0,2%	-0,11%	-0,14%	-0,17%	-0,25%	-0,3%
Piros	-0,045%	-0,008%	-0,01%	-0,03%	-0,08%	-0,12%
Elszenvedett lövések	-0,54%	-0,2%	-0,3%	-0,43%	-0,72%	-0,89%
Egyéni rating	1,93%	1,26%	1,46%	1,67%	2,35%	2,67%

11. táblázat: Az elaszticitási együtthatók az egyes célpontszámok szintjén.
Saját készítés.

A 11. táblázatban félkövér betűtípussal megjelölt rugalmassági együtthatókat fogom értelmezni a továbbiakban, ezzel is szemléltetve ezen mutatószám működését, illetve hatását a regressziós modellem eredményváltozójára. A klubok *legjobb játékosának értékelése* az átlagos szintről (7,24 pont) amennyiben 1%-kal növekedik, abban az esetben c.p. **1,93%**-kal fog a szerzett pontszám növekedni. Ez alapján, tehát amennyiben a Bundesliga csapatok valamely

játékosa képes kiemelkedni a mérkőzéseken és hétről hétre magas értékelést szerezni, abban az esetben csapata is képes lehet jobb teljesítményre. Egy klub elszenvedett lövéseinek száma amennyiben a bajnoki címhez szükséges szintről (7,975 lövés/mérkőzés) 1%-kal növekszik, abban az esetben c.p. **0,2%**-kal fog a várható pontszám csökkenni. Ez azt jelenti, hogy jobban teljesítő (bajnokesélyes) csapat esetén az elszenvedett lövések számának kis mértékű növekedése nem befolyásolja jelentős mértékben az adott csapat szezonvégi pontszámát. A labdabirtoklás amennyiben a Bajnokok Ligája indulást érő szintről (52%) 1%-kal növekedik, abban az esetben c.p. **0,38%**-kal fog növekedni a várható pontszám. Továbbá a 11. táblázat alapján az SGR mutató 1%-os növekedése a Bundesliga-ban való bentmaradást érő szintről c.p. **0,51%**-os növekedést eredményez a pontszámon.

7.4. A végső lineáris regressziós modell alkalmazása a 2020/2021-es szezon előrejelzésére:

Végső lineáris regressziós modellem szignifikáns változóit kigyűjtöttem a 2020/2021-es Bundesliga-szezonból egyaránt. Ez a szezon nem ért egyelőre véget, azonban 26 forduló már lejátszásra került az adatgyűjtés végzésekor, így a szezon nagyjából a háromnegyedénél tart. Kutatásom ezen fejezetében készítettem egy előrejelzést a négy korábbi szezont vizsgáló regressziós modellem segítségével a 2020/2021-es szezon végeredményére vonatkozóan. Az előrejelzés a korábbi kalkulációkkal ellentétben feltünteti az egyes Bundesliga-indulók neveit is, mivel ezáltal hatékonyabban összehasonlítható, a modell szerinti, illetve a tényleges bajnoki sorrendje a kluboknak. Ezen előrejelzés eredményeit az alábbi táblázatban foglaltam össze. A táblázat a modell előrejelzését hasonlítja össze egy olyan fiktív scenárióval, mely során a csapatok pontosan ugyanúgy teljesítenek az utolsó 8 fordulóban, mint azt az első 26 forduló során tették (ennek következtében a csapatok végső sorrendje teljesen ugyanaz marad, mint a szezon közbeni, 26 forduló utáni státuszban volt).

Hely.	Klub	Fiktív pontszám	Pontszám modellel	Pontszám változása	Új hely.	Hely. változása
1.	Bayern München	80	82	+2	1.	-
2.	RB Leipzig	75	65	-10	3.	↓ 1
3.	Wolfsburg	67	56	-11	6.	↓ 3
4.	Frankfurt	61	60	-1	4.	-
5.	Dortmund	56	70	+14	2.	↑ 3
6.	Bayer Leverkusen	52	60	+8	5.	↑ 1
7.	Union Berlin	50	55	+5	7.	-

8.	Freiburg	48	44	-4	12.	↓	4
9.	Stuttgart	47	46	-1	10.	↓	1
10.	Mönchengladbach	47	49	+2	8.	↑	2
11.	Hoffenheim	39	46	+7	11.		-
12.	Werder Bremen	39	40	+1	13.	↓	1
13.	Augsburg	38	37	-1	15.	↓	2
14.	Hertha Berlin	31	48	+17	9.	↑	5
15.	Mainz	31	35	+4	16.	↓	1
16.	Köln	30	38	+8	14.	↑	2
17.	Bielefeld	29	33	+4	17.		-
18.	Schalke	13	19	+6	18.		-

12. táblázat: Szezonvégi teljesítmény előrejelzésének eredményei.
Saját készítés.

A 12. táblázat fiktív pontszám oszlopa egy olyan becslés eredménye, mely során minden klub már megszerzett pontjait szoroztam meg a szezonbeli összes mérkőzés (34), illetve a már lejátszott mérkőzések (26) hányadosával. Ezzel elérve, hogy az ezen becsléssel készített tabella helyezései teljesen ekvivalensek a 26 forduló utáni tabellával, azonban mégis egy teljes szezonról kaptam képet ezáltal. Ezen becslés, akárcsak a regressziós modellem becslése tartalmazhat pontatlanságokat, hibákat. Ugyanis nem veszi azt figyelembe, hogy a bajnokság utolsó fordulóiban az osztályban maradásért küzdő csapatok gyakran nyújtanak jobb teljesítményt, szereznek több pontot. Ennek az eredménye az a tény is, hogy ezen becslésben a hátul található csapatok teljesítménye (főleg a 14. helytől) rendkívül alacsony, vagyis könnyedén felülbecsülhető lesz. A középmezőnyben található csapatok számára az utolsó néhány forduló mérkőzései sokszor már valódi téttel nem bírnak, így ezen csapatok teljesítménye hullámzó lehet a bajnokság utolsó fordulóira. A táblázat eredményeinek egy további korlátja, hogy nem veszi figyelembe a mérkőzésenként szerezhető pontok számát, kizárólag egy 34 mérkőzés utáni teljesítményt becsül a modell.

A 12. táblázatból továbbá látható, hogy a modellel alkotott becslés ugyan eltér, azonban a klubok helyezését a 18 csapatból 11 esetén csupán legfeljebb egyetlen helyezéssel becsülte felül, valamint alul. Fontos kiemelni, hogy a magyarázó változók értékei a 2020/2021-es szezonban a sárga lapok, illetve a piros lapok esetén becsült értékek, hiszen az eredeti modellemben ezek mennyisége minden esetben egy teljes szezon alatt szerzett büntetőlapok számát tartalmazta. Ezt a becslést a fiktív pontszám kalkulációjához hasonlóan a szezon alatt

lejátszott összes mérkőzés (34), illetve az adatgyűjtéskor lejátszott mérkőzések (26) hányadosával készített arányszámmal végeztem. A regressziós modell a fiktív pontszámot átlagosan csapatonként 2,78 ponttal becsülte felül csapatonként. A szórása a pontszámok különbségének a két becsült pontszám között azonban viszonylag magas, 6,93 pont volt.

Az általam korábban alkotott célpontszámok közül a 77 pont ezen előrejelzés szerint is bajnoki címet ért volna, a 64 szerzett pont BL-indulást. Az 55 pontos határ az előrejelzés esetén a 7. hely elérésére elég, azonban csupán egy ponttal lenne lemaradva az EL-helyet érő 6. helyezéstől. A biztos Bundesligában maradás szintjét 38 pontnál határoztam meg korábban, ez a szint a modellem becslése alapján is elegendő a ligában maradáshoz (15. helyhez) a következő szezonra. Az osztályozós helyezés szintjét korábban 33 pontban határoztam meg, azonban a 2020/2021-es szezon előrejelzése alapján 34 pontra lesz legalább szüksége annak a csapatnak, aki el akarja kerülni a kiesést majd, amely mérkőzésenként pontosan 1 pontos átlagot jelent a teljes szezon (34 forduló) során.

7.5. Regressziós modellem becslésének összevetése a tényleges 2020/2021-es teljesítménnyel:

Dolgozatomban ezen fejezetében azt fogom megvizsgálni, hogy a Bundesliga 2020/2021-es kiírásában az eltelt 26 forduló után regressziós modellem becslése hogyan alakul a bajnokság tényleges állásához képest. Fontos kihangsúlyozni, hogy modellem alapvetően 34 mérkőzéses mintára készült, így a kalkulált pontszám ismételtelen be lesz majd szorozva a lejátszott mérkőzések (26), valamint a lejátszandó mérkőzések (34) hányadosával. Az alábbi táblázatban foglaltam össze a tényleges, illetve a becsült teljesítmény összevetésének eredményeit:

Hely.	Klub	Tényleges pontszám	Pontszám modellel	Eltérés	Hely. a modellel	Hely. változása
1.	Bayern München	61	64	+3	1.	-
2.	RB Leipzig	57	51	-6	3.	↓ 1
3.	Wolfsburg	51	45	-6	6.	↓ 3
4.	Frankfurt	47	49	+2	4.	-
5.	Dortmund	43	55	+12	2.	↑ 3
6.	Bayer Leverkusen	40	48	+8	5.	↑ 1
7.	Union Berlin	38	44	+6	7.	-
8.	Freiburg	37	35	-2	12.	↓ 4
9.	Stuttgart	36	38	+2	10.	↓ 1
10.	Mönchengladbach	36	40	+4	8.	↑ 2

11.	Hoffenheim	30	38	+8	11.	-
12.	Werder Bremen	30	33	+3	13.	↓ 1
13.	Augsburg	29	31	+2	15.	↓ 2
14.	Hertha Berlin	24	39	+15	9.	↑ 5
15.	Mainz	24	29	+5	16.	↓ 1
16.	Köln	23	32	+9	14.	↑ 2
17.	Bielefeld	22	28	+6	17.	-
18.	Schalke	10	17	+7	18.	-

13. táblázat: Valós eredmények összevetése modellem becsléseivel a 2020/2021-es Bundesliga-szezon első 26 fordulójára alapján.
Saját készítés.

Abban az esetben, hogyha a kiindulási adatbázis részét képezte volna ez a szezon is, elképzelhető, hogy az eredeti modellből más változók magyarázták volna szignifikánsan a továbbiakban a teljesítményt, így a regressziós becslés kevésbé pontosan illeszkedik a ténylegesen elért pontszámokra a bajnokság első 26 fordulójára után. A 13. táblázat alapján látható továbbá az is, hogy az általam készített becslés többnyire felülről közelíti a tényleges eredményeket. Átlagosan csapatonként 4,33 ponttal becsülte a modell felül a tényleges eredményeket, a szórása az eltéréseknek 5,31 pontos volt, amely az előző becslés eltéréseinek szórásánál alacsonyabb lett. A felülbecslés további magyarázata, hogy regressziós modellem nem veszi figyelembe azt a tényt, hogy a mérkőzéseken minden esetben három, illetve egy pontot szerezhetnek a csapatok, ugyanis a modell alapvetően szezon végi pontszámok becslésére lett megalkotva. A 13. táblázat becslései pontszámokra tehát kevésbé megbízhatóak (hiszen nem 34 fordulót becsülnek), azonban a helyezések alakulását érdemes a továbbiakban is vizsgálni.

Regressziós modellem becslése alapján a Borussia Dortmund csapata a teljesített statisztikákhoz képest egyértelműen alulteljesített ezen szezonban, hiszen három helyezéssel teljesítették jelen pillanatban 26 forduló után a modell várakozásait alul. Ez a felülbecslés négy vizsgált mutatószámra is köszönhető. A Dortmund által a 2020/2021-es szezonban teljesített SGR a harmadik legmagasabb a ligában, egyébként a listavezető Bayern München után a második legtöbb szerzett góllal is rendelkeztek az adatgyűjtés készítésekor. A *sárga* (36 db), illetve *piros* (0 db) lapjaik számát alacsonyan tudták tartani, mely annak köszönhető, hogy egy alapvetően támadó felfogásban játszó csapat a Dortmund. A kiemelkedő játékos teljesítmény szintén jelen van a Dortmund esetén ebben a szezonban, ugyanis a csapat kezdőcsatára, Erling

Braut Haaland az első 26 fordulót 7,61-es átlagos értékeléssel zárta, amellyel a második legjobb volt a teljes ligában.

Egy másik csapat is található a mezőnyben, amelynek a teljesítményét modellem kiugróan felül becsülte, a Hertha regressziós modellem szerint öt hellyel kellene, hogy előrébb legyen. Ez a klub a jelenleg Dárdai Pál irányítása alatt levő Hertha Berlin csapata volt. A nagy pontatlanság oka elsősorban a csapat egy remek formában lévő játékosának köszönhető, ugyanis a csapat brazil támadója, Matheus Cunha 7,37-os *értékelést* szerzett a szezon során átlagosan mérkőzésenként, amellyel ezen szezonban az összes játékos közül 8. helyen állt. Továbbá igaz a Hertha-ra, hogy nem engedik az ellenfeleket különösen sokszor kapura löni, *elszenvedett kapura lövéseik* száma meccsenként körülbelül a ligaátlaggal egyezik meg (Hertha: 12,4; ligaátlag: 12,23). A modellem becslése alapján ezen mutatószámokkal a Hertha a 9. helyen állna. A modell alapján tehát kijelenthető, hogy a Hertha egy olyan szezonban is egyelőre stabilan bent maradó helyen tud teljesíteni, mely során a csapatot sorozatosan érik a balszerencsés mérkőzések és a statisztikák alapján vártnál alul teljesít.

A továbbiakban megpróbáltam egy indokot találni a Hertha, illetve a Dortmund alulteljesítésére a 2020/2021-es szezonban. Az *elszenvedett kapura lövések* változó, illetve egy csapat kapott góljainak segítségével kiszámítottam, hogy hány százalékban születik gól, az elszenvedett kapura lövéseik közül. Ezt a mutatószámot CSGR (Conceded Shot Goal Ratio) - nek fogom rövidíteni a továbbiakban. Ezen eredmények alapján sorba rendeztem a Bundesliga csapatait az alábbi táblázatban:

CSGR Hely.	Klub	CSGR
1.	Wolfsburg	8,13%
2.	Freiburg	10,07%
3.	Augsburg	10,68%
4.	RB Leipzig	10,77%
5.	Werder Bremen	10,99%
6.	Frankfurt	11,54%
7.	Stuttgart	11,95%
8.	Union Berlin	12,07%
9.	Mönchengladbach	12,11%
10.	Leverkusen	12,17%
11.	Bielefeld	12,36%

12.	Bayern München	13,2%
13.	Mainz	13,52%
14.	Dortmund	13,55%
15.	Köln	13,82%
16.	Hertha	13,96%
17.	Hoffenheim	15,05%
18.	Schalke	15,08%

14. táblázat: CSGR mutatók a 2020/2021-es szezon alapján.
Saját készítés.

A 14. táblázatban félkövérrel jelöltem azon csapatokat, amelyekről említést teszek ezen bekezdésben. A táblázatból jól látható, hogy a Hertha, illetve a Dortmund csapata is magas százalékban kap gólt az ellenfelek lövéseiből a Bundesliga többi csapatához képest. Tehát az ellenfelek ezen csapatok ellen jobban használják ki helyzeteiket. Ennek magyarázata lehet az is, hogy a 14. táblázatban a hátsó régiókban szereplő csapatok ellen az ellenfelek minőségibb, illetve nagyobb gólhelyzetekbe tudnak kerülni a mérkőzések során, azaz a csapatok védelme kevésbé működhet hatékonyan. Figyelemre méltó eredmény a Wolfsburg csupán 8,13%-os CSGR mutatója, mellyel toronymagasan listavezető, és ez a mutatószám azt mutatja, hogy az ellenfelek többször hibáznak ellenük. Ez betudható annak, hogy az ellenfelek lövései alacsony minőségűek, illetve akár kiemelkedő kapusteljesítmény is állhat a háttérben. További érdekesség, hogy a CSGR szerinti első öt csapat, sorrendben a Wolfsburg, a Freiburg, az Augsburg, a Leipzig, valamint a Werder Bremen helyezését alulbecsülte a modellem. Ennek magyarázata abban rejlik, hogy ezek a csapatok alapvetően hatékonyan védekeznek, vagy kevesebb lövést engednek az ellenfeleknek, vagy kevésbé veszélyes helyzeteket tudnak kialakítani ellenük.

Érdekesség, hogy a listavezető Bayern München csupán a 13. helyet foglalja el a listában. A Bayern azonban a támadójátékot jellemző mutatószámokban, illetve szerzett gólokban rendszerint toronymagasan elsők a bajnokságban, tehát a védelmi megingásaikat sikerül eredményes támadójátékkal kompenzálniuk. A Bayern esetén tehát mondhatni, hogy a klub teljesítményébe beleférnek védelmi hibák, hiszen jelen szezonban mérkőzésenként 3 gólt is tudtak szerezni átlagosan. Az adatbázisom adatai alapján egyébként ezt a trendet a Bayern már néhány éve fent tudja tartani, ugyanis a négy szezon alatt 369 gólt szereztek, mely 2,71 gólt jelent minden mérkőzésen. Hogyha az idei, ennél jobb statisztikájukat is hozzáadom ehhez, akkor abban az esetben 2,76 gólt kapnánk már mérkőzésenként. Nem meglepő tehát, hogy a

Bayern minden vizsgált, illetve nagy valószínűséggel az előrejelzett szezonban is bajnoki címet ünnepelhetett (fog ünnepelni).

A Schalke 2020/2021-es csapata a Bundesliga történetének egyik leggyengébb szereplését mutatja be. A legtöbb szignifikáns mutatószámában, illetve a CSGR-ben is teljesen utolsók. A 13. táblázat becslése alapján is legalább 11 ponttal le kellene, hogy legyenek maradva az utolsó előtti helyezéstől, míg a valóságban 26 forduló elteltével 12 pontos hátrányt kellene ledolgozniuk, hogy előzni tudjanak, illetve az osztályozós (16.) hely eléréséhez 13 pontot. Alighanem Németország egyik legnagyobb múltú futballklubja jövőre a Bundesliga2-ben fogja folytatni a küzdelmeket.

A Dortmund esetén azonban egy másik tényező vizsgálatát is elvégeztem. A Dortmund rendelkezik ugyanis az egyik leghíresebb és leghatalmasabb szurkolótáborral, mely a 2020/2021-es szezonban egyelőre végig csupán otthonról a televízió elől tudta támogatni kedvenceit. A Dortmund átlagos nézőszáma a nem zártkapus mérkőzések figyelmen kívül hagyásával mintegy 80298 fő volt, amely a legmagasabb a Bundesliga teljes mezőnyében. A csapat stadionjának, a Signal Iduna Parknak kihasználtsága pedig mintegy 98,7%-os volt átlagosan ezen szezonok során. Az alábbiakban a zártkapus, illetve az azt megelőző négy szezon hazai, illetve idegenbeli győzelmi rátáját láthatjuk a Borussia Dortmund csapatának:

Szezon	Hazai győzelmi ráta	Idegenbeli győzelmi ráta	Különbség (H-I)
2016/2017	76,47%	29,41%	47,06%
2017/2018	52,94%	35,29%	17,65%
2018/2019	82,35%	52,94%	29,41%
2019/2020	64,71%	58,82%	5,89%
2020/2021	58,33%	42,86%	15,47%
Átlag	66,96%	43,86%	23,1%

15. táblázat: A Dortmund győzelmi aránya az elmúlt öt év Bundesliga szezonjaiban.
Saját készítés.

A 15. táblázatból látszik, hogy az elmúlt öt szezon második legrosszabb hazai teljesítménye győzelmek tekintetében a 2020/2021-es Borussia Dortmundé. Idegenbeli teljesítmény szempontjából az idei szezon a középmezőnyben található, a harmadik helyen. Ugyan szignifikáns visszaesés nem mutatkozik meg az idei hazai teljesítményben egyértelműen, azonban az idegenbeli, illetve a hazai mérkőzések győzelmi rátái között a

legkisebb a különbség azon szezonokban, ahol már rendeztek zártkapus mérkőzéseket is. A szurkolók hiánya tehát a hazai pálya győzelmi arányát az átlagostól 8,63%-kal rontotta le, mely a 17 hazai mérkőzésből 1,47-dal jelent kevesebb győztes mérkőzést, mint az átlagos teljesítménye volt a Dortmundnak az utóbbi években. Ez nem mondható azonban szignifikáns romlásnak, hiszen a 2017/2018-as szezonban saját szurkolóik előtt is ennél jóval rosszabbul tudott a csapat teljesíteni.

8. A kutatás korlátatai, további kutatási irányok bemutatása:

8.1. A kutatás korlátai:

Kutatásom eredményei után a korlátokról is érdemes említést tenni. Az egyik legfontosabb korlát, amely munkám során felmerült, az a manuális adatgyűjtés időigénye volt. Eleinte a kutatást nagyobb mintával terveztem, azonban végül a változók mennyiségére, illetve minőségére, valamint megalapozottságára helyeztem a hangsúlyt. Ennek is köszönhető, hogy a kutatás során összesen huszonöt mutatószám került a végső adatbázisomba. További korlát lehet, hogy a lineáris regresszió nem alkalmas tökéletesen előrejelzésre, azonban a végső modell erős magyarázóereje, az előrejelzések viszonylagos pontossága is azt erősítette meg, hogy a kalkulációim összességében eredményesek lettek. Fontosnak tartom még kiemelni, hogy ezen modell kizárólag a Bundesliga-ra alkalmazható, másik labdarúgóbajnokságot vizsgálva elképzelhető, hogy más változók lettek volna szignifikánsak. Például a sárga lapok változó egy másik bajnokságban elképzelhető, hogy kevésbé fontos, ugyanis az ottani játékvezetők felfogása más módon a szabálytalanságokat illetően, illetve a védekezés mentalitása is eltérő lehet a világ, valamint Európa egyéb ligáiban egyaránt.

További korlátja volt a modellnek, hogy kizárólag teljes szezonok során szerzett pontszám becslésére alkalmas, ugyanis nem tartalmazta a semmilyen módon azt a feltételt, hogy egy csapat kizárólag 3, 1, illetve 0 pontot tud szerezni valamennyi mérkőzésén, így akár olyan eredmények is kialakulhatnak, ahol a csapatok lényegében mintha nem egész pontszámot szereztek volna. Ez kiküszöbölhető lehet a győzelmek, illetve döntetlenek számának a további vizsgálatával. Kiküszöbölhető lehet egy olyan regresszió vizsgálata esetén is, ahol pontszám helyett a csapatok győzelmi rátája, (esetleg más olyan mutatószám, amely a teljesítményt szintén jellemezheti) az eredményváltozó.

8.2. További kutatási irányok:

Érdemes lehet a kutatást a továbbiakban akár több szezonon keresztül is vizsgálni, amellyel az eredmények pontosabbá válhatnak, illetve a minta nagysága is megemelkedhetne. Fontos itt kihangsúlyozni, hogy amennyiben több szezonon keresztül vizsgáljuk egy labdarúgóbajnokság mutatószámait, abban az esetben változhat az eredményes játékfelfogás,

illetve taktikai változások is bekövetkezhetnek amennyiben tíz, vagy esetleg több éven keresztül vizsgálunk egy bajnokságot. Érdekes lehet a kutatás abban az esetben is, hogyha egyéb bajnokságokra végeznénk el a kutatást, vélhetően másik bajnokságokban egyéb mutatószámok mutatkoztak volna szignifikánsnak. Egy modellben több bajnokságot vizsgálni azonban ismételten torzított eredményhez vezethet, hiszen az egyes bajnokságok játékfelfogása, védekezési, illetve támadási stílusai, felállásai, valamint akár adott esetben a játékvezetők mentalitása is eltérő lehet.

További érdekességeket hordozhat, hogyha a kalkulációkat a logisztikus regresszió eszköztárával végeznénk, hiszen exponenciális, esetleg logisztikus kapcsolat is fennállhat bizonyos mutatószámok, illetve a jó teljesítmény között. A modellek bővíthetők további változókkal is, például a szignifikáns változók magasabb hatványainak vizsgálatával, esetleg változók interakcióival (szorzatváltozókkal) egyaránt.

Forrás- és irodalomjegyzék:

Amin J. B., Dr. Thakore, D. G., Dr. Patel, N. M. (2018): Soccer Player Tracking System

András, K. (2003): Üzleti elemek a sportban, a labdarúgás példáján

András, K., Havran, Zs., Jandó, Z. (2011): Üzleti globalizáció és a hivatásos sport: sportvállalatok nemzetközi szerepvállalása

Arastey, G. M. (2018): OPTA sports The leading sports data provider

elérési link: <https://www.sportperformanceanalysis.com/article/opta-leading-sport-data-provider> letöltési dátum: 2021.04.05.

Balaton, K., Hortoványi, L., Incze, E., Laczkó, M., Szabó Zs., R., Tari, E. (2007): Stratégia és üzleti tervezés

Bate, A. (2017): Expected assists: What does it mean and who tops the list?

elérési link: <https://www.skysports.com/football/news/11661/11119595/expected-assists-what-does-it-mean-and-who-tops-the-list> letöltési dátum: 2021.01.07.

Bate, A., Campbell, R. (2017): Expected goals explained: The analysis that is changing the game

elérési link: <https://www.skysports.com/football/news/11661/10907419/expected-goals-explained-the-analysis-that-is-changing-the-game> letöltési dátum: 2020.08.20.

Baysal, S., Duygulu, P. (2015): Sentioscope: A Soccer Player Tracking System using Modern Field Particles

Beasley, A. (2017): How to calculate expected goals in soccer

elérési link: <https://www.pinnacle.com/en/betting-articles/Soccer/how-to-calculate-expected-goals-for-soccer-matches/JESJH3RKXCMUF9TY> letöltési dátum: 2020.12.12.

Bigsoccer (2020):

elérési link: <https://www.bigsoccer.com/threads/what-precisely-is-a-shot-on-goal.2028210/> letöltési dátum: 2020.08.17.

Bíró, D., László, D. (2021): Unibet Gólterápia videós műsor 80. adás

elérési link: <https://www.youtube.com/watch?v=cCkY5bwYNgw> letöltési dátum: 2021. 02.09

Bundesliga (2018):

elérési link: <https://www.bundesliga.com/en/news/Bundesliga/german-soccer-rules-50-1-fifty-plus-one-explained-466583.jsp> letöltési dátum: 2020.09.21.

Bundesliga (2019):

elérési link: <https://www.bundesliga.com/en/bundesliga/news/expected-goals-xg-model-what-is-it-and-why-is-it-useful-sportec-solutions-3177> letöltési dátum: 2020.12.12.

Bundesliga (2020a):

elérési link: <https://www.bundesliga.com/en/bundesliga/table> letöltési dátum: 2021.04.05.

Bundesliga (2020b):

elérési link: <https://www.bundesliga.com/en/bundesliga> letöltési dátum: 2020.05.24.

Bundesliga (2020c):

elérési link: <https://www.bundesliga.com/en/bundesliga/news/average-attendance-bayern-dortmund-allianz-arena-signal-iduna-park-6379> letöltési dátum: 2020.05.24.

Bundesliga (2020d):

elérési link: <https://www.bundesliga.com/en/bundesliga/news/expected-goals-xg-and-goal-probability-explained-13847> letöltési dátum: 2020.12.12.

Catapultsports (2020):

elérési link: <https://www.catapultsports.com/sports/soccer> letöltési dátum: 2020.10.03.

Channels TV (2017): Soccer Teams Adopt Catapult Sports System To Achieve Improved Performance

elérési link: https://www.youtube.com/watch?v=qwV_KWjf6TE&t=150s letöltési dátum: 2021.02.10.

Charles, V., Gherman, T. (2013): Achieving Competitive Advantage Through Big Data. Strategic Implications.

Chikán, A. (2017): Vállalatgazdaságtan

Choueiry, G. (2019): Understand Forward and Backward Stepwise Regression

elérési link: <https://quantifyinghealth.com/stepwise-selection/> letöltési dátum: 2021.02.22.

Cityfootball-performance (2020):

elérési link: <https://www.cityfootball-performance.com/location/> letöltési dátum: 2020.08.20.

Claro, D. (2019): A 2018-as labdarúgó világbajnokság kvantitatív elemzése

Cricfooty (2017):

elérési link: <http://www.cricfooty.com/2017/06/14/manchester-city-trophies/> letöltési dátum: 2021.02.18.

Cronin, B. (2017): An analysis of different expected goals models

elérési link: <https://www.pinnacle.com/en/betting-articles/Soccer/expected-goals-model-analysis/MEP2N9VMG5CTW99D> letöltési dátum: 2020.12.12.

Dendir, S. (2016): When do soccer players peak? A note

Dimiean, E. (2019): Valerij Lobanovskiy the pioneer of science in football

elérési link: <https://www.footballcoin.io/blog/valeriy-lobanovskiy-the-pioneer-of-science-in-football/> letöltési dátum: 2020.05.03.

ESPN (2017)

elérési link: <https://www.espn.com/soccer/english-premier-league/story/3224746/premier-league-third-most-reliant-on-foreign-players-in-europe-study> letöltési dátum: 2020.05.16.

Fenton, T. (2019): Football's Pioneer – The Charles Reep story

elérési link: <https://thevideoanalyst.com/footballs-pioneer-charles-reep/> letöltési dátum: 2020.05.02.

Ferenci, T. (2018a): Modellszelekció és a multikollinearitás, Ökonometria előadás, 2018/2019 tavasz

Ferenci, T. (2018b): A standard modellfeltevések, modelldiagnosztika, Ökonometria előadás, 2018/2019 tavasz

Ferenci, T. (2018c): Út az ökonometriához, Ökonometria előadás, 2018/2019 tavasz

Fifa (2021):

elérési link: <https://www.fifa.com/worldcup/archive/uruguay1930/matches/#groupphase>

letöltési dátum: 2021.02.18.

Financefootball (2019):

elérési link: <https://financefootball.com/2019/09/24/sponsorship-deals-in-premier-league-2019-2020/> letöltési dátum: 2021.03.27.

Footballperformanceanalysis (2020)

elérési link: <https://footballperformanceanalysis.com/tag/prozone/> letöltési dátum: 2020.06.12.

Footballxg (2021):

elérési link: <https://footballxg.com/> letöltési dátum: 2021.04.05.

Forbes (2020):

elérési link: <https://www.forbes.com/sites/manuelteth/2020/06/22/bundesliga-495-billion-television-deal-provides-stability-and-perhaps-a-brighter-future/?sh=15a7e1a7141e> letöltési dátum: 2021.03.27.

Giulianotti, R., Robertson, R. (2012): Mapping the global football field: a sociological model of transnational forces within the world game

Goes, F. R., Kempe, M., Meerhoff, L. A., Lemmink, K. A. P. M. (2018): Not Every Pass Can Be an Assist: A Data-Driven Model to Measure Pass Effectiveness in Professional Soccer Matches

Gollan, S., Ferrar, K., Norton, K. (2018): Characterising game styles in the English Premier League using the “moments of play” framework

González-Ródenas, J., López-Bondía, I., Aranda-Malavés, R., Desantes, T. A., Sanz-Ramírez, E., Malaves, R. A. (2019): Technical, Tactical and Spatial Indicators Related to Goal Scoring in European Elite Soccer

Hanke, M., Kirchler, M. (2013): Football championships and jersey sponsors’ stock prices: an empirical investigation

Hatum, A. (2010): Next Generation Talent Management: Talent Management to Survive Turmoil

Horton, M., Gudmundsson, J., Chawla, S., Estephan, J. (2015): Automated Classification of Passing in Football

Hunyadi, L., Vita, L. (2008): Statisztika II.

Intermiamicf (2021):

elérési link: <https://www.intermiamicf.com/en/club/ownership> letöltési dátum: 2021.03.26

Jones, P. D., James, N., Mellalieu, S. D. (2004): Possession as a performance indicator in soccer.

Kajos, A. (2019): Sportolók Fogyasztói Márkaértéke – konceptuális modellje, dimenzióinak mérése és hatása a szurkolók fogyasztói magatartásra

Kálnoki Kis, A., Kele, J., Kocsis, B.: Ziccer, 24.hu

elérési

link:

https://open.spotify.com/episode/0g3JkKE2jlLcu6nWMyntiX?si=lcF_2JLKR0yJSV4UNiIPPA&fbclid=IwAR0F-YngBgkXjP4ISagg0A-I7tczGVzJS39GSqgowVP8NI-gxzhQAA9FSiM&nd=1 letöltési dátum: 2020. 04.12.

Kuper S., Szymanski, S. (2014): Fociológia

Lago-Peñas, C., Lago-Ballesteros, Dellal, A., J., Gómez, M. (2010): Game-Related Statistics that Discriminated Winning, Drawing and Losing Teams from the Spanish Soccer League

elérési link: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3761743/> letöltési dátum: 2021. 02. 12.

Larsson, O. (2010): Charles Reep: A Major Influence on British and Norwegian Football

Legalsportsbetting (2020):

elérési link: <https://www.legalsportsbetting.com/revenue-tracker/> betöltés dátuma: 2020.05.02.

Littkemann, J., Kleist, S. (2002): Sportlicher Erfolg in der Fußball-Bundesliga: eine Frage der Auf- oder der Einstellung?

Mao, L., Peng, Z., Liu, H., Gómez, M.-A. (2016): Identifying keys to win in the Chinese professional soccer league

Maqureira, J., M., Bruque, S., Uhrin, Á. (2019): Talent management: two pathways to glory? Lessons from the sports arena

Matesanz, D., Holzmayer, F., Torgler, B., Schmidt, S. L., Ortega, G. J. (2018): Transfer market activities and sportive performance in European first football leagues: A dynamic network approach

McGuinness, R. (2012): Spectators who are on the ball: A look inside the world of sports stats

elérési link: <https://metro.co.uk/2012/01/03/spectators-who-are-on-the-ball-a-look-inside-the-world-of-sports-statistics-272423/> letöltési dátum: 2020.06.06.

Medeiros, J. (2017): How data analytics killed the Premier League's long ball game

elérési link: <https://www.wired.co.uk/article/premier-league-stats-football-analytics-prozone-gegenpressing-tiki-taka> letöltési dátum: 2020.06.13.

Optasports (2021):

elérési link: <https://www.optasports.com/about/> letöltési dátum: 2021.04.05.

Optasportspro (2020):

elérési link: <https://www.optasportspro.com/> letöltési dátum: 2020.06.12.

Pappalardo, L., Cintia, P. (2017): Quantifying the relation between performance and success in soccer

Pethő, A. (2014a): Bundesliga: Az 50+1%-os szabály

elérési link: <http://www.teamgeist.hu/nemet-foci/bundesliga-az-50-1-es-szabaly/> letöltési dátum: 2020.09.21.

Pethő, A. (2014b): Bundesliga: Az 50+1%-os szabály:

elérési link: <https://sportjog.reblog.hu/bundesliga-az-50--1-es-szabaly> letöltési dátum: 2020.09.21.

Petty, L. (2018): What is expected goals? Expected goals explained

elérési link: <https://www.pinnacle.com/en/betting-articles/Soccer/expected-goals-explained/B8Q2HGGJ7XMJRZ58C> letöltési dátum: 2020.08.18.

Pollard, R. (2010): Journal of Sport Sciences: Charles Reep (1904-2002): pioneer of notational and performance analysis in football

Polt, R. (2018): Budapesti Corvinus Egyetem, 2018/19 őszi félév, Statisztika II. 12. Előadás (12. dia)

Pudasaini, P. (2014): Top 10 Sports Leagues With the Highest Attendance Per Game

elérési link: <https://www.therichest.com/sports/top-10-sports-leagues-with-the-highest-average-per-game-attendance-in-the-world/> letöltési dátum: 2020.05.17.

Ramanathan, R. (1993): Statistical Methods in Econometrics

Redwood-Brown, A. J., O'Donoghue, P. G., Nevill, A. M., Seward, C., Sunderland, C. (2019): Effects of playing position, pitch location, opposition ability and team ability on the technical performance of elite soccer players in different score line states

Reep, C. (1982). The great Magyar myth exploded; Times

Ronay, B (2012): Euro 2012: Valeriy Lobanovsky, king of Kiev who was before his time

elérési link: <https://www.theguardian.com/sport/2012/jun/15/1> letöltési dátum: 2021.04.06.

Royuela, V., Gásquez, R. (2018): On the Influence of Foreign Players on the Success of Football Clubs

Sagiroglu, S., Sinanc, D. (2013): Big data: A Review

Scrapehero (2017):

elérési link: <https://www.scrapehero.com/sports-the-rise-of-big-data-and-analytics/> letöltési dátum: 2020. 05.01.

Shepotylo, O. (2010): The law of unintended consequences in soccer: impact of three-point-a-win rule on strategies and outcomes

Siregar, C. (2018): What is Total Football? Famous tactics explained: the clubs, countries & players to use it

elérési link: <https://www.goal.com/en/news/what-is-total-football-famous-tactics-explained-the-clubs/w5yd5dzofn4p17ewmav4atk2b> letöltési dátum: 2020.05.08.

Sportskeeda (2021)

elérési link: <https://www.sportskeeda.com/football/expected-goals-xg-xa#:~:text=Expected%20Assists%20is%20calculated%20by%20relying%20on%20the,can%20still%20have%20a%20high%20value%20of%20xA>. letöltési dátum: 2021.01.07.

Sportsshow (2020):

elérési link: <https://sportsshow.net/richest-sponsorship-deals-of-soccer/> letöltési dátum: 2020.08.17.

Stadiumguide (2020):

<https://www.stadiumguide.com/bundesliga-stadiums/> letöltési dátum: 2020.10.03.

Statista (2017):

<https://www.statista.com/statistics/370560/worldwide-sports-market-revenue/> letöltési dátum: 2020.05.01.

Statista (2019a):

elérési link: <https://www.statista.com/statistics/261223/european-soccer-market-total-revenue/>
letöltési dátum: 2020.05.01.

Statista (2020a)

elérési link: <https://www.statista.com/statistics/1126480/sports-betting-revenue-us/> letöltési dátum: 2020.08.17.

Statista (2020b):

elérési link: <https://www.statista.com/statistics/282974/average-per-game-attendance-german-football-bundesliga/> letöltési dátum: 2020.12.05.

Statista (2020c):

elérési link: <https://www.statista.com/statistics/268582/premier-league-league-wide-average-per-game-attendance/> letöltési dátum: 2020.12.05.

Statista (2020d)

elérési link: <https://www.statista.com/statistics/1087391/global-sports-market-size/> letöltési dátum: 2021.03.26.

Theguardian (2020):

elérési link: <https://www.theguardian.com/football/ng-interactive/2020/jul/30/transfer-window-summer-2020-all-deals-europe-premier-league-la-liga-bundesliga-ligue-1-serie-a>

letöltési dátum: 2021.03.26.

Thepunterspage (2020):

<https://www.thepunterspage.com/expected-goals-explained/#author> letöltési dátum: 2020.12.12.

Torgler, B., Schmidt, S. L. (2013): What shapes player performance in soccer? Empirical findings from a panel analysis

Torrens (2020):

elérési link: <https://www.torrens.edu.au/blog/why-sports-industry-is-booming-in-2020-which-key-players-driving-growth#.YGX8J7BxeM9> letöltési dátum: 2021.03.26.

Transfermarkt (2020a):

elérési link: <https://www.transfermarkt.us/> letöltési dátum: 2020.12.05.

Transfermarkt (2020b):

<https://www.transfermarkt.co.uk/ligue-1/besuchertzahlen/wettbewerb/FR1> letöltési dátum: 2020.12.05.

Tunaru, S. R., Winey, H. P. (2010): Valuations of Soccer Players from Statistical Performance Data

Veth, M. (2020): Bundesliga's \$4.95 Billion Television Deal Provides Stability And Perhaps A Brighter Future

Whitney, C. (2010): Borussia Dortmund Sell a Record 51200 Season Tickets

elérési link: <https://www.goal.com/en/news/15/german-football/2010/08/20/2079557/borussia-dortmund-sell-a-record-51200-season-tickets> letöltési dátum: 2021.02.18.

Whoscored (2020a):

elérési link: <https://1xbet.whoscored.com/Statistics> letöltési dátum: 2020.05.08.

Whoscored (2020b)

elérési link: <https://www.whoscored.com/AboutUs> letöltési dátum: 2020.05.08.

Whoscored (2020c)

elérési link: <https://1xbet.whoscored.com/Regions/81/Tournaments/3/Germany-Bundesliga>
letöltési dátum: 2020.07.18.

Wicker, P., Whitehead, J. C., Johnson, B. K., Mason, D. S. (2016): Willingness to pay for sporting success of football – Bundesliga teams

Wilson, B. (2018): Manchester United are China's most popular digital team

elérési link: <https://www.bbc.com/news/business-42875703> letöltési dátum: 2021.04.05.

Wood, L. (2019): Sports - \$614 Billion Global Market Opportunities & Strategies to 2022

elérési link: <https://www.businesswire.com/news/home/20190514005472/en/Sports---614-Billion-Global-Market-Opportunities#:~:text=The%20global%20recreation%20market,%20of%20which%20the%20sports,in%202018,%20having%20grown%20at%204.5%%20since%202014.> letöltési dátum: 2021.03.26.

Worldfootball (2020):

elérési link: <https://www.worldfootball.net/> letöltési dátum: 2020.12.05.

Yue, Z., Broich, H., Mester, J. (2014): Statistical Analysis of the Soccer Matches of the First Bundesliga

Mellékletek:

A labdarúgás eszköztárának egyéb mutatószámai:

A legelső mutató, melyet érdemes megemlíteni további mutatószámok esetén, egy angol statisztikus, James Grayson (Kele, 2020) találmánya, az úgynevezett TSR (Total Shots Ratio) mutató, mely tulajdonképpen azt mutatja meg, hogy egy adott mérkőzésen esett lövések hány százalékát adta le valamelyik csapat. A TSR 0 és 1 közötti értéket vehet fel, százalékosan értelmezhető mutatószám. Grayson szerint az a csapat teljesített egy mérkőzésen jobban, amely magasabb TSR-rel rendelkezett, tehát több lövést adott le egy mérkőzés során. A TSR értelmezése megvalósulhat több mérkőzést vizsgálva is, így a TSR is alkalmas, hogy csapatok teljesítményét teljes szezonokon keresztül vizsgálja. Abban az esetben, ha egy csapat mérkőzésenkénti átlagos TSR mutatója egy futbalszezon során 50% feletti lenne, az azt jelentené, hogy a mérkőzéseik során a lövések több, mint felét ők adták le, vagyis többet tettek a győzelemért (ebben a tekintetben), mint ellenfeleik. A TSR egyik legnagyobb hiányossága az, hogy nem tesz különbséget a lövések minősége között, így ugyanakkora jelentősége van egy tizenegyesrúgásnak, mint egy távoli kapura lövési kísérletnek.

A TSR legnagyobb hibájának kiküszöbölésével, a lövések minőségének értékelése céljával jött létre a modern labdarúgás egyik legjelentősebb találmánya az úgynevezett xG mutatószám (expected goal ratio néven ismert a szakirodalomban). Az xG egy olyan mutatószám, mely a kapura lövéseket minősíti oly módon, hogy mindegyik kapura lövéshez egy adott mérkőzésen hozzárendeli annak a valószínűségét, hogy az adott pozícióból leadott lövésből gól születik (Bundesliga, 2020d). A labdarúgásban kiemelt jelentősége van az xG-nek, hiszen talán ez az a csapatsport az, amelyben a legkevesebb pontot/gólt érik el a játékosok, így egy ilyen mérkőzés esetén a végeredmény nem mindig tükrözi reálisan azt, hogy melyik csapat teljesített jobban, lőtt többet és veszélyesebb helyzetekből kapura egy adott mérkőzésen. A futballt természetesen mindezek ellenére gólrá játsszák, azonban rendkívül hasznos, hogy a csapatok teljesítménye ilyen módon is mérhető.

Az xG elméleti hátterét kicsit alaposabban is körbe kell járni, hiszen rendkívül összetetten kalkulált mutatószámról van szó. A futballpálya minden egyes pontjához tartozik egy valószínűség, hogy egy adott pontról mekkora a valószínűsége egy gól elérésének. Azonban ez a valószínűség nem mesterségesen van kalkulálva, illetve nem annak a valószínűsége, hogy az adott pontról egy üres futballpályán hányszor talált valamely személy a kapuba. Az xG kiszámítása az OPTA hatalmas méretű (többsszázezer valamikori kapura lövést tartalmaz az

adatbázis), lövéseket tartalmazó adatbázisát használja fel a valószínűségek meghatározására (Kálnoki-Kis, 2020). Egy adott lövés xG-jét tehát rengeteg, a gólszerzést befolyásoló tényező határozza meg. Ezen tényezők közé tartozik a szög (Bundesliga, 2020d), amelyben a kapura lövő játékos a kaput látja. Ez a szög minden esetben a két kapufa és a játékos által bezárt szöget jelenti, mely maximálisan 90 fokos lehet, éppen abban az esetben, hogyha egy játékos a kapuval pontosan szemben állva végzi el a kapura lövést. A szög mellett a kaputól való távolság (thepunterspage.com) is rendkívül fontos tényező, értelemszerűen nagyobb távolságból kisebb a valószínűsége a gólszerzésnek. A lövések távolsága alatt a kapu középső pontja, illetve a lövés pillanatában a labda közötti távolságot értjük. A valószínűség kiszámításának részét képezi az a tényező is, hogy hány játékos (legjellemzőbben ezek védőjátékosok) állt a kapura lövő játékos közelében, illetve a kapu és lövő játékos között, mint zavaró, a lövést esetlegesen blokkoló tényezőként (Bundesliga, 2020d). Mindezek mellett a kapura lövő játékos lövés pillanatában történő testhelyzetét (Thepunterspage, 2020) is figyelembe veszi az xG mutató, elvesztett egyensúllyal például nehezebb egy lövést kontrollálni/irányítani, mint teljesen stabilan állva. A testhelyzet mellett a testrészeket sem szabad figyelmen kívül hagyni az xG kiszámításakor, fejvel (Bundesliga, 2020d) például a legtöbb helyzetből nehezebb gólt szerzeni, mint lábbal, de akármilyen akrobatikus mozdulattal még kisebb a valószínűség a gólszerzésre.

Az xG nem csupán egyetlen kapura lövés jellemzésére, hanem teljes mérkőzések, sőt teljes labdarúgó szezonok értékelésére is gyakran használt mutatószám. Mérkőzéseken a csapatok teljesítményét a lövések xG-jének kumulálásával jellemezhetjük (Kocsis 2020), ezen a módon a mérkőzés végére mind a két csapat elér valamekkora várható gól értéket. Teljes labdarúgó szezonokat az úgynevezett xP (Thepunterspage, 2020), azaz expected points mutatószámmal lehet leírni. Ez közvetlenül az xG-ből származtatható mutatószám, mely lényegében a csapatok összes mérkőzését kizárólag xG alapján értékeli, a labdarúgás pontszerzési szabályai alapján a győztesnek három pont jár, a vesztesnek nulla, míg döntetlen esetén egy-egy ponton osztozkodnak a felek. Az xG mutató esetén fontos megemlíteni, hogy a sporttudományban gyakran találkozhatunk az úgynevezett npxG (non-penalty expected goals) mutatóval, mely a játékosok xG-jét úgy vizsgálja, hogy tizenegyesrúgásaikat egyáltalán nem veszi figyelembe. Hogyha egy játékos gyakran végez el tizenegyesrúgásokat, az a legtöbb esetben pozitív irányban torzíthat az xG mutatóján (Thepunterspage, 2020). Az npxG tehát egy torzító tényezőtől tisztított xG mutatószám, amely segítségével pontosabban összehasonlíthatóvá válik játékosok teljesítménye.

Az xG azonban egy kalkulált érték, gyakran előfordul, hogy nem ad reális képet a valóságról. Campbell (2017) szerint az egyik legjelentősebb hibája az xG modellnek, hogy a kapura lövő játékos személyét nem képes figyelembe venni. Egy támadójátékos többször kerül helyzetbe például egy védőnél, feltehetőleg jobban fejezi be helyzeteit védő poszton bevethető társainál. Azonban lehet különbség támadójátékosok között is igen gyakran, a világ legjobb támadói például több gólt szereznek annál, mint amennyi elvárható lett volna tőlük az xG-jük szerint. A Skyports (2017) szerint további hátránya a modellnek, hogy noha a lövő játékos környékén lévő védők számát figyelembe veszi a valószínűség kiszámításakor a modell, a védők tényleges pozícióját a lövés pillanatában egyelőre nem tudja teljes mértékben figyelembe venni. Az xG mindemellett ahogyan a lövést leadó játékosok, úgy az adott lövéseket fogadó kapus személyét sem képes értékelni (Beasley, 2017).

Az xG-hez nagyon hasonló mutató, az úgynevezett xA (expected assists ratio) is rendkívül elterjedt a modern labdarúgásban. Az xA a lövések helyett a passzok jellemzésével foglalkozik és azt vizsgálja, hogy a labdarúgópálya adott pontjáról leadott passz mekkora valószínűséggel lesz gólpasz (assziszt), a modellben minden egyes passz helyet kap, még azok is, amelyek után nem lett kapuralövés. Ez a mutatószám rendkívül hatékonyan mutatja meg, hogy mely játékosok képesek a legjobban passzaikkal helyzeteket teremteni csapattársaik számára (Bate, 2017).