



Közzététel: 2023. november 22.

A tanulmány címe:

Magyar szoftverfejlesztők bérkülönbségeinek elemzése a Stack Overflow kérdőíve alapján

Szerzők:

KOVÁCS LÁSZLÓ

a Budapesti Corvinus Egyetem adjunktusa

E-mail: laszlo.kovacs2@uni-corvinus.hu

VASTAG EMESE

az IFUA Horváth tanácsadója

E-mail: emese.vastag@horvath-partners.com

DOI: <https://doi.org/10.20311/stat2023.11.hu0947>

Az alábbi feltételek érvényesek minden, a Központi Statisztikai Hivatal (a továbbiakban: KSH) *Statisztikai Szemle* c. folyóiratában (a továbbiakban: Folyóirat) megjelenő tanulmányra. Felhasználó a tanulmány vagy annak részei felhasználásával egyidejűleg tudomásul veszi a jelen dokumentumban foglalt felhasználási feltételeket, és azokat magára nézve kötelezőnek fogadja el. Tudomásul veszi, hogy a jelen feltételek megszegéséből eredő valamennyi kárért felelősséggel tartozik.

1. A jogszabályi tartalom kivételével a tanulmányok a szerzői jogról szóló 1999. évi LXXVI. törvény (Szjt.) szerint szerzői műnek minősülnek. A szerzői jog jogosultja a KSH.
2. A KSH földrajzi és időbeli korlátozás nélküli, nem kizárólagos, nem átadható, térítésmentes felhasználási jogot biztosít a Felhasználó részére a tanulmány vonatkozásában.
3. A felhasználási jog keretében a Felhasználó jogosult a tanulmány:
 - a) oktatási és kutatási célú felhasználására (nyilvánosságra hozatalára és továbbítására a 4. pontban foglalt kivétellel) a Folyóirat és a szerző(k) feltüntetésével;
 - b) tartalmáról összefoglaló készítésére az írott és az elektronikus médiában a Folyóirat és a szerző(k) feltüntetésével;
 - c) részletének idézésére – az átvevő mű jellege és célja által indokolt terjedelemben és az eredetihez híven – a forrás, valamint az ott megjelölt szerző(k) megnevezésével.
4. A Felhasználó nem jogosult a tanulmány továbbértékesítésére, haszonszerzési célú felhasználására. Ez a korlátozás nem érinti a tanulmány felhasználásával előállított, de az Szjt. szerint önálló szerzői műnek minősülő mű ilyen célú felhasználását.
5. A tanulmány átdolgozása, újra publikálása tilos.
6. A 3. a)–c) pontban foglaltak alapján a Folyóiratot és a szerző(ke)t az alábbiak szerint kell feltüntetni:
„*Forrás: Statisztikai Szemle* c. folyóirat 101. évfolyam 11. számában megjelent, **Kovács László–Vastag Emese** által írt, **Magyar szoftverfejlesztők bérkülönbségeinek elemzése a Stack Overflow kérdőíve alapján** című tanulmány (link csatolása)”
7. A Folyóiratban megjelenő tanulmányok kutatói véleményeket tükröznek, amelyek nem feltétlenül esnek egybe a KSH vagy a szerzők által képviselt intézmények hivatalos álláspontjával.

Kovács László – Vastag Emese

Magyar szoftverfejlesztők bérkülönbségeinek elemzése a Stack Overflow kérdőíve alapján

Wage differences for Hungarian software developers based on Stack Overflow's questionnaire

Kovács László, a Budapesti Corvinus Egyetem adjunktusa
E-mail: laszlo.kovacs2@uni-corvinus.hu
Vastag Emese, az IFUA Horváth tanácsadója
E-mail: emese.vastag@horvath-partners.com

A szoftverfejlesztés a K+F-szektor legdinamikusabban fejlődő területei közé tartozik. Egy ország K+ F-szektorának fejlettségét többek között az ebben az ágazatban dolgozók bérezése is jellemzi. Ebből kiindulva tanulmányunkban azokat a tényezőket vizsgáljuk meg klasszikus ökonometriai modellekkel, amelyek hozzájárulnak egy magyar szoftverfejlesztő fizetéséhez. Blinder–Oaxaca-dekompozíció segítségével kiemelten elemezzük az egyetemi diploma bérezésre gyakorolt rész- és összetételhatásait, ehhez az évente megjelenő, Stack Overflow-fórum szoftverfejlesztői kérdőíve által összegyűjtött adatokat modellezzük. Tanulmányunkban programozónak tekintünk minden kitöltőt, aki a Stack Overflow felhasználói közösségének tagja, ők zömében szoftver-, illetve webfejlesztéssel foglalkozó vagy rendszerüzemeltető szakemberek. Eredményeink többek között rávilágítanak a kódolási tapasztalat és a korszerű technológiák ismerete közötti interakciós hatásnak a magyar programozók bérezésében betöltött kiemelt szerepére. Egy elavult technológiában szerzett nagy tapasztalatnak akár bércsökkenő hatása is lehet. Eredményeink továbbá azt is igazolják, hogy Magyarországon változatlanul fontos, hogy egy programozó legalább alapszakos diplomával rendelkezzen. Az egyetemet be nem fejező programozók átlagos bére modellünk alapján várhatóan mintegy 7%-kal kevesebb, mint az alapszakos diplomával rendelkező társaiké. Külön-külön vizsgálva lényegében egy programnyelvnek sem volt saját bérnövelő hatása, az viszont már szignifikáns volt, hogy hány programnyelvet ismer vagy szeretne tanulni az illető. Hazánkban az informatika területén mutatkozó magas munkaerőhiány következtében elképzelhető, hogy a munkáltatók számára az is elegendő, ha a leendő munkavállaló akár más nyelvek terén szerzett már kódolási tapasztalatokat, és rendelkezik az alapvető algoritmikus gondolkodással, tehát könnyen átképezhető más nyelven való programozásra is.

Tárgyszavak: programozók bérezése, Stack Overflow-adatok, Blinder–Oaxaca-dekompozíció

Software development can clearly be classified as one of the most dynamically developing areas of the R & D sector. The development of a country's R & D sector is well characterized by the wages of workers in this sector, among other things. That is the reason for examining the factors that contribute to the salary of a Hungarian software developer in our paper. We have analysed the data collected by the Stack Overflow Annual Developer Survey by applying traditional econometric

models. By utilizing Blinder–Oaxaca-decomposition, the endowment and discrimination effects of a university degree on programmer salaries are analysed in detail. In our paper, respondents who are members of the Stack Overflow user community are considered to be programmers. These specialists are mostly software or web developers or system operators. Our results, among others, highlight the important role of the interaction between coding experience and knowledge of modern technologies in the wages of Hungarian programmers. Extensive experience in an outdated technology can even have a wage-reducing effect. Furthermore, our results indicate that in Hungary it is still authoritative if a programmer has at least a bachelor's degree. Based on our model, the average salary of programmers who did not graduate from university is expected to be 7% lower than that of their counterparts with a bachelor's degree. Examined separately, almost no programming language has wage-increasing effect on its own. However, the number of programming languages a person knows or wants to learn has a significantly positive effect on salaries. Due to the shortage of IT workers in Hungary, it is probably sufficient for employers if an employee has already gained coding experience in an arbitrary language and has basic algorithmic thinking, so they can easily be retrained to program in the language desired by the employer.

Keywords: programmer salaries, Stack Overflow data, Blinder–Oaxaca-decomposition

Az informatika mint iparág rendkívül gyorsan fejlődik, az újszerű technológiák elterjedése a keresletet is serkenti a munkaerőpiacon a szoftverfejlesztők irányában. Hazánkban a tevékenységek egységes ágazati osztályozási rendszere (a továbbiakban TEÁOR'08) a szoftvertechnológiai tevékenységeket az információ, kommunikáció nemzetgazdasági ágba sorolja, ezen belül is a 62-es számhoz, az információtechnológiai szolgáltatások közé.

Az információ, kommunikáció nemzetgazdasági ágban kb. 168,5 ezer főt foglalkoztattak Magyarországon 2022-ben, a 2011. évinél majdnem 78%-kal többet (*KSH, 2023a*). Fontos megjegyezni, hogy ebbe az ágazatba a telekommunikációs vállalatok is beletartoznak, de más nemzetgazdasági területeken is dolgoznak szoftverfejlesztők, például a pénzügyi, biztosítási, ipari tevékenységeket magába ölelő szektorokban, úgyhogy az adat inkább csak a nagyságrend szemléltetésére szolgál. A betöltött állások számával párhuzamosan a betöltetleneké is folyamatosan növekszik. Míg az információ, kommunikáció ágban 2011-ben 1257 betöltetlen állás volt, addig 2022-ben 4000-nél is több, azaz több mint háromszorosára nőtt (*KSH, 2023b*). Mindezek alapján elmondható, hogy a nemzetgazdasági ág a munkaerőt tekintve korántsem telített, ami szintén alátámasztja azt, hogy a kereslet az informatikusok (és ezen belül a szoftverfejlesztők) iránt egyre növekszik hazánkban.

Ennek egyik legkézenfekvőbb magyarázata lehet, hogy Magyarországon a 2000-es évektől kezdve egyre inkább elterjedtek a multinacionális cégek leány-

vállalataihoz kötődő SSC-k, azaz *Shared Service Centerek*, amelyek az egyszerű kiszervezéssel ellentétben elkülöníthető, elszámoltatható egységek, és főként adminisztratív, informatikai és pénzügyi szolgáltatásokat nyújtanak a vállalat többi (akár más országban található) területeinek, valamint a vállalaton kívüli ügyfeleknek is (*Janssen–Joha, 2006*). A *KSH (2023c)* szerint 2022-ben hazánkban az információ, kommunikáció nemzetgazdasági ágban már 51 SSC működött, amit egyedül a szakmai, tudományos, műszaki tevékenységek nemzetgazdasági ág előzött meg. Az SSC-k által nyújtott nemzetközi informatikai szolgáltatások értelemszerűen nagy létszámú és képzett, szoftverfejlesztésben jártas munkaerőt követelnek, de szoftverfejlesztéssel foglalkozókra magas igény lehet például az SSC-k által nyújtott pénzügyi és adminisztratív szolgáltatásokban is.

Az informatikai terület munkaerőigényét tovább növelte a Covid19-járvány idején megvalósult tömeges távoli munkavégzéshez, illetve az arra való átálláshoz szükséges infrastruktúra és szoftveres megoldások kiépítése (*Pranggono–Arabo, 2021*). A pandémia következtében súlyosbodó informatikushiány megoldását a magyar állam is a feladatai között tartja számon. Ezt támasztja alá az Újratervezés Program elindítása, amelynek keretein belül a 18. életévüket betöltő magyar állampolgárok állami támogatással jelentkezhetnek informatika területen intenzív képzésekre. A magyar kormány célja ezzel a kezdeményezéssel olyan informatikai szakemberek gyors ki-, illetve átképzése, akik képesek feltölteni a munkaerőpiaci űrt az IT-szektorban. A kurzusok lebonyolításával az állam több független, technológiai képzéseket biztosító szervezetet bíz meg, amelyek az államilag támogatott oktatásokon kívül is számos önköltséges tanulási lehetőséget kínálnak (*Nemzeti Szakképzési és Felnőttképzési Hivatal, 2023*).

A szoftverfejlesztők iránti nagy kereslet és a munkaerőhiány okán a szoftverfejlesztés kiemelt jelentőségű területnek minősül a K+F-szektorban. Egy ország K+F-szektorának fejlettségét többek között az ebben az ágazatban dolgozók bérezése is jól jellemzi. Ebből adódóan fontos megállapítani a bérezést alakító tényezőket az adott ország szoftverfejlesztéssel foglalkozó munkavállalói körében. Jelen tanulmány célja, hogy ezt a kérdést a Stack Overflow-fórum szoftverfejlesztői kérdőíve által összegyűjtött adatok ökonometriai modellezése és Blinder–Oaxaca-dekompozíció alkalmazása segítségével megválaszolja. A Stack Overflow-fórum szoftverfejlesztői kérdőíve tartalmazza a kitöltők havi bruttó fizetését, és ennek modellezéséhez számos munkatapasztalatra, végzettségre, alkalmazott technológiákra és demográfiai tényezőre vonatkozó magyarázóváltozót szolgáltat.

Az általunk elemzett kérdőív adatait több tanulmány is felhasználta már a szoftverfejlesztők munkaerőpiaci elemzésére. *Arbidāne és Silickis (2018)* kifejezetten a litván szoftverfejlesztők munkaerőpiaci kompetenciáira szűkítve vetették össze a 2018-as Stack Overflow-kérdőív litván eredményeit a saját, vállalatoknak szóló kérdőívük kitöltéseivel, ezáltal ütköztetve a munkavállalói és a munkáltatói

munkaerőpiaci preferenciákat. *Silveria és munkatársai (2019)* szintén a 2018-ban publikált kérdőív adatai alapján globális szemszögből nézve elemezték a STEM-szektorban (*Science, Technology, Engineering and Mathematics*) működő vállalatoknál meglévő munkaerő diverzitását. Ugyanakkor mindkét tanulmány elsősorban leíró statisztikai elemzéseken alapult. Tudomásunk szerint a kérdőív adatait formális ökonometriai módszerekkel még nem elemezték a bérezést alakító tényezők feltárására, ebben tanulmányunk újdonságnak tekinthető.

Az 1. fejezetben szakirodalmi forrásokra támaszkodva megfogalmazzuk a magyar szoftverfejlesztők bérezését alakító tényezőkre vonatkozó hipotéziseinket. A 2. fejezetben röviden ismertetjük a bérezés modellezéséhez használt klasszikus ökonometriai módszereket, illetve részletesen bemutatjuk a Blinder–Oaxaca-dekompozíciót. A tanulmány 3. fejezete a Stack Overflow-fórum szoftverfejlesztői kérdőíve által gyűjtött adatok részletes bemutatását és leíró statisztikai elemzését tartalmazza. A 4. fejezetben a bérezést alakító tényezők azonosítására szolgáló modellezés folyamatát ismertetjük. Az 5. fejezetben a diploma bérekre gyakorolt hatásával kiemelten foglalkozunk a Blinder–Oaxaca-dekompozíció eredményein keresztül. Az utolsó fejezetben a preferált modellünk alapján igazoljuk vagy cáfoljuk az 1. fejezetben megfogalmazott hipotéziseket, és általános következtetéseket vonunk le a magyar szoftverfejlesztők bérezését alakító tényezőkről.

1. A magyar szoftverfejlesztők bérezését alakító tényezőkre vonatkozó hipotézisek

A magyar szoftverfejlesztők munkaerőpiacáról elérhető legfrissebb elemzés a *Hays Hungary Salary Guide (2023)* összeállítása, amely a magyar szoftverfejlesztők átlagos havi bruttó bérére vonatkozó kimutatásokat foglalja magában, több technológiai, demográfiai stb. tényező szerinti bontásban. Azonban ez sem tartalmaz formális ökonometriai modellezést, így a kimutatott csoportátlagok közötti különbségek tartalmazhatják más tényezők hatását is (*confounding-jelenség*). Például az években mért tapasztalat szerinti átlagos bérkülönbségekben megjelenhet az életkor és az ismert technológiák, programnyelvek hatása is, az elemzés erre nem kontrollál.

A tanulmányban a *Hays Hungary Salary Guide (2023)* elemzése alapján azonosítjuk a magyar fejlesztők bérezését potenciálisan meghatározó magyarázó tényezőket, de ezek hatását nem csupán részátlagok különbségeivel kívánjuk kimutatni, hanem azok önálló, marginális hatását akarjuk számszerűsíteni formális

ökonometriai modell segítségével, a Stack Overflow-fórum szoftverfejlesztői kérdőívének adatait felhasználva. Ezenkívül Blinder–Oaxaca-dekompozíció segítségével kiemelten vizsgáljuk a diploma meglétének rész- és összetételhatásait a magyar szoftverfejlesztők munkaerőpiacán.

A magyar szoftverfejlesztők bérezését alakító potenciális magyarázó tényezőkre vonatkozó hipotéziseinket – elsősorban a *Hays Hungary Salary Guide* adatai alapján – a következő bekezdésekben foglaljuk össze.

Magyarországon 2022-ben a szoftverfejlesztésben már a junior (0–3 év munkatapasztalattal rendelkező) szinten is havi 800–850 ezer forint között mozgott az átlagos bruttó bér.¹ Egy medior (3–5 év munkatapasztalatot szerzett) programozó esetében ugyanez a szám már az 1 millió forintot is meghaladja. Jelen tanulmányban megvizsgáljuk azt a hipotézist, miszerint bár a junior, medior, szenior karrierút szerint nézve a fizetések drasztikusan nőnek, a szeniorszint felett (kb. 15 év tapasztalat után) a programozási-fejlesztési tapasztalatnak kevesebb hozzáadott értéke van a bérhez. Az informatikai szektorban lezajló gyors változások következtében a fiatalabb generáció tudása sok esetben naprakészebb lehet, míg az idősebbek inkább projektmenedzsment-vezetési feladatokat és kevesebb operatív fejlesztési munkát látnak el.

Az informatika területén bérezés szempontjából nem elhanyagolható, hogy a munkavállalók pontosan milyen pozíciót töltenek be. A *Hays Salary Hungary Guide* eszerint is elkülöníti a fejlesztőket, a bérekben pedig látható szóródás a részcsoportok között. Nagyon látványos például, hogy az adattudósok és a gépi tanúlással foglalkozó specialisták átlagos bruttó bére már juniorszinten is elérheti az 1 millió forintot. Az iparág Magyarországon még gyerekcipőben jár, ezért a téma szakértői kifejezetten előnyös alkupozícióban vannak a bérezésüket illetőleg (*Mayer-Schönberger–Cukier, 2014*). Tanulmányunkban azt is vizsgáljuk, hogy a Stack Overflow-kérdőívben is azonosítható pozíciók között létezik-e olyan részcsoport, amelynek bérezése kimutathatóan eltér a többitől (pozitív vagy negatív irányban).

A *Hays Hungary Salary Guide (2023)* és a *KSH (2023b)* adataiból is látható, hogy a piacon még mindig a munkaerő túlkereslete a jellemző, ami 2022-ben tovább fokozódott. Emellett ugyanakkor a kiadvány adatai alapján ugyanebben az évben trenddé vált a digitális elvándorlás, amelynek oka valószínűleg a kecsgettető nemzetközi bérekből adódó jobb megélhetés reménye, valamint a forint nagymértékű gyengülése. A back-end fejlesztésben a Java és C#, míg a front-endben az Angular és a React programnyelvek ismerete a keresettebb. Ebből adódóan ezek ismeretének a bérré gyakorolt hatásait megvizsgáljuk a Stack Overflow-kérdőíve alapján is, külön-külön.

¹ Az adat nem tartalmazza a béren kívüli juttatásokat és prémiumokat, valamint csak a teljes munkaidőben dolgozó szoftverfejlesztőket vettük figyelembe.

Árgilán és Kelemen (2016) szerint a magyar informatikai oktatás inkább az irodai eszközök használatára készítette fel a tanulókat, és kevésbé a kreatív, algoritmikus gondolkodást igénylő problémák megoldására. Megállapították továbbá, hogy a középiskolában oktatott informatikai tananyag az ilyen irányban továbbtanuló érettségizőket nem készíti fel megfelelően az egyetemi elvárásokra. A 2020-as Nemzeti Alaptantervben már történt előrelépés a 2016-os állapothoz képest, például a programozás már középszinten is helyet kap (5/2020. {I. 31.} Korm. rendelet).

A közoktatáson túllépve, az egyetemi informatikatanítás kérdéskörének is jelentős szerepe van az IT-szektor kiszolgálásában. Az informatikai alapszakok országsszerte népszerűek önköltséges formában is. A 2023. szeptemberi képzések általános felvételi eljárásában az ELTE Informatikai Karára összesen 3066 fő jelentkezett, míg a Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Karára 3427 (*Felvi.hu, 2023*). A mesterszakokra történő jelentkezések száma viszont elmarad a népszerűnek számító gazdaságtudományi területhez képest. Az ELTE Informatikai Karának mesterképzéseire például összesen 388 fő jelentkezett, a Gazdaságtudományi Karra pedig 2022 (*Felvi.hu, 2023*). Az egyetemek hiányosságairól a nemzetközi szakirodalomban *Stroustrup (2010)* is értekezett, aki megállapította, hogy egyre inkább fellelhető egy „rés” aközött, amit az informatikát tanuló hallgatók az egyetemen tanulnak, és amit a munkaerőpiac elvárna a leendő informatikusoktól, szoftverfejlesztőktől.

Tanulmányunkban Blinder–Oaxaca-dekompozíciós modellezéssel mutatjuk ki az átlagbér eltéréseinek rész- és összetételhatásait a diplomás és nem diplomás, valamint a mesterszakot és alapszakot végzett szoftverfejlesztői csoportok között.

A szoftverfejlesztők kora értelemszerűen szoros korrelációban áll a munkatapasztalattal, ugyanakkor azon korábbi hipotézisünk korrekt megítéléséhez – miszerint szeniorszint felett a programozási tapasztalatnak kevesebb hozzáadott értéke van a bérhez – modellünkben fontos számszerűsíteni az életkor marginális (tehát egyéb tényezők mellett a tapasztalattól is független) hatását a szoftverfejlesztők bérezésére.

Megjegyzendő, hogy a demográfiai tényezők közül fontos lenne a nemek közötti bérkülönbségek vizsgálata is, ami több munkaerőpiaccal foglalkozó tanulmány témája. Ugyanakkor a vizsgált években (2018–2022) a Stack Overflow-kérdőív kitöltői között olyan nagy a férfiak túlréprezentáltsága, hogy a nemek közötti bérkülönbségek modellezéséhez nincs meg az elégséges elemszám a nem férfi kitöltők esetében. A részleteket a tanulmány 3. fejezetében mutatjuk be. A nemek közötti bérkülönbségek modellezésének hiánya a magyar Stack Overflow-kérdőív adatainak és tanulmányunk egy lényeges korlátja.

2. Alkalmazott módszerek

A Stack Overflow-kérdőív magyar kitöltőinek béreit mint eredményváltozót többváltozós lineáris regresszió segítségével modellezzük, amelynek együtthatóit a legkisebb négyzetek módszere (LNM) segítségével becsüljük meg (*Stock–Watson, 2020*).

A Stack Overflow-kérdőívben elérhető magyarázóváltozók között modellszelekciót végzünk az Akaike és Bayes–Schwarz információs kritériumok (*AIC* és *BIC*), valamint Wald-teszt segítségével. A Wald-teszt nullhipotézise, hogy egy szűkített modell és egy bővebb modell által magyarázott varianciahányad az eredményváltozóban nem szignifikánsan eltérő (*James et al., 2021*).

Az LNM-becslés feltételei közül a helyes modellspecifikációt Ramsey-féle RESET-teszttel, míg a reziduumok homoszkedaszticitását White-teszt segítségével vizsgáljuk. A RESET-teszt nullhipotézise a modell helyes specifikáltsága, a White-teszté a reziduumok homoszkedasztikus viselkedése (*Wooldridge, 2016*).

Regressziós modellünk additív struktúráját interakciós hatások bevonásával kísérreljük meg feloldani. Amikor interakcióba vonunk két változót egy regressziós modellben, feltételezzük, hogy az egyik marginális hatását befolyásolja a másik szintje, és fordítva (*James et al., 2021*). Ezt olyan változók esetében érdemes megtenni, amelyek hatása eltérhet egy másik magyarázóváltozó által felvett értékek függvényében (például kor és tapasztalat).

A káros multikollinearitást okozó magyarázóváltozókból korrelálatlan látens változókat hozunk létre nem metrikus többdimenziós skálázás segítségével (*Lavery et al., 2019*). A többdimenziós skálázás előnye a klasszikus főkomponens-elemzéshez képest az, hogy az eredeti változók közötti nem lineáris kapcsolatokat is képes figyelembe venni a látens változók előállításához (*Kruskal, 1964*). A módszert Gower-féle távolságmetrikával párosítjuk, így különböző mérési skálájú változók esetében is alkalmazható (*Gower, 1971*).

Mivel a Stack Overflow-kérdőív eredményei minden vizsgált évben (2018–2022) egy-egy keresztmetszeti tábla formájában állnak elő, akár paneladatbázisként is kezelhetnénk a kérdőív adatait. Ugyanakkor a panelökonometriai elemzés egyik alapja az a feltétel, hogy az egyedek azonosíthatók és megkülönböztethetők legyenek, azért, hogy ezzel számszerűsíteni tudjuk az fix egyedhatásokat. Az adatokat anonim gyűjtötték, ezért nem garantálható, hogy aki például 2018-ban kitöltő volt, az a többi évben is kitöltötte a kérdőívet, és az egyes felhasználók sem azonosíthatók. Ilyen értelemben az adatbázisnak az a korlátja, hogy nem használható paneladatbázisként a fix egyedhatások mérésére. Ugyanakkor az idő hatását megtartjuk a kitöltés évére vonatkozó dummyváltozók bevezetésével. Az

ilyen jellegű módszertant hívjuk „ömlesztett” LNM-becslésnek (Wooldridge, 2016).

A Blinder–Oaxaca-dekompozíció egy csoportok közti, egy konkrét változó átlagában mért különbség felbontására alkalmas eszköz, amelyet a társadalomtudományokban kifejezetten a bérdifferenciák szemléltetésére használnak (Oaxaca, 1973; Takács, 2021). Tanulmányunkban ezt a módszert az egyetemi diploma különböző szintjeivel rendelkezők közti átlagos bérkülönbségek mélyebb elemzésére alkalmazzuk.

Egy általános lineáris regressziós analízis során egy tetszőleges C csoport várható értékének becslése az alábbi:

$$\bar{Y}_C = b_C \bar{X}'_C \quad (1)$$

\bar{X}'_C a magyarázóváltozók átlagos értékeit (és a tengelymetszet miatt egy konstans 1-ből álló oszlopot) tartalmazó mátrixot, b_C pedig a regressziós együtthatók vektorát jelöli, valamint az $\bar{Y} = E(\hat{Y})$ az LNM becslése a megfigyelt mintában. Amennyiben A és B csoport átlagos béreinek különbségét szeretnénk meg tudni, a képleteket felírhatjuk mindkét csoportra. Ezután a két csoport becslült átlagainak különbsége a következő:

$$D\bar{Y} = \underbrace{(\bar{X}_A - \bar{X}_B)' b_B}_{\text{összetételhatás}} + \underbrace{\bar{X}'_B (b_A - b_B)}_{\text{részhatás}} + \underbrace{(\bar{X}_A - \bar{X}_B)' (b_A - b_B)}_{\text{interakció}} \quad (2)$$

A (3) összefüggést hívjuk háromtagú dekompozíciónak, amelynek első tagját a szakirodalom *endowment*-nek, azaz „adottságnak” nevezi. Ez a tag jeleníti meg a két csoport közti különbség azon részét, amely abból adódik, hogy a két csoport átlagos egyedei az egyéb megfigyelt magyarázóváltozóknak különböznek. Ez az ún. összetételhatás (Blinder, 1973). A bérezések példáját folytatva, amennyiben a férfiak és nők közti bérdifferenciákat vizsgáljuk, a kor szerinti összetétel hatása lehet például az, ha a mintában több pályakezdő – és emiatt kevesebbet kereső – nő van, mint férfi, ami a nők bérének átlagát csökkenti. A második tag a koefficiensek közti különbség, amely konkrétan azokat a csoportkülönbségeket jeleníti meg, amiket nem az összetétel magyaráz. Ezt más szóval részhatáskülönbségnek is nevezhetjük: az eredményváltozó átlagának különbsége a magyarázóváltozók értékei alapján adott csoportokon belül (pl. egy adott korcsoporton belüli átlagos bérkülönbség). A dekompozíció harmadik tagja az előző kettő interakciója, ami azt implikálja, hogy a rész- és összetételhatás-különbségek egyszerre is fennállhatnak a két csoport közt, egymást erősítve (Jann, 2008).

A (3) összefüggés átrendezésével megkapjuk a kéttagú dekompozíciót, amelyet jelen tanulmányban is alkalmazunk:

$$D\bar{Y} = \underbrace{(\bar{X}_A - \bar{X}_B)' b_R}_{\text{összetételhatás}} + \underbrace{\bar{X}'_A (b_A - b_R) + \bar{X}'_B (b_R - b_B)}_{\text{részhatás és nem megfigyelt hatások}} \quad (4)$$

A kéttagú dekompozíció egy adott R referenciaesetre épített regresszió becsült paramétereire viszonyítja az összehasonlított csoportok regresszióbéli változóinak koefficiensait. Ezt a referenciaesetet a munkaerőpiacot elemző szakirodalom a „diszkrimináció hiányának” nevezi. Ez az az eset, ha a két csoport közt semmilyen megkülönböztetés nem lenne. A dekompozíció azt vizsgálja, hogy a két csoport közti különbség milyen mértékben tér el ettől az elméleti értéktől, és az eltérés mennyire indokolható a csoportoknak a megfigyelt magyarázóváltozók szerinti összetételével, valamint a részhatáskülönbségekkel (*Blinder, 1973*).

A kéttagú dekompozíció első tagja a megmagyarázott különbség, ami jelentéstartalomban ekvivalens a fejezetben korábban tárgyalt összetételhatással. Ez az a differencia, amelyet a mintában található két csoport megfigyelt magyarázóváltozók szerinti összetétel-különbségei okoznak. A második tag a nem megmagyarázott komponens, amelyet a szakirodalom „diszkriminációnak” hív. Ugyanakkor ez a tag magában foglalja a modellben nem megfigyelt hatásokat is. Jelentését tekintve ez a kéttagú felbontás azon része, amely a részhatáskülönbséggel megegyező. A háromtagú felbontás interakciós tagja az egyenletrendezés során gyakorlatilag ebbe a komponensbe került, így ezen tényező megfelelő értelmezése nem mindig egyértelmű. Különösen azért nem, mert a modellben szereplő magyarázóváltozókat több, modellen kívüli hatás is befolyásolja (*Jann, 2008*).

A referenciaeset koefficiensének megválasztására többfajta lehetőség is adódik. Választhatjuk például konkrétan az egyik csoport regressziós együtthatóit, amennyiben megvan az előfeltételezésünk, hogy egyedül az a csoport tapasztal megkülönböztetést. Például a férfi-női bérkülönbségek esetében feltételezhetjük, hogy csak a nőket diszkriminálják (*Oaxaca, 1973*). Ugyanakkor léteznek matematikai megközelítések is. *Cotton (1988)* például a csoportok együtthatóinak mintabeli megoszlással súlyozott átlagát javasolja. Jelen tanulmányban a Neumark-féle megközelítést alkalmazzuk, ami olyan regressziót épít, amely egyáltalán nem foglalja magában a csoportképző ismérvet mint regresszort (*Neumark, 1988*).

3. Az adatok bemutatása és leíró statisztikája

A Stack Overflow weboldal fórumot és közösséget biztosít világszerte a szoftverfejlesztőknek, illetve mindenkinek, aki a programozással hobbi- vagy professzionális szinten foglalkozik. A fórum látogatói bázisa Magyarországon is kellően széles, mivel a programozók itt gyorsan és hitelesen megoldást tudnak találni a

munkájuk közben felmerülő problémákra. A weblap készítői minden évben közlésnek egy kérdőívet (a Stack Overflow éves szoftverfejlesztői kérdőívét), amelyben a kitöltő szoftverfejlesztők demográfiai adatain túl a bérezésükről, a munkahelyükről, a tapasztalataikról kérdezik őket (*Stack Overflow, n.d.*). Elemzésünk tárgyát a 2018–2022 közötti, magyar kitöltők által szolgáltatott adatok képezik. 2018-tól kezdődően áll rendelkezésre megfelelő számú (évente átlagosan 450–500) magyar kitöltő adata, továbbá a gyors technológiai változások miatt 5 évnél korábbi adatokat nem tartunk relevánsnak.

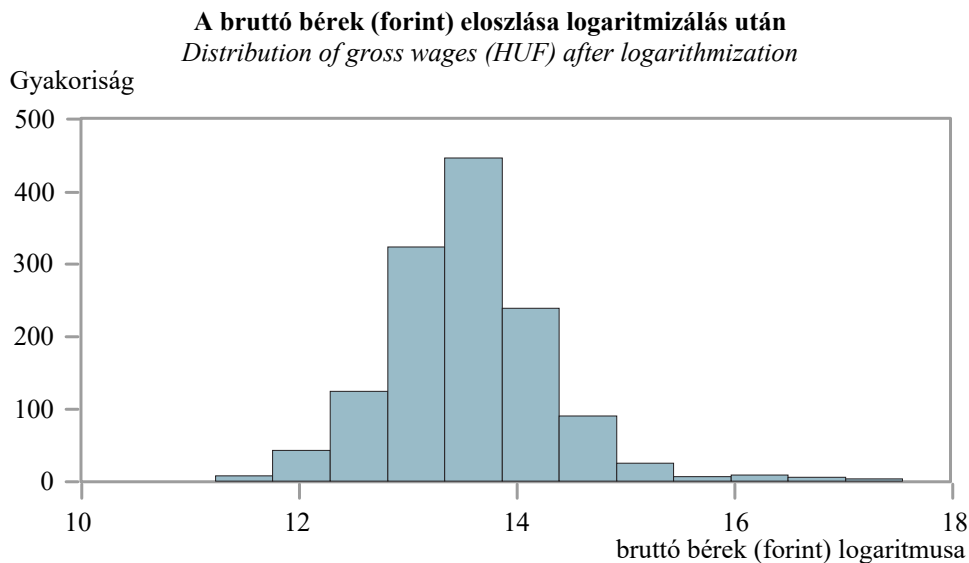
A kezdeti adatbázisban az egyes sorokban található megfigyelések maguk a válaszadók voltak, míg az oszlopok a tulajdonságaik. Egy-egy válaszadó jellemzői az alábbi nagyobb csoportokra voltak tagolhatók: demográfiai adatok, képzettséggel és munkahellyel kapcsolatos adatok, technológiai ismeretek és Stack Overflow-használati szokások. A kérdőívben szereplő kérdésekben évről évre történtek kisebb változások, így csak azon változók kerültek a végső adatbázisba, amelyek minden évre rendelkezésre álltak. Egy évnyi adatban átlagosan 60–70-féle ismérvtől állt rendelkezésre, az összefűzött adatbázisba végül 27 változó került. Ez a 2018–2022 közötti változók közös metszete. A kiinduló adatbázis összesen 2228 magyar válaszadó adatait tartalmazta.

Elemzésünk eredményváltozója a fizetés, így azt a hiányzó értékektől megtisztítottuk. Ezután 1295 fő maradt az adatbázisban, évenként átlagosan 250–300. Az ismeretlen fizetéseket azért nem célszerű semmilyen gépi tanulási módszerrel visszapótolni, mert ha a magyarázóváltozók alapján akarnánk a hiányzó értékeket pótolni az eredményváltozóban, azzal későbbi regressziós modellünkben indokolatlanul felnagyítanánk az egyes változók marginális hatásait (*Van Buuren, 2018*). A hiányzó fizetések kiszűrése nagyjából 42%-os adatvesztéssel járt, ugyanakkor az évek megoszlása egyenletes maradt. Ez azt is jelenti, hogy évente a kitöltők közel fele nem adta meg a bérét, jellemzően azok, akik több mint 30 év professzionális kódolási tapasztalattal rendelkeznek, és/vagy nem adták meg a korukat. Tehát, feltehetően a szűrt adatbázisunk a fiatalabb programozók felé némileg túlréprezentált lehet.

A szűrt adatbázisban a kitöltők nagyjából 75%-a jelölte, hogy forintban kapja a fizetését, ugyanakkor 102 ember az eurót, 42 az amerikai dollárt, 11 az angol fontot és 1 fő a svájci frankot jelölte bére devizájaként. Az átváltásokhoz a kitöltés évének átlagos árfolyamát használtuk a Magyar Nemzeti Bank honlapjáról. Ezzel a béreket folyó áron vesszük figyelembe, de az évek közti különbségekre a modellben a 2. fejezet szerint dummyváltozókkal kontrollálunk. A kiinduló adatbázisba bekerülő bérek tehát a forintban értendő bruttó havi bérek voltak, amelyeket a válaszadóknak a bónuszok és az egyéb, béren kívüli juttatások nélkül kellett megadni. 172 fő megadta a fizetését, de a deviza mezőjét nem töltötte ki. Mivel a két leggyakoribb, forinttól különböző deviza az euró és az amerikai dol-

lár volt, így ezekben az esetekben, ahol forinthez képest alacsony, azaz 3 vagy 4 számjegyű volt a fizetés, az adott évi USD- és euró-átlagárfolyamok átlagával váltottuk át a béreket. Végezetül a fizetést logaritmizáltuk, ugyanis – a béradatakra jellemző módon – jobbra rendkívül elnyúló volt. A logaritmikus skálára való váltás után az eloszlás közel szimmetrikus, ám némileg csúcsos lett (1. ábra).

1. ábra



Forrás: saját szerkesztés.

Regressziós modellünk magyarázóváltozói nominális változók, így azokat dummyváltozóval reprezentáljuk. Ahhoz, hogy az egyes kategóriák dummyváltozóinak standard hibái ne legyen túlságosan nagyok, a ritka (5% alatti arányú) kategóriákat összevontuk, például az életkor esetében a 18 év alatti és a 18–24 éves korcsoportot (1 fő volt 18 év alatt az 1295-ből).

A kérdőív minden évben tartalmaz kérdéseket az ismert és a munkában használt különböző technológiákról, amelyek a következő kategóriákba sorolhatók (mindhez 1-1 kérdés tartozott a kérdőívben): a munkában használt programnyelvek, adatbázis-technológiák, platformok és keretrendszerek. Mindegyik kategóriában a már elsajátított és használt technológiákon túl volt 1-1 olyan kérdés is, amely arra irányult, hogy melyek azok az adott kategórián belül, amelyeket a kitöltő a későbbiekben szeretne elsajátítani. Ezeknél a kérdéseknél a kitöltő bármennyi válaszlehetőséget megjelölhetett. Ebből fakadóan a technológiai magyarázóváltozók esetén mi is az említett négy kategória mentén képeztük meg az ismérveket. Azon legismertebb technológiákat – például programnyelvek esetén a

TIOBE-index² által 2022 szeptemberében 20 legnépszerűbbnek tartott nyelvet – válogattuk ki külön dummyváltozóba, amelyek az adatbázisban is kellően nagy számossággal (több mint 100) rendelkeztek. Emellett változókat vontunk itt össze, amennyiben logikailag közel álló technológiákról volt szó, és a kitöltők jellemzően együtt jelölték meg őket (pl. jellemzően front-end webfejlesztéshez használt HTML, CSS, JavaScript, PHP és TypeScript nyelv). Azon technológiákat, amelyek nem kaptak különálló dummyváltozót, egy „Egyéb” kategóriába soroltuk, amelynek jelentése, hogy az illető megjelölt egy, a külön dummyváltozóba sorolt technológiáktól különböző technológiát, vagy nem adott választ a kérdésre. Ezt az előzetes válogatási módszert a programnyelvek, adatbázis technológiák, platformok és keretrendszerek esetén is külön-külön elvégeztük. Minden, külön dummyváltozóba sorolt technológia kétféle változóval rendelkezett: a kitöltő dolgozott-e már az adott technológiával („*WorkedWith*”), vagy szeretné-e azt a későbbiekben ismerni, megtanulni („*Desire*”). Az adatbázis változóinak a fenti elvek mentén kialakított listáját az 1. táblázat tartalmazza.

Az adatbázis egy olyan, online felületen keresztül elérhető önkéntes kérdőív alapul, amely nem tekinthető véletlen mintának a magyar szoftverfejlesztők sokaságából, így számos ismérven a reprezentativitás nem biztosított. A következőkben összefoglaljuk a leglényegesebb reprezentativitást torzító tényezőket, amelyek elsősorban az adatforrás online jellegéből származnak.

A kérdőív – bár számos lényeges kérdéssel kapcsolatban összesít válaszokat – nem terjed ki például a fejlesztők egyéni attitűdjére. Az egyes évek eredeti kérdőíveiben szerepeltek ilyen jellegű kérdések (pl. munkahelyi elégedettség, szokások), ugyanakkor ezek a vizsgált évek között nem voltak egységesek, így nem volt lehetőségünk őket bevonni a végső, kiinduló adatbázisba.

A kérdőív kitöltéséhez nem szükséges Stack Overflow-felhasználónak lenni, ugyanakkor várhatóan azok töltötték ki, akik – még ha nem is felhasználók – naponta látogatják az oldalt. Ennél fogva a mintában felmerülhet egy torzítás efelé a csoport felé, ami például abban mutatkozhat meg, hogy elsősorban a fiatalabb korosztály van túlsúlyban a kitöltők között.

Mivel a kitöltés önbevallásos alapon történt, fennáll a veszélye, hogy egyes kitöltők nem a valós információkat adták meg magukról. Ez történhetett szándékos és véletlen módon is (például, ha félreértették a kérdést). A véletlen félreértések okozta torzítások egyike lehet, amennyiben az ismert technológiák esetén a kitöltők nemcsak azokat a jelölték meg, amelyekkel munkahelyen már dolgoztak, hanem minden olyan további, amelyet csak felületesen is, de ismernek (például egyetemi órákról), ám ténylegesen nem használnak. Ez különösen a fiatalabb, egyetemista vagy pályakezdő fejlesztők esetén tud zavaró hatással bírni az eredményekre.

² <https://www.tiobe.com/tiobe-index/>

1. táblázat

A feldolgozott Stack Overflow-kérdőívekből összeállított adatbázis változóinak listája
Variables of the dataset compiled from processed Stackoverflow questionnaires

Változó neve	Leírás
ID	Egyedi azonosító
Salary	Havi bruttó kereset (forint)
Currency	Havi bruttó kereset eredetileg megadott devizája
Year	Kitöltés éve
Gender	Nem
Sexuality	Szexualitás
Age	Életkor-kategória
Ethnicity	Etnikum
Data analyst_Data scientist_engineer	Data Analyst, Data Scientist, Data Engineer pozícióból legalább egyet betölt
Backend developer	Back-end fejlesztő?
Front_end developer	Front-end fejlesztő?
Full_stack developer	Full-stack fejlesztő?
Academic researcher_Educator	Egyetemi oktató vagy kutató?
Mobile_Game developer	Mobil- vagy alkalmazásfejlesztő?
DevOps specialist_Product manager	DevOps-specialista vagy -termékmenedzser?
Database_System administrator	Adatbázis- vagy rendszeradminisztrátor?
Other DevType	Megjelölt bármilyen egyéb informatikai állást?
Employed	Munkavégzés beosztása (pl. teljes állás, részmunkaidő stb.)
EdLevel	Iskolai végzettség
YearsCoding	Hány éve kódol? (Kategorikusan adott)
DjangoWorkedWith	Dolgozott már a Django-keretrendszerrel?
SpringWorkedWith	Dolgozott már a Spring-keretrendszerrel?
jQueryWorkedWith	Dolgozott már a jQuery-keretrendszerrel?
OtherFrameworkWorkedWith	Dolgozott már az előzőektől eltérő keretrendszerrel?
DotNET_AspNetDotNETDesire	Szeretné a .NET-, ASP.NET-technológiákat megismerni?
AngularDesire	Szeretné az Angulart megismerni?
Nodejs_Vuejs_ExpressJsDesire	Szeretné a Node.js, Vue.js, Express.js valamelyikét megismerni?
React_Desire	Szeretné a Reactot megismerni?
DjangoDesire	Szeretné a Django-t megismerni?
YearsCodingProf	Hány éve kódol professzionális szinten? (Kategorikusan adott)
OrgSize	Munkahely mérete (Létszám kategorikusan adott)
LanguageWorkedWith	Nyelvek, amellyel a kitöltő már dolgozott
LanguageDesire	Nyelvek, amelyeket a kitöltő a jövőben ismerni szeretne

(A táblázat folytatása a következő oldalon)

(folytatás)

Változó neve	Leírás
PlatformWorkedWith	Platformok, amelyekkel a kitöltő már dolgozott
PlatformDesireNextYear	Platformok, amelyeket a kitöltő a jövőben ismerni szeretne
SQLWorkedWith_Database	Dolgozott már SQL-alapú relációs adatbázisokkal?
MongoDBWorkedWith	Dolgozott már 0-alapú nem relációs adatbázisokkal?
RedisWorkedWith	Dolgozott már Redis-alapú nem relációs adatbázisokkal?
ElasticsearchWorkedWith	Dolgozott már Elasticsearch-alapú nem relációs adatbázisokkal?
OtherNoSQLWorkedWith	Dolgozott már egyéb, nem relációs adatbázis-technológiával?
SQLDesire_Database	Szeretné elsajátítani az SQL-alapú relációs adatbázisokat?
MongoDBDesire	Szeretné elsajátítani a MongoDB-adatbázis-technológiát?
RedisDesire	Szeretné elsajátítani a Redis-adatbázis-technológiát?
ElasticsearchDesire	Szeretné elsajátítani az Elasticsearch-adatbázis-technológiát?
OtherNoSQLDesire	Szeretne egyéb, nem relációs sémás adatbázis-technológiákat megismerni?
DotNET_AspNetDotNETWorkedWith	Dolgozott már .NET-, ASP.NET-keretrendszerekkel?
AngularWorkedWith	Dolgozott már Angular-keretrendszerrel?
Nodejs_Vuejs_ExpressJsWorkedWith	Dolgozott már a Node.js-, Vue.js, Express.js-keretrendszerek valamelyikével?
React_Worked_With	Dolgozott már a React-keretrendszerrel?
SpringDesire	Szeretné a Springet megismerni?
jQueryDesire	Szeretné a jQueryt megismerni?
OtherFrameworkDesire	Szeretne egyéb keretrendszereket megismerni?
OpSysWindows	Munkában Windowst használ?
OpSysLinux	Munkában Linuxot használ?
OpSysMac	Munkában Macet használ?
SOHasAccount	Van Stack Overflow-felhasználója?
SOVisit	Milyen gyakran látogatja a Stack Overflow-t?
SOParticipate	Milyen gyakran ír megjegyzéseket a Stack Overflow fórumán?

Forrás: saját szerkesztés.

Az adatbázis további reprezentativitást befolyásoló sajátosságaira pár, a kiemelt változó évekre lebontott leíró statisztikai elemzése világít rá (2. táblázat).

Az adatbázisban szereplő kitöltők gyakorisága az évek során stagnált, átlagosan 250 fő körül mozgott. Minden évben olyan kevés nő töltötte ki a kérdőívet, hogy az adatbázis a nemek közötti bérkülönbségek jellemzésére nem alkalmas, mivel a kis elemszám miatt a standard hibák túlságosan nagyok lennének. Ez azt is jelenti, hogy adatbázisunk lényegében csak a férfi szoftverfejlesztők bérezési

viszonyairól árulkodik. Magyarországon a szoftverfejlesztés területén is jelentős a férfiak felülreprezentáltsága, még ha nem is olyan mértékben, mint a jelen adatbázisban (KSH, 2023a). Az életkort tekintve viszonylag fiatalabb munkavállalók voltak a kitöltők, minden évben a 25–34 éves korosztály volt a leggyakoribb, őket rendre a 35–44 évesek követték. Ez egybevág azzal a ténnyel, hogy a kódolási tapasztalatokat tekintve például professzionális szinten a 3–5 évig terjedő kategória volt a leggyakoribb.

2. táblázat

A kiinduló adatbázis leíró statisztikai elemzése*Descriptive statistics for the initial dataset*

Megnevezés	2018	2019	2020	2021	2022
Gyakoriság	230	328	234	261	242
Nők, fő	9	19	4	5	4
Férfiak, fő	203	303	212	249	232
Leggyakoribb életkor-kategória, év	25–34 év	25–34	25–34	25–34	25–34
Fizetések átlaga, forint	2 786 156,62	2 930 089,27	3 040 425,80	3 004 724,51	3 843 516,99
Fizetések mediánja, forint	954 000	813 000	1 100 000	1 300 000	1 500 000
Leggyakoribb pozíció	Full-stack fejlesztő	Full-stack fejlesztő	Full-stack fejlesztő	Full-stack fejlesztő	Full-stack fejlesztő
Leggyakoribb iskolai végzettség	Msc. vagy ezzel egyenértékű szint	Bsc. vagy ezzel egyenértékű szint	Bsc. vagy ezzel egyenértékű szint	Bsc. vagy ezzel egyenértékű szint	Bsc. vagy ezzel egyenértékű szint
Leggyakoribb munkahelytípus	10–100 fős	10–100 fős	10–100 fős	10–100 fős	10–100 fős
Leggyakoribb ismert programnyelv	Webfejlesztéshez használt nyelvek	Webfejlesztéshez használt nyelvek	Webfejlesztéshez használt nyelvek	Webfejlesztéshez használt nyelvek	Webfejlesztéshez használt nyelvek
Leggyakoribb kódolási tapasztalat, év	9–11	6–8	9–11	9–11	6–8
Leggyakoribb professzionális kódolási tapasztalat, év	3–5	3–5	3–5	3–5	3–5

Forrás: saját szerkesztés.

Az összes évre jellemző az, hogy a legtöbb válaszadó full-stack fejlesztő volt, ami megmagyarázza azt, hogy miért a webfejlesztéshez használt nyelvek fordultak elő legtöbbször az adatbázisban. Ezek azok a nyelvek, amelyeket az adatbázis

előfeldolgozása során egy közös kategóriába soroltunk: HTML, CSS, JavaScript, PHP és TypeScript. Hozzá kell tenni, hogy mivel ez a kategória öt nyelvet is magába foglal, a nagyszámú gyakoriság az összevonásból is adódhat. A második leggyakrabban használt nyelv az SQL (évenként átlagosan 136 felhasználó adta meg), ezt a Python és Bash/Shell/PowerShell követte, évenként átlagosan 100 és 98 kitöltővel. Ezen jellemzők arra utalnak, hogy a kitöltők között a webfejlesztők felülreprezentáltak. Ennek a jelenségnek lehet olyan torzító hatása a későbbi eredményeink tekintetében, hogy az adatbázisban előforduló nem webfejlesztő pozíciók és az ezekhez köthető technológiák kevésbé tűnhetnek szignifikánsnak a valóságnál.

A leggyakoribb kategória a legmagasabb iskolai végzettség esetében az első vizsgált évet leszámítva az alapképzés vagy az azzal egyenértékű szint volt, és a mesterfokozat gyakorisága minden évben csak kevéssel lemaradva követte. Érdekes továbbá, hogy minden évben a kitöltők átlagosan 20%-át tették ki olyanok, akik elkezdtek valamilyen főiskolai vagy egyetemi képzést, de nem fejezték be.

Ami a fizetéseket illeti, az adatbázisban alapvetően a bruttó fizetések adottak, illetve az átlagot a kiugró értékek könnyen el tudják mozdítani, így ebből adódhatnak az elsőre irreálisan magasnak tűnő bérek. A mediánok már valósabb képet mutatnak. Megállapítható, hogy a fizetések évről évre növekvő tendenciát mutatnak. Ez egyfelől lehet az infláció, vagy az árfolyamok hatása forinttól különböző devizák esetén. A táblázatban feltüntetett fizetések a *Hays Hungary Salary Guide* szerinti átlagos medior-senior fizetést tükrözik. Az átlag és a medián elhelyezkedése minden évben jobbra elnyúló eloszlásra utal, logaritmizálás nélkül.

Mivel az adatbázisban a leggyakoribbak a webfejlesztéssel foglalkozó szoftverfejlesztők voltak, ezt a részcsoportot a 2. ábrán alaposabban is megvizsgálunk.

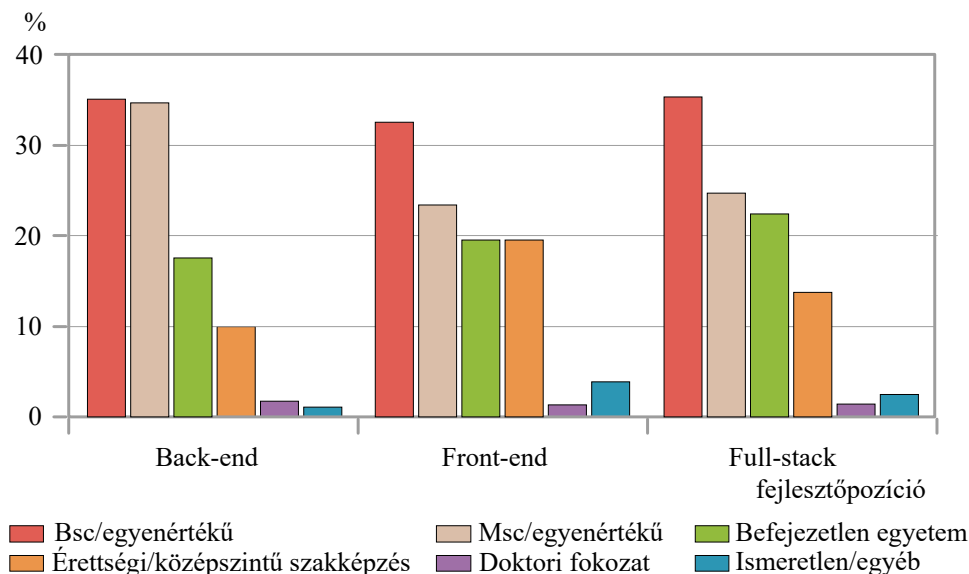
Az iskolai végzettségeket tekintve elmondható, hogy az adatbázisban lévő webfejlesztők közül arányaiban az egyetemi végzettséget megszerzettek fordulnak elő a leginkább, a legtöbb mesterfokozattal rendelkező fejlesztő pedig a back-end fejlesztők körében van. Érdekes információ továbbá, hogy egyik csoport körében sem elhanyagolható az egyetemet otthagynók és a középiskolai, középszintű szakképesítéssel rendelkezők aránya. Front-end és full-stack fejlesztők esetében a mesterfokozattal rendelkezők arányát jelentősen meg is közelítik.

A munkarendet megfigyelve mindhárom webfejlesztői állásnál a legjellemzőbb a teljes munkaidőben foglalkoztatott kategória, átlagosan 65% körül, ezt pedig rendre mindegyik csoportban az egyéni vállalkozói státusz követi, ami mindhárom fejlesztői részcsoportban körülbelül 31%-os arányban fordul elő.

2. ábra

A webfejlesztés területén dolgozók megoszlása iskolai végzettség szerint a Stack Overflow-adatbázisban

Distribution of web developers by educational level in the Stack Overflow dataset



Forrás: saját szerkesztés.

4. A modellépítés folyamata és eredménye

A kiinduló modellben a logaritmizált béreket becsljük az adatbázisban található összes egyéb változó függvényében. Először azokat a nominális változókat szelektáljuk, amelyeknek egyik ismérvkimeneteléhez tartozó dummyváltozó sem szignifikáns semmilyen szokásos szignifikanciaszinten a parciális t-próbák p-értékei szerint. Ebben az esetben a szelekciót megelőzően minden nominális változóra külön-külön ellenőrizzük az információs kritériumok és a Wald-teszt eredményeit, illetve amennyiben a dummyváltozók együttesen elhagyhatónak minősíthetők a három szelekciós eszköz alapján, akkor a nominális változót. A technológiai változók esetén úgy végeztük el a szelekciót, hogy az adott technológiai dummyváltozót a hozzá tartozó egyéb kategóriába soroltuk, majd kivettük a modellből.

A teljes szelekciós fázist megelőzően ellenőriztük minden változó VIF-mutatóját, amely csak 4 darab esetén jelzett káros multikollinearitást. Ezek a

YearsCoding, a *YearsCodingPro*, a *SOParticipate* és a *SOHasAccount* változók voltak. Az eredmény azért nem meglepő, mert feltételezhető az erős kapcsolat aközött, hogy valaki hány éve kódol és hány éve kódol professzionális szinten. Ehhez hasonlóan, aki Stack Overflow-felhasználó, az várhatóan nagyobb eséllyel vesz részt a fórumbeszélgetésekben, így a szoros összefüggés itt is érthető. A szelekció során ebből a négy változóból csak a *SOHasAccount*-ot hagytuk el. Ennek oka, hogy a *SOParticipate* rendelkezett olyan ismérvkimenetekkel, amelyek minden szokásos szignifikanciaszinten szignifikáns dummyváltozóval bírnak, illetve ez a tényező magában hordozza a *SOHasAccount* információtartalmát, olyan szempontból, hogy csak az tud hozzászólni a fórumon feltett kérdésekhez, akinek van felhasználói fiókja. A kódolási tapasztalatokra vonatkozó két változót nem hagytuk el, hanem a nem metrikus többdimenziós skálázás módszerét felhasználva látens változót alkottunk belőlük, ami a kitöltő kódolási tapasztalatának feleltethető meg. Ez onnan látható, hogy az eredményül kapott – 0-ra centrált – látens változó és az eredeti *YearsCoding* és *YearsCodingProf* változók közti korreláció pozitív, míg az egyéb változókkal a látens változó korrelációja 0 közeli.

3. táblázat

A modellszelekció főbb lépései
Main steps of model selection

Állapot	Modell változójának neve az R-kódban	Leírás	Változók száma (tengelymetszet nélkül)	AIC	BIC	Wald-teszt p-értéke, %
1	model_full_log	Kiinduló modell, amely az összes változóval (kivétele a deviza) becsülte a logaritmizált béreket	110	2391	2970	–
2	model_szelektalt_3	Első szűkített modell, amelyben csak a p-érték alapján nem szignifikáns változókat szelektáltam	93	2387	2878	4,77
3	vegso_szelektalt	Végső szűkített modell	70	2364	2736	55,62

Forrás: saját szerkesztés.

A modellépítés főbb fázisait a 3. táblázat foglalja össze. A Wald-teszt p-értékei a táblázatban mindig az adott állapotot hasonlítják az előző állapothoz, tehát a 3. táblázat harmadik sorában az 58,27% a *model_szelektalt_3* modell és

vegso_szelektalt modell közti 23 változó elhagyására kalkulált p-érték, ugyanakkor minden nominális változóhoz tartozó dummyk elhagyhatóságát külön-külön is teszteltük, mielőtt elhagytuk volna őket.

A nem metrikus többdimenziós skálázás után számos lehetséges interakciót és kvadratikus tag bevonását teszteltük, amelyekről – hasonlóan a szelekcióhoz – az információs kritériumok és a Wald-teszt alapján döntöttünk. A kor és a tapasztalat látens változójának interakciója volt az egyetlen, amely az információs kritériumok és a Wald-teszt alapján is javított a modellen. A RESET-teszt eredménye alapján az interakciókat tartalmazó modell minden szokásos szignifikanciaszinten jól specifikáltnak tekinthető.

A harmadik, végső szelektált modell együtthatóit, azok parciális t-próbájának p-értékét a 4. táblázat tartalmazza.

A végső modell különböző szinteken szignifikáns együtthatói alapján már számos következtetést tehetünk a magyar szoftverfejlesztők bérezését alakító tényezőkről.

Az iskolai végzettségek dummyváltozói közül az alapszakos vagy az azzal egyenértékű végzettséggel rendelkező kitöltők csoportja volt a referenciakategória. Egyedüli szignifikáns együtthatója a mesterszakot végzett fejlesztőknek volt, ugyanakkor csak 10%-os szignifikanciaszinten. A végső modell regressziós becslése szerint minden egyéb tényező változatlansága mellett a magyar szoftverfejlesztők közt átlagosan kb. 7,5%-kal keres többet egy mesterdiplomával rendelkező szakember egy alapszakot végzettnél. 5%-os szignifikanciaszint mellett az eredmények alapján továbbá az állítható, hogy azok a kitöltők, akik nem adták meg iskolai végzettségüket, nem végezték el a közoktatást, vagy csak általános iskolát végeztek, várhatóan körülbelül 24%-kal keresnek kevesebbet az alapszakkal rendelkezőknél, feltételezve, hogy minden egyéb tényezőt konstans értéken tartunk. Ez az eredmény ebben az értelemben igazolja azt, hogy a magyar IT-szektorban továbbra is van relevanciája az alapfokú egyetemi végzettségnek, a szakképzés és az egyéb, nem egyetemi kurzusok erőteljesebb megjelenése ellenére is.

Összességében elmondható, hogy a programnyelvekkel kapcsolatos változók közül szinte egyik sem volt szignifikáns, és elhagyásuk minden tekintetben javított a modellen. Ugyanakkor, ha konkrét nyelvek helyett két olyan változót vezetünk be, amelyek azt foglalják magukban, hogy az illető hány nyelvet ismer, vagy hányat szeretne a későbbiekben megtanulni, akkor már szignifikánsabb hatásokat tapasztalhatunk. Érdekes módon az ismert programnyelvek száma negatív előjellel jelent meg a végső modellben, míg a megismerni vágyott nyelveké pozitívvval. Az eredményre magyarázat lehet, hogy számos, ma már elavultnak tekinthető programnyelv létezik, amelyek ismerete nem feltétlenül növeli a bért, az viszont meghatározó tényező lehet, hogy egy fejlesztő új nyelvek szintaxisát és logikáját mennyiben sajátítja el. Az eredményeknek olyan értelmezése is lehetséges, miszerint az, hogy egy bizonyos programnyelvet vagy platformot is-

merjen a fejlesztő, nem megkülönböztető tulajdonság, hanem eleve belépési kritérium az adott vállalathoz. A magasabb kompenzációt a fejlesztők az ezen alapkompenciákon felül felmutatott teljesítményükért kaphatják.

4. táblázat

A végső modell változói, e -ra emelt β együtthatókkal és p -értékekkel

Exponentiated β coefficients of the final model with p -values

(Variables in bold are significant on at least 5%)

Változó neve	e^β	p -érték, %	Változó neve	e^β	p -érték, %
Languages_Worked_With	0,99	5,4	EdLevelSeco- ndarySchool_Associa	0,94	25,4
Languages_Desire	1,01	7,3	EdLevelUnknown_ Primary_NeverCo	0,76	2,0
Year2019	0,89	2,8	EdLevelProfessionalDoctoralDeg	0,96	69,8
Year2020	1,01	59,9	OrgSize1000To4999	0,98	76,7
Year2021	1,08	22,7	OrgSize100To1000	1,00	94,3
Year2022	1,04	64,5	OrgSize10To100	0,80	0,0
Age25 to 34 yrs	1,13	30,0	OrgSize5000To9999	1,09	55,8
Age35 to 44 yrs	1,06	48,4	OrgSizeFewerThan10	0,77	0,0
Age45 to 54 yrs	1,29	16,8	OrgSizeJustMe	0,96	70,1
Age55 to 64 yrs	0,74	63,2	OrgSizeNA_DontKnow	0,78	1,9
AgeUnknown	0,92	84,0	SQLWorkedWith_Database1	0,93	33,1
Database_ System.administrator1	0,88	0,3	ElasticsearchWorkedWith1	1,18	0,1
Other.DevType1	1,11	56,1	OtherNoSQLWorkedWith1	1,06	53,6
EmployedEmployedPartTime	0,70	0,0	SQLDesire_Database1	1,04	54,2
EmployedIndependentContractor	1,32	0,0	ElasticsearchDesire1	1,09	11,9
EmployedReti- red_NotEmployed_Un	2,01	1,5	OtherNoSQLDesire1	0,96	65,7
EdLevelMastersDegree	1,08	8,9	AngularWorkedWith1	1,16	0,3
EdLevelLeftUniversityOrCollege	0,91	12,2	DjangoWorkedWith1	1,18	0,6
jQueryWorkedWith1	0,86	0,8	SOParticipateFewTimesPer- Month	1,18	3,3
OtherFrameworkWorkedWith1	1,02	68,5	SOParticipateFewTimesPer- Week	1,29	0,0
AngularDesire1	0,96	35,4	SOParticipateMultipleTimes- PerD	1,15	39,7
DjangoDesire1	0,84	5,9	NMDS1	8,79	1,9
jQueryDesire1	0,97	38,2	Age25 to 34 yrs:NMDS1	0,79	93,3
OtherFrameworkDesire1	1,06	31,6	Age35 to 44 yrs:NMDS1	0,25	15,7
OpSysMac1	1,34	0,0	Age45 to 54 yrs:NMDS1	0,10	2,2
SOParticipateUnknown	1,10	18,6	Age55 to 64 yrs:NMDS1	0,15	34,6
SOParticipateLessThanOnce- PerMo	1,14	2,8	AgeUnknown:NMDS1	0,15	8,0

Megjegyzés: a félkövérrel szedett változók legalább 5%-on szignifikánsak.

Forrás: saját szerkesztés.

Platformok tekintetében sem a konkrét platformok változói, sem az ismert és megismerni kívánt platformok száma nem rendelkeztek szignifikáns marginális hatással a szokásos szignifikanciaszinteken. Az adatbáziskezeléssel kapcsolatos ismeretek esetén az Elasticsearch-csel kapcsolatos tapasztalat bizonyult egyedül szignifikánsnak a szokásos szinteken. Egy olyan fejlesztő, aki munkája során dolgozott már az Elasticsearch-csel, átlagosan kb. 22%-kal keres többet azoknál, akik még nem, minden egyéb tényező változatlansága mellett. Az SQL-alapú adatbázistechnológiák változója vélhetően hasonló okból nem lett szignifikáns, mint a programnyelvek: a legtöbb munkahelyen nagy valószínűséggel már eleve bemeneti követelmény az SQL-ismeret.

Az ismert keretrendszereket tekintve már több dummyváltozó is szignifikánsnak mutatkozott. Az Angularra, a Djangora és a jQueryre vonatkozó munkatapasztalat szignifikáns hatást gyakorol a bérezésre (ezek közül az első kettő pozitív, míg az utolsó negatív előjelű együtthatóval). Ugyanezen technológiák elsajátítása iránti igény csak a Django esetében volt szignifikáns, viszont a konzisztencia miatt a másik kettő keretrendszer „*Desire*” dummyváltozói is bent maradtak a modellben. Modellünk alapján ezen keretrendszerrel dolgozó fejlesztők minden más tényező változatlanságát feltételezve várhatóan kb. 11%-kal keresnek kevesebbet más keretrendszert alkalmazó fejlesztőknél. A jQuery ismeretének bércsökkenő hatását az magyarázhatja, hogy ez már egy alapvetően elavultabbnak tekinthető technológia. Ezt a technológiát még 2006-ban fejlesztették ki, azzal a céllal, hogy a különböző weboldalak közti korábban fennálló kompatibilitási problémákat áthidalja. Ennek oka az volt, hogy egyes JavaScript-motorok más-hogyan működtek például az Internet Explorerben vagy a Safari-ban. Ugyanakkor a JavaScript-könyvtárak fejlesztésével és a webböngészők egységesítésével ma már kevésbé van szükség egy jQuery-hez hasonló áthidaló megoldásra.

Az operációs rendszerek közül egyedül a macOS-re vonatkozó változó volt szignifikáns minden szokásos szignifikanciaszinten, így a Windows és a Linux változóit elhagytuk. Összességében a Macet használó fejlesztők bruttó jövedelme minden más tényező változatlansága mellett átlagosan kb. 32%-kal volt több azokénál, akik nem Macet használnak. Akik Macen dolgoznak szoftverfejlesztőként, feltehetően azért teszik ezt, mert olyan szoftvert vagy alkalmazást fejlesztenek, amely macOS-szel vagy iOS-szel is kompatibilis, és ez platformspecifikus ismereteket igényel (*MacSources, 2022*). Ugyanakkor fontos megemlíteni a kérdőív mint torzítóelem fontosságát. Bár a kérdésekben külön volt választva az otthoni és a munkahelyi környezetben használt operációs rendszer, előfordulhat, hogy voltak, akik az otthonit jelölték munkahelyinek is.

Pozíciók közül egyedül az adatbázisrendszer-adminisztrátori állást van értelme megkülönböztetni, ugyanis az ilyen munkakört betöltő fejlesztők lényegesen kevesebbet keresnek a többiekénél. Az eredmény olyan szempontból nem meglepő, hogy feltehetőleg egy ilyen állás betöltéséhez kevesebb programozási vagy

egyéb speciális technológiai ismeret szükséges. Ugyanakkor a kérdőív reprezentativitása is kérdéses ebben az esetben. A 3. fejezetben láttuk, hogy a kitöltők között a webfejlesztők a felülreprezentáltak, és a rendszeradminisztrátorok csupán az egyetlen, náluk átlagban kevesebbet kereső csoport, akik a kérdőív kitöltői között kellő számosságban reprezentáltak.

A munkahely méretét (létszám) tekintve a referenciakategória a több mint 10 ezer főt számláló multinacionális cégek voltak. A végső modellben az ismeretlen, illetve kisebb mint 10 és a 10–100 fő közötti vállalatok változójának regressziós paraméterei adódtak szignifikánsnak minden szokásos szinten. Az eredmények szerint a kevesebb mint 10 főt számláló (véltetően technológiai startup) cégek fejlesztői kb. 22%-kal keresnek kevesebbet a több mint 10 ezer embert foglalkoztató multinacionális cégek dolgozóinál, míg a 10–100 munkavállalót számláló kkv-kban dolgozó fejlesztők kb. 18%-kal kapnak kevesebb bért, mint a referenciakategóriában dolgozók. Előfordulhat olyan kitöltési hiba, hogy egyes figyelmetlen kitöltők a juttatásaikat, bónuszaikat is beleszámolták a havi bérükbe (a kérdőívben csak a bruttó bért kellett megadni). További indok lehet még, hogy az életciklusuk elején járó kisebb startupok vagy kisvállalkozások vevőköre még nem kiforrott, ezért az elkönyvelt bevételek még nagymértékben ingadoznak és bizonytalanabbak, mint egy érett vállalat esetében. Ez a nagyobb fokú bizonytalanság a beosztottak bérére is negatív hatással lehet.

A munkaviszony esetében a teljes állásban (heti 40 óra) foglalkoztatottak voltak a referenciakategória. Nem meglepő módon a részmunkaidőben dolgozók minden további tényező változatlansága mellett átlagosan kb. 30%-kal kevesebb bért kapnak a teljes időben foglalkoztatottaknál. Érdekes eredmény viszont, hogy a szerződéses munkaviszonyban dolgozók ugyanezen feltételezések mellett várhatóan 29%-kal magasabb jövedelemre számíthatnak egy teljes idejű munkavállalónál. A szerződéses keretek közt dolgozó kitöltők zöme valószínűleg a kata (kisadózó vállalkozók tételes adója)-konstrukcióban dolgozó egyéni vállalkozónak bejelentett szoftverfejlesztők voltak. A kata adónem ezen szakemberek számára nagyon kedvező feltételeket biztosított, a törvényt azonban 2022 júliusában módosították, így a 2022. évi, a kisadózó vállalkozók tételes adójáról szóló XIII. törvény értelmében a vállalatoknak való számlázás lehetősége gyakorlatilag megszűnt. Ez a 2022. évi Stack Overflow-kérdőívet még nem befolyásolta, mert a mintavétel az év első felében történt (*Stack Overflow, 2022*).

A Stack Overflow-használati szokásokra vonatkozó ismérvek közül egyedül a fórumbeszélgetések gyakoriságára vonatkozó dummyváltozók maradtak a modellben. Az ismérvek szinte minden ismérvkimenetele szignifikáns volt, mindegyik pozitív együttthatóval. Ezek közül is leginkább az, amelyik arra vonatkozott, hogy az illető heti néhány alkalommal kommentel, vagy tesz fel kérdést. Feltehetően ez a változó egyfajta indirekt kontrollváltozóként funkcionálhat a mo-

dellben a fejlesztő személyiségére. Feltehetően az, aki rendszeresen használja kérdésfeltevésre a Stack Overflow-t, hajlamosabb a kollaborációra, és nyitott arra, hogy a mások által javasolt, adott esetben optimálisabban működő kódokat vagy programcsomagokat kipróbálja, ezáltal pedig a munkában nyújtott teljesítménye is javul.

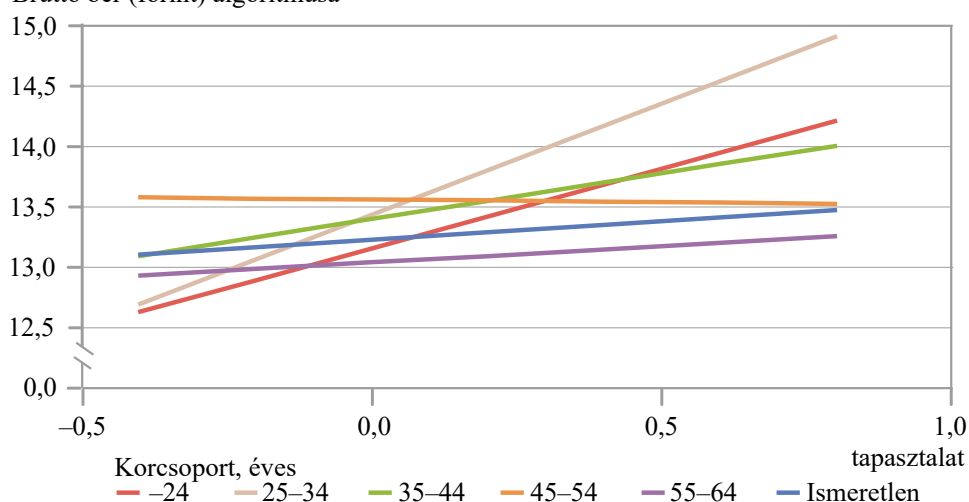
A végső modellben szereplő interakciós hatás alapján a magyar szoftverfejlesztők fizetései kor szerint nézve a tapasztalat függvényében eltérő mértékben nőnek vagy csökkennek. A marginális hatásokat a 3. ábrán szemléltjük.

3. ábra

Az életkor és a tapasztalat marginális hatása

Marginal effects of age and work experience

Bruttó bér (forint) algoritmus



Forrás: saját szerkesztés.

Az ábrán látható, hogy a kor és a tapasztalat nem áll teljesen lineáris kapcsolatban a bérral: a görbék meredeksége korcsoportonként jelentősen eltér, illetve összességében teljesen eltérő tengelymetszetekből induló egyeneseket látunk. A 24 év alatti, valamint a 35–44 éves korcsoportot összehasonlítva észrevehetjük, hogy az utóbbi bérezési görbéje alapvetően magasabb tengelymetszettel rendelkezik, de laposabb meredekséggel, míg a fiatalabb korosztályé alacsonyabb metszettel, ám meredekebb függvényel, egy adott tapasztalati szint felett leahagyva a 35–44 éveseket. Ennek az a jelentése, hogy a referenciacsoportot képező 24 év alatti fejlesztőkhöz viszonyítva a 35–44 éves korosztály jövedelme a tapasztalat szerint megfigyelve laposabb meredekséggel nő. A magyarázat erre az eredményre az lehet, hogy az évről évre változó technológia következtében az idő előrehaladtával a fejlesztők egy-egy, egyre inkább elavuló technológiában lesznek tapasztaltabbak (feltételezve, hogy nem képzik tovább magukat más

technológiákban). Az eredmény megerősíti a bevezetésben felállított hipotézist, amely szerint az ilyen korú vagy ennél idősebb fejlesztők kevesebb operatív fejlesztési és több vezetési, koordinálási feladatot látnak el.

5. A diploma hatásainak vizsgálata Blinder–Oaxaca-dekompozícióval

A diploma hatásainak szétbontásához bevezettünk egy technikai változót, amelynek értéke 1 volt, amennyiben az illető rendelkezett Bsc-, Msc- vagy PhD-fokozatú diplomával. Minden egyéb esetben (hiányzó érték esetén is) a változó a 0 értéket vette fel. A diplomával nem rendelkezők részcsoportja lett a referenciakategória. A Blinder–Oaxaca-dekompozíció eredményei összesítve a 5. táblázatban találhatóak.

5. táblázat

A diplomával rendelkezők és nem rendelkezők dekompozíciós vizsgálatának eredményei

Results from Blinder–Oaxaca-decomposition for groups with and without degrees

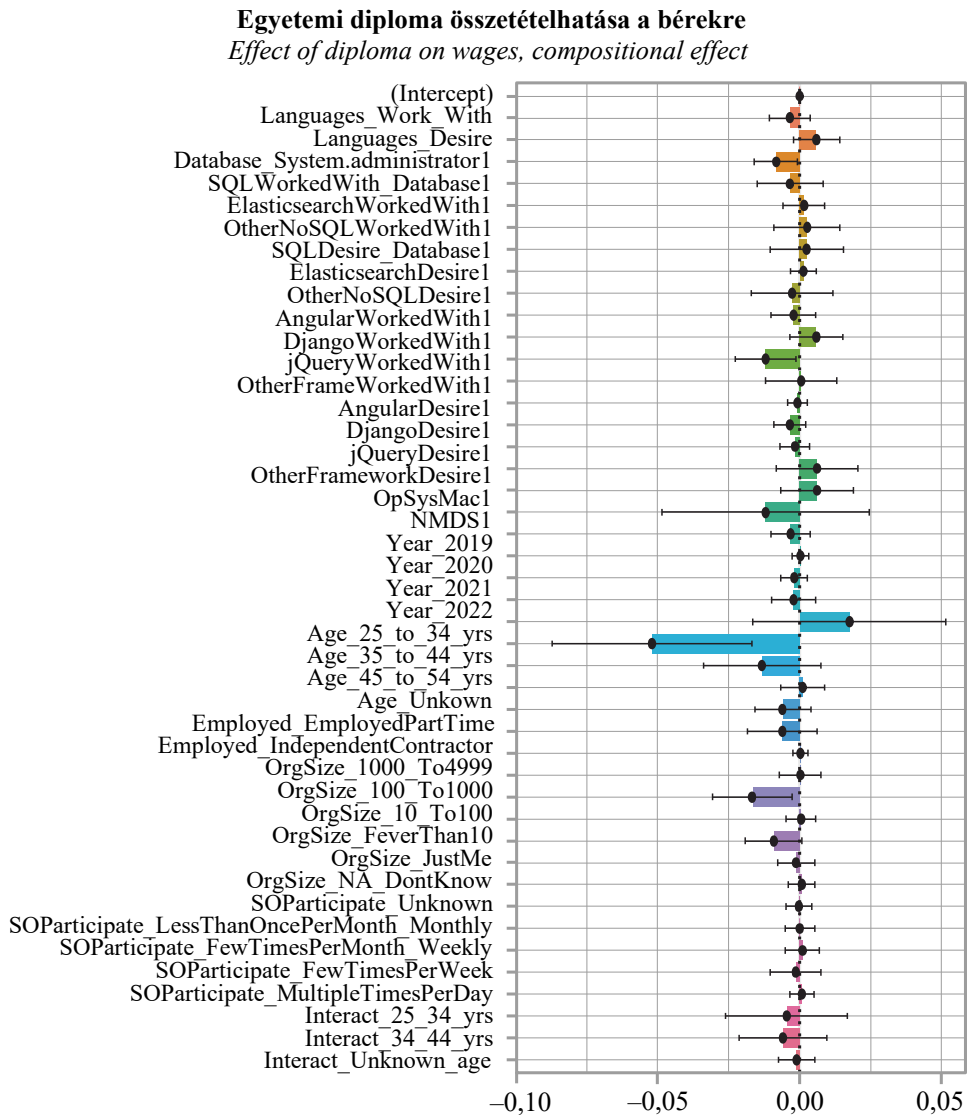
Részcsoportok átlagának bruttó bérkülönbsége, forint	Szignifikáns összetételhatások	Szignifikáns összetételhatások mértéke a bérre, %	Szignifikáns részhatások	Szignifikáns részhatások mértéke a bérre, %
~ -157 500	Dolgozott jQueryvel?	~ -1,14	SQL-lel szeretne a későbbiekben foglalkozni?	~ +22,18
	35–44 éves?	~ -5,15	Egyéb NoSQL nyelvekkel szeretne-e a későbbiekben foglalkozni?	~ -26
	10–100 fős cégnél dolgozik?	~ -1,6		

Forrás: saját szerkesztés.

Az 5. táblázat a dekompozíció eredményeit mindig a referenciacsoporthoz viszonyítva mutatja be. Jelen esetben például a bruttó bérkülönbség jelentése, hogy a diploma nélküli kitöltők részcsoportjának átlagbére kb. bruttó 157 500 forinttal volt kevesebb a diplomával rendelkezők csoportátlagánál. A különbség összetétel- és részhatáskülönbségekre bomlik, előbbi a magyarázóváltozók által megmagyarázott, utóbbi a megmagyarázatlan/nem megfigyelt hatások következtében bekövetkező differencia.

A 4. ábrán látható, hogy számos változó nem szignifikáns összetételhatással bír 5%-os szignifikanciaszint mellett, mivel 95%-os konfidencia-intervallumaik a 0-át is magukba foglalják. A szignifikánsak közül a következőket állapíthatjuk meg: a jQuery-vel dolgozók, a 35–44 éves korosztály, a 10–100 fős vállalatoknál dolgozók átlagosan kevesebbet keresnek, és a nem diplomások körében többen vannak, ezzel csökkentve a diploma nélküliek csoportátlagát.

4. ábra



Forrás: saját szerkesztés.

A jQuery elavultságát a 4. fejezetben tisztáztuk. Az 5. táblázatban összesített eredmények alapján az eleve kevesebbet kereső, jQueryvel dolgozók a diploma nélküliek körében többen fordulnak elő, ezzel átlagosan 1,14%-kal rontva az egyetemi képzést nem végzett kitöltők csoportjának átlagjövedelmét.

Az 5. táblázat szerint a diploma nélküliek körében jellemzőbben fordulnak elő a 4. fejezet alapján kevesebb jövedelemmel rendelkező 35–44 évesek, így a felsőoktatást otthagyo vagy nem végzett részcsoport átlagbére kb. 5,15%-kal romlik. Ez más aspektusból is megerősítette a korábban tett megállapítást, miszerint minden egyéb tényező változatlansága mellett – feltételezve, hogy a szoftverfejlesztők nem képzik magukat tovább – egy bizonyos pont után átlagosan kevesebbet kereshetnek, mint a fiatalabb, frissebb technológiákat ismerő korosztály.

A 10–100 főt számláló vállalatok a kkv-szektorba tartoznak, ahol a cég költségkeretei sokszor szűkebbek, mint a nagyvállalatoké, ezért feltehetően az alacsonyabb átlagbér ebből a tényezőtől is adódik. Azon fejlesztők, akik ilyen vállalatoknál dolgoznak, jellemzően kisebb jövedelemmel rendelkeznek, és a diplomával nem rendelkezők körében többen fordulnak elő, ezáltal kb. 1,6%-kal csökkentve ezen részcsoport bruttó átlagjövedelmét.

A részhatáskülönbségek megtalálhatók az 5. táblázatban. Ezen hatások között is – hasonlóan az összetételhatások esetéhez – azok tekinthetők 5%-os szignifikanciaszint mellett szignifikánsnak, amelyek 95%-os konfidencia-intervalluma nem tartalmazta a 0-át. Az eredmények szerint SQL-szabványon alapuló relációs sémás adatbázis-kezelő programok használatát elsajátítani akarók valamilyen nem megfigyelt hatás következtében többen keresnek a diplomások körében a diploma nélkülieknél, ezáltal a diplomával rendelkező részcsoport átlagbérét kb. 22%-kal növelik. Ugyanakkor az egyéb NoSQL-nyelveket megismerni vágyók átlagfizetése szintén meg nem figyelt hatások miatt alacsonyabb a diplomával nem rendelkezők körében a diplomásokénál, emiatt ennek a részcsoportnak a bruttó átlagbére kb. 26%-kal kevesebb a diplomával rendelkező kitöltők csoportjának átlagánál.

Az első állításra az egyetemi képzések struktúrájából kiindulva adható lehetséges választ. Általánosságban elmondható, hogy az informatikai képzést biztosító szakokon kötelező tárgyként szerepel az adatbázisrendszerek ismerete, aminek tananyaga elsősorban a hagyományos relációs sémák logikája és az ilyen környezetben történő lekérdezésekhez használt SQL-nyelv, míg a NoSQL-módszertanok kevésbé szerepelnek az informatikai képzések tantárgyi programjaiban (*Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki Kar, 2015; Eötvös Loránd Tudományegyetem Informatikai Kar, 2018*). Aki tehát egyetemi szinten végez el egy informatikai képzést, az várhatóan több jártasságot és rálátást szerez, illetve – akár a kötelező szakmai gyakorlat idején – olyan cégeknél helyezkedik el, ahol esetlegesen elvárt hasonló kompetencia. Érthető,

hogya ha valaki ilyen jellegű állással rendelkezik, a továbbiakban is szeretné tudását bővíteni a témában.

Az eredményeket az is alátámasztja, hogy Magyarországon valószínűleg még nem nagy a kereslet a NoSQL-technológiákat ismerő szakemberekre a hagyományos SQL-alapú adatbázis-kezelő kompetenciához képest. Ennek oka részben az ország technológiai szempontból perifériás helyzete, részben pedig a NoSQL mint megoldás sajátossága, hogy nem feltétlenül szükséges minden helyzetben alkalmazni, amikor a probléma egy egyszerűbb relációs sémás adatbázissal is megoldható. A végső modellben (4. fejezet) kitüntetett Elasticsearch például JSON-formátumhoz hasonló, nem strukturált adatok beágyazott lekérdezésére alkalmas keresőmotor. Ez a technológia alapvetően olyan vállalatok számára megfelelő, amelyeknél szenzoros adatfeldolgozás (is) folyik, ezért nagy mennyiségű (több ezer sor/perc) strukturálatlan adat keletkezik, amelyeket üzleti, üzletfolytonossági szempontból elsősorban a monitoring és az anomáliakeresés céljából érdemes ábrázolni (Meier–Kaufmann, 2019). A magyar piacon működő vállalatok inkább hagyományos adattárház-architektúrát használnak, ami az adatokat relációs sémában tárolja és amelyhez standard konnektorokon keresztül számos elemzéshez, riportinghoz stb. használható eszköz csatlakoztatható. Bár nem reprezentatív adatgyűjtés, a jelenséget alátámasztja, hogy Magyarország legnagyobb, álláshirdetéseket kezelő weboldalán, a profession.hu-n 2023. augusztus 7-én 15 állásajánlat tartalmazza a *NoSQL* kulcsszót, míg az SQL-kulcsszót 248. Egy további, ebből következő indok lehet még, hogy mivel a NoSQL-megoldások alkalmasabbak speciálisabb, nem strukturált adatok kezelésére, bevezetésük és üzemeltetésük pénzt és erőforrást tekintve is költségesebb lehet, így a kevésbé tőkeerős magyar kkv-szektorban eleve nem merülnek fel technológiaként (Vajda–Magda, 2020).

Összességében a dekompozíciós elemzésünk eredményei igazolják, hogy Magyarországon továbbra is mérvadó, ha egy programozó legalább alapszakos diplomával rendelkezik. Diploma birtokában átlagosan 157 500 forinttal magasabb havi bruttó bérre számíthat egy magyar szoftverfejlesztő. A különbség az összetételhatás-oldalon abban keresendő, hogy a diplomások között nagyobb a súlya a nagy, multinacionális vállalatoknál dolgozóknak, és a fiatalabb, 35 év alatti korcsoportoknak. Eredményeink alapján a diploma hozzáadott értéke a hagyományos, SQL-alapú technológiák alaposabb elsajátításának köszönhető, amik a magyar vállalati rendszerek adatkezelésében a mai napig meghatározók. A NoSQ-technológiák pedig a magyar piacon még viszonylag ritka esetekben megfelelő megoldások, így a jelek szerint a piac nem is díjazza, ha egy fejlesztő ilyen irányban képezné magát tovább.

6. Következtetések

Jelen tanulmányban a Stack Overflow éves fejlesztői kérdőíveinek 2018 és 2022 közötti adatait felhasználva, ökonometriai modellezés segítségével kíséreltük meg azonosítani a magyar szoftverfejlesztők havi bruttó béreit alakító tényezőket. A modellezés során egy 70 változót tartalmazó modellt tudunk felállítani a keresett magyarázó tényezők azonosítására. A modell segítségével az egyetemi diploma értékét is számszerűsíteni tudtuk a magyar szoftverfejlesztők munkaerőpiacán, Blinder–Oaxaca-dekompozíció segítségével.

A végső, diagnosztikailag megfelelőnek tekinthető modellünk alapján a következő marginális hatások azonosíthatók:

- Az egyik leginkább bérnövelő hatású tényező a magyar szoftverfejlesztők körében a kódolási tapasztalat, ami szoros összefüggésben áll a korrall és az iskolai végzettséggel is. A korrall kombinálva viszont a tapasztalat növekedésének adott esetben bércsökkenő hatása is lehet. Ez abból adódhat, hogy az adott szoftverfejlesztő egy elavult technológiában lesz jártasabb (például a jQuery, amelynek ismerete átlagosan rontja a bért), így bérének görbéje 35–44 éves kor felett olyan mértékben tud ellapulni, hogy minden egyéb tényező változatlansága mellett például a 18 év alatti és a 18–24 év közötti korcsoport átlagbér-növekedésének üteme meg tudja azt előzni.
- Összességében egyedül a rendszeradminisztrátor, rendszerüzemeltető munkakört érdemes megkülönböztetni, ezen részcsoporthoz átlagbére ugyanis szignifikánsan kevesebb a többi (adatbázisunkban elsősorban a webfejlesztői) munkakörtől. Ennek vélhetően olyan okai lehetnek, hogy egy ilyen pozíció betöltése kevesebb speciális tudást igényel technológiai szempontból.
- Megállapítható, hogy bár megjelentek szakképesítést biztosító, nem felsőfokú kurzusok, Magyarországon továbbra is mérvadó, ha egy szoftverfejlesztő legalább alapszakos diplomával rendelkezik és nem hagyta ott az egyetemet. Az egyetemet be nem fejező fejlesztők átlagos bére – feltéve, hogy minden további tényező változatlan – várhatóan 7%-kal kisebb egy alapszakot végzett fejlesztőkhöz képest. Ugyanakkor a leíró statisztikai elemzések alapján felfedezhető az a tendencia, miszerint a mesterképzést egyre kevesebben végzik el a területen. Dekompozíciós elemzésünk eredményei alapján elmondható, hogy az egyetemi diplomával rendelkezők elsősorban a magyar piacon továbbra is meghatározó klasszikus adatbázis-technológiák területén szerzett tudásuknak köszönhetően jutnak magasabb bérekhez. A diplomával rendelkezők jellemzően nagyobb arányban dolgoznak eleve jobban fizető nagyvállalatoknál és jellemzően a fejlesztők között bérezés szempontjából előnyösebb korcsoportokba is tartoznak.

- Külön–külön vizsgálva szinte egy programnyelvnek sem volt saját bérnövelő hatása, az viszont már szignifikáns volt, hogy hány programnyelvet ismer vagy szeretne tanulni az illető. Hazánkban a magas informatikai munkaerőhiány miatt elképzelhető, hogy a munkáltatók számára az is elegendő, ha a leendő munkavállaló akár más nyelvek terén szerzett már kódolási tapasztalatokat, és rendelkezik az alapvető algoritmikus gondolkodással, tehát könnyen átképezhető más nyelven való programozásra is. A platformok ismerete szintén nem volt szignifikáns, viszont ott még az sem, hogy a kitöltő hányat ismer vagy hányat szeretne megismerni. Nagy valószínűséggel elmondható, hogy amennyiben a leendő munkavállaló nem találkozott még egy adott platformmal, annak elsajátítása alapvető informatikai kompetenciákkal nem jelent nagy erőforrást a munka közben sem.

Fontos megjegyezni, hogy a levont következtetések nem teljesen tekinthetők reprezentatívnak a teljes magyar szoftverfejlesztői munkaerőpiacra nézve, mivel az adatfelvétel a Stack Overflow felhasználóira korlátozódik. Ez a tény önmagában nem lenne még nagy korlátozó tényező, hiszen a fórum felhasználói bázisa Magyarországon is elég széles, ám a kérdőív kitöltése önkéntes, így eredményeink általánosíthatósága egy szűkebb körre korlátozódik, mint a teljes magyar szoftverfejlesztői munkaerőpiac. A modell alapján azonosított szokásos szignifikanciaszinteken szignifikánsnak adódó marginális hatások elsősorban a 25–44 év közötti férfi webfejlesztők sokaságára tekinthető relevánsnak. Bár ez korántsem teljes körű munkaerőpiaci lefedettség, a magyar szoftverfejlesztők egy jelentős hányadának bérezését alakító tényezőket sikerült azonosítani modellünkkel. Mivel a kitöltés önbevallásos alapon történt, fennáll annak veszélye, hogy egyes kitöltők nem valós információkat adtak meg magukról. Ez történhetett szándékos és véletlen módon is, például, ha félreértették a kérdést.

Tanulmányunk eredményeit és a Stack Overflow-kérdőív adatait felhasználva tervezzük a magyar fejlesztők esetén azonosított bérezést alakító tényezőket összevetni a szomszédos országok és Nyugat-Európa szoftverfejlesztői esetében azonosítható tényezőkkal. Bár az eredményeink csupán a 25–44 év közötti férfi webfejlesztők sokaságára tekinthetők relevánsnak, felhívjuk a figyelmet arra, hogy a Stack Overflow éves fejlesztői kérdőívéhez hasonló struktúrájú szélesebb körű adatgyűjtések segítségével még teljes képet kaphatunk a magyar szoftverfejlesztők bérezési és munkaerőpiaci viszonyairól jelen tanulmány módszertanát és eredményeit alkalmazva. Hasonló eredmények a hazai K+F-szektor legdinamikusabban fejlődő területének fejlesztéséhez szolgáltathatnak kulcsfontosságú információkat.

Irodalom

2022. évi XIII. törvény a kisadózó vállalkozók tételes adójáról. (2022.07.18).
<https://net.jogtar.hu/jogszabaly?docid=a2200013.tv> (letöltés dátuma: 2023. április)
- 5/2020. (I. 31.) Kormányrendelet a Nemzeti Alaptanterv kiadásáról, bevezetéséről és alkalmazásáról szóló 110/2012. (VI. 4.) Kormányrendelet módosításáról. (2020.01.31.).
<https://net.jogtar.hu/jogszabaly?docid=a1200110.kor> (letöltés dátuma: 2023. április)
- Arbidāne, S. – Silickis, M. (2018): *How Latvian Developers Learn New Technologies. A Conclusive Review and Analysis of Survey Data*. Students International Scientific and Practical Conference. <http://dx.doi.org/10.17770/het2018.22.3644>
- Árgilán, V. S. – Kelemen, A. (2016): *Az elvárt digitális kompetenciák a XXI. században és az informatika oktatás gyakorlata a közoktatásban Magyarországon*. XVII. ENELKO – XXVI. SzámOkt Nemzetközi Energetikai-elektrotechnikai és Számítástechnikai Konferencia. Október 6–9. Kolozsvár. <http://publicatio.bibl.u-szeged.hu/id/eprint/14770>
- Blinder, A. S. (1973): Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*. Vol. 8. No. 4. pp. 436–455. <https://doi.org/10.2307/144855>
- Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki Kar (2015): *Adatbázisok. Tantárgyi adatlap*. Budapesti Műszaki és Gazdaságtudományi Egyetem, Budapest.
<https://portal.vik.bme.hu/kepzes/targyak/VITMA311/>
- Cotton, J. (1988): On the Decomposition of Wage Differentials. *The Review of Economics and Statistics*. Vol. 70. No. 2. pp. 236–243. <https://doi.org/10.2307/1928307>
- Eötvös Loránd Tudományegyetem Informatikai Kar (2018): *Adatbázisok I. Tantárgyi adatlap*. Eötvös Loránd Tudományegyetem, Budapest.
<https://www.inf.elte.hu/dstore/document/1051/Adatb%C3%A1zisok%20I..pdf>
- Felvi.hu. (2023): *2023 szeptemberében induló képzések*.
https://www.felvi.hu/felveteli/ponthatarok_statistikak/friss_statistikak!/FrissStatistikak/inde_x.php/friss_statistikak/karonkent?page=1&orderBy=sta_kar_kod&sortBy=ASC
- Gower, J. C. (1971): A General Coefficient of Similarity and Some of Its Properties. *Biometrics*. Vol. 27. No. 4. pp. 857–971. <https://doi.org/10.2307/2528823>
- Hays Hungary Salary Guide (2023) <https://www.hays.hu/salary-guide/trendek#salary-guide>
- James, G. – Witten, D. – Hastie, T. – Tibshirani, R. (2021): *An Introduction to Statistical Learning with Applications in R*. Springer. New York.
- Jann, B. (2008): The Blinder – Oaxaca-decomposition for Linear Regression Models. *The Stata Journal: Promoting Communications on Statistics and Stata*. Vol. 8. No. 4. pp. 453–479.
<https://doi.org/10.1177/1536867X0800800401>
- Janssen, M. – Joha, A. (2006): Motives for establishing shared service centers in public administrations. *International journal of information management*. Vol. 26. No. 2. pp. 102–115.
<https://doi.org/10.1016/j.ijinfomgt.2005.11.006>
- Kruskal, J. B. (1964): Nonmetric multidimensional scaling: A numerical method. *Psychometrika*. Vol. 29. No. 2. pp. 115–129. <https://doi.org/10.1007/BF02289694>
- Központi Statisztikai Hivatal (KSH) (2023a): 20.1.1.9. *A foglalkoztatottak száma nemzetgazdasági ágak, ágazatok szerint, nemenként – TEÁOR'08 [ezer fő]*.
https://www.ksh.hu/stadat_files/mun/hu/mun0009.html
- Központi Statisztikai Hivatal (KSH) (2023b): 20.1.1.72. *Üres álláshelyek száma és aránya nemzetgazdasági áganként*. https://www.ksh.hu/stadat_files/mun/hu/mun0072.html

- Központi Statisztikai Hivatal (KSH) (2023c): 9.1.1.18. *Magyarországon működő szolgáltató központok száma, teljes- és nem teljes munkaidőben alkalmazásban állók állományi létszáma, havi bruttó átlagkeresetük alakulása nemzetgazdasági ág szerint.*
https://www.ksh.hu/stadat_files/gsz/hu/gsz0064.html
- Lavery, M. R. – Acharya, P. – Sivo, S. A. – Xu, L. (2019): Number of predictors and multicollinearity: What are their effects on error and bias in regression? *Communications in Statistics – Simulation and Computation*. Vol. 48. No. 1. pp. 27–38.
<https://doi.org/10.1080/03610918.2017.1371750>
- MacSources (2022): *Why Is MacOS Better For Graphic Designing?*
<https://macsources.com/why-is-macos-better-for-graphic-designing/>
- Mayer-Schönberger, V. – Cukier, K. (2014): *Big Data: Foradalmi módszer, amely megváltoztatja munkánkat, gondolkodásunkat és egész életünket*. HVG Kiadó. Budapest.
- Meier, A. – Kaufmann, M. (2019): *SQL & NoSQL databases*. Springer. Berlin, Heidelberg.
- Nemzeti Szakképzési és Felnőttképzési Hivatal (2023): *2.sz. melléklet: képzési jegyzék GINOP-6.2.8-VEKOP-20-2020-00001 kiemelt projekt keretében 2023. februárban indítani tervezett képzések jegyzéke.*
https://ujratervezes.nive.hu/api/static/doc/kepzesi_jegyzek_3.0%20halad%C3%B3.pdf
- Neumark, D. (1988): Employers' Discriminatory Behavior and the Estimation of Wage Discrimination. *The Journal of Human Resources*. Vol. 23. No. 3. pp. 279–295.
<https://doi.org/10.2307/145830>
- Oaxaca, R. (1973): Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*. Vol. 14. No. 3. pp. 693–709. <https://doi.org/10.2307/2525981>
- Pranggono, B. – Arabo, A. (2021): COVID-19 pandemic cybersecurity issues. *Internet Technology Letters*. Vol. 4. No. 2. pp. e247. <https://doi.org/10.1002/itl2.247>
- Silveira, K. K. – Musse, S. – Manssour, I., Vieira, R. – Prikładnicki, R. (2019): *Reinforcing Diversity Company Policies: Insights from StackOverflow Developers Survey*. School of Technology. Pontifícia Universidade Católica do Rio Grande do Sul.
- Stack Overflow (n.d.): *Stack Overflow Annual Developer Survey*.
<https://insights.stackoverflow.com/survey> (letöltés dátuma: 2022. október)
- Stack Overflow (2022): *Stack Overflow Developer Survey 2022*.
<https://survey.stackoverflow.co/2022/> (letöltés dátuma: 2022. október)
- Stock, J. H. – Watson, M. W. (2020): *Introduction to econometrics*. Pearson. Boston.
- Stroustrup, B. (2010): What should we teach new software developers? Why? *Communications of the ACM*. Vol. 53. No. 1. pp. 40–42. <https://doi.org/10.1145/1629175.1629192>
- Takács O. (2021): Nemek közötti bérkülönbségek Magyarországon: a véletlenerdő- és az OLS-becslésen alapuló Blinder–Oaxaca-dekompozíció eredményeinek összehasonlítása. *Statisztikai Szemle*. 99. évf. 1. sz. 5–45. o. <https://doi.org/10.20311/stat2021.1.hu0005>
- Vajda A. – Magda R. (2020): A kkv-k szerepe a versenyképességben, a magyar kkv-k összehasonlítása az EU vállalkozásaival. *Logisztikai trendek és legjobb gyakorlatok kiadvány*. 6. évf. 2. sz. 50–54. o. <https://www.doi.org/10.21405/logtrend.2020.6.2.50>
- Van Buuren, S. (2018): *Flexible Imputation of Missing Data*. CRC Press. Boca Raton.
<https://stefvanbuuren.name/fimd/>
- Wooldridge, J. M. (2016): *Introductory econometrics: A modern approach*. Cengage Learning. Boston.