

Szakdolgozat

Hoang Péter

2015

**Budapesti Műszaki és Gazdaságtudományi Egyetm**

**Gazdaság- és Társadalomtudományi Kar**

**Nemzetközi gazdálkodás alap szak**

A közbringarendszerek bérlésének és felhasználóik elégedettségének változói

Középpontban a MOL Bubi-val

**Készítette:** Hoang Péter

Témavezető: Dr. Danyi Pál

Budapest

2015

Tartalmi kivonat

Mindenek előtt köszönetet szeretnék nyilvánítani a Budapesti Közlekedési Központ és a Közbringa Kft. munkatársainak, amiért betekintést nyerhettem Magyarország legnagyobb közösségi biciklikölcsönző rendszerének működésébe, és rendelkezésemre bocsátották a szükséges információkat a MOL Bubi-val kapcsolatban.

A dolgozat célja, hogy feltérképezze azokat az emberi tényezőket és a külső hatásokat, amelyek befolyással vannak a MOL Bubi felhasználóinak elégedettségére és a biciklibérlések számára.

Az elemzés megelőző részt az adatbányászatot mint önálló diszciplínát taglalja, annak fontosabb mechanizmusait, gyakorlatban való felhasználását és a biciklimegosztó rendszerek, és az adatbányászat kapcsolatát tekinti át. Mindezek mellett kitér a nemzetközi rendszerekre is és azok legfontosabb színtereire.

A fő részben a primer és a szekunder kutatásomat mutatom be, végezetül pedig ezek kimeneteleit.

Az eredmények magukért beszélnek, de reményeim szerint sikerült egy olyan dolgozatot létrehozni, amely alapot adhat későbbi kutatások elvégzésére.

Tartalomjegyzék

1. Bevezetés 5

2. Szakirodalmi áttekintés 8

2.1. Nemzetközi színtereken 9

2.2. Az adatbányászati áttekintés 10

2.3. CRISP-DM modell 11

2.4. Feladattípusok 13

2.4.1. Asszociációs szabályok 13

2.4.2. Osztályozás 13

2.4.3. Regresszió 14

2.4.4. Klasszterezés 14

2.4.5. Eltéréselemzés 14

2.1. Adatbányászat és a biciklimegosztó rendszerek kapcsolata 15

3. Kutatások 16

3.1. A primer kutatás 16

3.1.1. A feldolgozás módszertana 16

3.2. A szekunder kutatás 22

3.2.1. Az eredményváltozó meghatározása 22

3.2.2. A magyarázóváltozók kiválasztása 23

3.2.3. Adattisztítás, konvertálás, hiánypótlás 24

3.2.4. Korrelációs mátrix vizsgálata 24

3.2.5. Modellezés 28

3.2.6. Asszociációs vizsgálat 31

3.3. Prediktív előrejelzés 32

4. Összegzés 33

5. Irodalomjegyzék 35

6. Mellékletek 37

1. Bevezetés

A bicikli szerepe a városi közlekedésben nagy figyelmet kapott az elmúlt évek során, leginkább vagy legalábbis egy részben annak köszönhetően, hogy kezdi visszanyerni egykori népszerűségét mint alternatív közlekedési forma. A hatvanas években történt az első kezdeményezés biciklimegosztó, illetve nyilvános kerékpár programokra Európában, azonban működőképes formában csak az ezredforduló után alakult ki 2005-öt követően az információs technolológia fejlődésének köszönhetően. Napjainkra a kerékpármegosztó- vagy közbringarendszerek már mind az öt kontinensen jelen vannak, beleértve több, mint 700 várost világszerte, közel 800 ezer biciklivel és 37.500 állomással. A fejlődés még nem ért véget, egyre több városban terveznek újabb és újabb hasonló rendszereket. Az alapok lefektetéseként az Obis kézikönyv a következőképp definiálja biciklimegosztó rendszereket:

,,A közösségi kerékpározási rendszer (KRR, angolul: bike sharing system) egy önkiszolgáló, rövidtávú, akár egyirányú utazásra is alkalmas kerékpárkölcsönzési hálózat nyilvános helyeken, többféle célcsoport számára.”1 (OBIS kézikönyv: *A közösségi kerékpározási rendszerek optimalizálása az európai városokban*, 10. oldal)

A hozzáférhetőség és a megfizethetőség nagyban segítették, hogy a rövidtávú kerékpárbérlés azokat is meggyőzze, akik hajlandóak lecserélni autójukat egy kétkerekűre belső égésű motor nélkül. Az ingázók és a külvárosban lakók a parkolóban hagyhatják általa a belvárosi forgalmi dugók és parkolóhelyek hiánya okozta stresszt. A turisták keresztül-kasul szelhetik át a várost, megkímélve magukat a tömegközlekedésre jellemző többszöri átszállástól, a magas taxiköltségektől és a lábfájástól. Nem beszélve arról, hogy az anyatermészetnek is előnyére válik a hétköznapi szmog csökkentése.2

Budapest kerékpározási kultúrája az elmúlt években pozitív irányba változott. Egy dán felmérés szerint 150 városból jelenleg a 13. helyezést érte el. A Mol Bubi bevezetése pedig talán egy új löketet ad a fővárosi bicikliskultúra további növekedésére. Természetesen infrastrukturálisan és a kerékpározási kultúra szempontjából is van még mit fejleszteni, de a tendencia azt mutatja, hogy megvan a szándék a városlakókban, hogy személygépkocsijaikat biciklire cseréljék.3

A Mol Bubi Magyarország első biciklimegosztó rendszere, mely 2014 nyarán kezdte meg a próbatesztelést, és végérvényesen szeptemberben 8-án vált elérhetővé hivatalosan a nagyérdemű számára. Jelenleg több, mint 76 gyűjtőállomáson 1100 biciklije és több, mint 6000 regisztrált felhasználója van.

A Bubi lefedettsége a belvárosra terjed ki, a nagykörúton belül. Legdélibb pontja a Corvin negyed, majd végig a körúton találhatók a dokkolók. Az állomások elhelyekezdése sűrű, így ha nem is találunk biciklit egy állomáson - bár ez nagyon ritkán fordul elő -, a legközelebbi állomás is csak pár percre található. Budán is találhatóak egy vonalban, a Margit-híd budai hiídfőjétől indulva. Itt még rengeteg hely kínálkozik további állomások kiépítésére. Már folynak a tárgyalások a további bővítésekkel kapcsolatban, így 2015 nyarára már több, mint 97 állomással és 1150 biciklivel fog bővülni a Bubi.4

Budapest 2008-ban hozta meg a döntést a fővárosi közösségi kerékpármegosztás elindításáról. Kialakítását azonban többéves előkészületek, tervezés és szervezés előzték meg.  A dokkolóállomások helyének megtervezésére nagy figyelmet fordítottak, hogy a lehető legrészletesebben felmérhessék a legoptimálisabb elhelyezésüket.5

A projekt kedvezményezettje a Budapesti Közlekedési Központ, aminek feladata a teljes folyamat irányítása és levezénylése, a tervezéssel pedig a Parking Kft.-t bízták meg. 2013 augusztusában a BKK által kiírt tendert a MOL Bubi rendszer kialakítására és üzemeltetésére a T-Systems Magyarország Zrt. és a Csepel Kerékpárgyártó és Forgalmazó Zrt. nyerte meg. Utóbbi két vállalat együtt a Közbringa Kft.-t teszi ki, mely magukat a kerékpárokat és az azokhoz szükséges eszközöket gyártja, felel a karbantartásukért, illetve az allokációs folyamatokért. A technológiai feltételeket, a szoftvereket egy német vállalat, a Nextbike biztosítja.6

A kialakítások előtt a Budapesti Közlekedési Központ kutatásainak során fő céljuk az volt, hogy felderítsék, melyek azok a tényezők, amelyek a budapestiek utazási morálját, közlekedési szokásait alakítják, illetve melyek a leggyakoribb úti célok, mik az utazási mód kiválasztási szempontjai, továbbá a kerékpározástól visszatartó erőket is feltérképezték. Ezeket egy közvélemény-kutatással próbálták felderíteni, kérdőíves módszerrel, amelynek eredményeképpen a nagy utazássűrűségű helyszíneket találták a legmegfelelőbb gyűjtőpontoknak.7

A szakdolgozat témája a Magyarországon újonnan létrehozott közbringrarendszer, a MOL Bubi általános használatának feltérképezése: ezen belül annak vizsgálata, hogy milyen tényezők milyen mértékben befolyásolják a bérlések számát. A matematikai alapoktól, a statisztikán át a különböző tudományágakig, ökonometriai és adatbányászati alklamazásokat, eszközöket és módszereket használtunk az elemzés során. Mindezt különböző modellek segítésével próbáljuk feltérképezni, a lehetségek faktorokat, magyarázó tényezőket megtalálni.

A dolgozat írása során nehézséget jelentett a források és az adatok elérhetősége, és azok korlátozott száma is tovább bonyolította a munkát. Ennek magyarázata abban rejlik, hogy a Magyarországon működő közbringarendszerek még gyermekcipőben járnak, és még nem telt el kellő idő ahhoz, hogy több információ álljon rendelkezésünkre. A determinisztikus idősorok nem térképezhetők fel, nincs elegendő adat a szezonalítás és a peridoicitás elemzésére.

Igaz, hogy több külföldi portálon már elég sok kutatás és elemzés elérhető a tárgyalt témával kapcsolatban, de ez Magyarországon még várat magára, itt a közbringarendszerek analízise még egy szinte teljesen meghódításra váró terület. A teljesség igénye nélkül kezdtem el a vizsgálatokat, nyilvánaló volt, hogy fontos, de csak részleges eredményeket kaphatok. Jelen dolgozat alapja lehet további, jövőbeli kutatásoknak.

A vizsgált időszak 2014 júniusától 2015 áprilásáig terjed ki, alanyai kizárólag a MOL Bubi felhasználói: a tölük kapott információk segítették a kutatást.

A szakdolgozat három nagyobb részre osztható, amelyek kapcsolódnak a kutatásokhoz. Elsőként az adatbányászatot, mint önálló diszciplinát mutatom be és ennek kapcsolatát a biciklimegoszót rendszerekkel, majd az általam készített kérdőívek alapján a primer kutatásokkal dolgoztam, az elégedettség emberi tényezőit figyeltem és elemeztem, melynek eredményeit ismertetem a mellékelt kérdőívekkel és azok statisztikai és adatbányászati kiértékelésével. A harmadik rész a szekunder kutatásokon alapul, a már meglévő adatállományokra, illetve a MOL Bubi üzemeltetői által a rendelkezésemre bocsátot, náluk nyilvántarott adatokra épül. Végezetül mindezek kinyert eredémenyeit hasonlítom össze, majd a modelleket finomítom, és előrejelzéseket készítek a várható biciklibérlések számát illetően. Mindezentúl kitérek a dolgozat témájához kapcsolodó szakirodalomra, áttekintést adok a nemzetközi közbringarendszerekről és a sikerességüket befolyásoló tényezőkről.

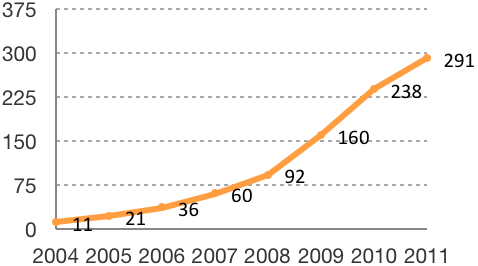
Arra kerestem a választ, hogy mitől függ a biciklibérlések száma. Melyek azok a meghatározó tényezők, különös tekintettel a szubjektív, emberi-felhasználói illetve a külső, objektív pl. időjárási tényezőkre, amelyek köthetők a közbringa használatához.

A dolgozat ***hipotézise***, hogy a biciklibérlők nemek szerinti valamint korcsoportos megoszlása befolyással van az elégedettségre, valamint ugyanez mondható el a szekunder kutatásban vizsgált objektív-külső tényezők és az elégedettségi mutatók között.

1. Szakirodalmi áttekintés

A következőkben más országok biciklimegosztó rendszereit vizsgálom, amellyekkel mélyebb áttekintést szeretnék nyújtani a témával kapcsolatban, és amiből a magyar MOL Bubi a külföldi tapasztalatok alapján még sikeresebben működhet, és kiküszöbölheti a külföldön megtapasztalt hátráltató faktorokat.

1. ábra: Közbringarendszerek számának alakulása 2004-től 2011-ig világszerte



Jelenleg körülbelül 300 biciklimegosztó rendszer működik világszerte. Európa jelentős biciklimegosztó tevékenységével vezeti a kontinensek közötti versenyt. Ázsiában viszont rohamosan növekednek és terjeszkednek ezek a rendszerek. Az 1. ábrán látható, hogy 2004-től kezdődően drasztikus növekedés történt a világ biciklimegosztó rendszereinek számát tekintve, és ez a tendencia még mindig folyamatosan növekszik.8

Valójában minden, a nagyvárosokban jelenlevő közbringarendszer hasonló alapokon nyugszik. Kisebb eltérések technológiai, technikai szempontok alapján lehetnek, mint például a bicikliállomások sűrűsége, vagy magukhoz a kerékpárokhoz kapcsolódó mechanikai zárak különbsége, a felhasználói hozzáférés, esetlegesen a biciklik felvételi és leadási protokolja. Alapvetően azonban hasonlítanak egymáshoz a közbringarendszerek, még a működéshez szükséges eszközökben is ugyanolyan elven működő állomáskezelő vagy információgyűjtő szoftvereket alkalmaznak. Eltérések attól függően alakulhatnak ki, hogy az adott város milyen kényelmi szempontokat részesít előnyben. Például: a MOL Bubi tömör gumikat használ, amik nagyban hozzájárulnak ahhoz, hogy a vandalizmust kiszorítsák, a használatból származó lehetséges károkat kiküszöböljék, míg más városokban a kényelmi szempontokat helyezték előtérbe, ahol a biciklik kerékgumija és váza könnyebb, így egyszerűbb vele a közlekedés.9

Véleményem szerint sok energiát és időt lehet spórolni azzal, ha a biciklik vandálbiztosak, így minden kerékpár hosszú távon működőképes. Kevesebb a gond a szervizelésével és karbantartásával. Nem kell az állomásoknál a rossz biciklikkel számolnia a felhasználóknak. Külföldi tapasztalataim azt mutatják, hogy sokszor nehezítette meg a bicikli bérlését, hogy egyetlen egy ép bicikli sem volt szabad a legtöbb közeli állomáson, és rengeteg időbe telt, mire megfelelő bicikli találni lehetett.

Gazdasági szempontokat vizsgálva nyilván lényeges kérdés, hogy az árakat hogyan szabják meg úgy, hogy közben profitot is termeljen a vállalat. Ha túl drága, kevesebben használják, a turistákat is elriasztja a magas ár, és nem lesz elég felhasználó, hogy fenntartható legyen a működés. Sokszor a városok törekednek arra, hogy minél szélesebb réteg számára elérhető legyen a közösségi kerékpározás, így olyan kedvezményeket ajánlanak fel, amelyek elfogadhatónak számítanak. Így például, mint ahogyan Budapesten is, az első félóra ingyenes. Bécsben is hasonló kedvezmény működik, azonban két biciklibérlés között el kell telnie egy kis időnek. Ezzel biztosítják azt, hogy a kerékpárok ne menjenek tönkre idő előtt, amit talán a magyar Bubinál sem ártana bevezetni óvintézkedésként.

A közösségi kerékpározás sikere alapvetően attól is függ, milyen éghajlattal, milyen földrajzi elhelyezkedéssel rendelkezik az adott ország. A szezonális változások nagyban hozzájárulnak a felhasználók számának gyors vagy éppen lassú növekedéséhez esetleg csökkenéséhez. Egy olyan országban, ahol sokat süt a nap, kevés a csapadék, mint például Izrael, kézenfekvő, hogy a drága tömegközlekedés helyett kipróbáljam a Tel-O-Funokat, a csapadékos országok azonban nem kedveznek a kerékpáros kultúrának.

A fentiek nagyban befolyásolják, hogy milyen a kerékpáros kultúra az adott országban, városban, és milyen lehetőségekkel indulnak a közbringarendszerek.

* 1. Nemzetközi színtereken

A bécsi székhelyű CityBike koncepcióját 12 éve dolgozták ki, és a mai napig hatalmas népszerűségnek örvend. A rendszernek több, mint 100 állomása és 1500 biciklije van. Nagyrészt hasonló adottságokkal rendelkezik, mint a Magyarországon is jelenlevő MOL Bubi.10

Amerikában a biciklis kultúra nem nagyon elterjedt, és ezért az ott fellelhető biciklirendszerek nem példaértékűek, nem túl relevánsak a mi vizsgáltunk szempontjából. Az amerikai biciklirendszerek sikerességéről nagyon sok tanulmány készült, amelyek azt taglalják, hogy lesz a kontintensen biciklis forradalom.

Kitekintés viszont mindenképp érdemes tenni az ázsiai országokra is, leginkább Kínára, ahol egy félreértelmezett politikának köszönhetően a bicikliket egyre jobban szorítják ki, és teret adnak a gépjárműveknek. Egyre csak tűnnek el a bicikliutak, és szélesednek a főutak. Mégis a világ egyik legnagyobb biciklimegosztó rendszere Kínában található, az érzékelhetőség végett több, mint 400 ezer közbicikli található meg, ezeket a rendszereket csak az elmúlt években létesítették. Nem rejtély, hogy ez az óriási számadat csak a lakosság létszámának függvénye, és a gyors urbanizáció mellett alakulhatott ki. Remélhetőleg a tendencia másik irányba megy, és nagyobb hangsúlyt fektetnek a közlekedési dugók csökkentésére és a levegő szennyezettség visszaszorítására.

Sajnos a világ többi részén, ahol hasonló urbanizációs folyamatokkal és kihívásokkal küzdenek, még el sem kezdődtek az ilyen programokban való tervezések. Indában vagy az egész Afrika kontinensen csak kísérleti tanulmányok vannak, eddig zéró aktivitással.11

* 1. Az adatbányászati áttekintés

Az 1960-as évektől az adatbázisok és az információs technológia fejlődése ütemesen felgyorsult, így a korábban fejletlen adatfeldolgozó rendszerek és hardverek sokkal hatékonyabbá és kifinomultabbá váltak. Az 1970-es években beindult a hálózati és relációs adatbázisok fejlődése, amely a relációs adatmodell elvén rendezte az adatokat, így biztosítva precízebb adatrendszerező technikák létrejöttét. Ezek közül kiemelkedik az On-line Transaction Processing (OLTP) informatikai rendszer, amely a könnyebb és dinamikusabb keresést, lekérdezést és adatfelhasználást nagymértékben segítette elő. Az 1980-as évek végétől megjelentek az adattárházak, melyek kiszorították az eddig használt forrásrendszereket.12

Napjainkra a gazdaságot meghatározó tőkeelemek fontossági sorrendje megváltozott, az emberi tudás és információ előtérbe került.13 Mi több, ez annyira determináló tényezővé vált, hogy egy vállalat előmenetelét is befolyásolni tudja. Viszont azzal, hogy hihetetlen mennyiségű adatot halmozunk fel életünk során – vásárláskor, internethasználatkor, mobiltelefonással, de akár egy vállalat szervezeti működése, kutatása vagy fejlesztése közben is -, a keresés és a rendszerezés is nehézségekbe ütközik.14 Továbbá, ha inkább az adatok összefüggéseire vagyunk kíváncsiak, akkor az OLAP (Online Analytical Proccessing) rendszerek használhatatlanná válnak számunkra, ugyanis ezek gyenge pontja az, hogy konkrét és meghatározott keresési feltételeket kell megadnunk, ami ilyen mennyiségű adatfelhalmozás mellett nem képes megfelelő hatékonysági mutatót produkálni.15 Ezzel szemben az adatbányászat az adatok összefüggéseinek feltárásában tud hasznunkra lenni. Összegzésképp tehát az adatbányászat definíciói közül néhány:

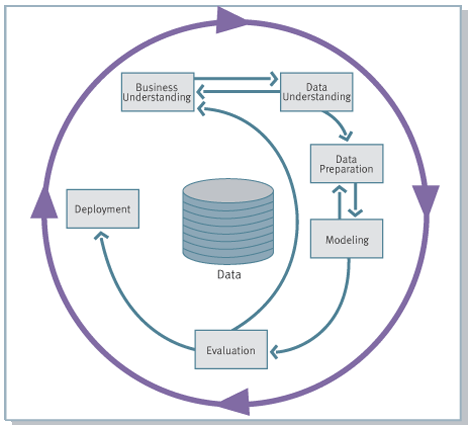
* „Az implicit, korábban ismeretlen és hasznos információk kinyerése különféle adatokból” 16(Frawley, Piatetsky-Shapiro és Matheus, 213-228. oldal)
* „Hasznos információk hatalmas adatsorokból vagy adatbázisokból való kiszűrésének tudománya”17(Hand, Mannila és Smyth)
* ,,Az adatbányászat nagy mennyiségű adatok kutatása és elemzése rejtett minták és szabályok felfedezése érdekében”18(Michael J. A. Berry and Gordon S. Linoff)

A definíciókból is látható, hogy nem mindegy, milyen összefüggéseket tárunk fel, ugyanis az általunk keresett számos információ mellett melléktermékként több haszontalan és felesleges információ is keletkezik. Ezért az információnak 3 fontos kritériumnak kell megfelelnie: újszerűség, hasznosság, magyarázhatóság. Az újszerűség az adatbányászok számára új, még ismeretlen összefüggést jelent. Hasznosság alatt értendő, hogy a kinyert adatok üzleti célokra és tartalmakra felhasználhatóak legyenek, ez a hasznosság csak vélhető, spekuláción alapul. A magyarázhatóság pedig a miért és hogyan kérdésekre ad választ, azaz a megszerzett adatok célját és folyamatát hivatott megvilágítani.19

* 1. CRISP-DM modell

A CRISP-DM, azaz CRoss Industry Standard Process for Data Mining modell első hivatalos verzióját 1999-ben hozták először nyilvánosságra. Ennek a nemzetközileg elismert szabványnak a célja az, hogy az adatbányászat folyamatát lépésről lépésre írja le, így biztosítva a kinyerendő adatok minőségi homogenitását. 20 (SSPS: Step-by-step Data mining guide 10-29. oldal)

Az 2. ábrán látható adatbányászati modell áttekintést nyújt az adatbányászati projektek életciklusát illetően, ami magában foglalja az ezeken belüli egyes szakaszokat és az ezekhez kapcsolódó feladatokat, és végezetül a feladatok közötti összefüggéseket. Ennél a szintnél már lehetetlen minden kapcsolat azonosítása. Mindenféle adat között létre jöhet kapcsolat, az adatbányászati cél a háttéradatok, a felhasználó érdeklődésétől függően változik, és természetesen az adattól magától is függ.



2. ábra: A CRISP modell

Forrás: Machinelearning.ru

Az adatbányászati projektek 6 ciklusból tevődnek össze, ami az 2. ábrán látható. A szakaszok sorrendje nem kötött, sőt kifejezetten ajánlatos a fázisok közötti ugrás. Minden szakasz eredménye befolyásolja, hogy melyik szakasz legyen a következő. Az ábrán a nyilak a leginkább gyakori és ajánlott kapcsolatokat illetve függőséget ábrázolják.

A külső kör az adatbányászás ciklikusságát jelzi, ha a folyamat véget ér, akkor újból kezdődik, elölről. Az eredmény a folyamat közben jön létre és formálódik. A következőkben röviden összefoglalom, mely szakaszokból tevődik össze e folyamat:21

* Business Understanding, azaz üzleti probléma értelmezése: A kezdeti szakaszban a projekt üzleti és adatbányászati céljait, követelményeit kell meghatároznunk, ehhez pedig a rendelkezésünkre álló források megfelelő használata és egy projektterv szükséges, aminek elkészítéséhez a kellő eszközök és technikák helyes kiértékelésére, használatára kell fókuszálnunk.
* Data Understanding, azaz adatértelmezés: ez a szakasz a kiindulási adatok összegyűjtésével kezdődik, majd a következő lépés ezen összegyűjtött adatok csoportosítása a főbb jellemzőik alapján. A csoportosítás egyben értékelést is jelent, amely lehetőséget nyújt arra, hogy rápillantást kapjunk az adatokra – minőségi szempontból.
* Data Preparation, azaz adatok előkészítése: ebben a fázisban a különböző modellezési technikák közül a megfelelő kerül kiválasztásra és alkalmazásra. Ugyanarra az adatbányászati problématípusra jellemzően több modellezési technika is létezik. Bár vannak olyan típusú problémák, melyek a folyamat során specifikus követelményeket állítanak. Így lehetséges, hogy az adatok sokszor visszakerülnek az előző szakaszba, ahol újra kiértékelik őket.
* Modeling, azaz modellezés: különböző modellezési technikák kerülnek kiválasztásra és alkalmazásra. Jellemzően számos modellezési technika áll rendelkezésre ugyanarra az adatbányászati problématípusra, bár néhány technikának specifikus követelményei vannak a folyamat során. Így lehetséges, hogy az adatok sokszor visszakerülnek az előző szakaszba, ahol újra kiértékelik.
* Evaluation, azaz üzleti értékelés: ennél a pontnál a projekt megfelelő minőségű adatokkal és egy már kiválasztott adatbányászati modellel rendelkezik. De a folyamat elindítása előtt nagyon fontos, hogy üzleti szempontból is kiértékeljük, és még egyszer a teljes elemzési folyamatot felülvizsgáljuk.
* Deployment, azaz alkalmazás: azzal, hogy kiválasztottuk a megfelelő modellt és az előző lépéseken végigmentünk, és kinyertük a kívánt információkat, a folyamatnak még nincs vége. Ugyanis a projekt így még felülvizsgáltra szorul, feedback-et kell készíteni, azaz a folyamat pozitív és negatív oldalait fel kell tárni, és dokumentálni kell azokat, ennek eszközéül a projekttanulmány szolgál (beszámoló, prezentáció stb.), továbbá a folyamatot továbbra is fent kell tartani, és irányítani kell.
  1. Feladattípusok
     1. Asszociációs szabályok

Az asszociációs szabályok az objektumok között keresik az összefüggéseket, és az ezek között fennálló kapcsolat erősségét vizsgálják.22 (J. Han & M. Kamber, 13-16. oldal) Így a vásárlási szokásokat is vizsgálat alá lehet venni ezzel a módszerrel. Ilyenkor a cél a gyakori termékegyüttesek/elemhalmazok megtalálása. A gyakoriság fontossága abban áll, hogy csak ennek az adatnak a segítségével lehetséges megbízható statisztikai információt adni. A kérdés pedig az, hogy ha A terméket megvesszük, akkor mekkora az esélye annak, hogy B terméket is veszünk mellé. Konkretizálva, ha babapelenkát veszünk, akkor annak az esélye, hogy veszünk babahintőport is 78%. Természetesen magától értetődik, hogy a két termék között szoros az összefüggés. Ilyen nagymértékű egybeesés olyan termékek között is létezik, amelyek között látszólag nincs összefüggés, erre példa az autókat és a romantikus könyveket reklámozó felületek. A kapcsolatok ilyen formáját lehet az asszociációs feladatokkal feltérképezni, ehhez persze elengedhetetlen, hogy magát az asszociációs szabályt is megvizsgáljuk és megértsük. Ily módon egy vállalat előnyre tehet szert, ha ezeket az adatokat felhasználva ügyel egy üzlethelyiségben a termékek elrendezésére, hogy a vevő elé kerüljön mindkét termék, az ő akaratától függetlenül.23 A reklámokkal foglalkozó weblapok is hasznát vehetik ennek az alkalmazásnak, mert így célirányosan tudják a potenciális vásárló várható érdeklődésének megfelelő termékeket népszerűsíteni. 24

* + 1. Osztályozás

Klasszifikáció alatt az értendő, hogy egy olyan modell kerül kialakításra, ami képes meghatározni és megkülönböztetni specifikus kategóriákat. Azaz a predikció célja, hogy a tulajdonságok milyenségét figyelembe véve osztályokat alkosson, továbbá az, hogy az osztályok attribútumait feltárja. Ilyen osztályként szolgálhatnak a (Ha-Akkor) szabályok, a döntési fák, meghatározott matematikai algoritmusok vagy a neurális hálók. 25

„A **döntési fa** egy olyan, a döntéshozatalban használt grafikus modell, amit az optimális tevékenység határoz meg olyan esetekben, amikor több választási lehetőség is rendelkezésre áll, és a kimeneteik bizonytalanok ebben az esetben maguk az osztályok.”26. (Wikipedia: Döntési fa első sor) „Az agykutatás és a mesterséges intelligencia területek egymást inspiráló és rohamosan fejlődő ágainak közös eredményeként határozható meg a **neurális hálók** kifejlesztése.”27 (Kovács Gy. 10. diasor)

Leggyakoribb alkalmazásuk a hitelelbírálásánál, égitestek besorolásánál vagy kampányelemzéseknél fordul elő.28

* + 1. Regresszió

„A regresszió úgy mutatja meg két változó kapcsolatát, hogy egyben az egyik változó (függő változó/eredményváltozó) a másik változótól (független változó/magyarázóváltozó) való függésének mértékét is kifejezi.”29 (Kovács Gy. 10. diasor) A leggyakrabban a gazdasági matematikában használják a regressziót. Itt idősorok jövőbeli állapotát tudják megállapítani egy már meglévő idősorból. Különböző termékek későbbi népszerűségének megjósolásához is használatos. Az úgynevezett sikerfüggvény alkalmazásai is helyet találnak a regresszióban, mert ez összefüggésbe hozza a célközönséget, a termék árát a reklámra szánt összeggel, így előrejelzi a várható profitot. Ezzel a módszerrel dolgoztam a vizsgálataim során, így ennek a pontosabb kifejtésére most nem kerül sor. 30

* + 1. Klasszterezés

A klaszterezés és az osztályozás egymástól nem túl távol eső folyamatok, mert a klaszterezés révén egy előre meghatározott minta felhasználása nélkül a halmaz elemei egy meghatározott csoportba kerülnek. Az algoritmus a működése során megállapítja, hogy a halmaz mely elemei tartozhatnak össze, majd a csoportokat ellátja a megfelelő címkével, amely az összetartozásukat hivatott jelezni. Egy klaszterbe azon elemek tartoznak, amelyek magas hasonlósági rátával rendelkeznek, azok pedig, amelyek egy adott elemhez alacsony rátával, míg egy másikhoz magas rátával kapcsolódnak, képeznek egy újabb csoportot. A csoportot alkotó elemek összességében (azaz tulajdonképpen magán a csoporton) szabályszerűséget lehet felfedezni. Tipikus példaként említhető a piacelemzés.

* + 1. Eltéréselemzés

Ezen feladattípus lényege, hogy kiszűrje azokat az adatokat, amelyek lényegesen eltérnek a többi adattól, vagy más mintát követnek, mint a többi. Ezeket az adatokat az adatbányász megvizsgálja és elemzi. Leggyakrabban a csalások és visszaélések felderítésében játszik szerepet, mint például bankkártyacsalás, adócsalás stb.31

* 1. Adatbányászat és a biciklimegosztó rendszerek kapcsolata

A hosszabb időszak alatt gyűjtött forgalmi adatok feldolgozása által képet kaphatunk a felhasználók viselkedési szokásairól, a gyűjtőállomások időbeli kihasználtságról, illetve a logisztikai folyamatokról. Mindezek mellett a kinyert adatokból ellenőrizhetjük az állomások kihasználtságának helyességét és a nem megfelelő kihasználtság mellett optimalizálhatjuk a logisztikai folyamatokat. Talán az egyik legfőbb feladat a közbringarendszereknél a folyamatos egyensúly megtalálása, hogy a megfelelő számú kerékpárok álljanak rendelkezésre, hogy kielégítsék a felhasználók keresletét. Adatbányászati szempontból prediktív előrejelzéseket lehet készíteni a felhasználók viselkedési mintáiról, így meg lehet határozni a bicikliknek a jövőbeni útvonalait és megkönnyíteni az optimalizálási folyamatokat. 32

1. Kutatások
   1. A primer kutatás

A primer kutatásom a téma humán tényezőire koncentrált: a felhasználói szokásokat mértem fel kérdőíves módszerrel. Az alanyok kiválasztásánál alapszempont volt, hogy biztosan MOL Bubi használók legyenek, ezért a felmérések helyszínei a legforgalmasabb dokkolóállomásokhoz kötödtek. A Deák téri, Kálvin téri, valamint az egyetemekhez közeli le-és felvevő pontoknál, a Budapesti Műszaki és Gazdaságtudományi Egyetemnél, a Budapesti Corvinus Egyetemnél és az Eötvös Lóránd Tudomány Egyetemnél. A megkérdezettek száma több, mint 50 fő volt. Mivel a felmérés ideje nem korlátozódottt egy napra, ezért nem egyetlen nap keresztmetszetét kapjuk, hanem három héten keresztül (2015. márciusban) egy-egy napot töltöttem a személyes megkédezéssel.

A felmérés (1. melléklet) kérdéseinek összeállításánál szempont volt a bickili bérlő profilja (természetesen név nélkül): neme, életkora, foglalkozása, lakhelye stb., amelyek alapján különböző egymással összevethető osztályokat alakítottam ki. A kérdések másik csoportja közvetlenül a közlekedési szokásokhoz, azonbelül is a MOL Bubi használatához köthetők. Megvizsgáltam a bérléshez vonatkozó motivációs tényezőket, valamint a felhaszálási célokat, a közlekedési eszközök és a MOL Bubi használatának kapcsolatát. Kiemelten figyeltem a rövid és hosszútávú jegyek illetve bérletek arányára, illetve lehetőséget adtam a megkérdezettek számára, hogy javaslataikat feltüntessék a kérdőíven. A felhasználói elégedettséget megnéztem az egyes összetevők (térkép elérhetősége és minősége, éjszakai világítás, állomások biztonsága, információk elérhetősége, call center stb.), de külön rákérdeztem az elégedettség mértékére a szolgáltatás összességének tekintetében is.

* + 1. A feldolgozás módszertana

A kérdőíven megkapott szöveges válaszokat számszerűsítettem, hogy az adatokkal a kiválaszotott módszerekkel dolgozni tudjak. A hipotézisek alátámasztására statisztikai módszereket használtam: két és három mintás t-próbát, ANOVA táblát és függetlenségi vizsgálatot. A kinyert adatok közül két lényeges nézőpont alapján válogattam: egyrészt, hogy kapcsolódjon az elégedettséghez, másrészt pedig, hogy a kiválasztott módszerekkel fel lehessen dolgozni őket. Minden egyes próba esetén 5% szignifikancia szinttel számoltam.

Előfeltevésimet az alábbiakban modelleztem:

1. Arra kerestem a választ, van-e szignifikáns eltérés az általános elégedettség terén a nemek között?

Mivel ebben az esetben két felhasználható érték adott (Férfi/Nő), így egy kételemű t-mintás próbát alkalmaztam a vizsgálat során. Ez a modell azt vizsgálja, hogy két eltérő mintában található átlagok szignifikánsan eltérnek-e egymástól.

Kiindulásként felállítottam két hipotézist, melyet a próba eredményei alapján elfogadtam vagy elutasítottam.

* H0 : A férfiak és nők elégedettségi foka nagy valószínűséggel azonos.
* H1: A férfiak és nők elégedettségi foka nem azonos.

A kétoldali t-próba előfeltétele a szórásegyezőség, emiatt F-próbával kell ezt megvizsgálni. Számolását az alábbiakban mutatom be:

3.ábra: Férfiak és nők elégedettségének átlaga és szórása

|  |  |  |  |
| --- | --- | --- | --- |
|  | elemszám | átlag | szórás |
| férfiak | 28 | 4,39 | 0,84 |
| nők | 23 | 4,43 | 0,79 |

A 3.táblázatban található, külön a két nemhez köthető elégedettségi átlag férfiaknál 4,39, míg a nőknél 4,43, nem láthatunk nagymértékű eltérést. Úgyanígy nem találtam szignifikáns eltérést a szórások között sem. F-próba segítségével viszont ellenőrizni kell, hogy mindkét szórás egyezik-e.



Fszámított így 1,130588047 és Fkritikus 2,00 (az F eloszlás táblázatból számolva). Előbbi nagyobb, mint utóbbi, így elfogadjuk, hogy a szórások megyeznek. Ezzel megkaptam a t-próba előfeltételét.

A Ho hipotézisünk elfogadásához a t-mintás próba esetén Tszámított értéknek nagyobbnak kell lennie a Tkritikus értéknél. Míg ellenkező esetben H1 hipotézist fogadjuk el.

T érték kiszámításához sp-t is számolni kellett.



Jelen esetben, n1 a férfiak elemszámát, n2 a nők elemszámát értjük. Az eredményünk pedig 0,669008 lett. Ebből számítunk a következő módon Fkritikus és Fszámított eredményt:

1. , ahol



x1felülvonás a férfiak átlagos elégedettségét, x2felülvonás a nők átlagos elégedettségét jelöli. Az így kapott érték 0,173781 ezzel szemben tkritikus 2,009575.

Az t értékek alapján a H1 hipotézisünket elfogadjuk, miszerint a férfiak és nők elégedettségi foka azonos.

*P* valószínűsági érték kiszámítása esetén megkapjuk, hogy mennyi az esélye annak, hogy a férfi és női elégedettségi szint megyezik. Ez alapján 86,30%-a annak, hogy a kettő nem tér el egymástól.

1. Hasonló módszerrel vizsgáltam azt is, hogy van-e eltérés a férfiak és nők között az egy főre eső javaslatok számában.

Kiindulásképpen ismét felállítottam mind a H0 hipotézist, mind a H1 hipotézist:

* H0: A férfiak és nők jellemzően azonos számban tesznek javaslatot
* H1: A férfiak és nők jellemzően nem azonos számban tesznek javaslatot

Első lépésként elkülönítettem egymástól a férfiakra illetve a nőkre vonatkozó adatokat, így képet kaptam a javaslatok alakulásáról az egyes személyek és a nemek szerinti megoszlásáról. Az így rendszerezett adatok között volt olyan érték (az egy főre eső javaslatok száma), amely 0, azaz a szolgáltatásokban semmilyen kivetnivalót nem talált, míg az is előfordult, hogy a javaslatok száma 7 volt a maximális 9-ből.

4. ábra: Egy főre eső javaslatok nemek szerinti átlaga és szórása

|  |  |  |  |
| --- | --- | --- | --- |
|  | elemszám | átlaga | szórása |
| Férfiak | 28 | 3,43 | 1,97 |
| Nők | 23 | 3,78 | 1,88 |

A 4.táblázatban található, külön a két nemhez köthető, egy főre jutó javaslatok átlaga a férfiaknál 3,43, míg a nőknél 3,78, ami nem számít szignifikáns eltérésnek. A szórások értéke is nagyjából azonos mértékű.

Ismét t-próbát alkalmaztam az elöző hipotézisvizsgálathoz hasonlóan. Az Fkritikus=0,91 Fszámolt=2,00, azaz kritikus érték kisebb, mint számolt érték, így megvan az előfeltétele a t-próbának. Ezért kiszámolhatjuk t értékeit, ami Tkritikus és Tszámolt esetén rendre 0,651407 és 2,009575.

Mivel Tkritikus kisebb, mint Tszámolt a H0 hipotést fogadjuk el, amely alapján az egyes személyek, nemtől függetlenül azonos számban tesznek javaslatot. Ennek a valószínűsge *P* érték alapján 51,78%. A férfi én női javaslattételek száma megegyezik.

1. A következő, engem érdeklő kérdés, hogy vajon a korcsoportok között van-e különbség az elégedettség mértékében?

Ebben az esetben a hipotézisek így alakultak:

* Ho: az egyes korosztályok azonos mértékben elégedettek
* H1: az egyes korcsoportok nem azonos mértékben elégedettek

A korcsoportok felállítását az alábbi intervallumokban határoztam meg: 18-24, 25-34, 35-44.

A teljes minta átlagos elégedettségi mutatója 4,41, külön-külön a korcsoportokra nézve a legmagasabb érték a 18-24 éves korosztályra jött ki, míg a legalacsonyabb a 35-44 éves korúakhoz kapcsolható.

5.ábra: Elégedettség alakulása korcsoportok szerint

|  |  |  |  |
| --- | --- | --- | --- |
|  | elemszám | átlag | szórás |
| 18-24 | 16 | 4,75 | 0,58 |
| 25-34 | 29 | 4,38 | 0,86 |
| 35-44 | 6 | 3,67 | 0,52 |
| összesen | 51 | 4,41 | 0,80 |

Az elemzéshez ANOVA táblát készítettem, melynek eredményei 6. Táblázatban láthatóak, ennek alapján számítottam ki a P értékét: 0,015 és Fszámított értéket: 4,56.

6. ábra: ANOVA tábla

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Négyzetösszeg | DF | Szórás^2 | F érték | p érték |
| Csoportok közötti | 5,161 | 2 | 2,58 | 4,56 | 0,015 |
| Csoporton belüli | 27,192 | 48 | 0,57 |  |  |
| Összes | 32,35294118 | 50 |  |  |  |



Ezen felül megkaptam az Fkrit = 3,19 értéket (az F-eloszlás táblája alapján). Mivel az Fszámított érték nagyobb, mint az Fkritikus értéke, ezért a H0 hipotézist elutasítom, és a H1 hipotézist tekintem igaznak. A *P* értéke alapján az is elmondható, hogy 98,5% esélllyel minél fiatalabb valaki, annál elégedettebb a MOL Bubi által nyújtott szolgáltatásokkal.

1. *Végezetül megvizsgáltam, hogy függ-e az életkortól a váltott jegy vagy bérlet típusa.*

Ebben a hipotézisvizsgálatban függetlenségi vizsgálattal kerestem, hogy melyik hipotézisem bizonyul igaznak, amelyek:

* H0: A korcsoportok és a váltott jegyek vagy bérletek típusai között nincs kapcsolat
* H1: A korcsoportok és a váltott jegyek vagy bérletek típusai között van kapcsolat

Első lépésben a tapasztalati gyakoriságokat rendszereztem, melyek az 7. táblázatban találhatóak meg:

7. ábra: Tapasztalati gyakoriságok

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | éves | féléves | hetijegy | összesen |
| 18-24 | 7 | 6 | 3 | 16 |
| 25-34 | 6 | 10 | 13 | 29 |
| 35-44 | 2 | 4 | 0 | 6 |
| Összesen | 15 | 20 | 16 | 51 |

Már itt látható, hogy a 18-24 éves korosztálynál az éves/féléves bérletek a dominánsabbak, a 25-34 éves korosztálynál a hetijegy, és legvégül a 35-44 éves felhasználók leginkább féléves jegyet váltanak.

8. ábra: Elméleti gyakoriság (a táblázatban lévő adatok számításának módja az 2.melléklet alapján)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | éves | féléves | hetijegy | összesen |
| 18-24 | 4,7 | 6,3 | 5,0 | 16 |
| 25-34 | 8,5 | 11,4 | 9,1 | 29 |
| 35-44 | 1,8 | 2,4 | 1,9 | 6 |
| összesen | 15 | 20 | 16 | 51 |

A táblázat alapján végzek khi-négyzet próbát, így számíottam ki az értékét a Khiszámított =39,2, Khikritikus =9,49 mutatókat.

1. , ahol



fij = az i-edik sor és a j-edik oszlop tényleges gyakorisága, az Eij = az i-edik sor és a j-edik oszlop várható gyakorisága, az *s* = sorok száma és végezetül az *o* = oszlopok száma.

,,*A Khi négyzetpróba (függetlenségi vizsgálatunk) annak a valószínűségét közli, hogy függetlenséget feltételezve véletlenszerűen előfordulhat olyan statisztikai érték*”, mely legalább akkora, mint a fenti képlettel kiszámított érték.33

Mivel a kritikus érték kisebb, mint a számított, így a függetlenségi vizsgálat azt az eredményt hozta, hogy H0 hipotézis elvetendő, H1 hipotézis pedig megtartandó, az életkor befolyással van a váltott jegyek/bérletek típusára. A vizsgálat szerint a *P* valószínűségi érték közel 100%-os lett, szinte biztos a köztük lévő kapcsolat.

A primer kutatás összefoglalásaként elmondható, hogy mindkét, a nemekkel kapcsolatos vizsgálat kimutatta, hogy sem az elégedettség mértéke, sem a javaslattételek számának szempontjából nem mutatható ki szignifikáns kapcsolat a változók között. A korcsoportokkal kapcsolatos elemzéseknél az elégedettségi szint egyértelműen függ a kortól: a 18-24 éves korosztálynak magasabb az elégedettségi szintje az idősebbekénél. Ugyancsak bizonyítható az összefüggés a korosztályi megoszlás és a jegyek/bérletek különböző típusainak vásárlása között. Az utóbbi két eredmény között feltételezhető összefüggés, amennyiben a legnagyobb elégedettséget mutató (18-24) korosztály a legnagyobb arányban veszi meg a hosszabb távra szóló bérleteket (éves/féléves).

* 1. A szekunder kutatás

Ebben a részben a Mol Bubi biciklibérlések számának alakulása kerül modellezésre azoknak a külső tényezőknek a segítségével, amelyek kiválasztásanak okait a bevezetésben említettem. Elsősorban az adatok elérhetőségén múlt a szelektálás, illetve megnehezítette a munkát az adatok/adatbázisok korlátozott száma. Az egyik legnagyobb kihívás, feladat, amivel az adatbányászoknak szembe kell néznie: az adatok megszerzése. Európában az adatbiztonság szigorúbb mivolta miatt az adatokhoz való hozzáférés sokkal korlátoltabb, mint pl. az Egyesült Államokban, ahol több forrás áll rendelkezésre.34

Mivel a kutatás témája a Mol Bubi biciklibérlések alakulása, a legkézenfekvőbb adatbázis közvetlenül kapcsolódik hozzá: az üzemeltetők betekintés engedtek saját belső rendszerükbe, amely egy adatbányász számára megannyi lehetőséget rejt.

A külső tényezők másik csoportját, az időjárási értékeket egy, a mindenki számára hozzáférhető adatbázisból (www.ogimet.com) kerestem ki.

A vizsgálat időtartama 2014. június 12-től 2015. április 28-ig tartott. Ez az időszak tehát jóval meghaladja a primer kutatás időtartamát, azaz a minta nagysága a szekunder kutatás eredményét reprezentatívvá teszi (több, mint 320 adattal dolgoztam).

A szekunder kutatás alapján arra kerestem a választ, hogy milyen külső befolyásoló tényezők hatottak a biciklibérlések számának alakulására.

Az adatok feldolgozásához a lineáris regressziós modellt alkalmaztam: a magyarázó változó és az eredményváltozó között lineáris kapcsolatot feltételezünk. A mi modellünk esetében nem egy egyszerű lineáris regressziós modellről van szó, hanem több magyarázó változóval dolgoztunk, amellyel a nemlinearitást is beépíthettük a modellbe.

1. , ahol



xn értékei a magyarázóváltozók értékeit, Bx pedig a hozzájuktartozó koefficienseket jelöli.

*A szekunder kutatás* ***hipotézise****, hogy az általam kiválasztott magyarázóváltozók hatással vannak a biciklibérlések számára.*

* + 1. Az eredményváltozó meghatározása

Szakdolgozatom a MOL Bubi-bérlők elégedettségét vizsgálja, mindazonáltal jelen esetben ennek definiciója úgy alakul, hogy a lineáris regressziós modell létrehozásához kijelentjük, hogy a bérlések száma az elégedett felhasználók számával arányos. Így a biciklibérlések számát jelöltük ki eredményváltozónak.

* + 1. A magyarázóváltozók kiválasztása

Kiindulópontként több olyan külső tényezőt is felsorakoztattunk, amelyek feltételezhetően befolyásolják a biciklibérlések számát. Az időjárási jelenségek, a hétköznapok megoszlása, a Ferihegyre érkező utasok száma, a fővárosi tömegközlekedési eszközök meghibásodásainak száma, mind megannyi felmerülő lehetőség volt. (Az utóbbiak végül azért maradtak ki a vizsgálatból, mert semmilyen adatbázis ezekkel kapcsolatban nem állt a rendelkezésemre.)

A külső, az időjáráshoz kapcsolódó tényezők közül az alábbi tizenhármat választottam ki magyarázó tényezőként:

* Napi maximum hőmérséklet 0C
* Napi minimum hőmérséklet 0C
* 24 órás átlaghőmérséklet 0C
* 24 órás átlag harmatpont
* 24 órás relatív nedvesség
* Szélirány COS-ban mérve
* Átlagszél
* Légnyomás
* Csapadék
* Felhőzet (x/8)
* Alacsony szintű felhők (1/8)
* Napos órák száma
* Látótávolság

Ezen kívül két dummy változót is beiktattunk. Ha az adott nap hétköznapra esett, az érték 1-et vett fel, míg a hétvége 0-át. A másik esetben külön választottuk a vizsgált időszakot, 1-es értéket társítottuk a tesztüzemeltetés napjaihoz, míg 0 értéket kaptak a teljes értékű üzemeltetési napok.

Lehetőségként felmerült a problémás bérlések száma is, de mint magyarázó változó, nem lehet egyértelműen kijelenteni, hogy az ok-okozati viszonyban a biciklibérések száma van-e hatással a problémás bérlések számra, vagy az összefüggés éppen ennek a fordítottja-e.

* + 1. Adattisztítás, konvertálás, hiánypótlás

Ahhoz, hogy kapjak egy saját adattáblát a fenti tényezőkből, az esetleges hibás adatokat kiszűrtem, a szöveges együtthatókat numerikus adatokká konvertáltam, valamint a hiányzó adatokat pótoltam. Példának említhetjük, hogy eredetileg a szél kódolása É,D,NY,K, és ezek kombinációiból tevődött össze a szélirány értéke, amit a regressziós modellünk sajnos nem tud értelmezni, így ezeket az értékeket fokokba váltottam át, majd így számoltam tovább. Másrészről a csapadék mennyisége sok esetben 0 vagy nagyon szélsőséges értéket vett fel, így a csapadék kódolásához 0, 1 és 2 értéket társítottam. (0, ha semmilyen csapadék nem volt az adott napon, 1, ha mérsékelten volt csapadék, és végül 2, ha sok csapadék esett.)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Nap | Teszt (0v1) | Hétköz-nap(0v1) | Hőm. Max 0C | Hőm. Min 0C | Átlagos hőm. 0C | Harmat- pont | Relatív Nedv. | ... |
| 2014.06.12 | CS | 0 | 1 | 34 | 18,6 | 27,1 | 11 | 38,8 | ... |
| 2014.06.13 | P | 0 | 1 | 31,9 | 18,2 | 25 | 11,7 | 45,5 | ... |
| 2014.06.14 | SZo | 0 | 0 | 26,9 | 17,5 | 21,9 | 6,5 | 38,5 | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

Az alábbiakban a végleges adattáblát mutatom be. Megtalálhatóak benne a modellhez tartozó eredmény- és magyarázóváltozók elemei. Az adatbázis egy 321 elemű minta, amely naponkénti eloszlásban 2014. juniús 12-tól 2015. április 28-ig terjedt 15 magyarázóváltozót tartalmazva. 3 modellváltozat közül az egyes modellek változóinak P értékei alapján választottam ki a legjobbat.

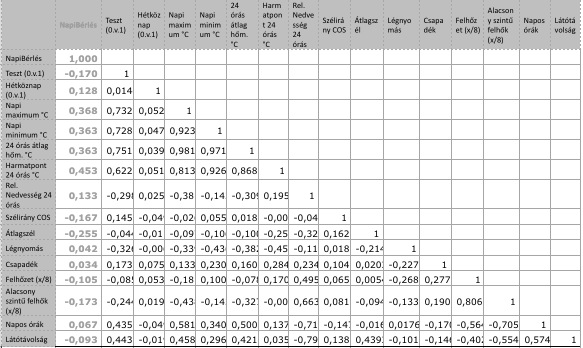
9.ábra: Adattábla

* + 1. Korrelációs mátrix vizsgálata

A korrelációs mátrix segítségével meghatározhatjuk, hogy az egyes változók milyen százalékos arányban magyarázzák egymást, a sorok és az oszolopok megfeleltetési pontjában található korrelációs mérőszámok alapján. A mi esetünkben egy többváltozós korrelációról van szó. Látható, hogy az egyes változók milyen mértékben magyarázzák egymást.

Többváltozós korrelációnál többfajta mérőszámot lehet meghatározni, amelyek megmutatják, mekkora a kapcsolatok szorossága. Minden korrelációs együtthatóra teljesül, hogy az értéke 0 és 1 között mozog. Minél erősebb a kapcsolat szorossága, annál jobban közelít a korrelációs együttható az 1 értékhez.

10. ábra korrelációs mátrix



Első lépésként az eredményválozó (biciklibérlések száma) és a magyarázóváltozók korrelációs együtthatóit vizsgáltam meg. Szignifikánsnak tekintttem azokat a tényezőket, amelyek leginkább közelítettek abszolút értékben az 1 értékhez.

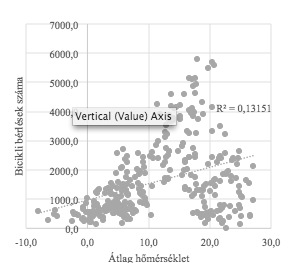
A táblázatban jól látható, hogy három tényező között (a minimum/maximum és a 24 órás átlaghőmérséklet) a korrelációs együtthatók alapján multikollinearitás van. 92, 97 illetve 98 százalékban magyarázzák egymást, ami rontja a modellünk becslésének pontosságát. Ezért megvizsgáltam, hogy a három tényező közül melyiknek van a legszorosabb kapcsolata az eredményváltozóval. Ez alapján a napi maximum hőmérséklet korrelációs együtthatója a legnagyobb, viszont, mivel ennek értékét nem 0-tól 24 óra átlag hőmérsékletéből számítják ki, hanem elcsúsztatják az időhatárokat (akár reggel 8-tól másnap reggel 8-ig), feltételezhetően torz eredményeket kapnánk, ha ezekkel az adatokkal dolgoznánk. Hogy ezt kiküszöböljük, a napi hőmérséklet helyett a napi 24 órás átlaghőmérsékletet választottam a magyarázótényezők közé.

A táblázat tartalmaz még egy olyan tényezőt, a csapadékmennyiséget, amihez olyan korrelációs együttható társul, amely indokolná ennek az elemnek a kihagyását. Figyelembe kellett venni azonban azt a tényt, hogy nyaranta a gyors lefolyású zivatarok eltorzítják az átlagértéket. A nyári időszakban a csapadék mennyisége épp annyira magas, mint a bérlések száma, ami hamis összeffüggésekre mutat rá. Ezt megfontolva a csapadékmennyiség, mint tényező nem hagyható ki a magyarázóváltozók sorából.

Így végül az alábbi tényzők kerültek a szignifikáns magyarázóváltozók közé: a két dummy változónk (hétköznapok és a tesztüzemeltetés időszaka), a 24 órás átlaghőmérséklet, a szélirány COS, az átlagszél, a csapadékmennyiség, a felhőzet és a napos órák száma.

A következő grafikonok értelmezései segítenek a magyarázóváltozók helyes kiválasztásának ellenőrzésében, amelyeken jól látható, hogy milyen tendenciák vannak az egyes változók között.

11. ábra: Összefüggés a biciklibérlések száma és az átlaghőmérséklet között

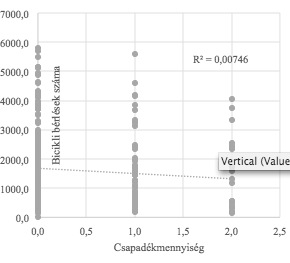


Amennyiben minden pont a regressziós egyenesre illeszkedik, a többszörös determinációs együttható R2=1. Ebben az esetben a modell tökéletes (szaturált vagy telített). Ezzel ellentétben, a null, azaz üres modellnél nulla a regressziós egyenes meredeksége is, véletlenszerűen helyezkednek el az adatpontok, nincs magyarázóerő.

Az 11.ábra az átlaghőmérséklet és a biciklibérlések számának viszonyát mutatja. Átlaghőmérséklet változása esetén jól láthatóan egy egyenes köré csoportosulnak a pontok, másrészről a bérlések száma változik: a hőmérséklet növekedésével biciklibérlések száma az egyenes mentén szóródik tovább. A grafikon felmutat részt, ahol a pontok tömörülnek, ennek alapján vélhetően van egy olyan rendszeresen a MOL Bubit használó felhasználói réteg, akik függetlenül az időjárástól közlekednek így. Feltételezhető tehát az összefüggés a két tényező között. R2 magyarázó tényező az 13,15%, amely szerint ez a faktor erősen magyarázza a bicikli bérlések számát. A függvényvizsgálat alapján pozítiv visszajelzést kaptam az átlaghőmérsékletnek, mint magyarázó tényező kiválasztására.

Ígyugyanígy megnéztük a csapadékmennyiség és a bérlések száma közötti összefüggést, ezt a 12.ábra mutatja. A csapadékmennyiség alakulásával fordítottan arányos a biciklikbérlések számának változása. A függvény megerősíti az általános feltételezéseket: alacsony csapadékmennyiség esetén többen használják a MOL Bubit, míg magasab csapadékmennyiség esetén egyre kevesebben.

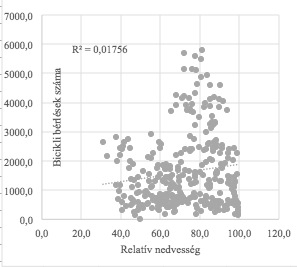
12. ábra A csapadékmennyiség és a kerékpárkölcsönzések függvénye



Minden, általám választott magyarázótényező függvényének alakulását a fentiekhez hasonlóan megvizsgáltam. A szignifikancia minden esetben ellenőrzésre került, mind a nyolc megmaradt magyarázótényező alkalmasnak bizonyonyult arra, hogy a további vizsgálatokat velük végezzem el.

A magyarázóváltozók szelektálásánál a vizsgálatból kikerült változókat is ábrázoltam a grafikonon. Az 13.ábra bizonyítja, hogy nincs markáns csoportosulás egy függvény köré, hanem a pontok szóródása véletlenszerűséget mutat. Ebben az esetben nem tudnánk használni a regressziós modellünket, mert rontaná a becslés pontosságát.

13. ábra A relatív nedvesség és a bérlések száma közötti függvény



A függvényekkel való visszaellenőrzés megerősítette, hogy a 15 feltételezett tényezőből a releváns nyolc, kiválasztott faktorok szignifikáns magyarázótényezőként vizsgálhatók a további modellezés során.

* + 1. Modellezés

A 8 magyarázótényezősre szelektált modellre alkalmaztam a linerás regressziós módszert. A következő eredmények születtek:

14. ábra: A 8-as modell regressziós statisztikája

|  |  |
| --- | --- |
| Regressziós stat | |
| Többszörös R | 0,816077092 |
| R négyzet | 0,66598182 |
| Korrigált R négyzet | 0,657417251 |
| Standard hiba | 734,3090431 |
| N = megfigyelt adatok | 321 |

Az R2, azaz a modell magyarázó ereje 0,66598182, azaz a biciklibérlések számának alakulását és időbeli szóródását 66,59%-ban sikerült magyarázni. A modell standard hibája 734,3090431, ami a valódi adatpontok eredményértékeinek előrejelzésekor elkövetett átlagos hiba.

A továbbiakban készíttem el a modell vizsgálatát, a releváns változók kiszűrését és a felesleges prediktorok szelektálását. Egyéni szinten a változók relevanciája t-próbával tesztelhető, ahol a p-érték a választott (jellemzően 5%-os) szignifikanciaszint alatt marad, ott legalább 95%-os biztonsággal kijelenthető, hogy az adott változó számunkra fontos.

15. ábra: A 8-as modell t statisztikája

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Koefficiens | Standard hiba | t statisztika | P érték |
| Intercept | 1564,685636 | 203,5760732 | 7,685999695 | 1,98925E-13 |
| Teszt | -2722,374035 | 144,3201047 | -18,86344277 | 1,69945E-53 |
| Hétköznapok | 264,1310058 | 88,76352959 | 2,97567038 | 0,003151966 |
| 24 órás átlag hőm. | 183,6465929 | 8,349447452 | 21,99505943 | 2,11778E-65 |
| Szélirány COS | -57,0517695 | 69,71447322 | -0,818363345 | 0,413774326 |
| Átlagszél | -58,03834129 | 11,68911349 | -4,965161928 | 1,13196E-06 |
| Csapadék | -161,9419293 | 82,47668071 | -1,963487472 | 0,050477413 |
| Felhőzet | -108,8683782 | 24,81611993 | -4,387002421 | 1,57395E-05 |
| Napos órák száma | -64,03539384 | 13,37526504 | -4,787598126 | 2,61108E-06 |

Az adattáblában kijelölt rész mutaja azt a P értéket, amelyik nagyobb, mint a kijelölt konfidenciaszint. Ez esetben a szélirány COS magyarázóváltozónk P értéke alapján nem tekinthető szignifikánsnak.

A megmaradt tényezőkre korlátozva újra felállítjuk a regressziós B modellt.

Ennek az eredménye az alábbi táblázatban van feltüntetve, melyet összehasonlítottam az eredeti modellel.

16.ábra: 2. Modell regressziós statisztikája

|  |  |
| --- | --- |
| Regressziós statisztika | |
| Többszörös R | 0,815637688 |
| R négyzet | 0,665264838 |
| Korrigált R négyzet | 0,657778748 |
| Standard hiba | 733,9215149 |
| N = megfigyelések | 321 |

A modellkészítés során az egyik leggyakrabban használt mutató az R2 determinációs együttható. A több modellváltozat közül a legjobb modell kiválasztsához ezt az értéket nézzük.

Az A modell esetében a R2 az 0,66598182, míg B modell esetén 0,695264838, így az újonnan kapott modellünk jobbnak bizonyult az eredetinél. Ráadásul minden magyarázó változó esetén a P értéke alapján a modellünk szempontjából szignifikáns. Így a második modell végső modellként szolgál a vizsgálatban. Végül ennek a modellnek veszem a részletesebb elemzését.

Az adattáblában található koefficiens értékével tudjuk a változókat ismertetni. A koefficiens azt jelenti a tesztüzem mint magyarázóváltozó esetében, hogy amikor a MOL Bubi tesztüzemben működött, akkor körülbelül 2748 biciklibérléssel volt kevesebb, mint a normál üzemeltetésnél. A teszt magyarázóváltozó emiatt egy nagyon szignifikáns tényező a becslésünk szempontjából.

17.ábra: 2. Modell t statisztikája

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | 1547,452062 | 202,3770988 | 7,646379316 | 2,56084E-13 |
| Tesztüzem | -2748,322091 | 140,7196687 | -19,53047585 | 4,16587E-56 |
| Hétköznapok | 268,9013052 | 88,5251827 | 3,037568486 | 0,00258567 |
| 24 órás átlag hőm. | 183,9409036 | 8,337296375 | 22,06241632 | 1,00252E-65 |
| Átlagszél | -59,64216443 | 11,51757168 | -5,178362774 | 4,01388E-07 |
| Csapadék | -163,393578 | 82,41408666 | -1,982592838 | 0,048286976 |
| Felhőzet | -107,4143186 | 24,73936774 | -4,341837663 | 1,91041E-05 |
| Napos órák száma | -61,75162847 | 13,07399832 | -4,723239742 | 3,50957E-06 |

Úgyanígy, ha a második magyarázóváltozónkat vizsgáljuk, a koefficiens értékén keresztül azt látjuk, hogy a hétköznapokon körülbelül 268 emberrel többen használják a kerékpárokat, mintha hétvége lenne, aminek magyarázata abban rejlik, hogy egyrészről a hétköznapok száma jóval meghaladja a hétvégi napok számát, másrészről pedig az emberek valóban hétköznap használják aktívabban a munkába, iskolába járáshoz a bicikliket.

Az átlaghőmérséklet egy fokos pozítiv változásával nő a biciklibérlések száma is, 183-mal. Negatív arány van a csapadék/felhőzet és a biciklibérlések száma között, ami azt jelenti, hogy a csapadék vagy felhőzet növekedésével csökken a kerépárkölcsönzések száma.

Érdekes módon – szemben a várt eredménnyel - a napos órák száma és a bérlések száma között is negatív korreláció tapasztalható, vagyis a bérlések száma csökken a napos órás számának növekedtével.

* + 1. Asszociációs vizsgálat

Mivel a hétköznapok mint magyarázóváltozók a MOL Bubi használata között negatív korreláció jött ki, ezért megvizsgáltam, hogy van-e különbség az egyes napokon történő biciklibérlések száma között külön-külön napokra lebontva.

A vizsgálathoz feltettem újabb két hipotézist az alábbiak szerint:

* H0: a biciklibérlések száma független a napoktól
* H1: a biciklibérlések száma függ a napoktól

Ennek kiderítésére asszociációs kapcsolatot vizsgáltam, melyhez az alábbi táblázatban található elemszámot, minimumértéket, maximumértéket, átlagot, szórást, mediánt, első kvanitilist és kilencedik kvantilist használtam fel. Ezen értékek alapján elmondható, hogy vasárnap és hétfőn volt a legkisebb a bérlések számának átlaga, a többi napon pedig hasonlóan alakultak a számok. A szórásra is hasonló értékek jöttek ki.

18.ábra: a biciklibérlések alakulása a napok megoszlása szerint

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | N | min | max | átlag | szórás | medián | d1 | d9 |
| 1 | 46 | 210 | 4150 | 1550 | 1213 | 1224 | 421 | 3760 |
| 2 | 46 | 367 | 5705 | 1725 | 1330 | 1351 | 536 | 3691 |
| 3 | 45 | 424 | 5599 | 1718 | 1287 | 1352 | 566 | 3653 |
| 4 | 46 | 284 | 5125 | 1684 | 1190 | 1404 | 507 | 3589 |
| 5 | 46 | 214 | 5136 | 1710 | 1280 | 1343 | 469 | 3615 |
| 6 | 46 | 176 | 5808 | 1577 | 1334 | 1174 | 391 | 3480 |
| 7 | 46 | 0 | 5508 | 1299 | 1166 | 916 | 269 | 2885 |
| Σ | 321 | 0 | 5808 | 1609 | 1255 | 1231 | 440 | 3701 |

Az elemzéshez ANOVA táblát készítettem, melynek eredményei 19. táblázatban láthatóak, ennek alapján számítottam ki az Fszámított értéket, amely a primer kutatásban szereplő Fszámított képlete alapján került kiszámításra.

Az Fkritikus érték 0,69 lett, így a számított érték kisebb, mint a kritikus érték, ezért a nullhipotézisünket fogadjuk el.

A P érték 0,658, azaz 65,8%-ban mutatja, hogy a kerékpárkölcsönzések száma hasonlóan alakul függetlenül a napoktól.

19. ábra: ANOVA tábla

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA | négyzetösszeg | DF | szórás-négyzet | F érték | P érték |
| Csoportok közötti | 6552332 | 6 | 1092055 | 0,69 | 0,658 |
| Csoporton belüli | 497113212 | 314 | 1583163 |  |  |
| Összes | 503665544 | 320 | 1573955 |  |  |

* 1. Prediktív előrejelzés

A kapott modellünk segítségével a kerékpárkölcsönzések alakulását 66,52%-ban sikerült magyarázni a 2014. június 12. és 2015. április 28-a közötti időszakban, naponkénti lebontásban. A lineáris regresszió által előrejelzéseket tehetünk, ha ismerjük a prediktorok értékeit.

Hogyha behelyettesítjük a magyarázóváltozók értékeit az xi helyeire, akkor egy előre meghatározott időpontra meg tudjuk becsülni a kerékpárbérlések várható számát.

1. y=1547,452062-2748,322091x1+268,9013052x2+183,9409036x3-59,64216443x4-163,393578x5-107,4143186x6-61,75162847x7

Összességében elmondható, hogy a MOL Bubi biciklibérlések forgalmát 66,52%-ban 7 releváns változóval sikerült magyaráznunk. A releváns változók által meghatározott értékek megfeleltek a várt eredményeknek.

1. Összegzés

A MOL Bubi a budapesti tömegközlekedést megkönnyítendő jött létre. Bár még nincs egy éve, hogy használatba helyezték a bérelhető kerékpárokat, a budapestiek és a turisták között is nagy népszerűségnek örvend. Egy frissen bevezetett rendszer fejlesztéséhez szükség van az addigi tapasztalatok összegyűjtésére, a felhasználók javaslatainak meghallgatására, valamint olyan, számszerűsíthető felmérésekre, amelyek a rendszer befolyásoló tényezőivel számolnak.

A szakdolgozat ezeket a befolyásoló tényezőket határozta meg (a bevezetőben említett okok miatt, a teljesség igénye nélkül).

A dolgozat egyes részeiben a hipotéziseink az alábbiak voltak:

A primer kutatás egyik célja annak bizonyítása volt, hogy a nemek megoszlása kapcsolatba hozható az elégedettségi mutatóval. Az eredmény egyértelműen azt mutatta, hogy a férfiak és nők biciklibérlésekkel kapcsolatos elégedettségi foka azonos. Ennek a valószínűsége 86,30% volt.

Ugyancsak a primer kutatásban merült fel annak a kérdése, hogy milyen arányban tesznek javaslatot a MOL Bubival kapcsolatban a férfiak és a nők. Az eredmény azt mutatta, hogy a felhasználók nemtől függetlenül, azonos mértékben tesznek javaslatokat.

A korosztályokkal kapcsolatban felmerülő feltételezések szintén a primer kutatás kérdéskörébe tartoztak. Az elégedettség és a korosztályos megoszlás összefüggéseit vizsgáltuk, az eredmény szerint 98,5%-os eséllyel minél fiatalabb valaki annál elégedettebb a szolgáltatással.

Az életkor és a bérletek/jegyek típusainak megvásárlása között összefüggést feltételeztem, a vizsgálatok pedig alátámasztották mindezt. Ebben az esetben a P valószínűségi érték közel 100%-os lett.

A szekunder kutatásban regressziós vizsgálattal dolgoztam. Ebben az esetben olyan külső tényezők hatását vizsgáltam, amelyek befolyással voltak a biciklibérlések számára. Ide soroltam bizonyos időjárási tényezőket valamint dummy változókat, mint a napok megoszlását és a tesztüzemeltetést. Az elemzés eredményeként azt kaptam, hogy szelektálnom kellett a magyarázóváltozók közül: tesztüzemeltetés, a hétköznapok, a csapadék mennyisége, a napos órák száma, a felhőzet és az átlagszél.

A szekunder kutatásban vizsgáltam azt is, hogy van-e összefüggés a kerékpárkölcsönzések száma és a napok megoszlása között. Eredményként azt kaptam, hogy a két tényező egymástól független.

A bevezőben a feltett összetett hipotézist, ,,hogy a biciklibérlők nemek szerinti valamint korcsoportos megoszlása befolyással van az elégedettségre, valamint ugyanez mondható el a szekunder kutatásban vizsgált objektív-külső tényezők és az elégedettségi mutatók között”, a primer és a szekunder kutatás során részenként megvizsgáltam, és arra a következtetésre jutottam, hogy az előfeltevés túlságosan komplex volt, csak a részleteket elemezve kaphattuk meg a megfelelő eredményeket, amelyeket dolgozatomban bemutattam.

1. Irodalomjegyzék

1 OBIS consortium: OBIS kézikönyv (2011): A közösségi kerékpározási rendszerek optimalizálása az európai városokban, 10. Oldal – Magyar nyelvű fordítás a Nemzeti Fejlesztési Minisztérium által

2 Budapesti Közlekedési Központ: A budapesti kerékpáros közlekedés fejlesztési koncepciója (2013) http://molbubi.bkk.hu/docs/BKK\_kerekparos\_koncepcio\_2013\_majus.pdf -6-10 oldal 2015.05.15

3 European Urban Knowledge Network honlapja, Gabriel Metz 2011 december: http://www.eukn.eu/e-library/project/bericht/eventDetail/catch-mr-cooperative-approaches-to-transport-challenges-in-metropolitan-regions/ 2015.05.13.

4 Turizmus online: Huszonegy gyűtjőállomással bővülhet a MOL Bubi rendszer (2014.nov.21): http://www.turizmusonline.hu/kozlekedes/cikk/huszonegy\_gyujtoallomassal\_bovulhet\_a\_mol\_bubi\_rendszer

5 Budapesti Közlekedési Központ: A budapesti kerékpáros közlekedés fejlesztési koncepciója (2013) http://molbubi.bkk.hu/docs/BKK\_kerekparos\_koncepcio\_2013\_majus.pdf - 2015.05.15

6 Klimon Péter: személyes interjú – a Közbringa Kft. Ügyvezetője és egyben a vállalati konzulensem, Dalos Péter: személyes interjú a Budapesti Közlekedési Központ, MOL Bubi üzemeltetésért felelős főmunkatárs

7 A budapesti kerékpáros közlekedés fejlesztési koncepciója - 2013, 32-36. Oldal

8 S. Shaheen, S. Guzman, and H. Zhang. 2010 Transportation Research Record (2010 mácius): BIKESHARING IN EUROPE, THE AMERICAS, AND ASIA: PAST, PRESENT, AND FUTURE 1-2 oldal

9 Klimon Péter: személyes interjú – a Közbringa Kft. Ügyvezetője és egyben a vállalati konzulensem, Dalos Péter: személyes interjú a Budapesti Közlekedési Központ, MOL Bubi üzemeltetésért felelős főmunkatárs

10 Luca di Caspero, Andrea Rendl, Tommaso Urli: Balancing bikesharing system with programming (2014): <http://www.tunnuz.net/documents/digaspero_rendl_urli_constraints2014.pdf> 2015.05.13.

11 Bloomerg Business: Flavia K.-Jackson: Chine races ahead of the pack as bike sharing goes viral (2014. Augusztus 22) : <http://www.bloomberg.com/news/2014-08-22/china-races-ahead-of-the-pack-as-bike-sharing-goes-viral.html> 2015.05.13

12 Jiawei Han & Micheline Kamber (1999): Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems) 3-5. oldal ( Szeptember 7, 1999)

13 Információs Társadalom folyóirat - A 2007. VII. évfolyam 4. szám : Székely Levente: Információgazdaság 146. oldal 2.bekezdés

14 Tikk Domkos előadása(2011): Adatbányászat 12. oldal

15 Technet Magazin: Balássy György : Csákányt a kézbe! – Adatbányászat SQL server 2008 Anaylis Sevices segítségével 23. oldal, 2008 szeptember 5

16 W. Frawley and G. Piatetsky-Shapiro and C. Matheus, Knowledge Discovery in Databases: An Overview. AI Magazine, Fall 1992, 213-228. oldal

17 D. Hand, H. Mannila, P. Smyth: Principles of Data Mining. MIT Press, Cambridge, MA, 2001. ISBN 0-262-08290-X

18 Michael J. A. Berry and Gordon S. Linoff : Data Mining Techniques, 2004 John Wiley & Sons

19 Wikipedia honlapja : Adatbányászat ; utolsó módosítás időpontja: 2013 május 7. A honlap URL címe: <http://hu.wikipedia.org/wiki/Adatbányászat>

20-21 Pete Chapman (NCR), Julian Clinton (SPSS), Randy Kerber (NCR), Thomas Khabaza (SPSS), Thomas Reinartz (DaimlerChrysler), Colin Shearer (SPSS) és Rüdiger Wirth (DaimlerChrysler) CRISP-DM 1.0 Step-by-step Data mining guide 10- 29. oldal, 2000

22 Jiawei Han & Micheline Kamber: Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems) 13-16. oldal ( Szeptember 7, 1999)

23 Dr. Bodon Ferenc : Adatbányászati algoritmusok 94-95. oldal, 2010 február 28.

24 Stephen Baker: The Numerati : A vásárló 65. oldal , 2008 ISBN 978-963-9765-97-9

25 Technet Magazin: Balássy György : Csákányt a kézbe! – Adatbányászat SQL server 2008 Anaylis Sevices segítségével 23. oldal, 2008 szeptember 5

26 Wikipedia honlapja: Döntési fa. A honlap URL címe: hu.wikipedia.org/wiki/Döntési\_fa‎ hivatkozás: Kovács Gyula előadása: Data Mining az Üzleti életben 10. diasor , Széchényi István Egyetem, Győr

27 Kovács Gyula előadása: Data Mining az Üzleti életben 10. diasor , Széchényi István Egyetem, Győr  
28 Tikk Domkos előadása (2011): Adatbányászat 28. Oldal

29 Kovács Gyula előadása: Data Mining az Üzleti életben 10. dia , Széchényi István Egyetem, Győr  
30 Dr. Bodon F. : Adatbányászati algoritmusok 111. utolsó bekezdés és 115. oldal, 2010 február 28

31 Tikk Domkos előadása (2011): Adatbányászat 31. Oldal

32 Wikipedia honlapja : Adatbányászat ; utolsó módosítás időpontja: 2013 május 7. A honlap URL címe: <http://hu.wikipedia.org/wiki/Adatbányászat>

33 BKK Zrt. :Bubi Projekt: Teljesítménymutatók gyűjtési, képzési módszertana 2014

34 Stephen Baker: The Numerati: Bevezető, 2008 ISBN 978-963-9765-97-9

Képeletek(1)(2)(3)(4)(5)(6)(7): A Budapesti Műszaki és Gazdaságtudományi Egyetem, Menedzsment és Vállalatgazdaságtan tanszék: Dr. Jónás Tamás és Dr.Tóth Zsuzsanna Eszter: Gazdaságstatisztika oktatási segédanyag (2014)

1. Mellékletek

1.számú melléklet

Kedves MOL Bubi használó!

Hoang Péter vagyok a Budapesti Műszaki és Gazdaságtudományi Egyetem végzős hallgatója. Jelenleg szakdolgozatomat írom a MOL Bubi felhasználók elégedettségi mértékét befolyásoló tényezőkről. A kérdőív kitöltésével arra szeretnék választ kapni, hogy az egyének mennyire vannak megelégedve a MOL Bubival. A kérdőív kitöltése 2-3 percet vesz igénybe. Köszönöm a segítséget!

1. Az Ön neme

|  |  |  |  |
| --- | --- | --- | --- |
| Férfi |  | Nő |  |

2. Az Ön életkora

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 18-24 |  | 24-34 |  | 34- |  |

3.a. Tanul-e jelenleg?

|  |  |  |  |
| --- | --- | --- | --- |
| Igen |  | Nem |  |

3.b. Dolgozik-e jelenleg?

|  |  |  |  |
| --- | --- | --- | --- |
| Igen |  | Nem |  |

4. Mikor csatlakozott a MOL Bubi-hoz?

|  |
| --- |
|  |

5. Milyen bérlettel rendelkezik jelenleg?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Éves |  | Féléves |  | Heti |  |
| Egyéb: |  | | | | |

6. Milyen mértékben motiválta Önt az alábbi tényezők közül, hogy csatlakozott a MOL Bubihoz? Kérem, 1-5-ig értékelje! 1-es az egyáltalán nem fontos, míg az 5-ös a nagyon fontos.

|  |  |
| --- | --- |
| Pénzt sporolni | 1 2 3 4 5 |
| A közlekedés könnyebb és gyorsabb | 1 2 3 4 5 |
| Szeret biciklizni, élvezetesebb utazás | 1 2 3 4 5 |
| Mozgás, egészségesebb életmód | 1 2 3 4 5 |
| Co2 visszaszorítása | 1 2 3 4 5 |
| Egyéb:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | 1 2 3 4 5 |

7. Milyen céllal használja a MOL Bubit, milyen mértékben fontos az Ön számára? Kérem, 1-5-ig értékelje! 1-es, egyaltalán nem, míg az 5-ös a mindig.

|  |  |
| --- | --- |
| Munkába | 1 2 3 4 5 |
| Iskolába | 1 2 3 4 5 |
| Találkozókra | 1 2 3 4 5 |
| Étterembe | 1 2 3 4 5 |
| Edzésre | 1 2 3 4 5 |
| Egyéb:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | 1 2 3 4 5 |

8. Hétköznaponként vagy inkább hétvégénként használja a MOL Bubit?

|  |  |  |  |
| --- | --- | --- | --- |
| Hétköznaponként |  | Hétvégénként |  |

9. Milyen gyakran használ más alternatív közlekedési formát a MOL Bubi mellett? Az 1-es a nagyon ritkán, míg az 5-ös a nagyon sűrűn.

|  |  |
| --- | --- |
| Busz | 1 2 3 4 5 |
| Taxi | 1 2 3 4 5 |
| Metro | 1 2 3 4 5 |
| Saját kerékpár | 1 2 3 4 5 |
| Gyaloglás | 1 2 3 4 5 |
| Egyéb:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | 1 2 3 4 5 |

10. Hogyan értékelné az alábbi MOL Bubi szolgáltatásokat egy 1-től 5-ös skálán. Az 1-es az egyáltalán nem elégedett, míg az 5-ös a teljes mértékben elégedett.

|  |  |
| --- | --- |
| Térkép | 1 2 3 4 5 |
| Éjszakai világítás | 1 2 3 4 5 |
| Állomások biztonsága | 1 2 3 4 5 |
| MOL Bubi honlap | 1 2 3 4 5 |
| Bicikli | 1 2 3 4 5 |
| Call Center | 1 2 3 4 5 |

11. 1-5-ös skálán hogyan értékelné összességében a MOL Bubi által nyújtott szolgáltatásokat? Az 1-es az egyáltalán nem elégedett, míg az 5-ös a teljes mértékben elégedett.

|  |
| --- |
| 1 2 3 4 5 |

12. Az alábbiak közül milyen javaslatokat tenne a MOL Bubival kapcsolatban?

|  |  |
| --- | --- |
| Még több dokkoló a már a meglévő állomásokban |  |
| Még több állomás külső kerületekben |  |
| Még több állomás a belvárosban |  |
| Még több állomás metro közelében |  |
| Még több állomás a már meglévő területeken, ahol a MOL Bubi van |  |
| Kibővíteni több állomással, ahol nincsenek MOL Bubi állomások |  |

13. Tapasztalt bármilyen egyéb problémát, vagy tenne egyéb javaslatokat a szolgáltatás javítása érdekében?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. számú melléklet

A Budapesti Műszaki és Gazdaságtudományi Egyetem, Menedzsment és Vállalatgazdaságtan tanszék: Dr. Jónás Tamás és Dr.Tóth Zsuzsanna Eszter: Gazdaságstatisztika oktatási segédanyag: képletgyűjtemény (2014)

