**Budapesti Corvinus Egyetem**

Szakdolgozat Projekt dolgozat

2020/21/tavasz



Szakdolgozat 1. fejezet

avagy az airbnb idősoros értelmezése 2015 és 2020 között

Készítette: Meleg András

*Projekt 1*  kurzus

Oktató: Sándor Máté Csaba

Tartalomjegyzék

[1. Elméleti áttekintés 3](#_Toc74514332)

[2. Adatok elemzéshez való előkészítése 4](#_Toc74514333)

[Adatok áttekintése, vizsgálata 5](#_Toc74514334)

[Scrapelés előkészítése és adatok scrapelése 5](#_Toc74514335)

[Kereszttáblás adatok kezeléséhez szükséges függvényműveletek előkészítése, és kereszttáblás adatokon a függvénytranszformációk végrehajtása 5](#_Toc74514336)

[3. Adatok elemzése és diszkusszió 6](#_Toc74514337)

[H0: Létezik szezonális periodicitás: 6](#_Toc74514338)

[H1: az olcsóbb szállások kínálata lassabban követi az árak emelkedését, mint a drágább szállásoké: 6](#_Toc74514339)

[H2: A „szezonban” több single-host adja drágábban a lakását: 6](#_Toc74514340)

[Továbbfejlesztési lehetőségek: 7](#_Toc74514341)

[Felhasznált irodalom 8](#_Toc74514342)

# Elméleti áttekintés

Ahhoz, hogy az Airbnb adatok elemzésekor megértsük, hogy pontosan milyen közgazdasági kérdéseket érdemes vizsálni, elengedhetetlen a tágabban vett szakirodalom átolvasása és megértése. A *gig economy* egy folyamatosan fejlődő és egyre népszerűbbé váló business-forma, amelyet alapvetően a közösségiség, illetve ennek ígérete tesz a felhasználók számára népszerűvé. Az Airbnb kutatások bibliometrikus vizsgálata azt mutatja, hogy a tudomány is hasonlóan gondolkodik erről, a legfontosabb mozgatórugói az airbnb-vel kapcsolatos kutatásoknak a bizalom, az innováció, és az elégedettség (Nunez-Tabales et al., 2020). Mindenképp érdemes észrevenni ezen a ponton, hogy az előbb felsorolt három fogalom viszonylag jól kvantifikálható, és sok tudományos elmélet létezik a leírásukra, ezért rögtön megállapítható, hogy az Airbnb különösen értékes táptalaja lehet az Airbnb-vel kapcsolatos kutatásoknak.

Az Airbnb-hez hasonló sharing economy modellek az elmúlt egy évben indultak jelentős fejlődésnek. Mivel ez a dolgozat egyfajta időkapszulaként íródik, remélem a rohamosan innoválódó magyar fővárosban olvasva ezeket a sorokat legalábbis megmosolyogtató lesz visszaolvasni a példákat, amik a mindennapokban körülvesznek minket éppen: a Lime rollerek, különböző gépjármű-sharing rendszerek, de tulajdonképpen a tágabban értelmezett definícióba minden olyan közgazdasági rendszert is belevehetünk, ami nem a klasszikus értelemben vett központi újraelosztási rendszerekben gondolkozik (pl. adók, állam, és segélyek rendszere), hanem arra helyezi a hangsúlyt, hogy az interneten áramló információk segítségével a különböző szolgáltatások nem egy helyen, és egy időben, permanensen birtokolva, hanem csak időlegesen használva vegyük igénybe. Mindez magát a munkaerőpacot is jelentősen átalakítja, ahogy arra többen is remekül rámutatnak (Aeppel, 2016; Ashford et al., 2018).

Az előző bekezdésben ismertetett *sharing economy* modell definíció munkaerőpiacra gyakorolt hatásaival foglalkoznak Dugganék (2021) is, illetve azzal, hogyan alakítja át mindez a társadalmat. A szerzők könyvükben öt fő faktort állapítanak meg, amiben a *gig economyk* eltérnek a klasszikus értelemben vett közgazdasági rendszerektől. Ezen öt faktorból az első maga a modell eltérő volta, ami nagyobb fokú szabadságot és függetlenséget biztosít a munkásoknak, a második az ebből keletkező munkaklasszifikációs problémák, a harmadik a technológia szingergiája az ilyen típusú rendszerekkel, a negyedik az eltérő gazdasági élmények és mentalitást amit a szolgáltatásban részt vevők tapasztalnak, és végül az ötödik a fenntarthatóság kérdésköre, vagyis az, hogy egészen más jövőbeli elvárásokkal viseltet a társadalom a *gig economyk* felé (Duggan et al., 2021).

Mindez azt is jelenti, hogy még mielőtt el is kezdenénk bármilyen, airbnb-vel kapcsolatos kvantitatív elemzésbe, megfigyelhetjük azt, amit az elemző közgazdaságtani gondolkodásmód nulladik tételének is nevezhetnénk: *„A közgazdaságtanban nincsenek értékítéletek, minden morális kérdésre válaszolhatjuk, hogy attól függ.”* De pontosan mit is jelent ez? Megfogalmazásom szerint nagyjából azt, hogy az airbnb-nek egyéni szintű (fogyasztó bérlő és szolgáltatást biztosító szállasadók) és rendszerszintű (kereslet-kínálat egyensúlya, és annak az egész társadalmi rendszerre gyakorolt hatásának) vannak pozitív és negatív hasznai. Ezt az álláspontomat tovább erősíti az is, hogy ehhez hasonló kiindulópont mellett foglalta össze az Airbnb elméleti előnyeit és hátrányait Bivens (2019). Ő végül arra a konklúzióra jut egyébként, hogy az Airbnb-nek hosszú távon bele kellene tagozódnia a standard szállásszolgáltató iparba, adók, kötelezettségek, és jutatások terén (Bivens, 2019).

Ez azonban nem jelenti azt, hogy hosszú ha ez így is történne (ahogy azt sok helyen igyekeznek szabályozásokkal elérni), az megváltoztatná az airbnb piac gig economy jellegét, tehát továbbra is egy viszonylag szabadabb, és kötetlen rendszerről beszélhetünk, és ez egy fontos alapelv a továbbiakban is: az airbnb szabályozása, adók, jutalékok, kötelezettségek kivetése nem jelenti azt, hogy a gig economy elvesztenék gig economy jellegüket, ezért közgazdasági elemzésünk szempontjából nagyon könnyen elképzelhető, hogy különböző szabályozási próbálkozások csak rövid-és középtávon tudnak hatással bírni a szálláspiac szerkezetére.

Az Airbnb-vel kapcsolatos friss kutatások egyik kiemelkedően érdekes összefoglalása és példaelemzése, melyből sok ihlet merítődött ebbe a dolgozatba is, az Osterwalder Business Model Canvast használja fel a sharing economy business modell elemzéséhez (Chua et al., 2019). Ez a módszertan alkalmas volt arra, hogy meghatározza a Fülöp-szigetek, Szingapúr, és Dél-Korea térségében levő Airbnb felhasználók demográfiai profilját, és a felhasználók feltételezet viselkedési mecjhanizmusát. Végső soron arra az eredményre jutnak, hogy a legnagyobb motiváció a modellre való áttérésre a hagyományos gazdasági modellű szolgáltatásokról legtöbbször pénzügyi, de jelentős szerepet tölt be a környezettudatosság is. Mindez pedig összefüggésben van azzal, hogy megváltoztak a fiatalok utazási szokásai, és a több utazást nagyobb igényt teremt a helyi kultúrákra is (Chua et al., 2019).

# Adatok elemzéshez való előkészítése

Ahhoz, hogy jobban megérthessük az airbnb jelenséget, további kutatásokra van szükség, éppen ezért vált relevánssá az egyik legfiatalosabb, legnagyobb, kulturálisan legsokszínűbb európai város, Berlin városának elemzése. Az elemzéshez az insideairbnb honlapról gyűjtöttem az adatokat, és Python programmal (3.7-es verzió) állítottam össze idősorokat, hogy azokból néhány egyszerűbb, átfogó kimutatást hozzak létre, ami egy viszonylagos képet mutathatott az Airbnb piac változásáról Berlin városában, 2015 és 2021 között. (Insideairbnb, 2021; Van Rossum & Drake, 2009)

## Adatok áttekintése, vizsgálata

A félév elején az adatok áttekintését kaptuk házi feladatnak, én az Inside Airbnb Berlinre vonatkozó adatait választottam, ahol megállapítottam, hogy az adatokat nagy pontossággal tartalmazó kereszttáblákkal foglalkozhatok: bérelhető szállásonként ID, host\_ID, ár, kihasználtság, szállásadó hány szállást hirdet, összes értékelés száma, havonta kapott új értékelések száma voltak a fő adatoszlopok.

## Scrapelés előkészítése és adatok scrapelése

Mivel az *insideairbnb* honlap több évre visszamenőleg elérhető csv fileokat tartalmazott, ezért úgy döntöttem, hogy a *Selenium* Python csomag segítségével scrapelni fogom a kereszttáblához az adott időközönkénti paneladatokat. *BeautifulSoup* Python csomag segítségével megkerestem az összes linket a honlapon, és bemeneti filter változóként *„Berlint”* adtam meg – ez egy hardcode input, ha átállítja a tisztelt olvasó pl. *„Amszterdamra”*, akkor a program a további elemzést Amszterdamra fogja automatikusan megcsinálni. Ezután a datatype (listings) a datayear, és a datamonth változók segítségével kiszűrtem a megfelelő linkeket, és ezeken iterálva azokra a linkekre kattintattam rá a *Seleniumot*, amik épp tartalmazták a megfelelő datatype, datayear, datamonth kombinációt. Így 2016 és 2021 között kb. 50 hónapnyi adatot gyűjtöttem össze, gyakorlatilag havi bontásban, 2017 és 2018 tavasza között azonban hiányos volt az adatsor. Ezzel el is készült a nyers adathalmazom.

## Kereszttáblás adatok kezeléséhez szükséges függvényműveletek előkészítése, és kereszttáblás adatokon a függvénytranszformációk végrehajtása

A nyers adathalmazon a *Pandas* Python csomag segítségével végeztem el a tényleges elemzést, az elvégzett műveleteket az egyes számú mellékletben összegzem. A mellékletben összegzett adattisztító és csoportosító műveleteket lefuttattam a scrapelés során előkészített csv fileokon végigiterálva, és három felé rendeztem az adatokat. (1) Metadata, ahol vizsgált időszakonként szedtem össze az összes adat outputjait, illetve (2) single\_hostok, és (3) multi\_hostok esetére külön adatbázist hoztam létre. .

A transzformált adatok felfedező elemzése, és ezek alapján H1, H2 megfogalmazása az alábbiak:

# Adatok elemzése és diszkusszió

## H0: Létezik szezonális periodicitás:

Az adatok alapján nagyjából kirajzolódik a nyaranta bekövetkező szezonálitási hatás. Sajnos az adatsor nem feltétlenül tartalmaz minden elérhető időpontot. Emiatt azt nem merem kijelenteni, hogy a hipotézis egyértelműen elfogadható, ehhez több év, több nyári szezon kiugró adataira lenne valószínűleg szükség. (Lásd még: II/3. ábra)

Ezután egy korrelációs táblát készítettem az adatok leíró statisztikábiól, arra voltam kíváncsi, hogy hol lelhetőek fel erős multikorrelációk a kereszttáblás adatokban. Sajnos az összegző leíró statisztikákból (átlag, medián, szórás, stb.) nem vehető ki egyértelmű erős korreláció, ezért végül az árak vizsgálatát választottam következő lépésnek. (Lásd még: II/4. ábra)

## H1: az olcsóbb szállások kínálata lassabban követi az árak emelkedését, mint a drágább szállásoké:

Ennek vizsgálatához az árakat sorbarendeztem, és kvartilisenkénti változását vizsgáltam meg. Ezeket a II. Számú melléklet 1-es és 2-tes számú ábráin csatoltam.

Bár a II/2 alapján nem lettünk sokkal okosabbak, de annyit látunk, hogy a magas árak jobban felhúzzák az átlagot, mint az alacsony árak. Továbbá a harmadik és az első kvartilis korrelál a legkevésbé, a drágább szállások drágulása nem feltétlenül okozza az olcsóbb szállások drágulását is. 4 tizedesjegyre kiszámolva ezt, a q1-q3 korreláció 0,7471, és a p értéke 0. (4 tizedes pontossággal), tehát meglehetősen biztos az eredményünk.

## H2: A „szezonban” több single-host adja drágábban a lakását:

Mivel a kumulált adatok kevés információt tartalmaztak, úgy döntöttem, hogy megpróbálom az adatok változását vizsgálni. Ehhez összesítve, single-host, és multi-host eseteket szétválasztva az alábbi relatív változásokat számoltam ki: (1) Szállások számának arányos változása, (2) Szállások átlagárának arányos változása, (3) Szállások friss értékeléseinek arányos változása.

Ezek közül aztán ismét kiderült, hogy nem igazán van sok korreláció az adatok között. (Lásd: II/5. ábra). Érdekesség viszont az, hogy mennyire nem mozog együtt a single-multi host szállások növekedésének / csökkenésének az üteme.

4 tizedesjegy pontossággal számolva 0,5654 a korrelációja, p=0 mellett a single-multi hostok korrelációja. Ez erős korrelációnak is mondható, de látható, hogy az összesített adatoktól egyébként ezek távol vannak. Azonban ez csak egy érdekesség, a hipotézis, amivel éltem ugyanis az, hogy pozitív a korreláció az árak növekedése, és az új single hostok megjelenése között a piacon. Ez azt jelentené ugyanis, hogy a keresletre tudatosan reagálnak azok a hirdetők, akik nem „professzionálisan” hirdetnek. Az új belépők hatása ugyanis nem ellensúlyozza a kereslet növekedésének a hatását, ezért még akkor is emelkedhetnek az árak, ha épp többen lépnek be a piacra.

Ezt a 0.47-es korrelációt 0,0234-es p érték mellett sikerült megállapítani, tehát viszonylag nagy biztonsággal. Hozzátartozik azonban ehhez az, hogy itt eredetileg nem találtam korrelációt, de levágtam az adatsor elejét, és csak azt az időszakot néztem, amikortól már minden adat rendelkezésre állt. ( 2018 májusától.)

## Továbbfejlesztési lehetőségek:

A fenti rövid összefoglaló csak egy nagyon szűk keresztmetszetét tudta bemutatni annak, hogy mivel is foglalkozott a felfedező adatelemzés, amivel a félév során foglalkoztam, és azt hiszem, főleg a kevésbé sikeres részét. A III. Mellékletként csatolt jupyter notebookban láthatóan próbálkoztam gini-együtthatóval is, és még kb. 15-20 egyéb korrelációval is. Természetesen lenne még tér mélyebb, átfogóbb elemzésre, amivel szívesen foglalkozom a jövőben.

A fenti módszerekhez képest előrelépés lehet a leíró statisztikai adatok helyett konkrét apartmanok idősorait vizsgálni. Ehhez egy 70-30 arányú test-train split felosztás tűnik a legideálisabbnak. Természetesen látható, hogy a makroadatokból is kimutathatóak érdekes összefüggések, de ez az adatelemzés átfogóságához képest csak nagyon kis mértékben fordul elő. Emellett az adatok alaposabb tisztítása lehet, hogy javítaná az eredményeket. A reprodukálhatóság, és az automatizáltság, és ezáltal a kutatás elméleti alapokon való nyugvása viszont nagyon hasznos és fontos része a munkának.

# Felhasznált irodalom

Aeppel, T. (2016). *How the On-Demand/Gig Economy is Redefining Work*. https://ide.mit.edu/wp-content/uploads/2016/06/IDE-Research-Brief-v06.pdf

Ashford, S. J., Caza, B. B., & Reid, E. M. (2018). From surviving to thriving in the gig economy: A research agenda for individuals in the new world of work. *Research in Organizational Behavior*, *38*, 23–41. https://doi.org/https://doi.org/10.1016/j.riob.2018.11.001

Bivens, J. (2019). *The economic costs and benefits of Airbnb*. Epi.Org. https://www.epi.org/publication/the-economic-costs-and-benefits-of-airbnb-no-reason-for-local-policymakers-to-let-airbnb-bypass-tax-or-regulatory-obligations/

Chua, E. L., Chiu, J. L., & Bool, N. C. (2019). Sharing Economy: An Analysis of Airbnb Business Model and the Factors that Influence Consumer Adoption. *Review of Integrative Business and Economic Research*, *8*(2), 19–35.

Duggan, J., McDonell, A., Sherman, U., & Carbery, R. (2021). *Work in the Gig Economy: A research overview*. Routledge.

Insideairbnb. (2021). *Inside Airbnb*. http://insideairbnb.com/get-the-data.html

Nunez-Tabales, J., Solano-Sanchez, M. A., & Caridad-y-Lopez-del-Rio, L. (2020). *Ten Years of Airbnb Phenomenon Research: A Bibliometric Approach*.

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*.

# I. Számú melléklet.

*Az adatok megfelelő feldolgozásához az alábbi függvényműveleteket írtam meg:*

**describe\_numberic\_columns:** a számadatok leíró statisztikai jellemzőit gyűjti össze, pl. minimum éjszakák átlaga, Q1, Q3, szórás

**bardiagram\_dtype\_review**: összeszedi a bevitt kategória adatoszlopról, hogy milyen egyedi elemeket tartalmaz, valamint oszlopdiagramot plotol az egyedi kategóriák egyes darabszámaiból, és visszaadja egy listában az egyedi kategóriákat listáját a lista első elemeként, és azonos sorrendben a darabszámokat a lista második elemként

**gethist\_prices\_return**: egy kereszttáblás adatsor inputból visszaadja az előforduló árakat listában, illetve kirajzolja az árak hisztogramjait

**gini:** megadja a Gini együtthatóját a bevitt adatoknak, például a szállások árainak

**lorenz görbe**: bevitt adatokból Lorenz-görbét plotol (ezt nem használtam végül)

**plotprices:** a gethist\_prices outputját belerakván, kiadja az árak kumulatív eloszlását ábrázoló függvényt, gyakorlatilag egy Lorenz-görbét. Pl.:

**collect\_data\_from\_timestamp\_df**: ez a legfontosabb függvény, a fenti függvények használatával, egy kereszttáblából és egy megadott időpontból, visszaadja az időpontot, a leíró statisztikai makroadatokat, a szállások típusait és darabszámait, a kerületek neveit és kerületenkénti darabszámot, az árakat ID szerint rendezve, az árakat növekvő sorrendbe rendezve, és a Gini együtthatót.

# II. számú melléklet

1. ábra: a leíró statisztikák változása (saját készítés)

A screenshot of a computer

Description automatically generated with low confidence

1. Ábra: az összesített árak leíró statisztikáinak korrelációs mátrixa

# Chart, treemap chart Description automatically generated

1. Ábra: Egy szállásra eső havi értékelések számának változása

Chart, line chart

Description automatically generated

1. Ábra: Korrelációs matrix multikollienaritás keresése céljából

**Chart, treemap chart

Description automatically generated**

1. Ábra: Új szállások önmagukhoz viszonyított arányának típusonkénti korrelációja:

Chart

Description automatically generated with medium confidence

# III. Melléklet: az elemzéshez használt Jupyter Notebook

Elérhető: <https://drive.google.com/file/d/1Ysd_9m948ufJ0YVeSAtPs2sYM_xyF2N0/view?usp=sharing>