

Az elemzéshez felhasznált adathalmazt amellyel dolgoztunk, több mint 49 200 nyilvános Facebook profilból kinyert 2015-ös évi adat jelentette. A vizsgált alanyok 3 megye (Hargita, Kovászna illetve Maros ) 17  különböző városából származtak, emellett Kolozs megyéből Kolozsvárról.

Ezekből 9500+ profilon szerepelt munkaköri leírás, melyeket 19 csoportba soroltunk, mint például: Orvos/Gyógyszerész, Tanár/Kutató, Informatikus, Fodrász/Kozmetikus stb… Továbbá több, mint 22 000 facebookozó tüntette fel nyilvános adatlapján azt az intézményt, ahol tanulmányait folytatta, melyeket 50+ kategóriába soroltunk be.

A feldolgozást az adatok tisztításával kezdtük, ami annyit jelentett, hogy kisbetűsítettük az adatokat, valamint ékezetmentesítettünk a könnyebb, hatékonyabb feldolgozás érdekében. A különböző kategóriák kialakítása időigényes feladatnak bizonyult, ugyanis a Facebook profilok adatai változatos nyelveken, főként magyarul, románul, angolul szerepeltek, ám találkoztunk számunkra ismeretlen, ázsiai írásjelekkel is.

A városok, intézmények, munkaköri leírások különböző megnevezései, az irreleváns információk (pl munkahely: diák) és az elírások is komoly problémát jelentettek. A leggyakoribbakat sikerült kiszűrni Python dictionary-k építésével és rendezésével: a kulcs jelentette az adatot, az értéke meg, hogy hányan írták ugyanazt a jellemzőt az illető mezőhöz. A következőket vizsgáltuk: születési hely, tanulmányok, munkakör, lakhely. A kód ezeken a linkeken tekinthető meg:

<https://github.com/retyidoro/data_analysis_competition/blob/master/Adatfeldolgoz%C3%A1s%20%20-%20kib%C5%91v%C3%ADtve.ipynb>

<https://github.com/retyidoro/data_analysis_competition/blob/master/Feldolgozas2-livesin.ipynb>

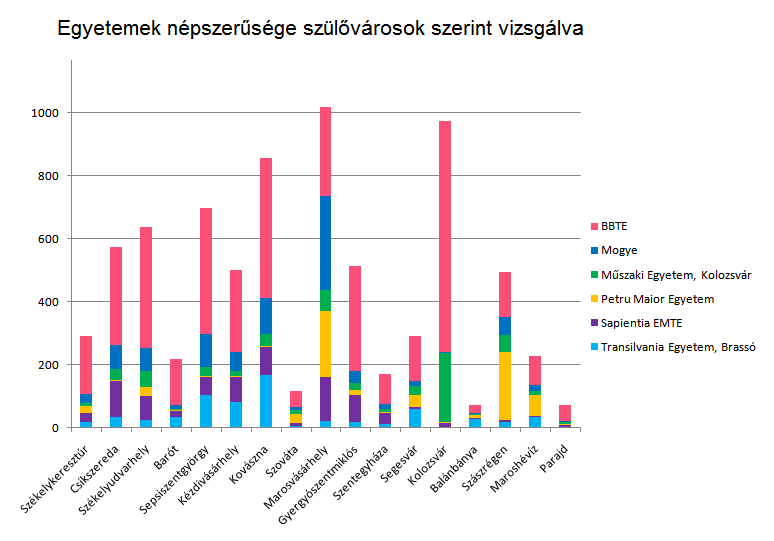
Érdekesnek találtuk volna vizsgálni az egyes munkakörökön belül a nemek eloszlását. A vezeték- és családnév mezők vizsgálata adott volna lehetőséget erre. Az észrevétel alapján, miszerint a vezetéknevek között elég ritka a női név, arra a következtetésre jutottunk, hogy elég jó becslést kaphatunk a nemek arányára, ha a női neveket szűrjük ki, egy előre összeállított halmaz alapján, melyek a legelterjedtebb magyar és román női neveket tartalmazzák. Ám tudtuk, hogy a neveket fiktív nevekre cserélték. Már az elkészített munkakörök esetén is figyelemfelkeltő volt a férfiak és nők közel egyenlő eloszlása, végül a házastárs mezők “véletlenszerűsége” beigazolta a sejtést, hogy nem tartották meg az egyedek nemét a nevek megváltoztatásánál.

Utolsó lépésben a kinyert adatok vizualizációja következett, melyhez a tisztított adathalmazunkat használtuk. Készítettünk egy statikus és egy interaktív vizualizációt is. Vizsgáltuk a különböző munkakörök városonkénti eloszlását, az egyes városok lakóinak továbbtanulási szokásait, valamint azt is, hogy az egyes egyetemeken végzett hallgatók a továbbiakban milyen munkakörökben tudnak elhelyezkedni, végül pedig a munkakörök és elvándorlás kapcsolatát.

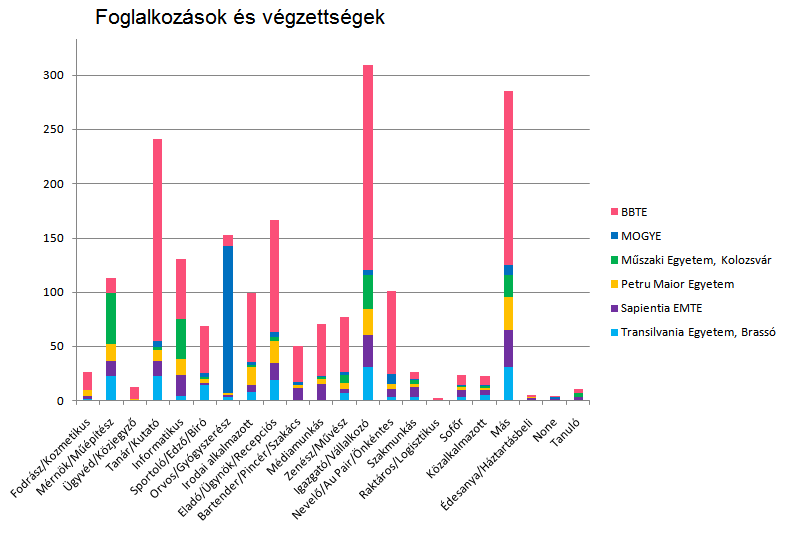
A legnépszerűbb munkakörnek az Igazgató/Vállalkozó csoport számított a “Más” után, amely a nem annyira elterjedt kategóriákat és az irreleváns adatokkal kitöltött mezőket is magába foglalja. Ez után következett az Eladó/Ügynök/Recepciósok csoportja, majd a Tanárok/Kutatók.

Ahogy sejthető volt, a legmeghatározóbb intézménynek számít a Babeș-Bolyai Tudományegyetem, a változatos képzési kínálatának köszönhetően. Ezt követi a Marosvásárhelyi Orvosi és Gyógyszerészeti Egyetem, valamint a Kolozsvári Műszaki Egyetem.

Megfigyelhető, hogy sokan választanak a születési helyükhöz közel eső egyetemet, például Marosvásárhelyről és környékéről a Marosvásárhelyi Orvosi és Gyógyszerészeti Egyetemet, a Petru Maior Egyetemet és a Sapientia Erdélyi Magyar Tudományegyetemet, Kolozsvárról a Babeș-Bolyai Tudományegyetemet és a Műszaki Egyetemet, Brassó környékéről a Transilvania Egyetemet.



A foglalkozások és vézgettségek közötti kapcsolat is szépen tükrözi, hogy aBabeș-Bolyai Tudományegyetemen végzettek sokszínű munkakörökben helyezkednek el (de főleg Igazgatói/Vállalkozói, Tanár/Kutatói pozícióban), míg a speciális képzést igénylő szakmákhoz főleg a megfelelő egyetem képzési programját választják (pl: Orvosok/Gyógyszerészek a Marosvásárhelyi Orvosi és Gyógyszerészeti Egyetemet, Mérnökök/Műépítészek a Kolozsvári Műszaki Egyetemet).

****

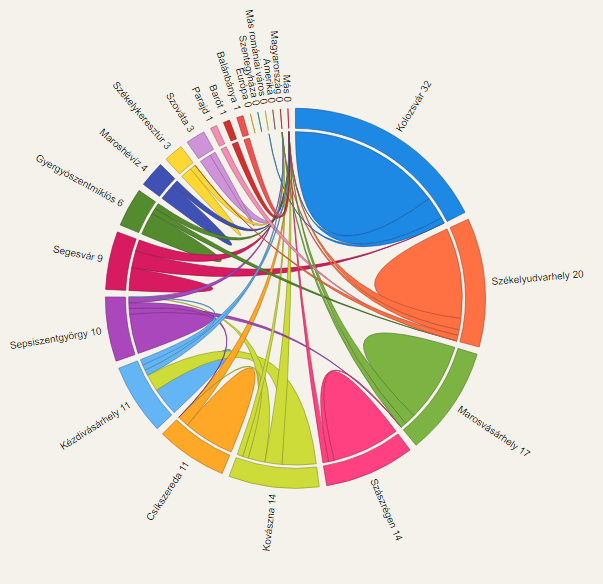
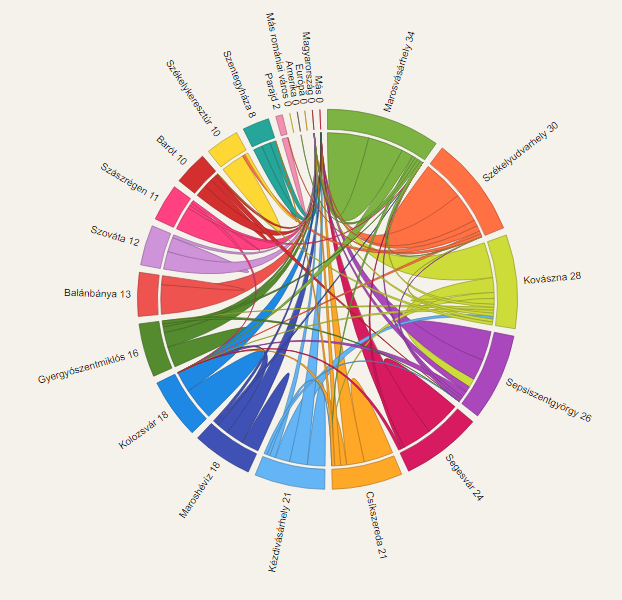
A teljes statikus vizualizációt itt lehet megtekinteni: <https://github.com/retyidoro/data_analysis_competition/blob/gh-pages/Vizualiz%C3%A1ci%C3%B3%202%20-%20d3js/viz.pdf>

Végül a munkakörök és elvándorlás közötti kapcsolatot figyeltük meg, azoknál az egyedeknél, ahol szerepelt a születési hely, lakhely és munkaköri leírás is. A lakhely mezők esetén a jelentősebb falvakat a hozzá közeli város osztályába soroltuk. A nagyobb székelyföldi városokon kívül a következő további osztályokat alakítottunk ki: Magyarország, más romániai város, Európa, és Amerika. Az interaktív vizualizáció itt tekinthető meg: <https://retyidoro.github.io/data_analysis_competition/>

Általános megfigyelés, hogy szinte minden munkakör képviselői közül sokan vándorolnak külföldre, a kisebb városokból meg nagyobb székelyföldi városokba, vagy éppen a közeli kisebb városokba. Például a háromszéki városok esetén a legszembetűnőbb a Kovászna, Kézdivásárhely, Sepsiszentgyörgy és Barót közötti kölcsönös átvándorlási jelenség. Ezek közül is Kovászna a legfeltűnőbb, (ahonnan körülbelül annyi adat származik, mint Kolozsvárról), itt ugyanis kevesen maradnak helyiek, viszont érkeznek más városokból, nem úgy mint Régen és Segesvár esetén, ahonnan főleg elvándorolnak.

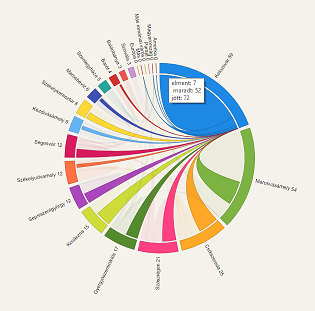
Az egyes munkakörökhöz tartozó sűrűbb diagramok a szakterületen belüli elvándorlás mértékét jelzik, mint például a Bartender/Pincér/Szakács, vagy Tanár/Kutató kategóriák esetén. Más szakmáknál meg nagyobb mértékű a szülővárosban való elhelyezkedés, mint például a Fodrász/Kozmetikusok vagy Irodai alkalmazottak esetében. Ám figyelnünk kell a feltüntetett számokat is, amik azt jelzik, hogy hányan származnak abból a városból az adott munkakörön belül.

**Bartender/Pincér/Szakács                                                           Fodrász/Kozmetikus**



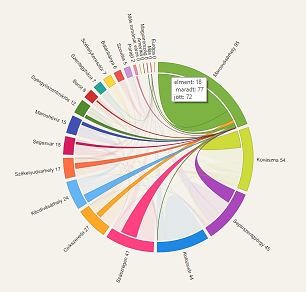
**Informatikusok:**

* Sokan maradnak a kolozsváriak és marosvásárhelyiek közül szülővárosukban.
* Ugyanakkor a kisebb városok informatikusai inkább elvándorolnak a nagyvárosok felé, kevesen maradnak otthon. Szívesen mennek Kolozsvárra és Vásárhelyre dolgozni.
* Továbbá megfigyelhető, hogy Szentgyörgy és Udvarhely informatikusainak körülbelül fele talál munkát szülővárosában.

****

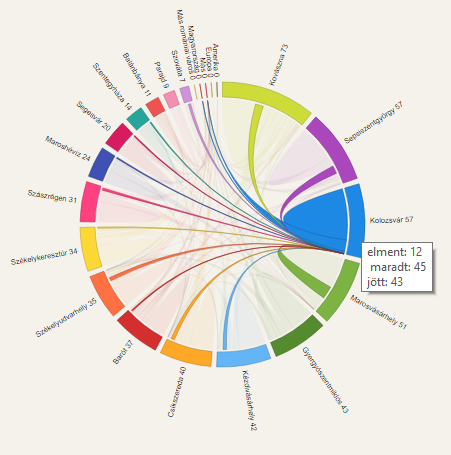
**Orvos/Gyógyszerész:**

* A sejtésünk beigazolódik, hiszen orvosok és gyógyszerészek szempontjából Marosvásárhely a leggazdagabb, ide vándorolnak Erdély kisebb városaiból is a legtöbben

****

**Tanár/Kutató**

* Sűrű diagram, tehát viszonylag nagymértékű vándorlásról van szó a szakmán belül.
* Ám sokan helyezkednek el szülőföldjükön is ezen a pályán.
* Kolozsvár és Marosvásárhely ismét befogadóként szerepel, hiszen ide többen jönnek.



**Összegzés**

Érdekes volt kipróbálni magunkat az adatvizualizáció terén, szívesen dolgoztunk a pályázaton, hiszen szeretjük a kihívásokat. Az általunk készített diagramok is rengeteg információt tartogatnak, de ezeken kívül még sok összefüggést lehetne figyelni, például a nemek és munkakörök kapcsolata vagy a tanulmányok és elvándorlás közti összefüggés, hogy el tudnak-e helyezkedni itthon egyetem után az emberek, vagy inkább elvándorolnak. Továbbá érdekesnek találtuk a d3.js nyújtotta lehetőségeket is. Köszönjük szépen a lehetőséget!

Földi Zsuzsánna-Boglárka & Rétyi Dorottya