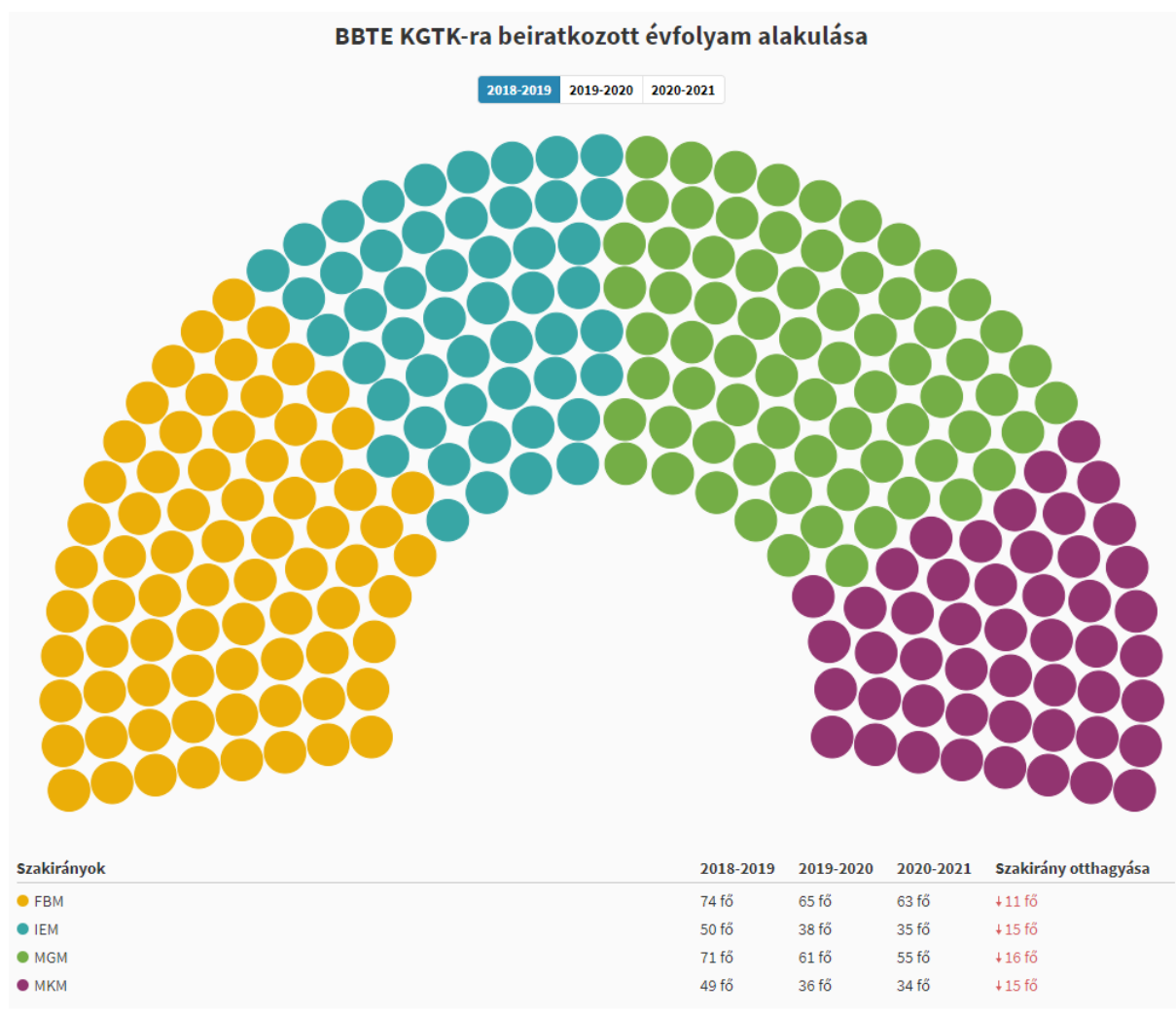


A BBTE KGTK magyar tagozatára 2018-ban beiratkozott hallgatókról



Farkas Bernadett

TÉMAVÁLASZTÁS

Számomra a témaválasztás szokott minden munkámnál a legnehezebb lenni, mert minden apró részletre szeretek gondolni és szeretem a különlegességeket is. Úgyhogy fordítva kezdtem el gondolkozni. Elképzeltem milyen vizualizációkat szeretnék elkészíteni. Nagyon megtetszett a SzékelyData blogon a Szabad Európa által készített parlamentes infografika, illetve az általam bemutatott repülőjáratos Arc map. Mindkét vizualizáció az egyszerűségéért és látványáért beszélt. Mivel a vizualizációkat hamarabb elterveztem és hozzuk megfelelő adatsorokra volt szükségem ezért döntöttem a saját téma mellett. Épp a Netflix-en néztem egy filmet (The Social Dilemma), amikor beugrott az ötlet, hogy Facebook-ról megpróbálnék adatokat gyűjteni, és onnan egy csoport adatait lenne érdemes kinyerni, mert viszonylag nagy minta. Lassan kirajzolódott a kép, hogy csinálhatnám a saját évfolyamomról a projekteket, a Kozgaz2018 Facebook csoport alapján. Illetve elképzeltem azt is, hogy a leíró adataim mellett lenne hely- és akár idődimenzióm is.

ADATGYŰJTÉS

Az adatgyűjtés nálam három fázisból állt. Először ki kellett derítenem melyik szakirányon hányan kezdtek, illetve kik mentek el az egyetemi évek során. Mivel törvény szerint semmilyen hivatalos dokumentumban nem szerepelhet a hallgatók neve csak a nr. matricol-ja, így kisebb kutatást igényelt ez a rész. Adatokat gyűjtöttem a hivatalos dokumentumok, eddigi parciális és szessziós jegyek alapján, illetve minden szakról kérdezősködtem a csoportfelelősöktől, arról ki mikor hagyta abba az egyetemet vagy váltott szakot. Így készült el egy összesítő lista évfolyamtársaimról. Ezután egy kémprogramra volt szükségem, ami minél hatékonyabb és kevesebb munkát is igényel kódolni. Addig keresgéltem, találtam egy adatkeresési és kinyerési alkalmazást (LetsExtract Email Studio), amely főként cégeknek szól, akik ügyfeleik médiaprofiljaikból szeretnék adatokat kinyerni. Sajnos a demoverzió elég butított, de tutoriálok segítségével sikerült átvernem és letöltenem az adatokat, amelyeket a Facebook profilomba bejelentkezve lekértem. Harmadik lépésben összesítenem kellett a két adathalmazom. Közgáz2018 csoportban ugyanis felsőbbévesek, tanárok is vannak, akiket ki kellett szűrnöm a nr.matricol-ok alapján beazonosított diákokkal.

ADATFORMÁZÁS

Véleményem szerint ide tartozik az adatok egységesítésének lépése is, hiszen a Facebook-ról kinyert adatok fele magyarul, fele románul, fele hiányosan, fele helytelenül írva volt. Miután meglett az összesített nagy adathalmazom a megfelelő számú adatpontokkal, Google Colab Notebooks segítségével formáztam Python-ban. Dimenziók formátumát állítottam be, új oszlopokat szűrtem be értékekkel (születési dátum alapján a hónapokat és évszakokat, neveket szétszedtem vezetékek és keresztnemekre). Többféle DataFrame-t exportáltam, hiszen csak azon adatsorokat tudtam megtartani, amelyek teljesek voltak és nem hiányosak. Nemutolsó sorban geokódoltam is az egyes helységeket és megyéket ahonnan a hallgatók érkeztek. Románia megyéit tartalmazó gejson fájlt Github-ról szedtem, és azt használtam fel. Illetve ha valami nem jött össze Excel segítségével megoldottam, mert abban otthonosabban mozgok.

VIZUALIZÁCIÓ VÁLASZTÁS

Miután a témával kapcsolatos adataimat gyűjtöttem, feljegyeztem, egységesítettem, elemezgettem jöhetett az elemzésük és megjelenítésük. Céлом az volt az általam összegyűjtött adatokat ne egyszerű diagramok segítségével (pár) mutassam be, hanem letisztult, mutatós és könnyen kihámozhatóakkal. 2018-ban beiratkozott hallgatók történetét szerettem volna elmesélni. A vizualizációk elkészítése során igyekeztem elkerülni a tengelylevágásokat vagy a tengelyértékek levágását, szövegméretezés és padding-ek által. Minimális komplexitásra törekedtem, főként a színek, méret és alakzat alapján terveztem.

Eszközválasztásnál próbálkoztam a PowerBI-al és a Flourish-al. Mindkettőben volt olyan vizualizáció, ami tetszett. Egyforma mértékben próbálgattam ki opciókat és beállításokat, végül a Flourish adatvizualizációs eszközt választottam. Nagyon megragadott az, hogy színpaletták is vannak és sajátot is lehet megadni, az is, ahogyan csoportosítja a vizualizációtípusokat illetve az, hogy példa adatsorokkal volt feltöltve annak érdekében, hogy tudjuk melyik típus milyen adatokat vár el és hogyan kell formázzuk adataink. Vizualizációimat egy story-ba tettem, melynek linkje a következő: <https://public.flourish.studio/story/708401/>

Az évfolyam hallgatóit nemekre lebontva szakirányonként hierarchiatérképpel ábrázoltam, mégpedig Packed circles diagrammal. Azért választottam ezt, mert nem volt olyan sok dimenzió értékem, hogy szembetűnőbben kiemeljem. A Tree map típusú vizualizációban minimális eltérések látszódtak. Ugyanezt a Packed circles nevű diagramot használtam ugyanazokkal a színekkel ugyancsak az évfolyam hallgatóit lebontva szakirányonként miután néhányan már abbahagyták az egyetemet annak érdekében, hogy látszódjon a vátozás nemek szerint.

Érdekességképpen betettem egy Word Cloud vizualizációt is a hallgatók keresztnéveivel.

A Parlament chart típusú vizualizációm az évfolyamunk alakulásáról szól, szakonként felbontva, mikor mennyi volt a létszám, illetve hányan hagyták abba a szakot 2018-tól 2021-ig. Hallgatókról lévén szó a Parliament chartot használtam, mert a helyek az aulára emlékeztettek, ahol a vírusnak köszönhetően rég jártam. Ha pontosabb adatokat tudtam volna gyűjteni ki melyik félévet nem kezdte meg stb. akkor lehetett volna ezt Bar chart race típusú vizualizációval is megjeleníteni.

Projection map-et használtam annak megjelenítésére melyik megyéből hányan jöttek 2018-ban a Közgáz karra. Azért választottam ezt (annak ellenére csak erdélyi megyék jellennek meg rajta), mert sokkal hamarabb értelmezzük a megyéket, ha látjuk, mint ha valamelyik nevét olvassuk és beazonosítjuk (elképzeljük), még ha tizedmásodpercekről is legyen szó. Helyek hallatán az emberek általában egyből próbálják beazonosítani a koordinátákat. Így a Choropleth térkép felgyorsítja az adatfeldolgozást a színek segítségével.

Arc map-et használtam pedig annak szemléltetésére ki milyen helységekből érkezett Kolozsvárra, amelyen látszik, hogy évfolyamunkra Erdély sokszínű helységeiből érkeztek hallgatók.

A legutolsó vizualizáció lett a kedvencem, mert összefoglalta mindazt, amiről a projektem szólt. Survey típusú vizualizációban megjelenítettem az összes adatpontomat. Így remekül összegezte az előtte lévő vizualizációim, és látszott témám limitációja is, a Facebook adatvédelme, hiszen külön csoportokban megjelentek azok a hatszögek (egyedek), akikről nem kaptam megadott dimenzióban adatot.

KÖVETKEZTETÉSEK

Vizualizációm egy szuper kis történet a 2018-as évfolyamról, vagyis évfolyamunkról. Eleinte 244-en iratkoztunk be (128 lány, 116 fiú) és a 2021-es évet már csak 187-en kezdtük meg, 36 fiú és 21 lány abbahagyta. Remekül látszik az is szakonként hogyan csökkent a létszám, illetve összben is. Fiúk inkább a menedzsment szakot hagyták ott a legtöbben, míg a lányok inkább a marketing szakot. A Facebook-on lévő nyilvános adatok alapján a legtöbb hallgató Kolozs megyéből jött majd utána van Hargita, Szilágy, Maros stb. megye is. 53 különböző helységről érkeztek a hallgatók.

Ami a mi szakirányunkat illeti, a gazdasági informatikát, 50-en en kezdtünk pont fele fiú és fele lány arányban illetve a harmadévre 35-en maradtunk. El szeretném mondani remekül álltuk a helyünk és kitartóak voltunk, hiszen sokszor volt olyan az elmúlt évek során hogyan nemcsak 15-en hagyták ott a szakirányunkat.

Ami az összefüggéseket illeti, mivel a hallgatók kb. fele nem osztja meg nyilvánosan azokat az adatokat, amelyekre nekem lett volna szükségem az idő és helydimenzióhoz, így nem tudtam reális következtetéseket levonni, a limitációk miatt. Ettől függetlenül elkészítettem őket csak nem találtam relevánsnak, hogy betegyem a bemutatómba is. Ilyen összefüggések például hogy melyik megyéből jöttek inkább lányok és inkább férfiak, azok között, akik abbahagyták hányan voltak ugyanabból a megyéből, melyik helységről jött hallgatók hagyják ott leginkább az egyetemet stb.