**Önálló kutatási feladat mérföldkő beszámoló**

**Cikkek tanulságai:**

A miénkhez hasonló tematikájú, de nem pontosan olyan célú projektekről vannak cikkek birtokunkban. BME-n folytatott projektek beszámolóit is megkaptuk, szintúgy Szilágyi Brigitta Tanárnő vezetésével készültek. Ezen tanulmányok fő feladata a gyakori tesztelés, illetve a nemrégiben bevezetett új oktatási módszerek eredményességének vizsgálata, illetve a kognitív és matematikai tudást mérő felmérés alapján ezen képességek közötti összefüggések feltárása. A konklúzió általánosságban a bevezetett módszerek (edubase, online konzultációk, segédanyagok stb.) és a gyakori, illetve gyakorlatias tesztelés jótékony hatását bizonyítja (ezt leginkább kérdőívek és végső eredmények alapján származtatják), emellett kimutatják a korrelációt a matematikai és a kognitív képességek között. Mi is ugyanezen felmérő adatait vizsgáljuk, azonban célunk az utóélet és további eredményesség vizsgálata lesz, a megfelelő összefüggések kimutatása mellett. A cikkekből merítettünk mérési és probléma megközelítési ötleteket, így hasznosnak mondhatjuk a tanulmányozásukat.

**Adattisztítás:**

Az adatokat Szilágyi Brigitta Tanárnő bocsátotta a rendelkezésünkre, amik szerencsére már egészen tiszta állapotban voltak, azonban végeznünk kellett még pár lépést, hogy valóban konzisztens, hiánytalan, használható adatokat kapjunk. Voltak adataink 2019, 2020 és 2021-ről, pontosabban az azévben kezdett diákok kognitív teszteredményeiről. Először el kellett távolítanunk néhány redundáns, illetve irreleváns oszlopot (pl. edubase jelszó, felhasználónév). Ezután volt egy kis inkonzisztencia abban, hogy egyes hallgatók hogyan nevezték meg a szakot, amin tanulnak (kis- és nagybetű, különbség a ’mérnök – mérnöki’ között), úgyhogy ezt pár Python kóddal kiküszöböltük és egységes nevére hoztuk őket (Vegyészmérnöki, Biomérnöki, Környezetmérnöki). Voltak olyan rekordok, ahol nem tudtuk egyértelműen beazonosítani, hogy mi az adott hallgató szakja (mert pl. VBK-t írt oda), így azt átírtuk UNKNOWN-ra, amit esetleg majd a szak alapján csoportosított elemzéseknél elhagyunk (ebből szerencsére évenként csak 2-3 db volt). Kicsit kellett még tisztítani az irányítószámokon, illetve a tankör megnevezésén, de mivel ebből kevés volt, ezt manuális is meg tudtuk tenni.

Az adatok nagyon alaposak voltak, minden hallgatónál külön fel volt jegyezve, hogy mely kérdésre milyen választ adott, és hogy az helyes volt-e. A nagy átfogó elemzéseknél ez nem annyira fontos, ezért azokat különválasztottuk, és áthelyeztük egy másik Excel táblába, amikkel majd a részletesebb elemzéseknél fogunk esetleg foglalkozni. Így ami maradt a 3 évi táblában: a hallgatók neve; Neptun kódja; irányítószáma; emeltezett-e matekból vagy sem; természettudományi tagozatos volt-e vagy sem, illetve a teszteredmények lebontva matekos és nyelvi részre egzakt pontszámmal és százalékos teljesítménnyel. A matek- és kognitíveredmények a 2019-es és 2020-as táblában explicit ki voltak írva, a 2021 táblában csupán az összteljesítmény volt, így ott még cellafüggvényekkel meg kellett határozni a kérdésekre adott válaszokból. A Covid-19 által sújtott években ezen felül még rendelkezésünkre állt egy olyan kategorikus változó is, ami azt fejezte ki, hogy 12. osztályban milyen formában folyt a matematika oktatás. Ezután csináltunk néhány táblát, amiben a különböző évek eredményeit összeillesztettük, nyilván hozzáadva minden rekordhoz egy új oszlopot, amiben az évszám áll, így majd évenkénti bontásban is meg tudjuk vizsgálni az adatokat.

Idő közben kaptunk még 0. Zh eredményeket és felvételi pontszámokat is egyes hallgatókhoz, akik az elmúlt két évben iratkoztak be, így azokat merge-eltük a megfelelő évi táblákkal, és kiszűrtük a többletoszlopokat (mert pl. minden sorban kétszer lenne a hallgató neve).

**Első Adatelemzési lépések:**

Először Tableau segítségével készíttettünk néhány kétdimenziós ábrát arra vonatkozóan, hogy milyen összefüggés van a matematikai és a kognitív pontszámok között. Azt tapasztaltuk, hogy egy minimális pozitív korreláció van a két eredmény között, ám nem túl szignifikáns. Ezután megnéztük évre és szakra lebontva is. Ekkor azt tapasztaltuk, hogy a 2020-as matematika eredmények ugyan jobbak a 2019-eshez képest minimumpontszám alapján, ám összességben nem sokkal. Ennek az lehet az oka, hogy nem nagyon lehet kipuskázni a kérdéseket, mert nem lexikális tudást mérnek. Ugyanakkor a 2021-es eredmények rosszabbak lettek, mint a 2020-as és 2019-es, és ez valószínűleg a digitális időszak következménye. A szakonkénti bontásnál pedig azt tapasztaltuk, hogy a vegyészmérnökök átlagosan jobban teljesítenek mindkét területen, mint a bio- vagy környezetmérnökök. Azt is észrevettük, hogy a pontszámok közti szórás 2021-ben a legnagyobb. Ezután azt vizsgáltuk, hogy mekkora kihatása van a teszteredményekre annak, hogy az adott hallgató emelt matematika érettségit tett-e vagy sem, illetve, hogy természettudományos tagozaton volt-e. Nem meglepő módon az emelt vizsgát tevők sokkal jobban teljesítettek a matekos részen, a kognitív részen viszont sok különbség nem volt. Hasonló volt a helyzet a természettudományos tagozati vizsgálatkor. Az sem lepett meg minket, hogy a felvételi pontszám, a 0. ZH eredménye és a többi teszteredmény között valamekkora pozitív korreláció azért van (a számolt R^2 statisztika 0.2-0.4 közé esett).

Ezekből az szűrhető le, hogy nem véletlenszerűek az eredmények, vannak összefüggések: érdemes külön vizsgálni az emeltes és nem emeltes csoportot, illetve szakonkénti bontásban is.

**Modellezés:**