**SVEUČILIŠTE U RIJECI**

**ODJEL ZA INFORMATIKU**

**Ime i prezime studenta**: Nola Čumlievski

**Smjer**: Informacijski i komunikacijski sustavi

**Studij**: Diplomski studij

**SUSTAVI ZA PODRŠKU ODLUČIVANJU**

*Odnos diplomiranih i upisanih studenata u UK prema području obrazovanja*

**PRIKUPLJANJE, PRIPREMA I VIZUALIZACIJA PODATAKA**

**Nositelj kolegija:** dr. dc. Sanda Martinčić-Ipšić

**Asistent:** Karlo Babić

**Rijeka, lipanj 2020.**

**Sadržaj**

[1. Uvod 1](#_Toc42519788)

[2. Prikupljanje podataka 2](#_Toc42519789)

[3. Priprema podataka 6](#_Toc42519790)

[4. Vizualizacija podataka 12](#_Toc42519791)

[4.1.Prvostupnici ženskog spola po području studiranja 12](#_Toc42519792)

[4.2. Prvostupnici muškog spola prema području studiranja 13](#_Toc42519793)

[4.3. Razlika između upisanog područja studiranja među spolovima za 2014. godinu 14](#_Toc42519794)

[4.4. Razlika između magistara područja studiranja među spolovima za 2014. godinu 15](#_Toc42519795)

[4.5. Omjer upisanih, diplomiranih i doktoranda za 2015.godinu 16](#_Toc42519796)

[4.6. Omjer upisanih, prvostupnika, magistara i doktoranda za 2017. godinu 17](#_Toc42519797)

[4.7. Razlika između broja aplikanata u razdoblju od 2014 – 2017. godine 18](#_Toc42519798)

[4.8. Razlika među prvostupnicima u razdoblju od 2014 – 2017. godine 19](#_Toc42519799)

[4.9. Prosjek magistara prema područjima studiranja 20](#_Toc42519800)

[4.10. Prosjek prvostupnika prema područjima studiranja 21](#_Toc42519801)

[4.11.Ukupan prosjek diplomiranih prema području studiranja u razdoblju od 2014. do 2017. godine 22](#_Toc42519802)

[4.12. Odnos upisanih i diplomiranih studenata u 2017.godini 23](#_Toc42519803)

[5. Literatura 24](#_Toc42519804)

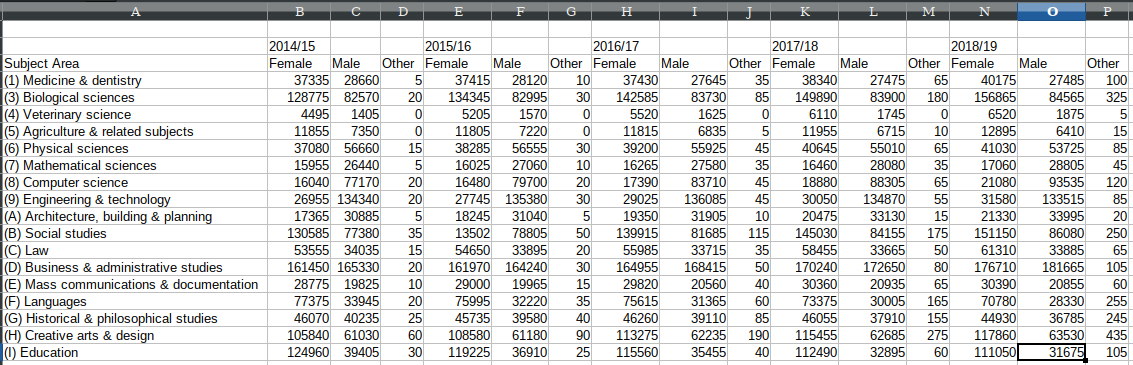
# **Uvod**

Seminar se sastoji od pisanog izvješća u Wordu koje opisuje sve potrebne korake za izradu ovog seminara te praktičnog djela gdje se obavlja prikupljanje podataka, priprema te vizualizacija podataka. Tema seminarskog rada je *„Odnos diplomiranih i upisanih studenata u UK prema području obrazovanja“*. U drugom poglavlju opisani su izvorni podaci korišteni za izradu vizualizacija te je naveden njihov izvor. U trećem poglavlju opisan je način spajanja tablica pomoću alata *„Talend Data Integration“.* Četvrto poglavlje sadrži interpretaciju i vizualizacije koje su izrađene pomoću alata *„Tableau“* i *„Microstrategy“.*

# **Prikupljanje podataka**

Za potrebe seminara trebalo je odabrati minimalno 2 izvora podataka s različitim podacima.

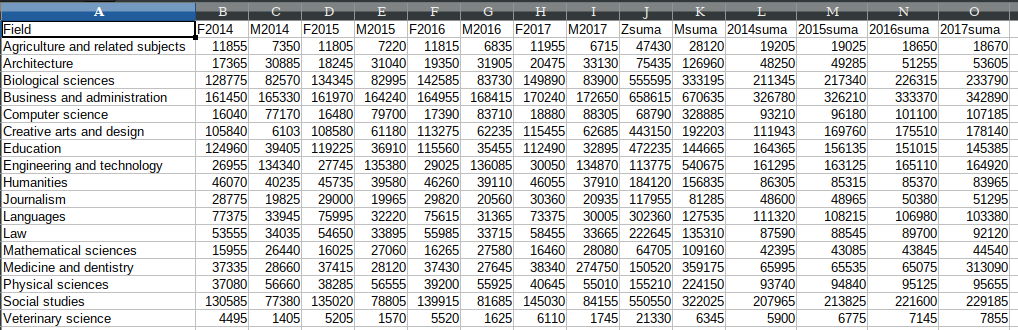
Kako bi mogli usporediti upisane i diplomirane studente, potrebne su minimalno dvije tablice. Prva tablica (tablica upisanih studenata u UK) skinuta je sa ovog izvora: <https://www.hesa.ac.uk/data-and-analysis/students/what-study>. Ova tablica sadrži broj upisanih studenata prema spolu i području studiranja u razdoblju od 2014 – 2018. godine. Na slici 1. prikazan je izvorni izgled tablice.



Slika 1. Izvorna tablica upisanih studenata

Tablica je uređena na način da su ostavljeni samo podaci za muški i ženski spol te je pomoću funkcije *„sum“* određen ukupan broj studenata ženskog i muškog spola po određenom području, kao i ukupan broj studenata (muških i ženskih) za određenu godinu. Također su izbačeni podaci za 2018. godinu, s obzirom da u drugim tablicama nemamo podataka za istu. Nakon uređivanja tablica izgleda kao na slici 2.

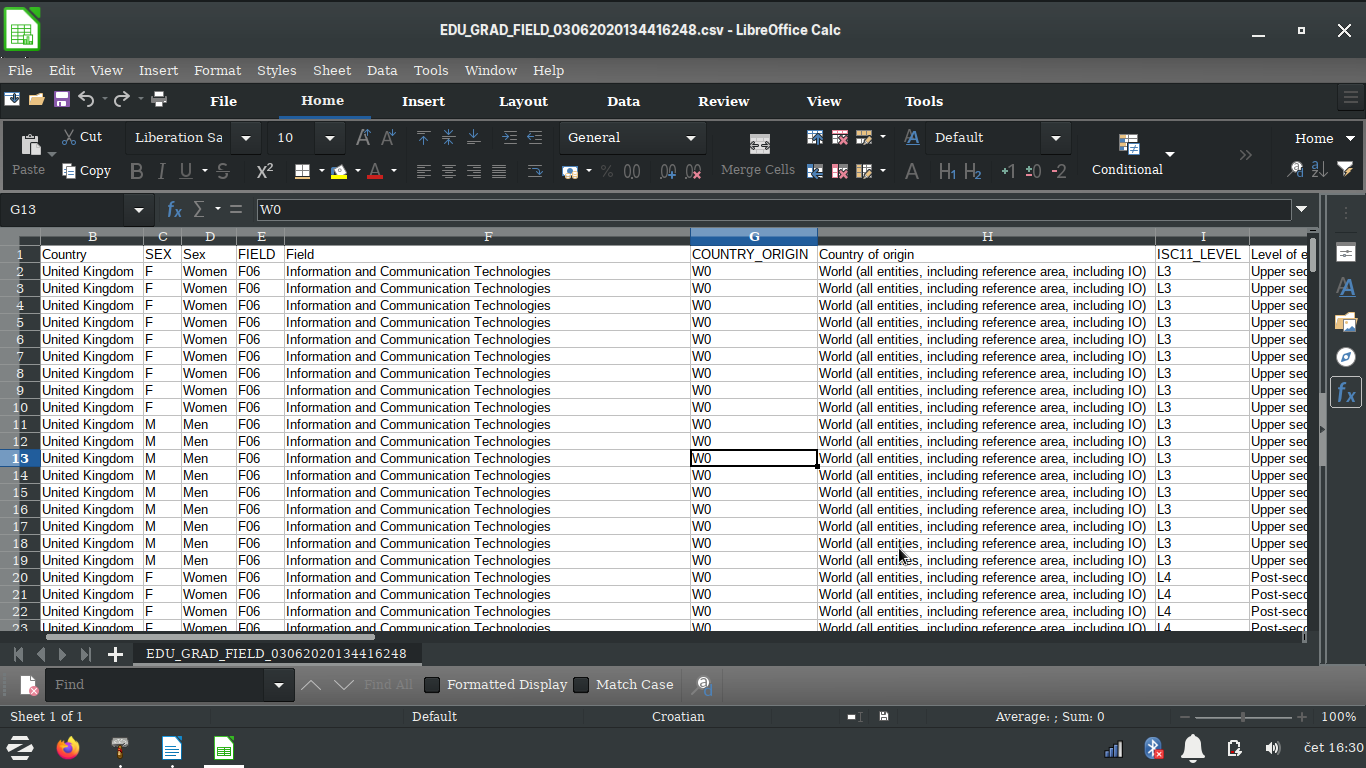
Stupci koji čine uređenu tablicu su: *„Field“* (područje studiranja)*, „F2014“* (broj studentica upisanih u 2014. godini)*, „M2014“* (broj studenata upisanih u 2014. godini) *„F2015“* (kao i *M2015*, broj studenata/ica upisanih u 2015. godini)*, „F2016“* (kao i *M2016*, broj studenata/ica upisanih u 2016. godini)*, „F2017“* (*M2017*, broj studenata/ica upisanih u 2017. godini)*, „Zsuma“* (kao i *Msuma*, ukupan broj studenata/ica upisanih u određenom području za sve godine)*,, „2014-2017suma“* (ukupan broj studenata/ica upisanih određene godine za određeno područje)*.*



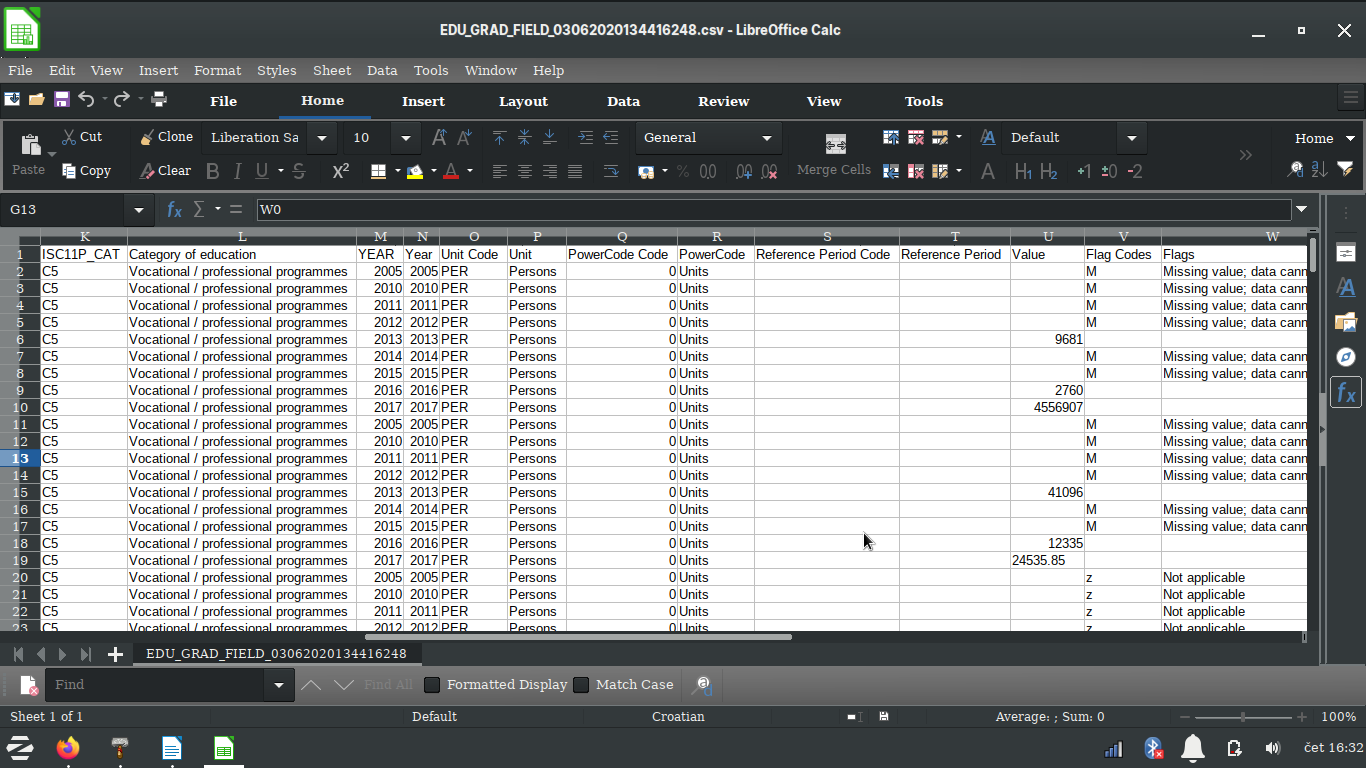
Slika 2. Uređena tablica upisanih studenata

Druge tri tablice (tablica prvostupnika, tablica magistara i doktorata) skinuta je sa navedenog izvora: [https://stats.oecd.org/viewhtml.aspx?datasetcode=EDU\_GRAD\_FIELD&lang=en#](https://stats.oecd.org/viewhtml.aspx?datasetcode=EDU_GRAD_FIELD&lang=en).

Izvorna tablica sadrži veliku količinu podataka, kao što su područje, spol, država (tablica je napravljena za veći broj država, točnije njih 47 od kojih je potrebna jedna), razinu obrazovanja, kategoriju obrazovanja, godinu (2005, zatim 2010 - 2017), i mnoge druge atribute. Izvorna tablica prikazana je na slikama 3. i 4.

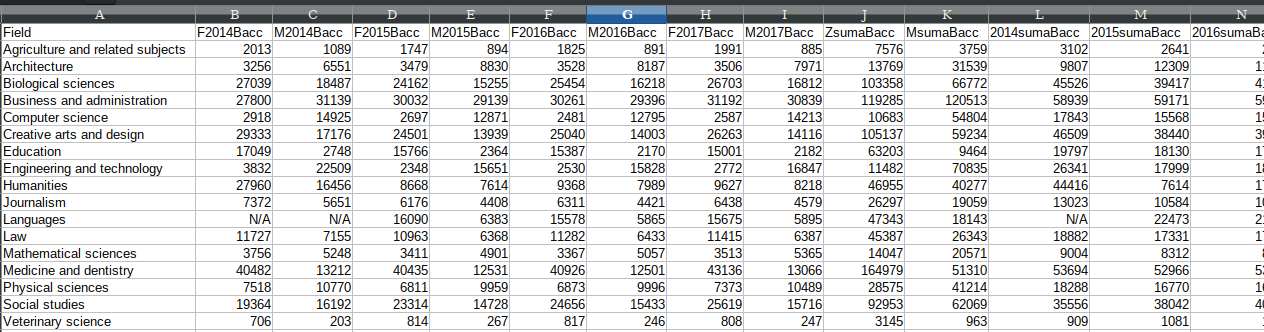


Slika 3. Prvi dio izvorne tablice

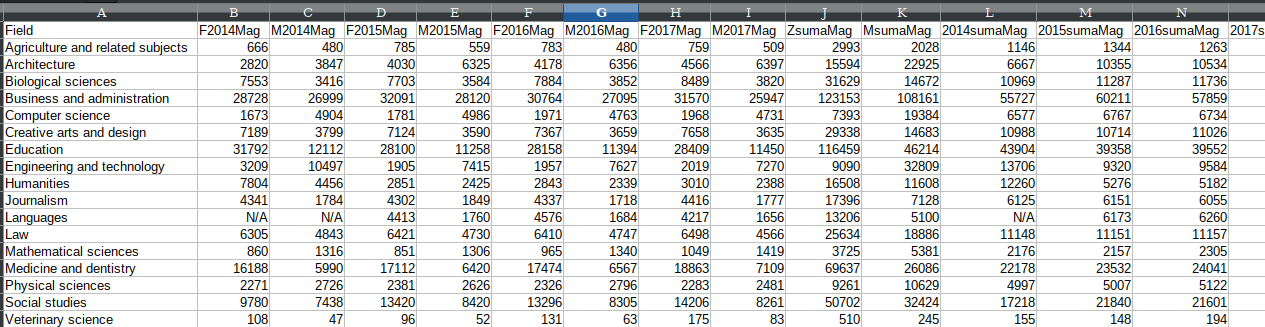


Slika 4. Drugi dio izvorne tablice

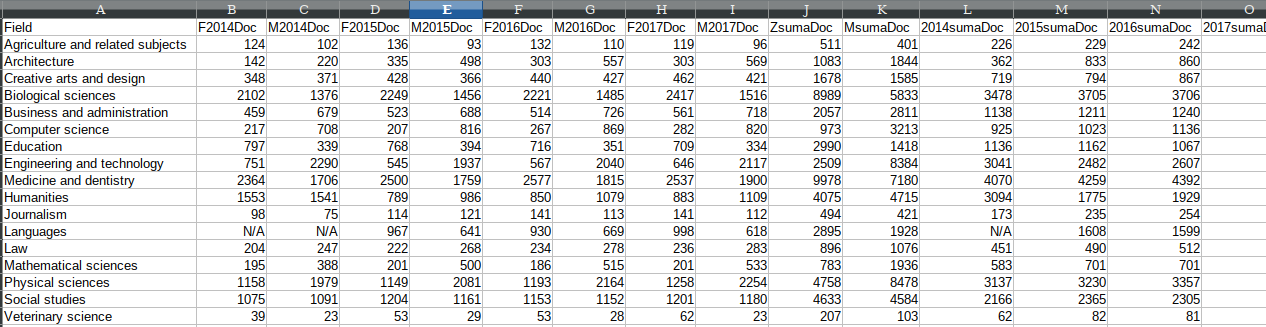
Kao što je vidljivo iz priloženih slika, tablice sadrže za ovaj projekt mnogo nerelevantnih pa i istih informacija (atribut *YEAR* i *year* npr., *FIELD* i *field* gdje jedan sadrži pun naziv područja, a drugi klasifikacijsku kategoriju, *SEX* i *sex* i slično). Tablica je uređena i razdvojena u 3 različite tablice sa tri različite razine obrazovanja (prvostupnici, magistri i doktorati) koje sadrže iste atribute (stupce) kao i tablica sa brojem upisanih studenata po području. Prikazi uređenih tablica vidljivi su na slikama 5, 6 i 7.



Slika 5. Uređena tablica prvostupnika



Slika 6. Uređena tablica magistara

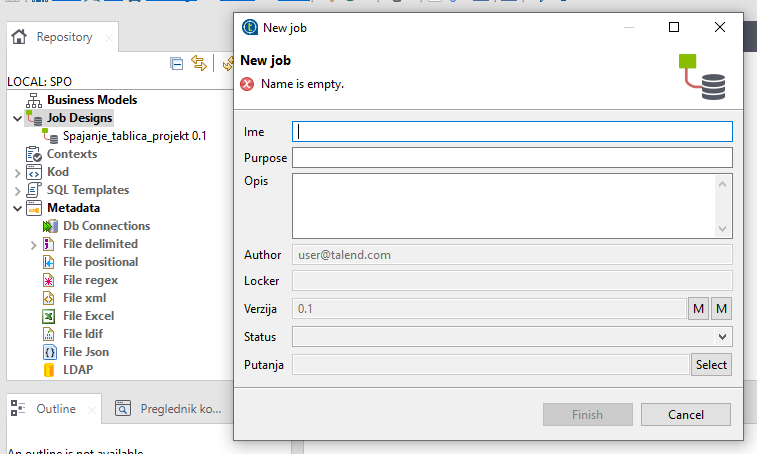


Slika 7. Uređena tablica doktoranda

Iz gore navedenih slika, vidljivo je da sve uređene tablice sadrže iste atribute, uz dodatak određene ekstenzije u nazivu stupca kako bi se moglo razlikovati međusobno tablice u konačnoj spojenoj tablici (u tablici prvostupnika pridodana je ekstenzija *„Bacc“* , u tablici magistara *„Mag“* i u tablici doktoranda ekstenzija *„Doc“*). Suma je izračunata na isti način kao i kod tablice s brojem upisanih studenata/ica po području obrazovanja. Iz priloženih tablica za prvostupnike, magistre i doktorande vidljivo je da postoje nepoznati podaci vezani uz 2014. godinu za područje jezika (kod stupaca *F2014, M2014 i 2014suma*) koji nisu bili dostupni u izvornim tablicama iz nepoznatih razloga.

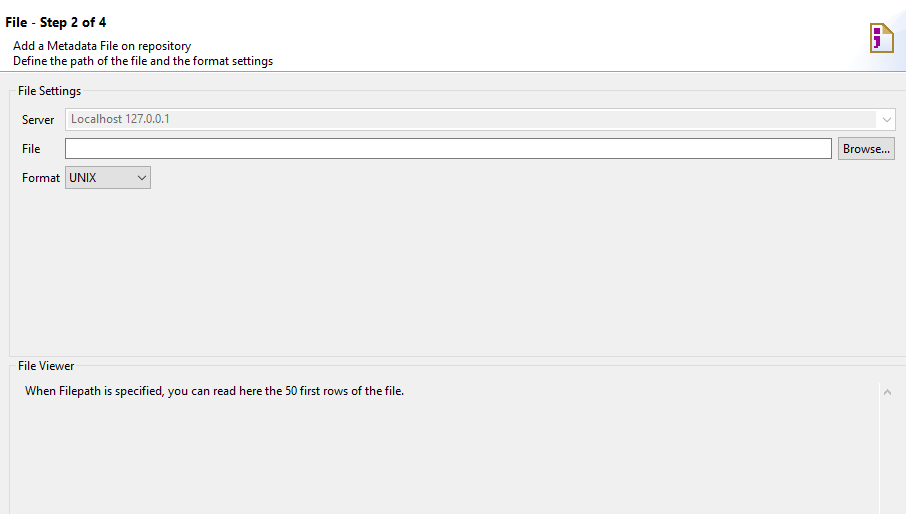
# **Priprema podataka**

Prvi korak prilikom otvaranja alata *„Talend Data Interogation“* kako bi se spojile tablice jest stvoriti/kreirati novi *job.* Pod *Repository* nalazi se opcija *Job design* te desnim klikom na tu opciju odaberemo *Create job.* Otvori se prozor prikazan na slici 8. gdje je potrebno upisati ime job-a.



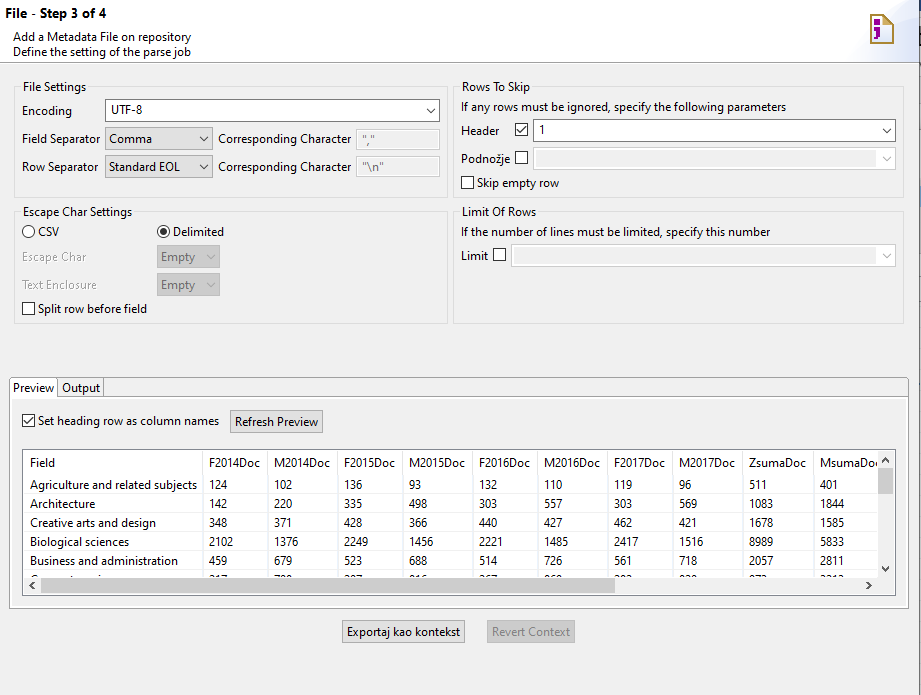
Slika 8. Prozor za stvaranje job-a

Nakon upisa imena i svrhe *job-a* kliknemo na *„Finish“* te smo tim postupkom kreirali novi *job*. Nakon kreiranja *job-a*, u izborniku pod *Metadata* desnim klikom na opciju *File delimited* otvaramo opciju *Create file delimited.* Nakon upisa imena datoteke otvara se prozor na slici 9. gdje je potrebno uploadati datoteku (tablicu) sa podacima. Kao format u ovom seminaru je odabran format „*WINDOWS*“.



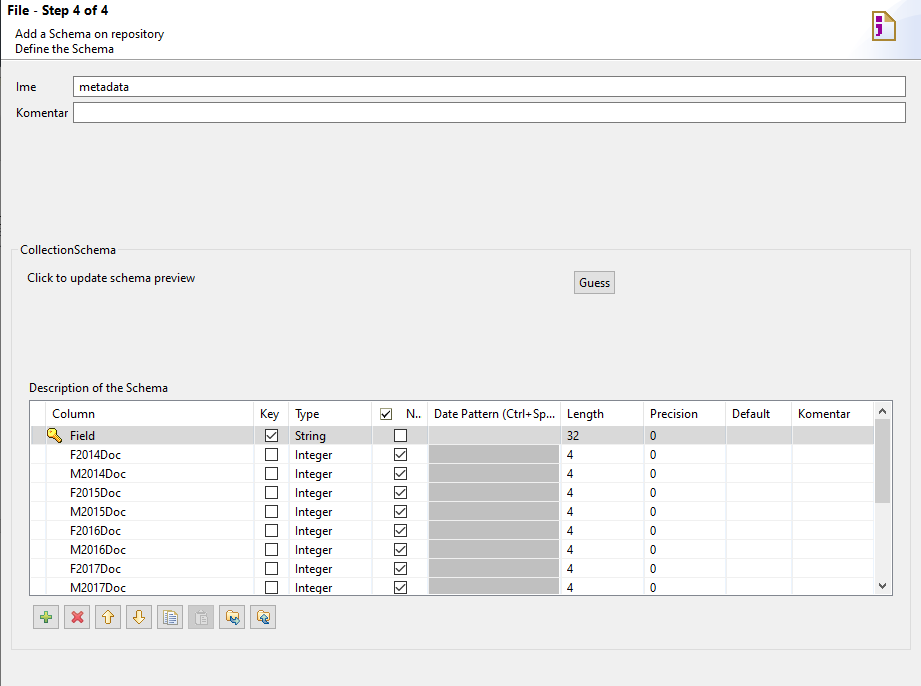
Slika 9. Odabir datoteke za upload

Nakon odabira formata i datoteke, potrebno je postaviti *encoding* UTF-8, *field separator* kao *„comma“* s obzirom da se radi o *csv* datoteci te označiti „*set heading row as column names“* kako bi alat prepoznao prvi red tablice kao imena stupaca. Nakon odabira svih potrebnih opcija potrebno je osvježiti pregled podataka kako bi provjerili da li su podaci dobro navedeni i razdvojeni. Prozor sa potrebnim označenim postavkama prikazan je na slici 10.



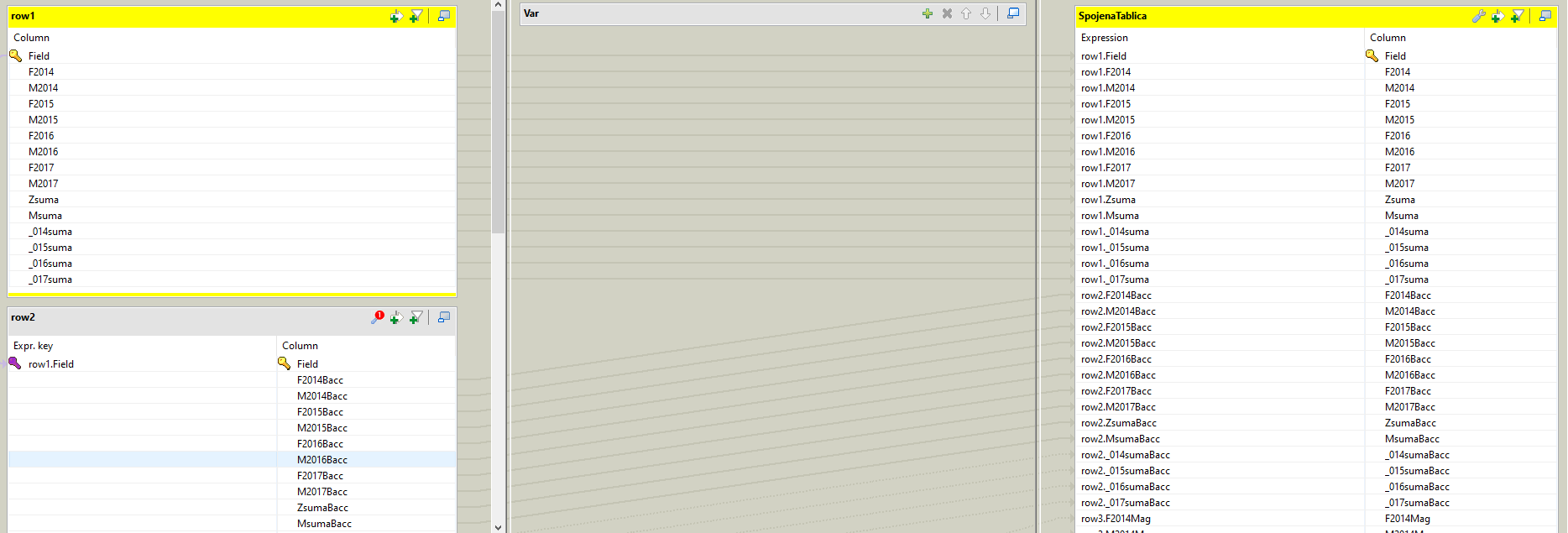
Slika 10. Prozor za izbor separatora i postavljanje naslova

Nakon što je osvježen pregled datoteke, potrebno je postaviti primarni ključ tablice, kao i vrste podataka za svaki stupac. Primarni ključ u ovom primjeru je „*field“* tipa *string,* a svi ostali stupci su tipa *integer.* Također, potrebno je odznačiti stavku  *„nullable“* za primarni ključ jer on ne smije nedostajati i napisati ime sheme. Na slici 11. prikazan je prozor izrade. Klikom na *finish* izrađen je prvi *fileInputdelimited* te je isti postupak potrebno ponoviti sa preostale 3 tablice. Po izboru, u postavkama datoteke (*component*) može se odabrati „*trim all collumns“* i „*die on error“*.



Slika 11. Određivanje vrste podataka i primarnog ključa

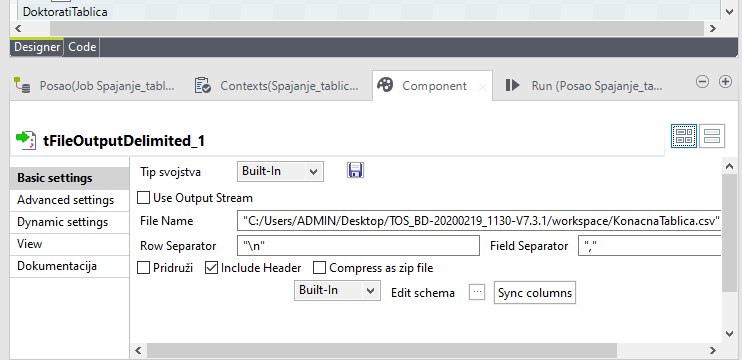
Nakon izrade *4 fileInputdelimited* datoteka (za svaku tablicu po jedna) potrebno je pomoću „*tMap“* komponente spojiti navedene tablice. Desnim klikom na „*fileInputdelimited*“ te odabirom Row > Main spajamo datoteke sa *tMap* komponentom. Na slici 12. nalazi se prozor *tMap* komponente.



Slika 12. tMap funkcija - spajanje tablica

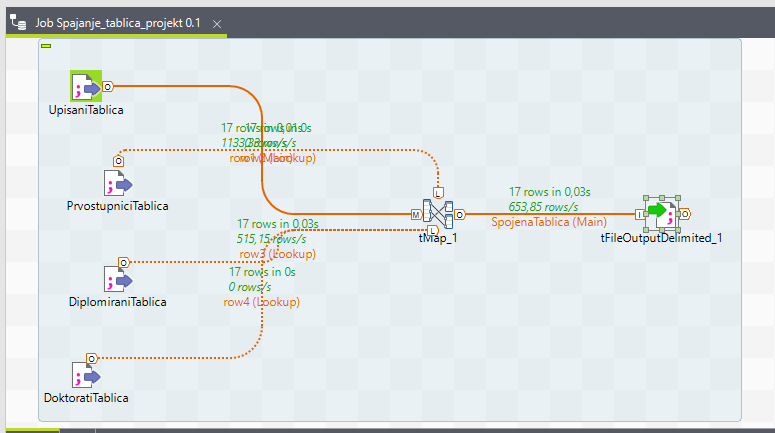
Prije spajanja tablica u jednu, potrebno je povezati njihove primarne ključeve (tablice s lijeve strane prozora) na način da primarni ključ prve tablice povučemo prema primarnom ključu druge, primarni ključ druge povučemo prema primarnom ključu treće te primarni ključ treće povučemo do primarnog ključa četvrte tablice. Nakon povezivanja primarnih ključeva, potrebno je povući sve tablice s desne strane u zajedničku tablicu na desnoj strani (uz uvjet da se primarni ključ povlači samo iz prve tablice, iz ostale tri tablice se povlače sva polja osim primarnog ključa). Nakon spajanja kliknemo na *„auto Map“,* „*apply“ i „ok“.*

Nakon spajanja tablica potrebno je izraditi *fileOutputdelimited* koji izlazni rezultat sprema u novu datoteku.



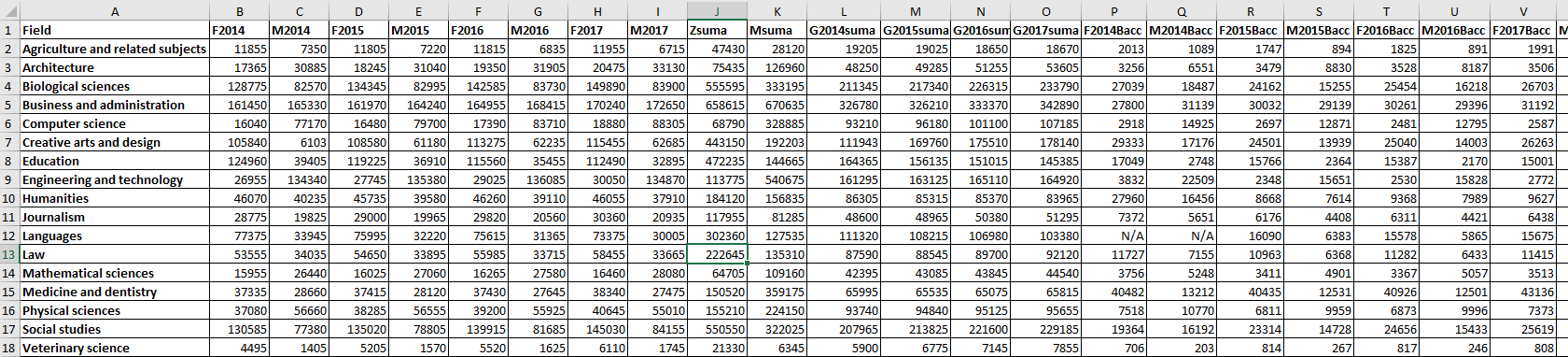
Slika 13. Postavke fileOutputdelimited-a

U komponentama izlazne datoteke potrebno je navesti put do datoteke (*path*)*, field separator* (u ovom slučaju zarez) te označiti postavku *include header.* Nakon postavljanja izlazne datoteke potrebno ju je spojiti sa *tMap* komponentom i pokrenuti *job*. Na slici 14. prikazano je uspješno izvršavanje *job-a*.

**

Slika 14. Uspješno odrađen job

Kao rezultat izvršavanja *job-a* imamo spojeno tablicu koja sadrži sve atribute prethodno navedene četiri tablice. Za potrebe vizualizacijskih alata „*csv“* datoteka konvertirana je u „*xlsx“* format. Na slici 15. prikazan je dio zajedničke tablice.



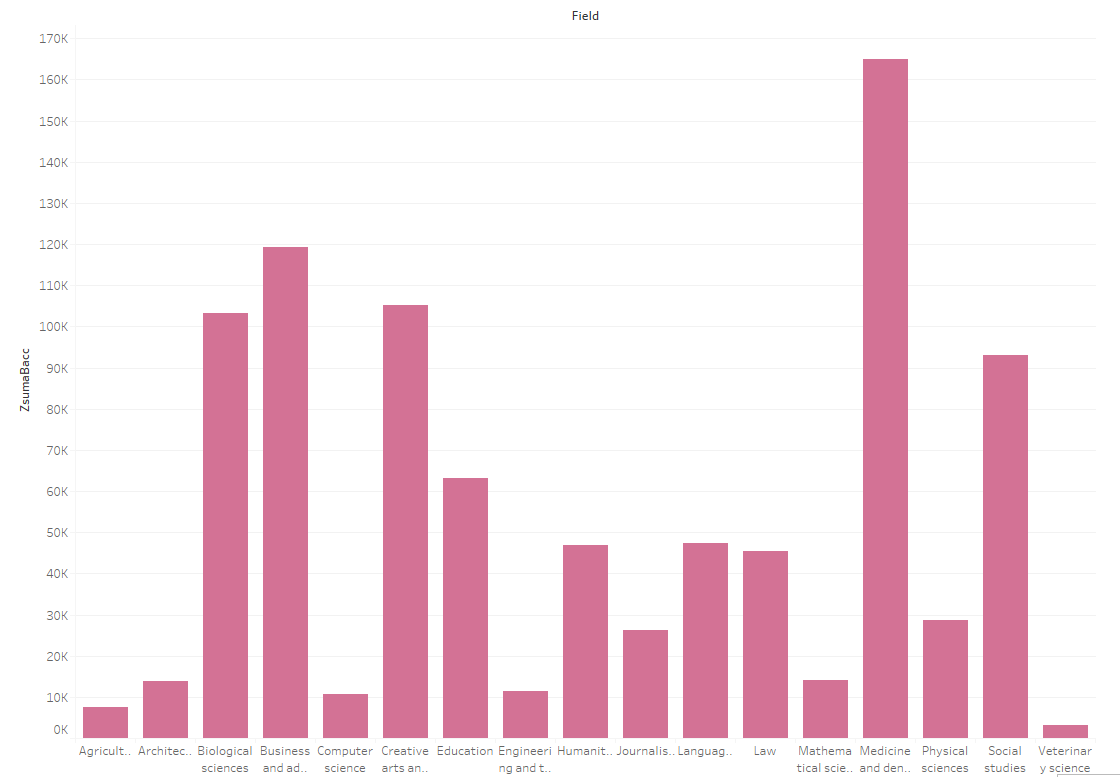
Slika 15. Zajednička tablica

# **Vizualizacija podataka**

U ovom poglavlju navedena su pitanja i pripadajuće vizualizacije te prikladna interpretacija rezultata. Za svaku vizualizaciju navedeno je pomoću kojeg alata se dobila ista. S ovim podacima moguće je modelirati velik broj pitanja koja nije moguće sva obraditi te je u ovom seminaru odabrano desetak.

## **4.1.Prvostupnici ženskog spola po području studiranja**

Pitanje: Koja su područja sa najviše ženskih prvostupnika?

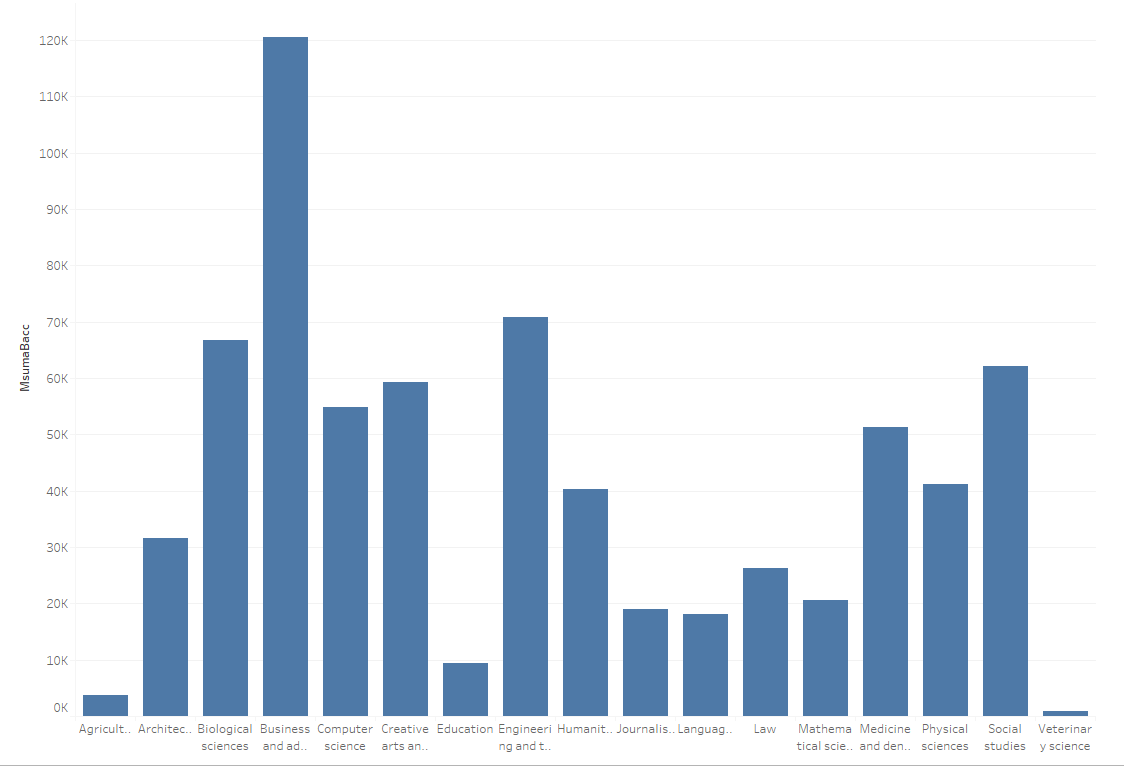


Slika 16. Studentice sa diplomom prvostupnika

Interpretacija: Ova vizualizacija dobivena je pomoću *tableau* alata. Iz navedene vizualizacije vidljivo je da najveći broj ženskih prvostupnika u Velikoj Britaniji (u razdoblju od 2014 – 2017. godine) pripada području medicine, zatim ekonomiji, društvenim znanostima, likovnoj umjetnosti i biologiji. Najmanji broj ženskih studenata diplome prvostupnika nalazi se u području agrikulture, inženjerstva, arhitekture, matematike i veterine (za koja je općenito i manji broj upisanih studentica).

## **4.2. Prvostupnici muškog spola prema području studiranja**

Pitanje: Koja su područja sa najvećim brojem muških prvostupnika?

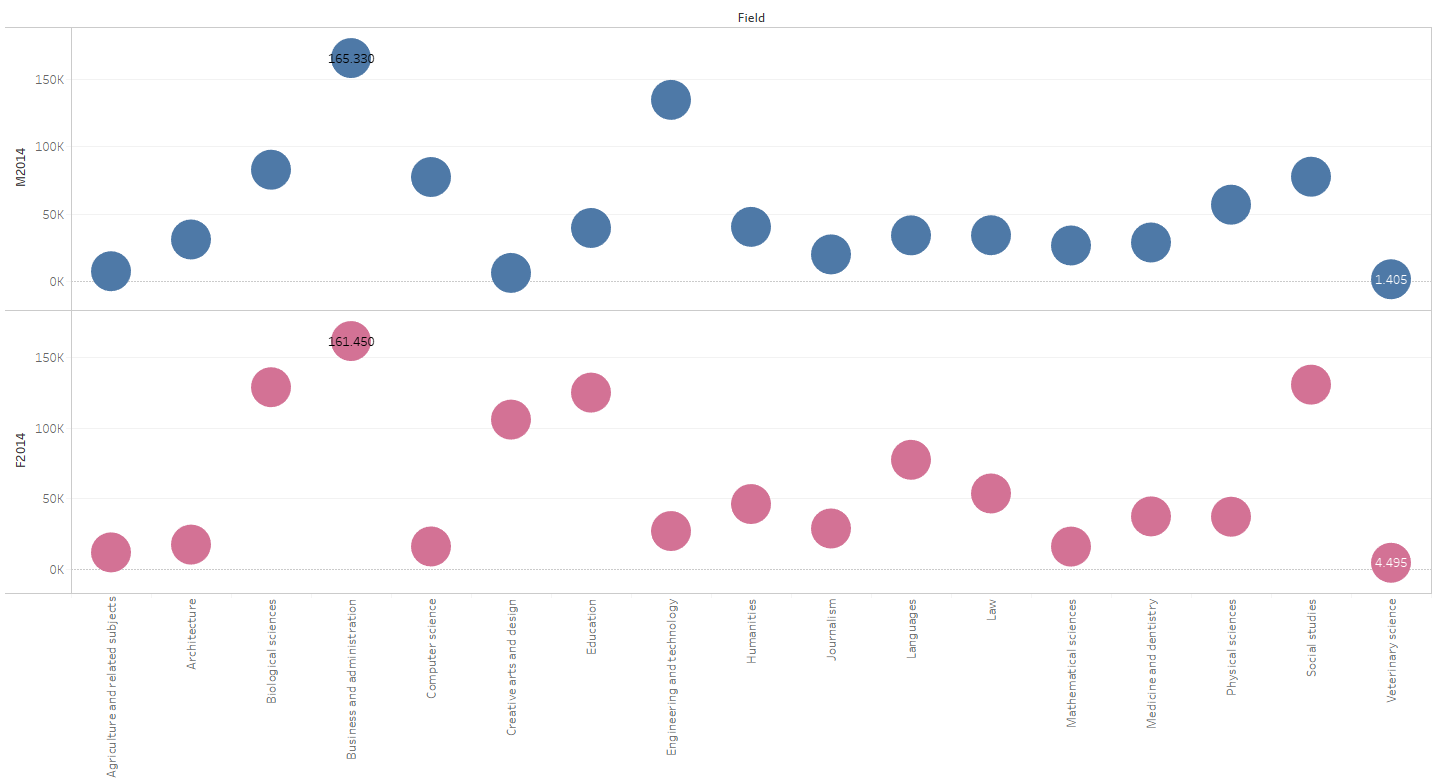


Slika 17. Studenti sa diplomom prvostupnika

Interpretacija: Navedena vizualizacija izrađena je pomoću alata *tableau*. Vidljivo je da najveći broj muških prvostupnika pripada području ekonomije, zatim biologije, arhitekture, inženjerstva te matematike, fizike i društvenim znanostima. Područja s najmanjim brojem muških prvostupnika su agrikultura, edukacija i veterina (također gledano za sve navedene godine – ukupan broj prvostupnika).

## **4.3. Razlika između upisanog područja studiranja među spolovima za 2014. godinu**

Pitanje: Koja su najveća odstupanja među upisanim područjima kod spolova 2014. godine?

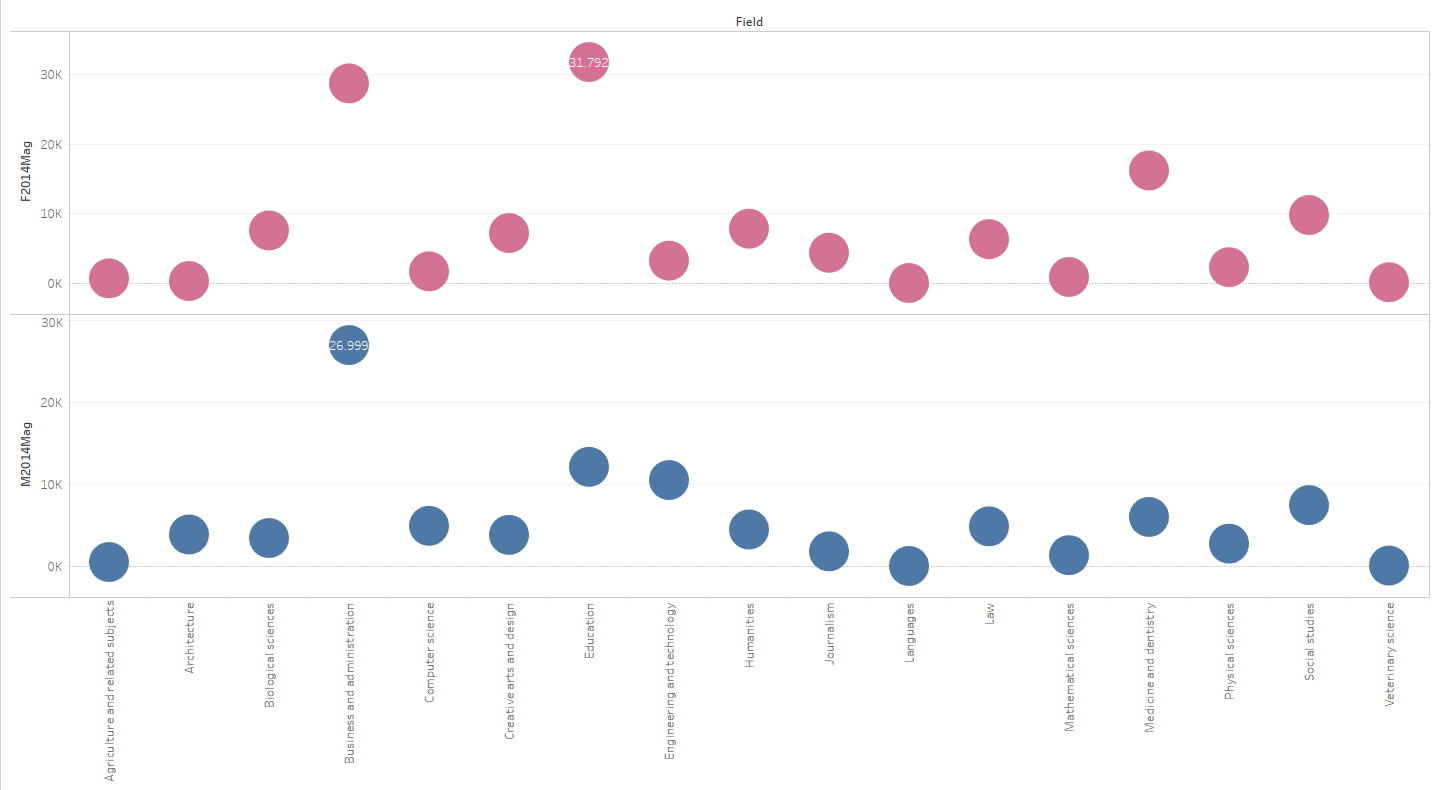


Slika 18. Odstupanja za 2014. godinu

Interpretacija: Vizualizacija je izrađena pomoću alata *tableau*. Iz slike je vidljivo da se područje sa najmanjim i najvećim brojem upisa ne razlikuje (za oba spola vrijedi da je najveći broj upisalo ekonomiju, a najmanji veterinu). Veće razlike vidljive su kod područja likovne umjetnosti, edukacije i društvenih znanosti gdje je znatno veći broj ženskih aplikanata, te u područjima inženjerstva i informatike gdje je veći broj muških aplikanata.

## **4.4. Razlika između magistara područja studiranja među spolovima za 2014. godinu**

Pitanje: Koja su najveća odstupanja među diplomiranim (magistar) područjima kod spolova 2014. godine?

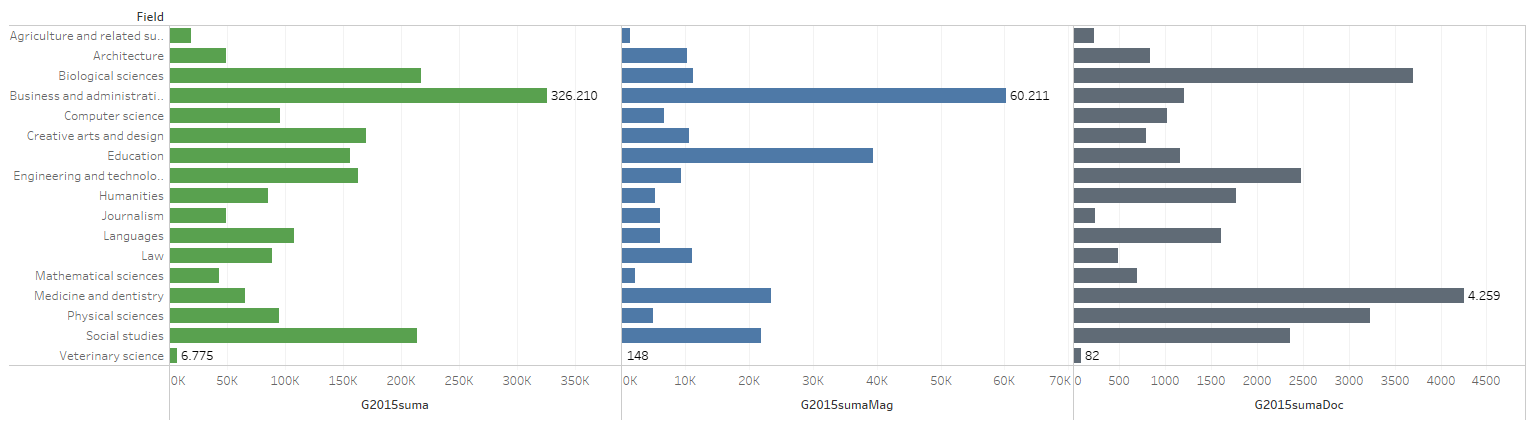


Slika 19. Odstupanja u diplomiranim područjima

Interpretacija: Vizualizacija je izrađena pomoću *tableau* alata. Iz navedene vizualizacije vidljivo je da najveći broj ženskih magistara pripada području edukacije (učiteljski fakultet) dok najveći broj muških magistara pripada području ekonomije. Ostala područja veoma su slična, uz manja odstupanja (npr. u području medicine kao i biologije ima nešto više ženskih magistara, u području informatike ima neznatno više muških magistara itd.).

## **4.5. Omjer upisanih, diplomiranih i doktoranda za 2015.godinu**

Pitanje: Kakav je omjer upisanih, diplomiranih i doktoranda po područjima za 2015.godinu?

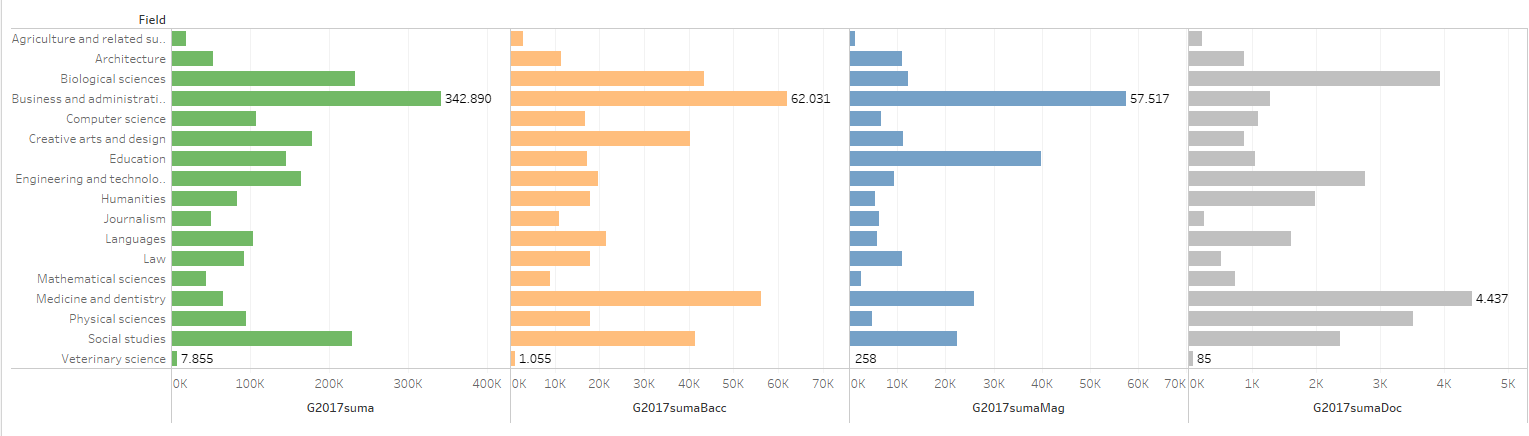


Slika 20. Omjer upisanih, diplomiranih i doktoranda

Interpretacija: Vizualizacija je napravljena pomoću alata *tableau.* Iz priložene slike vidljivo je da je najveći broj upisanih kao i diplomiranih studenata unutar ekonomskog područja, dok najveći broj doktoranda pripada području medicine. Kod ekonomije (sa otprilike 300 000 upisanih po godini) doktorira nešto više od 1000 studenata (ovisi o prirodi područja), dok npr. u područjima biologije, medicine i fizike nalazimo najviše doktoranda.

## **4.6. Omjer upisanih, prvostupnika, magistara i doktoranda za 2017. godinu**

Pitanje: Kakav je omjer aplikanata, prvostupnika, magistara i doktoranda po područjima za 2017.godinu?

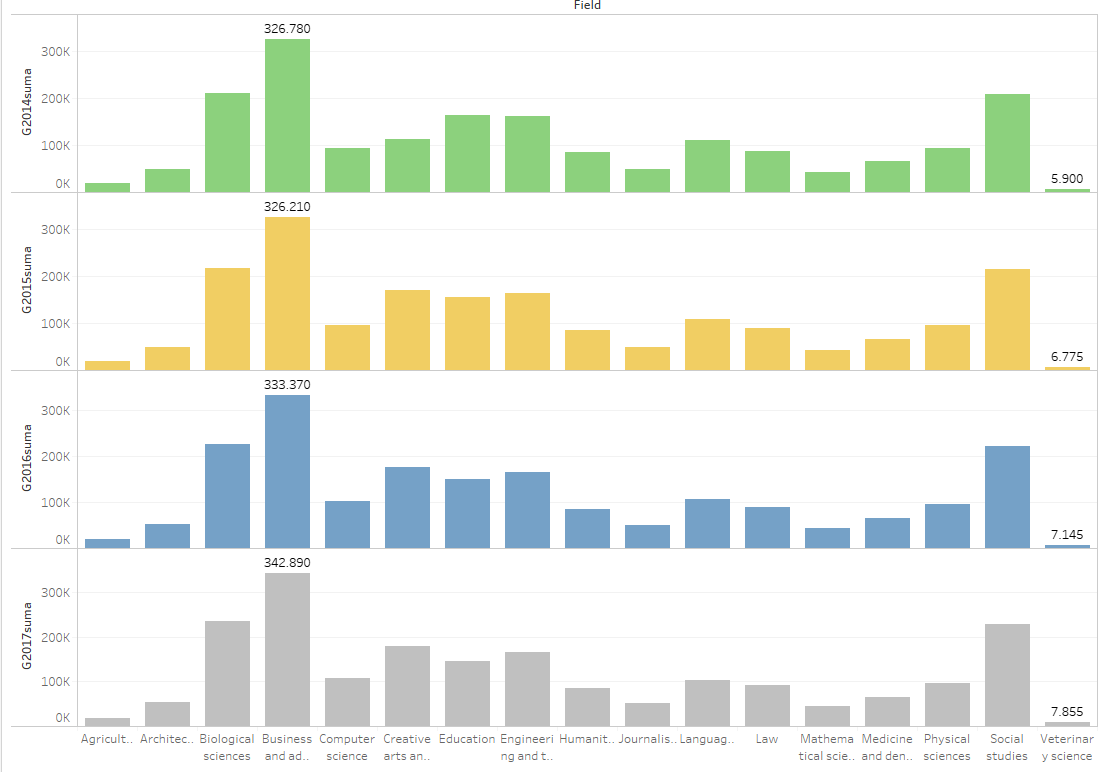


Slika 21. Omjer aplikanata, prvostupnika, magistara i doktoranda

Interpretacija: Vizualizacija je izrađena pomoću alata *tableau*. Kao i kod prošle vizualizacije, vidimo da najveći broj aplikanata pripada području ekonomije (kao i prvostupnika i magistara) dok najveći broj doktoranda pripada području medicine. U razdoblju od 2015 godine do 2017. uzorak nije previše promijenjen. Kod upisa 2017. godine može se primijetiti malen rast broja aplikanata u području edukacije i fizike u odnosu na 2015. godinu. Omjeri magistara i doktoranda po područjima u odnosu na 2015. godinu veoma su slični.

## **4.7. Razlika između broja aplikanata u razdoblju od 2014 – 2017. godine**

Pitanje: Postoji li veća razlika između broja upisanih studenata u razdoblju od 2014 do 2017. godine?

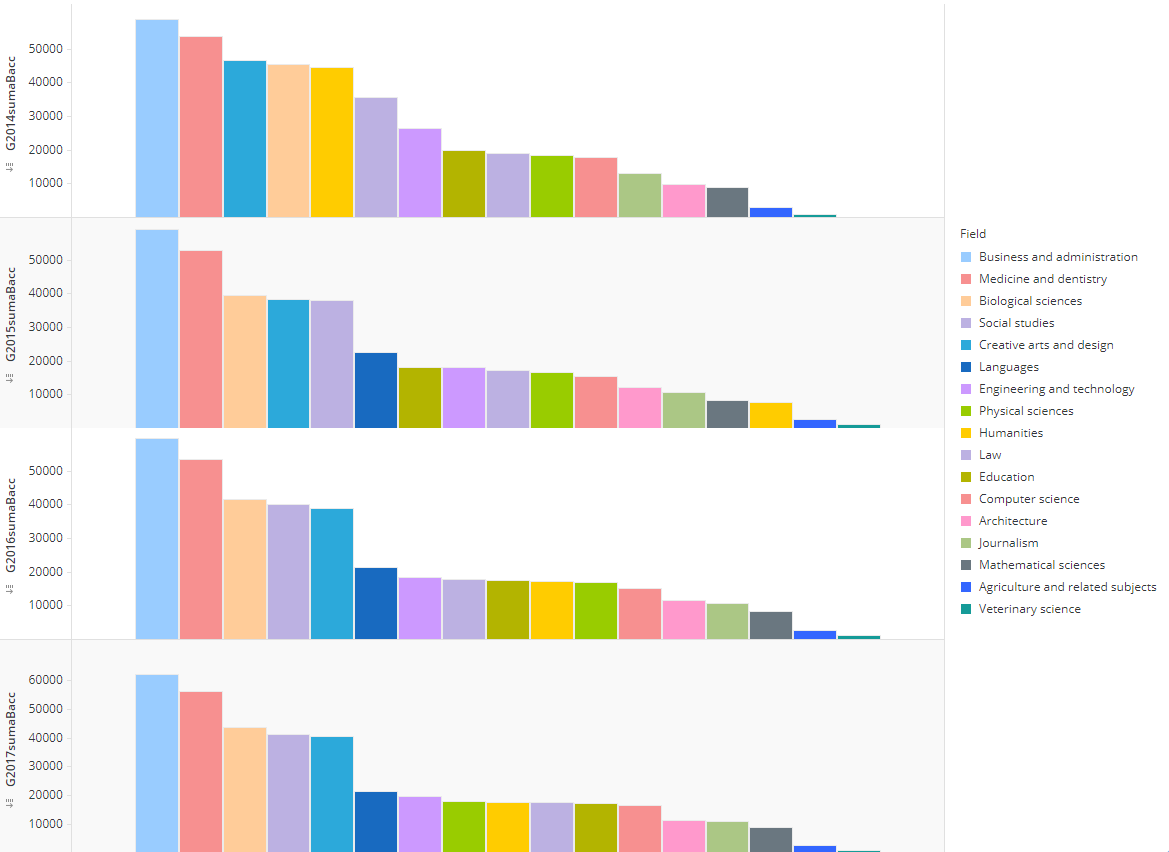


Slika 22. Upisi po godinama

Interpretacija: Navedena vizualizacija izrađena je pomoću alata *tableau.* Iz priložene slike vidljivo je da ne postoje neke veće promjene kod upisa studenata/ica u razdoblju od 2014 – 2017 (možda bi se primijetili veći skokovi i padovi u broju upisa kad bi se analiziralo duže razdoblje s obzirom da se u ovom seminaru analiziraju promjene u razdoblju od 4 godine).

## **4.8. Razlika među prvostupnicima u razdoblju od 2014 – 2017. godine**

Pitanje: Postoji li veća razlika u broju prvostupnika u navedenim područjima u razdoblju od 2014 – 2017. godine?

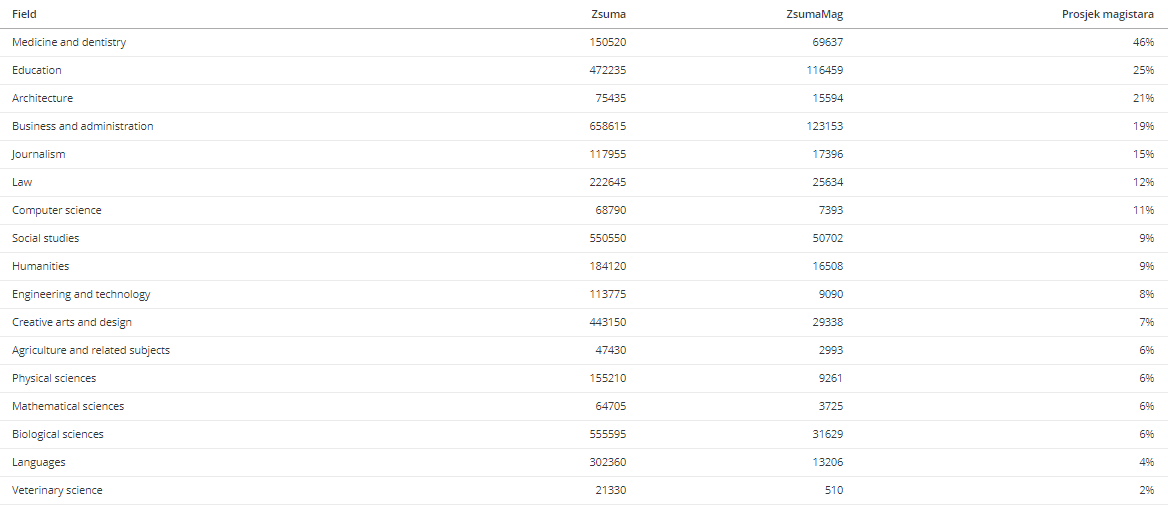


Slika 23. Razlika između broja prvostupnika

Interpretacija: Ova vizualizacija izrađena je pomoću alata *microstrategy.* Iz vizualizacije vidljivo je da kroz navedeno razdoblje najveći broj prvostupnika pripada području ekonomije i medicine. Zatim 2014. godine treće područje sa najvećim brojem prvostupnika je likovna umjetnost dok za ostale godine taj broj pripada području biologije (biološke znanosti). Također se 2014. godine područje humanističkih znanosti nalazi među područjima s najvećim brojem prvostupnika, dok je isto područje na sredini ili kraju skale u razdoblju od 2015. – 2017. godine. Jezici su 2014. pri dnu skale po broju prvostupnika dok su u razdoblju od 2015. – 2017. godine na šestom mjestu (jedno od područja s većim brojem prvostupnika). Ostala područja su više-manje konzistentna kroz navedeno razdoblje, uz mala odstupanja.

## **4.9. Prosjek magistara prema područjima studiranja**

Pitanje: Koliko iznosi prosjek magistara (ženskog spola) prema područjima studiranja u odnosu na broj upisanih studenata/ica u razdoblju od 2014. – 2017. godine?

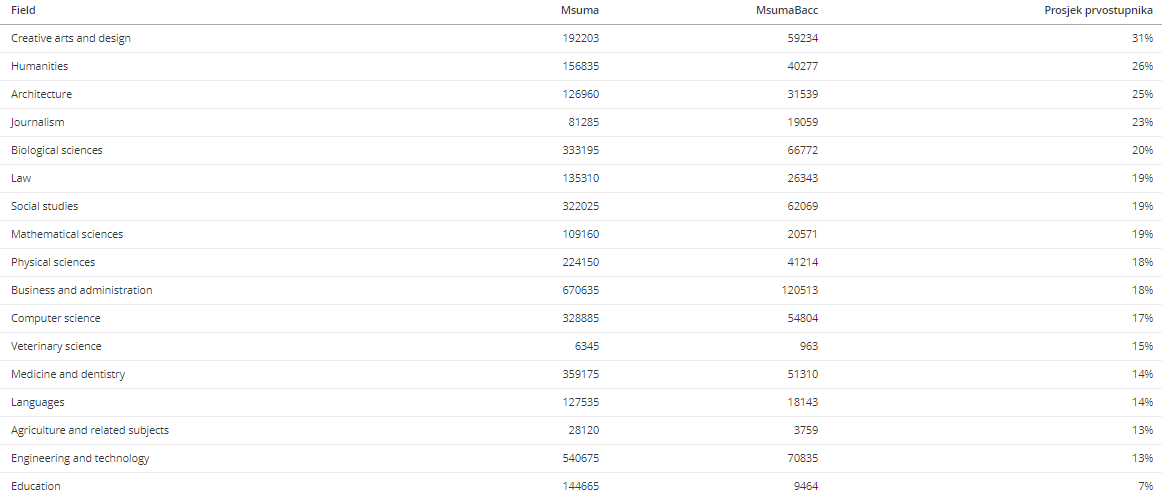


Slika 24. Prosjek magistara

Interpretacija: Vizualizacija je izrađena pomoću alata *microstrategy*. Prosjek magistara dobiven je na način da ukupan broj magistara područja podijelimo sa ukupnim brojem upisanih studenata/ica za određeno područje. Vidimo da je najveći prosjek magistara u odnosu na broj upisanih studenata redom u području medicine, edukacije i arhitekture (46 – 21 %). Najmanji prosjek magistara u odnosu na broj upisanih studenata/ice pripada područjima biologije, jezika i veterine.

## **4.10. Prosjek prvostupnika prema područjima studiranja**

Pitanje: Koliki je prosjek prvostupnika (muškog spola) prema područjima studiranja u odnosu na broj upisanih studenata/ica u razdoblju od 2014. – 2017. godine?



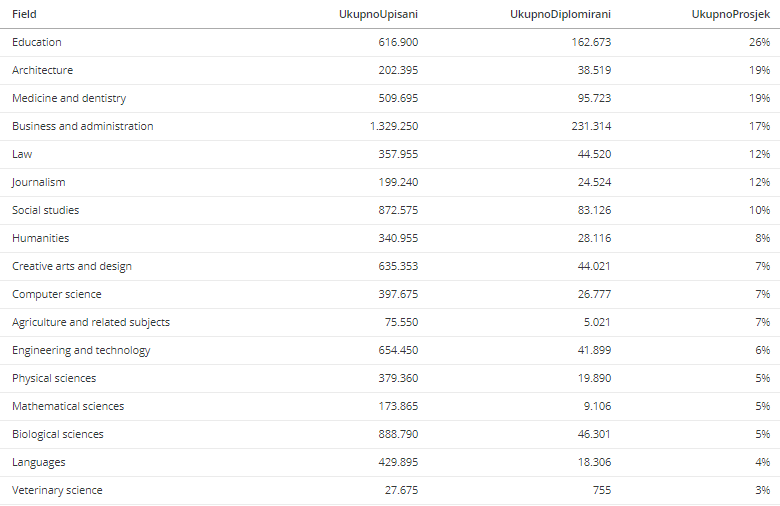
Slika 25. Prosjek prvostupnika

Interpretacija: Vizualizacija je izrađena pomoću alata *microstrategy*. Prosjek prvostupnika, kao i magistara u prošlom poglavlju, dobiven je na način da ukupan broj prvostupnika područja podijelimo sa ukupnim brojem upisanih studenata/ica za određeno područje. Vidimo da je najveći prosjek prvostupnika u odnosu na broj upisanih studenata redom u području likovne umjetnosti, humanističkih znanosti, arhitekture i novinarstva (31 – 25 %). U odnosu na prosjek magistara, jedino područje koje pripada području s višim prosjekom magistara i prvostupnika jest područje arhitekture.

Najmanji prosjek prvostupnika u odnosu na broj upisanih studenata/ice pripada područjima agrikulture, inženjerstva i edukacije (područje edukacije je s prosjekom prvostupnika pri dnu liste, dok je s prosjekom magistara jedno od područja s većim prosjekom magistara u odnosu na broj upisanih studenata/ica).

## **4.11.Ukupan prosjek diplomiranih prema području studiranja u razdoblju od 2014. do 2017. godine**

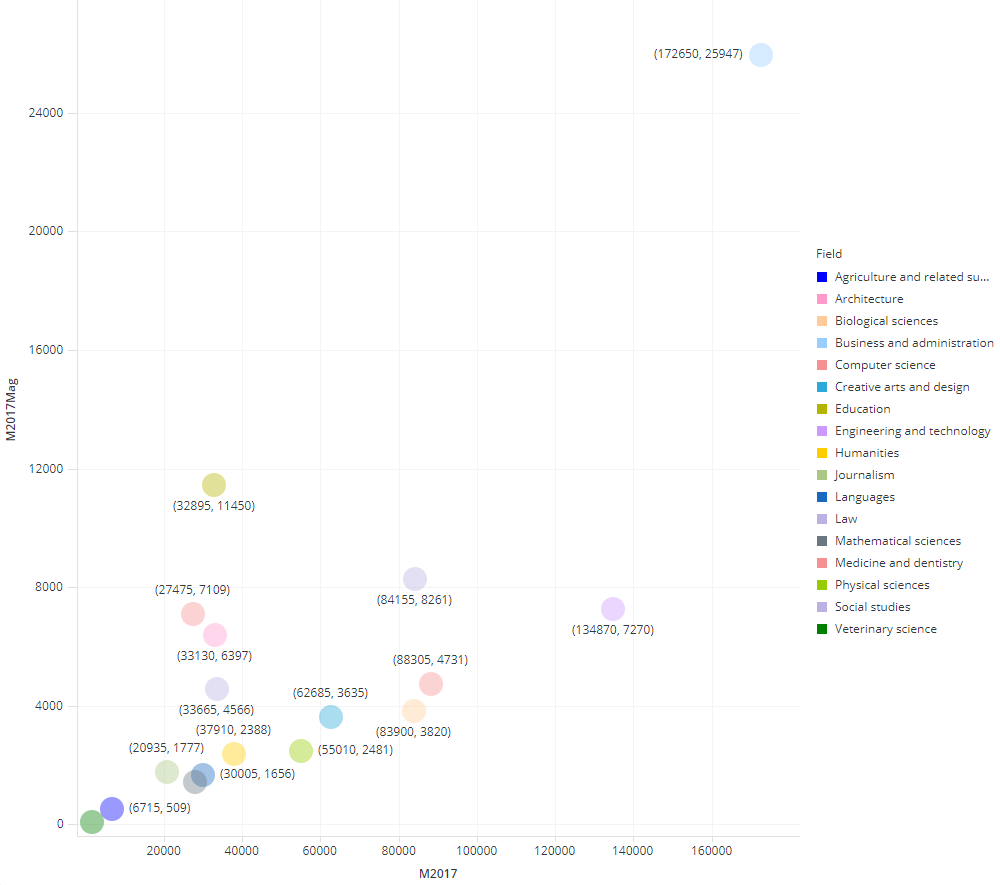
Pitanje: Koliko iznosi ukupan prosjek magistara (muških i ženskih) u odnosu na ukupan broj upisanih studenata u razdoblju od 2014. do 2017. godine?



Interpretacija: Vizualizacija je izrađena pomoću alata *microstrategy.* Kako u tablici nemamo ukupan broj upisanih i diplomiranih već su odvojeni prema spolovima, metrika *UkupnoUpisani* izračunata je na način da sumu upisanih ženskih studenata zbrojimo sa sumom upisanih muških studenata. Sukladno tome, metrika *UkupnoDiplomirani* izračunata je na način da sumu ženskih magistara za naredno razdoblje zbrojimo sa sumom muških magistara u navedenom periodu. Iz vizualizacije je vidljivo da je ukupan prosjek magistara najveći u područjima edukacije, arhitekture i medicine (26 – 19%) dok je isti najmanji u područjima jezika, bioloških znanosti i veterine što znači da u ova tri navedena područja, s obzirom na broj upisanih studenata, diplomira najmanji broj istih u odnosu na druga područja studiranja.

## **4.12. Odnos upisanih i diplomiranih studenata u 2017.godini**

Pitanje: Kakav je odnos između upisanih i diplomiranih **studenata** u 2017. godini?



Slika 26. Odnos upisanih i diplomiranih studenata za 2017. godinu

Interpretacija: Vizualizacija je izrađena pomoću alata *microstrategy*. Na x osi vizualizacije nalazi se ukupan broj upisanih studenata (isključivo muških) za 2017. godinu, dok se na y osi nalazi ukupan broj diplomiranih studenata u 2017. godini.

Analiza po područjima: U području studiranja edukacija diplomirao je dosta velik broj studenata u odnosu na broj upisanih (34% - kad podijelimo 11 450 sa 32 895). Ekonomiju je upisalo i diplomiralo velik broj studenata no prosjek magistara u području u odnosu na broj upisanih studenata iznosi tek 15%. Broj od 25 947 magistara jest velik sam po sebi i u usporedbi s brojem magistara unutar drugih područja ali čini samo 15% ukupnog broja upisanih studenata u području ekonomije. Inženjerstvo i tehnologiju upisao je velik broj studenata, no iz vizualizacije vidimo da je malen broj diplomirao (5% ukupnog broja upisanih te godine). Što se područje nalazi niže na y osi i teži desnoj strani x osi na grafikonu to je manji prosjek magistara u području i obrnuto (što se područje nalazi više na y osi i teži lijevoj strani x osi to je prosjek magistara veći).

# **Literatura**

1. HESA: *What to HE students study?.* Dostupno na: <https://www.hesa.ac.uk/data-and-analysis/students/what-study> [05.06.2020.]
2. MindMajix : *Microstrategy tutorial.* Dostupno na: <https://mindmajix.com/microstrategy-tutorial> [05.06.2020.]
3. OECD.Stat: *Graduates by field.* Dostupno na: [https://stats.oecd.org/viewhtml.aspx?datasetcode=EDU\_GRAD\_FIELD&lang=en#](https://stats.oecd.org/viewhtml.aspx?datasetcode=EDU_GRAD_FIELD&lang=en) [05.06.2020]
4. *Open Studio for Data Integration User Guide* (skinuto sa merlin stranice kolegija). Dostupno na: <https://moodle.srce.hr/2019-2020/pluginfile.php/3364532/mod_resource/content/3/Vjezbe/TalendOpenStudio_DI_UG_6.1.1_EN.pdf> [05.06.2020]
5. Tutorialspoint: *Tableau tutorial.* Dostupno na: <https://www.tutorialspoint.com/tableau/index.htm> [05.06.2020.]