

# Prikupljanje podataka o upisu u srednje škole

LUKA JOVIČIĆ, 4E

MATEMATIČKA GIMNAZIJA, BEOGRAD

# Šta?

- ▶ Podaci o osnovcima, od 2015.
- ▶ Ocene, bodovi na završnom, prijemnima, nagrade na takmičenjima, liste želja, pohađana i upisana škola, mesto/opština/okrug
- ▶ Inicijalno u klasičnim fajlovima
- ▶ Portovano u bazu podataka (PostgreSQL)
- ▶ Pristupanje bazi kroz Javu i Python
- ▶ Značaj: postoje izveštaji, ali sami podaci *nisu otvoreni*
- ▶ Mogućnost nezavisnih analiza sistema

# Kako preuzeti podatke?

- ▶ Skreper u Javi
- ▶ Projekat koji je evoluirao od 2015. do 2018.
- ▶ Imitira zahteve koje pravi pretraživač
- ▶ Manipulacija DOMom
- ▶ Stara varijanta – izvlači podatke iz tabela
- ▶ Nova varijanta – parsira vrednosti promenljivih iz Javascripta
- ▶ Deljeni kod – pristup fajlovima, upravljanje preuzimanjem

# Kako napraviti bazu?

- ▶ Play! Framework (serverski okvir)
- ▶ Skreper uključen kao biblioteka – učitava podatke iz fajlova
- ▶ Prebacivanje u model pogodan za ORM
- ▶ Automatsko mapiranje u SQL tabele i generisanje upita
- ▶ Lako za održavanje i proširivanje
- ▶ Mogućnost pravljenja serverskih aplikacija
- ▶ Primer: query language/plotter

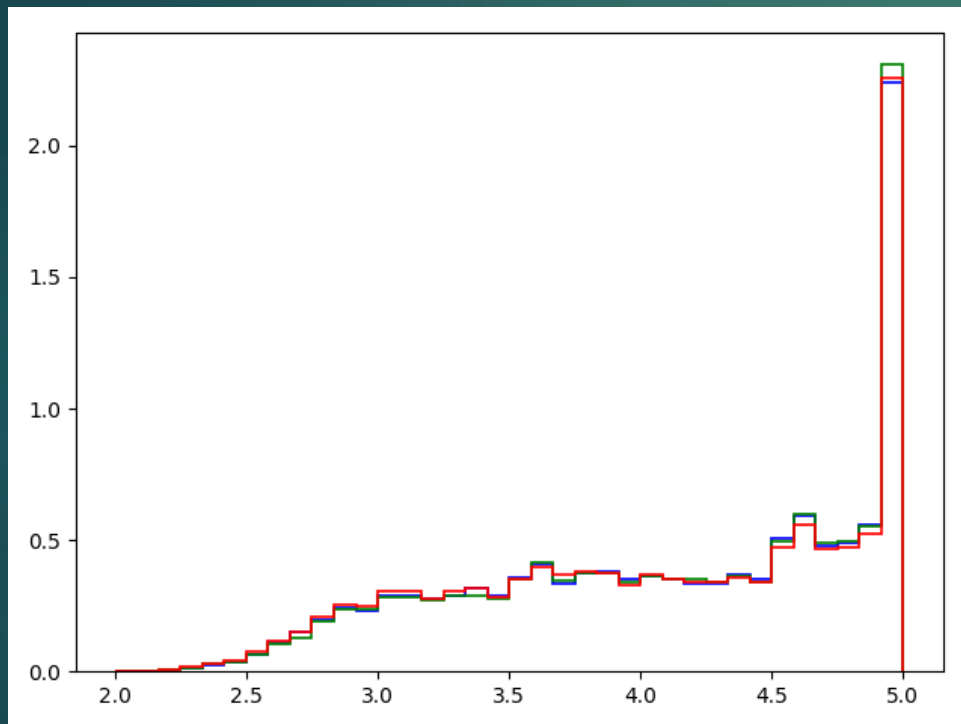
# Kako iskoristiti podatke?

- ▶ Skreper se može proširiti metodama koje se pozivaju iz komandne linije
- ▶ Pristup bazi kroz server (Java/Scala)
- ▶ Direktan pristup bazi (SQL, ...)
- ▶ Simulacije
- ▶ Replikacije Izveštaja o realizaciji završnog ispita
- ▶ Međusobni odnosi i raspodele ocena, bodova, veličine/mesta osnovne škole, upisanog profila, područja rada, ...
- ▶ Traženje outliera

# Primeri obrade

Histogram raspodele ocena;

Prosečna ocena i devijacija po području rada smera



područje	prosek	stddev	učenici
zdravstvoisocijalnazastita	4.7323156585995	0.29769133688249	5633
gimnazija	4.6868037142754	0.25320496107733	16181
kulturaumetnostijavnoinformisanje	4.2438174603175	0.23829378987084	420
ekonomijapravoadministracija	4.2073564247453	0.34827775342308	7785
elektrotehnika	4.1305000672621	0.54592496773668	6974
hidrometeorologija	4.0672777777778	0.26186521129941	60
geodezijaigradjevinarstvo	3.8367583101852	0.38468573121657	1592
saobracaj	3.8296052562641	0.53588049696118	3603
hemijanemetaliigrafičarstvo	3.8238149869744	0.3018139359520	2175
trgovinaugostiteljstvoiturizam	3.6454255158916	0.49247371033996	4470
poljoprivredaproizvodnjaipreradahrane	3.5587516428161	0.3493371019462	2869
masinstvoibradametala	3.5322565671368	0.39116804671372	4887
sumarstvoibradadrвета	3.4383038566738	0.4009867386585	579
geologijarudarstvoimetalurgija	3.3724983164983	0.35095700883105	327
ostaladelatnostlicnihusluga	3.3655417363095	0.30471046454634	679
tekstilstvoikozarstvo	3.2847209182599	0.30670704991301	710

# I šta s tim?

- ▶ Pokazano kako više jezika i ekosistema intereaguju
  - ▶ Projekat koji je "izdržao" tri godine i nekoliko nadogradnji
  - ▶ Real-world uslovi i podaci
- 
- ▶ Baza je javno dostupna
  - ▶ Mogućnost detaljnijih i rigoroznijih analiza
  - ▶ Podaci su po prvi put *otvoreni*

Hvala na pažnji