

Analiza podataka iz Stack Overflow 2024 ankete

Ranka Radulović

Predmet: Sistemi baza podataka

Sara Svrkota

- Dataset:

Stack Overflow Annual Developer Survey 2024

- Broj učesnika:

~65,000 developera iz celog sveta

- Broj kolona: **144**

- Format: **CSV**

- Cilj ankete:

Prikupiti informacije o developerima, njihovim veštinama, iskustvu, obrazovanju, platama, tehnologijama i zadovoljstvu poslom

Opis skupa podataka

Link do dataseta:

<https://www.kaggle.com/datasets/berkayalan/stack-overflow-annual-developer-survey-2024/data>

SEMANTIKA KOLONA

- **Country** - Zemlja developera
- **Age** - Godine developera (po kategorijama npr. 35-44)
- **YearsCodePro** - Godine profesionalnog iskustva
- **EdLevel** - Obrazovanje (Bachelor/Master/PhD)
- **MainBranch** - Glavna branša
- **OrgSize** - Veličina kompanije
- **ConvertedCompYearly** - Godišnja plata u USD
- **Employment** - Tip zaposlenja
- **RemoteWork** - Rad na daljinu / onsite
- **LearnCode** - Kako je učio da programira?
- **SOVisitFreq** - Koliko često posećuje SO
- **TimeSearching** - Vreme utrošeno na pretragu rešenja
- **JobSat** - Zadovoljstvo na poslu
- **Industry** - Kojoj industriji pripada developer

SEMANTIKA KOLONA

- **LanguageHaveWorkedWith** – Programski jezici koje developer koristi
- **PlatformHaveWorkedWith** – Platforme koje developer koristi
- **AISelect** – Plan korišćenja AI
- **DatabaseHaveWorkedWith** – Baze podataka koje developer koristi
- **AISelectDevHaveWorkedWith** – AI koje developer koristi
- **AISelectDevWantToWorkWith** – AI koje koje bi developer koristio
- **WebframeHaveWorkedWith** – WebFrame koji developer koristi

Kolekcija developers

	Value	Type
<code>id</code> _id	69763507ea6dae34db5190f9	ObjectId
<code>i32</code> ResponsId	380	Int32
<code>"</code> MainBranch	I am a developer by profession	String
<code>"</code> Age	35-44 years old	String
<code>[]</code> Employment	[1 elements]	Array
<code>"</code> 0	Employed, full-time	String
<code>"</code> RemoteWork	Remote	String
<code>"</code> EdLevel	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	String
<code>"</code> Country	Turkey	String
<code>i32</code> YearsCodePro	6.0	Double
<code>[]</code> LearnCode	[3 elements]	Array
<code>"</code> 0	Books / Physical media	String
<code>"</code> 1	Other online resources (e.g., videos, blogs, forum, online community)	String
<code>"</code> 2	Online Courses or Certification	String
<code>"</code> OrgSize	20 to 99 employees	String
<code>"</code> SOVisitFreq	A few times per month or weekly	String
<code>i32</code> ConvertedCompYearly	91295.0	Double
<code>"</code> TimeSearching	Less than 15 minutes a day	String
<code>i32</code> JobSat	10.0	Double
<code>"</code> Industry	Computer Systems Design and Services	String

Kolekcija technologies

Key	Value	Type
ResponsesId	41	Int32
LanguageHaveWorkedWith	[4 elements]	Array
0	HTML/CSS	String
1	JavaScript	String
2	Ruby	String
3	SQL	String
DatabaseHaveWorkedWith	[3 elements]	Array
0	PostgreSQL	String
1	Redis	String
2	SQLite	String
PlatformHaveWorkedWith	[4 elements]	Array
0	Amazon Web Services (AWS)	String
1	Fly.io	String
2	Google Cloud	String
3	Render	String
AISelect	Yes	String
AISearchDevHaveWorkedWith	[1 elements]	Array
0	GitHub Copilot	String
AISearchDevWantToWorkWith	[1 elements]	Array
0	GitHub Copilot	String
WebframeHaveWorkedWith	[2 elements]	Array
0	React	String
1	Ruby on Rails	String

Predlozi agregacija

1. Analiza plata po radnom iskustvu i obrazovanju:

Izračunajte prosečnu platu za developere sa 5 do 10 godina iskustva (YearsCodePro) u svakoj zemlji, ali samo za one koji su završili fakultet (Edlevel polje). Sortirajte rezultate od najviše do najniže prosečne plate i prikažite samo prvih 10 zemalja.

2. Korelacija popularnosti programskog jezika i plate:

Koja je prosečna godišnja plata (ConvertedCompYearly) za developere koji koriste i Python i JavaScript u poređenju sa onima koji koriste samo Python ili samo JavaScript?

3. Analiza trendova po tipu zaposlenja:

Izračunajte prosečno radno iskustvo (YearsCodePro) za svaku vrstu zaposlenja (Employment), ali samo za one koji rade isključivo na daljinu (RemoteWork = 'Remote'). Pored toga, izračunajte i minimalnu i maksimalnu platu za svaku od tih grupa. Prikazati i po grupama developere koji koriste, ne koriste i ne planiraju da koriste AI (AISelect = "No....")

4. Popularnost baza podataka u odnosu na ispitanike koji su na studijama:

Koje su tri najčešće korišćene baze podataka (DatabaseHaveWorkedWith) među ispitanicima koji još uvek nisu završili fakultet, odnosno koji se nalaze na nivou obrazovanja „Some college/university study without earning a degree“ u odnosu na one koji se nalaze na nivou obrazovanja "Secondary school (e.g. American high school, German Realschule or Gymnasium, etc.)" (hobi/profesionalno) ?

5. Dubinska analiza veštačke inteligencije (AI):

Grupišite developere po tome da li trenutno koriste alate za veštačku inteligenciju (AISelect = 'Yes') i izračunajte prosečnu platu za svaku od te dve grupe (Yes/No) u Srbiji i izračunati procenat onih koji uče preko knjiga i onih koji su stekli znanje od prijatelja (LearnCode). Sortirajte rezultate od najviše do najniže prosečne plate.

Predlozi agregacija

6. Analiza trendova AI alata u odnosu na plate:

Koji su AI alati (AISeachDevHaveWorkedWith) povezani sa najvećom prosečnom platom (prikazati prvih 5) i prikazati i one developere koji su označili da su ucili *Other online* (LearnCode) u odnosu na one koji su označili da su učili preko Coding Bootcamp (LearnCode)?

7. Korelacija tipa posla sa radnim okruženjem i tehnologijom:

Izračunajte prosečno radno iskustvo (YearsCodePro) za developere u kompanijama sa različitim brojem zaposlenih (OrgSize) (npr. '20-99', '100-499', '1,000-4,999'). Zatim, za naj grupu, prikažite tri najpopularnija programska jezika.

8. Analiza migracije radne snage:

Pronađi programere koji posećuju StackOverflow sajt svakodnevno (SOVisitFreq = 'Daily or almost daily') u odnosu na one koji posećuju SOVisitFreq = A few times per month i uporediti njihov prosečni plato i prikazi njihovu prosečnu platu i radni staz i grupisati po zemljama.

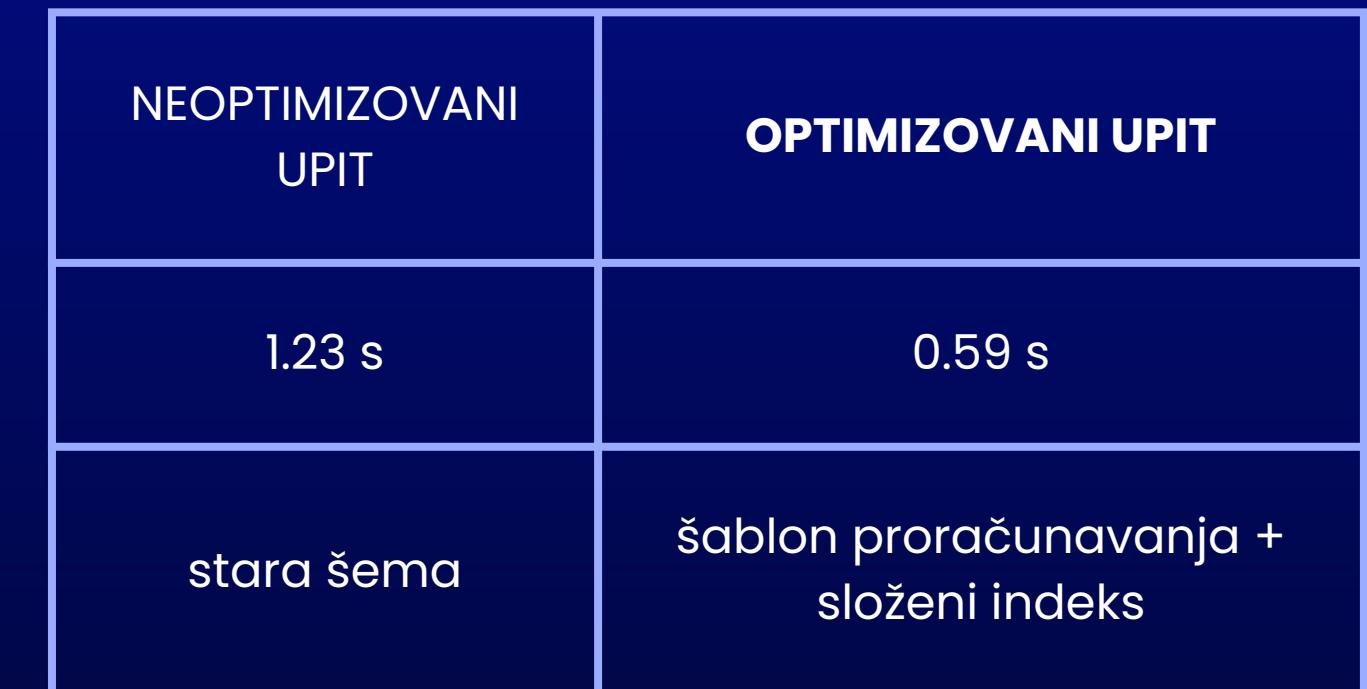
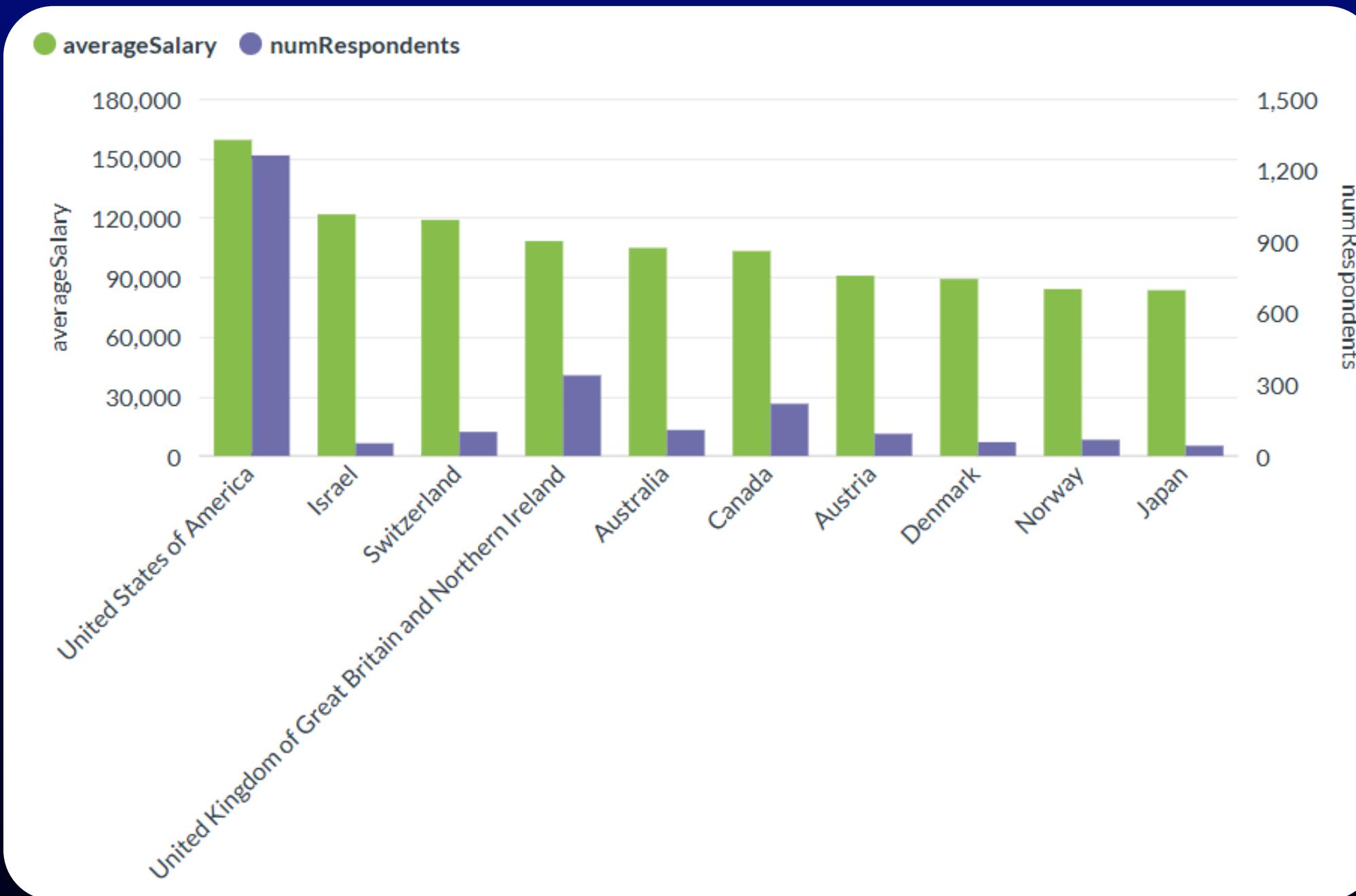
9. Iskustvo i popularnost frameworka:

Napravite upit kako bi se ispitala korelacija između nivoa radnog iskustva, vremena utrošenog na pretragu rešenja (TimeSearching) i učestalosti korišćenja Stack Overflow platforme (SOVisitFreq).

10. Dubinska analiza odnosa između industrije i zadovoljstva:

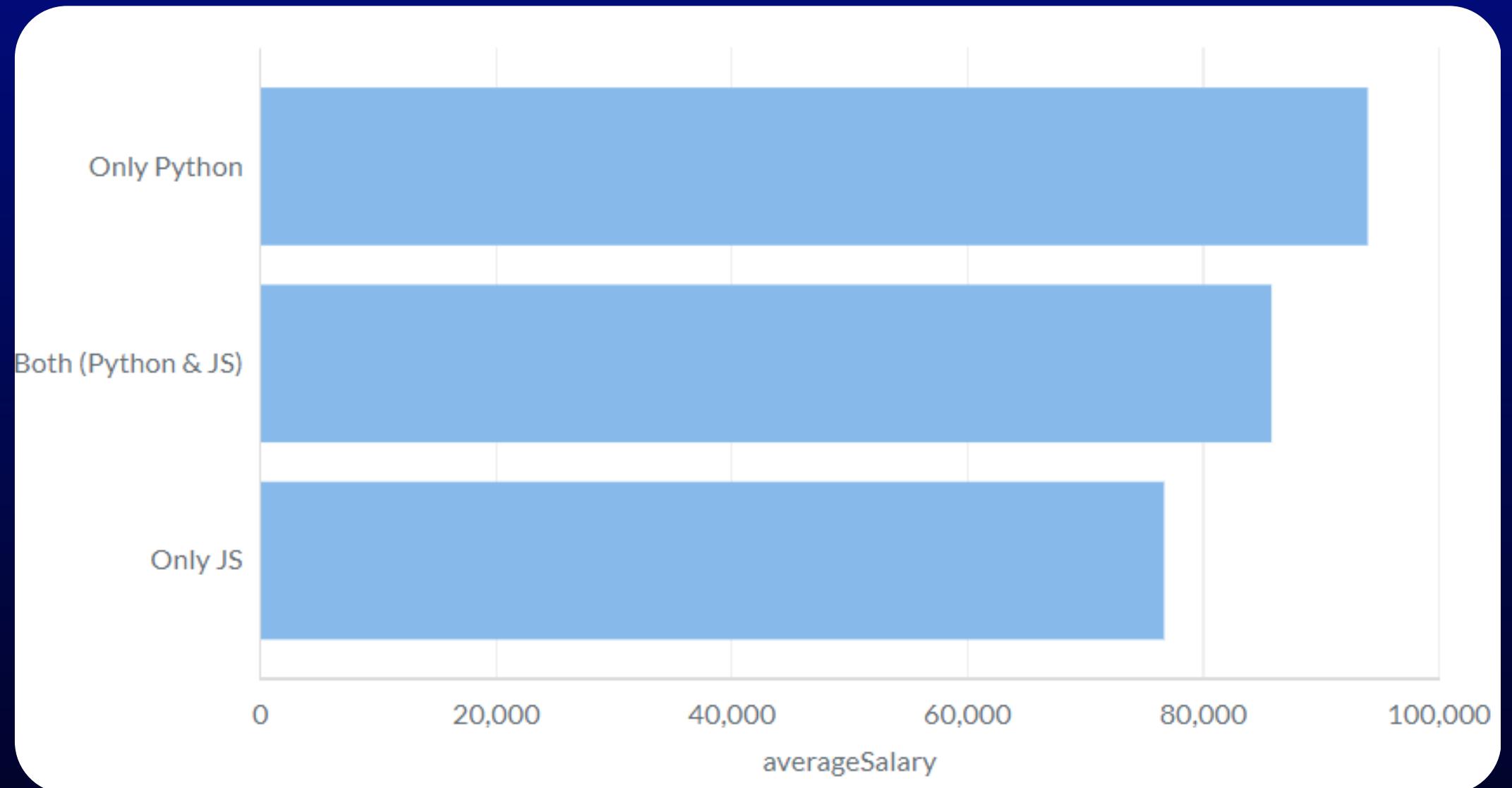
Izračunati za svaku industriju prosečnu platu, prosečno zadovoljstvo poslom. Prikazati najkorisćeniju bazu podataka i procenat zaposlenih koji su veoma zadovoljni poslom (JobSat - 9 ili 10).

1 Top 10 zemalja sa najvećom prosečnom platom gde su izabrani samo developeri koji su fakultetski obrazovani sa 5 – 10 godina ravnog iskustva.



country	averageSalary	numRespondents
United States of America	159,134.84	1,261
Israel	121,648.49	53
Switzerland	118,852.84	101
United Kingdom of Great Britain and Northern Ireland	108,181.93	338
Australia	104,775.03	109
Canada	103,190.75	219
Austria	90,766.84	93
Denmark	89,135.22	58
Norway	84,043.62	68
Japan	83,451.14	43

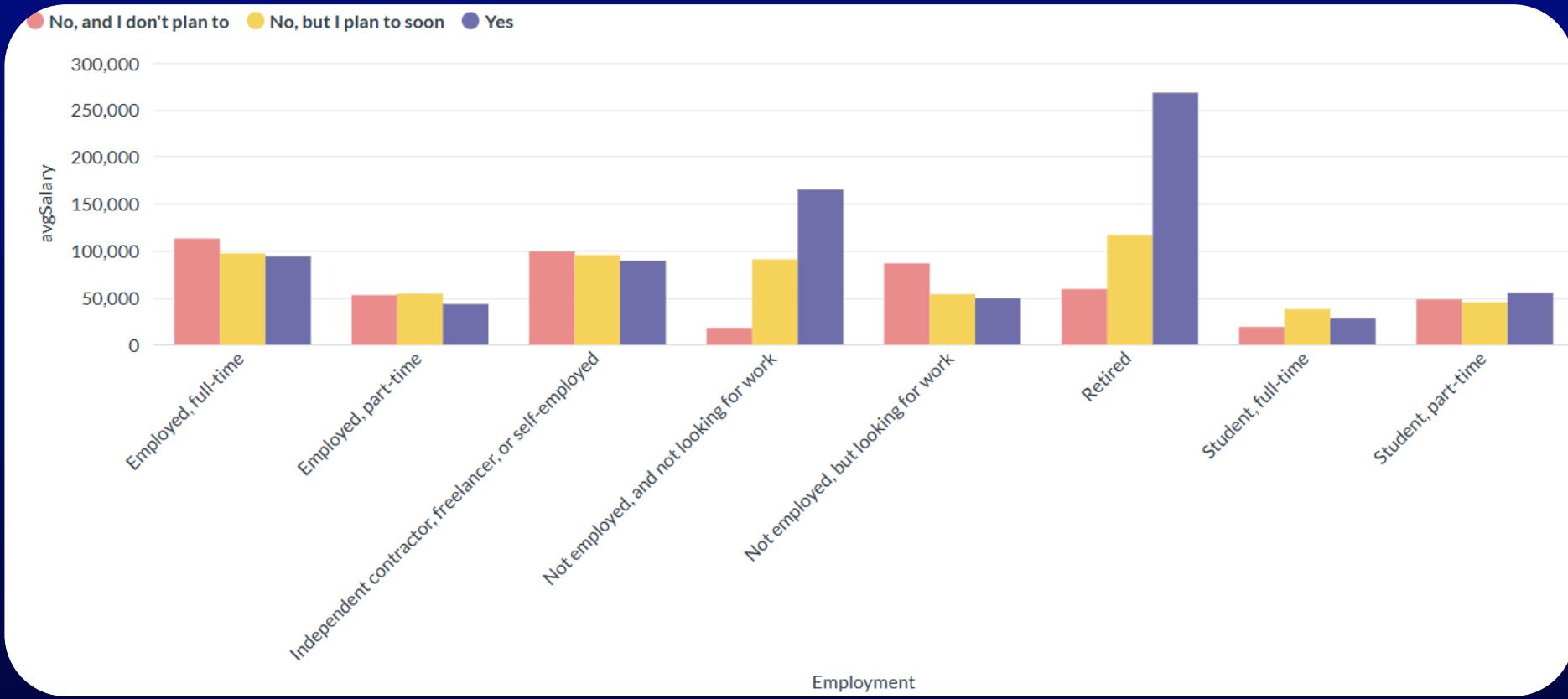
2. Prikaz prosečne godišnje plate za developere koji koriste samo Python, samo JS i za one koji koriste oba jezika.



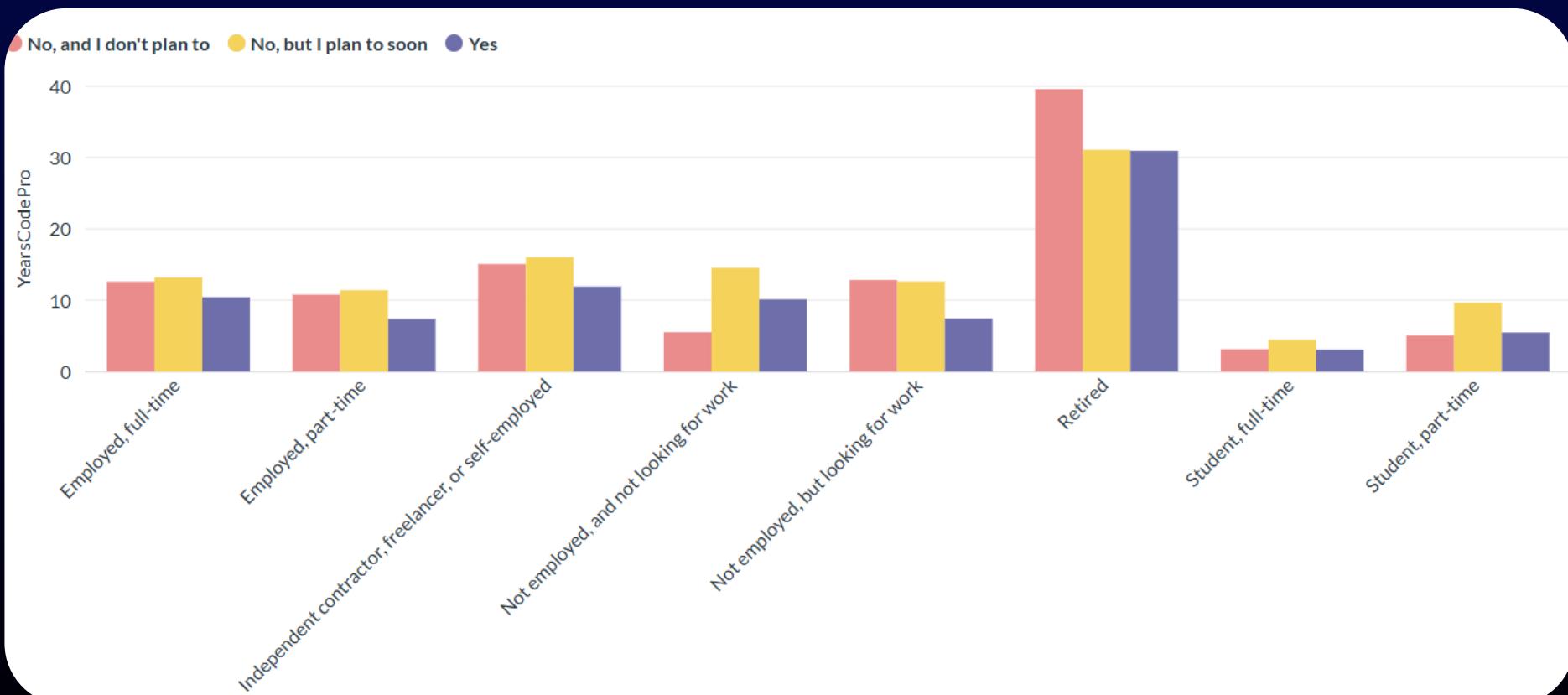
NEOPTIMIZOVANI UPIT	OPTIMIZOVANI UPIT
8.67 s	0.40 s
stara šema	šablon proširene reference

_id	averageSalary	count
Only Python	93,954.71	4,451
Both (Python & JS)	85,780.47	6,956
Only JS	76,669.65	7,798

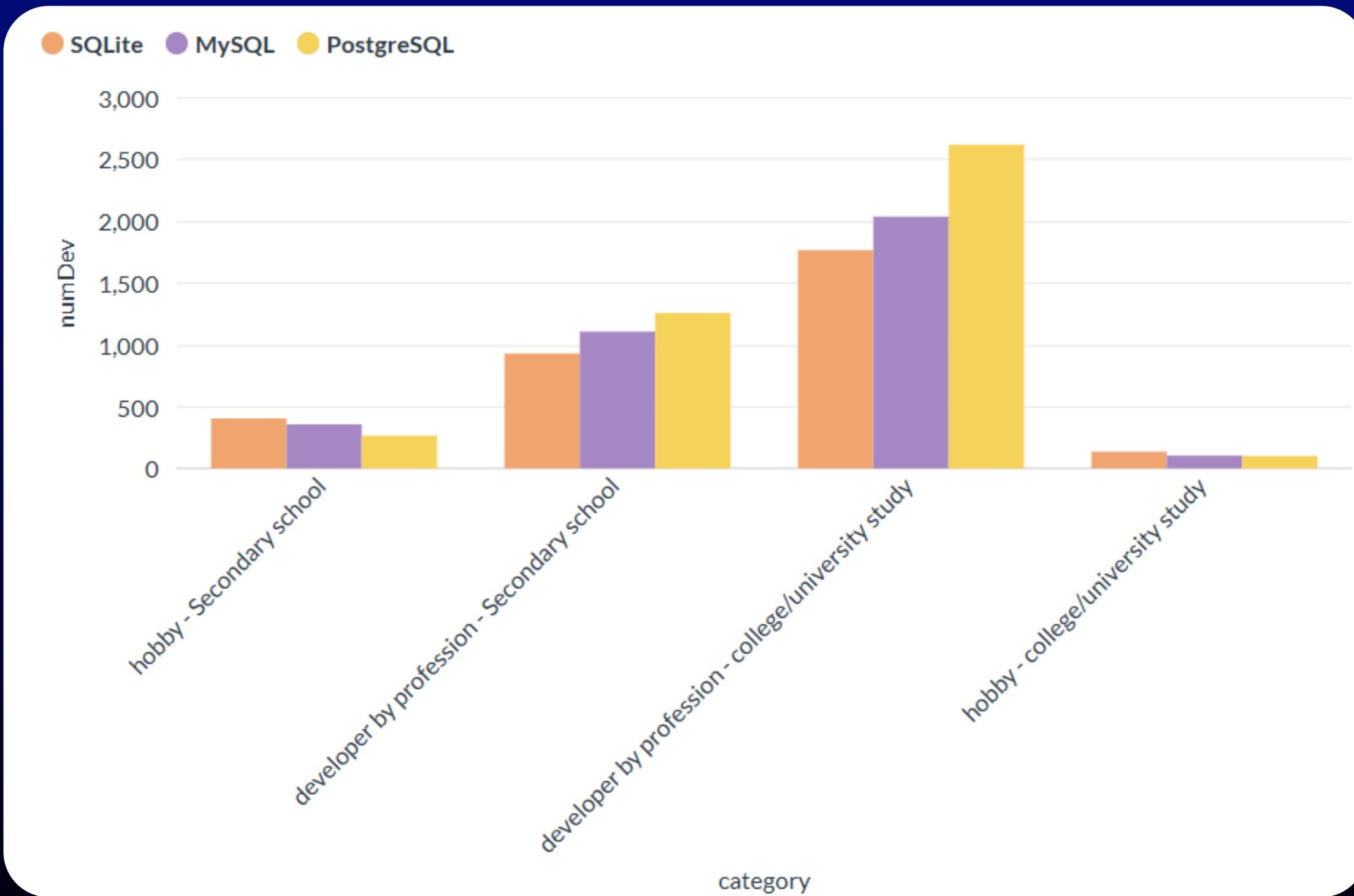
3. Prosečno radno iskustvo za svaku vrstu zaposlenja ali samo za one koji rade isključivo na daljinu (RemoteWork = 'Remote'). Prosečna plata za svaku od tih grupa. Po grupama developeri koji koriste, ne koriste i ne planiraju da koriste AI.



NEOPTIMIZOVANI UPIT	OPTIMIZOVANI UPIT
3.20 s	0.29 s
stara šema	šablon proširene reference



4. Top 3 najčešće korišćene baze podataka među ispitanicima koji još uvek nisu završili fakultet u odnosu na one koji se nalaze na nivou obrazovanja Srednje škole.



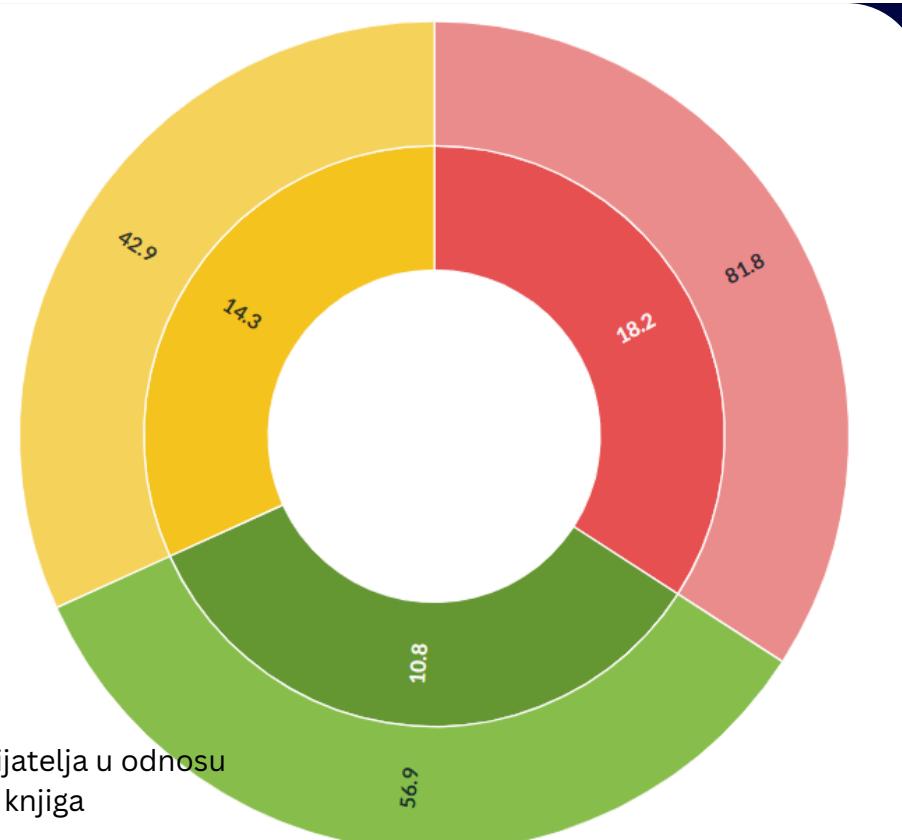
NEOPTIMIZOVANI UPIT	OPTIMIZOVANI UPIT
2.66 s	0.28 s
stara šema	šablon proširene reference

5. Prikaz prosečne plate u Srbiji za developere koji (ne) koriste AI kao i procenat onih koji uče od prijatelja i onih koji uče iz knjiga.



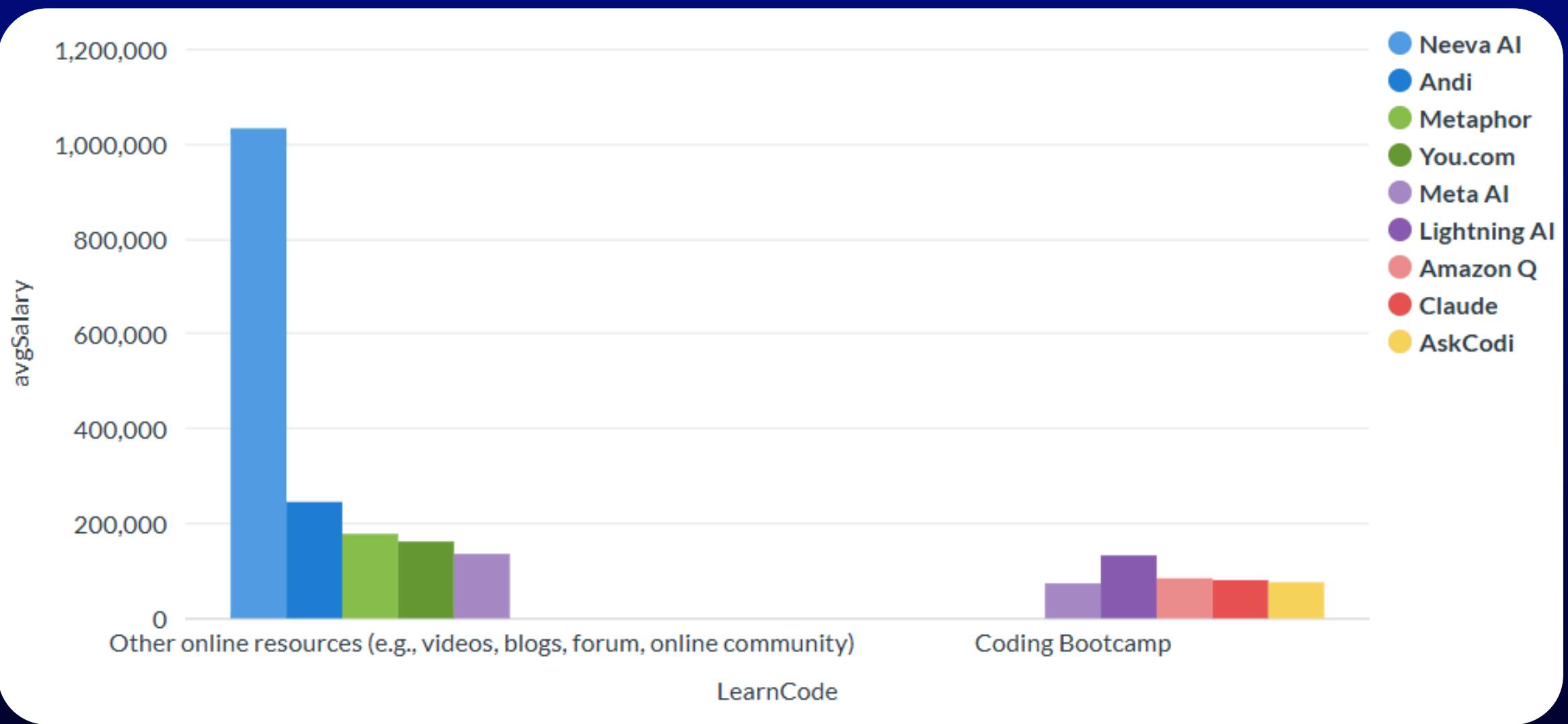
No, and I don't plan to
22 46,448 100 %

NEOPTIMIZOVANI UPIT	OPTIMIZOVANI UPIT
6.93 s	1.94 s
stara šema	šablon proširene reference



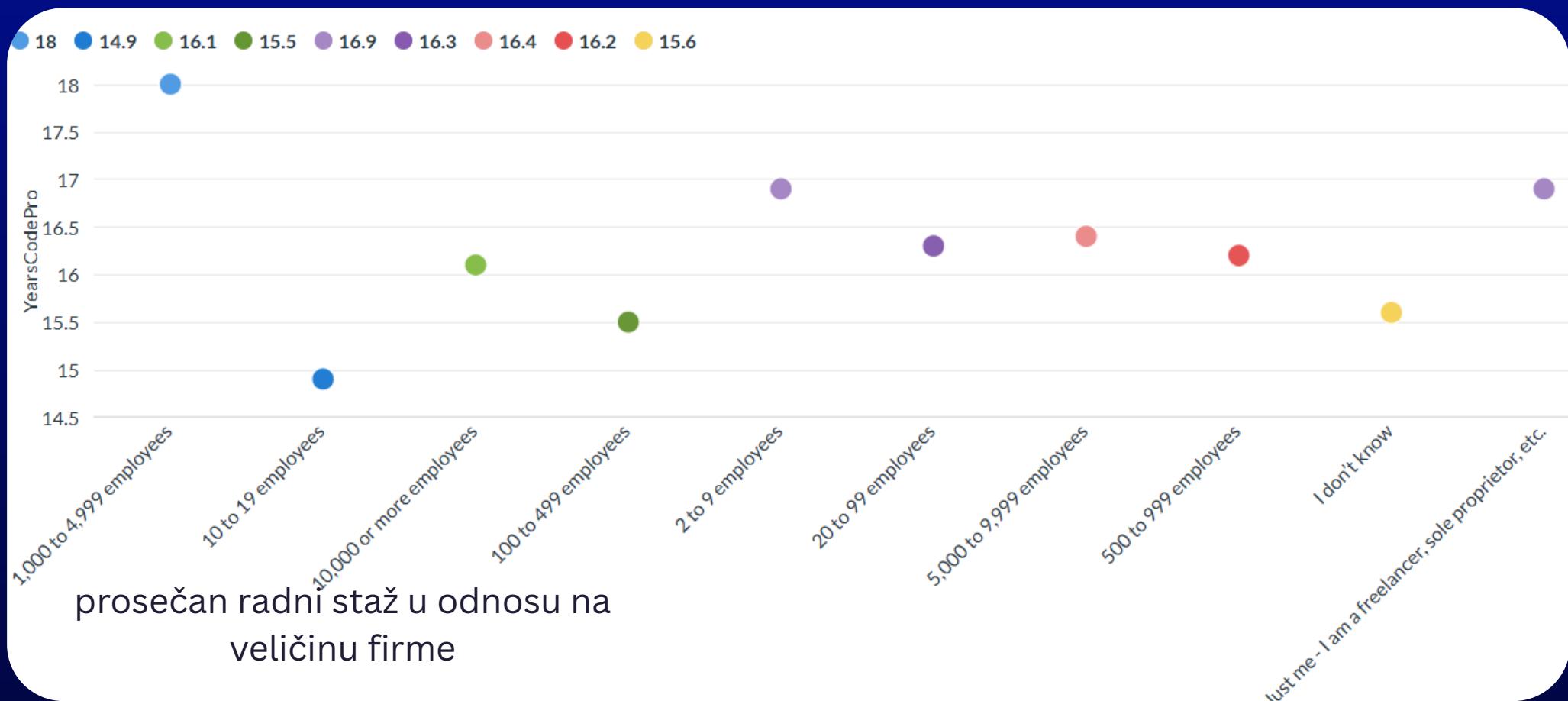
KoristiAI	Plata	Procet_Knjige	Procenat_Prijatelji	BrojIspitanika
No, and I don't plan to	46,448	81.8	18.2	22
Yes	46,363	56.9	10.8	65
No, but I plan to soon	43,231	42.9	14.3	7

6. Top 5 AI alata koji su povezani sa najvećom prosečnom platom (prikazati prvih 5) i koliko zavisi da li su developeri učili preko Online-a ili preko kampa.

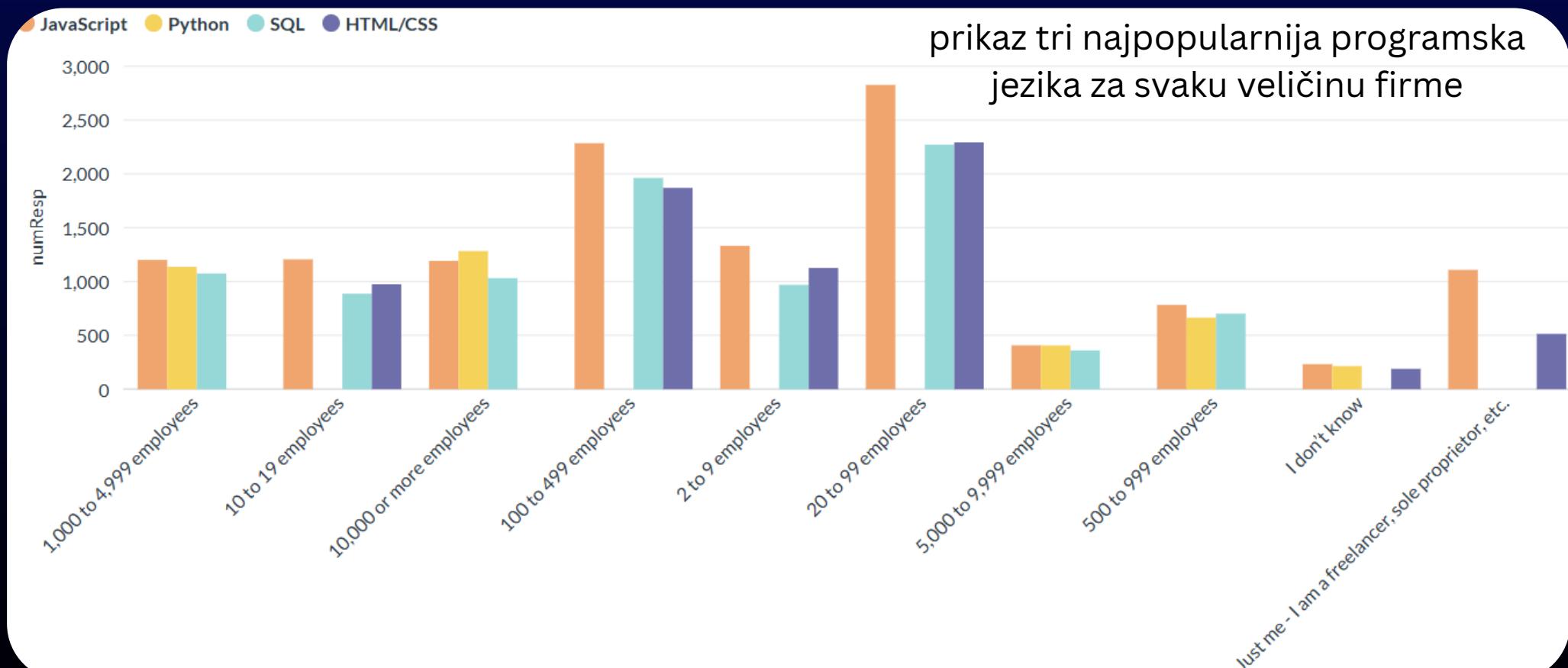


NEOPTIMIZOVANI UPIT	OPTIMIZOVANI UPIT
8.00 s	0.94 s
stara šema	šablon proširene reference

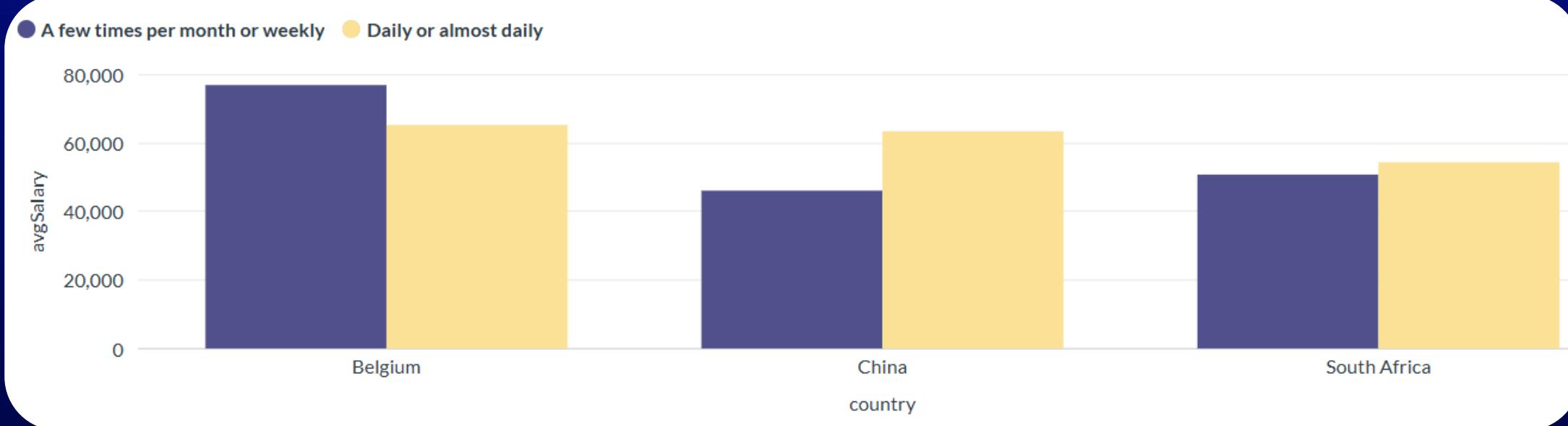
7. Prosečno radno iskustvo za developere u kompanijama sa različitim brojem zaposlenih prikažite tri najpopularnija programska jezika za grupu i koja je najčešća po godinama.



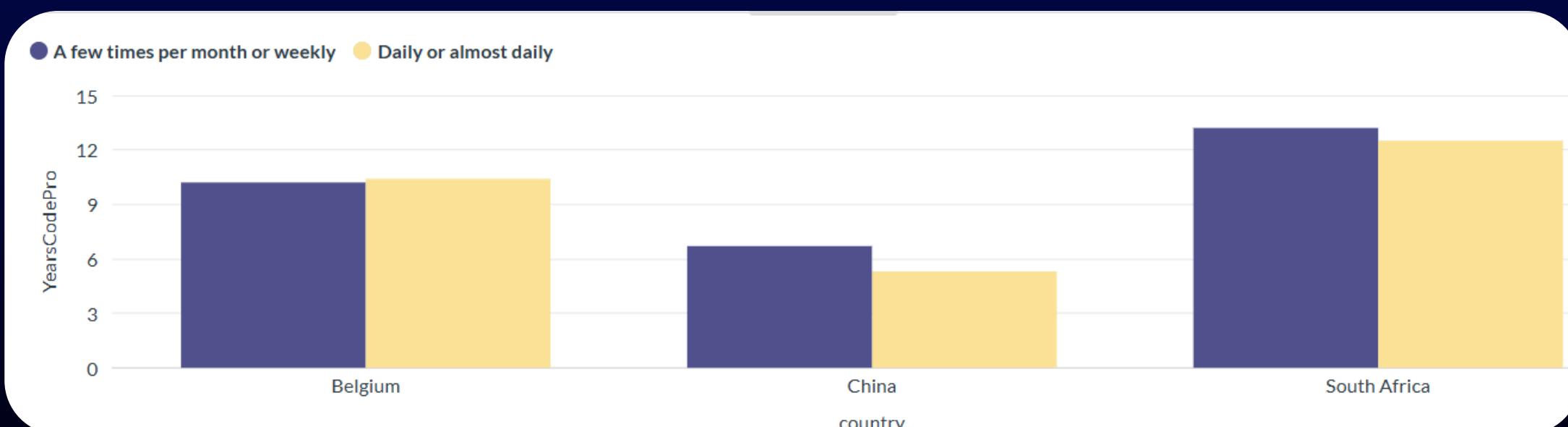
NEOPTIMIZOVANI UPIT	OPTIMIZOVANI UPIT
16.937 s	0.04 s
stara šema	šablon proračunavanja



8. Prikaz i poređenje programera iz Belgije, Kine i Južnoafričke Republike koji posećuju StackOverflow sajt svakodnevno u odnosu na one koji posećuju par puta mesečno i poređenje proseka plata i radnog staža .

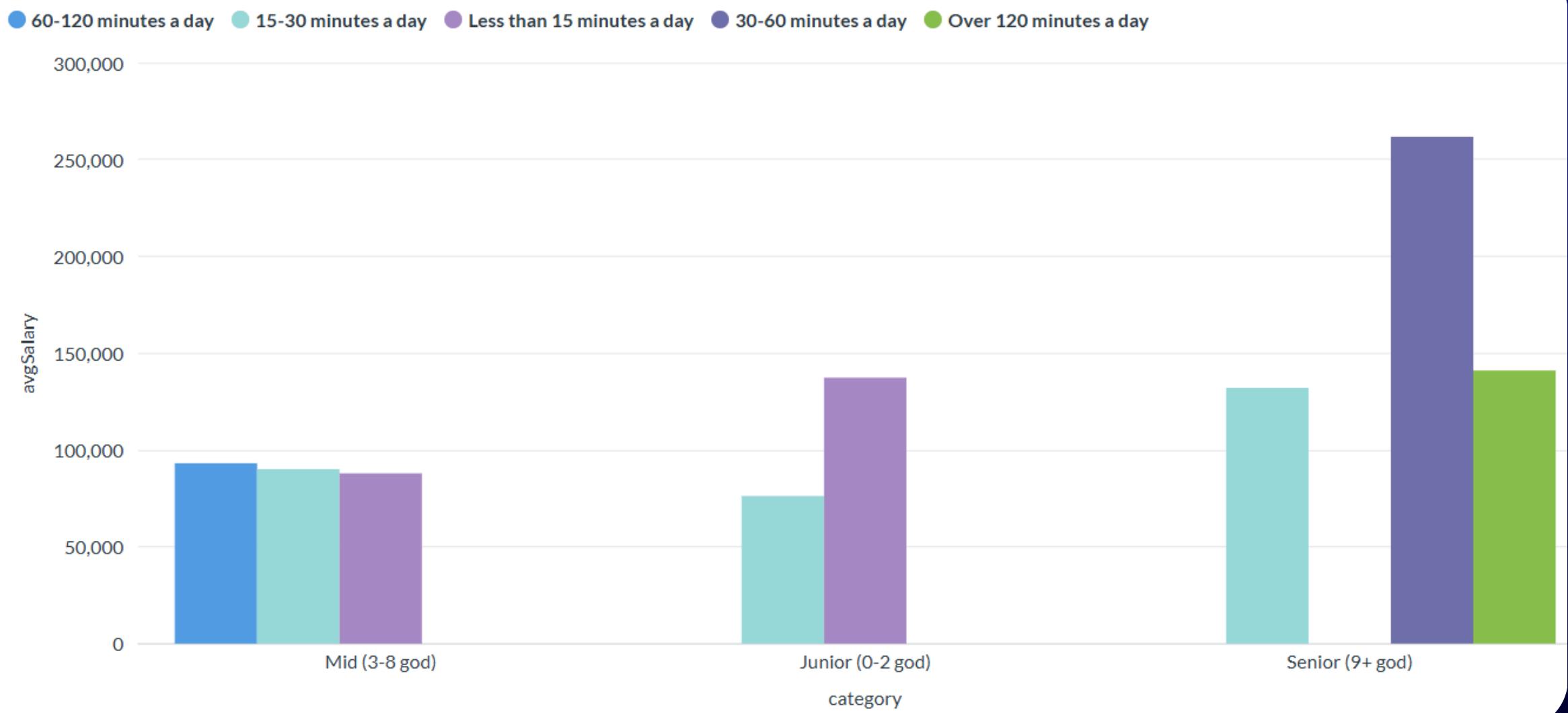


NEOPTIMIZOVANI UPIT	OPTIMIZOVANI UPIT
0.34 s	0.08 s
stara šema	šablon proračunavanja



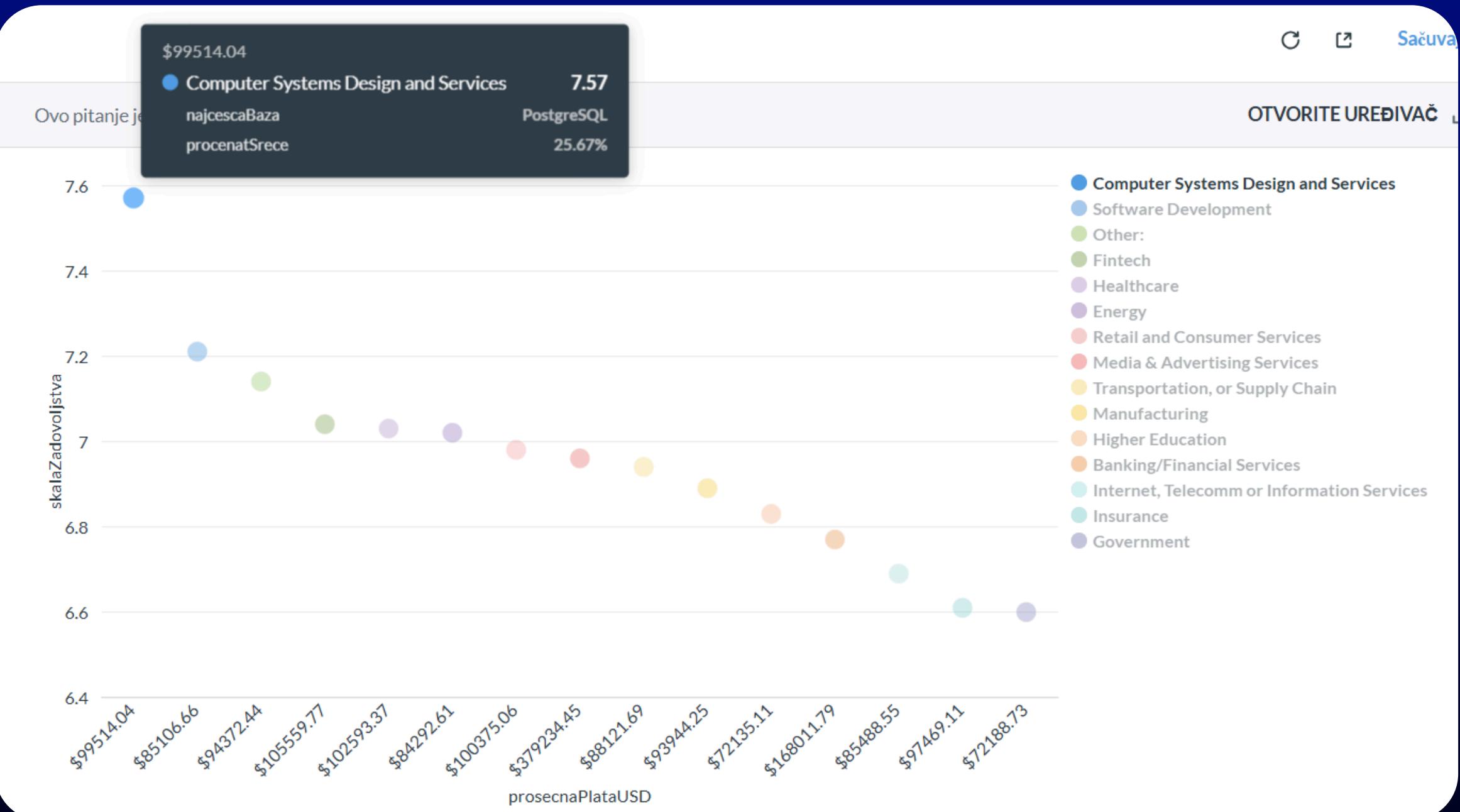
drzava	tipPosetioca	plata	staz	brojLjudi
Belgium	A few times per month or weekly	76,789.82	10.2	45
Belgium	Daily or almost daily	65,137.64	10.4	56
China	A few times per month or weekly	46,007.38	6.7	13
China	Daily or almost daily	63,273.65	5.3	23
South Africa	A few times per month or weekly	50,690.12	13.2	41
South Africa	Daily or almost daily	54,266.03	12.5	32

9. Prikaz developera po senioritetu i prosečnoj plati u zavisnosti od toga koliko koriste StackOverflow.



NEOPTIMIZOVANI UPIT	OPTIMIZOVANI UPIT
0.75 s	0.33 s
stara šema	novo polje (šablon proračunavanja)

10. Prikaz prosečne plate za svaku industriju i prosečno zadovoljstvo poslom. Prikazane najkorisćenija baza podataka i procenat zaposlenih koji su veoma zadovoljni poslom (JobSat - 9 ili 10).



NEOPTIMIZOVANI UPIT	OPTIMIZOVANI UPIT
3.98 s	0.30 s
stara šema	šablon proširene reference

REZULTATI

