**UVOD U PODATKOVNU ZNANOST**

**plan seminarskog rada**

**akademska godina 2023./2024.**

**Autori seminarskog rada:** Laura Boras i Sunčica Prgomet

**Naslov seminarskog rada:** Analiza plaća u području podatkovne znanosti

**Skup podataka:** <data-science-salaries-2023.htm>

**Opis analize:**

Skup podataka sadrži informacije o godini rada, razini iskustva, tipu zaposlenja, vrsti zaposlenja, visini plaće, zemlji boravka zaposlenika, postotku obavljanja rada od kuće, lokaciji te veličini tvrtke.

S obzirom na podatke koji se nalaze u skupu podataka, smatramo da bi bilo zanimljivo pronaći odgovore na iduća pitanja:

* **Pitanja:**

1. Plaćaju li veće tvrtke više te kako prosječna plaća ovisi o veličini tvrtke?
2. U kojoj zemlji tražiti posao s ciljem maksimizacije plaće ovisno o razini iskustva zaposlenika?
3. U kojoj su zemlji plaće iz područja podatkovne znanosti najviše?
4. U kojoj se zemlji većinski radi puno radno vrijeme?

Također smatramo da je zanimljivo promatrati:

1. Prosječne plaće ovisno o iskustvu, lokaciji tvrtke, tipu zaposlenja, …
2. Kretanje prosječne plaće i potražnje posla po godinama
3. Top 20 zemalja po prosječnoj plaći i potražnji posla
4. Distribucije određenih varijabli
5. Veza između varijabli ( matrica korelacije te hipoteze )
6. Grupiranje danih podataka
7. Predviđanje plaća pomoću odabranih varijabli

* **Tehnike/algoritmi:**

1. EDA
   1. vizualizacije (lineplot, histplot, boxplot, countplot, pie...)
   2. deskriptivna i inferencijalna statistika
   3. matrica korelacije
2. Regresija
   1. linearna
   2. polinomijalna
3. Grupiranje
   1. K-means algoritam
   2. DBSCAN algoritam