

Analiza studentskog dataset-a

Cilj: povezati bazu podataka, obraditi dataset, očistiti ga i vizualizovati rezultate.

Rad sa bazom (MySQL Workbench, Python)

1. Instalirajte **MySQL Workbench**.
2. Kreirajte bazu i tabelu koristeći sledeći SQL kod:

SQL

```
CREATE TABLE studenti (  
    ID INT,  
    Ime VARCHAR(100),  
    Godine VARCHAR(10),  
    Grad VARCHAR(100),  
    Prosek VARCHAR(10),  
    Datum_upisa VARCHAR(20),  
    Datum_diplomiranja VARCHAR(20),  
    ESPB VARCHAR(10),  
    Email VARCHAR(150),  
    Telefon VARCHAR(50)  
);
```

3. Učitajte dataset `studenti_dataset_700.csv` u tabelu.
4. Povežite se iz Pythona na MySQL bazu (npr. preko `mysql-connector-python` ili `SQLAlchemy`).

Osnovna analiza podataka (Pandas)

1. Učitajte dataset i prikažite prvih 10 redova.
2. Prikažite `.info()` i `.describe()`.
3. Izračunajte koliko dataset ima redova i kolona.
4. Identifikujte kolone koje sadrže **NaN** vrijednosti.

Čišćenje podataka

1. Dopunite nedostajuće vrijednosti:
 - **Godine** → zamijeniti prosječnom vrijednošću.
 - **Prosek** → zamijeniti median vrijednošću.
 - **Grad** → zamijeniti stringom "nepoznat".
2. Standardizujte nazive gradova (npr. "pg", "Podgorica " → "Podgorica").
3. Izbacite ili zamijenite nelogične vrijednosti u svim kolonama.
4. Očistite mejlove: obrišite redove gdje email ne sadrži "@" i ".".
5. Standardizujte brojeve telefona u obliku: **+381XXXXXXXXXX**.
6. Provjerite duplikate po kolonama i uklonite ih tako da ostane samo prvi unos.
7. Sredite datume (upis/diplomiranje) i pretvorite ih u jedan format (npr. **YYYY-MM-DD**).
8. Napravite novu kolonu **Trajanje_studija = Datum_diplomiranja - Datum_upisa** (u godinama).
9. Pronađite studente sa najdužim i najkraćim trajanjem studija.

Vizualizacija (Python Matplotlib/Seaborn ili Looker Studio)

1. Prikažite raspodjelu **godina studenata** na histogramu.
2. Prikažite raspodjelu prosječnog prosjeka na boxplot-u.
3. Napravite bar chart: prosječan prosjek po gradu.
4. Prikažite raspodjelu ESPB po gradovima.
5. Napravite scatter plot: Godine vs. Prosjek.
6. Napravite line chart: prosječan prosjek po godini upisa.

Napredna analiza

1. Napravite pivot tabelu: prosječan prosjek po gradu i godini upisa.
2. Pronađite top 5 studenata sa najvišim prosjekom.
3. Izračunajte prosječan broj ESPB bodova po gradu.

Looker Studio (Google Data Studio)

1. Uvezite očišćeni CSV u Looker Studio.
2. Napravite dashboard sa nekoliko vizualizacija (bar chart, line chart, scatter plot).
3. Pripremite kratak zaključak: šta se može naučiti iz podataka.