

## Report zo zimného semestra

Počas tohto semestra som sa oboznámil s technológiami ukladania dát do súborov vo formátoch CSV, Parquet a JSON a s ich importovaním/ukladaním z/do databáz. Taktiež, keďže som nevedel Python, tak som sa ho naštudoval. Počas štúdia som sa naučil a v praxi použil, ako pracovať s dátami v Pythone pomocou knižnice Pandas a s databázami. Ako databázový systém som zvolil PostgreSQL.

Bol vytvorený objektový a dátový návrh, ako aj návrh používateľského rozhrania (GUI). V konečnom dôsledku som napísal základnú štruktúru kódu, ktorá vykonáva nasledujúce:

- Monitoruje konkrétny súbor.
- Spracováva nové súbory, ktoré nahráva používateľ.
- Počíta metriky pre tieto súbory.
- Ukladá výsledky do databázy.
- Umožňuje načítať uložené hodnoty z databázy.

Metriky:

- Všeobecné
  - Počet všetkých záznamov
  - Počet prázdnych záznamov
    - CSV, Parquet – počet prázdnych riadkov
    - JSON – počet NULL objektov a prázdnych objektov
  - Počet duplicitných záznamov
- Parametrizované stĺpcom/JSON cestou
  - Počet NULL/blank hodnôt v danom stĺpci / na danej ceste
  - JSON - Počet záznamov, ktoré majú definovanú danú cestu
  - Počet unikátnych hodnôt v danom stĺpci / na danej ceste
  - Priemer pre stĺpce / cesty s číselnými hodnotami

Všetky všeobecné metriky boli implementované.

Okrem toho bola vytvorená testovacia trieda pre jednoduchšie testovanie.