

Stručná dokumentace zprovoznění

Ukládanie a príprava dát

Názov tímu: xryban00

Členovia tímu: Dávid Pukanec, Adam Rybanský, Michal Sandanus

Github repozitár: <https://github.com/adaryb123/UPAproject>

Zvolená téma: Zdravotníctvo

Pre spustenie programu je potrebné urobiť nasledovné:

- Stiahnuť si Github repozitár
- Mať nainštalovaný Python, ideálne verziu 3.8 alebo vyššiu
- Importovať knižnice uvedené v requirements.txt (pandas, numpy, jupyter, mongoengine, certifi, seaborn, matplotlib, pathlib, scipy)
- spustiť súbor main.py – POZOR- tento krok už netreba vykonávať, dáta sa už v databáze nachádzajú
- otvoriť jupyter notebook **Part2_ipynb**. Všetky úlohy sú riešené v ňom

Notebook je rozdelený na 2 časti: **vytvorenie csv súborov** a **vizualizácia**. Časť **vytvorenie csv súborov** vyberá potrebné dáta z databázy a uloží ich do csv súborov, ktoré vytvorí. CSV súbory sú pre každú úlohu rozdielne, preto sme ich pomenovali podľa názvov úloh („A1.csv“, „A2.csv“ a tak ďalej).

V časti **vizualizácia** sa tieto csv súbory načítavajú do dataframov, a ďalej sa upravujú, aby mali správny tvar pre potreby danej úlohy. Potom sa pre každú úlohu vykreslí graf.

V časti **vizualizácia** je riešená aj úloha C, kde sa najprv načíta csv súbor, upraví sa do potrebného tvaru a uloží pod názvom „**C1_OUTPUT.csv**“.

Potom sa csv súbor načíta znova, vykoná sa nahradenie odľahlých hodnôt, normalizácia a diskretizácia, a výsledok sa uloží do súboru „**C1_OUTPUT_2.csv**“.

Výsledky a vykreslené grafy sú v súbore **Vysledky.pdf**

Dokumentácia k 1. Odovzdávaniu projektu je v súbore Dokumentacia1odovzdanie.pdf