

Zvolené téma: 4 - COVID-19

Řešitelé: Kateřina Fořtová (xforto00), Jakub Kolb (xkolbj00), Ondřej Pavela (xpavel34)

Zvolené dotazy a formulace vlastního dotazu:

- **Dotaz skupiny A** - vytvořte popisné charakteristiky pro alespoň 4 údaje (např. věk, pohlaví, okres, zdroj nákazy) z datové sady COVID-19: Přehled osob s prokázanou nákazou dle hlášení krajských hygienických stanic (využijte krabicové grafy, histogramy, atd.)
- **Dotaz skupiny B** - určete vliv epidemie COVID-19 na počet zemřelých v porovnání dle počtu nemocných, počtu hlášených úmrtí na nemoc COVID-19 a v porovnání s minulými lety
- **Vlastní dotaz** -

Stručná charakteristika zvolené datové sady:

(Zde konkrétně popište jaké soubory budou představovat zdroj dat pro zvolené úlohy. Dále popište, jakým způsobem budou tato data získána a stručně charakterizujte strukturu souborů vybraných pro řešení projektu. Zaměřte se na části souborů, které jsou důležité pro zodpovězení zvolených dotazů.)

Pro dotaz skupiny A bude využito datové sady, kterou poskytuje Ministerstvo zdravotnictví ČR (<https://onemocneni-aktualne.mzcr.cz/api/v2/covid-19/osoby.json>). Tato datová sada nám poskytuje všechny potřebné údaje o věku, pohlaví a okrese nakažených. Také je zde uvedeno, zda se daná osoba nakazila v zahraničí.

Pro dotaz skupiny B je nutno využít více datových sad. První z nich je opět sada poskytovaná MZ ČR (<https://onemocneni-aktualne.mzcr.cz/api/v2/covid-19/umrti.json>). Jedná se o informace o datu, věku, pohlaví, kraji a okrese zemřelých. Dále je nutné získat datovou sadu z jiného zdroje, která nám bude podávat informace o vývoji mortality v předešlých letech v ČR.

Při práci s daty jsme upřednostňovali data ve formátu JSON díky snadnému stahování a získávání potřebných údajů pomocí skriptu napsaném v jazyce Python. Níže se nachází ukázka formy ukládání dat v datové sadě, která udává informace o aktuálních zemřelých na COVID-19:

```
"modified": "2020-09-23T17:50:04+02:00",
"source": "https://onemocneni-aktualne.mzcr.cz/v",
"data": [
  {
    "datum": "2020-03-22",
    "vek": 94,
    "pohlavi": "M",
    "kraj_nuts_kod": "CZ010",
    "okres_lau_kod": "CZ0100"
  }
]
```

Zvolený způsob uložení uložení surových dat:

(Zde stručně charakterizujte NoSQL databázi, která bude využita pro uložení zvolených zdrojových dat.)