Projekt World Population

## Projektmål

Lær at forstå, udvide/forbedre og visualisere en database. Lær at finde data og integrere dem fra forskellige kilder i forskellige formater i din egen database.

### Forberedelser

Download denne database: <https://www.kaggle.com/datasets/iamsouravbanerjee/world-population-dataset?select=world_population.csv>

Alternativt finder du filen i samme mappe som denne opgavebeskrivelse. (ML1110\_World\_Population.csv)

Find ud af, hvilke oplysninger databasen indeholder, og hvordan databasen er struktureret.

### Udvid/forbedr databasen

Berig databasen med yderligere befolkningstal fra det 20. århundrede ved at tilføje kolonner for befolkningstallet for årene 1900, 1910, ..., 1960. Indsaml data fra alle mulige forskellige steder, du kan finde. Dette kan f.eks. være

* CSV-filer eller SQL-databaser
* du modtager data i JSON-struktur via en web-API
* du henter data direkte fra hjemmesider ved hjælp af "web scraping".
* du henter data via en forespørgsel fra Wikidata.

Gem disse data først i din egen separate database (f.eks. CSV eller SQL). Kolonnerne i denne database kan være:

* "CCA3" (<https://en.wikipedia.org/wiki/ISO_3166-1_alpha-3>)
* "Year"
* "Population"
* "Kilde" (hvor du har fået tallet fra, f.eks. worldometers.info eller wikidata)

CCA3 bruges også i den anden database og vil hjælpe dig med at flette data fra begge databaser.

### Konsolider data

Hvad gør du, hvis

* du ikke kan finde data for en celle i databasen?
* data fra forskellige kilder modsiger hinanden?
* størrelsen af et land har ændret sig markant i den betragtede periode?
* du ikke kan finde data for en celle i databasen?

Beregn for hvert land ud fra dine indsamlede data et skøn for befolkningen i årene 1900, 1910, ..., 1960, og gem disse data som yderligere kolonner i den database, hvor du allerede har data fra 1970 og frem.

### Visualiser væsentligt indhold af databasen

Måske finder du inspiration i de notebooks, som andre brugere har gemt offentligt til denne database: <https://www.kaggle.com/datasets/iamsouravbanerjee/world-population-dataset/code> Her kan du f.eks. se, hvordan man med få linjer kode kan farvelægge landene på et verdenskort afhængigt af størrelsen af en nøgletal: <https://www.kaggle.com/code/wumanandpat/world-population-trends> Alternativt kunne Python-bibliotekerne Matplotlib, Seaborn og Folium (sidstnævnte især til landkort) være nyttige til visualisering.