

# Labo 1 Data Foundations:

## Doel van de opdracht

In deze extra opdracht zullen we de basisanalyse uitbreiden naar meer geavanceerde data-analysetechnieken. Naast het verkennen van één variabele, zullen we nu meerdere variabelen onderzoeken en relaties daartussen analyseren. Je leert hoe je correlaties kunt ontdekken en nieuwe inzichten kunt verkrijgen door middel van data-manipulatie en visualisatie in Python.

## Stappen

### Stap 1: Data verzamelen en importeren

Net als bij de originele opdracht zullen we werken met de bevolkingsdataset van Australische staten (austpop) uit de AER package. Download en importeer de dataset in Python.

Extra uitdaging: je mag werken met dezelfde dataset, of je gaat op zoek naar een andere dataset uit dezelfde datalibrary. Importeer deze ook in Python.

Tip: gebruik Pandas om eventuele ontbrekende waarden te verwerken.

[Link naar de data-library](#)

Van deze datalibrary kan je het .csv-bestand downloaden.

Tip: Zorg ervoor dat je het .csv-bestand hebt gedownload en dat het in dezelfde map staat als je Python-script. Vervolgens kan je de dataset importeren met Pandas.

### Stap 2: Data verkennen

Voer de volgende analyses uit op een variabele naar keuze in de dataset:

a. Bivariate analyse:

Bereken de relatie tussen twee variabelen uit de verschillende datasets. Bijvoorbeeld, kijk naar de relatie tussen bevolkingsgroei en economische indicatoren (zoals BBP of werkloosheid). Visualiseer deze relatie met een scatter plot.

b. Tijdreeksanalyse:

Selecteer een variabele die over de tijd heen varieert, zoals bevolkingsaantal of BBP. Visualiseer deze variabele met een lijndiagram om trends door de tijd te ontdekken. Voeg meerdere lijnen toe als je meerdere variabelen in de tijd wilt vergelijken (bijvoorbeeld bevolkingsgroei versus economische groei).

c. Boxplots:

Genereer boxplots om de verdeling van variabelen tussen verschillende staten te vergelijken. Analyseer of bepaalde staten afwijkende trends vertonen in de dataset.

### Stap 3: Gegevens vrijken (Extra uitdaging)

Gebruik feature engineering om nieuwe variabelen te creëren op basis van bestaande data. Bijvoorbeeld, je kunt de relatieve bevolkingsgroei berekenen als percentage per jaar of het gemiddelde BBP per hoofd van de bevolking per staat.

**\*\*Stap 4: Interpretatie en conclusies** Analyseer de resultaten van de bivariate en tijdreeksanalyses. Geef een uitgebreide interpretatie van de relaties tussen de variabelen en de trends die je hebt ontdekt. Probeer verbanden te leggen tussen bevolkingsgroei en economische prestaties van staten.

- Zijn er duidelijke correlaties?
- Hoe zijn de trends in bevolkingsgroei door de jaren heen?
- Hebben economische indicatoren invloed op de bevolkingsgroei?

**Indienen:**

Net als bij de oorspronkelijke opdracht: verzamel je Python-code en CSV-bestanden in een .zip-bestand. Zorg ervoor dat je script meteen uitvoerbaar is en dat eventuele externe datasets correct worden ingeladen.

Tip: vergeet je portfolio niet!