



## Thuisproject 2024 - Voorbereiding

### Team-samenstelling (max 2 personen)

Vul hier uw gegevens in

Student 1	Lara Mestdagh
-----------	---------------

### Dataset

Ga naar <https://www.kaggle.com/> en zoek twee datasets waarrond je jouw project wenst op te bouwen. De docent geeft later door op welke dataset er verder kan gewerkt worden.

**Opgelet:** zoek géén 'kant-en-klare' dataset (zie projectopgave voor meer info)

	Naam	URL
Dataset 1	Animal Crossing New Horizons Catalog	<a href="https://www.kaggle.com/datasets/jessicali9530/animal-crossing-new-horizons-nookplaza-dataset">https://www.kaggle.com/datasets/jessicali9530/animal-crossing-new-horizons-nookplaza-dataset</a> (meerdere datasets, ik zou villagers.csv willen gebruiken)
Dataset 2	Video Game Sales	<a href="https://www.kaggle.com/datasets/gregorut/videogamesales">https://www.kaggle.com/datasets/gregorut/videogamesales</a>

### Dataset preprocessing

Geef per dataset minimum **vier** preprocessing stappen die je zou doorlopen. Geef ook op zijn minst **vier zinnvolle plotjes** die je zou willen genereren en wat je er hoopt uit te leren. Let op, hou er rekening mee dat sommige waarden categorisch zijn (en bepaalde plotjes dan niet zinvol zijn).

#### Dataset 1

	Beschrijving
Preprocessing stap 1	Ontbrekende waarden: Controleer op ontbrekende waarden in de dataset en beslis hoe je hiermee om wilt gaan (bv. verwijderen, invullen).
Preprocessing stap 2	Datatypes: Verzeker dat de datatypes correct zijn ingesteld voor elke kolom (bijvoorbeeld, jaar als integer, naam als string).
Preprocessing stap 3	Categorische gegevens: Converteer categorische gegevens naar een formaat dat geschikt is voor analyse (bv. one-hot encoding voor persoonlijkheden en hobby's).
Preprocessing stap 4	Opschonen van tekst: Normaliseer en schoon tekstvelden op, zoals 'Name' en 'Catchphrase', om consistentie te waarborgen.
Plot 1	Verdeling van Personages per Soort: Een staafdiagram om de verdeling van personages over verschillende soorten te tonen.
Plot 2	Verdeling van Personages per Persoonlijkheid: Een taartdiagram om de verdeling van personages over verschillende persoonlijkheden te tonen.
Plot 3	Verdeling van Personages per Hobby: Een staafdiagram om te laten zien hoe hobby's zijn verdeeld onder de personages.
Plot 4	Verjaardagen per Maand: Een histogram om te zien in welke maanden de meeste personageverjaardagen vallen.

**Dataset 2**

	Beschrijving
Preprocessing stap 1	Ontbrekende waarden: Identificeer en behandel ontbrekende waarden in het 'Year' en 'Publisher' veld.
Preprocessing stap 2	Data consistentie: Zorg ervoor dat de namen van games en uitgevers consistent zijn (bijvoorbeeld geen typfouten of verschillende namen voor hetzelfde spel of uitgever).
Preprocessing stap 3	Uitschieters: Identificeer en onderzoek uitschieters in de verkoopgegevens.
Preprocessing stap 4	Categorisatie: Mogelijk categoriseren van de 'Year' veld in tijdsperiodes (bijv. per decennium) voor eenvoudigere analyse.
Plot 1	Top 10 Verkochte Games: Een staafdiagram met de top 10 van meest verkochte games wereldwijd.
Plot 2	Verkoop per Regio: Een reeks taartdiagrammen die de verkoopverdeling per regio tonen.
Plot 3	Verkoop per Genre: Een staafdiagram met de totale verkoop per genre.
Plot 4	Trends over Tijd: Een lijngrafiek die de totale gameverkoop per jaar toont.

**Vier zinvolle zoekopdrachten**

Geef per dataset **vier** zinvolle zoekopdrachten door. Zie projectopgave voor meer details

- Bij minimum drie opdrachten moeten er vooraf één of meerdere parameters opgevraagd worden om zo de zoekopdracht te verfijnen.
- Bij minimum één zoekopdracht moet er een grafiek kunnen gegenereerd worden.

**Dataset 1**

	Beschrijving zoekopdracht
Zoekopdracht 1	Vind de meest voorkomende persoonlijkheid onder de Animal Crossing-personages. Parameter: Geen. Verwachte Uitvoer: Naam van de meest voorkomende persoonlijkheid.
Zoekopdracht 2	Zoek alle personages die een specifieke hobby hebben. Parameter: Hobby (bijv. Fitness, Natuur). Verwachte Uitvoer: Lijst van personages met de opgegeven hobby.
Zoekopdracht 3	Bepaal de verdeling van mannelijke versus vrouwelijke personages. Parameter: Geen. Verwachte Uitvoer: Aantal mannelijke en vrouwelijke personages.
Zoekopdracht 4	Genereer een grafiek die de verdeling van personages over de verschillende soorten toont. Parameter: Geen. Verwachte Uitvoer: Staafdiagram van het aantal personages per soort.

**Dataset 2**

	Beschrijving zoekopdracht
Zoekopdracht 1	Zoek de game met de hoogste wereldwijde verkoop. Parameter: Geen. Verwachte Uitvoer: Naam van de game met de hoogste wereldwijde verkoop.
Zoekopdracht 2	Vind alle games uitgegeven door een specifieke uitgever. Parameter: Uitgever (bijv. Nintendo). Verwachte Uitvoer: Lijst van games uitgegeven door de opgegeven uitgever.
Zoekopdracht 3	Bepaal het totale verkoopvolume per regio voor een specifiek jaar. Parameter: Jaar (bijv. 2020). Verwachte Uitvoer: Totale verkoop per regio voor het opgegeven jaar.
Zoekopdracht 4	Genereer een grafiek die de trend van de wereldwijde gameverkoop over de jaren heen toont. Parameter: Geen. Verwachte Uitvoer: Lijngrafiek met de wereldwijde verkoop per jaar.