

# Toelichting op voorbeeld-bestanden

---

## 1. Ziekenhuizen

---

Dit data-bestand is afkomstig van het Centraal Bureau voor de Statistiek (CBS) en bevat voor alle 431 Nederlandse gemeenten (in 2010) de gemiddelde afstand van de inwoners (in km) tot het dichtstbijzijnde ziekenhuis en tot de dichtstbijzijnde huisarts. Het bestand bevat ook voor elke gemeente de naam ervan, de omvang van de bevolking, de bevolkingsdichtheid en het landsdeel.

Bestand: Ziekenhuizen.xlsx; ziekenhuizen.json  
Aantal variabelen: 6  
Aantal records: 431  
Variabelen: Gemeente, Jaar, Bevolking, Dichtheid, Huisarts, Ziekenhuis, Landsdeel

## 2. Exitpoll in Leiden

---

Dit data-bestand bevat de gegevens van een peiling op de dag van de gemeenteraadsverkiezingen in Leiden op 21 maart 2018. Voor de peiling werd bij de uitgang van stembureaus de mening gevraagd van personen die net hadden gestemd. Ook de stem bij de vorige verkiezingen werd gevraagd. Een dergelijke peiling noemen we ook wel een exitpoll.

Bestand: Exitpoll.xlsx; exitpoll.json  
Aantal variabelen: 8  
Aantal records: 5130  
Variabelen: Stem2018, Stem2014, Lokaal, Referendum, Geslacht, Leeftijd, Opleiding, Stembureau.

## 3. Strandtent “Strandzicht”

---

Dit data-bestand gaat over de omzet van een strandtent met de naam ‘Strandzicht’. Dit bestand bevat fictieve gegevens over de omzet en de weersomstandigheden. Deze gegevens zijn een maandlang bijgehouden (31 dagen),

Bestand: Strandzicht.xlsx; standzicht.json  
Aantal variabelen: 5  
Aantal records: 31  
Variabelen: Omzet, Temp (temperatuur), Vtemp (voorspelde temperatuur), Zon (aantal uren zon), Vrij (vrije dag of werkdag).

## 4. Samplonië

---

Dit data-bestand gaat over het denkbeeldige land Samplonië. Dat is een eilandje met 1000 inwoners. Het heeft slechts twee provincies: Agrië, waar met name landbouw en veeteelt wordt bedreven, en Indusië, waar vooral industriële bedrijvigheid is en de grotere steden te vinden zijn.

Bestand: Samplon.xlsx; samplon.json  
Aantal variabelen: 6  
Aantal records: 1000  
Variabelen: Gemeente, Provincie, Geslacht, Leeftijd, Werk, Inkomen.

## 5. Hotdogs

---

Amerikanen houden van het eten van hotdogs. Amerikanen die zich bezighouden met hun eigen gezondheid, geven de voorkeur aan hotdogs die weinig calorieën bevatten. Daarom is een onderzoek uitgevoerd naar de samenstelling van 54 belangrijke merken hotdogs. Elke hotdog werd ingedeeld in een van drie typen: rundvlees, varkensvlees en kip. Verder werd van elk van 54 hotdogs vastgesteld hoeveel categorieën deze bevatte.

Bestand: Hotdogs.xlsx; hotdogs.json  
Aantal variabelen: 2  
Aantal records: 54  
Variabelen: Calorie (aantal calorieën), Vlees (rund, varken, kip).

## 6. Pinguïns

---

Bestand: Pinguins.xlsx; pinguins.json  
Aantal variabelen: 7  
Aantal records: 344  
Variabelen: species, island, bill\_length, bill\_depth, flipper\_length, body\_mass, sex

In het continent Antarctica rond de zuidpool van de Aarde leven naar schatting 5,9 miljoen pinguïns. Australische wetenschappers o.a. Kirsten Gorman hebben onderzoek gedaan naar de kenmerken van deze pinguïns. Daarvoor hebben ze van ruim 300 pinguïns de soortnaam genoteerd, het eiland waarop ze voorkomen, de snavelengte en -breedte, de flipperlengte, het lichaamsgewicht en het geslacht.