

# Inleiding

Toen de VOC in de 16e eeuw begon kruiden en specerijen vanuit Indonesië naar Nederland te vervoeren stond het niet alleen aan het begin van de huidige Nederlandse welvaart, maar luidde het ook een nieuw tijdperk in voor de rest van de wereld: steeds meer landen volgden het Nederlandse voorbeeld en begonnen handel met elkaar te drijven. Zo ontstond langzaam maar zeker de huidige wereldeconomie waar de prijzen van producten worden bepaald door wereldwijde vraag en aanbod [Bron].

De toename in handel tussen landen zorgde in het algemeen voor een daling in prijzen en een grotere verscheidenheid aan beschikbare producten. Maar het zorgde ook voor een toenemende mate van afhankelijkheid tussen landen, zoals duidelijk werd tijdens de wereldwijde voedselcrises in 2007-2008: in twee jaar tijd maakten veel breed-geconsumeerde producten een onvoorspelbare hoge stijging door. Zo steeg de gemiddelde prijs van rijst met 217% en de prijs van graan met 136% [Bron]. Het gevolg was een globale crisis waarbij zowel ontwikkelde - als ontwikkelingslanden door economische en sociale onrust werden getroffen [bron].

Daarom is er veel onderzoek gedaan naar de oorzaken van stijgingen in voedselprijzen, met als doel een herhaling van de voedselcrisis te voorkomen. Zo bleek onder andere de toename in het gebruik van biobrandstoffen de voedselprijzen te verhogen: gewassen die worden gebruikt voor de productie van deze brandstoffen kunnen niet worden gebruikt voor het produceren van voedsel [bron]. Ook blijkt dat de olieprijs positief gecorreleerd is aan voedselprijzen, omdat het zowel de energiebron is voor mechanische voedselproductie als voor het transport ervan, maar ook omdat een stijging van de olieprijs het gebruik van biobrandstof weer aantrekkelijker maakt [bron].

Hier wordt met behulp van verschillende datasets getracht deze oorzaken te bevestigen, wat leidt tot de volgende deelvragen:

- Bestaat er een verband tussen de olieprijs en voedselprijzen?
- bestaat er een verband tussen bio-brandstofprijzen en voedselprijzen?

Verwacht wordt dat er een positieve correlatie bestaat tussen de olieprijs en alle voedselproducten, maar dat de correlatie sterker is naarmate het betreffende product veel mechanische verwerking vereist (bijv). Dit is aannemelijk omdat...

Er wordt hier echter verondersteld dat nog meer factoren van invloed zijn op de prijsstijgingen, zoals het klimaat en de complexiteit van het voedselproduct. Daarom wordt ook geprobeerd de volgende deelvragen te beantwoorden:

- Vertonen complexe producten dezelfde prijsverandering als de ingrediënten waaruit het bestaat?
- Als dat zo is, is die correlatie dan altijd aanwezig, of alleen tijdens bepaalde periodes?
- Vertonen landen in dezelfde regio's vergelijkbare prijsverschillen?
- bestaat er een verband tussen regenval en voedselprijzen?

Er wordt een correlatie verwacht tussen producten die dezelfde ingrediënten bevatten of een ingrediënt zijn van een ander product , zoals ...

Ook wordt verwacht dat landen in vergelijkbare klimaatstreken soortgelijke fluctuaties in de voedselprijs laten zien en dat een lange periode van droogte leidt tot het stijgen van de prijzen van bepaalde gewassen zoals...

Bovendien wordt verondersteld dat voedsel prijsstijgingen niet alleen sociale onrust veroorzaken, maar dat het ook leidt tot meer ziektes en vluchtelingen. Daarom wordt tot slot geprobeerd ook de volgende deelvragen te beantwoorden:

- bestaat er een verband tussen de voedselprijzen en het sterftecijfer in een land?
- bestaat er een verband tussen vluchtelingen-stromingen en voedselprijzen?

Er wordt verwacht dat er een positief verband bestaat tussen de vluchtelingenstromen, sterftecijfers voedselprijzen.

Om deze deelvragen te kunnen beantwoorden is gebruikt gemaakt van de *Global Food Prices Database (GFPD)* van het *World Food Programme* (<http://www1.wfp.org/>). In de database zijn voor 76 verschillende landen de maandelijkse voedselprijzen gegeven voor veel geconsumeerde voedsel waren zoals bonen, rijst en olie. Deze database is uitgebreid genormaliseerd zodat zowel de voedselprijzen van verschillende producten met elkaar vergeleken konden worden als de prijsveranderingen *tussen* landen en gebieden.

\*meer

# Methode

## pre-processing

Ten eerste zijn het is het aantal dimensies van de dataset terugbracht: de kolom met de provincies is verwijderd omdat bij het merendeel van de provincies enkel data was verzameld van één stad, waardoor deze kolom geen extra informatie toevoegde dan de kolom met steden zelf. Ook waren in de oorspronkelijke dataset de jaren en de maanden gegeven in aparte kolommen; deze zijn samengevoegd tot één kolom.

Vervolgens zijn identieke stads-namen die bij verschillende landen hoorden gedisambigueerd door de afkorting van het betreffende land erachter te zetten. Zo werd San Vicente veranderd naar Sante Vicente (Sal). Dit was wenselijk omdat er later bij de visualisatie per stad data opgevraagd zou worden.

Daarna zijn alle eenheden genormaliseerd. Daarvoor zijn alle gewichten naar kilogram (KG) omgezet en alle volumes naar liter (L). Zo was bijvoorbeeld de graanprijs zowel per 1,5 KG als per KG gegeven. Ook dit zou het vergelijken van productprijzen later vergemakkelijken. Bij eenheden waarvoor geen vaste conversion-rate gevonden kon worden zijn de betreffende rijen verwijderd, zoals bijvoorbeeld 'quartillas' en (goat) 'heads'.

Ook zijn alle wisselkoersen en de bijbehorende prijzen genormaliseerd naar US Dollars (USD). Hiervoor is gebruik gemaakt van de .... dataset. Rijen met munteenheden waarvoor geen officiële wisselkoers kon worden gevonden zijn verwijderd. Zo bleek de Somaliland Shilling (SOS) geen officieel erkende munteenheid te zijn en was er voor de Armenian Dram (AMD) pas wisselkoers-informatie beschikbaar vanaf februari 2003; alle rijen met de SOS zijn dus verwijderd, en alle rijen met de AMD vóór februari 2003 ook.

Omdat nu alle prijzen in USD waren gegeven was het ook mogelijk prijzen te normaliseren op koopkracht (GDP) met behulp van de ... dataset. Door telkens de GDP van het land door de betreffende prijs van het product te delen ontstond een maat die aangeeft hoeveel er van

dat product in het betreffende land gekocht kan worden; hoe hoger het getal hoe beter dat product dus beschikbaar is voor de inwoners van dat land.

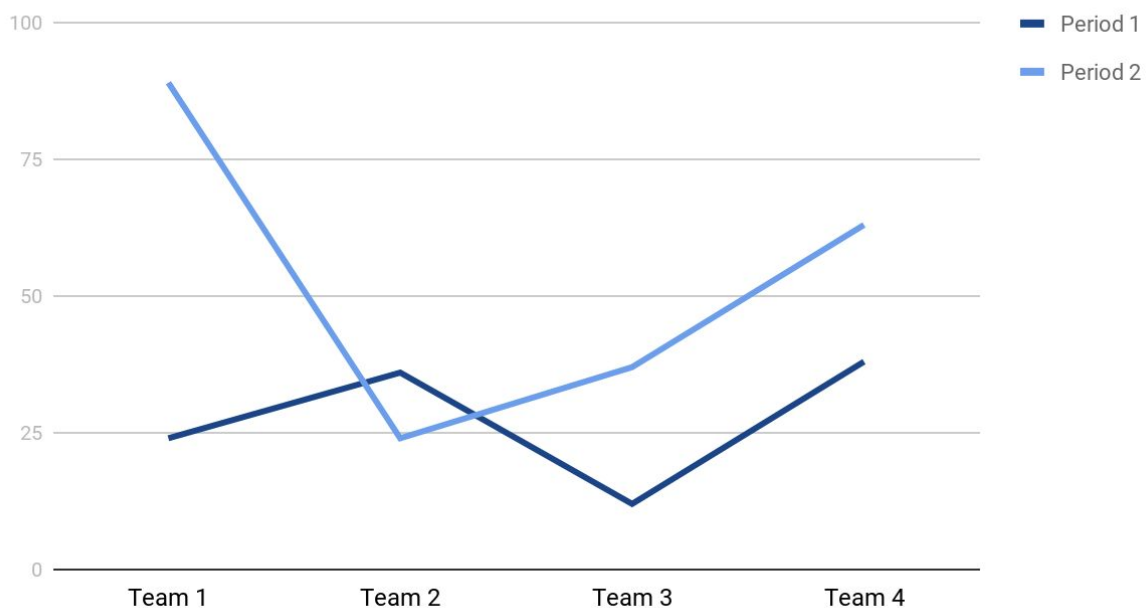
En omdat het uiteindelijke doel het vergelijken van verschillende voedsel-prijs-grafieken was, en er hiervoor onafgebroken grafieken nodig zijn, zijn alle producten waarvan er meer dan twee opeenvolgende maanden aan data misten verwijderd. Gaps van één of twee jaar zijn ingevuld met behulp van lineaire regressie, waarbij alleen de eerste productprijs vóór en de eerste productprijs ná de gap als referentiepunten zijn gebruikt.

Tot slot is voor elke productprijs de afgeleide bepaald zodat de daadwerkelijke veranderingen in voedselprijzen tussen steden en landen vergeleken konden worden, en is met behulp van de ... dataset elk land ingedeeld in sub-regio's zodat ook de voedselprijzen tussen gebieden zoals Midden-Oosten en Indo-China vergeleken konden worden.

### Exploratory Data Analysis?

op het eerste gezicht leek dit en dit en dit... het geval.

#### Points scored



- uitleg en interpretatie

### **K-means**

Vervolgens is er een K-means algoritme geïmplementeerd zodat producten met gelijke prijsveranderingen automatisch bij elkaar gegroepeerd kunnen worden. Dat gebeurt door willekeurig een door de gebruiker gekozen aantal cluster te initiëren en vervolgens elk datapunt toe te wijzen aan het cluster met de kleinste euclidische afstand tot dat punt. En door vervolgens in elk cluster alle datapunten bij elkaar op te tellen en de nieuwe cluster-centra op het resulterende gemiddelde te plaatsen, kan het proces van het toewijzen van datapunten aan clusters en het opnieuw berekenen van cluster centra iteratief worden herhaald tot de totale euclidische afstand van alle datapunten tot hun cluster-centrum niet of nauwelijks meer afneemt.

## **Resultaten**

-

