

”The only thing I like better than talking about
food is eating”

- John Walters

Technisch Rapport Data Analyse en Visualisatie
Kunstmatige Intelligentie
Willemijn Beks

Mirka Schoute, 11868643
Anne Boomsma, 11769556
Aram Elias, 11807911
Tim Müller, 11774606

28 juni 2018

1 Inleiding

In tegenstelling tot het beeld dat algemeen heerst, wordt de hoeveelheid mensen die leeft in armoede steeds kleiner. Een steeds groter gedeelte van de wereldbevolking leeft in een middenklasse. Laten we voor de duidelijkheid het welzijn van de wereldbevolking opdelen in vier niveaus: de mensen met extreme armoede leven op niveau 1. Mensen op niveau 1 leven van minder dan 1,90[8] dollar per dag, hebben geen schoenen en besteden het grootste gedeelte van de dag aan het vergaren van water en brandhout. Op niveau 2 leef je op ongeveer 3.10 [8] dollar per dag: je hebt je eigen kippen en een fiets, en kinderen kunnen huiswerk maken bij kunstlicht. Het leven op niveau 3 is al wat welvarender: je verdient zo'n 16 dollar per dag, waar stabiele elektriciteit, stromend water en bijvoorbeeld een motor om naar je werk te gaan bij horen. En dan niveau 4, het hoogste niveau, waar wij in de westerse wereld in leven.

Het heersende beeld is dat het grootste gedeelte van de bevolking op niveau 1 leeft. Immers, dat is het beeld dat wordt geschetst door ontwikkelingsorganisaties in de hoop jouw medelijden te wekken en je over te halen een donatie te doen. Echter, waar tweehonderd jaar geleden 84% [8, 3], is dat tegenwoordig nog maar ongeveer 9%[8, 1]. Driekwart van de wereldbevolking leeft op niveau 2 en 3, vergelijkbaar met de welvaart in West-Europa in de jaren '50. We zijn er met zijn allen dus flink op vooruit gegaan. Toch?

Enkele tijd terug kwam de WHO, de Wereldgezondheidsorganisatie, met het bericht in 2030 een obesitas-epidemie te verwachten[9]. Johan Cruijff zei het al eens: ieder voordeel heb z'n nadeel. Zo ook het terugdringen van armoede. Want dankzij die verbeterde welvaart, doen de mens zich ook steeds vaker te goed aan zaken die niet per sé goed voor hen zijn. Is hier een verband in te ontdekken? Kan er een verband ontdekt worden tussen voedselprijzen en het gemiddelde BMI van de bevolking in een land? En heeft de geografische ligging van een land dan ook invloed op de ontwikkeling van het gemiddelde BMI? Om hier een volledig antwoord op te kunnen geven is ook het antwoord op de vraag 'Zijn er verbanden te zien in voedselprijs ontwikkeling tussen verschillende landen binnen een mondiale regio?' van belang. Deze vraag doet op zijn beurt dan weer de vraag opkomen hoe dat zit met de prijsontwikkeling in verschillende regio's binnen een land. Ten slotte ontstond ook de behoefte verbanden tussen prijzen van producten te onderzoeken, dit heeft geresulteerd in de vraag: 'Als de prijs van graan stijgt, stijgt de prijs van brood dan ook?'

2 Methode

Om een antwoord te kunnen formuleren op deze vragen zijn de voedselprijzen, de BMI indices en de dollar koersen van verschillende landen door de tijd heen nodig. De voedselprijzen: er is beschikking over 738.788 voedselprijzen, afkomstig uit 1531 verschillende steden in 74 landen van januari 1992 tot aan oktober 2017. In deze periode zijn de prijzen van 331 verschillende producten vast-

gesteld, van een schaap tot aan een kilo maïs. De gegeven voedselprijzen zijn gegeven in de eigen valuta. Om prijzen van verschillende landen beter te kunnen vergelijken, zijn alle prijzen omgerekend met de toenmalige dollar koersen, de koers afkomstig uit de maand waarin de voedselprijs is opgenomen. De voedselprijzen zijn vervolgens omgerekend naar dezelfde eenheid. Prijzen in megaton of per 5 kilogram werden omgerekend naar prijs per kilogram. Ten slotte is ook de dataset voor BMI aangepast: kolommen werden hernoemd zodat deze bruikbaar werden voor Pandas.

Om de voedselprijzen uit zoveel mogelijk verschillende landen te kunnen vergelijken, is ervoor gekozen om een standaard prijs te bepalen. Er is gekeken welk soort voedsel bij zoveel mogelijk landen voorkomt. Dit bleek rijst te zijn, in verschillende vormen en maten. Denk dan aan verschillende kwaliteiten, verschillende soorten of de voedzaamheid. Dit werd allemaal, om zoveel mogelijk landen te kunnen vergelijken, onder het product rijst gezet. Van de 74 landen hadden 66 landen prijsinformatie over deze groep. Dit was meer dan het aantal landen met prijsinformatie over tarwe (43) of suiker (31). De landen zonder rijst zijn Azerbaidjan, Georgia, Kenya, Lesotho, Uganda, Zimbabwe, Sudan en South Sudan. De keuze voor rijst is verantwoord vanwege de grote hoeveelheid data over dit product en de relatieve onafhankelijkheid van economische factoren, in tegenstelling tot bijvoorbeeld olie.

De voedselprijzen bleken grote *outliers* te hebben in Syrië en Myanmar. In Myanmar is in april 2012 besloten de waarde van de Kyat ten opzichte van de Dollar te veranderen om een koers te hebben die zich dichterbij bevindt bij de waarde van de Kyat op de zwarte markt. De voedselprijzen in Kyat in de database veranderden echter niet, waardoor een kilo rijst in één maand tijd veranderde van 0,36 naar 43,02.

In Syrië zijn door de burgeroorlog die daar sinds 2011 woedt de voedselprijzen zo van slag, dat een kilo rijst daar gemiddeld 20 tot 60 dollar kost, met een piek van 200 dollar voor een kilo in 2017. Deze data is zodanig afwijkend van de andere data, en beïnvloeden de andere data ook als het gaat om gemiddelden, dat is gekozen om de rijstprijzen uit Myanmar en Syrië te verwijderen uit de dataset. (Tabel 1 & figuur 1)

Table 1: Rijstprijzen met en zonder Syrië en Myanmar

	Gemiddelde rijstprijs per kilogram	Gemiddelde afwijking van de gemiddelde rijstprijs
Met Syrië en Myanmar	\$2.38	\$3.09
Zonder Syrië en Myanmar	\$0.81	\$0.30

Het verwerken van de dataset is gebeurd gebruikmakend van *Pandas*, een *library* voor *Python*. Ter beantwoording van de opgestelde vragen is gekozen om grafieken te genereren. Op basis van de benoemde info in de vraag is de juiste data geselecteerd, waar met behulp van *Bokeh*, eveneens een *Python library*, grafieken uit zijn gemaakt. Deze grafieken visualiseerden verbanden en

verschillen tussen verschillende prijzen. Vervolgens is via internet gezocht in oude nieuwsberichten naar verklaringen voor gevonden resultaten.

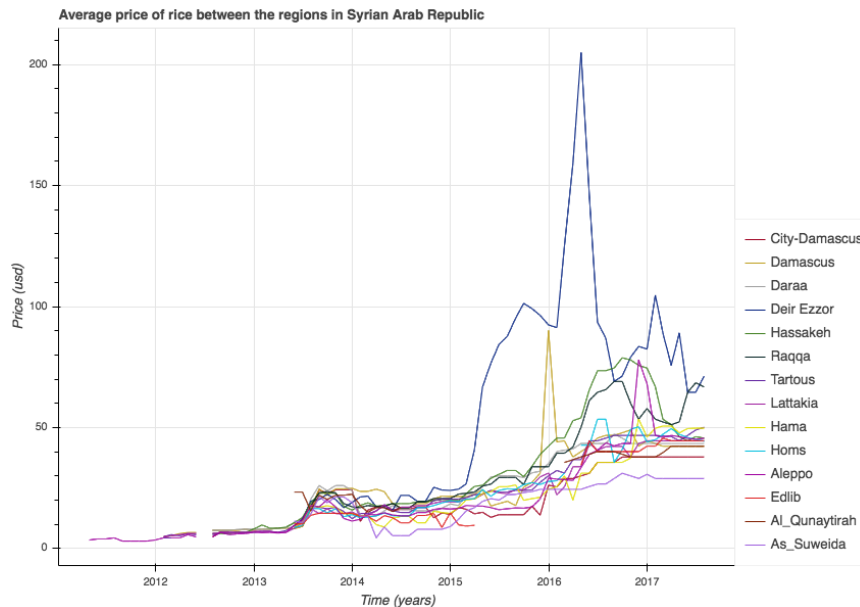


Figure 1: rijstprijs in Syrie per regio

3 Resultaten

Zijn er voedselprijzen die een negatieve of een positieve correlatie laten zien? Is deze correlatie consistent, of komt deze alleen in bepaalde periodes voor? Kun je misschien ingrediënten voor een ander voedselproduct identificeren?

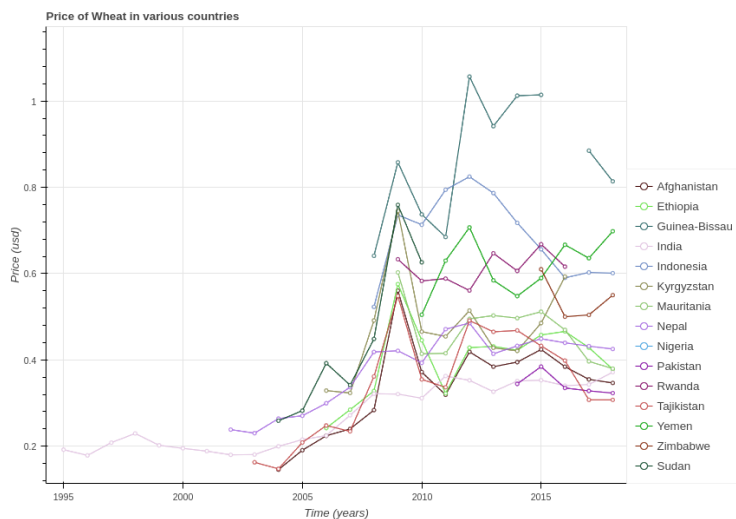
In negen landen is een duidelijke samenhang gezien tussen de prijzen van verschillende soorten voedsel. Zoals in figuur 3a te zien is, stijgen in Benin van 2006 tot 2008 alle voedselprijzen. Vanaf 2009 is samenhang minder zichtbaar, waar de prijs van rijst en bonen beide stijgt, daalt de prijs van maïs en sorghum. De samenhang in El Salvador is duidelijker: zoals in figuur 3b te zien is, hangen de prijzen van de rode bonen zeer sterk samen. Deze hebben zelfs een correlatie van 0,999162. Deze prijzen stijgen en dalen bijna even sterk. De prijzen van rijst en maïs zijn in dit land constanter. In 2008 zijn in de prijzen voor bonen en rijst pieken te zien, in 2011 zijn in de prijzen voor bonen en maïs pieken te zien.

Ook de voedselprijzen in Nepal correleren sterk, met name die van rijst en tarwe. Zo sterk zelfs dat ze een correlatie hebben van 0.9982. Waar de rijstprijs in 2004 stopt met stijgen, doet de tarweprijs dat ook. Zoals in figuur 3c te zien is,

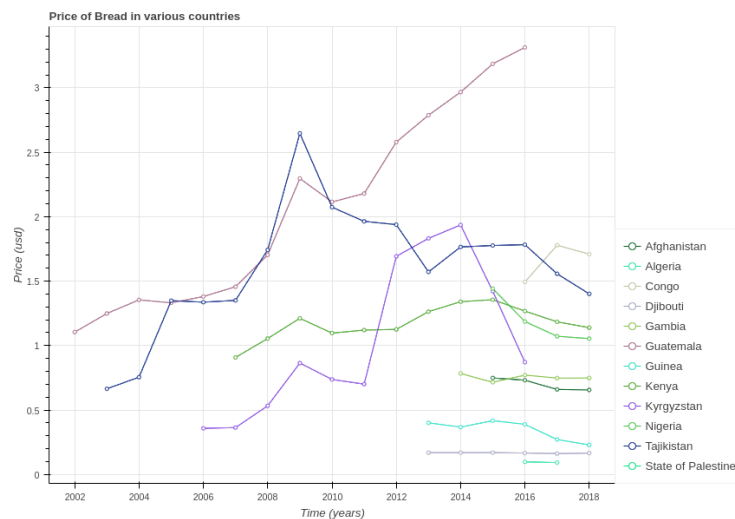
stijgen en dalen ze samen tot aan 2008. Zoals in Nepal de rijstprijis en tarweprijis samenhangen, zo is dat in Niger het geval voor maïs, sorghum en millet. Deze producten correleren sterk met elkaar. Bonen laten een minder sterke correlatie zien met maïs, sorghum en millet dan die producten onderling. Door de tijd heen zijn in figuur 3d pieken in voedselprijzen te zien in 1999 en 2001, waarna de prijzen aan de hoge kant blijven. Over de voedselprijzen in Senegal is te zeggen dat deze sterk samenhangen. Vooral de prijzen van sorghum en millet, allebei soorten koren, komen sterk overeen. Deze hebben een correlatie van 0.996852. Prijspieken zijn in figuur 4a voor alle producten te zien in 2004, en een daling voor alle producten is te zien in 2005. In Sudan hangen de prijzen van de verschillende producten, allemaal granen, zeer sterk samen. De gelijklopende lijnen in figuur 4b bevestigen dit. Vooral millet en sorghum dalen en stijgen gelijk, een correlatie van 0.99557. In Zuid Sudan is een soortgelijke correlatie te zien: van 2012 tot 2015 zijn de voedselprijzen redelijk gelijk. Zoals te zien is in figuur 4c stijgen in 2015 de prijzen van alle producten en in 2016 dalen deze prijzen weer, om weer licht te stijgen richting 2017. De voedselprijzen in Tanzania correleren ten slotte ook: rijst en maïs hebben een correlatie van 0,98118. In figuur 4d zijn in 2009 en 2012-2013 vergelijkbare pieken te zien.

Als de prijs van graan stijgt, stijgt de prijs van brood dan ook?

Terwijl tarwe duurder wordt, zie figuur 2a, is in figuur 2b te zien dat de prijs van brood over de jaren heen ongeveer gelijk blijft. Ze stijgen niet even hard en in ieder geval niet op hetzelfde moment, dus er is geen verband.

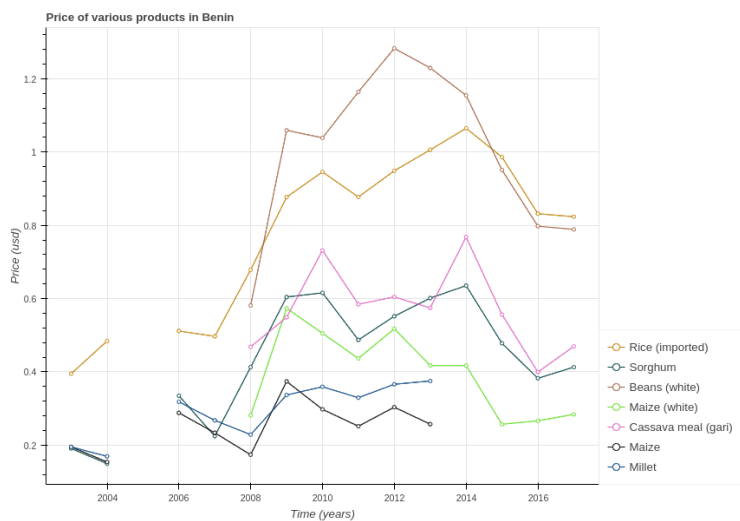


(a) Prijs van tarwe in verschillende landen.

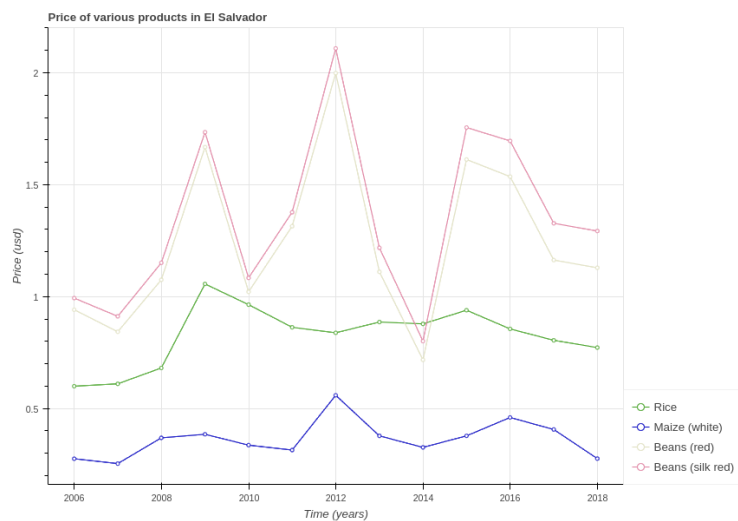


(b) de prijs van brood in verschillende landen

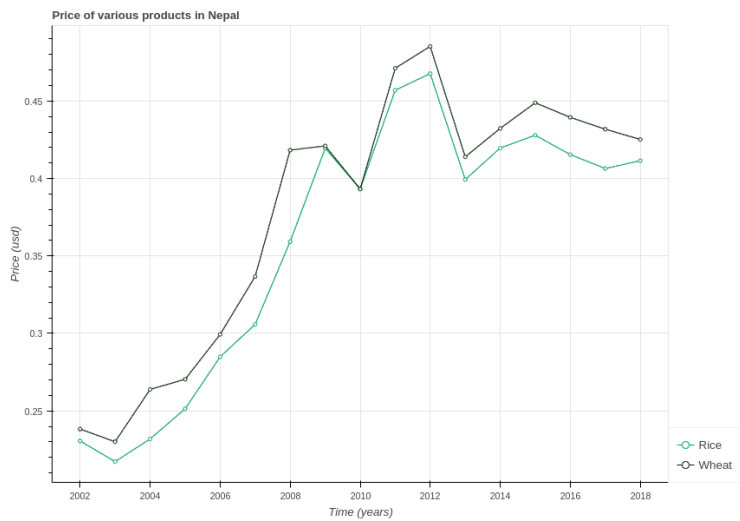
Figure 2: Prijzen van brood en graan



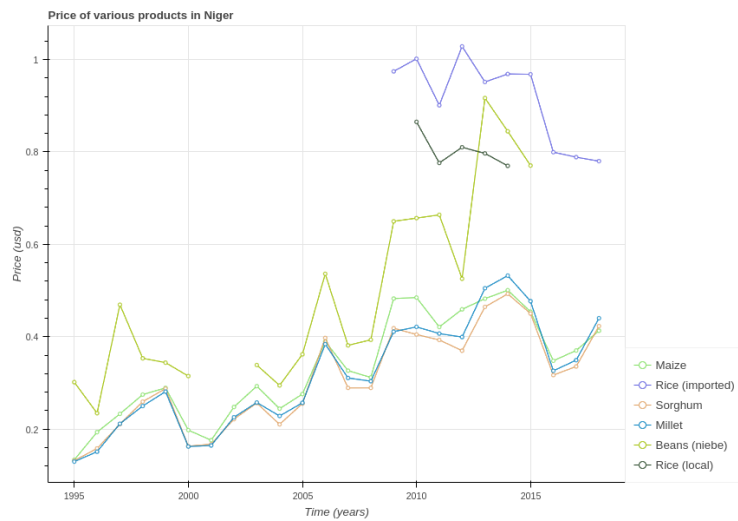
(a) voedselprijzen in Benin



(b) voedselprijzen in El Salvador

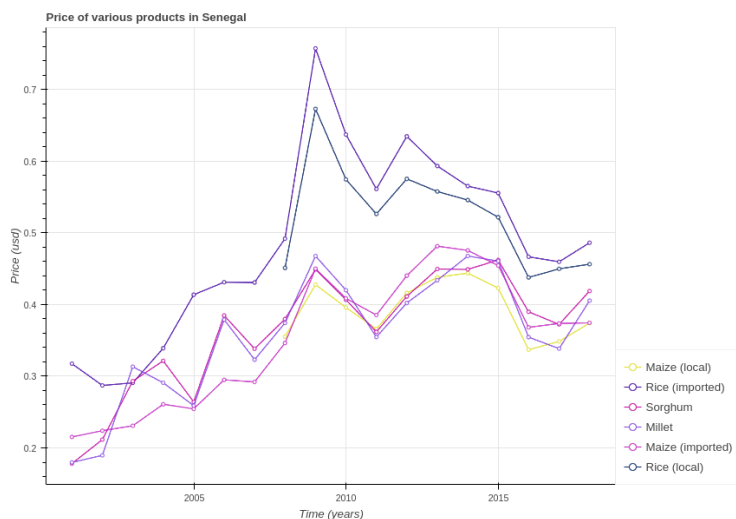


(c) voedselprijzen in Nepal

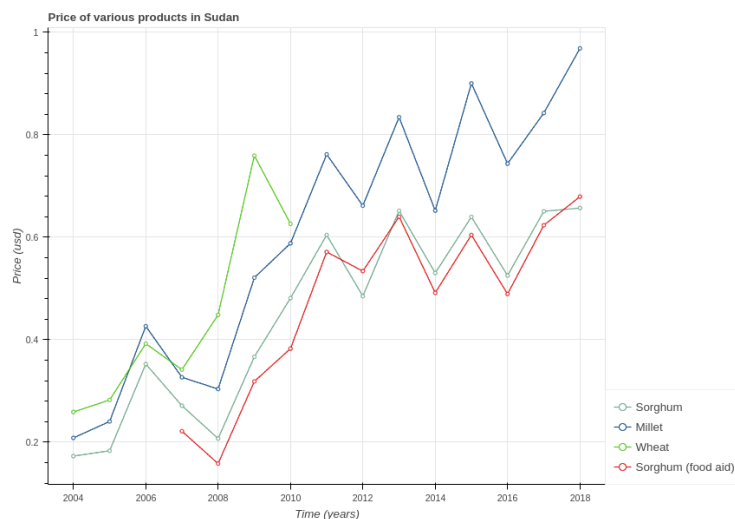


(d) voedselprijzen in Niger

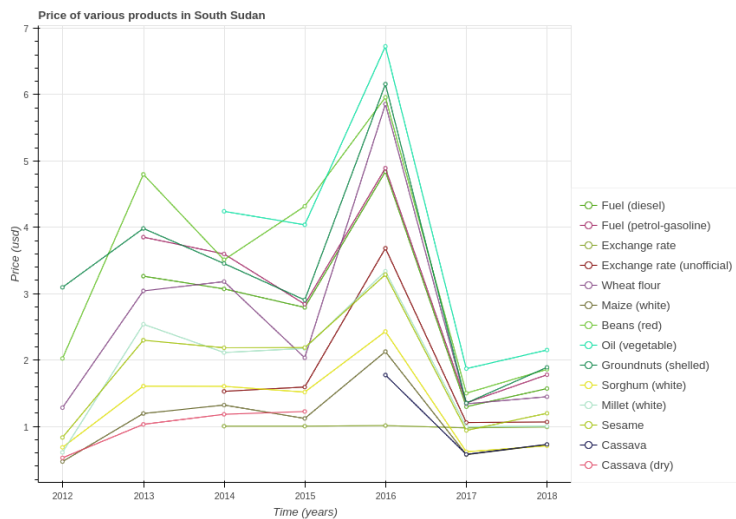
Figure 3: Grafieken bij de vraag "Zijn er voedselprijzen die een negatieve of een positieve correlatie laten zien?"



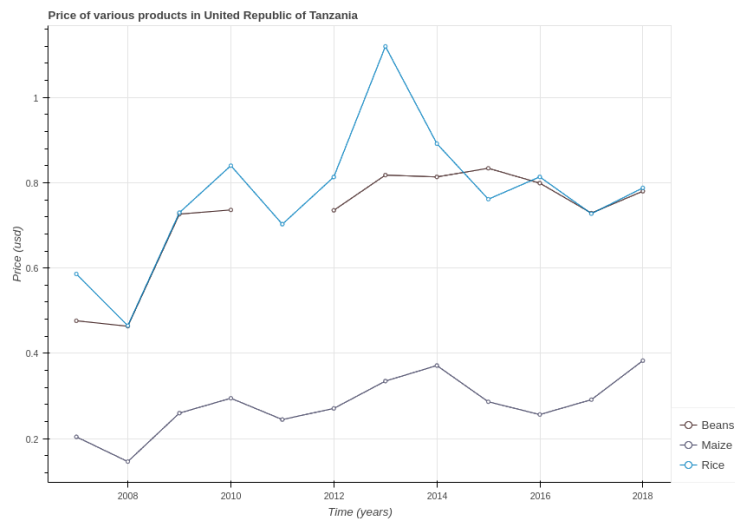
(a) voedselprijzen in Senegal



(b) voedselprijzen in Soedan



(c) voedselprijzen in Zuid Soedan



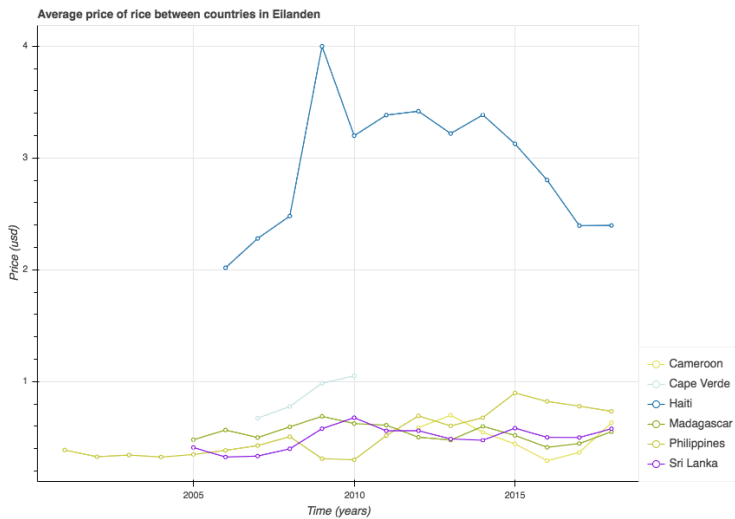
(d) voedselprijzen in Tanzania

Figure 4: Grafieken bij de vraag "Zijn er voedselprijzen die een negatieve of een positieve correlatie laten zien?"

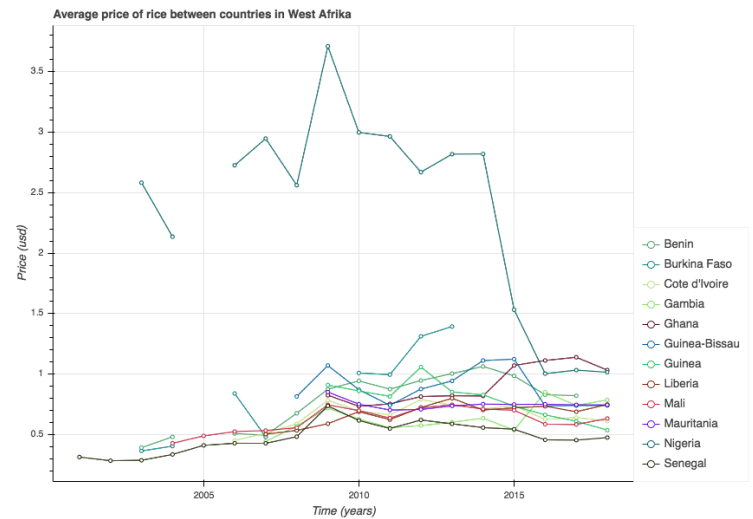
Laten landen in dezelfde regio ook soortgelijke prijsstijgingen en -dalingen zien? Zo nee, is daar een mogelijke verklaring voor te vinden? Over het algemeen hebben landen in dezelfde regio's vergelijkbare prijsverschillen. Sterker nog, de meeste trends in de rijstprijs zijn door meerdere regio's heen te zien. Een voorbeeld hiervan is een piek rond 2009-2010, die in de meeste landen in verschillende maten terug te vinden is, ongeacht in welke

regio's ze zitten. Deze piek is ook terug te vinden in het overzicht met daarin de gemiddelde rijstprijs per regio (Figuur 6. Maar er zijn ook regio-specifieke overeenkomsten te vinden: zo heeft West-Afrika een piek in de rijstprijs rond 2003, en is deze vrijwel niet terug te vinden in andere regio's. De regio die de minste correlatie tussen landen vertoont is de Eilanden regio, waarvan de leden overal op de wereld verspreid liggen.

Niet ieder land volgt deze regel. In figuur 5 zijn twee voorbeelden van zo'n landen te zien.



(a) Haïti is hier de duidelijke outlier



(b) Nigeria is hier de duidelijke outlier

Figure 5: Voorbeelden van afwijkende landen ten opzichte van hun regio's

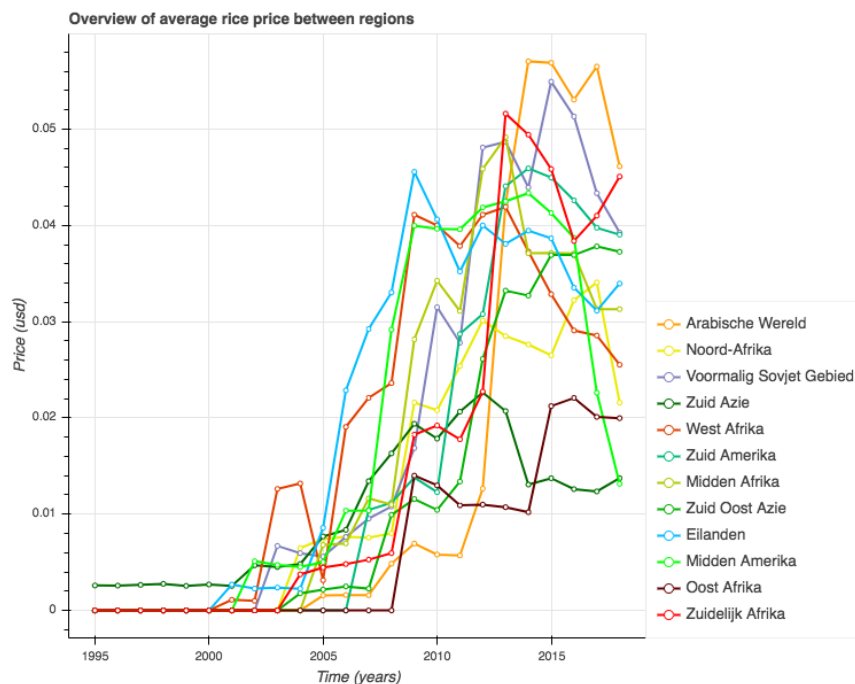
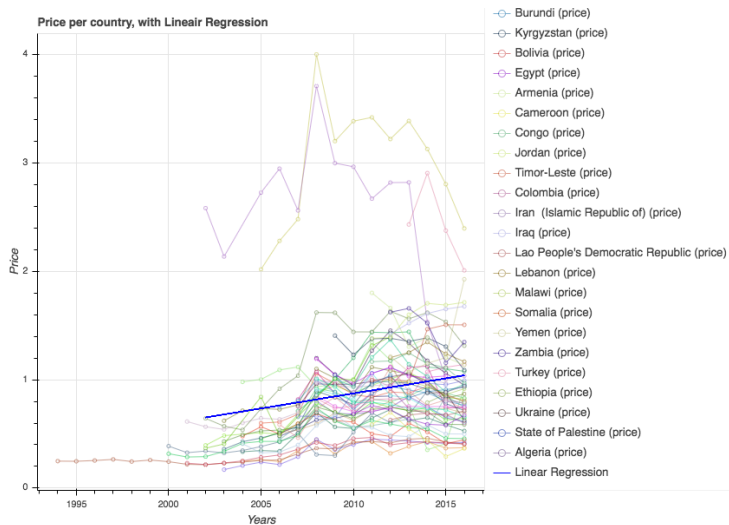
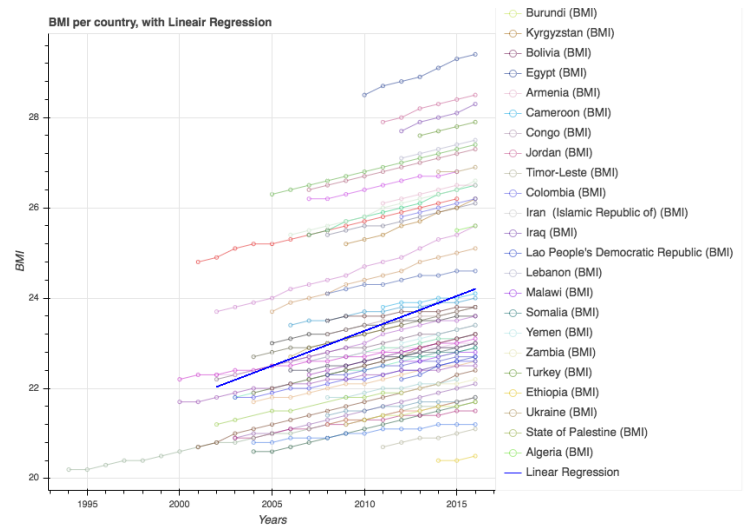


Figure 6: Overzicht van gemiddelde rijsprijzen per regio

Kunnen er correlaties worden gevonden tussen de gemiddelde BMI van een land en de gemiddelde prijs van rijst in dat land? Uit figuur 7 blijkt dat zowel de BMI als de prijs van rijst een stijgende lijn vormt. Om dit te verduidelijken, is met behulp van lineaire regressie (en dan in het specifiek *Least Squares*) een trendlijn geplot van de data. Zoals te zien is, zijn beide regressies stijgend, en dus is er een correlatie tussen de gemiddelde BMI en de gemiddelde rijstprijz van een land.



(a) Gemiddelde rijstprijzen per land, door de tijd heen



(b) BMI per land, door de tijd heen

Figure 7: Lineaire regressie

Zijn er verbanden tussen voedselprijzen tussen regio's binnen hetzelfde land? In algemene zin: ja, deze verbanden zijn aanwezig. Deze zijn niet in alle landen zichtbaar, maar in 35 van de 74 landen zijn verbanden te ontdekken in de rijstprijzen tussen verschillende regio's binnen een land. De landen waar verbanden tussen prijsverschillen waarneembaar waren zijn:

- | | | |
|---------------|--------------|---------------|
| • Afghanistan | • Guinea | • Niger |
| • Armenië | • Haïti | • Pakistan |
| • Bangladesh | • India | • Philippines |
| • Benin | • Kirgyzstan | • Rwanda |
| • Bolivia | • Laos | • Senegal |
| • Burundi | • Lebanon | • Syrië |
| • Cambodja | • Liberia | • Ukraine |
| • Colombia | • Madagascar | • Tanzania |
| • Djibouti | • Mali | • Zambia |
| • El Salvador | • Mozambique | |
| • Gambia | • Nepal | |

Deze lijst met landen is te rangschikken in drie groepen: landen waarin een

sterke samenhang zichtbaar is tussen de regio's. De prijzen liggen zeer dicht bij elkaar en stijgen en dalen gelijk. Zie ter illustratie figuur 8a. Tot deze groep behoren de volgende landen:

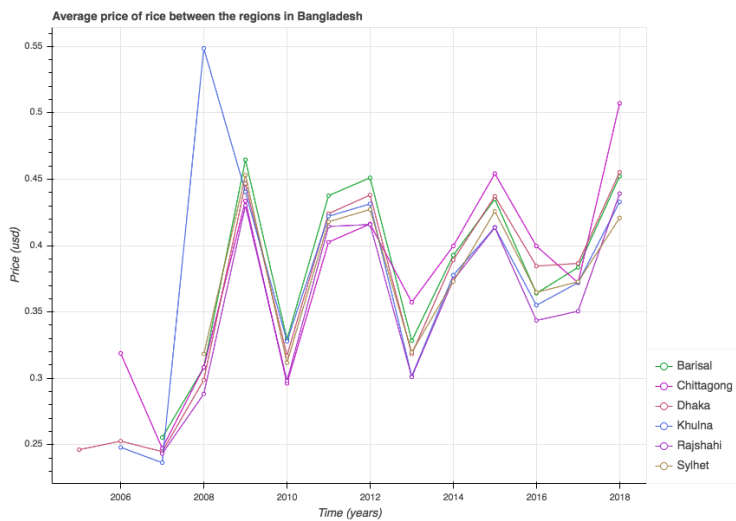
- Bangladesh
- Madagascar
- Tanzania
- Cambodja
- Niger
- El Salvador
- Senegal
- Oekraïne

De tweede groep bestaat uit de landen waartussen zeker algemene verbanden te zien zijn, maar niet zo sterk als in bovengenoemde groep. De algemene trend is gelijk, maar ten opzichte van elkaar kan de trend in de rijstprijz verschillen. Ter illustratie, zie figuur Voorbeelden-Regio's (B). Deze trend is aanwezig in de volgende landen:

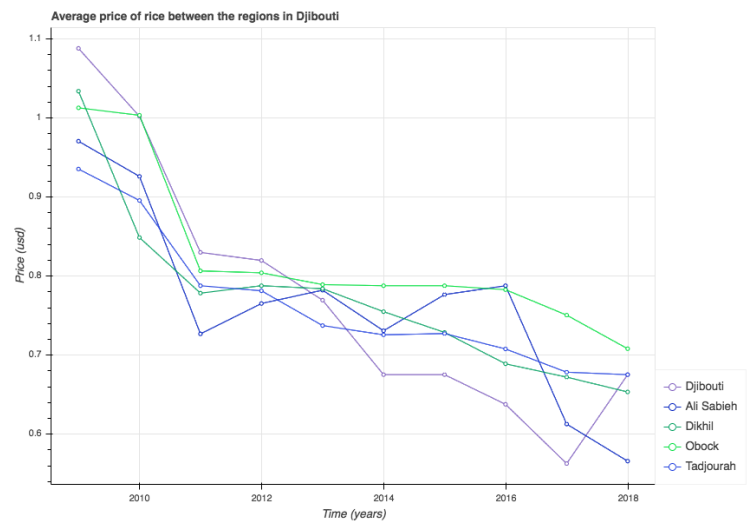
- Afghanistan
- Djibouti
- Liberia
- Armenië
- Guinea
- Mali
- Benin
- India
- Mozambique
- Bolivia
- Laos
- Nepal
- Colombia
- Lebanon
- Zambia

Ten slotte is er de groep landen die overeenkomsten heeft met zowel de eerst als de tweede genoemde groep. Na verloop van tijd veranderde de sterke samenhang in rijstprijz tussen de regio's naar een verband waarin de algemene trend gelijk blijft, maar waarin de trend van de rijstprijz bij regio's onderling van elkaar verschillen. Een voorbeeld van deze trend is te zien in figuur Voorbeelden-Regio's (C). Onder deze groep vallen de volgende landen:

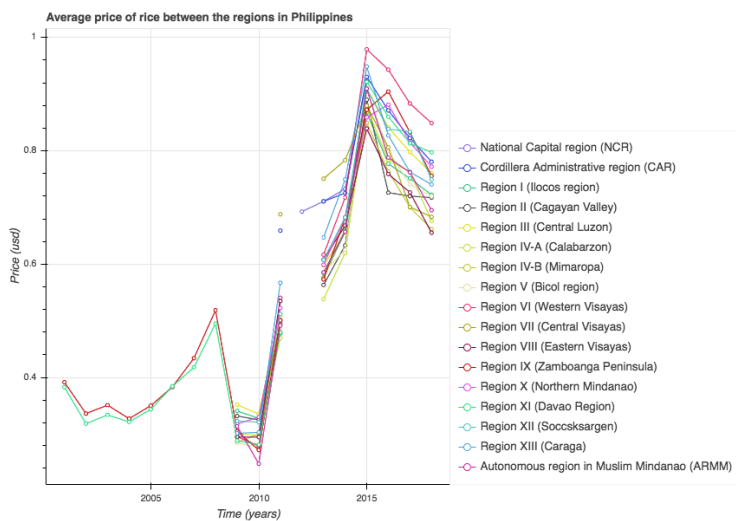
- Gambia
- Kirgyzstan
- Philippines
- Haïti
- Pakistan
- Rwanda



(a) Regio's binnen Bangladesh, als voorbeeld van een sterk verband



(b) Regio's binnen Djibouti, als voorbeeld van een globaal verband, maar specifieke verschillen



(c) Regio's binnen de Filipijnen, als voorbeeld van een deel sterk verband, deel globaal verband

Figure 8: Voorbeelden van verbanden tussen regio's binnen een land

Als de waarde van de plaatselijke valuta vergeleken met de dollar omhoog/omlaag gaat, gebeurt dat dan ook met de voedselprijs, hier de rijstprijs, in dollars? Uit de data blijkt dat dit voor veruit de meeste valuta's geldt dat als de valuta ten opzichte van de dollar verandert, de voedselprijzen een soortgelijke wijziging ondergaan. Echter, het omgekeerde lijkt

niet te gelden: bijvoorbeeld BDT, of de Bangladeshi Taka, wordt gebruikt in Bangladesh. Zoals in figuur 9 te zien is, ondervinden de rijstprijzen in dat land stijgingen en dalingen die elk ongeveer twee jaar duren. Echter, de waarde van de Bangladeshi Taka lijkt hierdoor te zijn onaangetast. Maar het omgekeerde geldt wel: de algemene trend van de Taka wordt ook gevolgd door de rijstprijz.

Er zijn ook een paar valuta's waar dit niet voor geldt. Daarvan zijn AMD (Armenian Dram) en PHP (Philippine Peso) de valuta's die de minste correlatie vertonen met de bijbehorende voedselprijz. Verder zijn er ook nog een aantal munteenheden, zoals BIF (Burundian Franc), die in het grijze gebied vallen: het eerste of het laatste deel vertoont geen of een omgekeerd verband met de rijstprijz, maar de andere helft weer wel.

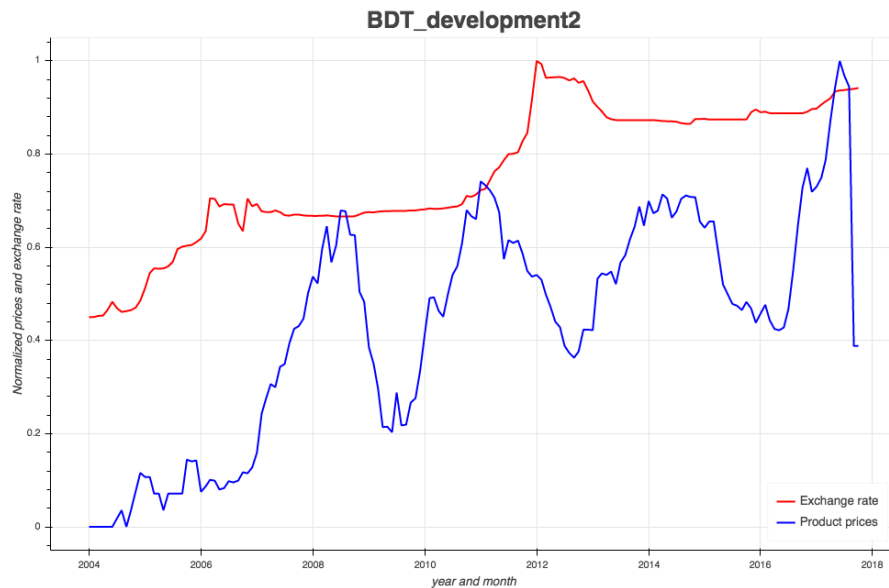


Figure 9: gemiddelde rijstprijz VS gemiddelde wisselkoers van BDT. Wisselkoers is berekend ten opzichte van USD, en de rijstprijz is genormaliseerd om goed te kunnen vergelijken met de exchange rate.

Is er een relatie te zien tussen enerzijds geografische ligging van een land en anderzijds de verhouding tussen BMI en voedselprijz, door de tijd heen? Ja, een relatie tussen deze eigenschappen is waarneembaar: in algemene zin kan worden gezegd dat zowel de gemiddelde voedselprijz, als het gemiddelde BMI omhoog gaat. Stijgingen van deze eigenschappen verschillen per regio en per tijdvak. Een voorbeeld: de rijstprijz in West Afrika ligt voor de verschillende landen binnen die regio in 2008-2010 dicht bij elkaar: tussen de 0,50 en 1,00. Kijk je echter naar de periode 2011-2014 voor dezelfde regio, dan is het verschil in voedselprijzen groter: tussen de 0,50 en 1,50 voor een kilo rijst in de verschillende landen in de regio. Ook het BMI ligt in deze periode ver uit

elkaar: waar in 2010 de gemiddelde BMI's nog tussen de 22,0 en 23,6 lagen, is dit in 2012 veranderd naar gemiddelde BMI's tussen de 21,8 en 24,4. Oftewel: een groter verschil door de tijd tussen de verschillende landen uit dezelfde regio qua prijs én BMI. Daarentegen is wel een verband te zien tussen gemiddeld BMI en geografische ligging: landen in dezelfde geografische regio hebben vaak ook een vergelijkbaar BMI: het gemiddelde BMI van landen in Midden-Afrika door de tijd heen is bijvoorbeeld 21,53, het gemiddelde BMI van landen in voormalig Sovjet Gebied is 25,52. De landen van voormalig Sovjet Gebied hebben een hogere BMI dan de landen in Midden Afrika.

Samenvattend: er zijn verbanden te ontdekken tussen verschillende eigenschappen, het is alleen niet zo dat landen binnen dezelfde regio een vergelijkbare verandering door de tijd in verhouding prijs/BMI hebben.

Om de analyse van de data compleet te maken, is ook gebruik gemaakt van *machine learning*-algoritmes: *linear regression* en *K-Means*. Lineaire regressie is al aangestipt bij de vraag "Kunnen er correlaties worden gevonden tussen de gemiddelde BMI van een land en de gemiddelde prijs van rijst in dat land?". K-Means, echter, nog niet. In figuur 10 staat de categorisatie van de prijzen-database, gedaan door K-Means. Omdat de data die K-Means heeft verwerkt tien dimensies heeft, is na het runnen van K-Means *t-SNE* op de data losgelaten. Dit algoritme heeft de tien dimensies teruggedrongen naar twee, zodat het resultaat visualiseerbaar is.

Verder heeft K-Means niet de hele database gecategoriseerd. Van de meer dan 700000 rijen, zijn er 20000 willekeurig gekozen. Dit is gedaan omwille van tijd- en geheugenbeperkingen. De data is ook gecategoriseerd in vier clusters, wat volgens figuur 11 het optimale aantal clusters is. Dit is af te lezen, omdat de richtingscoëfficiënt van de lijn bij vier clusters het sterkste veranderd.

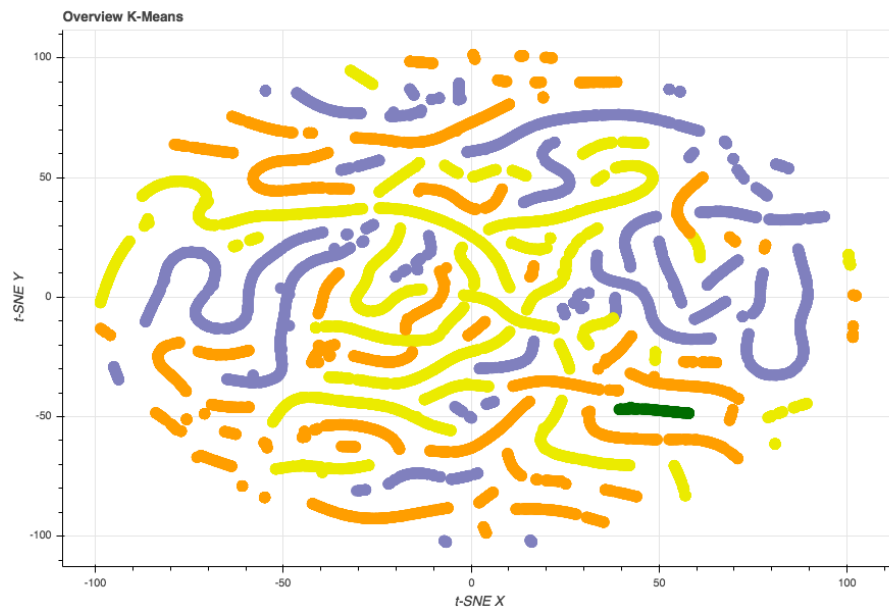


Figure 10: Het resultaat van het *K-Means* clusterings algoritme.

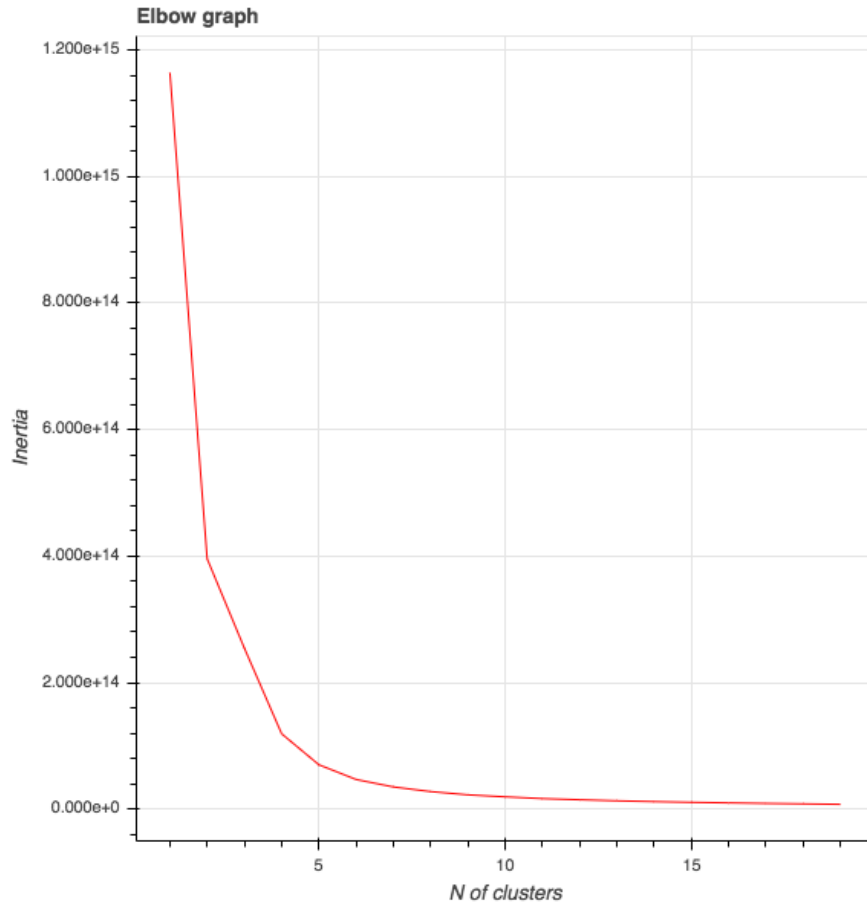


Figure 11: Een plot van de error (Inertia) VS het aantal clusters waarmee K-Means heeft gewerkt (n_clusters). Het optimale aantal clusters voor de desbetreffende database is af te lezen op het punt waar de richtingscoëfficiënt van de lijn het sterkst veranderd (n_clusters=4).

4 Discussie

In figuur 3a is in Benin vanaf 2006 een sterke prijsstijging te zien. Deze zou veroorzaakt kunnen zijn door de presidentsverkiezingen in 2006 in het land, die voor maatschappelijke onrust zorgden[2]. In El Salvador kunnen de prijsverschillen verklaard worden door de extreme regenval die dat land kent [7], waardoor de piek in dat jaar voor de prijs voor bonen en maïs, zichtbaar in figuur 3b te verklaren is: het was zo nat dat de bonen niet meer konden groeien. Dit kan ook verklaren waarom de prijs van rijst wel constant blijft, rijst groeit namelijk in natte grond. De pieken en dalen in voedselprijzen in Niger in figuur 3d zijn

te verklaren op basis van staatsinrichting. Van 1997 tot 1999 was het leger aan de macht. Aan het begin van deze periode dalen de prijzen: de gevolgen van het veranderen van de macht. Van 1997 tot 1999 worden de voedselprijzen echter steeds hoger, met een grote piek in 1999-2000. Eind 1999 vonden er presidentsverkiezingen plaats in Niger die leidden tot de vijfde republiek die stand hield tot 2009 [6].

Ook in Senegal zouden maatschappelijke ontwikkelingen reden kunnen zijn voor prijswijzigingen. In Senegal zorgt het Casamanceconflict [10] namelijk voor een burgeroorlog die al jaren woedt. Het Casamanceconflict is een conflict tussen de regering en the Movement of Democratic Forces of Casamance, een beweging die de zuidelijk gelegen regio Casamance onafhankelijk van Senegal wil hebben. De prijsstijgingen tot aan 2004 in figuur 4a zijn het gevolg van een steeds hoger oplopend conflict tussen beide partijen. In november 2004 sloten beide partijen vrede, met als waarschijnlijk gevolg de dalende voedselprijzen in 2005. Het MDFC bleef echter strijden voor een autonoom Casamance, wat de wisselende en hoge voedselprijzen na die tijd kan verklaren.

In Zuid Sudan zijn de maatschappelijke ontwikkelingen nog duidelijker zichtbaar: Er is in figuur 4c een enorme piek te zien in 2015. Op dat moment bereikte de burgeroorlog in dat land zijn hoogtepunt. Het uithongeren van de bevolking bleek een strategie van de strijdende partijen[5]. In augustus 2015 werd echter (een relatieve) vrede gesloten, waardoor de voedselprijzen na dat jaar weer konden dalen.

In Tanzania ten slotte, zijn de pieken in figuur 4d richting 2012 te verklaren door de extreme droogte die er toen in de Afrikaanse regio heerste[4], wat de afname in de hoogte van de prijs kan verklaren.

Voor het resultaat dat sommige landen soortgelijke rijstprijzen hebben ten opzichte van andere landen binnen dezelfde regio, is een erg voor de hand liggende verklaring te vinden. Landen binnen dezelfde regio hebben te maken met (over het algemeen) soortgelijk klimaat, en met soortgelijke variaties in regen, wind of andere, natuurlijke invloeden. Ook hebben naburige landen een soortgelijke cultuur, wat bijvoorbeeld het verschil in vraag naar een bepaald product tussen twee landen beperkt. Kort gezegd hebben landen binnen dezelfde regio dus geografische, ecologische en culturele overeenkomsten, wat dus de verschillen in voedselprijzen minimaliseert. Toch zijn hier uitzonderingen op te vinden, zichtbaar in figuur 5. De specifieke reden hiervoor verschilt natuurlijk per land, maar er zijn wel algemene redenen aan te wijzen. De meeste afwijkingen komen door externe factoren, zoals ecologische en politieke gebeurtenissen of rampen. Ook verschilt ieder land natuurlijk in meer of mindere mate van haar burens. Deze verschillen zijn ook zeker terug te vinden in de grafieken.

Helaas is lang niet ieder resultaat in dit onderzoek even betrouwbaar. Het is goed de conclusies wat betreft de vergelijking van de prijzen van brood en tarwe trekken in twijfel te trekken. In de resultaten wordt gesteld dat terwijl tarwe duurder wordt, brood ongeveer even duur blijft. Nu zijn deze conclusies beiden waar, maar het is de vraag in hoeverre deze gegevens met elkaar in verband staan. De prijsgegevens van brood zijn namelijk van andere landen uit andere

mondiale regio's dan de prijsgegevens van tarwe. Hierdoor is een betrouwbare vergelijking niet mogelijk.

Tussen de resultaten van de valuta's tegenover de gemiddelde rijstprijzen zitten ook een paar landen waar weinig goede conclusies over getrokken konden worden. Aangezien de koers van de valuta's ten opzichte van de Amerikaanse dollar zijn berekend, hebben sommige munteenheden een koers die teveel lijkt op de dollar om ook een merkbare stijging of daling te zien. Ook kan het zijn dat sommige van deze dollars misschien fouten in de data zijn: bijvoorbeeld DJF (Djiboutian Franc) is tot en met 2015 precies één dollar waard, waar het daarna op eens van de ene op de andere maand keldert naar 0 dollar. Dit is te onwaarschijnlijk om waar te zijn, en daarom waarschijnlijk een fout.

Bij het analyseren van de database zijn twee soorten *machine learning*-algoritmes gebruikt: *Linear Regression* en *K-Means*. Beide hadden zo hun problemen bij de implementatie en daarna de analyse. Bij de lineaire regressie was dit dat de meeste data in die grafieken pas rond 2002 begint (figuur 7). Dit houdt in dat, als de lineaire regressie berekend wordt over alle data, het gemiddelde vóór 2002 een stuk lager ligt dan het gemiddelde van na 2002, terwijl dit misschien niet zo is als men alle data zou hebben. De lineaire regressie stijgt daardoor een stuk sterker dan dat deze zou doen als men zelf een regressie zou tekenen over de bekende data. Vandaar dat ervoor is gekozen om de lineaire regressie pas vanaf 2002 te berekenen, zodat deze een stuk beter bij de data klopt. Zeker bij de grafiek over de BMI is te zien dat de regressie de lijnen goed volgt.

Bij K-Means was er echter een heel ander soort probleem. Zoals te zien (figuur 10) is het resultaat nogal ongewoon, en zijn er geen duidelijke clusters te onderscheiden in de dataset. Dit probleem is ook terug te zien in de *elbow graph* van de data. De inertia is bij het optimale aantal clusters nog steeds torenhoog (1.5 biljoen) (figure 11). Dit kan verschillende oorzaken hebben. Zo werkt het clustering algoritme veel minder goed als de clusters van de databases niet rond of niet zo symmetrisch zijn. Verder werkt het ook niet goed als het extreem high-dimensional data is. Maar dat is waarschijnlijk niet het geval, aangezien er maar 10 dimensies waren. Wel kan het dat de *euclidian distances*, die het algoritme gebruikt om de clusters te bepalen, misschien net onhandig werken voor dit specifieke aantal dimensies. In ieder geval werkt K-Means clustering niet zo goed op de database als gedacht.

De resultaten van dit onderzoek kunnen maatschappelijk gezien een grote invloed hebben. Met grote hoeveelheden data kunnen modellen worden gemaakt en kan voorspeld worden hoe voedselprijzen zich in de toekomst zullen gedragen. Neem de grafiek van de rijstprijz in Bangladesh, te zien in figuur 8a. Hierin zijn periodieke pieken en dalen te vinden. Deze pieken en dalen vallen samen met El Niño en La Niña. Door deze pieken en dalen globaal door te trekken is het mogelijk te voorspellen wanneer de volgende prijsstijgingen in het land plaatsvinden. Hier zou door overheden op geanticipeerd kunnen worden door in tijden met lage voedselprijzen veel op te slaan voor tijden dat rijst schaarser is.

Ook is gebleken dat het gemiddelde BMI inderdaad stijgt, in lijn met de

verwachting van de Wereldgezondheidsorganisatie. Daarbij is ook de gemiddelde voedselprijs gestegen sinds het begin van de metingen. Het is waarschijnlijk dat beide een gevolg is van de groeiende mondiale welvaart, maar op basis van de gegevens en resultaten niet met zekerheid te zeggen. Vervolgonderzoek zou dit moeten uitwijzen. Nog lastiger is het met het verband tussen maatschappelijke ontwikkelingen. In de discussie worden redenen aangedragen voor stijgende voedselprijzen, waaronder staatsgrepen en burgeroorlogen. Of deze gebeurtenissen ook daadwerkelijk de aanleiding zijn voor stijgende voedselprijzen, of dat daar ook andere oorzaken aan ten grondslag liggen zal moeten blijken uit vervolgonderzoek.

References

- [1] World Bank. *Monitoring Global Poverty Monitoring Global Poverty Report of the Commission on Global Poverty*. World Bank, 2017.
- [2] AFRICAN ELECTIONS DATABASE. Elections in benin.
- [3] François Bourguignon, Christian Morrisson. Inequality Among World Citizens: 1820-1992. *American Economic Review*, 92(4):727–744, 2002.
- [4] Adriana Homolova. Droogte in afrika in 7 grafieken.
- [5] Koert Lindijer. Honger is in zuid-soedan een doel van de oorlog geworden.
- [6] Monty G. Marshall. polity4x.
- [7] Het Parool. Veel doden en vermisten door regen midden-amerika.
- [8] Max Roser and Eesteban Ortiz-Ospina. Global extreme poverty. *Our world in data*, 2017.
- [9] Sterre van der Hee. Who: in 2030 is er een obesitas-epidemie, maar niet in nederland.
- [10] wikileaks. Casamance: The 2004 truce has ended.