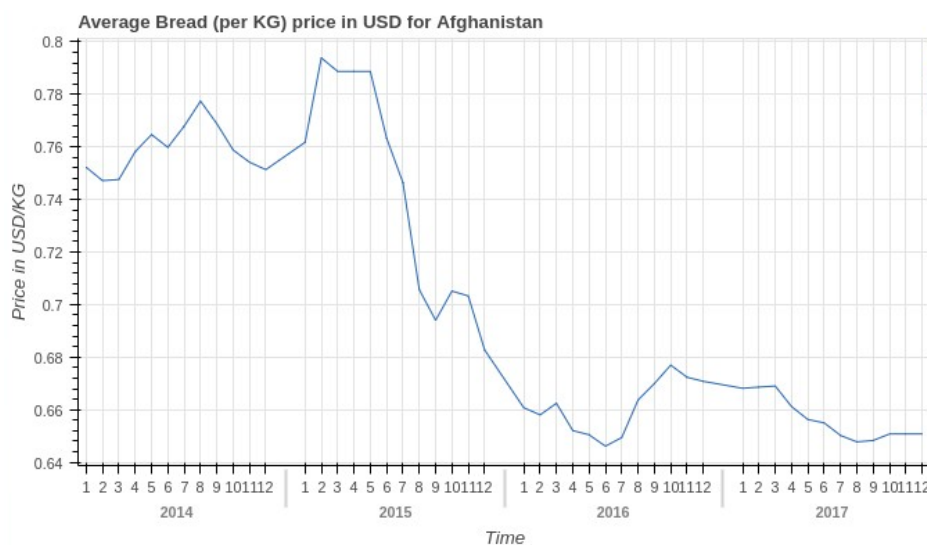


Exploratory Data Analysis Rapport - Thijs Kuipers

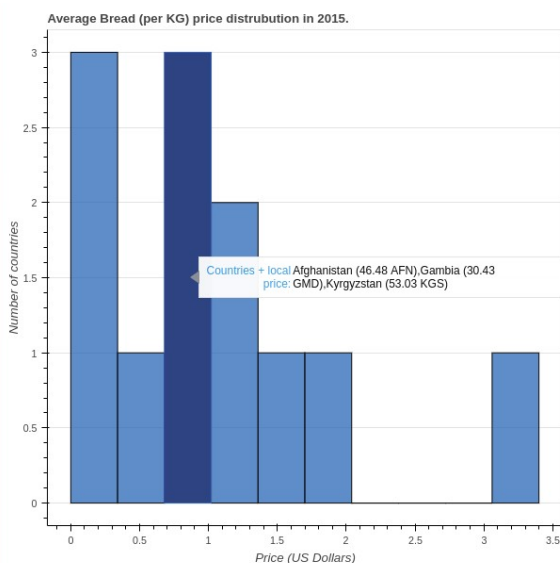
Het belangrijkste aspect van de gebruikte dataset is het in kaart brengen van de prijzen (in de lokale valuta) van (voornamelijk) voedselproducten binnen een groot aantal landen. Voor elk product is er per maand bijgehouden wat de prijs is. Omdat er veel verschillende producten en landen in de database staan, en niet alle producten prijzen hebben op de zelfde datum, is er voor gekozen om bij de EDA voornamelijk gebruik te maken van het land Afghanistan, en voor het product zelf zal brood (Bread) worden gekozen. Dit, omdat Afghanistan veel data bevat over dit product, en brood zelf een product is dat iedereen kent, waardoor de gegevens veel herkenbaarder zullen zijn. Dit deel van het EDA rapport zal zich bezig houden met de broodprijs van Afghanistan en hoe deze zich corraleert aan de volgend externe gegevens: binnenkomende vluchtelingen en de waarde van de Afghaanse valuta: de Afghani (AFD), ten opzichte van de USD.

Univariate Graphical EDA:



Figuur 1:

Om een overzicht te krijgen van de broodprijs van afghanistan is er een histogram gemaakt van de gemiddelde wereldwijde broodprijs van alle bekende landen. Het histogram beeldt de prijzen van het jaar 2015 uit, omdat dit het jaar was met de meeste informatie over de broodprijzen van verschillende landen in de database.



Univariate Non-Graphical EDA:

Uit het histogram valt meteen op dat de broodprijzen grotendeels rond en onder de 1 USD/KG liggen. De volgende gegevens blijken uit de data:

mean : 1.0
median : 0.78

Voor de *spread*:

variance : 0.84
sd (r) : 0.92

Figuur 2:

Uit de non-graphical EDA van de broodprijzen van Afghanistan zelf (zie figuur 1) blijken de volgende gegevens:

mean : 0.71
median : 0.68

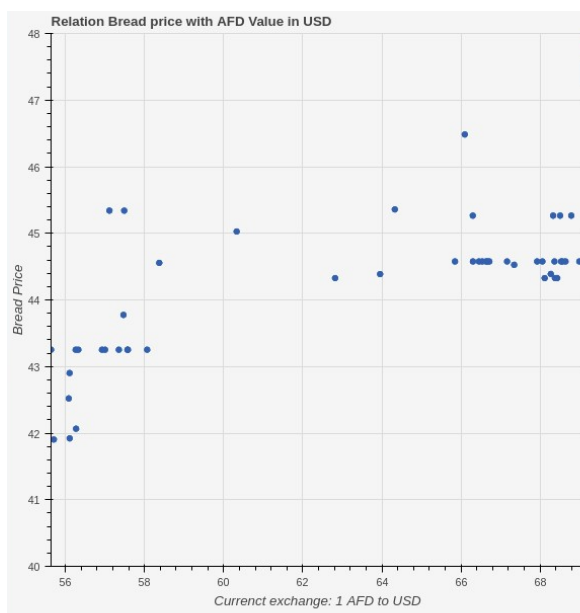
Voor de *spread*:

variance : 0.003
sd : 0.05

Deze gegevens komen uiteraard niet helemaal overeen met die van de globale broodprijzen. Toch, het gemiddelde en de mediaan van de globale prijs komen vrij dicht overeen met die van de Afghaanse broodprijs, zeker als de enorme uitschieter die te zien is bij de globale prijzen wordt laten vallen. Deze uitschieter zorgt ook voor de hogere variantie en standaardafwijking bij de globale prijzen. In figuur 1 is te zien dat Afghaanse broodprijs inderdaad veel minder varieert.

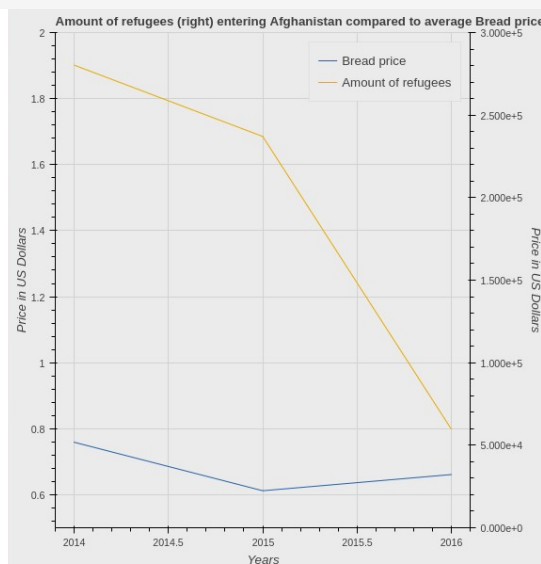
Multivariate Graphical EDA:

Zoals eerder werd beschreven, zijn de volgende twee correlaties bepaald: de correlatie tussen de waarde van de Afghaanse valuta (AFD) in USD en de broodprijs, alsmede de correlatie van het aantal vluchtelingen dat Afghanistan heeft opgenomen en de broodprijs.



Figuur 3:

Uit de scatterplot hiernaast valt meteen op dat er twee clusters zijn; een cluster met lagere broodprijzen en een lagere AFD waarde, en een cluster met hogere broodprijzen en een hogere AFD waarde. De twee clusters liggen dermate ver uit elkaar dat dit een vrij sterke correlatie tussen de broodprijs en AFD waarde suggereert. Dit komt ook overeen met de verwachting, immers, als de valuta meer waard wordt, zullen de producten ook in waarde stijgen.



Figuur 4:

De figuur hiernaast beeldt de relatie tussen het aantal vluchtelingen dat Afghanistan is binnengekomen en de broodprijs af. Omdat de dataset over vluchtelingen het aantal vluchtelingen dat een land binnenkomt, of uitgaat, per jaar bijhoudt, is het moeilijk om dit goed te vergelijken met de maandelijkse voedselprijzen, omdat er maar weinig data (als in datums) tussen de twee dataset overeenkomt. Toch blijkt uit de volgende grafiek dat er nauwelijks correlatie is tussen het aantal vluchtelingen en de broodprijs.

Multivariate Non-Graphical EDA:

De correlatie tussen de waarde van de afghaanse valuta en de broodprijzen is 0.69. Er is dus inderdaad een (vrij) sterke positieve correlatie tussen deze twee gegevens, zoals ook al werd verwacht, en bleek uit figuur 3.

De correlatie tussen het aantal vluchtelingen dat het land binnenkomt en de broodprijs is 0.37. Er is dus een erg lage positieve correlatie, zoals ook al bleek uit figuur 4. Blijkbaar hebben grote aantallen vluchtelingen dus geen effect op de broodprijs.

Hypothesen Interessante Kolommen:

Uit de bovenstaande resultaten blijkt dus dat er een sterke positieve correlatie is tussen de waarde van de Afghaanse Valuta AFD, en de prijs van brood. Het zal interessant zijn om naar de product-, prijs- en landkolommen te blijven kijken, om te onderzoeken of er meerdere producten en landen zijn waarbij een sterke positieve correlatie tussen de valuta waarde en product prijs kan worden vastgesteld.

Hiernaast kan worden gekeken naar of het aantal mensen dat juist uit een land wegvlucht invloed heeft op de productprijzen.

Er zal ook nog worden gekeken naar de invloed van bepaalde conflicten binnen een subset van landen uit de dataset op de product prijzen, en hoe deze conflicten wellicht relateren tot vluchtelingenstromen.