

Logboek

Wanneer?	Wat?
Week 1	
Maandag	<ul style="list-style-type: none">- We hebben ons verdiept in de opdracht.
Dinsdag	<ul style="list-style-type: none">- 2 extra datasets gekozen, 'migratiestromen' en 'temperatuur en neerslag'. We hebben voor deze datasets gekozen omdat je hiermee kunt vergelijken of het aantal emigranten verandert door veranderende voedselprijzen. Ook kun je eventueel afleiden aan de temperatuur en neerslag waar deze prijsverandering vandaan komt. <p>Meeting:</p> <ul style="list-style-type: none">- We hebben algemene informatie opgedaan over de opdracht met als belangrijkste informatie dat we meerdere datasets moeten kiezen.
Woensdag	<ul style="list-style-type: none">- We hebben landen gekozen waar we ons in verdiepen en alle losse datasets opgevraagd bij de desbetreffende websites. Deze datasets konden we per land los downloaden. Deze landen in de datasets van 'migratiestromen' en 'temperatuur en neerslag' zijn samengevoegd en staan genoemd in het technisch rapport.
Donderdag	<ul style="list-style-type: none">- We hebben een hoofdvraag, deelvragen en hypothese opgesteld. Zie technisch rapport.- Er is een functie geschreven om data van migratiestromen makkelijk per land en jaar op te vragen.- Iedereen heeft zich verdiept in de werking van Pandas <p>Meeting:</p> <ul style="list-style-type: none">- De voornaamste uitkomst van de meeting was dat we een hoofdvraag en deelvragen op moesten stellen. Hiernaast moesten we ook hypothesen opstellen. Ook moesten we beter documenteren waarom we welke keuzes hebben gemaakt.
Vrijdag	<ul style="list-style-type: none">- Alex heeft de verantwoordelijkheid gekregen voor de voedselprijzen dataset, Midas voor de migratiestromen en Take voor de temperatuur en neerslag dataset. Bob houdt het geheel in de gaten.- Jaren en maanden zijn voor elke dataset aan elkaar vastgemaakt. Mei 2017 wordt 2017.333 Dit hebben we besloten omdat het lastig bleek de data goed chronologisch te plotten.- Alle maanden waar een streepje stond in de data van migratiestromen zijn vervangen door de integer 0. Hierdoor kunnen we nu rekenen met alle waarden in kolom.
Week 2	
Maandag	<ul style="list-style-type: none">- Midas heeft Europese landen in de data van migratiestromen zijn samengevoegd zodat we het aantal vluchtelingen van landen uit Afrika en het Midden Oosten naar Europa gemakkelijk kunnen plotten..- Bob en Alex hebben een functie geschreven om alle markten in de foodprices dataset samen te voegen
Dinsdag	<ul style="list-style-type: none">- Take heeft de Github page aangemaakt.

	<ul style="list-style-type: none"> - Take heeft een functie geschreven om data van temperatuur en neerslag makkelijk per land en jaar op te vragen <p>Meeting:</p> <ul style="list-style-type: none"> - Ten eerste hebben we controlevragen gesteld of we op de goede weg zijn. Onze hoofdvraag en deelvragen zijn goedgekeurd. Onze hypothesen moeten beter onderbouwd worden met bronnen. Ook moet de functie om alle markten per land samen te voegen beter worden onderbouwd. Ten slotte kregen we als tip om de prijzen te normaliseren zodat je landen makkelijker kan vergelijken op procentuele prijsveranderingen.
Woensdag	<ul style="list-style-type: none"> - Veel namen zijn door Alex en Bob in alle datasets gelijkgetrokken. zo zijn in de migration dataset is Central African Rep. vervangen door Central African Republic. Hierdoor hebben nu alle landen in alle databases gelijke namen. - Midas heeft een normalize functie gemaakt. Met behulp van de 'Feature scaling' formule kan nu een kolom van een dataframe genormaliseerd worden
Donderdag	<ul style="list-style-type: none"> - Take en Alex hebben besloten welke data we gaan vergelijken. En uitgebreid aangevuld in het technisch rapport. - Bob en Midas zijn verder gegaan en alle functies om de verschillende data selectief te kunnen plotten. Zo kun je nu onder andere per land, per regio, een selectie van jaren en per product plotten <p>Meeting:</p> <ul style="list-style-type: none"> - Er is overlegd welke voorbereidingen er voor de machine learning algoritmes getroffen moet worden. Alle nominale waardes moeten vervangen worden door ints door middel van een dictionary. - Er moeten meer verschillende soorten grafieken geplot worden.
Vrijdag	<ul style="list-style-type: none"> - Er zijn voor elk afzonderlijk land grafieken geplot voor de prijs van elk lokaal product tegenover het aantal vluchtelingen in een land. De landen en producten met de meest opvallende resultaten zijn: Senegal:Maize:alles, Gambia:groundnuts:millet:palmoil:rice:sorghum, Madagascar:rice, Mali:rice, Niger:rice, Nigeria:yam, Rwanda 4 maanden waarin heel veel mensen gevlucht zijn, South-Sudan, Bij lage prijzen zijn er toch heel veel mensen gevlucht, Sudan alles, 4 maanden heel afwijkend - De afwijkende waardes vergelijken met het weer. - Controleren of er fouten in data kenya zijn. Malawi:maize outlier? Mauritania outlier? Mozambique 2 maanden met outliers? Zimbabwe heeft steeds 2 outliers. - Er is onderzocht hoe we interactieve grafieken op github pages kunnen maken.
Week 3	
Maandag	<ul style="list-style-type: none"> - Gebleken is dat de data die afgelopen vrijdag is geplot incorrect is. De grafieken hielden geen rekening met eventuele missing data waardoor de grafieken misleidend waren. Bob heeft een start gemaakt dit te corrigeren. - We lopen vast op het plotten van grafieken op website die data uit de database halen. Alex en Midas zijn zich hierin aan het verdiepen. Het plan is om een interactieve kaart van Afrika en het midden oosten te maken.
Dinsdag	<ul style="list-style-type: none"> - Take heeft een begin gemaakt aan het plotten van de correlaties tussen temperatuur en neerslag en foodprices. Het is echter nog erg lastig om dit goed in holoviews te krijgen. We kiezen ervoor dit in holoviews te plotten aangezien je hier in tegenstelling

	<p>tot bokeh met bar charts kunt plotten.</p> <ul style="list-style-type: none"> - We kwamen erachter dat we erg veel outliers hebben gemist. Midas heeft de voedseldata opnieuw geanalyseerd en data aangevuld/verbeterd. - Bob heeft outliers weggehaald uit de migratie data, zie het technische rapport. - Alex heeft verder gewerkt aan de Github page <p>Meeting:</p> <ul style="list-style-type: none"> - Tips gevraagd hoe te werken met github pages. Als tip gekregen te werken met D3. - Advies gekregen hoe de regression en clustering toe te passen
Woensdag	<ul style="list-style-type: none"> - Het is gelukt om door middel van least squares een lijn te plotten die een met een bepaalde error representeert - Er is een begin gemaakt met
Donderdag	<ul style="list-style-type: none"> - Bob heeft onderzocht hoe je de root mean square error moet plotten. - Het totaal van alles in Afrika met het midden oosten is vergeleken d.m.v stacked area chart <p>Meeting (10:00) - G3.02</p> <ul style="list-style-type: none"> - Clustering, meerdere voedselgroepen in een grafiek. En zoeken of het algoritme ze herkent. - Foodprices verschillen tussen landen. - Resultaten, geen grafieken voor 1 land.
Vrijdag	<ul style="list-style-type: none"> - Bob heeft verder gewerkt aan de indeling van het technisch rapport - Het is Midas en Alex gelukt om een interactieve wereldkaart te maken om op een handige manier data in kaart te brengen - Take heeft geanalyseerd wat de invloed is van temperatuur en regenval op de voedselprijzen en dit voor een deel aan het technisch rapport toegevoegd
Week 4	
Maandag	<ul style="list-style-type: none"> - Midas heeft veranderingen in temperatuur en neerslag interactief gemaakt op de page - Bob een chord chart gemaakt om inzichtelijk te maken welke landen dat van welke producten hebben - Alex heeft de opzet van de Github page geschreven -
Dinsdag	<ul style="list-style-type: none"> - Alex heeft een style voor de github page gekozen. De breedte van de plots stemmen we hier op af op 832 pixels. - De keuze is gemaakt om een extra deelvraag toe te voegen: In hoeverre zijn er verbanden tussen producten in verschillende regio's. Bob zal deze vraag uitzoeken en beantwoorden <p>Meeting (10:00) - G3.02</p> <p>Tips voor het technisch rapport:</p> <ul style="list-style-type: none"> - Titelblad - Table of contents - Vraag hypothese vraag hypothese - Waar komen de verwachtingen versie - Landen appendix - Pav schrijfgeregels volgen - Titels van grafieken in Latex zelf - Ook grafieken noemen waar niks aan te zien valt.

	<ul style="list-style-type: none"> - Deelvragen omzetten naar variabelen. - Dieper ingaan op prijzen per regio (Deelvraag toevoegen)
Woensdag	<ul style="list-style-type: none"> - Inleiding en methode afgerond voor het technisch rapport. - Bob heeft grafieken geplot die lokale met geïmporteerde producten vergelijken om zo de extra deelvraag te kunnen beantwoorden. - Iedereen heeft zijn eigen dataset beschreven in het technisch rapport.
Donderdag	<ul style="list-style-type: none"> - Iedereen heeft gewerkt aan het afronden van het technisch rapport, logboek en de eindpresentatie. <p>Meeting (10:00) - A1.24</p> <p>Tips voor het technisch rapport</p> <ul style="list-style-type: none"> - titel - verslag doornemen - ... - bijlagen ook in de tekst - bijlagen versus op 1 pagina - code verbeteren - conclusie en discussie
Vrijdag	<ul style="list-style-type: none"> - Eindpresentatie