# Logboek

|  |  |
| --- | --- |
| **Wanneer?** | **Wat?** |
| **Week 1** |  |
| Maandag | * We hebben ons verdiept in de opdracht. |
| Dinsdag | * 2 extra datasets gekozen, ‘migratiestromen’ en ‘temperatuur en neerslag   **Meeting:**   * We hebben algemene informatie opgedaan over de opdracht met als belangrijkste informatie dat we meerdere datasets moeten kiezen. |
| Woensdag | * 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. |
| Donderdag | * 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.   **Meeting:**   * De voornaamste uitkomst van de meeting was dat we een hoofdvraag en deelvragen op moesten stellen. Hiernaast moesten we ook hypotheses opstellen. Ook moesten we beter documenteren waarom we welke keuzes hebben gemaakt. |
| Vrijdag | * Jaren en maanden 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 waardes in kolom. |
| **Week 2** |  |
| Maandag | * Alle 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.. * Er is een functie geschreven om alle markten in de foodprices dataset samen te voegen |
| Dinsdag | * Er is Github page aangemaakt. * Er is een functie geschreven om data van temperatuur en neerslag makkelijk per land en jaar op te vragen   **Meeting:**   * Ten eerste hebben we controlevragen gesteld of we op de goede weg zijn. Onze hoofdvraag en deelvragen zijn goedgekeurd. Onze hypotheses 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 | * In de migration dataset is Central African Rep. vervangen door Central African Republic. Hierdoor hebben nu alle landen in alle databases gelijke namen. * Er is een normalize functie gemaakt. Met behulp van de ‘Feature scaling’ formule kan nu een kolom van een dataframe genormaliseerd worden |
| Donderdag | * Er is besloten welke data we gaan vergelijken. Zie technisch rapport. * Alle namen van landen in alle datasets zijn gelijk getrokken zodat deze eenvoudig te vergelijken zijn.   **Meeting:**   * 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 | * 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 | * We lopen vast op het plotten van grafieken op website die data uit de database halen. Parameters veranderen wordt ook lastig. |
| Dinsdag | * Er is een begin gemaakt aan het plotten van de correlaties tussen temperatuur en neerslag en foodprices. * We kwamen erachter dat er een fout in de functie voor het plotten zat. Aangezien niet elke dataset constant over dezelfde periode data heeft. Er is een extra functie geschreven om dit probleem te verhelpen. * Er zijn outliers weggehaald uit de migratie data, zie het technische rapport. * Er zijn landen en producten om ons verder op te focussen.   **Meeting:**   * 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 | * 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 |  |
| Vrijdag | * Er zijn stacked area charts geplot voor het Afrika en het Midden-Oosten * Er is verder gewerkt aan het technisch rapport * Het is gelukt om een interactieve wereldkaart te maken om op een handige manier data in kaart te brengen |
| **Week 4** |  |
| Maandag |  |
| Dinsdag |  |
| Woensdag |  |
| Donderdag |  |
| Vrijdag |  |