# 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 |  |
| Donderdag |  |
| Vrijdag |  |
| **Week 4** |  |
| Maandag |  |
| Dinsdag |  |
| Woensdag |  |
| Donderdag |  |
| Vrijdag |  |

# Planning

|  |  |  |  |
| --- | --- | --- | --- |
| **Wie?** | **Wat?** | **Wanneer?** | **....** |
|  |  | **Week 1** |  |
| Iedereen | Packages installeren | dinsdag |  |
| Iedereen | Nieuwe datasets zoeken, (sterftecijfer, regenval, droogte, natuurrampen, burgeroorlog, oorlog, ziekte uitbraken, wisselkoers (afhankelijkheid van andere landen), bevolkingsgrootte/dichtheid, vluchtelingen etc) en goed beargumenteren waarom jullie deze gekozen hebben | dinsdag |  |
| Iedereen | **meeting (10:00) - G3.02**   * Vragen wat we moeten doen * Algemene informatie opdoen | dinsdag |  |
| Iedereen | #1-(dataset global foodprices) invoeren in een csv | dinsdag |  |
| Bob | #2-(dataset europese migranten) invoeren in een csv | woensdag |  |
| Take | #3-(dataset temperatuur en regen) invoeren in een csv | woensdag |  |
| Take | (dataset temperatuur en regen) in 1 enkele csv zetten | woensdag |  |
| Alex | (dataset voedselprijzen midden-oosten) in 1 csv zetten | woensdag |  |
| Alex | #1-(gegeven dataset) Controleren / bepalen wat we doen met missing values | donderdag |  |
| Alex | #1 Verbeter verkeerde data | donderdag |  |
| Alex | #1 Normaliseer de data en bepaal wat voor waardes het zijn | donderdag |  |
| Midas | #2-(dataset europese migranten) Controleren / bepalen wat we doen met missing values | donderdag |  |
| Midas | #2 Verbeter verkeerde data | donderdag |  |
| Midas | #2 Normaliseer de data en bepaal wat voor waardes het zijn | donderdag |  |
| Take | #3- Controleren / bepalen wat we doen met missing values | donderdag |  |
| Take | #3 Verbeter verkeerde data | donderdag |  |
| Take | #3 Normaliseer de data en bepaal wat voor waardes het zijn | donderdag |  |
| Iedereen | **Meeting (10:00) - B2.01**   * Vragen welke dataset we toevoegen (klimaat en/of refugees) * Vraag bedenken over de gegeven dataset * Process boek met wanneer je keuzes maakt * Weten wat je attributen zijn * Hoofd- en deelvragen * Planning in github zetten * hypotheses, wat je denkt en weet uit bronnen | donderdag |  |
| Bob | Huidige informatie toevoegen aan het Technisch rapport | vrijdag |  |
| Iedereen | **Milestone** Data pre-processing | vrijdag |  |
|  |  | **Week 2** |  |
| Midas | * samenvoegen van alle Europese landen | maandag |  |
| Alex | * samenvoegen van de informatie van alle markten in een land | maandag |  |
| Take | * Functie om gemakkelijk data op te vragen af hebben | maandag |  |
| Iedereen | **Meeting (10:00) - C1.112**   * Zijn onze hoofdvraag en deelvragen okee? Ja   + Bronnen zoeken * Vragen of we ons het best alleen op Afrika kunnen richten - Nee * Was het samenvoegen van alle markten per land een goed idee? Ja maar moet wel onderbouwd worden * Hoe kunnen we het best data van verschillende datasets vergelijken? * Prijzen normaliseren? Procentuele toename mag * Github branches? * Correlatie geeft aan hoe verwant grafieken zijn, ook gebruiken. Spearman correlatie * Gemiddelde markten per land onderbouwen, error rate. Outlier eruit halen en kunnen toelichten. | dinsdag |  |
|  | Foodprices van verschillende producten per land kunnen vergelijken | dinsdag |  |
| Midas | Migratiestromen Midden Oosten naar Europa toevoegen | dinsdag |  |
| Bob | Bronnen en artikelen vinden | woensdag |  |
| Midas | Algemene normalise functie per kolom | woensdag |  |
| Iedereen | Opstellen welke plots we nodig hebben |  |  |
|  | namen landen gelijk maken |  |  |
| Iedereen | Exploratory Data Analysis pdf lezen (let op 40 pagina’s!) | woensdag |  |
| ... | Univariate non-graphical | ... |  |
| ... | Multivariate non-graphical | ... |  |
| ... | Univariate graphical | ... |  |
| ... | Multivariate graphical | ... |  |
| Iedereen | **Meeting (10:00) - G3.02**   * Vragen of het een goed idee was geweest om alles in een grote csv te zetten. * Welke voorbereidingen moeten er precies getroffen worden voor machine learning? Alle nominale waardes naar ints maar verder? sklearn. * Focussen op een aantal landen en daar dieper op ingaan? * Wanneer histogram wanneer lijn? kijken naar de grafiek op canvas * Wat valt er precies clusteren aangezien we nu vooral naar lijnen zoeken? * Interactie op website | donderdag |  |
| Bob | Logboek aanvullen | donderdag |  |
| Bob | functie om datasets aan te roepen. | donderdag |  |
| Bob | Onderbouwen waarom we averages gebruiken | donderdag |  |
| Alex en Take | Bepalen welke functies we gaan plotten | donderdag |  |
| Midas | verantwoordelijk om alle namen gelijk te trekken. |  |  |
| Alex | Kijken hoe je de GitHub page interactief kunt maken | vrijdag |  |
| Midas | plotten voedselprijzen → vluchtelingen | vrijdag |  |
| Take | Plotten neerslag en temperatuur → voedselprijzen | vrijdag |  |
| Bob | Huidige informatie toevoegen aan het Technisch rapport | vrijdag |  |
| Iedereen | **Milestone** EDA basis | vrijdag |  |
|  |  | **Week 3** |  |
| Bob | foutje in normalize functie fixen, geeft af en toe errors in de minimale data | maandag |  |
| Alex | Grafieken in github pages vanaf de datasets. Parameters kunnen aanpassen | maandag |  |
| Alex | Niet alles van de website in de master branch | maandag |  |
| Bob | compare moet dezelfde YearMonths hebben | dinsdag |  |
| Bob | Verder aan het technisch rapport | dinsdag |  |
| Take | Daadwerkelijk neerslag en temperatuur vs foodprices plotten | dinsdag |  |
| Alex | Chillen met de github pages | dinsdag |  |
| Iedereen | **Meeting (10:00) - G3.02**   * Kunnen we zeggen, deze grafiek zag er interessant uit dus gingen we deze nader onderzoeken? * wat zijn de vereisten voor de website? * hoe krijgen we de data van onze resultaten in een plot? * hoe moeten we een interactive data visualization doen zonder php? * Er moeten dynamische grafieken inzitten * D3 voor interactieve grafieken plotten | dinsdag |  |
| Bob | Regression dmv least squares | woensdag |  |
| Midas Take | Temperatuur en neerslag → voedselprijzen plotten + de interessante zaken er tussenuit halen | woensdag |  |
| Alex | Dmv een wereldkaart kunnen kiezen van welk land je een grafiek wilt | woensdag |  |
| Iedereen | **Meeting (10:00) - G3.02**   * ... * ... | donderdag |  |
| Iedereen | Opzet technisch rapport sturen | donderdag |  |
| Bob | Huidige informatie toevoegen aan het Technisch rapport | vrijdag |  |
| ... | Verslag in LaTeX zetten | ... |  |
| Iedereen | **Milestone** In-depth analysis | vrijdag |  |
|  |  | **Week 4** |  |
| Iedereen | **Meeting (10:00) - G3.02**   * ... * ... | dinsdag |  |
|  |  |  |  |
| Iedereen | AI EXPO | woensdag |  |
| Bob | Huidige informatie toevoegen aan het Technisch rapport | woensdag |  |
| Iedereen | **Meeting (10:00) - A1.24**   * ... * ... | donderdag |  |
| Iedereen | **Milestone** Final report   * Start with a short description of your problem * Describe methodology * Clearly describe the technical design * Clearly describe challenges that your have met during the project. * Defend your decisions by writing an argument of a most a single paragraph. * Make sure the document is complete and reflects the final state of your analysis. | donderdag |  |
| Iedereen | **Milestone** Final code   * the website: index.html * the README.md with link to your github pages * the logbook pdf * a report * a docs folder for the images in your documents * a separate folder containing your project's code and data * any configuration files like .gitignore | donderdag |  |
| Iedereen | **Milestone** Final demo   * Start with an introduction of the goals of your project * Answers to the questions supported by their visualization. * Top 3 most interesting finds, apart from the answers to the questions. | vrijdag |  |
| **Iedereen** | **Vakantie!!!** | 9 weken |  |

# Rollen

* Take zorgt ervoor dat Github netjes blijft
* Bob zorgt ervoor dat iedereen zich aan de planning houdt en doet zelf niks
* Alex zorgt voor de sfeer
* Midas zorgt voor voldoende vocht

# Opzet

* #1- De voedselprijzen stijgen/dalen → #2-Komen er meer/minder vluchtelingen naar Europa?
  + #1-Gegeven Dataset
  + #2-<http://popstats.unhcr.org/en/asylum_seekers_monthly>
  + Focus op het midden oosten en Afrika
  + landen: Afghanistan, Irak, Iran, Libanon, Palestina, Pakistan
* #3-Als de temperatuur/regenval verandert → #1-Stijgen/dalen de voedselprijzen?
  + #3-<http://sdwebx.worldbank.org/climateportal/index.cfm?page=downscaled_data_download&menu=historical>
  + #1-Gegeven Dataset
  + Focus op Afrika
  + Landen: Burkina faso, Cabo Verde, Kameroen, Centrale Afrikaanse Republiek, Burundi, Democratic Republic of the Congo, Djibouti, Ehtiopië, Gambië, Guinea, Kenia, Lesotho, Liberië, Madagascar, Malawi, Mali, Mauritanië, Mozambique, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, Somalië, Zuid-Sudan, Sudan, Timor-Leste, Uganda, Tanzania, Zambia, Zimbabwe