1. **Doelstelling:**  
   De eerste twee weken hebben wij besteed aan de data bestuderen en een hoofddoel formuleren. We wilden de data inkrimpen, dus besloten alleen naar avondeten te kijken. Het hoofddoel die we toen geformuleerd hebben was:  
     
   - Hoe kunnen we een app maken waarmee gebruikers een persoonlijk advies krijgen wat ze deze week kunnen eten en die zich kan aanpassen op voorkeuren van gebruikers?

- Op basis van favorieten van gebruiker, avondmaaltijden voor een week suggereren kijkend naar de bereidingstijd.  
  
Gedurende het project is dit nog veranderd en we hebben nooit meer iets met de bereidingstijd gedaan, omdat we niet zo ver kwamen en ons meer gefocust hebben op de keukens voorspellen. De doelstelling voor de tweede periode van het project was:  
  
- Kunnen we op basis van de ingrediënten die in een recept zit voorspellen tot welke keuken een recept behoord en zo recepten aanraden aan gebruikers met keukens als tags ?

1. **Data Preprocessing**De eerste periode bestond voornamelijk uit de data bestuderen en bewerken. Er waren 4 verschillende datasets: Tags (met alle recepten en bijbehorende tags, zoals “italiaans”, “vegetarisch” etc.), Recipes (met plaatjes en linkjes naar de recepten en een paar gegevens, zoals voor hoeveel personen en hoeveel calorieën), Nutritions (de nutritions per recept, dus hoeveel verzadigd vet, hoeveel eiwit, etc.), Ingredients (met van elk recept alle ingrediënten en hoeveel daarvan)  
     
     
   Voorbeelden van wat ik met de data heb gedaan in de ontdekkingsfase (In het 3e hoofdstuk ‘Notebooks’ zijn daar sommige voorbeelden van te vinden):
   1. Alle datasets ingeladen en bestudeerd. Gekeken naar de grootte en de gegevens.
   2. Nutritions gepivot en gegevens numeriek gemaakt om er visualisaties bij te kunnen maken en een simpel Lineair Regressie model te leren fitten op een echte dataset.

*Met de groep besloten we aan de slag te gaan met de datasets Tags & Ingredients. Als we voorspellingen (Classification models) wilde maken over recepten op basis van welke tags bij een recept hoorde of welke ingrediënten erin zaten, moest de data eerst bewerkt worden (Dummy data). Het moest numeriek gemaakt worden.*

* 1. Ik heb Tags omgezet naar een Dataframe met de recepten als index in de rijen en de tags als kolommen. Vervolgens liet ik in de dataframe een 1 invullen als een tag bij een recept voorkwam en een 0 als dit niet het geval was. Op deze manier is de data bruikbaar voor een model. Hetzelfde heb ik voor Ingredients gedaan.

*We ontdekte dat er in de dataset Ingredients veel ingrediënten voorkwamen die sterk op elkaar leken. Soms door haakjes, meervoud of verkleinwoordjes en soms doordat er meerdere soorten van iets zijn die heel erg op elkaar lijken. We wilden toen de dataset zo bewerken dat ingrediënten die sterk op elkaar leken, gecategoriseerd waren tot 1 ingrediënt.*

* 1. Ik heb in een notebook laten zien, hoeveel ingrediënten sterk op elkaar leken. Met de library ‘’ heb ik een functie geschreven die ingredienten die sterk op elkaar leken samen voegde. Micheal was hier ook mee bezig en we kwamen er samen achter dat er af en toe ingrediënten waren die onterecht samengevoegd werden, zoals “spruitjes” en “uitjes”. Daarom hebben we de functie zo uitgebreid dat de dataset naar het engels vertaald werd en vervolgens nog een keer het proces werd uitgevoerd. Alleen ingrediënten die in beide talen samengevoegd werden, werden ook echt aangepast  
     *Uiteindelijk is dit nooit gebruikt, aangezien we na aanleiding van de interne presentatie en een gesprek met Jeroen besloten dat dit niks zou toevoegen aan ons model. Jeroen sprak uit ervaring en legde uit dat het vaak niet voor verbeteringen zorgde en dat er ook recepten waren waarvoor het juist nadelig was om die kleine verschillen eruit te halen.*

*Voor dit project wilden we gaan werken met gesimuleerde gebruikers. We wilde de gebruikers in eerste instantie 10 favorieten meegeven en dan voorspellen met classificatie of die een recept wel of niet lekker zou vinden. Daarnaast wilde we in de train en testset zorgen dat er recepten zaten die een gebruiker lekker vond en juist niet, zodat die 0en (niet lekker) en 1en (wel lekker) kon voorspellen tijdens het trainen en testen.*

* 1. We hebben het lang gehad over hoe de input data er nou uit zou zien en dat was moeilijk helder te krijgen. Ik heb toen een visualistie gemaakt van hoe de trainset er ongeveer uit zou zien, die we ook tijdens presentaties gebruikt hebben:  
       
     Afbeelding met tafel

     Automatisch gegenereerde beschrijving  
       
      Ik heb, net als andere groepsgenoten, gepuzzeld met het schrijven van functies voor gesimuleerde gebruikers en voor het maken van zo’n dataframe als hierboven. Uiteindelijk had Bonno als eerst een snelle handige manier om dit te maken en we hebben toen de rest van het project zijn functies en gesimuleerde gebruikers gebruikt.

*Op een gegeven moment hadden we bedacht dat we dus naar de ingredienten kijken om te voorspellen welke keuken een recept is, zo kan het model herkennen of een recept bij een bepaalde smaak van een gebruiker past. Stel de gebruiker houdt van italiaanse recepten, dan zal het model snel een recept aanraden waarvan het denkt dat het uit dezelfde keuken komt.*

* 1. Er zijn nog meer verschillende ingrediënten dan tags, dus leek het ons handig om wat te krimpen en niet alle ingredienten mee te nemen in het model. Ik heb de dataset ingekort en alleen de meest gebruikte

1. **Predictive Analyses**  
   Als de data klaargemaakt was om te gebruiken in classifiactie modellen, konden we werken aan het voorspellingsmodel. We begonnen simpel en los van elkaar, maar hebben de laatste paar (4) weken gezamenlijk gewerkt aan het eindmodel. Bonno heeft grotendeels de code daarvoor geschreven, op een geven moment vond ik het moeilijk veel te helpen, omdat er steeds opeens superveel code bijkwam en ik dat niet zo snel kon begrijpen. Daarom heb ik soms meer m’n eigen ding gedaan en het wat simpeler gehouden. Daarnaast heb ik tijdens de meetings wel goed nagedacht over wat handig was en de resultaten van modellen geanalyseerd. Ik vind zelf dat ik met nuttige toevoegingen kwam tijdens de meeting met docenten en zo veel heb geholpen aan het verbeteren van het model. Ook heb ik tussendoor kleine functies geschreven die anderen toegevoegd hebben aan het grote eindmodel.  
     
   Voorbeelden van wat ik gedaan heb tijdens het maken van voorspellingen en werken aan ons model (In het 3e hoofdstuk ‘Notebooks’ zijn daar sommige voorbeelden van te vinden):
   1. Simpele Lineaire Regressie – df Nutritions.  
      Om te oefenen met het trainen van een model op een echte dataset, heb ik gekeken naar koolhydraten en energie (in kcal) in recepten. Ik heb een scatterplot gemaakt en vervolgens een simpel Lineair Regressie model gefit, die voorspeld hoeveel calorieën er in een recept zitten op basis van de hoeveelheid koolhydraten. Dit is gelukt, maar het model had wel een hele hoge RMSE dus is zeker nog niet optimaal.
   2. Tomaten classificatie model – df Ingredients  
      Omdat wij in ons project werken met classificatiemodellen heb ik vervolgens ook een simpel model gemaakt wat van 1 ingredient voorspelt of het wel of niet in het recept zit op basis van andere ingrediënten. Ik heb drie verschillende modellen gefit (Logistic Regression, K-Nearest Neighbour en Desicion Tree) en deze vergeleken. Ook hier zijn de resultaten niet zo goed, maar finetunen van de modellen kan helpen.
   3. Werken aan het eindmodel samen met de groep.
2. **Resultaten**  
   Resultaten model. Gelukt recepten aanraden op basis van ingrediënten en keuken bij gebruikers  
   Applicatie
3. **Conclusie/terugkoppeling doelstelling**  
   Uiteindelijk is onze doelstelling gedurende het project dus veranderd. Het doel wat wij voor de eerste periode hadden is niet behaald. Het model wat wij wilde ontwikkelen dat op basis van favorieten van gesimuleerde gebruikers recepten aan raadde kreeg niet een goed accuracy score en na veel finetuning en het proberen van veel verschillende modellen besloten we te beginnen met een andere aanpak.  
     
   tweede doel beter
4. **Aanbevelingen**Applicatie op model aansluiten  
   rekening houden met dieet + gevarieerd weekmenu  
   model verbeteren?