**Process Book Group 39**

June 5:

* Met for the first time as a group
* Made sure everyone had git and python3 installed and created git repository
* Chose dataset: Global Food Prices Database (WFP)
* Had first meeting with TA Houda
* Came up with subquestions:
  + Are there any food prices that are show negative/positive correlation, and is this correlation present throughout the years, or perhaps only in certain period? Can you perhaps detect possible ingredients of a certain other food product?
  + Do countries in similar regions, also show similar price differences? And if differences occur, can you find a potential explanation?
  + What are the correlations in production numbers of crops and the price of certain types of food? Are there correlations in production dominance of certain countries and the price of certain types of food?
* Created this process book
* Created report latex file
* Found database on crop production numbers per country
* Wrote hypotheses to subquestions

June 7

* Have a litte chat about the progress we made each the previous day.
* Write down the problems and prepare for the TA meeting.
* Problems we have:
  + Data cleaning problems, should we convert the prices to dollars or euro’s
  + Missing data in the price data set, we encounter a lot of data which is measured over a short period of time
  + Data problems with the production data set
* Solutions we have for our problems:
  + We decided to make a decision later in the process if we are going to convert the currency. It depends on the outcome of the tests.
  + We decided to try to eliminate data which is measured for a short period of time, still work in progress
  + We reduced the production dataset, we only kept the pure production data and filtered it along the countries we have in the price dataset. Note: the production dataset only contains production data per country per year.
* We have little with the source control on github, but we will be fine
* Matthijs is going to make a start with the eda part of this assignment. Trying to make some pots and do some pioneering work.
* Roel is going to check if the data of the production dataset is complete and determine what is missing. He is also going to find out which production product to link with the price products. (They differ a little bit)
* Tessel is going to finish the data cleaning, so we are ready for next week.
* Jesper is also going to start with the eda part of this assignment, and function as back-up for other problems which will arise.