Introduction

In this report, I will be discussing clustering and fitting of my chosen dataset. My task will be using various clustering techniques such as k-means or agglomerative clustering. Using the results from my clusters, I then perform exponential curve fitting in order to predict the values in the future whether it is trending up or down and the country of my choice is France, and the indicator is the cereal yield in the units of kg per hectare. The reason for my pick was to see whether a first world country such as France is still using agrarian methods in the future.

1. Clustering (K-means and Agglomerative)

I started by doing exploratory plots to see if there are any expected clusters from my chosen dataset of greenhouse gases, cereal yield and fertilizer. From those 3 data frames, I found that greenhouse against cereal yield looks the most interesting, so I made my decision to use that for clustering.

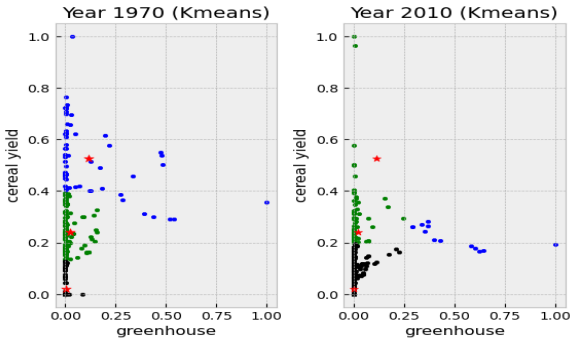

Figure 1 K-means plot cereal v greenhouse

For all method (Figure 1 and 2), it seems at least 2 clusters lean near 0. This could be due to a lot of the data that are Nan that was converted to 0 but we do see in (Figure 1) 1970 K-means there are values around 0.5 that contains a cluster in blue colour which is like the Agglomerative plot (Figure 2) where there are clusters around the value 0.4 at centre. However not all methods on each dataset are accurate. For example, on the year 2010 k-means on Figure 1 shows that the centre for green cluster is way off this may due that K-means do not work well with non-spherical result while centroids are also set at random.

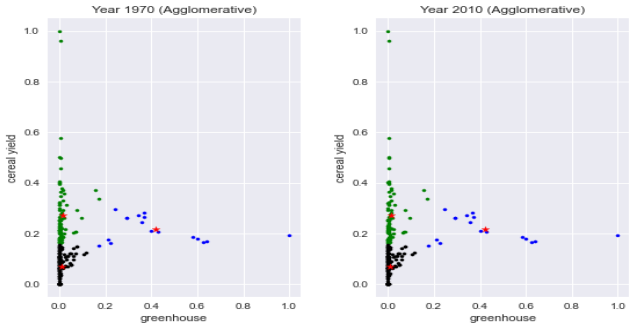

Figure 2: Agglomerative plot cereal v greenhouse

2. Curve Fitting

Fitting is used so that I can predict what my data would look like in the future. The indicator I picked was cereal yield in France, an agrarian society which relies on agricultural for their economy. On (Figure 3) I used the method of exponential curve fitting and I observe that the values are going down for France and in 10 years from the latest data 2018, I found that the value for year 2028 will be valued at 6814.9 (kg per hectare) but let's say 10 years from now at year 2032 the value will drop to 6776.08 (kg per hectare).

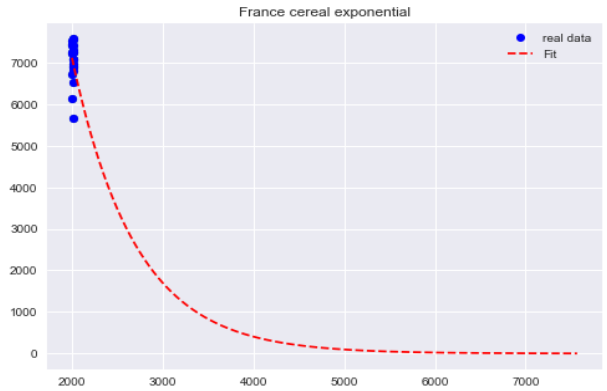

Figure 3 exponential curve fit of cereal yield for France

While the code was successful for my cereal yield. I found difficulty in generating result for my indicator greenhouse with the exponent curve fitting (Figure 4). However, one might expect the value to drop in a similar fashion as they do on (Figure 3) because I found on my dataset that while cereal yield is decreasing, so is the greenhouse which could mean that France might be slowly moving away from the cereal production.

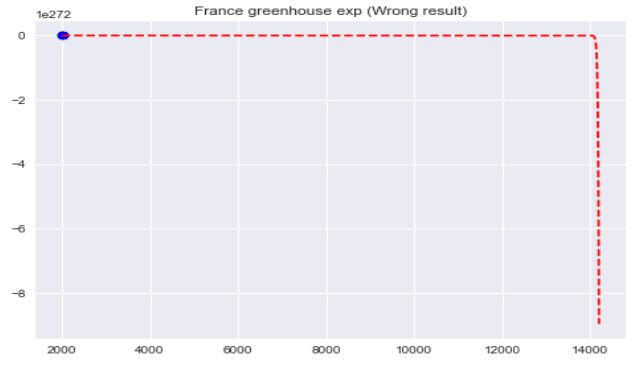

Figure 4 exponential curve fit greenhouse for France (INCORRECT)