Prozmi Biolabs group1 data analysis

Abstract:

The analysis surprisingly found that the food materials actually are quite similar regarding their impacts to the environment, as all the food groups were grouped into the same group while using the k-means clustering algorithm, and the result has determined that they are relatively similar overall when considering the combined land use, biodiversity, GHG emissions, and water use impacts.

Therefore, the main focus and efforts for the improvements on the sustainability aspect probably will be put on the issues related to the development of innovative technologies, the effective supply chain solutions, and further with the business and marketing strategies. But probably more data can be gathered in this case and to further obtain some more insightful conclusions.

Part 1

The statistics information of each indicator was calculated based on the food groups' environmental impact profiles, and a clustering algorithm was implemented after all the numerical values were standardised as well. And it's quite surprising that all the food groups were grouped into the same cluster, and this means the different kind of food groups actually have similar environmental impacts regarding the monitored aspects and suggests that despite the differences in the numerical values of the environmental impact metrics across the food groups, the result has determined that they are relatively similar overall when considering the combined land use, biodiversity, and water use impacts. The statistical information of the food groups are as below:

The Food Group: starch_rich
Feature: Land Use (m2*year)
Statistics (mean, min, std, median, max): [2.70269444 0.28 3.02964422 1.725 29.28]
Feature: Biodiversity (sp*yr*10^14)
Statistics (mean, min, std, median, max): [26.96611111 0.42 51.62498901 7.38 388.73]
Feature: GHG Emissions (kg CO2eq, IPCC2013 incl CC feedbacks)
Statistics (mean, min, std, median, max): [1.473 0.1 1.57815958 1.045 14.39]
Feature: Water Use (L)
Statistics (mean, min, std, median, max): [688.63116667 0. 1697.29830166 12.78 12406.61]

The Food Group: alcohols
Feature: Land Use (m2*year)
Statistics (mean, min, std, median, max): [1.29718182 0.21 0.96550706 1.285 4.01]
Feature: Biodiversity (sp*yr*10^14)
Statistics (mean, min, std, median, max): [15.92681818 0.15 30.20368024 5.3 145.33]
Feature: GHG Emissions (kg CO2eq, IPCC2013 incl CC feedbacks)
Statistics (mean, min, std, median, max): [1.33590909 0.44 0.59420121 1.1 4.7]

Feature: Water Use (L)
Statistics (mean, min, std, median, max): [28.80818182 1.63 69.03898446 6.96
336.19]

The Food Group: sugars
Feature: Land Use (m2*year)
Statistics (mean, min, std, median, max): [1.99611111 0.79 0.99709656 1.65 5.33]
Feature: Biodiversity (sp*yr*10^14)
Statistics (mean, min, std, median, max): [83.95138889 0. 114.88725853 27.055
425.32]
Feature: GHG Emissions (kg CO2eq, IPCC2013 incl CC feedbacks)
Statistics (mean, min, std, median, max): [1.93888889 0.55 0.85101174 1.9 5.19]
Feature: Water Use (L)
Statistics (mean, min, std, median, max): [544.48888889 8.14 820.01764498 17.745
3567.12]

The Food Group: protein_rich
Feature: Land Use (m2*year)
Statistics (mean, min, std, median, max): [3.65503378e+01 8.00000000e-02 1.37644168e+02
1.10100000e+01
 1.97194000e+03]
Feature: Biodiversity (sp*yr*10^14)
Statistics (mean, min, std, median, max): [4.86925338e+02 1.27000000e+00 2.31441312e+03
9.09100000e+01
 3.14593100e+04]
Feature: GHG Emissions (kg CO2eq, IPCC2013 incl CC feedbacks)
Statistics (mean, min, std, median, max): [19.81814189 -8.09 35.1064988 10.69
374.22]
Feature: Water Use (L)
Statistics (mean, min, std, median, max): [3673.48212838 0. 10808.41913735
1158.97
 110939.56]

The Food Group: milks
Feature: Land Use (m2*year)
Statistics (mean, min, std, median, max): [1.88980392 0.28 3.75530396 1.21 27.08
]
Feature: Biodiversity (sp*yr*10^14)
Statistics (mean, min, std, median, max): [6.28405882e+01 1.20000000e+00 3.43497132e+02
5.58000000e+00
 2.48510000e+03]
Feature: GHG Emissions (kg CO2eq, IPCC2013 incl CC feedbacks)
Statistics (mean, min, std, median, max): [2.12686275 0.72 2.10066795 1.9 14.58]
Feature: Water Use (L)
Statistics (mean, min, std, median, max): [2.87756078e+02 1.22000000e+00 6.02466169e+02
1.90410000e+02
 4.23362000e+03]

The Food Group: oils
Feature: Land Use (m2*year)
Statistics (mean, min, std, median, max): [13.03488189 2.02 15.38315194 7.91
135.05]

Feature: Biodiversity (sp*yr*10^14)
Statistics (mean, min, std, median, max): [4.82787559e+02 3.25000000e+00 1.76764970e+03 9.42200000e+01
 1.69235500e+04]
Feature: GHG Emissions (kg CO2eq, IPCC2013 incl CC feedbacks)
Statistics (mean, min, std, median, max): [4.46858268 -0.27 2.74370883 3.67 21.31]
Feature: Water Use (L)
Statistics (mean, min, std, median, max): [5.72934803e+02 1.43000000e+00 1.59293224e+03 2.99000000e+00
 1.21102500e+04]

The Food Group: vegetables
Feature: Land Use (m2*year)
Statistics (mean, min, std, median, max): [0.92373239 0.02 1.90436955 0.28 13.81]
Feature: Biodiversity (sp*yr*10^14)
Statistics (mean, min, std, median, max): [7.87014085e+00 3.00000000e-02 1.48483732e+01 1.55500000e+00
 9.07100000e+01]
Feature: GHG Emissions (kg CO2eq, IPCC2013 incl CC feedbacks)
Statistics (mean, min, std, median, max): [2.34535211 0.09 4.38217173 0.735 28.96]
Feature: Water Use (L)
Statistics (mean, min, std, median, max): [250.61485915 0. 606.03236912 67.115 4542.28]

The Food Group: fruit
Feature: Land Use (m2*year)
Statistics (mean, min, std, median, max): [0.99834783 0.19 1.96513742 0.53 17.64]
Feature: Biodiversity (sp*yr*10^14)
Statistics (mean, min, std, median, max): [6.38032174e+01 9.70000000e-01 2.48483958e+02 1.40200000e+01
 1.76408000e+03]
Feature: GHG Emissions (kg CO2eq, IPCC2013 incl CC feedbacks)
Statistics (mean, min, std, median, max): [0.72278261 0.03 1.09974586 0.47 8.41]
Feature: Water Use (L)
Statistics (mean, min, std, median, max): [196.648 0. 265.70339749 112.25 1540.24]

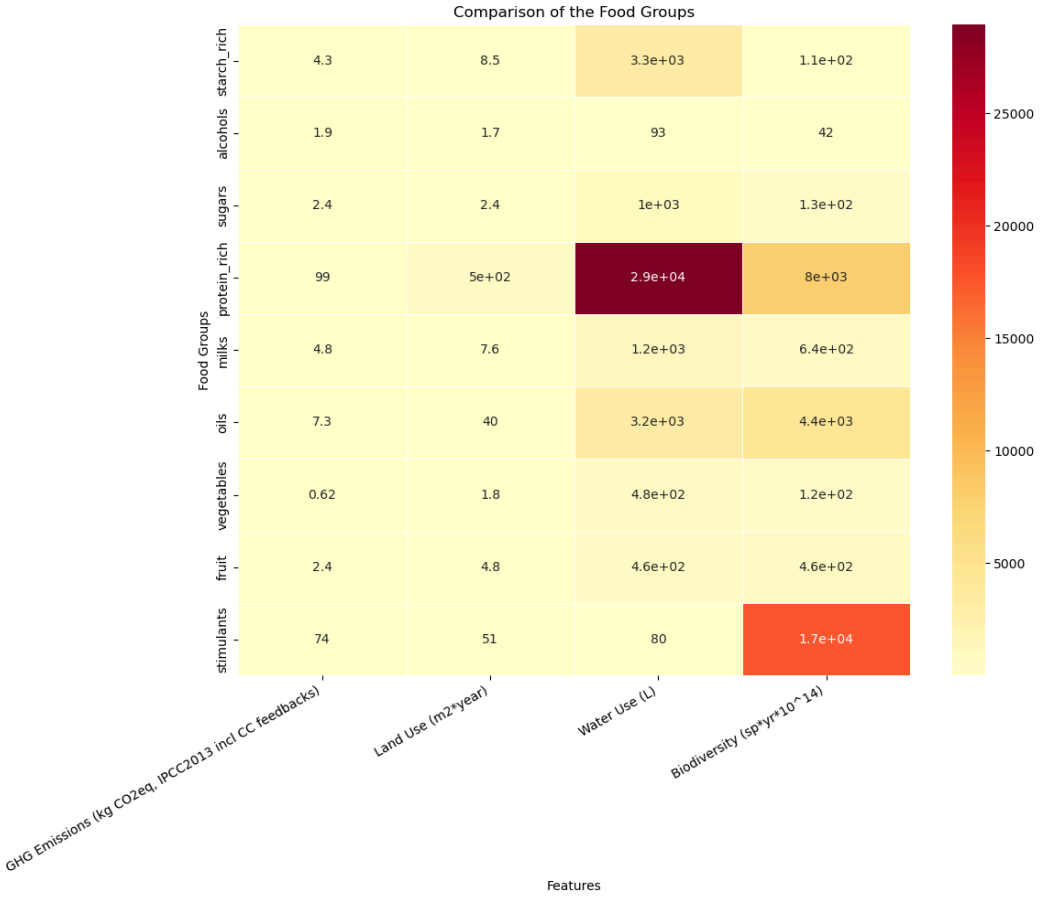
The Food Group: stimulants
Feature: Land Use (m2*year)
Statistics (mean, min, std, median, max): [42.26142857 8.4 35.61490783 22.53 130.2]
Feature: Biodiversity (sp*yr*10^14)
Statistics (mean, min, std, median, max): [7097.81666667 276.64 13704.01147449 2264.42
 60168.59]
Feature: GHG Emissions (kg CO2eq, IPCC2013 incl CC feedbacks)
Statistics (mean, min, std, median, max): [32.48714286 -3.66 59.5669011 7.37 257.95]

Feature: Water Use (L)
Statistics (mean, min, std, median, max): [36.5552381 7.56 58.46364402 10.41 265.8]

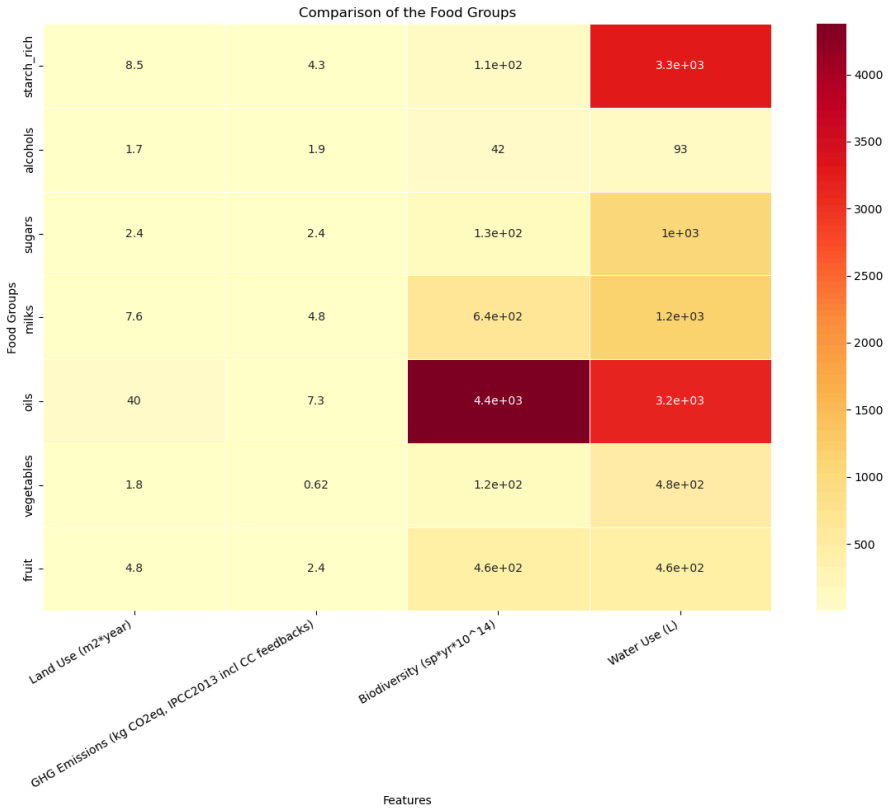
The result of the clustering:
Food Group starch_rich belongs to Cluster 1
Food Group alcohols belongs to Cluster 1
Food Group sugars belongs to Cluster 1
Food Group protein_rich belongs to Cluster 1
Food Group milks belongs to Cluster 1
Food Group oils belongs to Cluster 1
Food Group vegetables belongs to Cluster 1
Food Group fruit belongs to Cluster 1
Food Group stimulants belongs to Cluster 1

(have tried several different numbers of the clusters, changed the indicators, etc., and the results were the same. Also, the values were normalised for avoiding bias as much as possible.)

Furthermore, if to plot the statistical information into heatmap, it displays as this:



Generally, the protein rich and stimulants are the most two food groups which have a big impacts to the environment, and the protein rich, oils and starch rich food groups consume a lot of water during the production process; If to remove the two food groups protein rich and the stimulants, the heatmap may like this:

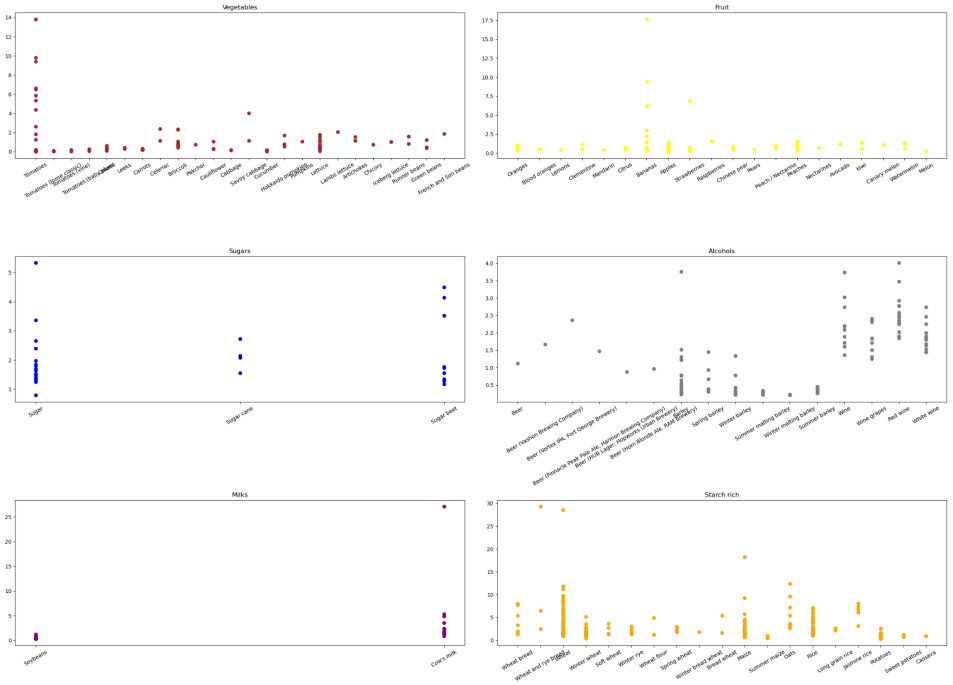


This gives the insight that the oils and the starch rich food groups have bigger environmental impacts than others. Therefore, this analysis report will focus on the fruit, vegetables, milks, sugars, alcohols food group, plus one more which is the starch rich food group, as it is a quite important material in the bread / baker industry, and will dig deeper into them to obtain the details about which resource may be better, which may have bigger impacts on the environment.

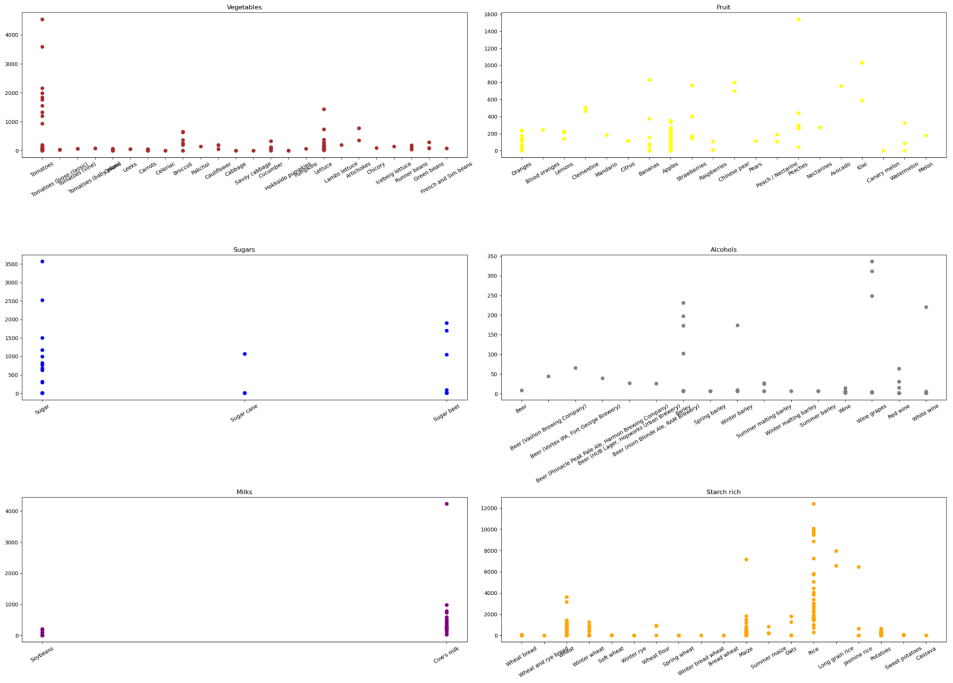
Part 2

Apart from that, each of the four main indicators was investigated and visualised. Additionally, the indicator value of each food material within its food group was shown respectively.

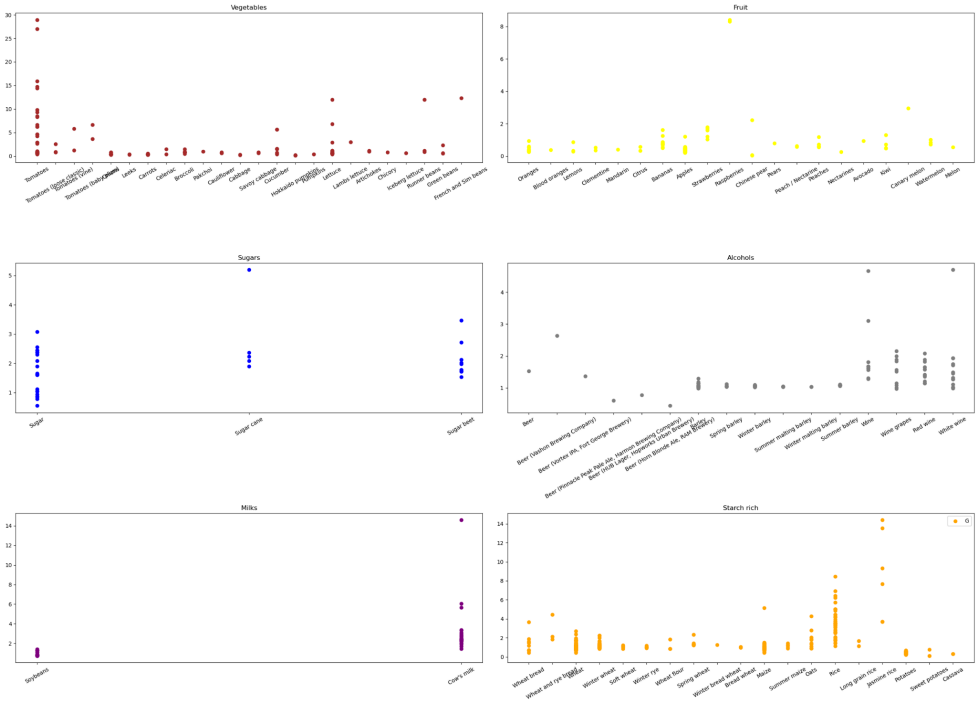
Land use:



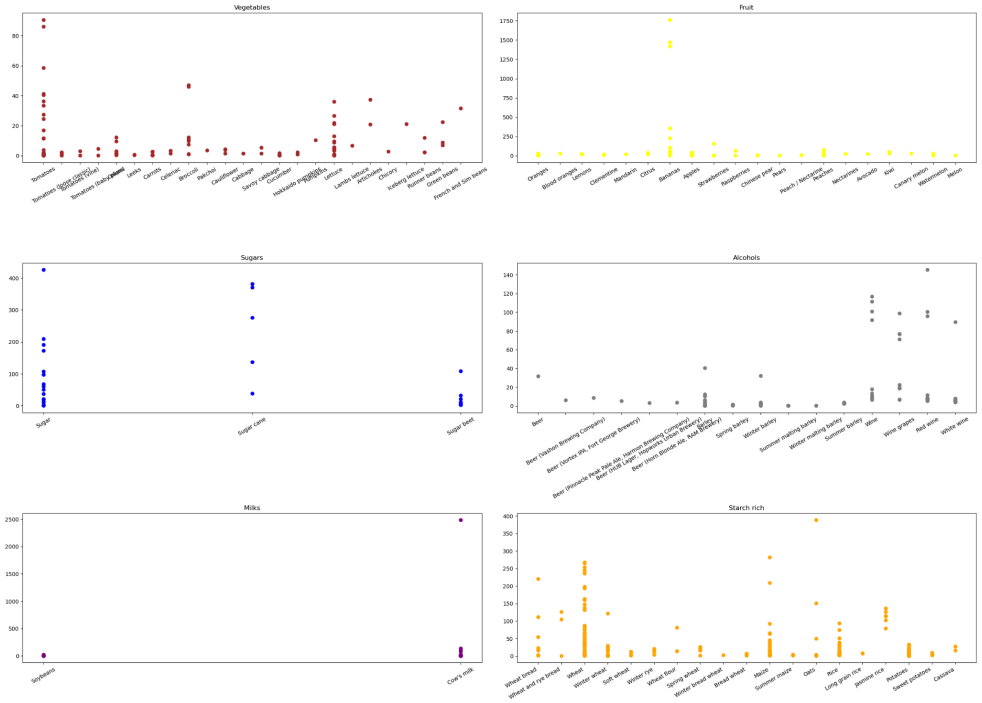
Water use:



GHG emissions:



Biodiversity:



These plots have given the insights that in terms of the Land use, Sugars and Alcohols have the least land use, and vegetables, fruit are the friendly resources as well, except for tomatoes and bananas though, under rigorous standards, they are a bit harmful compare to the others among the two categorise; And then, most milks products are quite good, the same as the starch rich, but not all of them have the same level of land using, so have to select the certain better ones which have lower land use and probably with better performance regarding other sustainable indicators, like water use, GHG Emissions, generation country (related to supply chain and carbon footprint), etc.

Regarding Water use, starch rich food consumes quite a lot of water, specifically the rice can consume water hugely, and then the vegetables, sugars, milks consume less water, after these, there is fruit, and the alcohols consume least water.

And if talk about the GHG emissions, surprisingly the tomatoes in the vegetables group have the most emissions, and some other vegetables as well, like lettuce, and some food from the starch rich category like rice, wheat actually have lots of emissions as well, but the food like wheat flour, oats, maize are much better. Lastly, the fruit, sugars, and alcohols have the least emissions.

At last, some bananas have the most biodiversity impact, and then the wheat, sugar and wine.

At the end

Moreover, in some conditions, it's crucial to consider the country or regional origin of the food, as it's related to the supply chain logistics. The transportation distance required to move foods from the original production point to the consumers can make significant environmental impacts and emissions overall. So it is better to have the locally-sourced foods if possible, and to raise the local community engagement (like improve the local farming condition, raise the number of local gardens, etc.), develop the innovative cost-saving technology and improve with the proper considerations on marketing and making the products as much sustainable as possible.