**DBSCAN 16 CLUSTERS ANALYSIS / REGRESSION ANALYSIS**

Cluster 0 holds wheat data from 126 countries

Cluster 1 holds rice data from 113 countries

Cluster 2 holds barley data from 77 countries

Cluster 3 holds corn data from 119 countries

Cluster 4 hold rye data from 32 countries

Cluster 5 holds oats data from 43 countries

Cluster 6 holds millet data from 47 countries

Cluster 7 holds sorghum data from 73 countries

Cluster 8 holds data from 42 countries that didn’t report agricultural data

Cluster 9 holds mixed grain(The categorical attribute mixed grain) data from 6 countries

Cluster 10 holds rice data from China – More Production than Cluster 1

Cluster 11 holds corn data from China – More Production and CO2 Emissions than China in Cluster 3

Cluster 12 holds wheat data from United States

Cluster 13 holds rice data from United States – Data with higher CO2 Emissions

Cluster 14 holds corn data from United States

Cluster 15 holds rye data from United States, and 1 year of rye data from China

**OBSERVATIONS ABOUT GRAIN PRODUCTION IN RELATION TO TOTAL DATA:**

Corn Production Values range 100% of entire data

Wheat and Rice Production Values range approximately 99.8% of the entire data

Barley and Sorghum Production Values range approximately 99% of the entire data

Rye, Oats, and Millet Production Values range 98% of entire data

Mixed Grains Production Values range 80% of entire data

**This simply suggests that in our dataset, Corn is the grain with the highest production value, it doesn’t necessarily mean it’s the most popular grain in our data.**

**OBSERVATIONS ABOUT RELATIONS WITHIN CLUSTERS OR WITHIN DATA**

**Production and Exports**

Not a very strong correlation between Wheat Production and Exports. Cluster 0 shows data that splits, where some countries have exports increase as production increases, and other countries don’t export Wheat as much even as production increases.

Rice falls under a similar case as wheat, where some countries drastically increase exports when production increases, and other countries may have high production but low exports in comparison.

Barley falls under a similar case as Corn, where Barley production increases alongside Exports.

Strong positive correlation between Corn Production and Exports – Shows that Corn is a popular grain that’s exported.

Rye is more scattered, with a portion of the countries producing Rye and not exporting it. Other countries seem to have a positive correlation between Rye production and exports

Oats falls under a similar case like Rye

There is no real correlation between Millet production and exports, since most countries seem to produce Millet, but not export it. This could mean that Millet is not a grain that’s high in demand compared to other grains like Wheat and Corn.

Sorghum production and exports seems to be split, where a portion of countries grow the grain but do not export it, and another portion of countries exports more Sorghum as production increases.

When it comes to mixed grains, there are no exports in this category of grain, possibly because of how mixed grain consists of grains that are already included in this data, so exports for mixed grain could be split into its respective grains. (Wheat and Sorghum together is mixed grain, but if exported, Wheat goes to wheat exports, Sorghum goes to sorghum exports).

China produces more rice than any of the countries that reported rice data in Cluster 0, and export more than the average amount of rice compared to those countries. This shows the contribution that they have made to the global food supply with regards to rice alone, rivaled by India, Thailand, and at one point, Pakistan

China is one of the top producers and exporters of Corn.

The United States is one of the top producers and exporters of Wheat

The United States tends to export more rice as rice production increases. With that said, while the United States is not a major producer of rice, compared with the likes of China and India. With that said however, The United States is a major exporter of Rice, alongside Thailand, India, China, and Vietnam.

The United States is the biggest producer and exporter of Corn in this dataset, and in general, the major outlier in this dataset, which is why there is such a wide range of production values.

The United States is one of the smallest producers and exporters of Rye, and China only reported exporting Rye once in 2001.

**Production and Imports**

One observation regarding the relationship between Production and Imports throughout the clusters, is that there seems to be a 3-way split within the clusters. You’ll have data where there are low production values and high import values, which suggest countries rely on importing for their grain supply heavily. Data where there are high production values and high import values suggests that countries may import grains even when producing them, as a means of either supplementing their production, or stockpiling for the future. Data where there are high production values and low import values suggests that countries may not rely on imports for their food supply.

Oats is a specific grain where a good majority of it is not imported a lot in general, despite low or high production values.

Millet isn’t a grain that’s highly imported by many countries, despite its production

Sorghum follows a similar trend to rice, corn wheat, but in a more extreme case, where there are countries that either do not produce, and only import this grain, or there are countries that only produce this grain and do not import at all, with a small mix of countries both importing and producing.

Mixed Grains are either produced but not imported, or vice versa. Considering the nature of this attribute, not much can be determined from this data, especially considering it’s a small minority compared to the individual grains.

China is a major importer of rice, which is understandable, considering how much they contribute to rice exports, Just because a country may be a top producer, doesn’t mean they won’t need to import, situations such as trade and consumption come into consideration when it comes to a country’s food supply, and whether they need to import more food.

China is a major importer of Corn

The United States is a major importer of Wheat and Rice.

**GDP and Production**

Throughout the clusters there seems to be a positive relation between GDP and Grain Production, but it’s difficult to really assess the relationship because while Production is going to be impacted by GDP, there are other factors that impact whether production of a certain grain will even occur. This can be due to environmental conditions, where certain crops won’t be grown there.

**Area Harvested and Production**

Throughout almost every cluster there was a strong positive relation between the Area Harvested, and Grain Production. This shows that countries that have more harvestable land for crops and grains, the more grains they’ll be capable, and probably grow. The countries with some of the highest amounts of harvested area is the U.S, China, India, and Russia, which are some of the biggest countries when it comes to exports of certain grains like Corn, Wheat, Rice, which are the crops that are most imported, exported, and produced.

**Area Harvested and Exports**

A bit of a split, but data seems to suggest that as Area Harvested increases, Exports increases. Considering that there is a strong positive relation between Area Harvested and Production, this relationship makes sense. If a country can harvest more area for growing grains, chances are those countries are going to have higher production, which means those countries are going to probably be exporting more. For wheat grains, this mostly applies, but there are a few countries that don’t export grains despite how high area harvested is.

For rice, there is a split, where there are countries that have low area harvested but high exports, and there are countries with high area harvested and exports, along with countries with low exports and high area harvested.

Barley follows a similar trend to Wheat in terms of the relationship between Area Harvested and Exports.

A lot of countries that grow corn as a crop seem to not export it as much, regardless of the area harvested. There’s a small cluster of countries that contribute quite a bit to corn exports, despite the low area harvested. However, the clear outlier in this cluster is the United States, which has harvested more land than most countries in this cluster, and exports more corn than most, if not all the countries in this cluster.

A large chunk of the rye exports relates to low area harvested, with some countries either not exporting rye, or some countries having higher area harvested as well as higher rye exports.

Oats follows a similar case to rye, where high oats exports relate to low area harvested. There are also countries that do not export oats, despite area harvested. Some of the countries do export more as area harvested, but not as much as some of the countries that have low area harvested.

Millet is not a highly exported grain, considering that the only countries that export Millet in the cluster are Argentina, Australia, China, India, Nigeria, Sudan, and Tanzania, which is about 1/7th of the countries that have reported Millet agricultural data. Compared to other grains, Millet is not a major player in the global food cycle, but was imported in the past by Chad, Japan, Mali, Senegal, and Taiwan.

Sorghum is split in 2 ways, with a mix of countries with increasing harvested area, and exports, and another mix of countries that barely, if at all, participate in the export of Sorghum, despite Area Harvested increasing.

China is among one of the countries that harvest the most land but is mixed with regards to their level of corn exports, where some years they’ll export a large amount of corn, and other years they’ll barely export any in comparison. However, China remains one of the top exporters of Corn.

The United States seems to follow a positive trend between Area Harvested and Wheat/Rice/Corn Exports, where exports tend to increase as Area Harvested increases.

**GDP and Area Harvested**

There isn’t a clear relationship between GDP and Area Harvested throughout the clusters, which makes sense because the area harvested for grains will depend heavily on the amount of land that is available in the first place, regardless of GDP. If the land isn’t there to be harvested, it doesn’t really matter how much money you have, you aren’t going to be able to harvest that land, unless you are able to spend that money to make the land harvestable.

**GDP and Imports**

With regards to GDP and Imports, there isn’t a strong relationship that exists between them. While trade and GDP are heavily connected, it also comes down to the need that a country must import certain grains. There are countries with high GDPs that do not have high imports, one reason being that they could use that money to grow crops themselves. There are other countries where imports are high and GDP can remain low, which can be due to a mix of factors. Countries that fall under this category might not have the resources to grow their own grains in a quantity that is sufficient for the population, so they import grains from other countries to supplement their food supply.

**GDP and Exports**

The distribution of GDP and Exports seems to match closely to GDP and Imports’ relationship. There are countries that will not participate in Exports, despite their GDP, and there are countries that tend to participate in more exports, the higher their GDP is. As a result of this, we can’ necessarily use GDP as a marker to determine whether a country will export more or import more, there are more important factors that determine whether a country will import or export grains.

**Total CO2 and Production**

With regards to Wheat, as there is a positive relationship between Total CO2 and Production of Wheat.

For countries with rice production there is a two-way split. There are countries that display a positive relationship between co2 emissions and rice production, while other countries that simply are bigger producers of co2 emissions, do not export as much rice, such as the United States and Russia.

For countries with barley production there is a positive relation between co2 emissions and barley production

There is a stronger positive relation between co2 emissions and corn production, with some countries not producing high amounts of corn despite co2 emissions, and other countries producing more corn even with lower co2 emissions.

**Year and Yearly AVG Temps**

In China, from 1990 to 2012, as the year progressed, the average temperature for the year increased slowly.

In the United States as well, as the year progresses, the average temperature for the year slowly increased as well, according to the line of best fit

**Year and Total CO2**

In China, from 1990 to 2012, as time progressed, CO2 emissions slowly increased over time.

In the United States, CO2 emissions generally increased from 1990 to 2007, and then started decreasing towards 2012.

**Milling Rate and Rough Production**

For Rice Data, Countries that are incapable of milling threshed rice, do not produce threshed rice, they only interact with rice through exports of broken rice, and imports of rice in general.

**Issues regarding regression analysis of the clusters/data**

Regarding the idea that increasing co2 emissions is a major factor in rising global temperatures, this isn’t something that can be analyzed throughout the entire dataset. Due to the wide range of countries and temperatures over the past 60 years, we aren’t going to be able to get anything conclusive regarding the relation between temperatures and co2 emissions. When we look at the clusters for individual countries, then we get a different perspective.

**General Conclusions based off Cluster and Data Observations:**

Rye, Oats, and Millet are the least popular grains in terms of data reported by countries, each grain being reported by less than 50 countries out of the 100+ countries.

Wheat, Rice, and Corn are the most popular grains in terms of data reported, and production overall.

Barley falls along the middle, more towards Rye, Oats and Millet

For Rice Data, Countries that are incapable of milling threshed rice, do not produce threshed rice, they only interact with rice through exports of broken rice, and imports of rice in general. As a result of this, there is higher dependence of the countries that are capable of milling rice, because countries that do not have those capabilities simply do not bother producing threshed rice in any noticeable quantities, and they will be heavily reliant on other countries to get a supply of milled rice. This also means that countries with milling capabilities are the ones involved in the trade of milled rice.

Also, exports of rice mainly consist of milled rice, and broken rice is only a minor portion of rice exports.

Countries that have not harvested any area, does not produce any grains.

No consumption noted for rice data