**HDSC August ’22 Stage D Project Documentation: Data storytelling using Pattern recognition.**

By Sakshi Jha (Data Storytelling track)

****

**Introduction:**

Food shortage is a serious problem facing the world and is prevalent in sub-Saharan Africa. There are a number of social factors causing food shortages.

It has far reached long and short-term negative impacts which include starvation, malnutrition, increased mortality and political unrest. There is a need to collectively address the issue of food insecurity and look at trends of supply and production to avoid over-production and wastage and actually address the people in need.

**Aim of this project**: To visualize the given datasets, gather insights and pattern and construct a story on how the world food shortage problem can be solved using these patterns and insights from the data.

**Given**: under this project two datasets have been provided:

* Africa Food Production (2004 - 2013)
* Africa Food Supply (2004 - 2013)

**Dataset description:**

The datasets are African food production dataset from 2004 to 2013 and African food supply dataset from 2004 to 2013; used for this project include information about the production and supply data in various countries of Africa from 2004 to 2013.

The food production dataset contains 4 columns namely; country, item (item produced), year and value and 23110 rows. The unit of the food production value is in Kilotons (kt).

The food supply dataset contains 3 columns namely; country, year, value and 450 rows. The unit of the food supply value is in kcal/person/day.

Ultimately, both datasets contain data for 45 African countries.

**Tasks to be performed:**

Visualisations and narratives should include;

* Trends in food production and consumption over the years.
* average and median food production, the outliers, quartiles and interquartile ranges (boxplot).
* A comparison between average food consumption and production for each year.

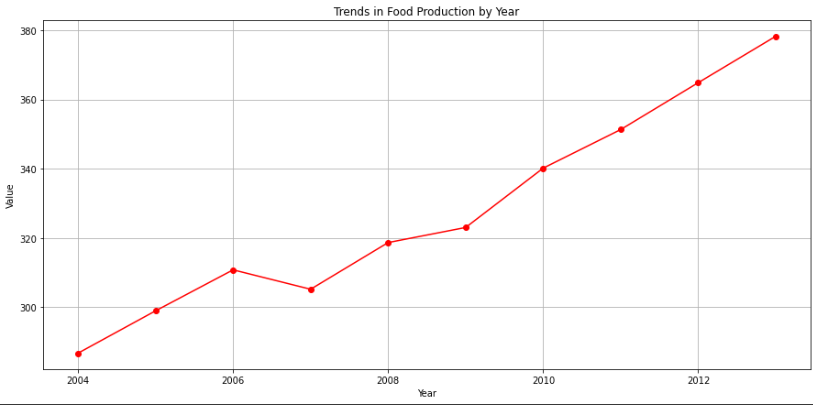
**Libraries used for Pattern Recognition:**

As we know, Python is used for various applications and there are different libraries for different purposes. In my further demonstration, the following libraries have been used:

* **Pandas**: Pandas is a library used for data manipulation and analysis. It is used to extract the data and store it in the desired format.
* Use **matplotlib** and **seaborn** for data visualization

**Trends in Food Production and Consumption in various countries of Africa from 2004 to 2013.**

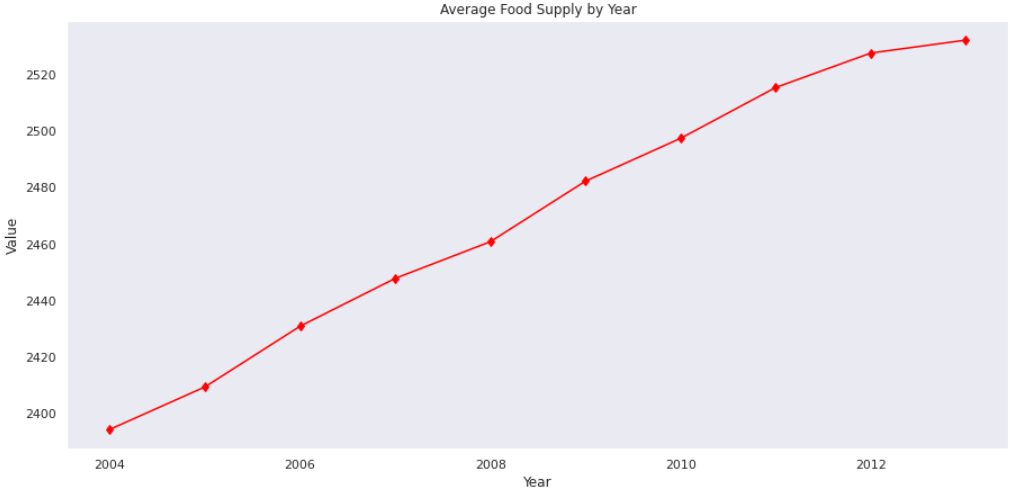
1. **Food production**

****

Food production in Africa increased significantly from 2004 to 2006 until the quantity of food hit rock bottom in 2007 causing the global food crisis that lasted till 2008. (a)

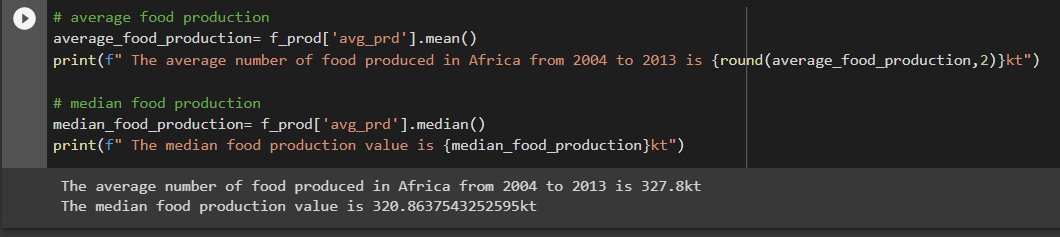
As a result, there was an increase in food prices and cost of food production in 2007. Post this, awareness was raised to suffering countries to produce more food and ensure food security. This explains the increasing trends in food production from 2009 till 2013.

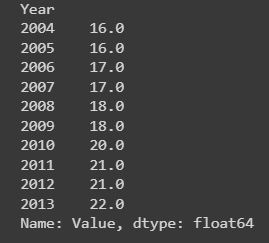
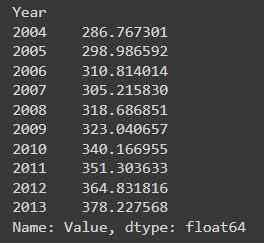
1. **Food consumption**

****

Food supply has increased exponentially over the years due to various reasons like increase in population, increased demands for supply due to food security etc.

**Average and Median Food Production**



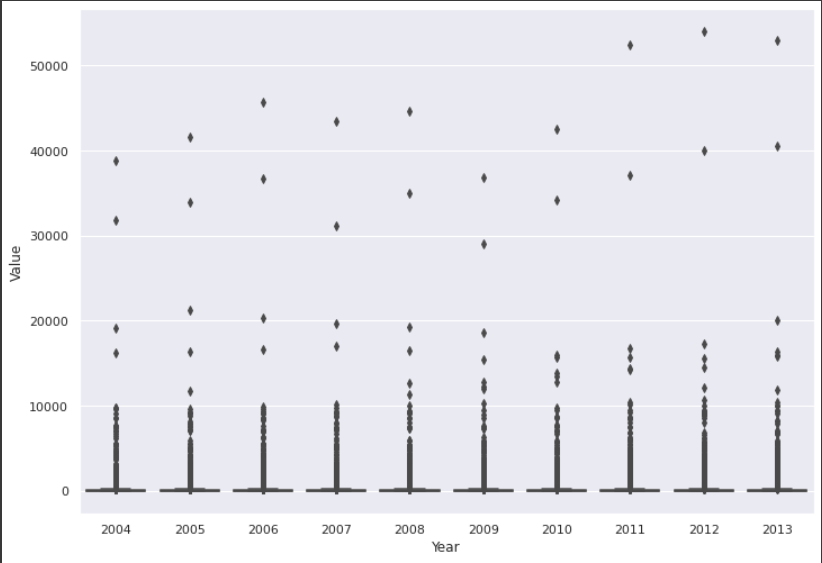


Hence the average food produced is 327.8kt and median food production value is 320.86kt (2004-2013).

**Outliers’ investigation**

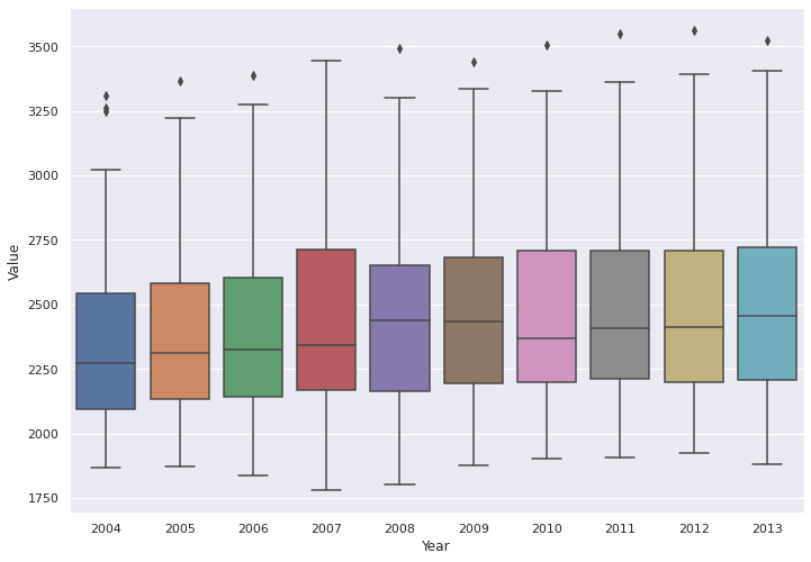
Boxplot analysis gave away extreme values occurrences in both the datasets as depicted below.

Outliers in production data

****

Nigeria was seen to produce as much as 54000kt of cassava and its products thereby showcasing Nigeria as the highest food producing country in Africa. (Median- 18kt)

Outliers in food consumption data

****

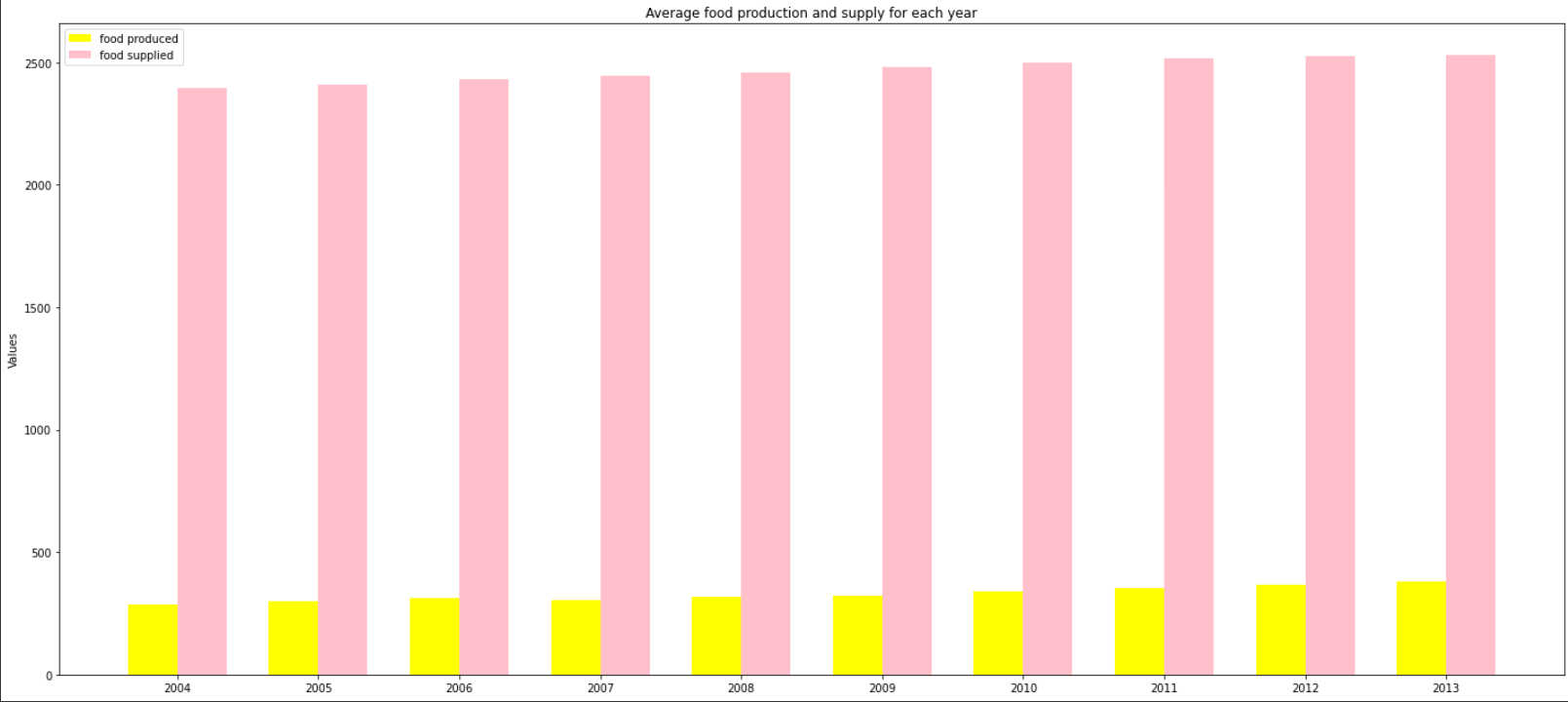
Trends in the supply data reveal Egypt as the country with highest food consumption from 2004 to 2013.

**Comparison between average food consumption and production for each year**

It can be seen from the graph below that food production rate in Africa is much higher than the food consumption rates in all its countries:

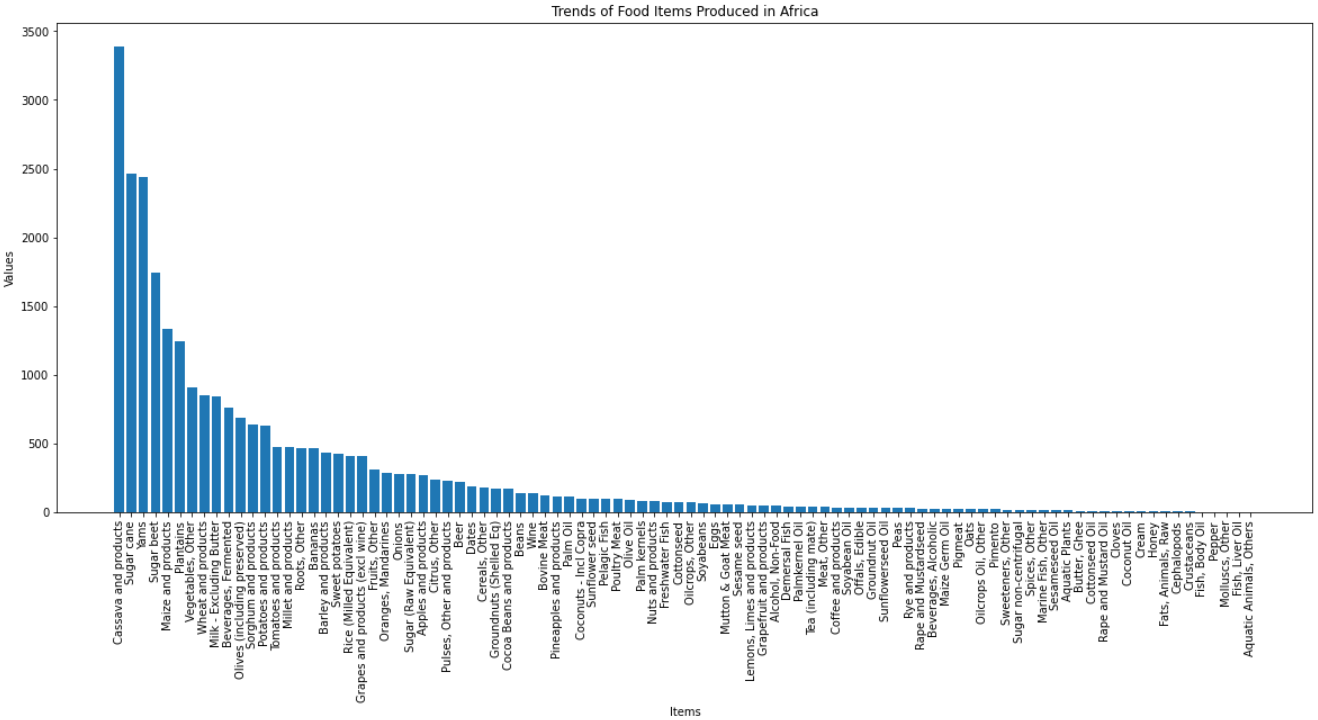
This is also resulting in food wastage in high supply area and indirect food shortage in low supply areas.

Continuous food overproduction trends don’t deny the possibility of another wave of food scarcity in few years.

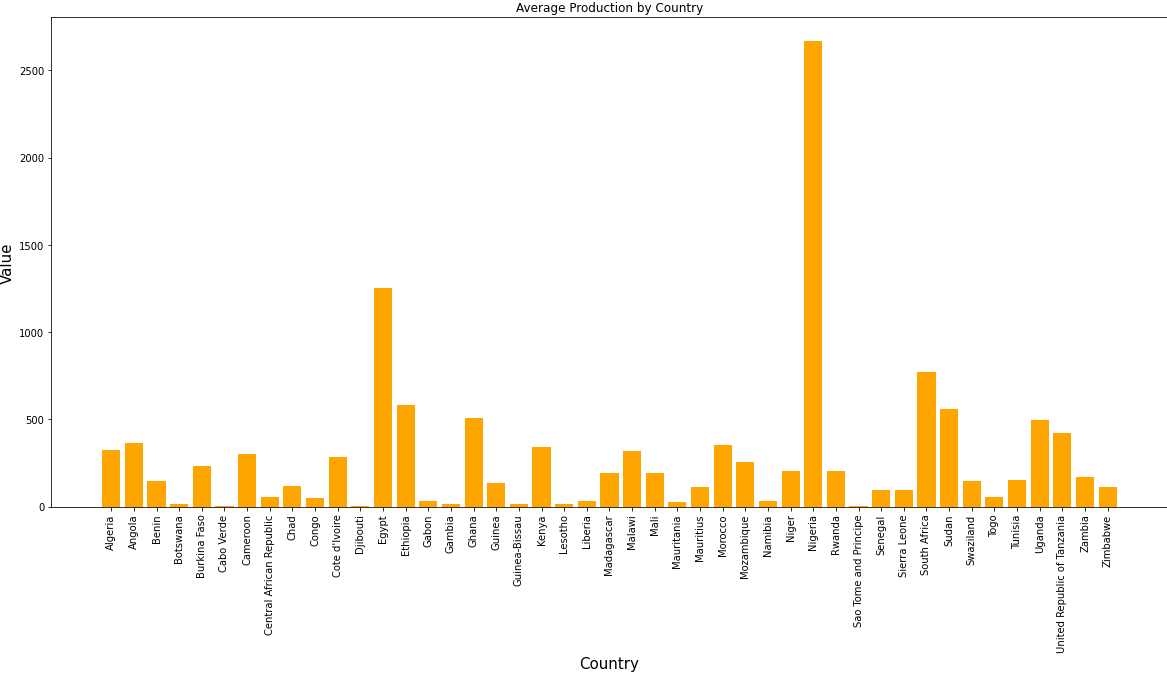
****

**Food Items Produced in Africa**

The top 3 food items produced in Africa are: Cassava and products, Sugar cane, Yams. Least are Fats and marine animals and raw products are amongst least produced certain quantities are negligible

****

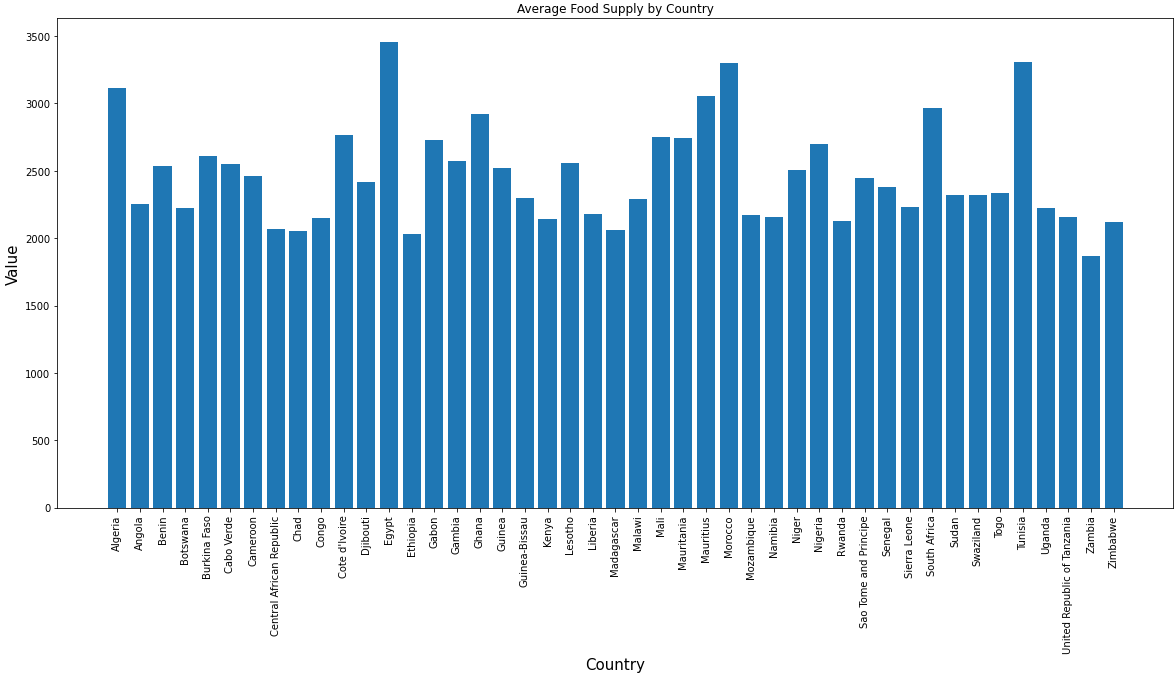
**Average Food Production by Country**



It is observed from the plot that Egypt, Ethiopia, Ghana, South Africa and Sudan are only countries with production values >500.

Evidently; Nigeria, Egypt and South Africa are amongst the top food producing countries in Africa.

**Average Food Consumption by Country**



Egypt, Morocco and Tunisia are countries that supply the most food.

**Conclusion**

From the trends, it was observed that only a few countries in Africa produce food in large quantities but there is the rising issue of over production and food wastage.

Whereas in other parts there is a general food demand and supply which is more or less in balance.

It is necessary for the government to look into the trends and manage balance the ratios of production and supply, as the population increase, there will be more demands for food globally.

Therefore, in order to solve the world's food shortage problem, there should be increased food production across countries of Africa and there should be a balance in food distribution and supply with more focus towards the weaker sections.