Detailed Report on Exploratory Analysis of African Agriculture (2004 - 2013)

https://www.africanexponent.com

Introduction

Agriculture is Africa's most important economic activity, employing about two-thirds of its working population. To gain insights into Africa's agricultural activities from 2004 to 2013, an exploratory data analysis was conducted. This report presents the findings of the investigation, which involved cleaning, describing, analyzing, and visualizing the data on food production and supply in Africa.

Methodology

The analysis was performed using Python programming language and various libraries and modules, including pandas, matplotlib, seaborn, numpy, datetime, geopandas, plotly express, and dataframe_image. The data was loaded from two CSV files: Africa Food Supply (2004 - 2013).csv and Africa Food Production (2004 - 2013).csv.

During my analysis, I discovered that I needed more data to convert kcal/person/day to kilotons visualize data and also test my hypothesis on the increase in rice production with an increase in population; as a result, I researched and loaded two extra datasets: population.csv and countries.geojson.

Data Exploration

The analysis began with exploring the data. The food supply data was examined first using the head() and info() functions. Similarly, the food production data was explored using the same functions. The describe() function was used to get statistical summaries of all datasets;

these include minimum, maximum, count, mean, median, standard deviation, etc of numerical and non-numerical values.

Data Transformation

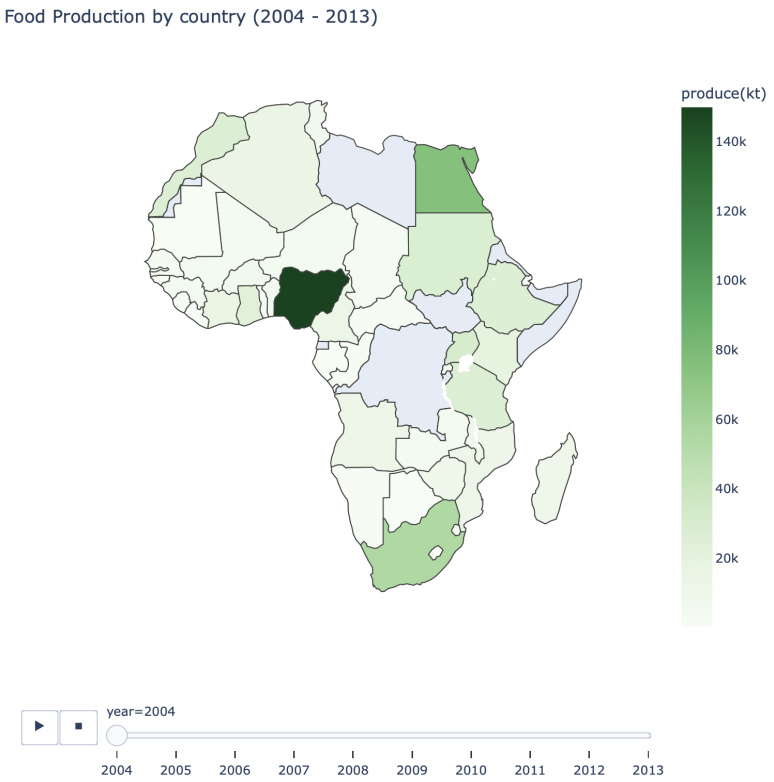
The data was transformed to prepare it for analysis; some common operations include renaming, subsetting, filtering, sorting, grouping, reshaping, and merging data. In the food production data, the column names were converted to lowercase, the "year" column was converted to a datetime format, and the year values were extracted and formatted as strings. Additionally, the column name "value" in the food production dataset was renamed to "produce(kt)" to provide a more precise representation of the data. The data was then grouped by country and year, and the production values were summed up for each year. Similar operations were performed on the food supply data.

The population data were subset to only contain data from 2004 - 2013. Furthermore, it was transposed from wide to long data before being merged with the food supply data to convert the supply unit from kcal/person/day to kilotons(kt).

The analysis also involved visualizing the food production and supply data on a map. The Geopandas library was used to read a GeoJSON file containing country boundaries, and the food production data was merged with this geospatial data. The resulting data was then used to create a choropleth map showing the total quantity of food produced by each country from 2004 to 2013. The map revealed that Nigeria, Egypt, and South Africa were the largest food producers, while the least-producing countries were mainly located in Western and Southern Africa.

Insights into Food Production

The analysis provided insights into the top food producers in Africa. The map revealed that Nigeria, Egypt, and South Africa were the largest food producers, while the least-producing countries were mainly located in Western and Southern Africa. Nigeria remained the number one producer every year.



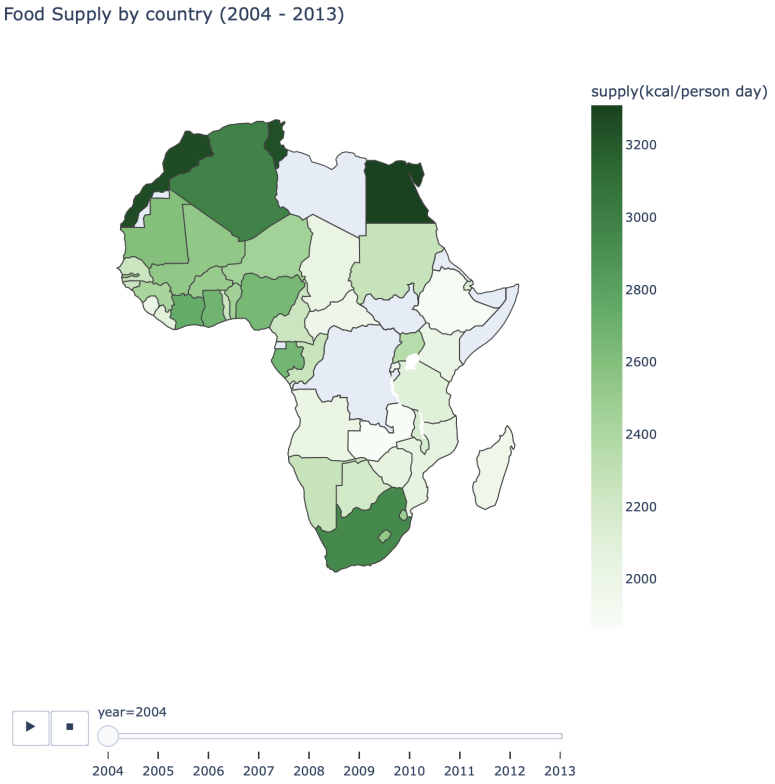
All interactive features may be accessed in the project's notebook

Cassava and products, Yams, Beverages, Vegetables, Sorghum, and products were the 5 most produced items in Nigeria.

Insights into Food Supply

To compare food production with food supply, the food supply data was transformed similarly to the food production data. The data was then merged with the geospatial data to create a choropleth map showing the food supply by country from 2004 to 2013.

The map indicated that Egypt, Morocco, and Tunisia had the highest food supply per capita, despite not being the top food producers. Egypt remained the number one producer for 10 years.

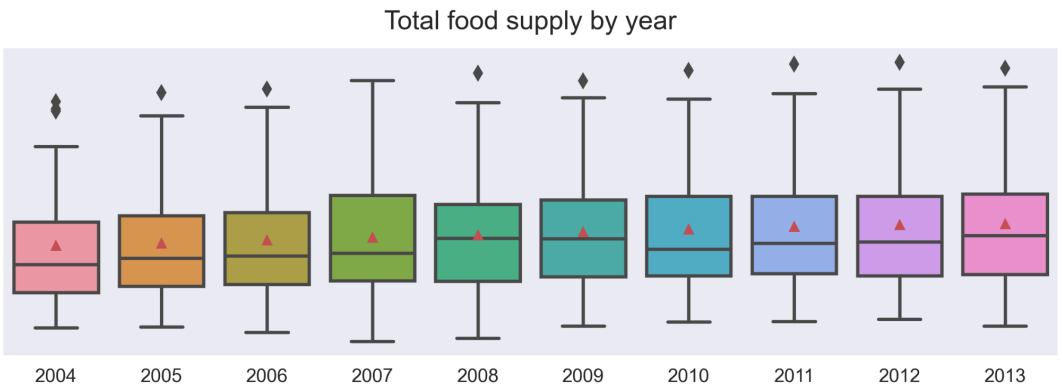


All interactive features may be accessed in the project's notebook

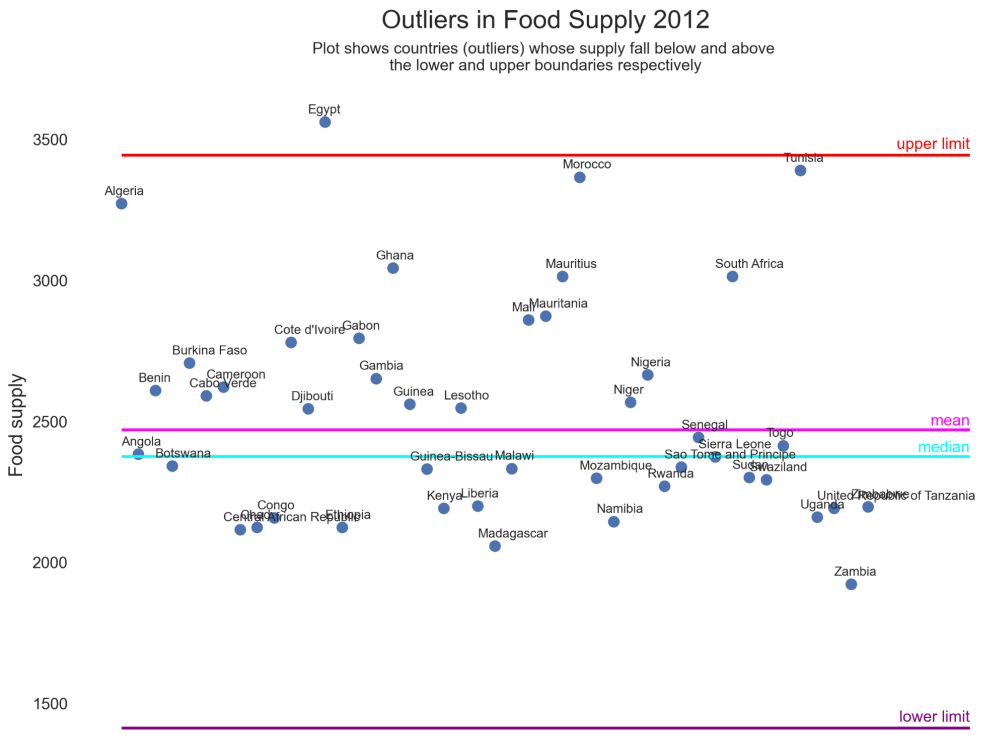
Sugar cane, Tomatoes and products, Vegetables, Wheat and products, and Maize and products are the top 5 most produced items in Egypt within the given time frame.

Outliers in Food Supply

The analysis also identified outliers in the food supply data. Box plots were used to visualize the distribution of food supply values over the years.



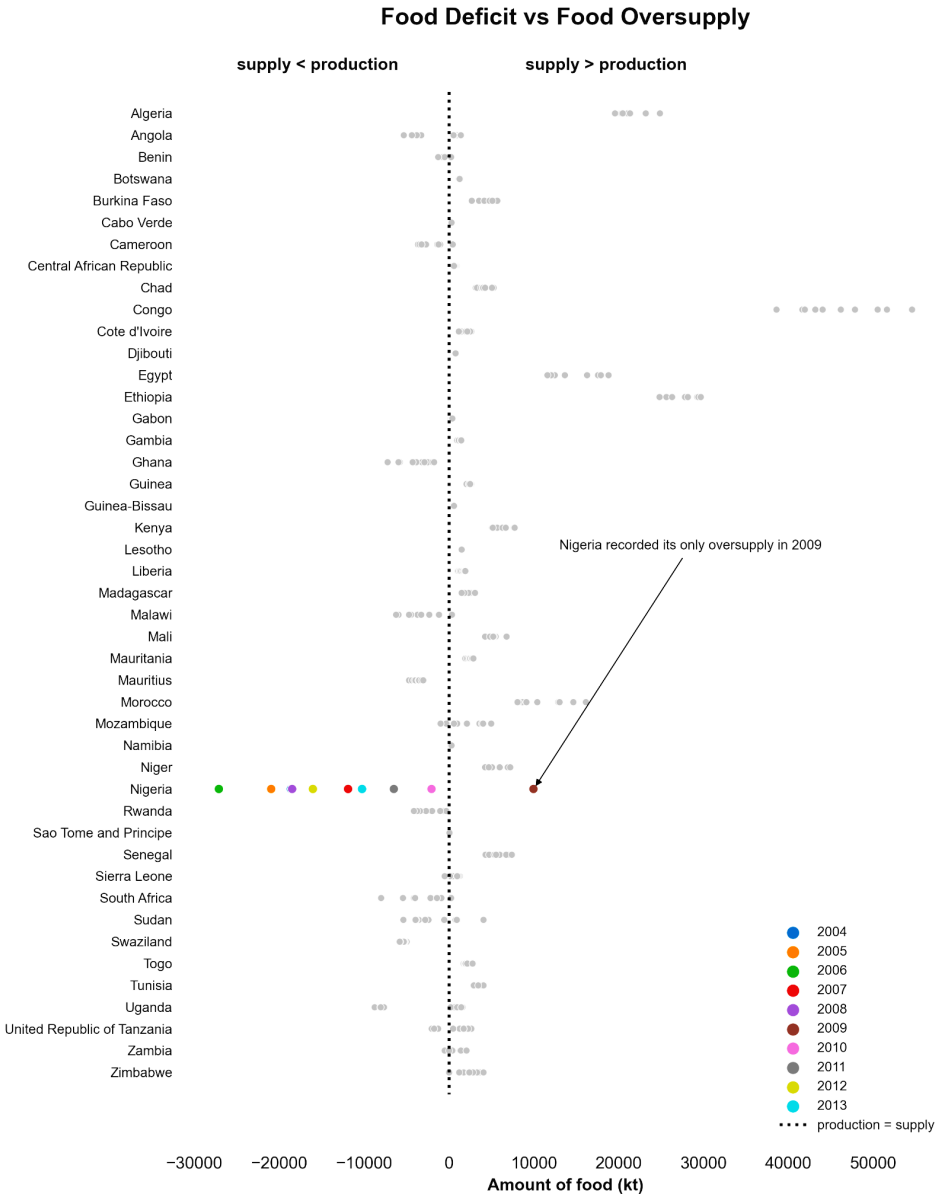
Egypt was identified as the only outlier in the data, as it recorded the highest food supply in 2012. A function was created to visualize the outliers, mean, and median values for different years, providing a clearer understanding of the data.



Food Deficit and Oversupply

To compare food production and supply, the food supply data was converted from kcal/person/day to kilotons. The average food production and supply values were calculated by country, providing insights into the countries with the highest production and supply. However, during the analysis, it was discovered that five countries were missing from the population dataset, which caused a mismatch in the data. The missing countries were identified and their names were synchronized to avoid data loss during the merge process.

Despite recording the highest amount of produce each year, Nigeria records the highest food deficit, having its only oversupply in 2009.



Test of hypothesis

Rice is a carbohydrate food eaten in most parts of the world. One could hypothetically say rice feeds the world. Rice was the third most-produced food worldwide in 2014, according to the FAO. However, I was curious to know if there happens to be a relationship between rice production in Africa, and its population. I carried out this test using the Pearson Correlation Test.

Null hypothesis:
There will be no significant correlation between the mean amount of rice produced in a year (X) in African countries and the mean population of African countries in that year (Y).

Alternative Hypothesis:
There will be a significant correlation between the mean amount of rice produced in a year (X) in African countries and the mean population of African countries in that year (Y).

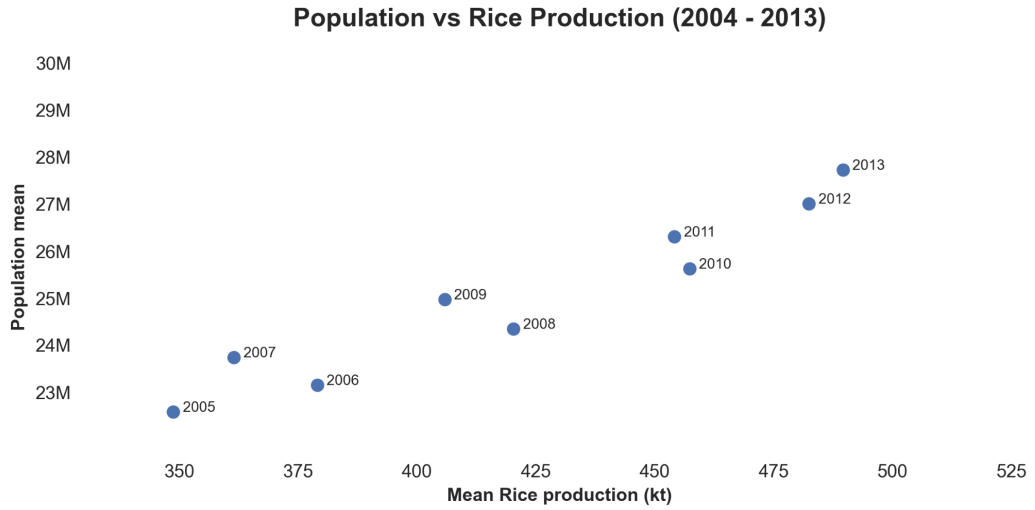
X = The mean amount of rice produced in a year (in kt)
Y = the mean population of African countries in that year

From my analysis, I obtained the following results:

- *A correlation coefficient of 0.97 indicates there is a strong correlation between the mean rice production in a year and the mean population of African countries for that year.*
- *The probability value (which is less than zero) indicates there is no likelihood for the null hypothesis to be true.*

It's often erratic to simply accept the values obtained from the Pearson correlation test without visualizing the data, given that this correlation test assumes that the relationship is linear, which is not always the case.

I created a scatterplot to visualize the relationship of my data points, to help me validate my chosen correlation test. Below is the result I obtained:



My chart showed a linear relationship which validates the results obtained from my correlation test.

Test Conclusion:

The null hypothesis that the two variables are independent is rejected based on the Pearson Correlation Test results.

NB: It is important to note that correlation does not indicate causation.

Conclusion

- Nigeria was the leading producer of agricultural products in Africa between 2004 and 2013.
- Nigeria has the biggest deficit in terms of production and supply compared to other countries.
- Egypt had the highest food supply per capita each year for the period.
- Generally, bar population size, Congo, Ethiopia, and Algeria had the highest supply.
- North African countries had the highest ratio of supply-to-product.
- West African countries had the least ratio of supply-to-product.

There's more to be discovered from this data; the relationship between countries in terms of common products for example. I wish to provide further analysis on that after I acquire adequate knowledge on how to create compelling visuals to support such findings.

Thank you for reading.