DATA STORYTELLING IN PATTERN RECOGNITION

Introduction

Story of how the world food shortage problem can be solved using appropriate visuals makers to show relationship between Food producing and Food Consumption.

Overview

For the analysis of this project, three (3) dataset were provided, Africa Production Dataset, Africa Supply dataset and World Population Dataset.

Africa Production Dataset consist of 23110 rows and 4 columns, Africa Supply Dataset consist of 450 rows and 3 columns and World population had 2640 rows and 3 columns.

Step 1

1. Analyze and plot for Food production and Food Supply Between 2004 to 2013.
2. Plot map to show visualization of food production and Food Supply.
3. Solve for some statistical measures (quartiles and interquartile).
4. Test for Hypothesis
5. How rice production across Africa differs from one year to another.

Step 2

Data Wangling, the loading of the dataset, modifying the data, data cleaning, removing outliers, removing duplicates data.

Step 3

Data Exploration; finding patterns and creating better features for exploration.

Questions to derive

1. Top 3 countries producing food in Africa from 2004 to 2013?
2. Top 3 countries supplying food in Africa from 2004 to 2013?
3. Solve for Quartiles, Interquartile and Outliers.
4. Country with Highest food supply in 2012?

Data Wangling

Importation of Libraries packages using Jupyter(Python) as IDE tool.

Loading the dataset

```
proddataset = pd.read_csv('Africa Food Production.csv')
proddataset.head()
```

The code above loads the data into a dataframe, to assess the quality of the data. From the dataframe, the properties of the data and descriptive statistics was generated.

Same Method was used to generate African Food supply and Population dataset.

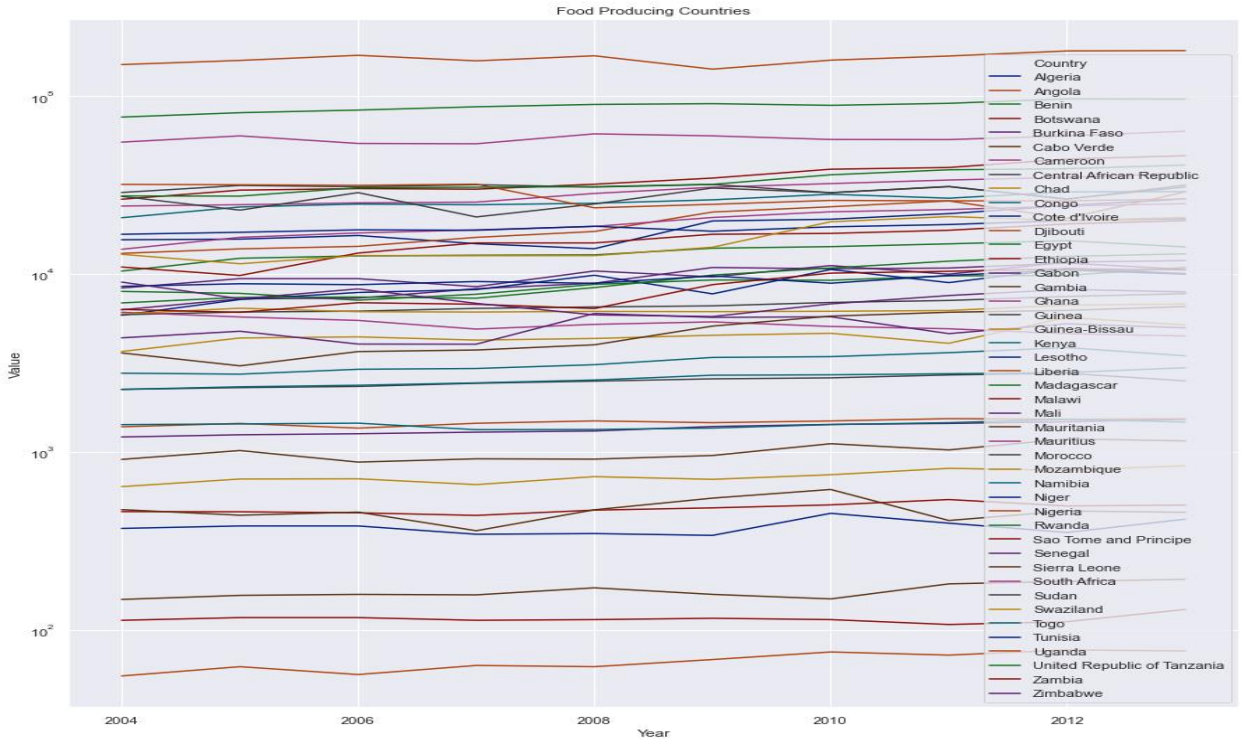
Data Cleaning

Modification of the data set, removal of extraneous data and duplicates, and adding new information to the data.

Exploratory Analysis

Africa Productions Dataset

The Data set is grouped by Selecting Country and Year from the data set. By doing so, a line plot is derived by plotting Year against Values as shown below.



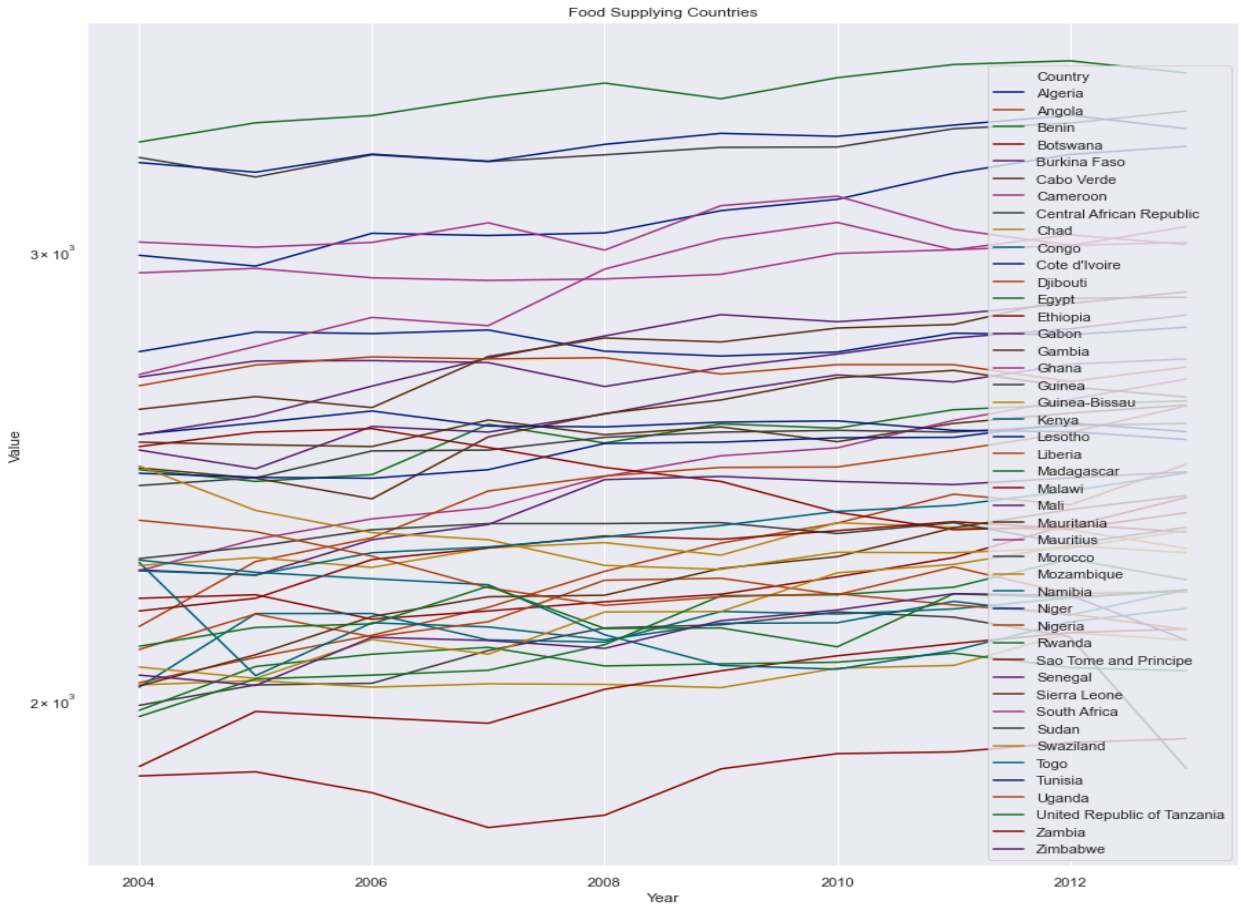
The Figure above shows the production of food by each African country. From the diagram above you can derive countries that have the highest production values from 2004 to 2014. Top 3 incudes:

- Nigeria
- Egypt
- South Africa

Map to illustrate the Producing food across Africa from 2004 to 2013 shows the interaction and change in production of food.

Africa Supply Production

The Data set is grouped by Selecting Country and Year from the data set. By doing so, a line plot is derived by plotting Year against Values as shown below.



The Figure above shows the supply of food per captial by each African country. From the diagram above you can derive countries that have the highest production values from 2004 to 2014. Top 3 incudes:

- Egypt
- Tunisia
- Morocco

Map to illustrate the supply of food across Africa from 2004 to 2013 shows the interaction and change in supply of food.

Some Statistical Measure

Quartile

First Quartile = 2174.0
Third Quartile = 2681.75

Interquartile Range

Interquartile Range = Third quartile – First quartile
2681.75 – 2174.0 = 507.75
The shows the dispersion of the dataset.

OUTLIERS

Lower_limit = Q3 - 1.5*IQR = 1920.125
Upper_limit = Q3 + 1.5*IQR = 3443.375

Scatter plot

A scatterplot to show the relationship between Africa countries in 2012, from the scatter plot w e can derive countries that supply per capital is within and outside the outliers. The scatterplot shoes that Egypt is above the upper fence and Zambia is below the lower fence, so it's an outlie r. The shows Egypt have a high food supply for consumption than other countries.

Rice Production vs Population

Merging of rice values from Africa production dataset and Population was done to compare the change in rice production to population from 2004 to 2013.

Test of Hypothesis

A correlation is a measure of dependence which describes the linear relationship between two variables X and Y. It is a number which lies between -1 (negative correlation) and +1 (positive correlation).

Using Pearson correlation. The coefficient of this coefficient measure indicates how correlated two variables are. It is calculated as the covariance of the two variables divided by the product of their standard deviations.

Result showed a correlation of 0.57, the test is Significant since it's far from 0, The P value for th e Test = 1.2234548112829303e-34.

How rice production across Africa differs from one year to another

A map was plotted to illustrate the change in rice production affect population of African countries from 2004 to 2013.

The map shows in increase of rice production all across Africa Production from 2004 to 2008, then a dip in production in 2009, from 2010 to 2013 a steady rice again.

Egypt, Nigeria and Madagascar are the top 3 rice producing countries.

Intersection

It show countries that produce the same type of food, Nigeria and Algeria was use a set example.