GLOBAL FOOD PRICES

IMPACT ANALYSIS ON POPULATION, MORTALITY, BIRTH-RATE AND GDP

Group Assignment – Data Science 2 – Introduction to Statistics

PRESENTED BY - GROUP 2

Jobbin Samuel
Ricky Lam
Maxim Smetin
Ramila Mudarth
Avery Lizhong Wang
James Lising

Table of Contents

iNTRODUCTION 5
data preparation 5
Global Food Prices ..... **Error! Bookmark not defined.**
Gross domestic product (GDP) 6
MORTALITY RATE ..... **Error! Bookmark not defined.**
POPULATION TREND ..... **Error! Bookmark not defined.**
CURRENCY EXCHANGE ..... **Error! Bookmark not defined.**
COST OF LIVING ..... **Error! Bookmark not defined.**
FEATURE ENGINEERING ..... **Error! Bookmark not defined.**
data ANALYSIS 7
HYPHOTSIS TESTING ..... **Error! Bookmark not defined.**
CONCLUSION 10
Bibliography 11
ApPENDIX 12

INTRODUCTION

The global food prices have always been subjected to external influences like fuel prices, natural disasters as a result of global warming activities. Some countries have been affected more than others depending on their ability to endure the fluctuations and food availability. There are assumptions made that the food prices influence the mortality, the GDP, affordability and income which in turn effects the population trends. Producers benefit from rise in prices where are consumers benefits from lower food prices. Any fluctuation in prices will have an effect especially on the lower income individuals as a result causing food shortages.

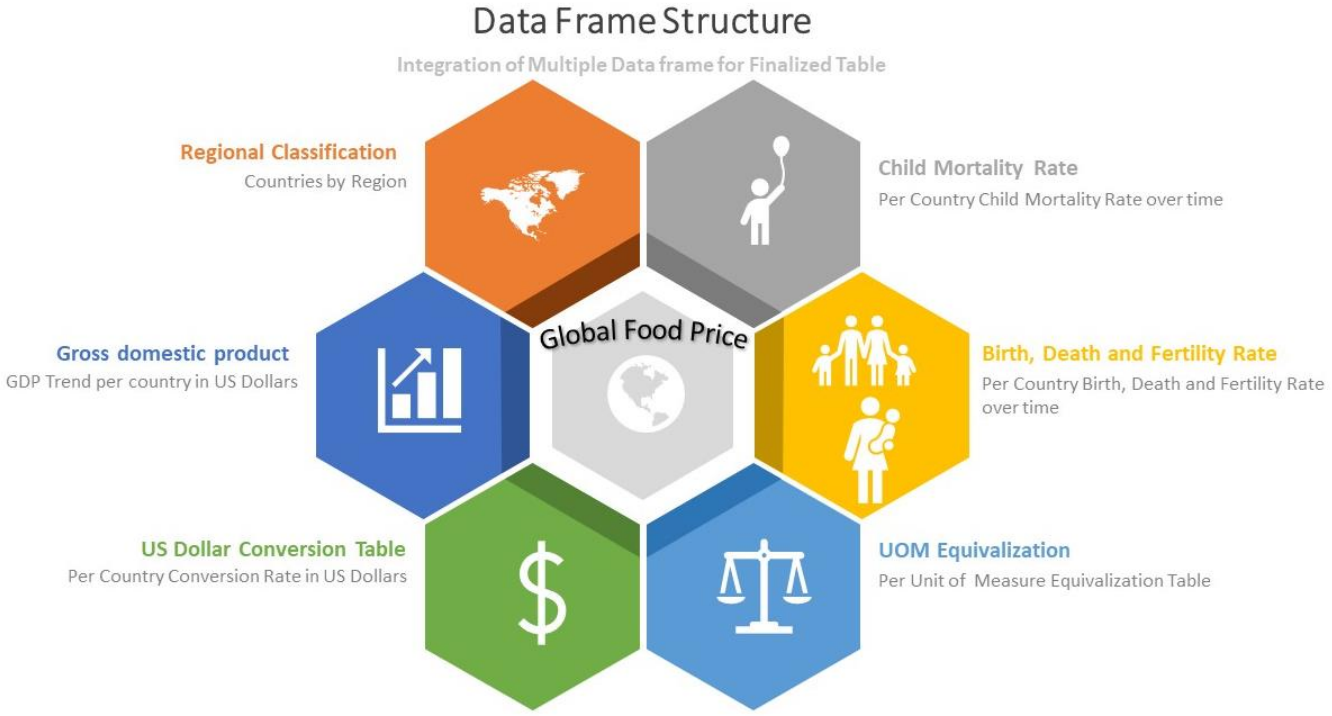
The objective of our assignment was to determine the impact of food prices of specific commodity in developing countries and compare their GDP, Mortality Rate to determine any kind of correlation that might exists using last twenty years of data. Additionally, we intend to pair it with restaurant index data to see if there are any insights derived.

We intend to resolve the two hypotheses

- **Hypothesis 1:** How much food price influence Population? Null Hypothesis is food price isn't a key driver of population. The alternative Hypothesis is that food price somewhat affect population.
- **Hypothesis 2:** How much do food prices impact all parameters? Null Hypothesis is that food prices impact all parameters equally. The alternative hypothesis is that there are some differences between some parameters affected by food prices.

DATA PREPARATION

Preparation of the data set required compiling and sourcing from multiple location. Each data set had to be solved for challenges presented and transformed for the required analysis.



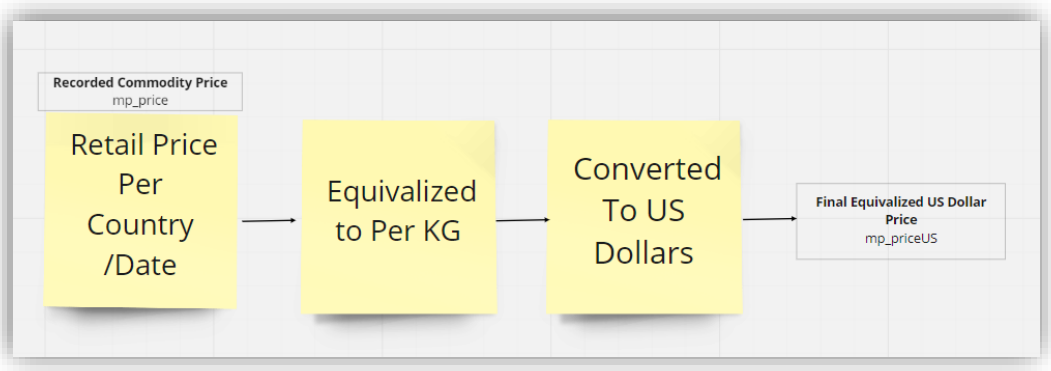
The core dataset contains Global Food Prices data from the World Food Programmed covering foods such as maize, rice, beans, fish, and sugar for 98 developing countries and some 1,500 markets. The data goes back as far as 1992 for a few countries, although many countries started reporting from 2003 or thereafter. The Data is collected by WFP(The World Food Program) and the dataset was distributed by HDX . Data includes developing countries, locality, market, goods purchased, price & currency used, quantity exchanged, and month/year of purchase.

The Food prices were in local currency. The classification of the commodity category was too granular, so we had to aggregate it at a boarder level. *e.g. Rice commodity name has 82 different versions. But we created a category by using the first word in the string.*

```
rice = df[df['cm_name'].str.contains("Rice")]
rice['cm_name'].unique()

'Rice (white, imported) - Retail', 'Rice (white) - Retail',
'Rice (coarse) - Retail', 'Rice (medium grain) - Wholesale',
'Rice (medium grain) - Retail',
'Rice (coarse, BR-8/ 11/, Guti Sharna) - Wholesale',
'Rice (coarse, BR-8/ 11/, Guti Sharna) - Retail',
'Rice (coarse, Guti Sharna) - Wholesale',
'Rice (coarse, Guti Sharna) - Retail', 'Rice - Wholesale',
'Rice (imported) - Wholesale', 'Rice (imported) - Retail',
```

The unit of measure (UOM) for the commodities were not consistent per observation. We created a table for equivalizing the retail value. As a result, when the aggregation per Country, Commodity and Year would be realistic. A conversion table was manually created to do the math. No conversion was applied to fuel, and commodity item such as toothbrush, toothpaste, and other liquid that is unmeasurable. UOM in file has not been changed in the file. But those that have been change will be reflect in retail price. *e.g. where the retail was 2500 dollar for 12 kg was converted to 208 dollars per KG.*



The other features merged into the table are

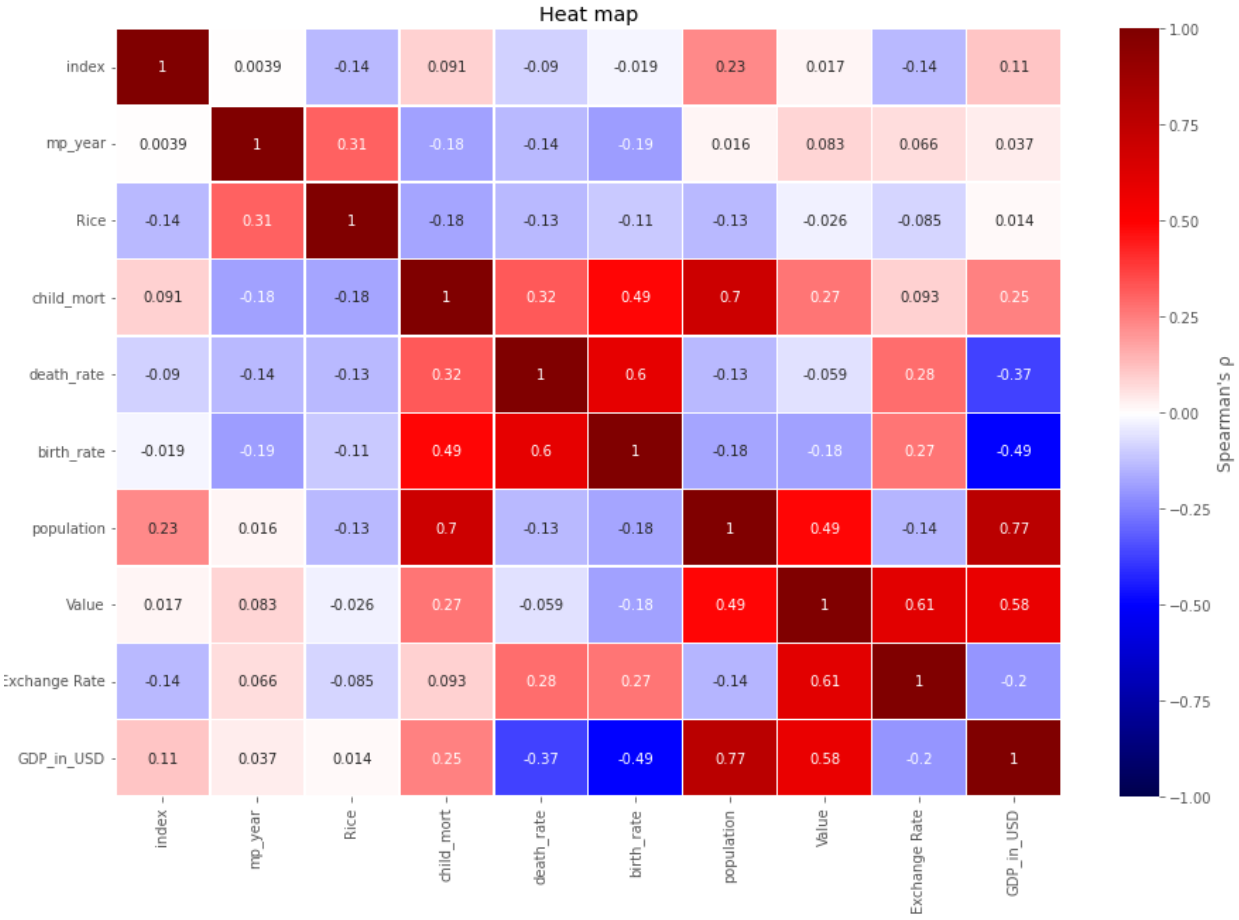
- Regional Classification - Countries were reclassified by the region for higher level analysis and to obtain greater number of observations eg. Northern Africa', 'Eastern Africa', 'Middle Africa', 'Southern Africa',
- Gross domestic product (GDP) – Twenty years of GDP information trend per country
- Child Mortality Rate - Child Mortality counts since 1967 per country
- Birth and Death Rate – Birth and Death count per country since 1960
- Fertility Rate - Fertility Rate since 1950 per country

We added another layer of classification for the commodities

- Raw - Milk, Eggs, Rice
- Processed – Bread, Curds
- Other - Fuel, Internet

DATA ANALYSIS

The prepared data set was further compressed and aggregated to give commodity per country per area. In the prepared data set, we observed that Rice was one commodity that had sufficient data point. We created a subset data frame to run an analysis on. A heat Map was generated using this data set.



Initial high-level observations from the Heat Map shows strong positive correlation between the following parameters

- Population and GDP in USD
- Birth Rate and Death Rate
- Population and Child Mortality

The negative correlation exists in the following parameters

- GDP versus Birth Rate as well as Death Rate
- Price of Rice versus Child Mortality, Birth and Death

The parameters that are more likely to be independent are

- Shelf Price of Rice and GDP is mutually exclusive.

LINEAR REGRESSION ANALYSIS

We performed Linear regression modeling on Birth Rate, Death Rate and Mortality at a regional level to see the relationship between the parameters in the data

BIRTH RATE ANALYSIS

In the Birth Rate OLS Model, we observed adj R Squared is 68% making the model reasonable fit.
All the p values are zero as a result we can say, we accept the null hypothesis. All parameters are acceptable in the model .

OLS Regression Results

Dep. Variable:	birth_rate	**R-squared:**	0.685
Model:	OLS	**Adj. R-squared:**	0.683
Method:	Least Squares	**F-statistic:**	237.8
Date:	Mon, 06 Dec 2021	**Prob (F-statistic):**	4.78e-187
Time:	04:17:05	**Log-Likelihood:**	-2442.8
No. Observations:	772	**AIC:**	4902.
Df Residuals:	764	**BIC:**	4939.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	28.3020	0.931	30.395	0.000	26.474	30.130
Continent[T.Americas]	-10.2610	0.675	-15.212	0.000	-11.585	-8.937
Continent[T.Asia]	-11.0315	0.579	-19.063	0.000	-12.168	-9.895
Continent[T.Europe]	-30.2254	2.924	-10.335	0.000	-35.966	-24.485
GDP_in_USD	-0.0017	0.000	-9.273	0.000	-0.002	-0.001
death_rate	0.8506	0.084	10.067	0.000	0.685	1.016
child_mort	9.309e-06	1.56e-06	5.986	0.000	6.26e-06	1.24e-05
Rice	-0.2908	0.069	-4.189	0.000	-0.427	-0.155

Omnibus:	0.537	**Durbin-Watson:**	0.204
Prob(Omnibus):	0.765	**Jarque-Bera (JB):**	0.633
Skew:	0.032	**Prob(JB):**	0.729
Kurtosis:	2.875	**Cond. No.**	2.23e+06

CHILD MORTALITY ANALYSIS

In the Child Mortality analysis, the Adjusted Rsquare is 15% as a result htemodel is unfilt.
Our Defaul Alpha is 0.05. So any we can remove continent of America and Europe

OLS Regression Results

Dep. Variable:	child_mort	**R-squared:**	0.163
Model:	OLS	**Adj. R-squared:**	0.156
Method:	Least Squares	**F-statistic:**	21.31
Date:	Mon, 06 Dec 2021	**Prob (F-statistic):**	2.52e-26
Time:	04:17:23	**Log-Likelihood:**	-10187.
No. Observations:	772	**AIC:**	2.039e+04
Df Residuals:	764	**BIC:**	2.043e+04
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-1.958e+05	3.07e+04	-6.386	0.000	-2.56e+05	-1.36e+05
Continent[T.Americas]	3.384e+04	1.75e+04	1.937	0.053	-448.449	6.81e+04
Continent[T.Asia]	8.24e+04	1.57e+04	5.247	0.000	5.16e+04	1.13e+05
Continent[T.Europe]	7565.5993	7.1e+04	0.107	0.915	-1.32e+05	1.47e+05
GDP_in_USD	10.9276	4.319	2.530	0.012	2.448	19.407
death_rate	9368.6059	2016.469	4.646	0.000	5410.128	1.33e+04
birth_rate	4812.8530	803.995	5.986	0.000	3234.551	6391.155
Rice	7318.5372	1574.478	4.648	0.000	4227.720	1.04e+04

Omnibus:	659.667	**Durbin-Watson:**	0.093
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	12562.181
Skew:	3.956	**Prob(JB):**	0.00
Kurtosis:	21.109	**Cond. No.**	1.81e+04

DEATH RATE ANALYSIS

The model is at 52%. The price of Rice is not relevant as we need to remove the parameter as the p value is greater than 0.05

OLS Regression Results

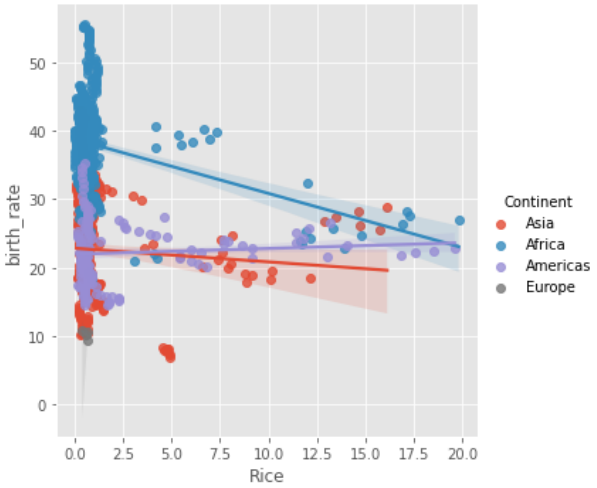
Dep. Variable:	death_rate	**R-squared:**	0.527
Model:	OLS	**Adj. R-squared:**	0.523
Method:	Least Squares	**F-statistic:**	121.6
Date:	Mon, 06 Dec 2021	**Prob (F-statistic):**	1.20e-119
Time:	04:17:35	**Log-Likelihood:**	-1739.8
No. Observations:	772	**AIC:**	3496.
Df Residuals:	764	**BIC:**	3533.
Df Model:	7		
Covariance Type:	nonrobust		

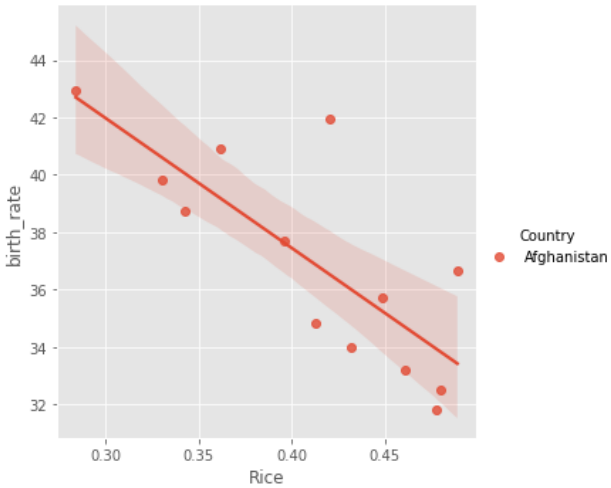
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	5.1025	0.525	9.713	0.000	4.071	6.134
Continent[T.Americas]	-2.4887	0.296	-8.397	0.000	-3.071	-1.907
Continent[T.Asia]	-1.8045	0.275	-6.557	0.000	-2.345	-1.264
Continent[T.Europe]	8.1572	1.221	6.681	0.000	5.760	10.554
GDP_in_USD	0.0004	7.56e-05	4.746	0.000	0.000	0.001
birth_rate	0.1377	0.014	10.067	0.000	0.111	0.165
Rice	-0.0382	0.028	-1.353	0.177	-0.094	0.017
child_mort	2.933e-06	6.31e-07	4.646	0.000	1.69e-06	4.17e-06

Omnibus:	168.052	**Durbin-Watson:**	0.236
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	356.695
Skew:	1.200	**Prob(JB):**	3.51e-78
Kurtosis:	5.309	**Cond. No.**	2.36e+06

RICE CONTRIBUTION IN AFGHANISTAN AND REGIONAL LEVEL

We have observed negative correlation between rice price and birth rate in Africa and a little bit in Asia





CONCLUSION

What did we infer from our analysis?

<< NOTABLE HIGHLIGHTS

Closing Comments three important notes

BIBLIOGRAPHY

Exchange Rates. (n.d.). Retrieved from https://data.oecd.org/conversion/exchange-rates.htm.
Global Food Prices. (n.d.). Retrieved from https://data.humdata.org/dataset/wfp-food-prices.
Gross Domestic Product. (n.d.). Retrieved from https://stats.oecd.org/Index.aspx?DatasetCode=SNA_TABLE1.
Population Trends. (n.d.). Retrieved from https://data.worldbank.org/indicator/SP.POP.TOTL.

APPENDIX