National College of Ireland

Student Name: Nikolaos Gkmpenompa
Student Number: 22213708

Course: Higher Diploma in Science in Data Analytics - HDSDA_JAN23
Module: Programming for Data Analytics
Lecturer: Symeon Charalabides
Submission Date: 07/05/2023
Project Title: The impact of COVID-19 pandemic on the food prices in the “developing world”

I hereby certify that the datasets used in this (my) project are publicly available, and the information provided comes from research I conducted specifically for this project. All the material and code other than my own is fully referenced in the relevant section.

Signature: *Nikolaos Gkmpenompa*
Date: *07/05/2023*

The impact of COVID-19 pandemic on the food prices in the "developing world".

Introduction

The COVID-19 pandemic undisputedly made a significant impact on all aspects of human life. Socially, mentally, physically, economically. Economically speaking, due to the national lockdowns that the governments put in place to stop the spread of the pandemic, one of the sectors that was heavily affected was the food retail supply chain. Due to the restriction measures, the international travel, global logistics, supplier production and the overall lack of staff in food production and distribution, resulted in shortages at retailers which increased prices for basic food commodities.

The COVID-19 pandemic has had and is still having different impacts and effects on food supply, food retail and food processing. A quick look into data of the Food and Agriculture Organization Corporate Statistical Database (FAOSTAT) shows an abnormal change in the producer prices in the years before, during and after the main COVID-19 hit.

But were all countries impacted at the same level? The answer is no, as unfortunately there is huge inequality in the countries' economies, HDIs, and GNI per Capita. This report is going to focus on the developing countries as their economies are found to be more fragile and they are more exposed to food insecurity.

The objectives of my analysis are to investigate the relationship between the HDI and GNI per Capita of the developing countries, investigate the impact of COVID-19 on the prices of basic commodities, the relationship between the HDI and the food prices after the COVID-19 hit, and the geographic distribution of food prices across the world. To achieve reaching these objectives I found and organised 2 different datasets. One with data from the developing countries and one with producer prices of basic food commodities from these countries, in the years before and after the COVID-19 pandemic. Eventually I end up merging these two datasets.

1st Dataset: Developing counties.

The United Nations uses the Human Development Index (HDI) metric to determine whether a country is fully developed or still developing. The HDI considers a broad range of factors, including economic growth, life expectancy, health, education, and quality of life. The highest possible HDI score is a 1.0, and any country that scores less than .80 is considered developing. Of the 191 countries analysed in the 2021/22 Human Development Report, 125 scored below .80 and were considered developing.

Another frequently used method of determining whether a country is developed or developing is to examine that country's nominal gross national income (GNI) per capita, which is an estimation a country's overall standard of living. Countries who's nominal GNI rises above a certain threshold (which changes slightly each year) are classified as developed, while those whose GNI falls below that amount are considered still developing. (worldpopulationreview.com)

The first dataset was downloaded from "worldpopulationreview.com" and its name is "Developing Countries 2023". It is saved as "dev_countries_data.csv" in my project folder. It consists of 169 rows and 19 columns, indicating the Place, Population in 2023, Growth Rate, Area, Country Name, ISO Codes: cca3, cca2, ccn3, Region, Subregion, Land Area (Km), Density, Density (Mi), Rank, HDI Tier, HDI in 2021, Income in 2022, GNI Per Capita in 2020, and Rank.

I imported the dataset as dataframe using the Pandas library. Since I didn't need all columns, I did a subset using only the columns "country", "cca2", "region", "subregion", "landAreaKm", "hdiTier", "hdi2021", "wbig2022", "gniPerCapita2020", which I renamed to "Country", "Code", "Region", "Subregion", "Land Area (Km)', "HDI Tier", "HDI", "Income", "GNI per Capita (USD)".

I renamed the missing values which was named as "not rated" to "nan" and located all rows (countries) with "nan" values. Removed the countries that was missing both HDI and GNI per Capita values. Replaced the "nan" values in the HDI Tier column according to the HDI values, using a combination of loops and conditional statements. For the last 7 countries with missing values in GNI per Capita, since there was no way to calculate the correct values based on the rest of the data in my dataset. I replaced them with values I found online (Google). Then I used again loops and conditional statements to replace the "nan" values in the Income column according to the GNI per Capita values. Finally, I removed the very small countries (one island: Saint Kitts and Nevis) with Land Area below 300 km2 and GNI per Capita above 15000 USD, as they are not really considered developing. Eventually, the clean dataset was saved as **"country_clean.csv"** in my project folder.

Insight 1: Relationship between GNI per Capita and HDI

First, I created a scatter plot to visualise the relationship between the GNI per Capita and HDI of the developing countries. To be able to visually distinguish the HDI level of the 123 countries, I used colour coded dots and loops, to assign them according to the 3 categories (High, Medium, and Low) of the HDI Tier column.

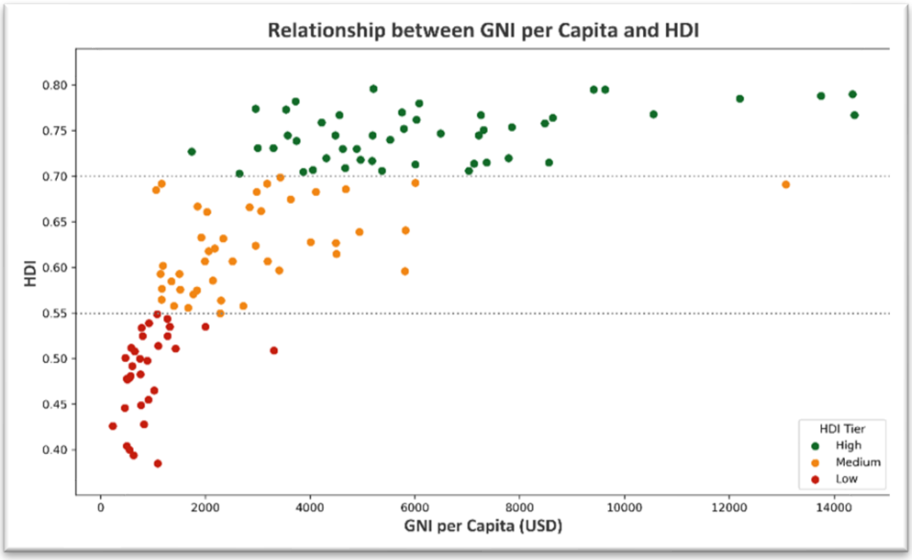

Fig. 1: Relationship between GNI per Capita and HDI in the developing countries.

Visually we can tell that as the GNI per Capita increases, so the HDI does. To strengthen my impression, I calculated the covariance between the "HDI" and "GNI per Capita". It was found to be 212.66, indicating that the two variables tend to vary together in a positive direction. When "GNI per Capita" increases, there is a tendency for the "HDI" to also increase. However, since the covariance is influenced by the scales and units of the variables. To have a more complete view, I also calculated the correlation coefficient, which is a more consistent measure of the strength and direction of their relationship. The correlation coefficient was 0.7461 which also indicates that there is a strong positive correlation between the "GNI per Capita" and "HDI" variables since it is closer to 1.

Finally, I calculated the linear regression and added the regression line in the previous graph.

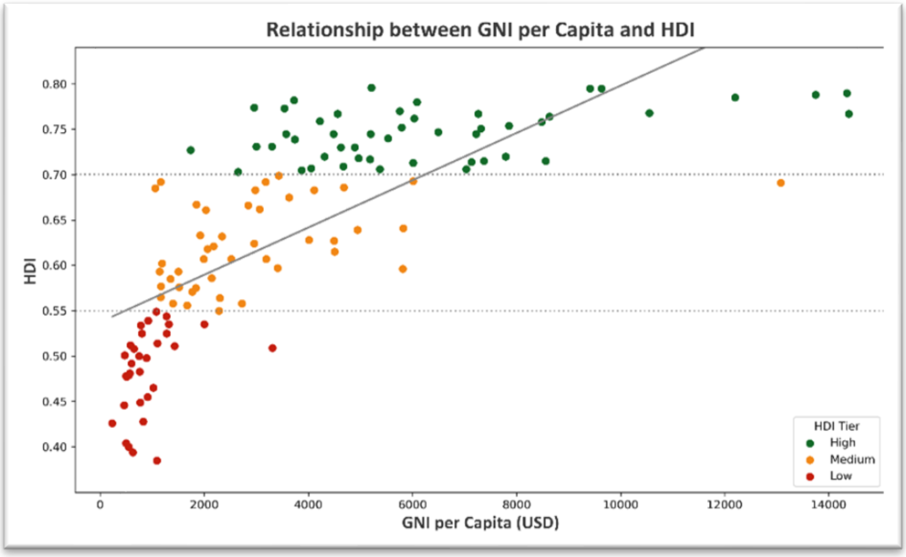

Fig. 1.1: Relationship between GNI per Capita and HDI with the regression line.

In conclusion, there is indeed a relationship between the GNI per Capita and HDI increase.

2nd Dataset: Basic food commodities

The basic food commodities for a country can vary depending on the geography, climate, culture, and economic conditions. There are several common food commodities that are considered basic for most countries:

- Grains: wheat, rice, corn, oats, and barley.
- Vegetables: potatoes, carrots, onions, tomatoes, and beans
- Fruits: apples, bananas, oranges, and grapes
- Meat: from animals such as beef, pork, chicken, and fish
- Animal products: Such as milk from cattle, goat, sheep and eggs
- Fats and oils: such as butter, margarine, and cooking oils.

These basic food commodities are supplemented by other foods depending on the cuisine of a particular country.

My second dataset was downloaded from fao.org/faostat ⇨ Producer Prices ⇨ Special: Land Locked Developing Countries + Elements: Producer Price (USD/tonne) + Years: 2020, 2021, 2022 + Months: Annual Value + Items: The list of items that was finally selected were the ones with the most data across as many developing countries possible. The downloaded file is saved as “FAOSTAT_data.csv” in my project folder. It consists of 2462 rows and 8 columns, indicating the “Domain”, “Area”, “Element”, “Item”, “Year”, “Months”, “Unit”, and “Value”.

I imported the dataset as dataframe using the Pandas library. Then, subset using only the columns "Area", "Item", "Year", "Value" and renamed them to “Country”, “Commodity”, “Year”, “Producer Price (USD”. I created a list of the unique values (Countries) of the Country column because it consists of many repeated values.

From now on I will work only with the countries I managed to found data on their food commodities. Therefore, I created a list of the countries with data from my second dataframe (food_clean), created a list of the countries from my previous dataframe (country_clean), and compared the names of the countries using list comprehension. The result was 3 countries whose names were replaced with the correct ones from the first dataframe. ‘'Iran (Islamic Republic of)' was replaced by 'Iran', 'Republic of Moldova' was replaced by 'Moldova', and 'Viet Nam' ' was replaced by 'Vietnam'. Then I replaced the commodities with names longer than 25 characters with a shorter version.

Because the current structure of my dataframe was not practical enough to work with, I did a pivot table and then reset the index to keep the new structure. The “Producer Price (USD/tonne)” column containing the values for the years 2019, 2020, 2021 was split into columns 2019, 2020, and 2021.

The commodities with missing prices in both 2019 and 2020 were removed, and the ones that were missing price in 2021, I calculated the percentage of difference between 2019 and 2020 and multiplied the result with the 2020 price to replace the “nan” value in 2021. Eventually I removed the countries with less than 5 unique commodities as I would not have enough data to properly represent their overall commodity price change.

Last thing was to add “Covid Impact” column representing the change percentage between 2019 and 2021 (pre and post COVID-19) foe every commodity price using this equation: Price Percentage Change = ((2021 Price - 2019 Price) / 2019 Price) * 100.

Eventually, the clean dataset was saved as **“food_clean.csv”** in my project folder.

Merging my datasets

From this point I was going to work with data from both datasets, therefore I merged them. The new merged dataset included the information on the countries in the “country_clean” dataframe, which prices were present in the “food_clean” dataframe. The final dataset was saved as **“final_country_food.csv”** in my project folder.

Insight 2: COVID-19 impact on the food prices.

To investigate the impact of COVID-19 on the food prices I grouped the 'Covid Impact' values of the commodities for each country and calculated the mean value. To visualise the number of countries that COVID-19 affected negatively and positively. I created a bar plot with colour coded bars, green for positive or no impact (mean ≤ 0), and red for negative impact(mean > 0), according to the mean value of each country (code).

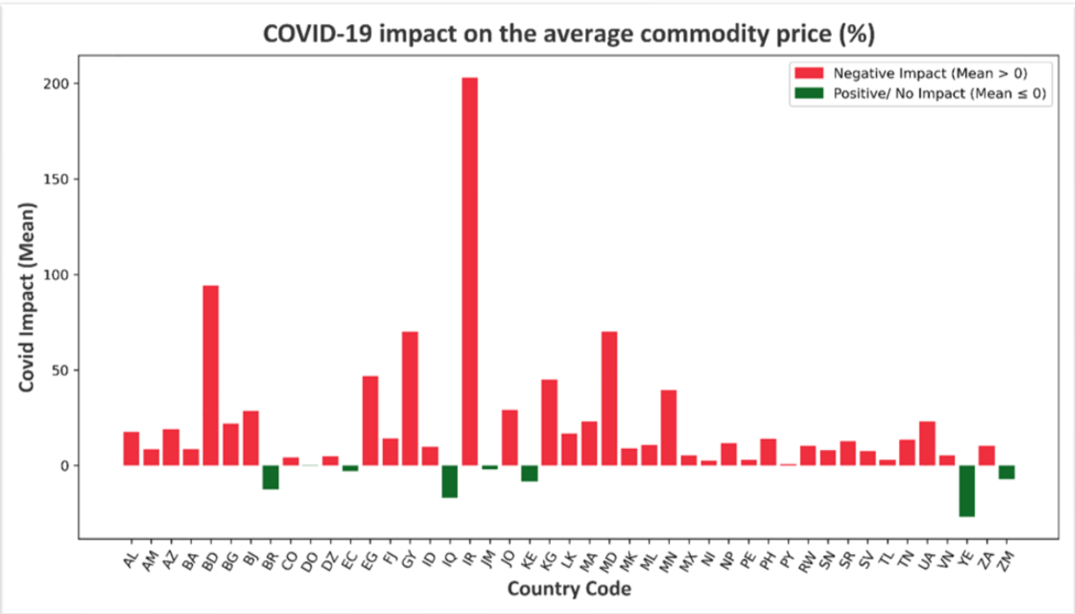

Fig.2: COVID-19 impact on the average price of the commodities per country

It is obvious that most of the countries were affected negatively from the COVID-19 pandemic. Meaning that their basic food prices went up between the years before and after COVID-19.

Insight 3: Relationship between the HDI and the food prices in 2021

My third objective is to investigate the relationship between the countries' HDI and the post COVID-19 food prices. I did subset the merged dataframe (final_country_food) to include the "HDI", the "HDI index" and the food prices for 2021.

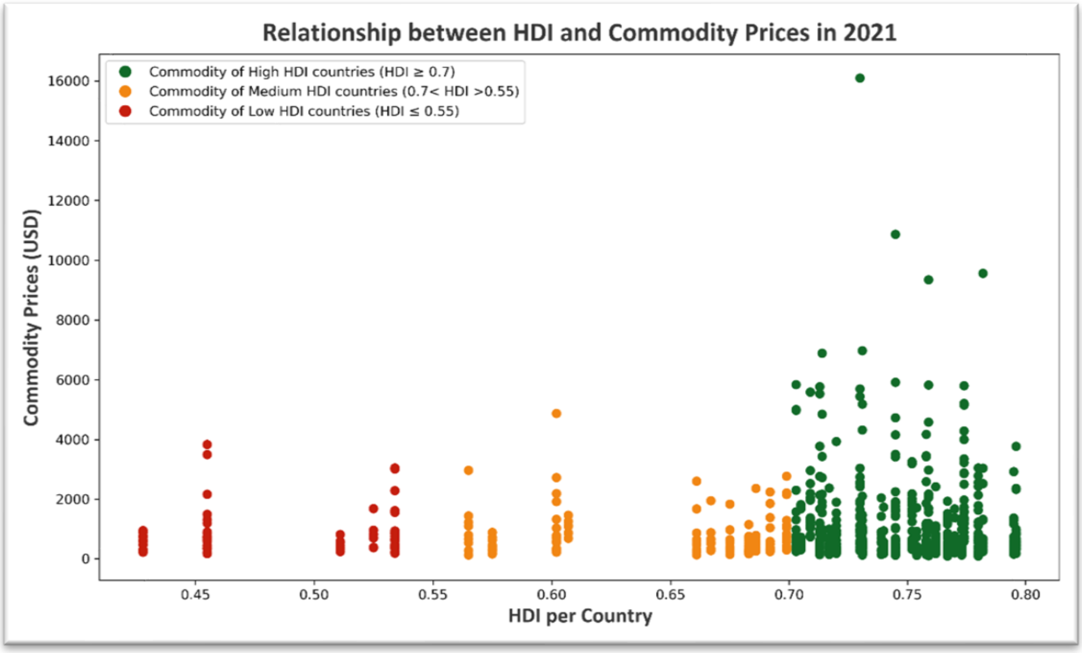

Fig. 3: Relationship between the HDI and the food prices in 2021

There is no visible connection between the HDI and the 2021 food prices. To further strengthen my conclusion, I calculated the correlation coefficient for each HDI group, using the corrcoef() method, for every HDI tier group. #corrcoef() takes two arrays as arguments and returns a sequence of correlation coefficients. #groupby() method is used to group the dataframe by HDI Tier. #corr() method calculates the correlation coefficient between HDI and food prices for each group and returns a multi-level index dataframe with correlation coefficients for all possible pairs of columns. Since I only need the correlation coefficient between HDI and food prices, I select those with .iloc[0::2, -1], where 0::2 notation selects every other row starting from the 1st one (the correlation coefficients for HDI). The -1 notation selects the last column (the correlation coefficients for the food prices). The correlation coefficient between,

- High HDI countries and their 2021 food prices = -0.10724
- Medium HDI countries and their 2021 food prices = 0.0612376
- Low HDI countries and their 2021 food prices = -0.1209

HDI Tier	Index	2021
High	HDI	-0.10724
Low	HDI	0.0612376
Medium	HDI	-0.1209

The correlation coefficient values indicate the strength and direction of the relationship between HDI and food prices for each HDI tier group. A value of 1 is a perfect positive correlation and a value of 0 indicates no correlation, meaning that there is no relationship between the two variables. In this case, all three correlation coefficients show that the relationship between HDI and food prices is not significant in any of the HDI tier groups. That means that other factors play a more significant role in determining food prices in these countries. This indicates that in these countries, other factors are more important in determining food prices.

Insight 4: Geographical distribution of the food prices on the map

For the last insight I chose to display the percentage of price change for basic commodities in my on the world map. This would make researching the geographical distribution of the food prices around the world easier, by using a colour coded impression of the countries on the map, according to the COVID-19 impact on their food prices.

I imported the required geopandas and geolocator libraries. Created a geocoder object (stackoverflow) and created a function to get the latitude and longitude of the locations in my dataset (Week 6 lesson). I created “countrycode_list” dataframe with my country names and their cc2 codes, to be used for geo-coordinates, and then added two columns for the latitude and longitude values.

Control structures were used to loop through each row in “countrycode_list” and get the latitude and longitude for my countries. Then “countrycode_list” containing the geocoordinates was merged with “covimp_average” containing the percentage of food price change, between the year before (2019) and year after the pandemic (2021).

Then I used the function geopandas.read_file() to get a shapefile of the world map. This was saved under “geo_world_map” dataframe. Then, I renamed the column region to country, so I can merge it with my “geo_covimp” dataframe on the "Country" name. The final dataframe was saved as “geo_covimp_map”.

Finally, I created a choropleth map using the geo_covimp_map and the function matplotlib.pyplot(). The colour gradient is used to represent the percentage from the column "Covid Impact".

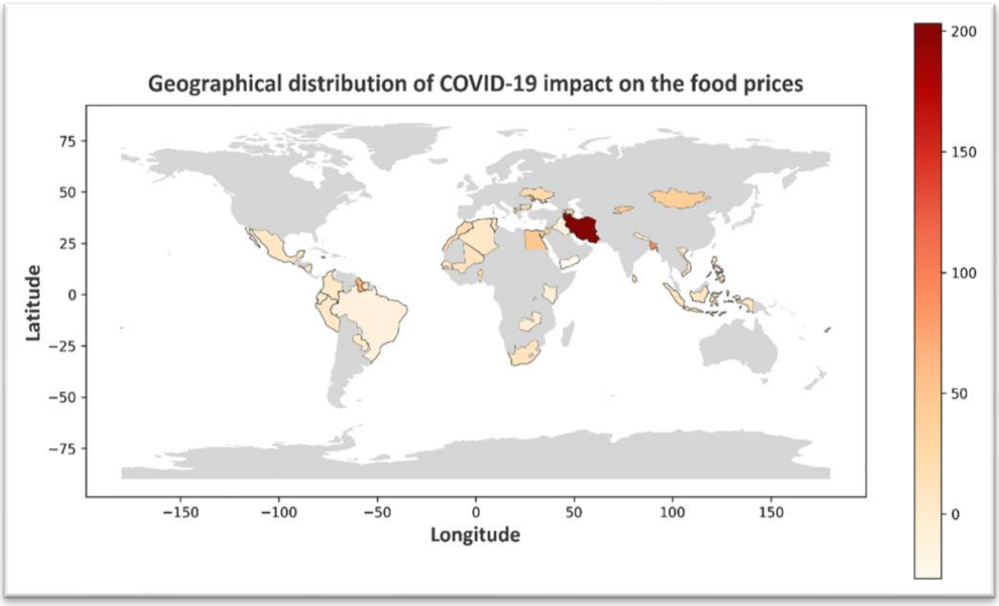

Fig. 4: Geographical distribution of the COVID-19 impact on food price.

The colour legend represents the percentage of average price change between 2019 and 2021. Significant values can be found in Mongolia, Kyrgyzstan, Egypt, Guyana, Moldova, Bangladesh and especially in Iran with 200% increase. Four of the most impacted countries are in Asia, however they are not in the same neighbourhood and therefore geographically have not much in common.

A larger dataset with prices of more commodities would have given better results. The fact that there was not even 1 commodity used by all countries in the datasets I could find, shows that it is difficult to come to conclusions when comparing different number of commodities per country and different types of food.

Dataset Sources

List of the developing countries and their data: https://worldpopulationreview.com/country-rankings/developing-countries

List of commodities and its producer prices: FAOSTAT: Food and Agriculture Organization of the United Nations: https://www.fao.org/faostat/en/#data/PP

References

Developing country, Wikipedia: https://en.wikipedia.org/wiki/Developing_country

COVID-19 and the food and agriculture sector: Issues and policy responses. OECD. https://www.oecd.org/coronavirus/%20policy-responses/covid-19-and-the-food-and-agriculture-sector-issues-andpolicy-responses-a23f764b/

Appendix

Week 6: Functions and Object-oriented Programming

- Function use

Week 7: Introduction to Dataframes

- **Cheat Sheet: The pandas DataFrame Object (used vastly throughout the project)**
- Importing and saving to csv file
- DataFrame iteration methods

Week 5: Control Structures

- Combination of controlled structures and Dataframe iteration methods
- List comprehension

Week 8: Data pre-processing

- Pandas library

Week 9: Data Visualisation

- Scatter plot
- Bar plot

Week 10: Data Analysis

- Correlation
- Correlation coefficient
- Linear regression

How to manually create a legend

- https://stackoverflow.com/questions/39500265/how-to-manually-create-a-legend

Add custom legend in Scatterplot with matplotlib

- https://dataplotplus.com/add-custom-legend-in-scatterplot-with-matplotlib-and-python/