Task 6.1

Section -1: Data Source

Consumer Price Index (CPI):

Consumer Price Indexes (CPIs) are index numbers that are index numbers that measure the changes in the prices of goods and services purchased or otherwise acquired by households for direct use or consumption ([link](#)).

Source of data: The data has been collected from external data source.

Owner/collector/trustworthiness of data: The data has been collected by IMF (International Monetary Fund) and . Organization for Economic Cooperation and Development (OECD). The dataset is open source and available at: [https://data.imf.org/.](https://data.imf.org/) Therefore, the data is trustworthy.

Data collection method: The data has been collected manually through surveys conducted by the government of each country and then reported to IMF and OECD. IMF also verifies the data to ensure reliability.

Contents of data: The dataset includes country wise monthly Consumer Price Index (CPI) data from 1990 to 2022, for overall and 12 subgroups of consumption expenditure. It also includes actual weights, weight percentage and percentage change from previous period for each subgroup.

Limitations of data: A lot of data is missing from this data set. As IMF depends of independent countries to report their data, the data might be vulnerable to misrepresentation by the reporting countries.

Relevance of data: The data in the data set is relevant for this project as it has country wise monthly consumer expenditure data for different goods and services from 1990-2022.

Primary Commodity Pricing System (PCPS):

Average monthly global price indexes and actual prices for fuel and non-fuel commodities ([link](#)).

Source of data: The data has been collected from external data source.

Owner/collector/trustworthiness of data: The data has been collected by IMF (International Monetary Fund). The dataset is open source and available at: [https://data.imf.org/.](https://data.imf.org/) Therefore, the data is trustworthy.

Data collection method: The benchmark prices for commodities are determined by the largest markets of a given commodity.

Contents of data: Average monthly global price indexes and actual prices for fuel and non-fuel commodities from 1992-2022.

Limitations of data: The price indexes are estimates determined based on the largest markets of a given commodity. Hence, the actual market prices might differ from country to country

Relevance of data: The data is relevant to the project as it will help me determine correlation between global commodity prices and Consumer Price Indexes.

Reason for choosing these datasets:

I chose these datasets for the following reasons:

    i)   <u>Reliable source:</u> the source of the data is IMF which is one of the most reliable sources of data
    ii)   <u>Timeliness:</u> This dataset contains recent data up to second quarter of 2022.
    iii)   <u>Detailed data:</u> We have monthly data for more than 100 countries since late 1990s

Section -2: Data Profile

Limitations:

Consumer Price Index (CPI):

1) The CPI data is reported by the countries and then verified by the IMF. Hence, the data is vulnerable to misrepresentation by the countries
2) The CPI only considers prices in urban areas. Harmonized price indexes, indicators that consider rural areas also, are available for EU countries only.

Primary Commodity Pricing System (PCPS):

1) The price indexes are estimates determined based on the largest markets of a given commodity. Hence, the actual market prices might differ from country to country
2) We do not have country wise monthly commodity prices. We have monthly global price indexes for different commodities in the PCPS dataset.

Data ethics related issues:

There are no data ethics related concerns as the data used in the project is readily available for public at IMF website and all the data points are at country level. Hence, there is no concern related to personal information of any individual.

Data Cleaning

CPI dataset:

1. Removed unnecessary columns like "Common Reference Priod" and "Unnamed" columns
2. Unpivoted the CPI data frame
3. Changed the column names to uniform format
4. Removed unnecessary records like records having 'Reference Period' in 'attribute' column and all the harmonized indicators.
5. Removed months having excessive number of null values (>23%).
6. Replacing string values in 'value' column with 'null' values
7. Transposed CPI dataframe to convert the values of 'indicator_name' column to separate variables
8. Created 'year' and 'month' columns based on 'month_code' column
9. Imputed null values in all indicator columns with country wise yearly averages
10. Removed records form countries with more than 1 indicators missing data from whole year
11. Imputed missing values for 'Sweden', 'Suriname', Lebanon, Vietnam, China, P.R.: Macao and Brunei Darussalam through calculating missing values based on other indicators.
12. Re-calculated the weight percent columns based on the indicator values

PCPS dataset:

1. Unpivoted the PCPS dataframe.
2. Removed unnecessary columns like "Country name", "Country Code", "Commodity Code" and "Unit Code" from the PCPS dataframe
3. Renamed the PCPS dataframe columns to uniform format
4. Removed unnecessary records related to previous period comparisons and USD units
5. Checked for null values, but no null value was detected.
6. Transposed the PCPS dataframe to convert "commodity_name" column in to separate variables

Merging the datasets and consistency checks:

1. Merged the CPI dataset and PCPS dataset on "year" and "month" columns.

2. No null values was detected
3. No mixed data type column was detected
4. The CPI variables had some massive outliers. The records from countries with massive outlier values have been removed.

Relevant variables and data sources:

Variables	Description	Source dataset	Ultimate source
country_name	Name of the country	CPI dataset	IMF
month_code	Month code composed of year and month sequence	CPI dataset	IMF
year	Year	CPI dataset	IMF
month_number	Sequence of the month	CPI dataset	IMF
month	Name of the month	CPI dataset	IMF
alcoholic_beverages_tobacco_narcotics	Consumer price index of alcoholic beverages tobacco narcotics	CPI dataset	IMF
alcoholic_beverages_tobacco_narcotics_weight_percent	Weight percentage of the consumer price index of alcoholic beverages tobacco narcotics	CPI dataset	IMF
clothing_footwear	Consumer price index of clothing footwear	CPI dataset	IMF
clothing_footwear_weight_percent	Weight percentage of the consumer price index of clothing footwear	CPI dataset	IMF
communication	Consumer price index of communication	CPI dataset	IMF

communication_weight_percent	Weight percentage of the consumer price index of communication	CPI dataset	IMF
consumer_price_index_all_items	Consumer price index of consumer price index all items	CPI dataset	IMF
education	Consumer price index of education	CPI dataset	IMF
education_weight_percent	Weight percentage of the consumer price index of education	CPI dataset	IMF
food_non-alcoholic_beverages	Consumer price index of food non-alcoholic beverages	CPI dataset	IMF
food_non-alcoholic_beverages_weight_percent	Weight percentage of the consumer price index of food non-alcoholic beverages	CPI dataset	IMF
furnishings_household_equipment_routine_household_maintenance	Consumer price index of furnishings household equipment routine household maintenance	CPI dataset	IMF
furnishings_household_equipment_routine_household_maintenance_weight_percent	Weight percentage of the consumer price index of furnishings household equipment routine household maintenance	CPI dataset	IMF
health	Consumer price index of health	CPI dataset	IMF
health_weight_percent	Weight percentage of the consumer price index of health	CPI dataset	IMF
housing_water_electricity_gas_other_fuels	Consumer price index of housing water electricity gas other fuels	CPI dataset	IMF
housing_water_electricity_gas_other_fuels_weight_percent	Weight percentage of the consumer price index of	CPI dataset	IMF

	housing water electricity gas other fuels		
miscellaneous_goods_services	Consumer price index of miscellaneous goods services	CPI dataset	IMF
miscellaneous_goods_services_weight_percent	Weight percentage of the consumer price index of miscellaneous goods services	CPI dataset	IMF
recreation_culture	Consumer price index of recreation culture	CPI dataset	IMF
recreation_culture_weight_percent	Weight percentage of the consumer price index of recreation culture	CPI dataset	IMF
restaurants_hotels	Consumer price index of restaurants hotels	CPI dataset	IMF
restaurants_hotels_weight_percent	Weight percentage of the consumer price index of restaurants hotels	CPI dataset	IMF
transport	Consumer price index of transport	CPI dataset	IMF
transport_weight_percent	Weight percentage of the consumer price index of transport	CPI dataset	IMF
agr_raw_material_index	Global commodity price index of agr raw material	PCPS dataset	IMF
agriculture	Global commodity price index of agriculture	PCPS dataset	IMF
all_index	Global commodity price index of all	PCPS dataset	IMF
all_metals_ex_gold_index	Global commodity price index of all metals ex gold	PCPS dataset	IMF
all_metals_index	Global commodity price index of all metals	PCPS dataset	IMF

apsp_crude_oil_us_dollar_per_bbl	Global commodity price index of apsp crude oil us dollar per bbl	PCPS dataset	IMF
beverages_index	Global commodity price index of beverages	PCPS dataset	IMF
coal_index	Global commodity price index of coal	PCPS dataset	IMF
commodities_for_index_all_excluding_gold	Global commodity price index of commodities for all excluding gold	PCPS dataset	IMF
energy_index	Global commodity price index of energy	PCPS dataset	IMF
fertilizer	Global commodity price index of fertilizer	PCPS dataset	IMF
food_and_beverage_index	Global commodity price index of food and beverage	PCPS dataset	IMF
food_index	Global commodity price index of food	PCPS dataset	IMF
industrial_materials_index	Global commodity price index of industrial materials	PCPS dataset	IMF
metal_index	Global commodity price index of metal	PCPS dataset	IMF
natural_gas_index	Global commodity price index of natural gas	PCPS dataset	IMF
non_fuel_index	Global commodity price index of non fuel	PCPS dataset	IMF
precious_metals_price_index	Global commodity price index of precious metals	PCPS dataset	IMF

Data Profile:

Variables	Time component	Structured/unstructured	Qualitative/quantitative
country_name	Time_invariant	Structured	Qualitative (Nominal)

month_code	Time_invariant	Structured	Qualitative (Ordinal)
year	Time_invariant	Structured	Qualitative (Ordinal)
month_number	Time_invariant	Structured	Qualitative (Ordinal)
month	Time_invariant	Structured	Qualitative (Ordinal)
alcoholic_beverages_tobacco_narcotics	Time_invariant	Structured	Quantitative (Continuous)
alcoholic_beverages_tobacco_narcotics_weight_percent	Time_invariant	Structured	Quantitative (Continuous)
clothing_footwear	Time_invariant	Structured	Quantitative (Continuous)
clothing_footwear_weight_percent	Time_invariant	Structured	Quantitative (Continuous)
communication	Time_invariant	Structured	Quantitative (Continuous)
communication_weight_percent	Time_invariant	Structured	Quantitative (Continuous)
consumer_price_index_all_items	Time_invariant	Structured	Quantitative (Continuous)
education	Time_invariant	Structured	Quantitative (Continuous)
education_weight_percent	Time_invariant	Structured	Quantitative (Continuous)
food_non-alcoholic_beverages	Time_invariant	Structured	Quantitative (Continuous)
food_non-alcoholic_beverages_weight_percent	Time_invariant	Structured	Quantitative (Continuous)
furnishings_household_equipment_routine_household_maintenance	Time_invariant	Structured	Quantitative (Continuous)
furnishings_household_equipment_routine_household_maintenance_weight_percent	Time_invariant	Structured	Quantitative (Continuous)
health	Time_invariant	Structured	Quantitative (Continuous)
health_weight_percent	Time_invariant	Structured	Quantitative(Continuous)
housing_water_electricity_gas_other_fuels	Time_invariant	Structured	Quantitative(Continuous)
housing_water_electricity_gas_other_fuels_weight_percent	Time_invariant	Structured	Quantitative(Continuous)
miscellaneous_goods_services	Time_invariant	Structured	Quantitative(Continuous)
miscellaneous_goods_services_weight_percent	Time_invariant	Structured	Quantitative(Continuous)
recreation_culture	Time_invariant	Structured	Quantitative(Continuous)
recreation_culture_weight_percent	Time_invariant	Structured	Quantitative(Continuous)
restaurants_hotels	Time_invariant	Structured	Quantitative(Continuous)
restaurants_hotels_weight_percent	Time_invariant	Structured	Quantitative(Continuous)

transport	Time_invariant	Structured	Quantitative(Continuous)
transport_weight_percent	Time_invariant	Structured	Quantitative(Continuous)
agr_raw_material_index	Time_invariant	Structured	Quantitative(Continuous)
agriculture	Time_invariant	Structured	Quantitative(Continuous)
all_index	Time_invariant	Structured	Quantitative(Continuous)
all_metals_ex_gold_index	Time_invariant	Structured	Quantitative(Continuous)
all_metals_index	Time_invariant	Structured	Quantitative(Continuous)
apsp_crude_oil_us_dollar_per_bbl	Time_invariant	Structured	Quantitative(Continuous)
beverages_index	Time_invariant	Structured	Quantitative(Continuous)
coal_index	Time_invariant	Structured	Quantitative(Continuous)
commodities_for_index_all_excluding_gold	Time_invariant	Structured	Quantitative(Continuous)
energy_index	Time_invariant	Structured	Quantitative(Continuous)
fertilizer	Time_invariant	Structured	Quantitative(Continuous)
food_and_beverage_index	Time_invariant	Structured	Quantitative(Continuous)
food_index	Time_invariant	Structured	Quantitative(Continuous)
industrial_materials_index	Time_invariant	Structured	Quantitative(Continuous)
metal_index	Time_invariant	Structured	Quantitative(Continuous)
natural_gas_index	Time_invariant	Structured	Quantitative(Continuous)
non_fuel_index	Time_invariant	Structured	Quantitative(Continuous)
precious_metals_price_index	Time_invariant	Structured	Quantitative(Continuous)

Completeness, Uniqueness and Timeliness:

Variables	Completeness	Uniqueness	Data Grain	Timeliness
country_name	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
month_code	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available

year	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
month_number	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
month	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
alcoholic_beverages_tobacco_narcotics	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
alcoholic_beverages_tobacco_narcotics_weight_percent	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
clothing_footwear	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
clothing_footwear_weight_percent	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
communication	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
communication_weight_percent	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
consumer_price_index_all_items	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available

education	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
education_weight_percent	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
food_non-alcoholic_beverages	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
food_non-alcoholic_beverages_weight_percent	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
furnishings_household_equipment_routine_household_maintenance	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
furnishings_household_equipment_routine_household_maintenance_weight_percent	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
health	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
health_weight_percent	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
housing_water_electricity_gas_other_fuels	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
housing_water_electricity_gas_other_fuels_weight_percent	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available

miscellaneous_goods_services	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
miscellaneous_goods_services_weight_percent	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
recreation_culture	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
recreation_culture_weight_percent	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
restaurants_hotels	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
restaurants_hotels_weight_percent	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
transport	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
transport_weight_percent	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
agr_raw_material_index	No mising data	No duplicates	year, month	Data until December 2021 is available
agriculture	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
all_index	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available

all_metals_ex_gold_index	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
all_metals_index	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
apsp_crude_oil_us_dollar_per_bbl	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
beverages_index	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
coal_index	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
commodities_for_index_all_excluding_gold	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
energy_index	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
fertilizer	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
food_and_beverage_index	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
food_index	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available

industrial_materials_index	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
metal_index	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
natural_gas_index	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
non_fuel_index	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available
precious_metals_price_index	No mising data	No duplicates	country_name, year, month	Data until December 2021 is available

Section-3: Questions to Explore

Questions to explore:

1) How do the fluctuations in global commodity prices impact the consumer price indexes in different countries?
2) Which commodity price impacts the CPIs the most?
3) Does the impact of commodity prices on CPIs vary across the countries?