

IMPACT OF COVID-19 ON FOOD SECURITY

UNDERSTANDING THE DATASET:

The first step to learn about the data is to thoroughly **understand the data**, So let us first see how our dataset looks like:

This is just a small overview of the original dataset:

	Date	Food Price Index	Meat Price Index	Dairy Price Index	Cereals Price Index	Oils Price Index	Sugar Price Index
0	Jan-90	108.7	112.3	94.3	106.4	73.0	201.5
1	Feb-90	109.9	117.7	91.9	104.0	72.5	207.9
2	Mar-90	107.9	119.6	73.6	102.1	74.6	218.0
3	Apr-90	114.2	131.0	85.2	105.1	71.8	216.3
4	May-90	111.2	130.5	70.0	105.3	74.4	207.2

We can clearly understand the data is about Food Price Index with change in time and we need to understand how the price is changing?

The attributes are clearly some examples of daily commodities like Meat, Dairy, Cereals, Oil and Sugar which is the main source of food worldwide.

To get **insight of the changing pattern** it is therefore advised to use Data Visualization Tool.

IMPACT ON THE DATASET:

Due to covid 19 the world is under huge economic crisis and surely we are interested to look at 2019-2020 data instead of what happened in the 90's. Surely we would do that but in this attempt we are trying to predict what could have been the price if there were nothing Called COVID that affected the mass. For that reason we are looking into the

2019-2020 data separately. **Finally we would compare the real price index and the predicted price index to check the impact.**

Let us look at a small overview of The same dataset after lockdown was imposed worldwide. Our dataset actually contain **365 rows with 9 columns.**

```
df.shape
```

```
(365, 9)
```

	Date	Food Price Index	Meat Price Index	Dairy Price Index	Cereals Price Index	Oils Price Index	Sugar Price Index
0	19-Aug	169.7	179.6	194.5	157.8	133.9	174.8
1	19-Sep	169.2	179.6	193.4	157.4	135.7	168.6
2	19-Oct	172.0	180.7	192.0	164.3	136.4	178.3
3	19-Nov	176.8	189.7	192.6	162.1	150.6	181.6
4	19-Dec	181.5	190.8	198.9	164.4	164.7	190.3
5	20-Jan	183.0	183.8	200.6	169.3	176.3	200.7
6	20-Feb	179.5	175.4	209.8	167.9	158.1	209.7
7	20-Mar	171.1	173.2	203.5	164.2	139.1	169.6
8	20-Apr	165.6	169.3	196.2	163.7	131.8	144.9
9	20-May	162.5	168.0	181.8	162.2	128.1	155.6

It is clearly evident from the data it **depicts the details of the same attribute after COVID was suspected in the mid November** in the city of Wuhan(CHINA).

We would closely analyze the change of the pattern of this dataset using various analyzing tools.

Let us check the summary of this dataset:

	Food Price Index	Meat Price Index	Dairy Price Index	Cereals Price Index	Oils Price Index	Sugar Price Index	Average	DATE_VAL
count	365.000000	365.000000	365.000000	365.000000	365.000000	365.000000	365.000000	365.000000
mean	142.161096	138.213699	143.945205	139.193699	134.906027	181.725479	146.690868	182.000000
std	42.523955	30.904489	56.995830	49.266329	51.583398	71.960425	46.452825	105.510663
min	85.100000	84.500000	66.100000	80.200000	57.100000	72.900000	84.283333	0.000000
25%	106.400000	115.200000	94.500000	99.100000	95.300000	130.600000	107.316667	91.000000
50%	125.500000	130.600000	126.600000	117.800000	126.500000	168.900000	133.550000	182.000000
75%	172.900000	167.500000	193.500000	166.700000	164.400000	215.700000	181.216667	273.000000
max	240.100000	212.000000	275.400000	267.700000	286.500000	420.200000	268.883333	364.000000

From the above summary we can go into the details of each attribute for example, The minimum food price index was 85.0 and maximum was 240.0

The standard deviation and average is also important for the attributes as it tells the range in which the data belongs to.

SUPPLY AND UTILIZATION:

The intake and the production of food worldwide is of utmost importance,

hence we need to keep in mind the Supply and the Utilization in the changing period.

The Supply and utilization of the same is depicted in the following dataset.

Let us for example have a look at the Wheat supply and utilization worldwide:

WORLD WHEAT MARKET							
	Production ^{1/}	Supply ^{2/}	Utilization	Trade ^{3/}	Ending stocks ^{4/}	World stock-to-use ratio	Major exporters' stock-to-disappearance ratio ^{5/}
	(..... million tonnes)					(..... percent)	
2011/12	699.0	902.1	693.2	149.2	203.9	29.9	18.8
2012/13	658.6	862.5	682.4	143.6	185.8	26.8	15.1
2013/14	715.3	901.1	692.2	159.4	200.4	28.3	16.0
2014/15	735.2	935.6	708.2	156.6	228.7	31.9	18.8
2015/16	737.2	965.9	717.0	167.5	243.2	33.0	18.0
2016/17	764.9	1,008.0	737.3	176.9	267.0	36.1	19.8
2017/18	761.6	1,028.5	739.1	177.4	288.1	38.4	21.0
2018/19	732.1	1,020.2	751.1	168.2	271.9	35.9	18.1
2019/20	762.2	1,034.1	757.5	175.1	276.2	36.6	16.1
2020/21	758.3	1,034.5	754.3	177.5	280.3	36.3	15.7

The above dataset refers to an insight of production , supply and utilization worldwide which are some of the important parameters we will be dealing with in the coming procedures.

Similarly like this Wheat market dataset, there are similar certain product dataset such as Rice, Coarse Grain , Cereals, etc. which we have combined to form a product supply dataset.

Many attributes such as **Trade and Ending stocks are irrelevant** to our project hence it is suitable for us to omit the irrelevant data.

Now as we are aware of all the details and attributes as well as what the dataset depicts we are all set to analyze our data using necessary tools.

DATA MINING:

Removing of the NaN values :

In the mining process, we have seen that whether there are any null values present in our datasets or not. It is important enough to know the actual values and data's of the datasets and replace the non necessary ones.

So, through our first dataset of "food price indices_data" , we found out that :

```
df.isnull().sum()
```

```
Date          0
Food Price Index  0
Meat Price Index  0
Dairy Price Index  0
Cereals Price Index  0
Oils Price Index  0
Sugar Price Index  0
dtype: int64
```

This depicts that there are **no Null Values** present in this dataset of "food price_indices" .

As the shape of our dataset is very minimal there is less scope to have NaN Values inside the attributes present.

The dataset attributes in the supply and utilization part is little critical hence we come down **reducing to the minimal important attributes** required and hence the dataset now turns out to be maximum relevance.

This is what the dataset after the data cleaning looks like

This is just a small overview of the complete dataset present:

	PRODUCT	DATE	Production	Supply	Utilization
0	CEREAL	2011/12	2357.6	2922.1	2321.0
1	CEREAL	2012/13	2317.9	2915.0	2332.2
2	CEREAL	2013/14	2557.5	3151.5	2449.1
3	CEREAL	2014/15	2607.9	3281.5	2507.8
4	CEREAL	2015/16	2583.2	3353.9	2550.5

We have cut down the irrelevant attributes and this procedure is known as data cleaning.

It is important to note the datatypes of the attributes we are working with:

The datatypes are as follows:

```
df.dtypes
```

```
PRODUCT      object
DATE          object
Production    float64
Supply        float64
Utilization   float64
dtype: object
```

```
df.dtypes
```

```
Date                object
Food Price Index    float64
Meat Price Index    float64
Dairy Price Index   float64
Cereals Price Index float64
Oils Price Index    float64
Sugar Price Index   float64
dtype: object
```

As we are now aware of all the types of data that are present now **this is the time to understand the pattern of the attributes categorically** and hence we go for visualization to understand how to treat each attribute and come up with a perfect algorithm for training our model.

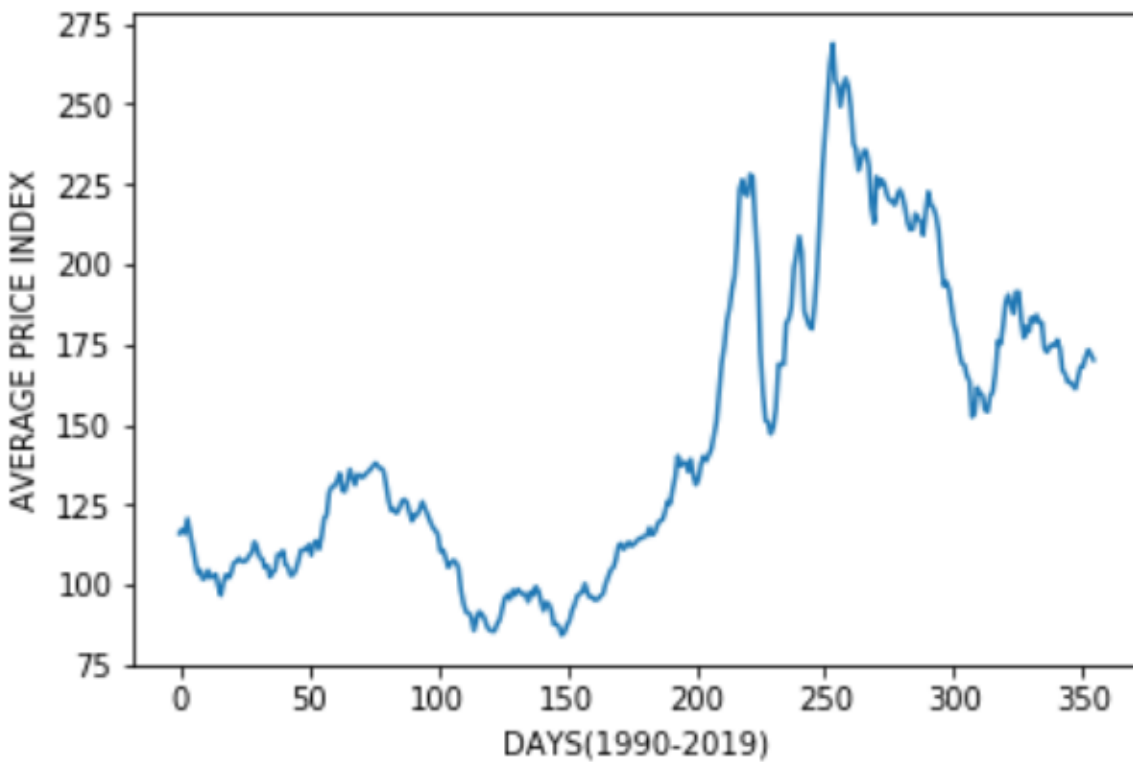
Data Mining and cleaning is one of the most relevant step while dealing with large datas and CSV files.

DATA VISUALIZATION:

With increase in the Days the price of Food index is necessary to **track the chnages of the pattern** in which the price rises or falls.

Considering from **the year 1990-2019** the average price index is calculated and then is plotted against the year with corresponding month.

```
Text(0, 0.5, 'AVERAGE PRICE INDEX')
```



As we can see there is a discontinuous trough against the price hence we need to be more specific towards our Products.

So in order to be specific we are now visualizing four products which are the most important commodities and vary that with **three criteria**:

1) Supply

2) Utilization

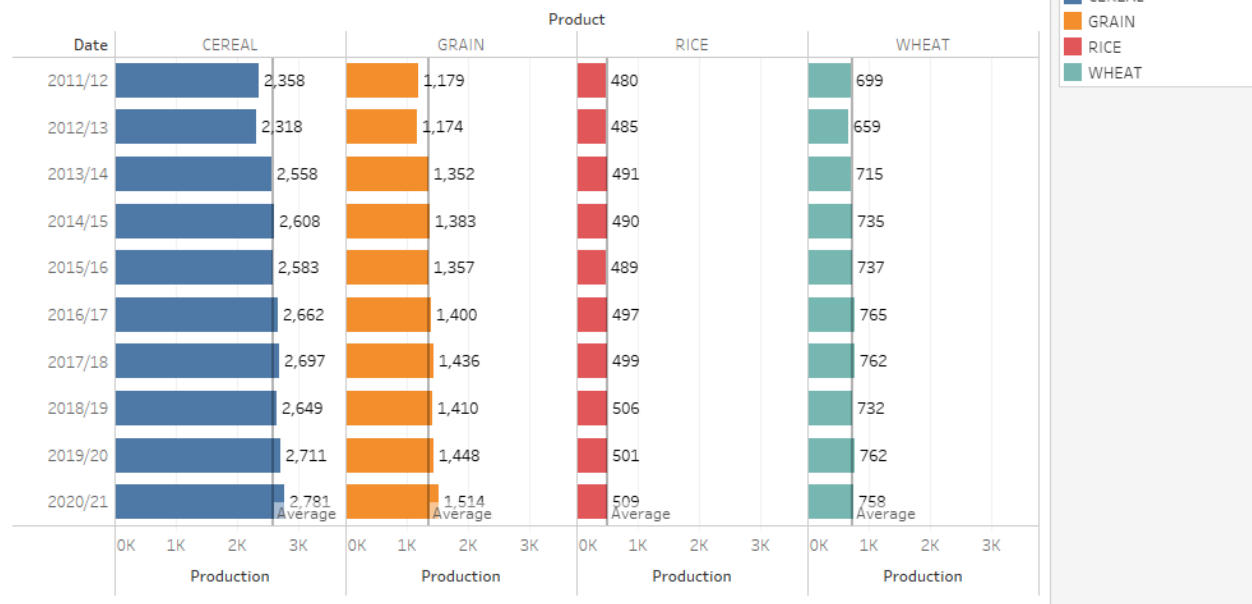
3) Production



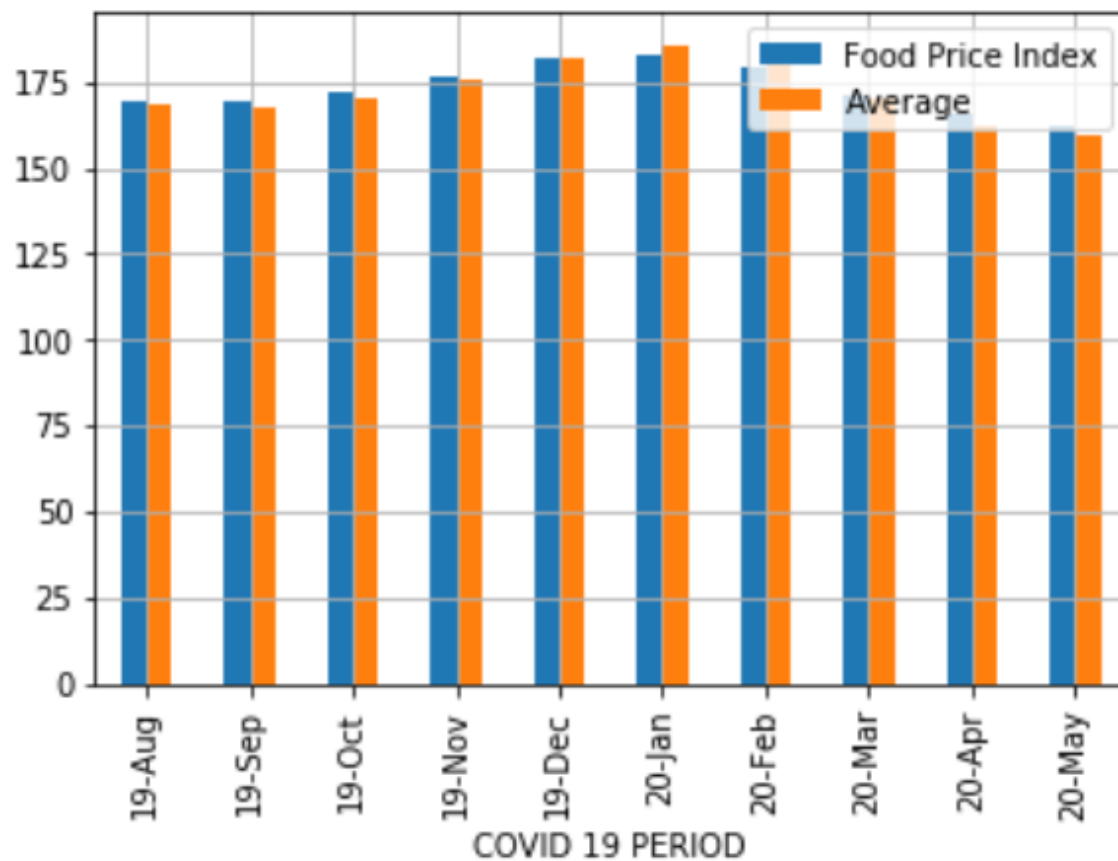
Tableau is used to visualize the above data and this depicts clearly how each product has its production/ supply and utilization.

Now it is also very crucial to note the supply and other parameters with respect to change of year hence the following visual depicts the four criteria among four essential commodity over the year 2011-2020

product VS production

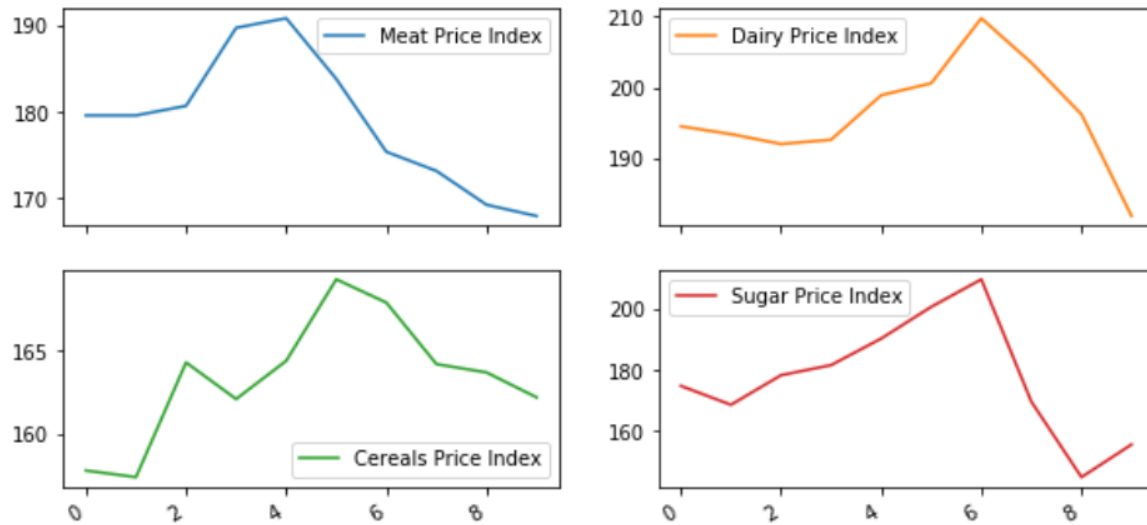


Impact of the covid 19 worldwide is one thing we are supposed to see hence magnifying the dataset and checking the part of Food price index as well as Average from the period of 2019 August-2020 May.



As we can see the food price has not much changed since the world went into lockdown, although there is a little increase in the price index as compared to the initial lockdown period.

The categorical plot of sugar, cereal, meat and dairy is also necessary. The same is visualized using a subplot



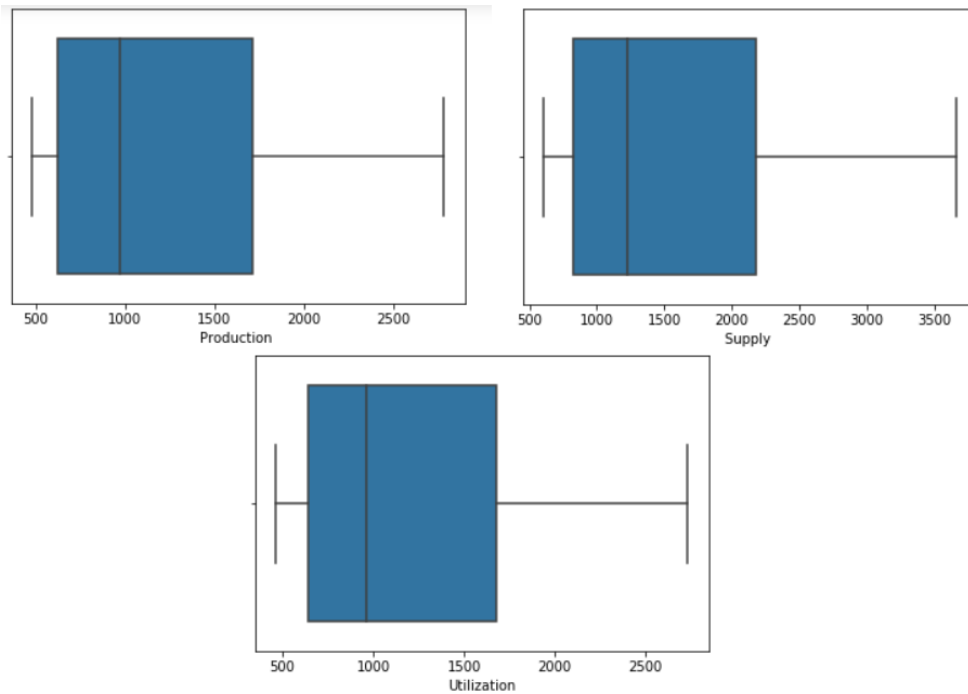
Data Visualization depicted the use of Attributes to correlate and depict the change with respect to some important pattern.

OUTLIER DETECTION

The presence of outlier in a data attribute means a data point which is far more bigger than the average value of all the data points.

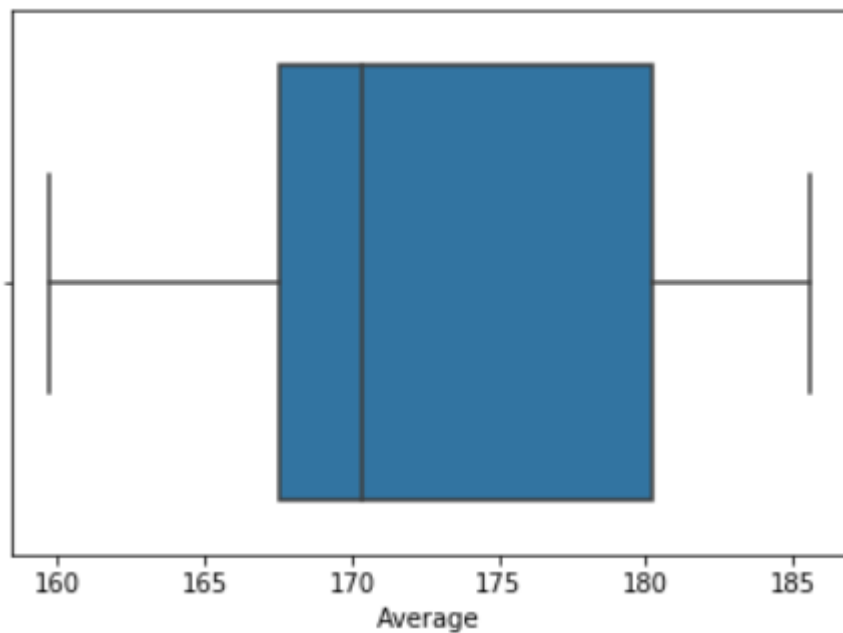
Presence of this outlier actually decreases the accuracy of the predicted value hence it is advisable to remove the outlier before hand to avoid any disturbance.

Let us check the outliers of our data using a box plot.



It is clearly evident from the boxplot that there are **no outliers** in the supply,utilization and production dataset.

Average food price indices boxplot is as follows:



Hence there are no outliers.

TRAINING AND TESTING

The Training and testing of the data is done to ensure the accuracy calculated is genuine and is compared or predicted with unknown data.

```
x=df_lock[["Lockdown_Month"]]  
y=df_lock[["Average"]]
```

```
from sklearn.model_selection import train_test_split  
x_train,x_test,y_train,y_test = train_test_split(x, y, test_size=0.3, random_state=42)
```

Keeping the Lockdown Period Month starting from August 2019 and keeping the Average Food Price Indices as the independent variable we separate the 2 variable.

We do the splitting of the dataset in 3(Test) : 10(Train) Ratio to ensure more data points are available for training hence better accuracy is achieved.

The Train Test dataset are as follows

Lockdown_Month		Average	
0	0	0	168.383333
7	7	7	170.116667
2	2	2	170.616667
9	9	9	159.700000
4	4	4	181.766667
3	3	3	175.566667
6	6	6	183.400000

x-Train **y-Train**

Lockdown_Month		Average	
8	8	8	161.916667
1	1	1	167.316667
5	5	5	185.616667

x-Test **y-Test**

As the dataset is now split into the training and testing model it can be now fit with specific algorithm to get the desired accuracy.

MODEL FIT

While Training the model we found out that the two axis of training needs to be integer.

Using the index as the dependent variable was vague and monotonous hence it is better to convert the date month column into a specific integer which can be used further easily.

The logic is the date month column is present as 95-Sep we want this to be converted to something like 199509 where 1995 is the year and 09 is the month specified.

This conversion will help to input the data to be predicted easily like March 2017 will be 201703 hence using various mathematical and pandas tools we come up with the dateMon Column which looks something like:

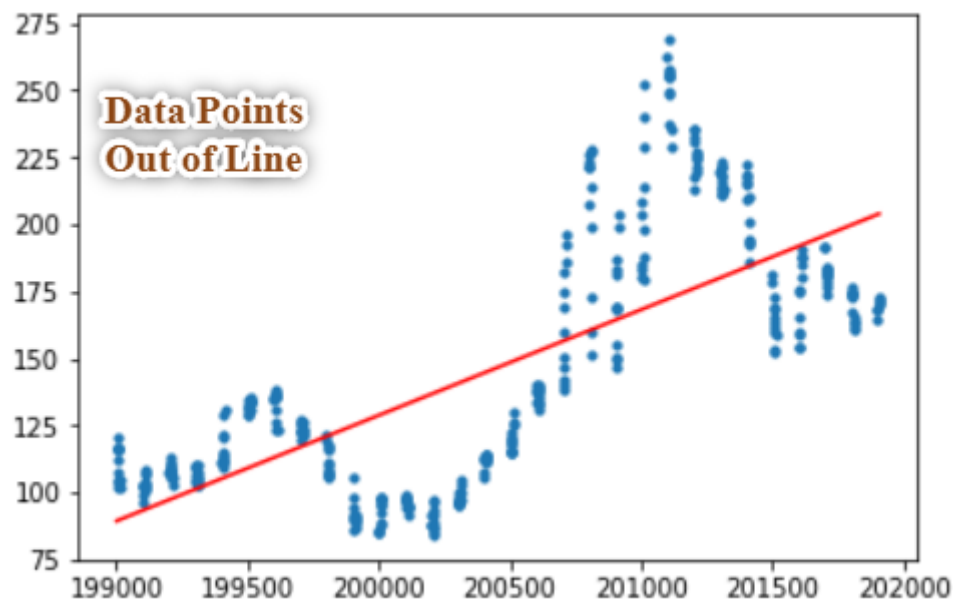
	Date	Food Price Index	Meat Price Index	Dairy Price Index	Cereals Price Index	Oils Price Index	Sugar Price Index	DateMod	Average
0	Jan-90	108.7	112.3	94.3	106.4	73.0	201.5	199001	116.033333
1	Feb-90	109.9	117.7	91.9	104.0	72.5	207.9	199002	117.316667
2	Mar-90	107.9	119.6	73.6	102.1	74.6	218.0	199003	115.966667
3	Apr-90	114.2	131.0	85.2	105.1	71.8	216.3	199004	120.600000
4	May-90	111.2	130.5	70.0	105.3	74.4	207.2	199005	116.433333

See the transformation which is the DateMod Attribute.

Now we are going to use it to train the model with respect to the dependent variable Average.

Using Linear Regression the line was not able to fit the data Points:

Something like this:



Hence we need to implement regression with much higher order more specifically, polynomial regression.

We know linear regression is : $Y = Mx + c$

While Polynomial Regression line is : $Y = Mx + Nx^2 + Ox^3 + Px^4 + \dots + Zx^n$

where n is the order of line.

We are using order 2 in order to fit the data points in a line :

After implementing the polynomial regression model the data looks something like:



Due to much change in the uniformity of the data the accuracy of the model is comparatively low although the predicted Price index is within the Range Of The Average of All the datapoints for all the Year.

Now the model is ready to predict The Food Price Index for certain month of Certain year, Which may be incase affected after the lockdown.

Suppose We want to predict the Food price index for the Coming October Month:

It comes out to be,

```
pipe.predict([[202010]])
```

```
array([[216.42567533]])
```

,where 2020 is the year and 10 depicts october.

It can be seen from the dataset that the average of the food price index is 160 hence the food price index is seen to be very high.

The **FAO Food Price Index (FFPI)** averaged **162.5 points in May 2020, down 3.1 points (1.9 percent) from April** and reaching the lowest monthly average since December 2018. With the continued negative economic effects of COVID-19, the FFPI has been on a **downward trend for four consecutive months**. The latest drop in May reflects falling values of all the sub-indices with the exception of sugar, which rose for the first time in three months.

Consumer Price Index (CPI) is designed to measure the changes over time in general level of retail prices of selected goods and services that households purchase for the purpose of consumption. Such changes affect the real purchasing power of consumers' income and their welfare. The CPI measures price changes by comparing, through time, the cost of a fixed basket of commodities.

The Annual percentage change in a CPI is used as a measure of inflation.

MODEL DEPLOYMENT:

The model fit is now ready for user to use and interpret through relevant references.

The model is deployed in the **IBM Cloud** using **Watson Deployment API Client**.

The UI is made using NODE RED with a form having the Year as input and in correspondance the output is the Food price index for that year.

The Deployed Model can now be designed via various media nodes and can be used in preference with all tools required.

FINAL PROJECT REPORT

1. INTRODUCTION :

1.1. **OVERVIEW** : The Team made the project about the **Impact on Food Security** due to the unalarming pandemic all over the world spreading with huge grasp. It is of utmost importance to study the food data and interpret with it the observed data after pandemic. Thus, we could be aware of any affected portion if any in the production, supply, utilization of food trade worldwide. People all over the world will surely face a recession and hence there will be scarcity of the available food. Getting alarmed through facts will help the government to take any action and be prepared of any circumstances in further future (if any).

***All the sources in making the project are authentic and depicts original values for the observed dataset.**

With advancement and introduction of machine learning tools it is clearly possible to predict any future any downfall or increase in demand supply or utilization, The project aim revolves around every possible way to make a clear understanding of the visual data and compare the prediction data with actual data.

For further more detail visit the following link below:

Visit the FAOSAT site for more info

1.2. **PURPOSE** : The goal and usage of a machine learning project is to build new or leverage existing algorithms to learn from data, in order to build generalizable models that give accurate predictions, or to find patterns, particularly with new and unseen similar data.

The purpose of this project is to set up a mathematically efficient predictive model which will further do predictive data analytics to predict future events on the food security field. And , moreover , the use of predictive models always gives a better optimization in risk of decision to further analyse the future problems more carefully.

2. LITERATURE SURVEY:

2.1. EXISTING PROBLEMS: Alarmed by a **potential rise in food insecurity** during the COVID-19 pandemic, many countries and organizations are mounting special efforts to keep agriculture safely running as an essential business, markets well supplied in affordable and nutritious food, and consumers still able to access and purchase food despite movement restrictions and income losses.

The primary risks to food security are at the country level: as the coronavirus crisis unfolds, disruptions in domestic food supply chains, other shocks affecting food production, and loss of incomes and remittances are creating strong tensions and food security risks in many countries.

The **United Nations World Food Programme** has warned that an estimated [265 million people could face acute food insecurity by the end of 2020](#), up from 135 million people before the crisis, because of income and remittance losses.

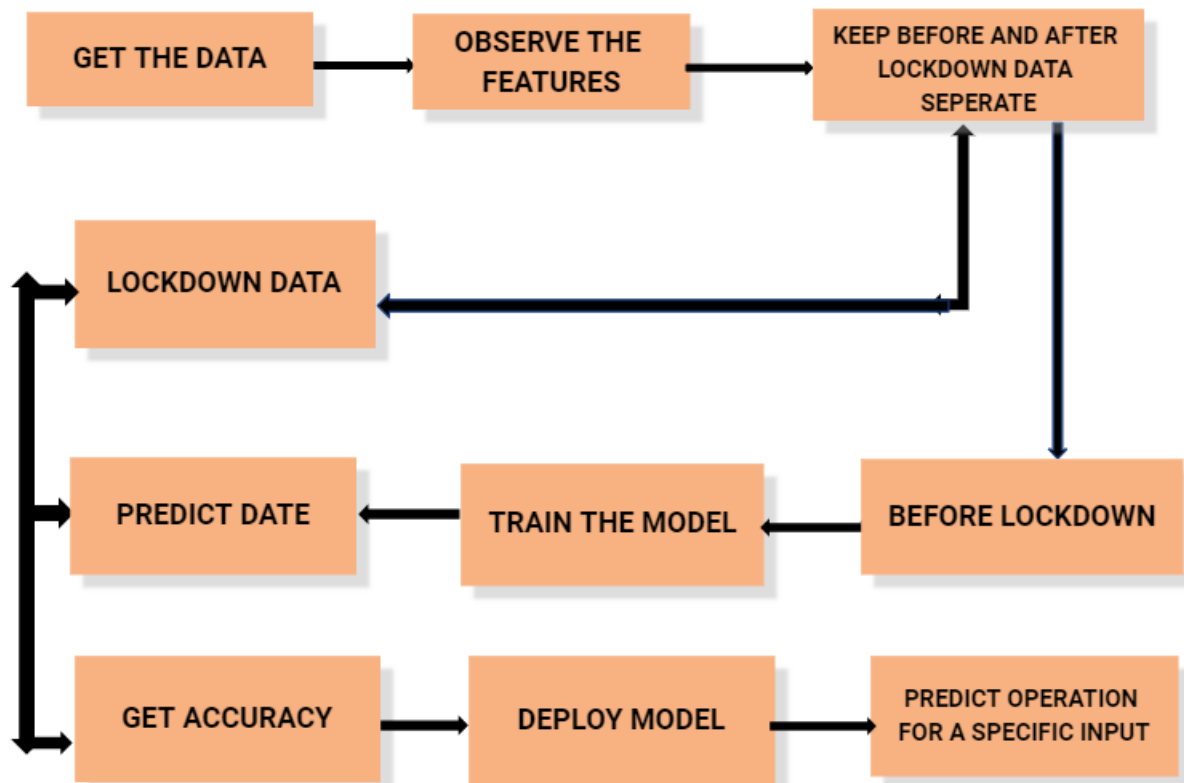
2.2. PROPOSED SOLUTION: The proposed solution for this existing problem can be diminished by the modern implementation of **predictive models of Machine Learning and Deep Learning** .

The Team's solution is to create a machine learning model to predict supply – demand gaps, Supply chain impact food security of people and provide a dashboard to show how the extended lockdowns / post lockdown situations can impact the food security of people .

This will further help the countries to be ready or predict any certain pandemics if come in near future which could hamper the economy of the nation.

3. THEORETICAL ANALYSIS:

3.1. **BLOCK DIAGRAM** : The Block Diagram of our project simply depicts our proposed solution to deal with the problem :

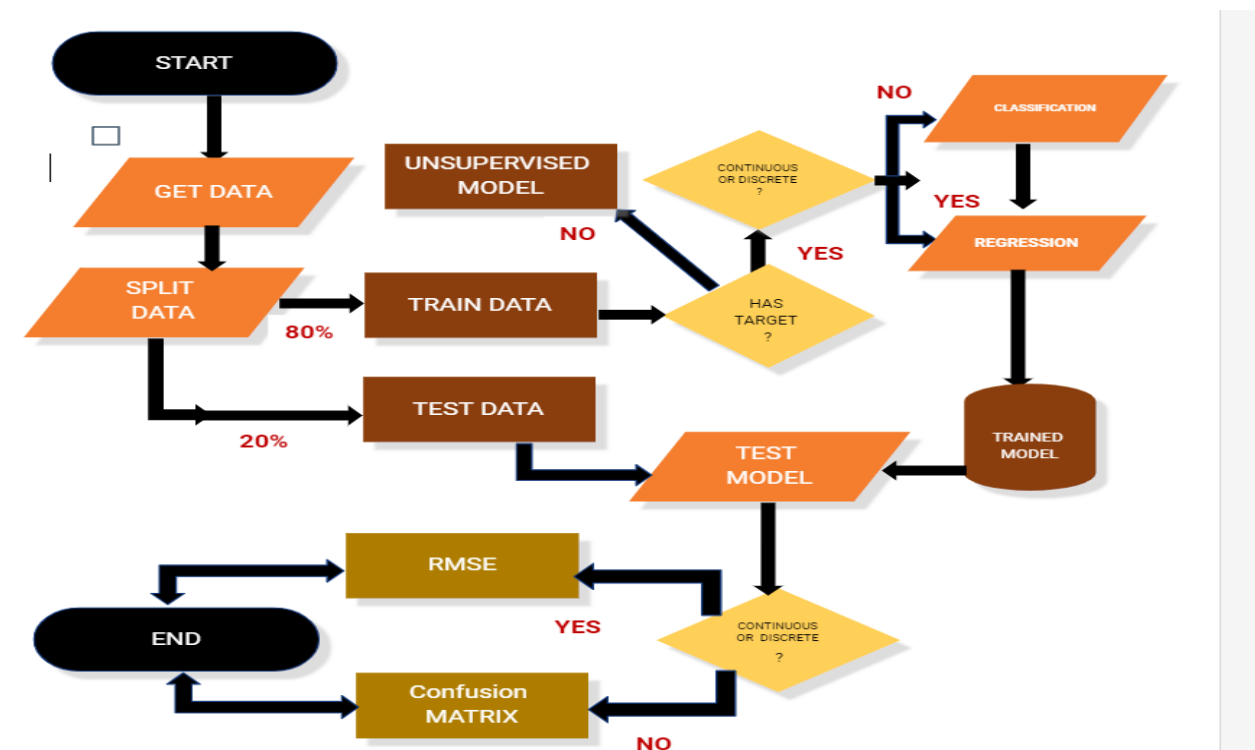


3.2. **SOFTWARE DESIGNING** : For the software designing purpose , this project uses the **NODE-RED** programming software to create a Visualization Dashboard.

Node-RED is a programming tool for flow based wiring together of Internet of Things, APIs and online services in new and interesting ways. It provides a browser-based editor that makes it easy to wire together flows using the wide range of **nodes** in the palette that can be deployed to its runtime in a single-click.

Integrating Machine Learning with Node Red enables to make our **IoT devices** learn from their environment or to build advanced analytics solutions leveraging the wealth of data coming from our IoT device mesh.

5. FLOWCHART: The project Machine Learning Model Flowchart comes as follows:



4.EXPERIMENTAL INVESTIGATION :

During the commencement of investigation we got to know about various parameters on which the food security depends worldwide.

The parameters are of utmost importance and need to be analyzed upon a certain period of time.

The major parameters we noticed and used in our project are:

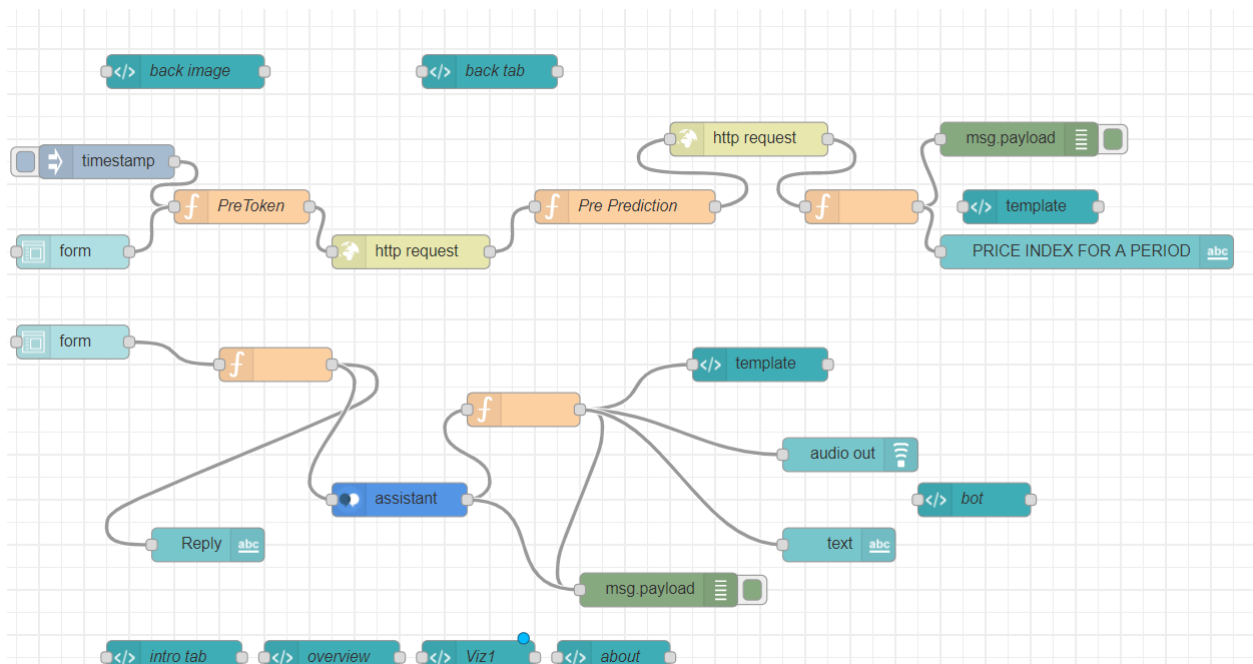
- a) Supply
- b)Utilization
- c)Demand

Apart from that , further more investigation was required on the several softwares that were used for the experiment to perform such as:

- IBM Cloud
- Watson Studio
- Watson Assistant
- Node Red Application

The Node RED Application is of great help for desinging the UI for our application.

The flow of the of all the nodes are given below:



There are in total 26 nodes for different purpose:

1. Template nodes are used for UI design using html.
2. Assistant node is integrating the Watson Assistant with the model for chatbot.
3. Function nodes are used to define the work of the inject nodes.
4. Http nodes are for providing http server into the Bluemix.
5. The msg payload is used to check for debug options in the node red palette.
6. Form node is used to take input from user in a format

6. RESULT:

Prediction Model :

The Food Price Index for a given Month of a given year is predicted using our trained model.

What does the Predicted Value Say?

The Average as seen is a Value(146) marked in the chart. Keep a note of the same.

- If Predicted Value > Average Value : Supply Production Needs to be Improved to meet demand.
- If Predicted Value = Average Value : Supply Production And Utilization is balanced.
- If Predicted Value < Average Value : Supply is enough but utilization needs to increase.

7. ADVANTAGES:

- ✓ The Model is capable of Training any dataset with input as a period specified hence it is independent of any different data point.
- ✓ Outliers are minimized hence out of range data points will not hamper the accuracy
- ✓ The Model has visualization dashboard and can be used to be aware of any changes in the price index of the major commodity.

- ✓ The Chatbot in the dashboard will be of great help for any frequently asked question answers.

8. DISADVANTAGES:

- ✓ The Biggest Disadvantage of the Model is it does not keep in mind any external factor while predicting the Food Indices price hence any change in the external issue the model will be of no help.
- ✓ The Model has a range of data trained and will give inaccurate result after a certain period of time but this problem can be solved by training the model distinctly every year or every month.
- ✓ The input format of date is not reliable and can take any value (like month is always ending by 12 but it would take any random value)

9. APPLICATIONS:

- ✓ The web application can be deployed with any software or app and can be used to predict the price index if required.
- ✓ The Government worldwide can use the dashboard in order to make its citizen no about the demand supply gap and to counter about it.
- ✓ The Dealers and factory producers will know about any change in the index and can implement their investment in business likewise.

10. CONCLUSION:

Looking back at what we have learnt and discussed about during this project we are ready to begin understanding the global food

security issue. We should remember that we want di-verse perspective of the issue , as well as global and local perspective . This model for which will be a reason for the global improvement in the food security levels.

Further , any kind of risk which is related to the food index could be earlier resulted by the model and UI made.

11.FUTURE SCOPE:

This project can be integrated to any other web application for further better use purpose as it was be a sub module for any other big prediction based application. It also will help in understanding the future of food security in the whole world.

The project review addresses the food science and technology roles in meeting current challenges and investigates possible solutions to feed the world in the near future. Its never too late to discover anything , this model will perhaps proof to deliver better results if any such global crises will hit the world in the near future.

12. BIBLIOGRAPHY:

- References:

1. Food Security Model Dataset:
<http://www.fao.org/home/en/>

2. IBM Cloud : <https://www.ibm.com/cloud/get-started>

3. IBM Watson Studio:

<https://cloud.ibm.com/catalog/services/watson-studio>

4. Watson Assistant:

<https://www.youtube.com/watch?v=6t8C0YRUGec&t=11218s>

5. Node Red : <https://nodered.org/>

13. APPENDIX:

- **Source Code:**

-

- **MODEL FIT CODE:**

```
# importing libraries for polynomial transform
from sklearn.preprocessing import PolynomialFeatures

# for creating pipeline
from sklearn.pipeline import Pipeline

# creating pipeline and fitting it on data
Input=[('polynomial',PolynomialFeatures(degree=2)),('modal',LinearRegression())]
```

```
pipe=Pipeline(Input)

pipe.fit(df[["DateMod"]],df[["Average"]])

poly_pred=pipe.predict(df[["DateMod"]])

#sorting predicted values with respect to predictor

sorted_zip = sorted(zip(df[["DateMod"]],poly_pred))

x_poly, poly_pred = zip(*sorted_zip)

#plotting predictions

plt.figure(figsize=(10,6))

plt.scatter(df[["DateMod"]],df[["Average"]],s=20)

plt.plot(df[["DateMod"]],y_pred,color='r',label='Linear Regression')

plt.plot(x_poly,poly_pred,color='g',label='Polynomial Regression')

plt.xlabel('Food Price',fontsize=16)

plt.ylabel('Average',fontsize=16)

plt.legend()

plt.show()
```

DEPLOYMENT CODE:

```
from watson_machine_learning_client import
    WatsonMachineLearningAPIClient

wml_credentials = {

    "apikey": "Ca9TDE1mjgkyP9p7COh7QLNeE7Uo3i0Jhbu9cGsO8n9v",

    "iam_apikey_description": "Auto-generated for key 8f43c563-b908-4e04-9988-f0ab893e935d",

    "iam_apikey_name": "Service credentials-2",

    "iam_role_crn": "crn:v1:bluemix:public:iam::::serviceRole:Writer",

    "iam_serviceid_crn": "crn:v1:bluemix:public:iam-identity::a/fd70f36980954d8aa1acf09207e59fde::serviceid:ServiceId-b391904d-fccc-4a9f-937c-54f68944fdc5",

    "instance_id": "5da4f613-4719-4a60-becb-a5b349a8bb24",

    "url": "https://eu-gb.ml.cloud.ibm.com"

}

client = WatsonMachineLearningAPIClient(wml_credentials)

model_props = {

    client.repository.ModelMetaNames.AUTHOR_NAME : "Srijani",

    client.repository.ModelMetaNames.AUTHOR_EMAIL
    : "srijanibhattacharjee123@gmail.com",

    client.repository.ModelMetaNames.NAME : " Food Data"

}
```

```
model_artifact = client.repository.store_model(pipe, meta_props =  
    model_props)
```

```
guid = client.repository.get_model_uid(model_artifact)
```

```
deploy = client.deployments.create(guid,name="Food Security  
    Prediction")
```
