

Abstract:

In this, what we are going to learn is how much worth of food is getting imported from other countries. In this dataset we are working on their food choices. They want to try different types of food and better quality. By seeing this Demand the US has started importing food from other countries. In the year the US imported worth \$177 billion in the year 2021. Every year the US imports more worth of food so to find out which type of food they are importing what they are importing and from where and how much worth it is. And from each country how much worth of food they are importing all that we can find from this dataset. And we can see this project will help us to understand all the situations.

In this, what we are going to learn is how much worth of food is getting imported from other countries. In this dataset we are going to see Visualization information, Data mining for business decision, Business Forecasting and Visualization, Quantitative methods for business, Business Intelligence and data analysis, Database management, Business data warehousing, Managerial Finance, Quantitative technique for financial Market, Investment analysis.

Outline of the Project

- Cover
- Abstract
- Introduction
- Data Description
- Methodology
- Findings and Analysis
 - Quantitative Methods For Business
 - Business Intelligence and Data Analysis
 - Database Management
 - Business data Warehousing
 - Business forecasting and Visualization
 - Visualizing Information
 - Data Mining for Business Decision
 - Elective 1 (Managerial Finance)
 - Elective 2 (Quantitative technique for financial Market)
 - Elective 3 (Investment Analysis)

Conclusion and recommendation

Introduction:

The US has so much demand for a variety of food products. As we know the US is the wealthiest country as they have many demands for a variety of food products. By this imports has risen 50%. Most of the imports are coming from Mexico, Canada and China.

As per the record the US has exported agricultural products worth \$177 billion in 2021 if we compare to previous year 2020 there is an increase of 18 percent. If we see 2012-2021 the imports are increasing more than exports. Because the US has more agricultural imports than exports.

We see that there is a rapid increase in the grains and feeds category by year to year. There is a 40 percent increase compared to 2020 in the form of money \$43 billion increased in 2021.

Animals, meat products also increased by 2020 there is 22 percent increase in 2021 in the form of money \$39 billion.

In 2021 from China the US exported more than 25 percent compared to 2020 and followed by Canada and Mexico. We can see that there are some disputes going on between China and the USA then also there is an increase in exports from China. Some other countries, Japan and other European countries are the top 5 countries of exports of the USA. The top 5 US import countries will share a value of 61 percent of trading.

US consumers totally depend on the imported agricultural products as there is high demand for the imported agricultural products . 95 percent spices, coffee, cocoa and fish are imported following fresh fruits and vegetables.

Data Description:

US food imports data that we are using has every detail of data of the past 21 years from 1999 to 2021. By this data we can compare the data with every year or for five years, ten years or from comparing for the year 1999 and 2021 for better understanding. In this data set we have the

Total Food:

In this data we can see every year how much worth of food is being imported from all over the world and total price for all types of food products.

Foods:

In this we have sub categories which are, meat, live meat animals, fish, dairy products, vegetables, oil, fruits, nuts and some more. In this we can see each type of food worth of it they imported. And we can see that for each type of food where they imported and how much worth of amount they imported from each country.

Subtotal:

So in this we have three categories and they are Animals, Plants and beverages. In each category we can see the total amount of the products cost they imported. For example in the Animal category it will be the combination of animal related products that will be there and they are like fish, meat and live meat animals.

Methodology:

The main aim of this dataset is to identify the type of product that the US is importing and the reason behind it and finding the solution for reducing the import from other countries. And to find the import increasing in each year and in which type of product. By this we need to compare the every year dataset. Though we aren't able to compare and analyze the data manually, for that purpose we are using some tools for the dataset to make things easy and for better understanding. And the software tools that we are using are:

Tableau:

With this software tool we can visualize the data very clearly and we can understand things easily. This software can help to compare the every year values and to analyze the problem or to identify the solution.

SQL:

In this SQL software will help to get the data that we are looking for in a very less time. And in this software we can work on updating the data and we can divide the data in groups and we can work on that group data for better understanding the data.

WEKA:

This software tool helps us to find the connections between dataset and we find how it is related to other data. By this we can find the solution in an easy process.

Excel:

This Excel tool we can use for the calculation and for the regression analysis. This tool helps us to analyze the financial situation. And how much has to be invested. Exactly it will help us to find the demand in the US for imported food.

Internal data :

The facts and information that come directly from the company's systems and are specific to the company in question.

External data:

External data is information that originates from outside the company and is readily available to the public. External data is used to help a company develop a better understanding of the world in which they are operating.

Quantitative Methods Business

Quantitative methods in business refer to processes and algorithms used to make decisions and predict outcomes. They are purely numbers-driven and do not account for individual employee capabilities or any soft calculations that vary based on departmental performance or personnel 1

Common quantitative methods include regression analysis, the use of probabilities, analyzing statistical data, linear programming, and data mining

Quantitative analysis is considered an indispensable tool in project management, production planning, marketing, finance, purchase, and inventory

The screenshot displays an Excel spreadsheet titled 'Alltables.xlsx - Read-Only'. The spreadsheet contains a table of U.S. food import values by food group from 2003 to 2021. The table is organized with columns for years and rows for food groups. A 'Solver Parameters' dialog box is open, showing the 'Set Objective' field set to '\$L\$11' and the 'By Changing Variable Cells' field set to '\$L\$11'. The 'Solving Method' is set to 'GRG Nonlinear'. The table data is as follows:

Food group	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003
Total foods 1/	166,946.5	146,406.1	141,720.7	139,587.6	131,143.3	123,153.3	121,117.9	119,666.2	117,861.5	114,391.6	110,428.9	106,473.6	102,518.7	98,563.8	94,608.9	90,654.0	86,699.1	82,744.2	78,789.3
Foods	2,299.9	2,158.0	2,253.4	2,029.4	2,016.3	2,103.3	2,774.9	3,009.9	2,173.9	1,673.3	1,139.4	1,278.0	1,243.7	1,187.6	1,138.7	1,089.9	1,041.1	992.2	943.3
Meats	13,196.0	10,389.4	9,668.7	9,251.2	8,875.2	8,587.2	9,991.7	8,940.3	5,243.7	5,752.0	5,718.5	4,426.9	4,243.7	4,060.5	3,877.3	3,694.1	3,510.9	3,327.7	3,144.5
Fish and shellfish 2/	24,198.5	21,285.7	21,797.7	22,272.6	21,324.2	19,257.8	18,513.6	20,051.5	13,112.3	11,840.2	11,106.3	10,859.9	10,613.5	10,367.1	10,120.7	9,874.3	9,627.9	9,381.5	9,135.1
Dairy	2,461.9	2,139.8	2,195.6	2,063.1	1,876.9	1,907.2	1,911.5	1,871.6	1,429.5	1,404.5	1,303.7	1,118.7	1,093.3	1,067.9	1,042.5	1,017.1	991.7	966.3	940.9
Vegetables	15,808.2	15,593.9	13,975.6	13,419.1	12,670.5	12,409.2	11,227.1	10,880.9	6,617.9	6,044.3	5,735.8	5,094.8	4,803.3	4,511.8	4,220.3	3,928.8	3,637.3	3,345.8	3,054.3
Fruits	22,695.9	20,497.1	20,315.5	19,539.3	18,384.6	17,157.8	15,954.8	14,807.7	7,833.2	7,003.6	6,047.3	5,631.4	5,215.5	4,799.6	4,383.7	3,967.8	3,551.9	3,136.0	2,720.1
Nuts	3,156.5	2,785.9	3,191.0	3,502.1	3,312.2	2,877.9	2,776.1	2,370.0	1,100.4	1,122.1	1,079.7	781.3	756.9	732.5	708.1	683.7	659.3	634.9	610.5
Coffee, tea, and spices	9,825.9	8,346.2	8,427.4	8,454.7	9,066.7	8,311.3	8,536.2	8,298.8	4,200.4	3,775.7	3,148.8	2,875.5	2,642.1	2,408.7	2,175.3	1,941.9	1,708.5	1,475.1	1,241.7
Grains	15,250.2	13,926.1	13,042.5	12,562.1	11,257.6	10,558.4	10,360.1	10,363.7	5,076.4	4,412.2	4,176.2	3,759.2	3,525.8	3,292.4	3,059.0	2,825.6	2,592.2	2,358.8	2,125.4
Vegetable oils	9,573.6	6,888.0	6,303.9	6,989.2	7,216.9	6,523.3	6,304.1	7,336.6	2,818.3	2,362.6	2,241.0	1,506.5	1,373.1	1,240.7	1,108.3	975.9	843.5	711.1	578.7
Sugar and candy	5,845.7	5,321.2	4,809.3	4,803.9	4,805.1	4,800.6	4,828.6	4,705.4	3,145.5	2,579.3	2,193.0	2,204.9	2,011.5	1,818.1	1,624.7	1,431.3	1,237.9	1,044.5	851.1
Cocoa and chocolate	5,633.2	5,051.8	4,977.1	4,712.6	5,012.6	5,080.9	4,859.9	4,728.6	2,659.7	2,751.2	2,485.2	2,440.2	2,287.8	2,134.4	1,980.9	1,827.5	1,674.1	1,520.7	1,367.3
Other edible products	18,833.7	16,553.8	15,266.6	14,929.3	10,871.7	10,019.5	10,237.6	10,285.5	6,212.1	5,759.4	5,006.5	3,875.2	3,592.8	3,319.4	3,046.0	2,772.6	2,499.2	2,225.8	1,952.4
Beverages 3/	18,167.3	15,669.2	15,496.4	15,059.0	14,453.7	13,558.9	12,841.7	12,013.7	9,238.2	7,911.2	7,047.5	6,621.1	6,196.6	5,771.7	5,346.8	4,921.9	4,497.0	4,072.1	3,647.2
Subtotal foods 4/	42,156.3	35,972.9	35,915.4	35,616.3	34,092.6	31,855.5	33,191.7	33,875.3	21,959.4	20,670.0	19,267.9	17,683.5	16,299.1	14,914.7	13,530.3	12,145.9	10,761.5	9,377.1	7,992.7
Animals	10,622.9	94,764.0	90,308.9	88,912.3	82,597.0	77,738.9	75,084.5	73,772.7	39,663.9	35,810.4	32,113.5	28,169.0	24,224.5	20,279.9	16,335.4	12,390.9	8,446.4	4,501.9	2,557.4
Plants	18,167.3	15,669.2	15,496.4	15,059.0	14,453.7	13,558.9	12,841.7	12,013.7	9,238.2	7,911.2	7,047.5	6,621.1	6,196.6	5,771.7	5,346.8	4,921.9	4,497.0	4,072.1	3,647.2
Beverages 3/																			

We use solver for the solving of complex problems it has several advantages such as:

- Linear Algebra: The process of solving equations with a single unknown variable is relatively straightforward.

However, when it comes to solving equations with multiple unknown variables, it can become quite complicated. Solver can quickly process scenarios involving multiple unknown variables.

- Optimization: Optimization is a key purpose of Solver as used in the corporate world. Solver uses a computer's fast processing power to subject any mathematical scenario to rapid data analysis for purposes of finding a solution to complicated formulas.

Business Intelligence and Data Analysis

In this we need to work on the data for to understand the data and for the analysis. In this when we are analyzing the data it help us to take the decision clearly. This analysis shows us how we can go further with the analysis. Here to analyze the data we need to have raw data by raw data we can't able to work on that so we need to clear the cluster according to our needs and here in Excel we need to remove the unnecessary columns or rows. After that we need to change excel into csv formate.

The screenshot displays the Weka Explorer interface. The 'Classify' tab is active, and the 'ZeroR' classifier is selected. The 'Test options' section shows 'Use training set' selected. The 'Classifier output' pane displays the following information:

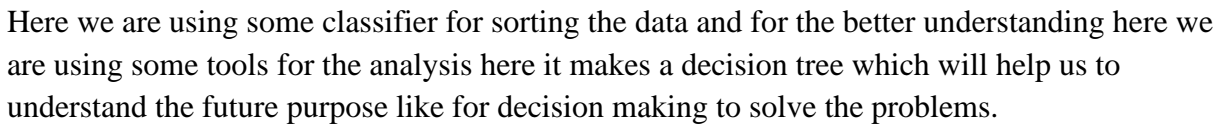
```
Test mode: evaluate on training data
=== Classifier model (full training set) ===
ZeroR predicts class value: 4,433.3
Time taken to build model: 0 seconds
=== Evaluation on training set ===
Time taken to test model on training data: 0 seconds
=== Summary ===
Correctly Classified Instances      2      11.1111 %
Incorrectly Classified Instances    16      88.8889 %
Kappa statistic                    0
Mean absolute error                 0.1106
Root mean squared error             0.235
Relative absolute error             100 %
Root relative squared error         100 %
Total Number of Instances          18

=== Detailed Accuracy By Class ===
```

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.000	0.000	?	0.000	?	?	0.500	0.056	41,400.9
0.000	0.000	?	0.000	?	?	0.500	0.056	1,190.6
0.000	0.000	?	0.000	?	?	0.500	0.056	3,260.5
0.000	0.000	?	0.000	?	?	0.500	0.056	8,859.8
0.000	0.000	?	0.000	?	?	0.500	0.056	931.4
0.000	0.000	?	0.000	?	?	0.500	0.056	3,631.8
0.000	0.000	?	0.000	?	?	0.500	0.056	4,827.4
0.000	0.000	?	0.000	?	?	0.500	0.056	833.3
0.000	0.000	?	0.000	?	?	0.500	0.056	3,605.9
0.000	0.000	?	0.000	?	?	0.500	0.056	2,755.6
0.000	0.000	?	0.000	?	?	0.500	0.056	1,518.7

The status bar at the bottom shows 'Status OK' and a 'Log' button. The Windows taskbar at the very bottom indicates the system time as 12:09 AM on 4/7/2023.

Here we are using WEKA software to analyze the data for better understanding and for the analysis which makes us make decisions. Here we are using excel for the cluster to sort the data and for better analysis.



Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose **RandomForest** -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options

☒ Use training set ☐ Supplied test set ☐ Cross-validation ☐ Percentage split

Folds: 10 %: 66

More options...

(Nom) 1999

Start Stop

Result list (right-click for options)

- 00:08:15 - rules.ZeroR
- 00:12:06 - trees.LMT
- 00:12:25 - trees.RandomForest**

Classifier output

	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	0.00000
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	931.4
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	3,631.8
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	4,827.4
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	833.3
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	3,605.9
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	2,755.6
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1,518.7
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1,661.8
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1,523.1
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	2,367.7
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	4,433.3
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	14,242.3
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	22,725.3
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

==== Confusion Matrix ====

a b c d e f g h i j k l m n o p q <-- classified as

1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | a = 41,400.9

0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | b = 1,190.6

0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | c = 3,260.5

0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 | d = 8,859.8

0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 | e = 931.4

0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 | f = 3,631.8

0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 | g = 4,827.4

0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 | h = 833.3

0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | i = 3,605.9

0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 | j = 2,755.6

0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 | k = 1,518.7

0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 | l = 1,661.8

0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 | m = 1,523.1

0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 | n = 2,367.7

0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 | o = 4,433.3

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 | p = 14,242.3

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 | q = 22,725.3

Status

OK

Log

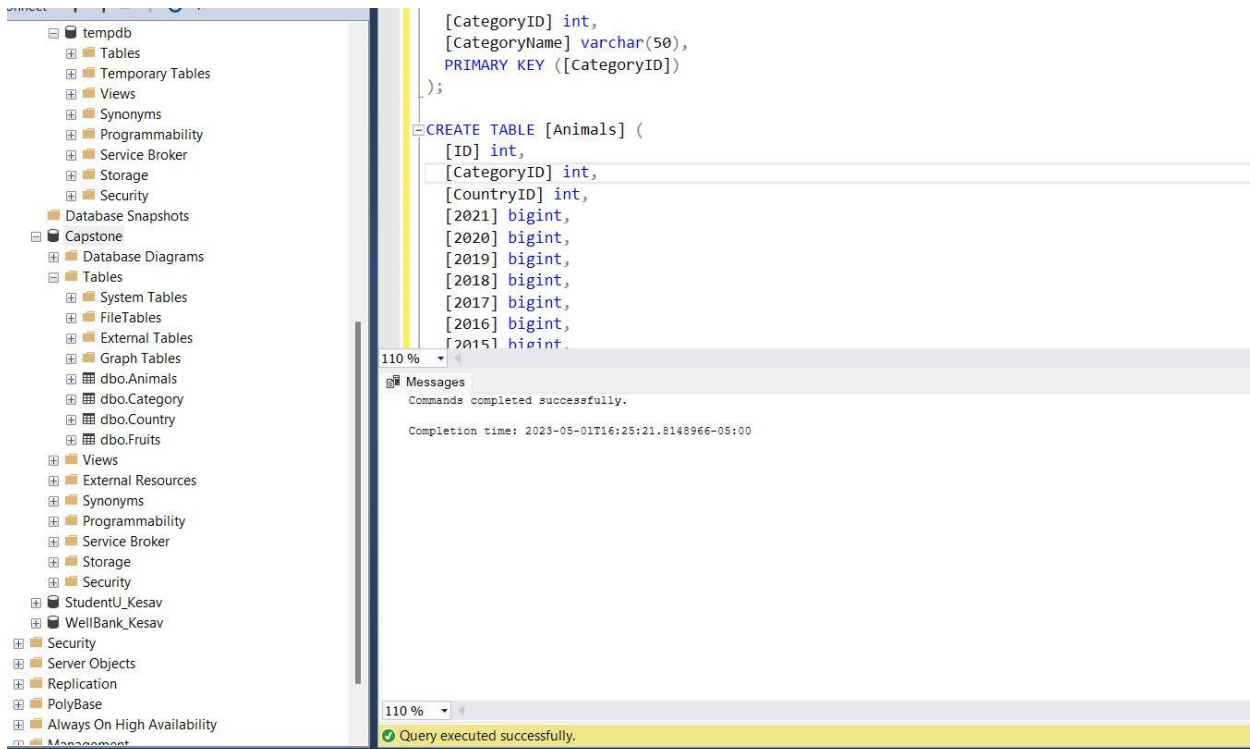
38°F Clear

Search

12:12 AM 4/7/2023

Database Management

In database management we learnt SQL ,also how it works .By executing the SQL query I transformed the excel data into table form, also below I'm attaching the process of conversion of data into tabular form.

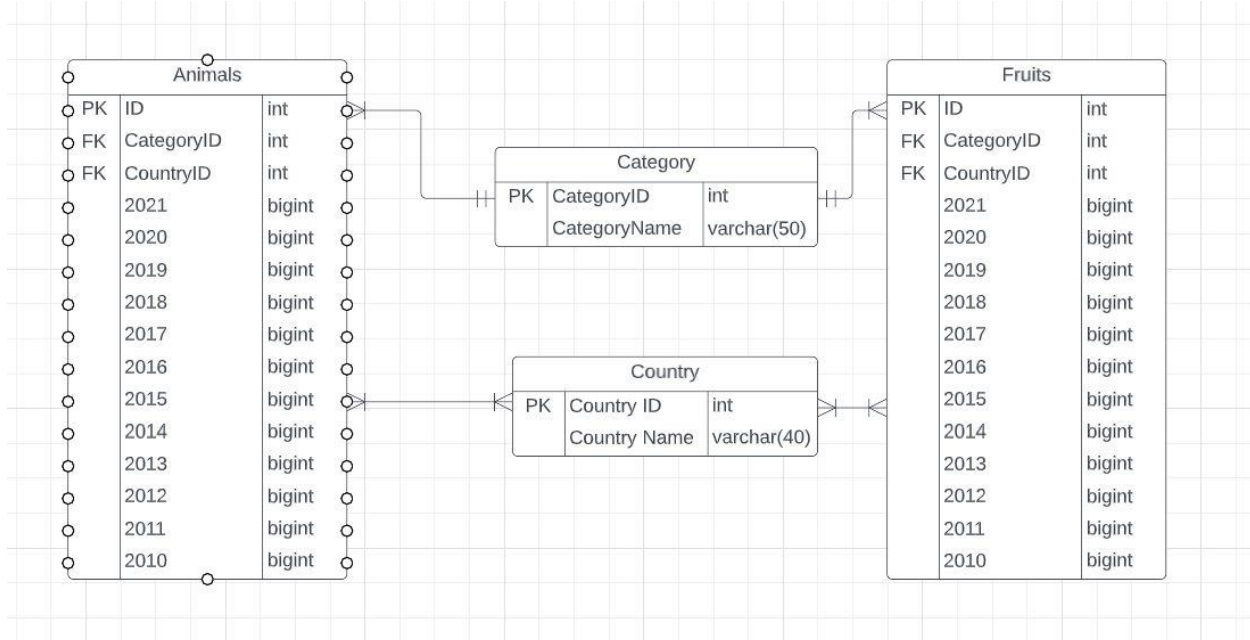


```
[CategoryID] int,
[CategoryName] varchar(50),
PRIMARY KEY ([CategoryID])
);

CREATE TABLE [Animals] (
[ID] int,
[CategoryID] int,
[CountryID] int,
[2021] bigint,
[2020] bigint,
[2019] bigint,
[2018] bigint,
[2017] bigint,
[2016] bigint,
[2015] bigint
);
```

Messages
Commands completed successfully.
Completion time: 2023-05-01T16:25:21.8148966-05:00

Query executed successfully.




```
CREATE TABLE [Category] (  
    [CategoryID] int,  
    [CategoryName] varchar(50),  
    PRIMARY KEY ([CategoryID])  
);
```

```
CREATE TABLE [Animals] (  
    [ID] int,  
    [CategoryID] int,  
    [CountryID] int,  
    [2021] bigint,  
    [2020] bigint,  
    [2019] bigint,  
    [2018] bigint,  
    [2017] bigint,  
    [2016] bigint,  
    [2015] bigint,  
    [2014] bigint,  
    [2013] bigint,  
    [2012] bigint,  
    [2011] bigint,  
    [2010] bigint,  
    PRIMARY KEY ([ID]),  
    CONSTRAINT [FK_Animals.ID]  
    FOREIGN KEY ([ID])  
    REFERENCES [Category]([CategoryID])  
);
```

```
CREATE TABLE [Fruits] (  
    [ID] int,  
    [CategoryID] int,  
    [CountryID] int,  
    [2021] bigint,  
    [2020] bigint,  
    [2019] bigint,  
    [2018] bigint,  
    [2017] bigint,  
    [2016] bigint,  
    [2015] bigint,  
    [2014] bigint,
```

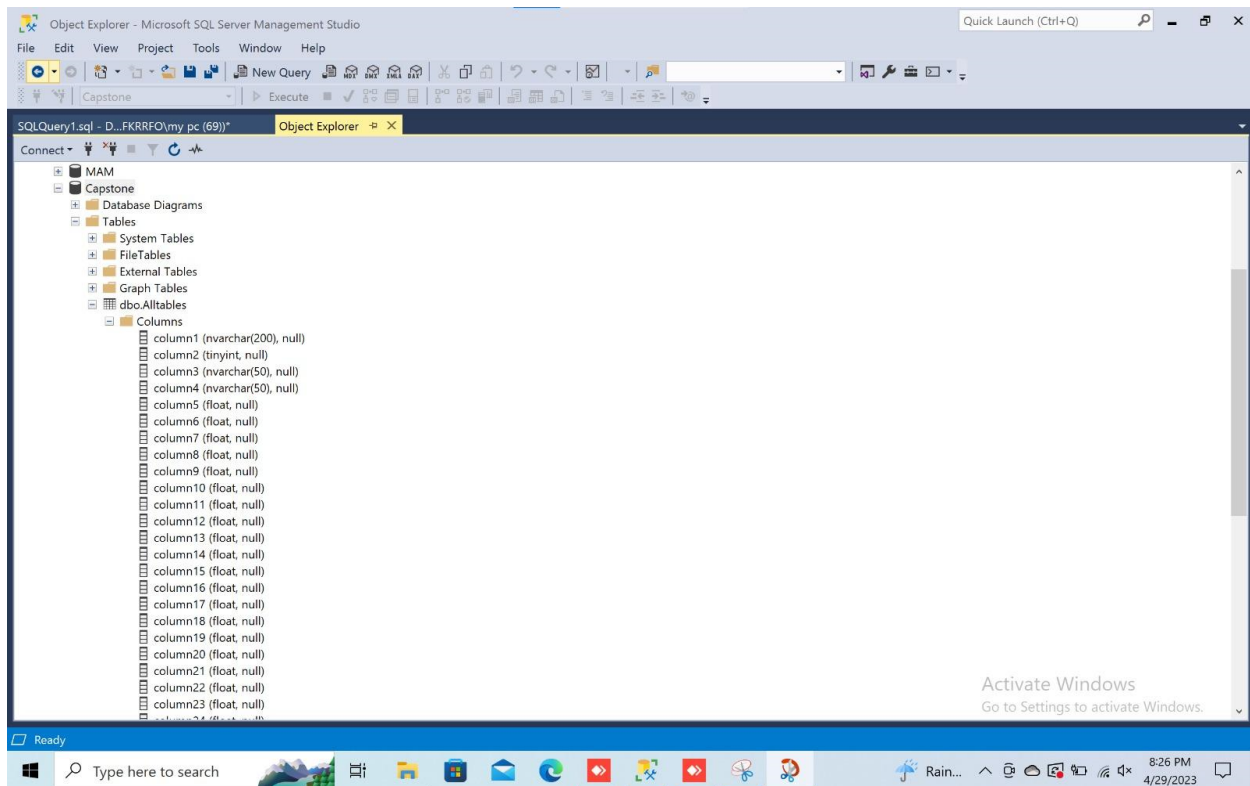
```
[2013] bigint,  
[2012] bigint,  
[2011] bigint,  
[2010] bigint,  
PRIMARY KEY ([ID]),  
CONSTRAINT [FK_Fruits.ID]  
FOREIGN KEY ([ID])  
REFERENCES [Category]([CategoryID])  
);
```

```
CREATE TABLE [Country] (  
[Country ID] int,  
[Country Name] varchar(40),  
PRIMARY KEY ([Country ID])  
);
```

Business Data Warehousing

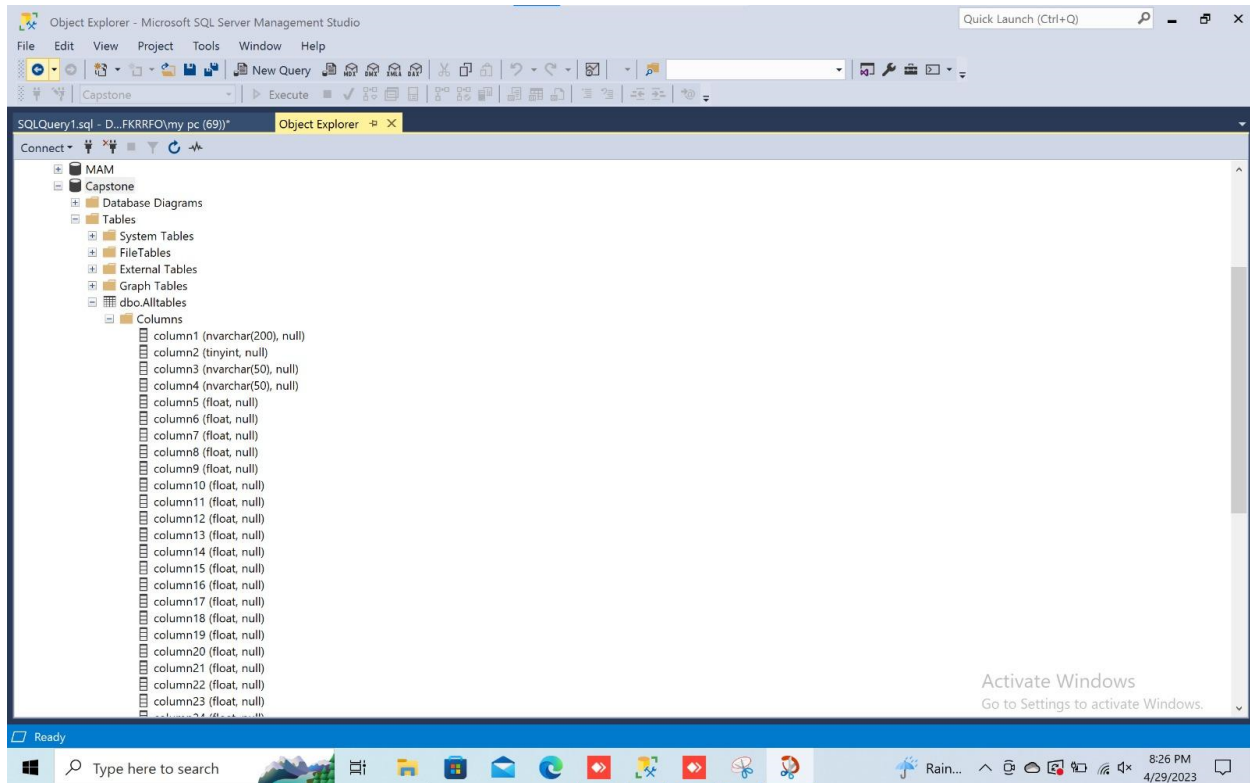
Large volumes of business data from numerous sources are collected, sorted, and then analyzed as part of the business data warehousing process, which usually involves a data warehouse. Querying, reporting, and data analysis are all business intelligence (BI) tasks that may be supported by a data warehouse, which is a database. A data warehouse's function is to offer a single, complete picture of all the data within an organization so that strategic business choices may be made.

Data extraction, transformation, and loading (ETL) from multiple sources, including transactional databases, flat files, and external systems, is one of the many phases involved in data warehousing. The data is often grouped into a dimensional model once it has been fed into the data warehouse in order to allow the generation of BI reports and dashboards.



Businesses that need to make data-driven choices should pay particular attention to business data warehousing. Businesses may enhance data quality, eliminate data silos, and acquire insights that

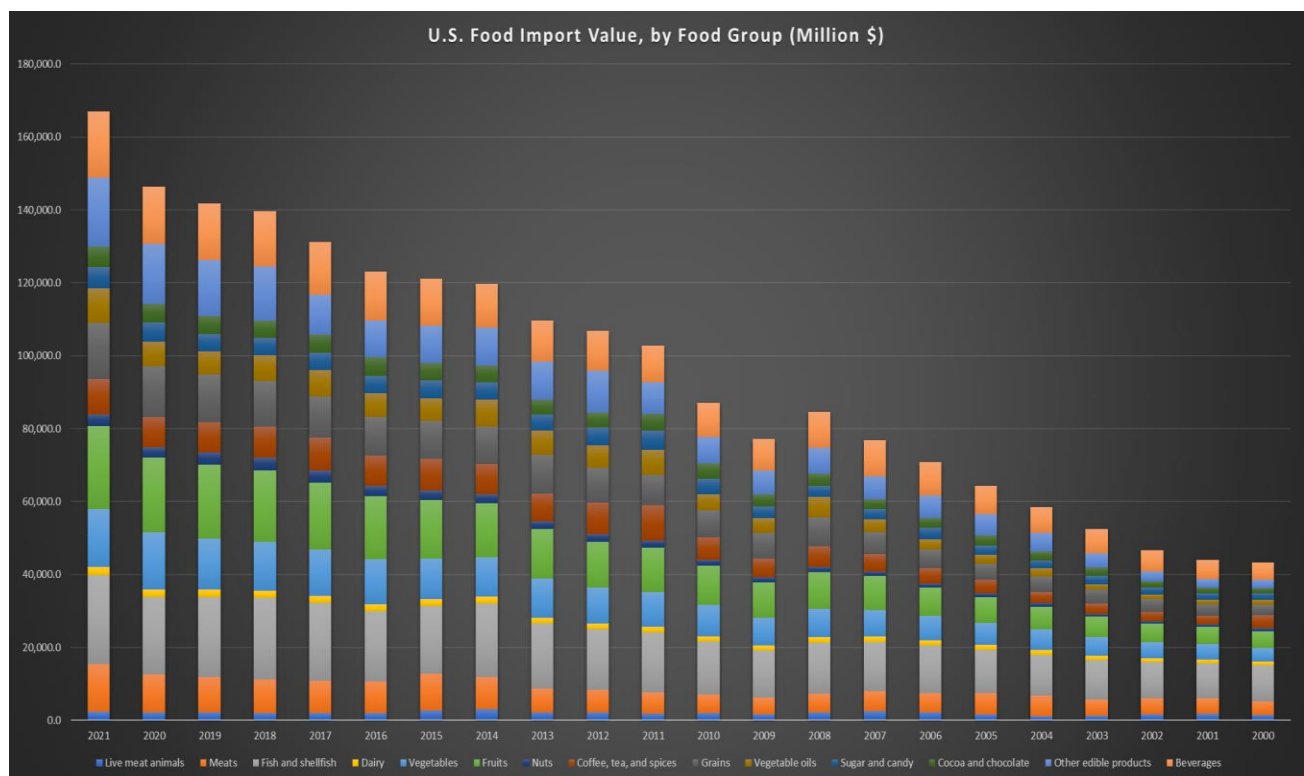
can aid them in identifying trends, opportunities, and areas for improvement by centralizing and organizing data in a data warehouse.



Business Forecasting & Visualization

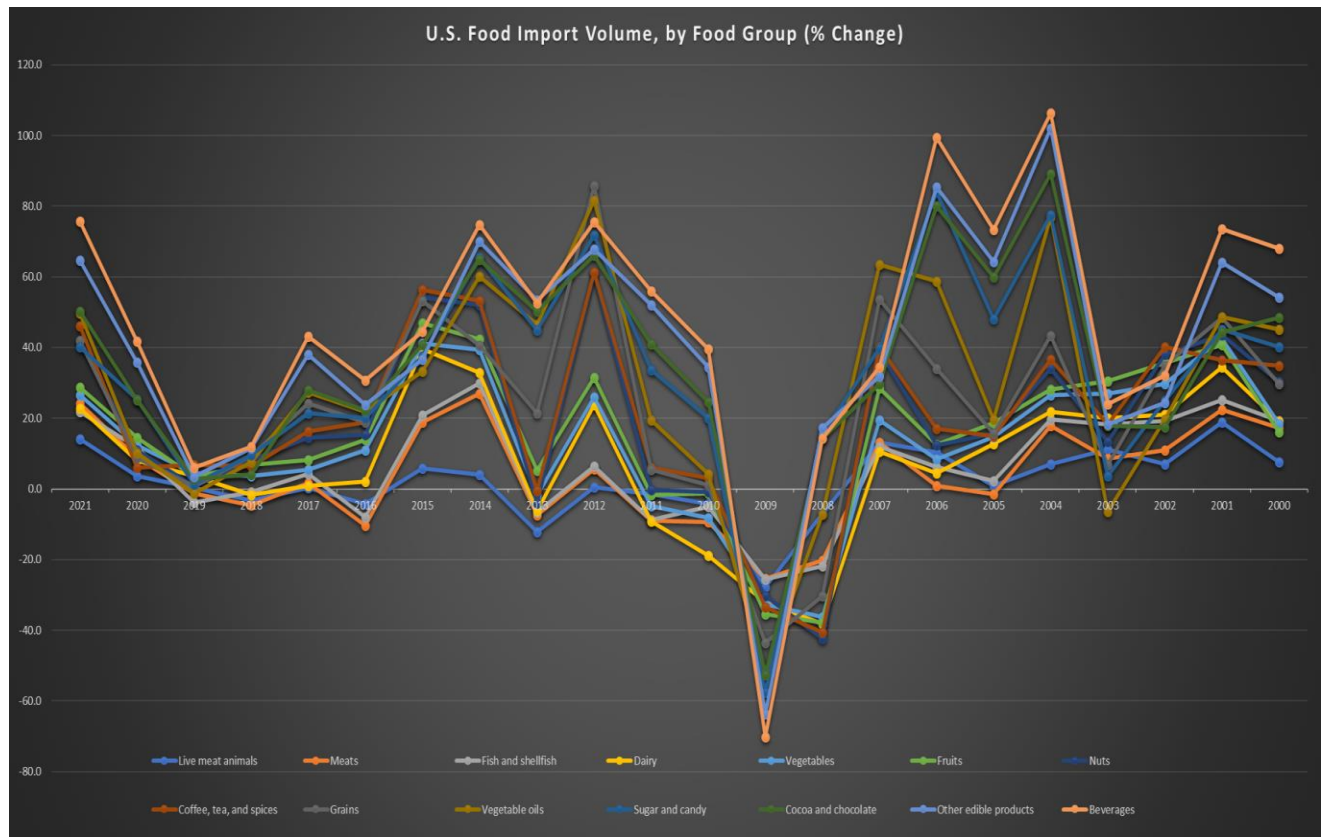
Business Forecasting & Visualization is all about having a clear understanding of the data and being able to visualize it in a way that you can help derive past trends to enable better forward thinking decisions through the use of forecasting. The method used to assemble these charts was done through Excel, using the recommended charts to allow a clear visualization of the data being analyzed.

The graph below shows the U.S. Food Import Value using a column chart, broken down by each particular food group (shown through color code on each of the columns), from the year 2000 to 2021.



As we can clearly see, the trend volume of Food Import Value increases with time, with the exception of 2009 and 2010. These are the years succeeding the 2008 global financial crisis, which look to have had an impact on the value of food the U.S. imported, due to the clear decrease in value being imported during these years.

In order to understand this impact further, it is important to analyze more data to see its true impact so we can help navigate any future consequences of a future financial crisis on U.S. Food Imports. To do so, a line chart was created to provide a different perspective, as shown below.



The line graph shows the U.S. Food Import Volume by food group, and is shown as the percentage change year on year using the same time frame as the previous graph (2000 to 2021). This graph clearly indicates that the largest percentage decrease in the volume of imported food from 2000 to 2021, was in the year 2009, which aligns with the drop in value in the previous graph. This visualization shows the dramatic percentage drop from the previous year (2008), which as mentioned previously was the year of the global financial crisis.

Having these data points allows us to look further into the knock on effect of a global financial crisis and hypothesize on the reasons why both volume and value dropped in the following years. Firstly, the crisis led to a sharp reduction in consumer spending and a general economic downturn, which in turn affected demand for food imports. As people tightened their budgets, they may have chosen to purchase less expensive, domestically-produced food items rather than imported products.

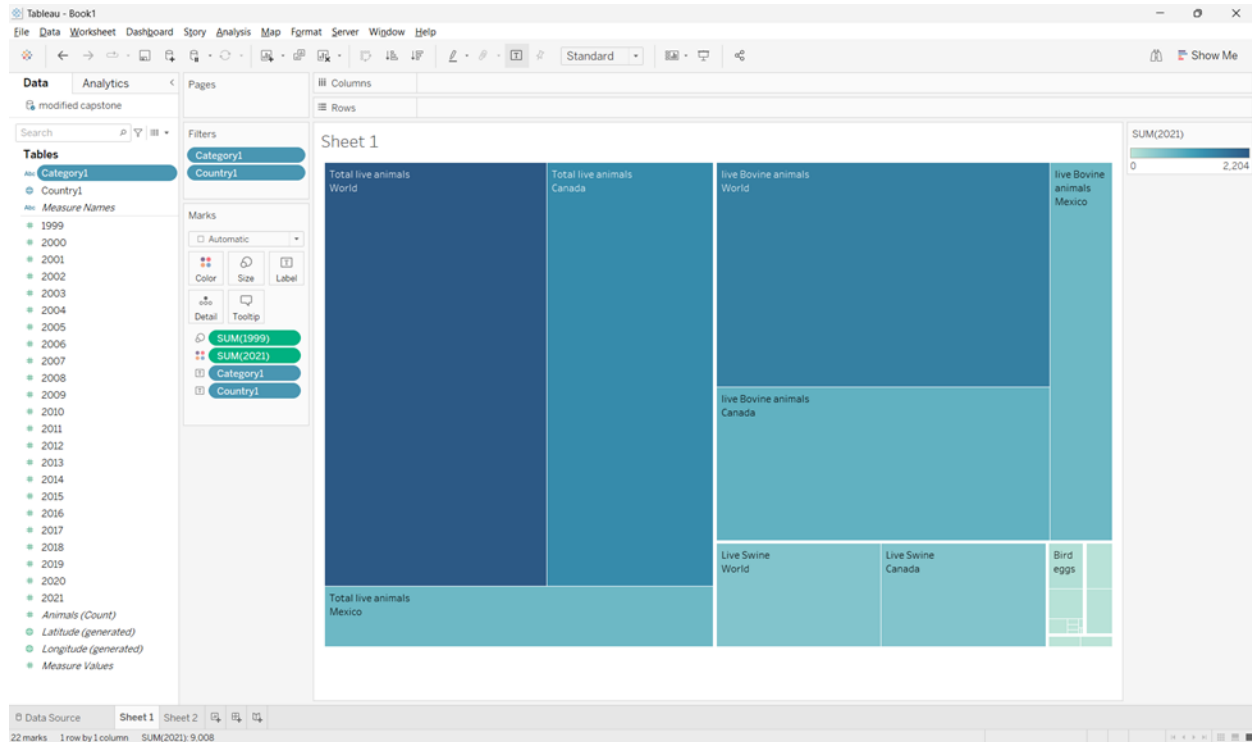
Additionally, the crisis led to a contraction of credit markets, which made it more difficult for businesses to obtain financing. This would have impacted importers, who may have had difficulty securing the necessary funds to purchase and transport goods from overseas.

The 'global' crisis also had a wider impact on the global food supply chain. It disrupted trade flows, which would have led to disruptions in the transportation of food products. It also resulted in a decrease in demand for food imports in many countries, including the US, which could have led to a decrease in production and exports from food-producing countries.

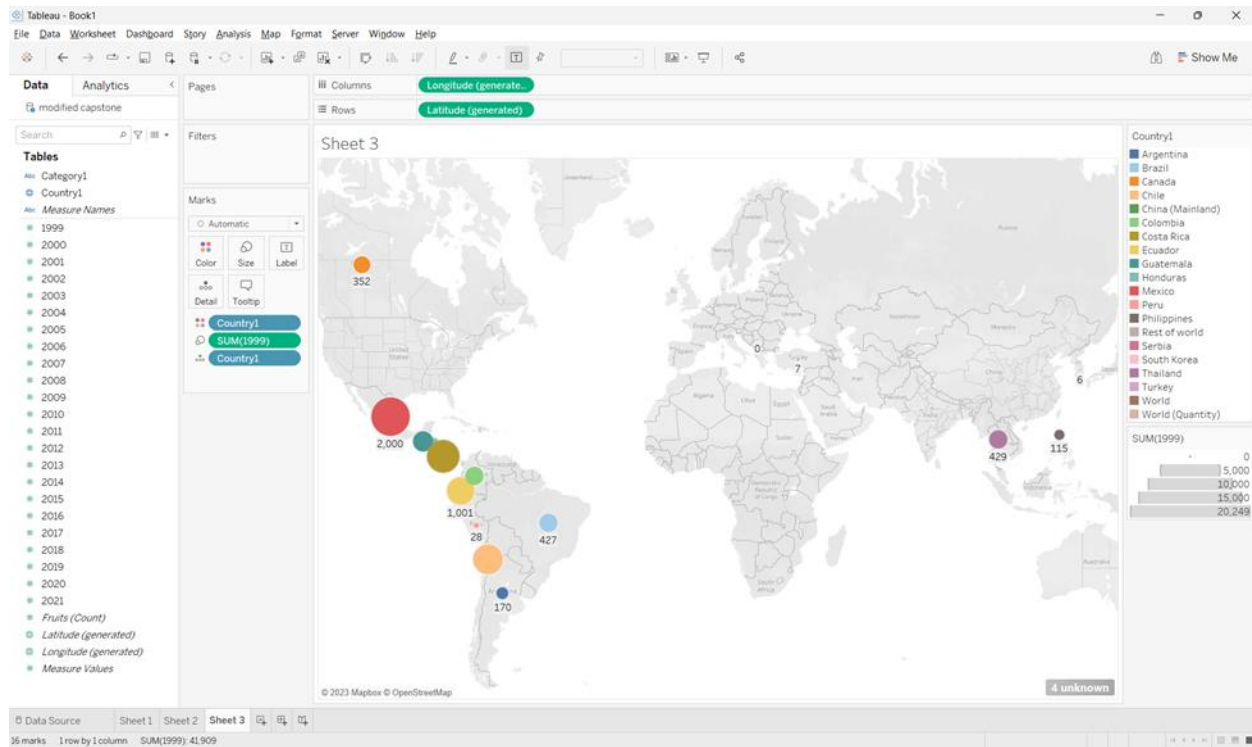
Visualizing Information

Visualizing Information is the graphical representation of information and data .We can interpret a visualization faster and accurate by using chart titles,Labeling,Annotations.We can represent time series data by using Bar graph,Line graph,Dot plot,Dot-bar graph.Here the data is extracted from

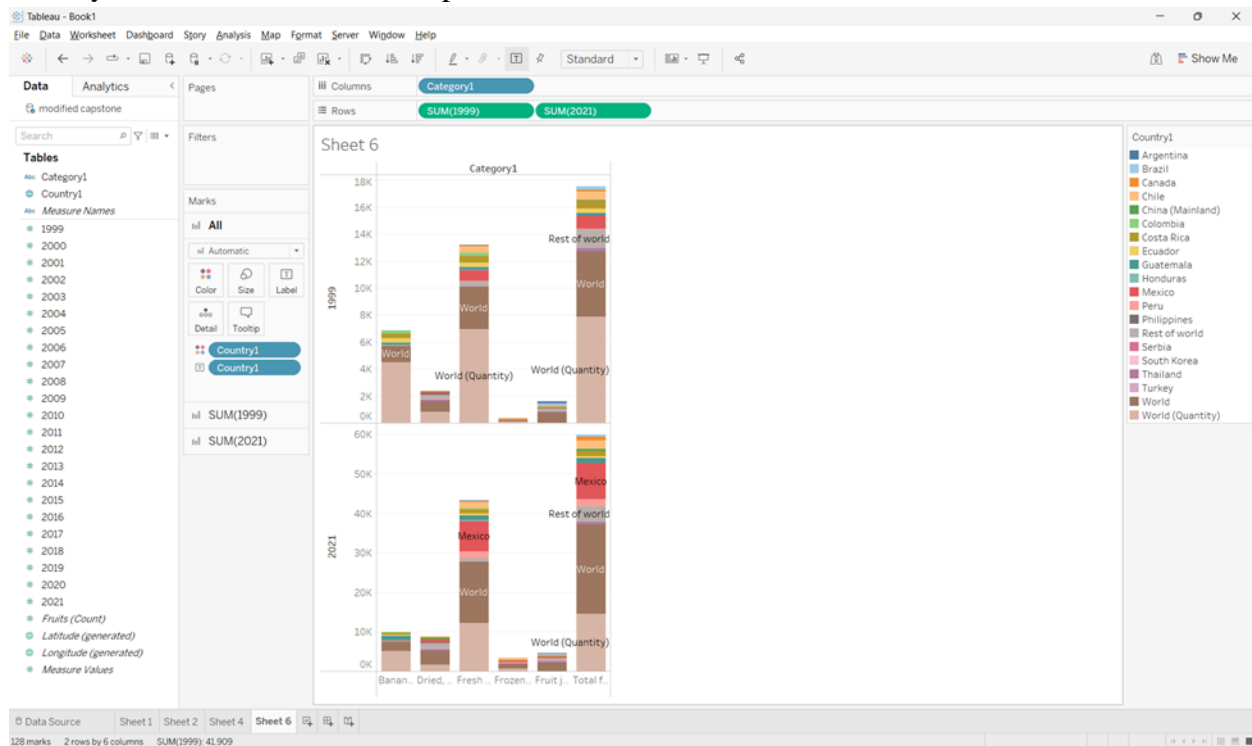
<https://www.ers.usda.gov/data-products/ag-and-food-statistics-charting-the-essentials/agricultural-trade/#:~:text=Nearly%20two%2Dthirds%20of%20U.S.,%2C%20cut%20flowers%2C%20and%20hops.>Also here we are using Tableau to represent data in a graphical format.



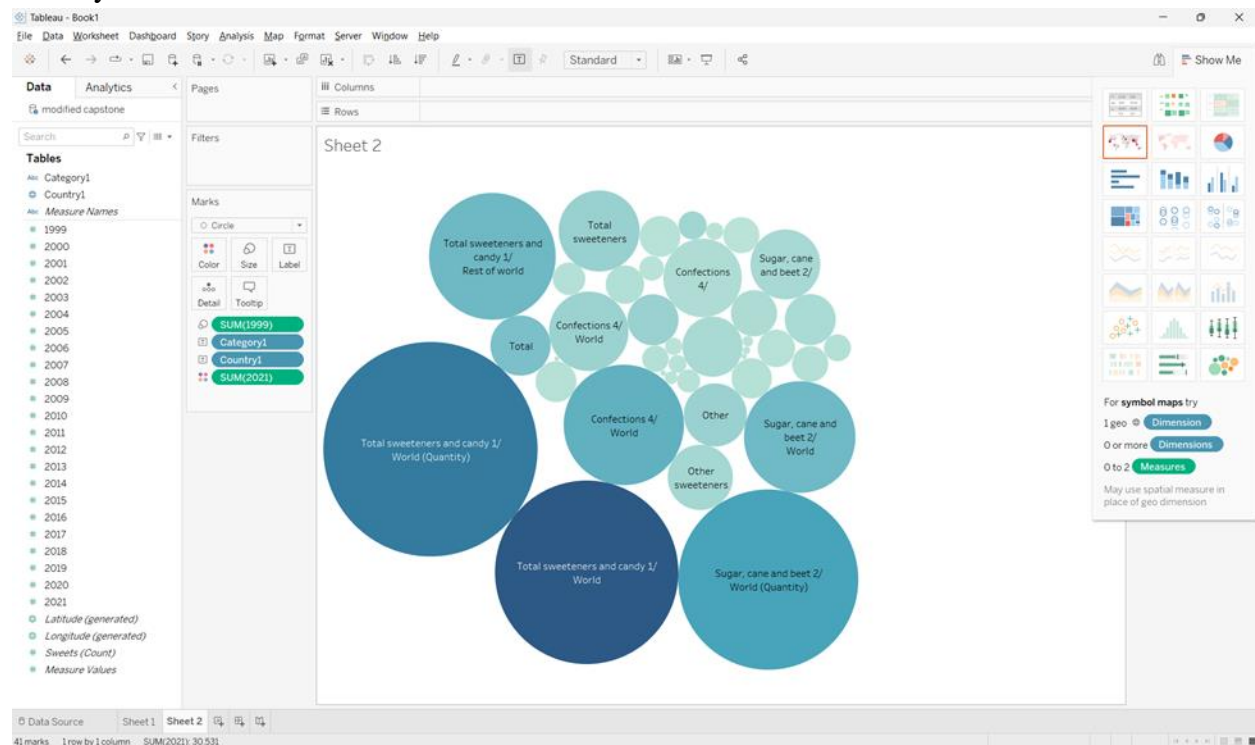
Here By using Tableau ,from the dataset I have chosen animals from the category and applied filters to country.By using Treemap here we can notice that the count of number of animals imported from the world is high compared to the remaining .By using marks enabled show markup labels so that we can able to see the count if we tap on it.Also applied color filters to year 2021 .Here in the years 2021 and 1999 Total live animals imported from the world is high.



Here from the dataset ,by using Tableau I can able to plot the points in the maps, in the year 1999 US imported the greater quantity of fruits from Mexico .By using tableau in the columns i have chosen longitude and in the rows I have chosen latitude ,by using marks enabled colors to country,also increased size of the plot .



Here from the dataset I have drawn a bar chart by using Tableau .Here I'm comparing the quantity of fruits imported from the different countries by the US in the year 1999 and year 2021.On the columns I have chosen category and on the rows I have chosen years sum 1999,sum2021.Example here from the graph frozen fruits importing from outside US is increased little bit from the year 1999 to 2021.Aslo we can able to notice lot of changes in the graph comparing to two years.Also enabled show markup labels to country and enabled colors to country.

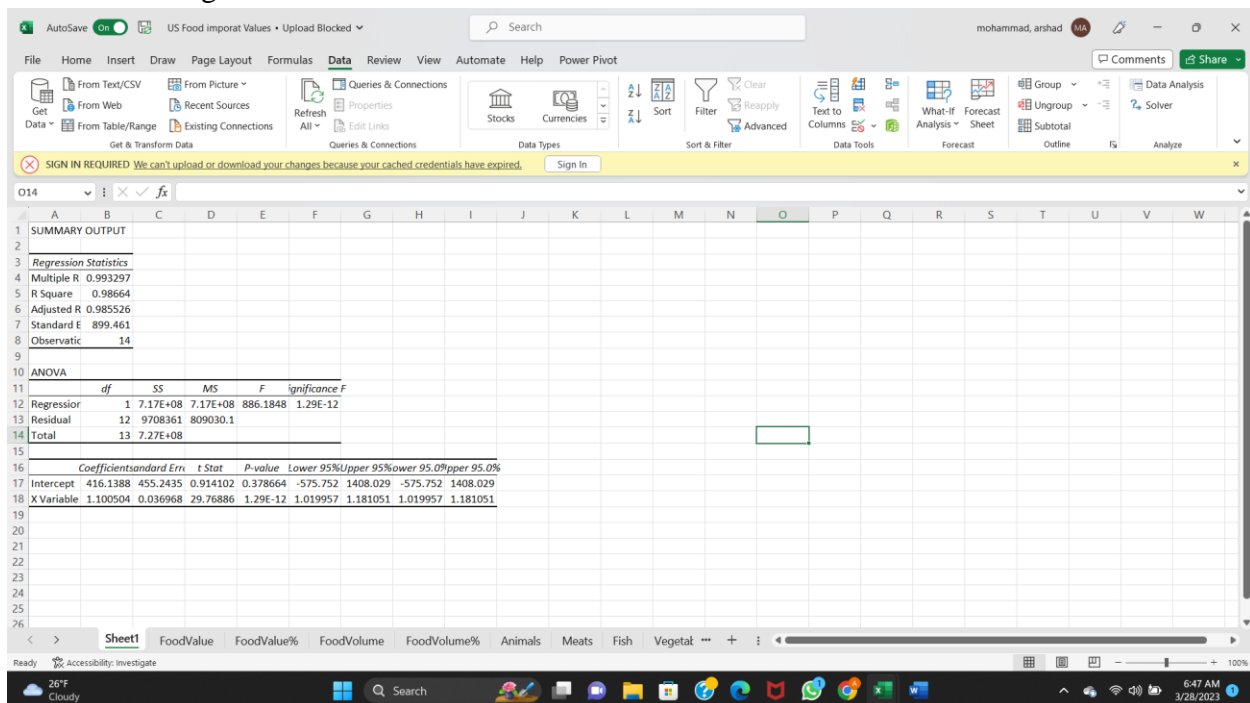


Here from the dataset I have drawn a Bubble plot by using Tableau.Here im comparing the number of sweets imported from the different countries by the US in the year 1999 and 2021.Also by using marks enabled show markup labels to the country and category.Also enabled color filters for identification.

Data Mining for Business Decision

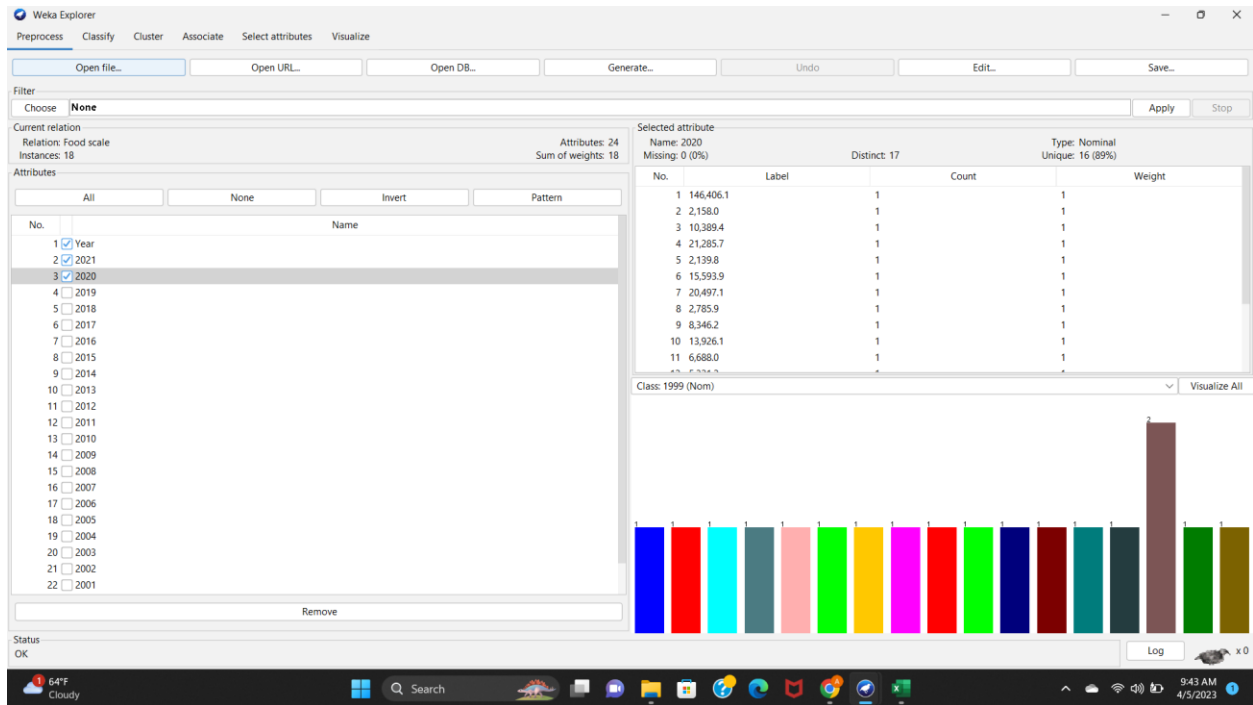
What is a Business Decision? If a company gets a problem we need to have a solution to clear the problem. For that we need exact data to solve the problem. Once we analyze the data it will help the organization to solve the problem and by this they take a business decision. In this we need to concentrate on Cost, Revenue, Break even here Break even means there is no profit and no loss which means what we are invested is that coming back to us without any loss and profit. Every organization wants to know that when they are going to achieve their Break even because after touching Break even then they are going to earn the profits.

In this below Image we are using regression analysis in Excel for better understanding and for decision making.

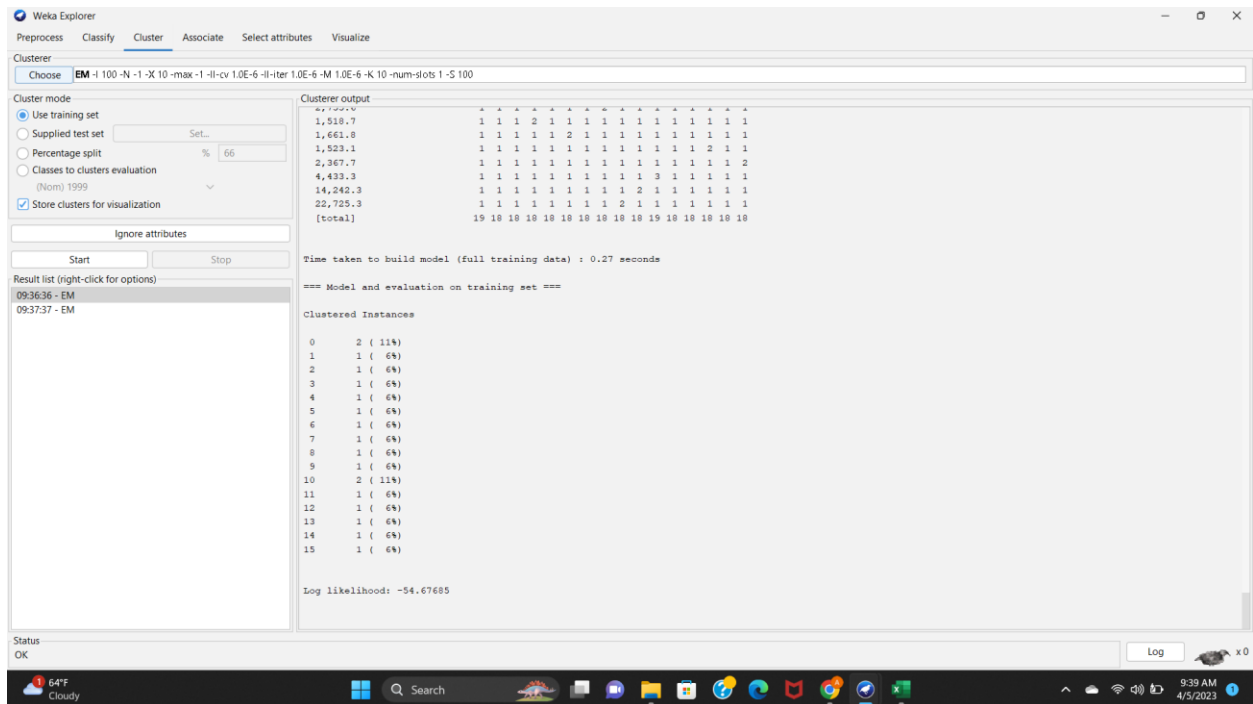


Here R square represents the coefficient of determination which means it indicates the percentage of variation in the dependent variable that gives the result to the independent variable. Here we got R square value 0.98664 which shows a perfect relationship between dependent and independent variables.

By data mining it will make the decision making easy. Here for better understanding we are using WEKA software for better understanding and to make easy business decisions. We are using Cluster, J48, Randomtree. First we need to add the data in weka after that we need to cluster the data if we positive result then we need to check in J48 and randomtree.



Here the cluster is used for similar instances in whole data by this we can find the common data and it help us to remove the data.



Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds: **10**
- ☐ Percentage split %: **66**

More options...

(Nom) 1999

Start Stop

Result list (right-click for options)

- 09:39:16 - trees.RandomTree
- 09:40:59 - trees.J48**

Classifier output

```

J48
2000
1999
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===

J48 pruned tree
-----
: 4,433.3 (18.0/16.0)

Number of Leaves : 1
Size of the tree : 1

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      0      0  %
Incorrectly Classified Instances    18     100 %
Kappa statistic                    -0.1096
Mean absolute error                 0.1168
Root mean squared error             0.249
Relative absolute error            102.7551 %
Root relative squared error        102.5896 %
Total Number of Instances          18

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
0.000  0.235  0.000  0.000  0.000  -0.130  0.029  0.056  41,400.9
0.000  0.000  ?  0.000  ?  ?  0.000  0.056  1,190.6
0.000  0.000  ?  0.000  ?  ?  0.029  0.056  3,260.5
0.000  0.000  ?  0.000  ?  ?  0.029  0.056  8,859.8
0.000  0.000  ?  0.000  ?  ?  0.029  0.056  931.4
  
```

Status: OK

Log

64°F Cloudy Search 9:41 AM 4/5/2023

Here we used Randomtree which will make a supervised classifier. This tool helps to create a decision tree. With the help of this decision tree it makes things easy to make decisions. And it gives a vision on which way we are going.

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options:
☐ Use training set
☐ Supplied test set **Set...**
☒ Cross-validation Folds: **10**
☐ Percentage split %: **66**
More options...

(Nom) 1999
Start **Stop**

Result list (right-click for options):
09:39:16 - trees.RandomTree
09:40:59 - trees.J48

Classifier output:

```

2021 = 22,695.9 : 4,827.4 (1/0)
2021 = 3,156.5 : 833.3 (1/0)
2021 = 9,825.9 : 3,605.9 (1/0)
2021 = 15,250.2 : 2,755.6 (1/0)
2021 = 9,573.6 : 1,519.7 (1/0)
2021 = 5,845.7 : 1,661.8 (1/0)
2021 = 5,633.2 : 1,523.1 (1/0)
2021 = 18,833.7 : 2,367.7 (1/0)
2021 = 10,167.3 : 4,433.3 (2/0)
2021 = 42,156.3 : 14,242.3 (1/0)
2021 = 106,622.9 : 22,725.3 (1/0)

```

Size of the tree : 18
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	2	11.1111 %
Incorrectly Classified Instances	16	88.8889 %
Kappa statistic	0.0069	
Mean absolute error	0.1046	
Root mean squared error	0.2364	
Relative absolute error	91.9766 %	
Root relative squared error	97.7689 %	
Total Number of Instances	18	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.000	0.118	0.000	0.000	0.000	0.000	-0.086	0.088	0.056	41,400.9
0.000	0.000	?	0.000	?	?	?	0.059	0.056	1,190.6
0.000	0.000	?	0.000	?	?	?	0.088	0.056	3,260.5
0.000	0.000	?	0.000	?	?	?	0.088	0.056	8,859.8
0.000	0.000	?	0.000	?	?	?	0.088	0.056	931.4
0.000	0.000	?	0.000	?	?	?	0.088	0.056	3,631.8
0.000	0.000	?	0.000	?	?	?	0.088	0.056	4,827.4

Status: OK **Log** x0

64°F Cloudy Search 9:42 AM 4/5/2023

Challenges:

Here we have raw data. Weka is not going to accept the raw data we need to clear the data, remove the blank cells and the excel file should have only one sheet. It does not accept multiple sheets.

Managerial Finance

Managerial Finance is mostly concentrated on figures, data, and outcomes. In this managerial finance will help us to reduce the losses and mostly it gives a good understanding for the decision maker. Managerial Finance is a combination of accounting and corporate finance. In this course we learned on Corporate governance, Dividends, stock repurchase, capital structure, working capital structure & multinational finance, hybrid financing & mergers and acquisitions, and financial derivatives. This course by above topic we are able to understand and take a right decision about bonds which one we are going to buy or to invest in it. This will help to see that the correct calculation and the value of the company of the mergers and acquisition.

In our data set we have US food imports throughout the world we can see the below images for better understanding.

The screenshot displays an Excel spreadsheet titled "US Food Import Values". The data is organized into columns for years from 2000 to 2021, with an additional column for the "average" of the last 10 years. The rows are categorized by food groups: "Total foods 1/", "Foods", "Subtotal foods 4/", and "Plants". The "Foods" category includes "Live meat animals", "Meats", "Fish and shellfish 2/", "Dairy", "Vegetables", "Fruits", "Nuts", "Coffee, tea, and spices", "Grains", "Vegetable oils", "Sugar and candy", "Cocoa and chocolate", "Other edible products", and "Beverages 3/". The "Plants" category includes "Animals", "Beverages 3/", and "Plants". The data shows a general upward trend in import values over the period, with a significant increase in the "Foods" category.

			average	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000
Total foods 1/	U.S. imports	Percent	5.0	14.0	3.3	1.5	6.4	6.5	1.7	1.2	9.1	2.7	3.8	18.2	12.6	-8.6	10.0	8.4	10.0	10.2	11.3	12.4	6.3	1.4	4.7
Foods	Live meat animals	Percent	3.1	6.6	-4.2	11.0	0.6	-4.1	-24.2	-7.8	37.3	-0.2	16.0	-6.0	21.2	-27.0	-12.3	19.4	29.9	46.9	-10.8	-25.9	-2.6	24.8	19.3
	Meats	Percent	9.4	27.0	7.5	4.5	4.2	3.4	-14.1	11.8	36.9	4.6	8.5	13.1	10.3	-8.8	-5.7	2.4	-8.8	0.6	29.2	3.3	0.6	11.2	17.4
	Fish and shellfish 2/	Percent	4.2	13.7	-2.3	-2.1	4.4	10.7	4.0	-7.7	12.8	8.0	0.0	13.4	12.2	-7.0	3.6	2.5	10.7	6.6	2.3	9.0	3.1	-2.2	11.5
	Dairy	Percent	5.0	15.1	-2.5	6.4	9.9	-1.6	-0.2	2.1	9.3	4.6	7.4	11.8	-0.2	-15.8	5.3	7.8	1.8	7.7	16.5	10.3	1.3	8.1	-0.6
	Vegetables	Percent	5.1	1.4	11.6	4.1	5.9	2.1	10.5	3.2	1.8	8.0	2.6	11.0	15.7	-3.7	7.5	9.6	9.5	5.4	12.6	15.9	5.7	10.3	3.9
	Fruits	Percent	6.5	10.7	0.9	4.0	6.3	7.2	7.5	7.7	8.9	8.5	3.1	12.4	10.3	-2.6	7.3	19.8	11.8	15.8	7.4	9.6	8.4	1.1	-2.8
	Nuts	Percent	5.8	13.3	-12.7	-8.9	5.7	15.1	3.7	17.1	17.4	-2.8	10.4	28.1	14.3	-5.4	14.1	8.2	-1.9	3.9	38.2	5.6	3.2	-15.6	1.9
	Coffee, tea, and spices	Percent	0.5	17.7	-1.0	-0.3	-6.7	9.1	-2.6	2.9	8.7	-12.5	-10.2	55.1	21.4	-7.5	16.5	14.1	11.2	19.9	9.5	17.0	2.3	-30.2	-4.5
	Grains	Percent	6.4	9.5	6.8	3.8	11.6	6.6	1.9	0.0	-2.1	12.7	13.1	13.0	4.4	-11.0	29.4	20.3	15.1	5.7	11.1	8.1	12.1	9.4	2.8
	Vegetable oils	Percent	4.5	43.1	6.1	-9.8	-3.1	10.6	3.5	-14.1	10.1	6.9	-8.2	50.7	11.6	-29.1	61.9	24.8	19.3	5.4	48.8	15.7	-0.6	-12.7	-1.2
	Sugar and candy	Percent	1.2	9.9	10.6	0.1	0.0	0.1	-0.6	2.6	5.4	-9.5	-7.0	26.8	30.5	3.1	14.8	-13.9	22.0	17.6	-0.5	14.4	14.4	2.6	-1.2
	Cocoa and chocolate	Percent	2.1	11.5	1.5	5.6	-6.0	-1.3	4.5	2.8	13.7	1.5	-12.5	9.0	23.6	5.4	24.0	0.1	-3.3	10.7	1.8	38.5	14.6	9.3	-7.7
	Other edible products	Percent	9.1	13.8	8.4	2.3	37.3	8.5	-2.1	-0.5	-1.0	-9.5	34.0	20.1	11.0	-8.3	10.3	2.3	7.9	15.0	29.2	44.9	8.7	6.6	-2.6
	Beverages 3/	Percent	6.0	15.9	1.1	2.9	4.2	6.6	5.6	6.9	6.4	3.6	7.1	9.5	6.2	-11.3	-0.8	7.6	16.8	12.3	6.4	13.8	13.6	5.9	9.1
Subtotal foods 4/	Animals	Percent	5.4	17.2	0.2	0.8	4.5	7.0	-4.0	-2.0	20.0	6.3	3.6	11.5	11.7	-10.0	-0.3	4.5	6.2	7.3	9.0	4.1	1.8	4.0	12.7
	Beverages 3/	Percent	6.0	15.9	1.1	2.9	4.2	6.6	5.6	6.9	6.4	3.6	7.1	9.5	6.2	-11.3	-0.8	7.6	16.8	12.3	6.4	13.8	13.6	5.9	9.1
	Plants	Percent	4.8	12.5	4.9	1.6	7.6	6.2	3.5	1.8	5.1	1.2	3.5	22.6	14.1	-7.5	17.8	10.8	10.8	11.5	14.0	18.0	8.0	-1.5	-1.2

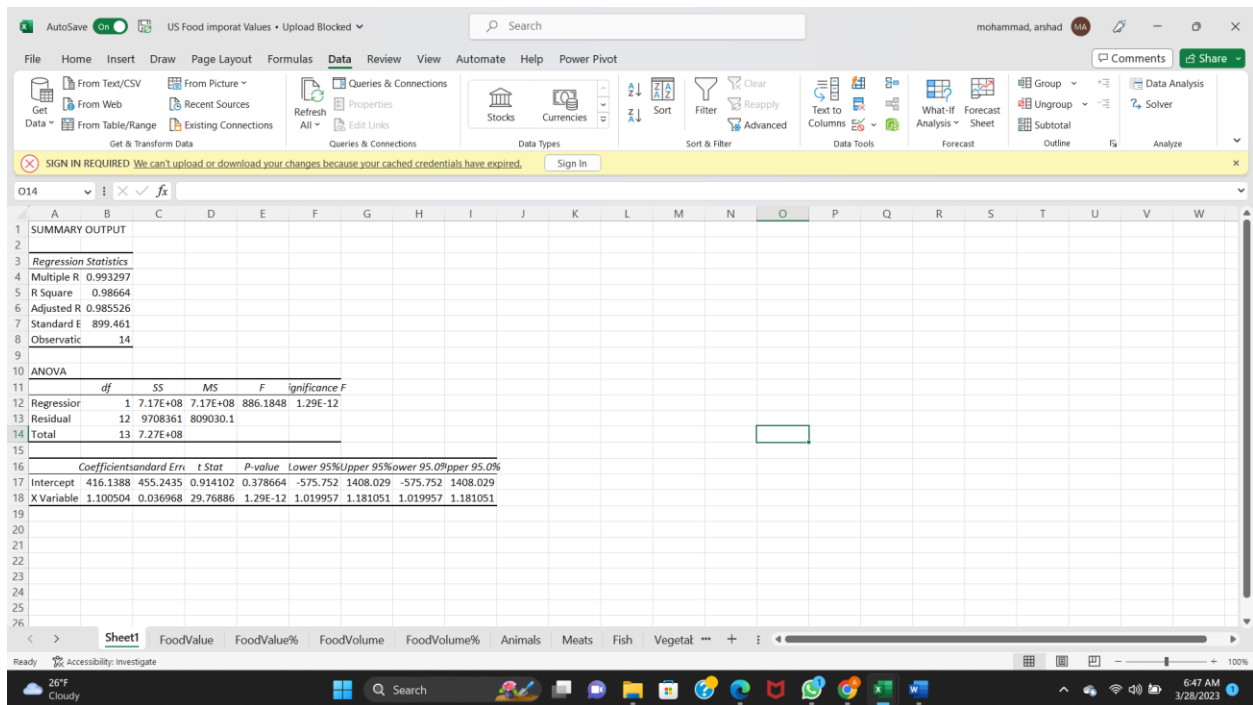
In the above image we can see that there is a change in percentage every year. Live meat animals, meats, fish. Dairy, vegetables, fruits, nuts, grains, oils, sugar and some other food. All of this is an increase in the percentage. As we know, the US is wealthiest country and by this the demand for different types of food also increasing rapidly so the import size also increasing every year.

</

We know that every year inflation will be there. When we are buying X product today for \$10 maybe next year that same X product have to buy at \$12. It happens because of inflation. When we are buying anything, we need to understand inflation also. In the above image it shows the changes in Inflation for better understanding. With this data we keep in mind how much the price is going to go up or how much the prices are going to reduce.

For better Understanding this procedure we need regression analysis this makes better understanding the procedure. And for decision making. In below are the results for regression analysis.

For this data, coming to our subject, we felt that regression between all types of imports, whether they are perfectly correlated in last two years, so we had taken two years of data for all types of imports and coming to summary of regression, regression is shows us relationship between dependent and independent variables, so summary of our analysis shows:



For regression, if the R-square is greater than 0.5, it tells us the perfect fit line between variables, for our data while we compare the last 2 years it has 0.98, it tells us 98% of correlation between variables and perfectly correlated and fit lines.

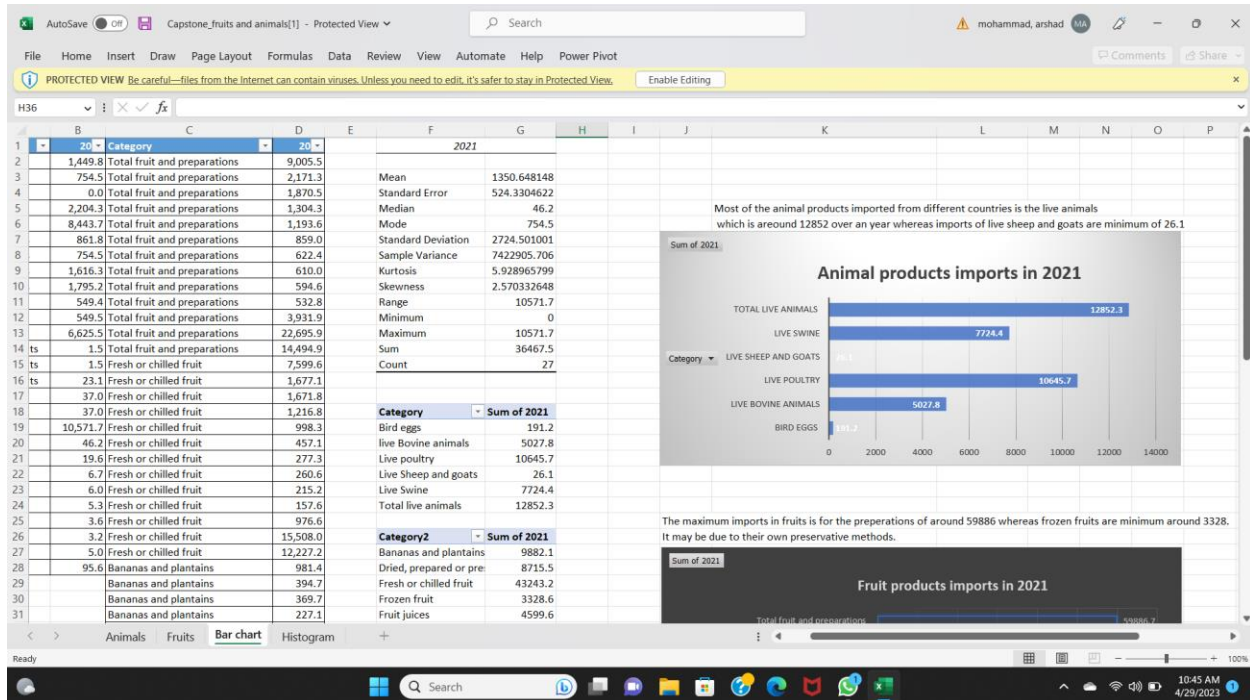
Quantitative technique for financial Market

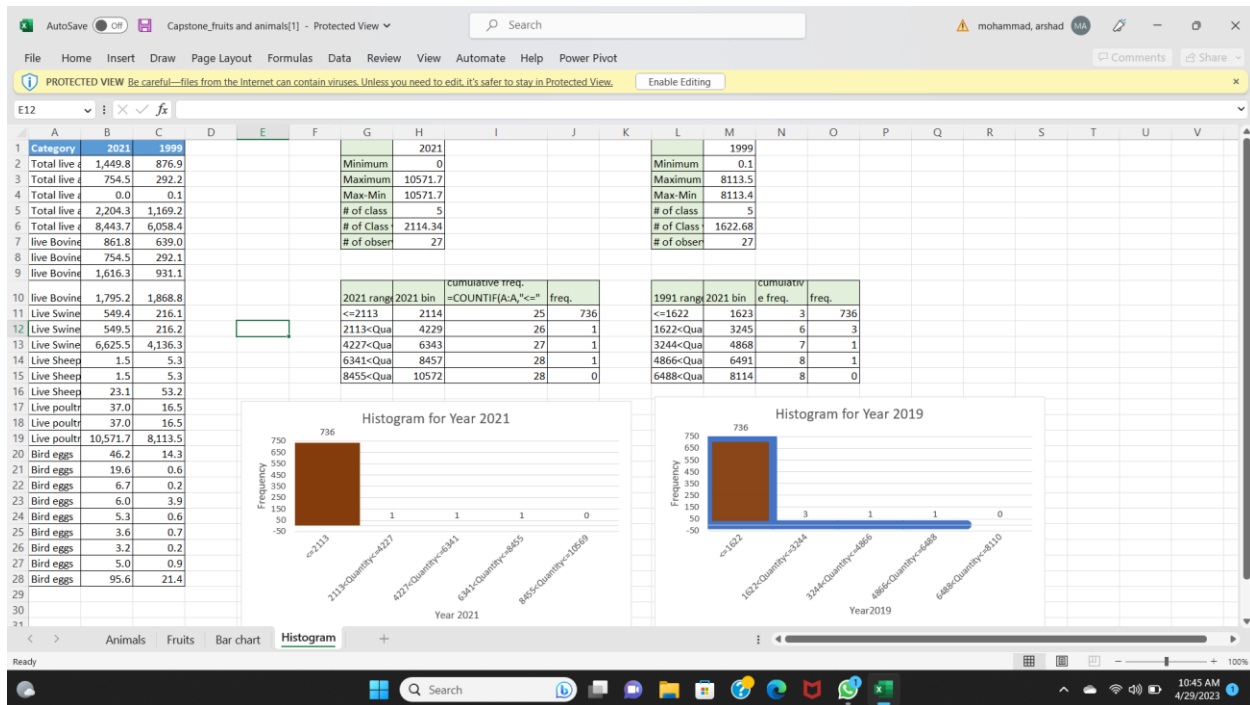
Data Analysis

Data analysis is a process of finding insights from information. The process of analyzing raw data to get insights and find patterns about the information by using various techniques and processes where the raw data can be transformed into a format which could be further visualized and be made human readable.

Excel provides the information about imports of animals and fruits for different categories of different parts of the world. It provides the figures from the period to 1999 to 2021.

I have taken the data for the year 2021 because as the developments increase in this era, we can identify which is having more imports from that category.





Investment Analysis

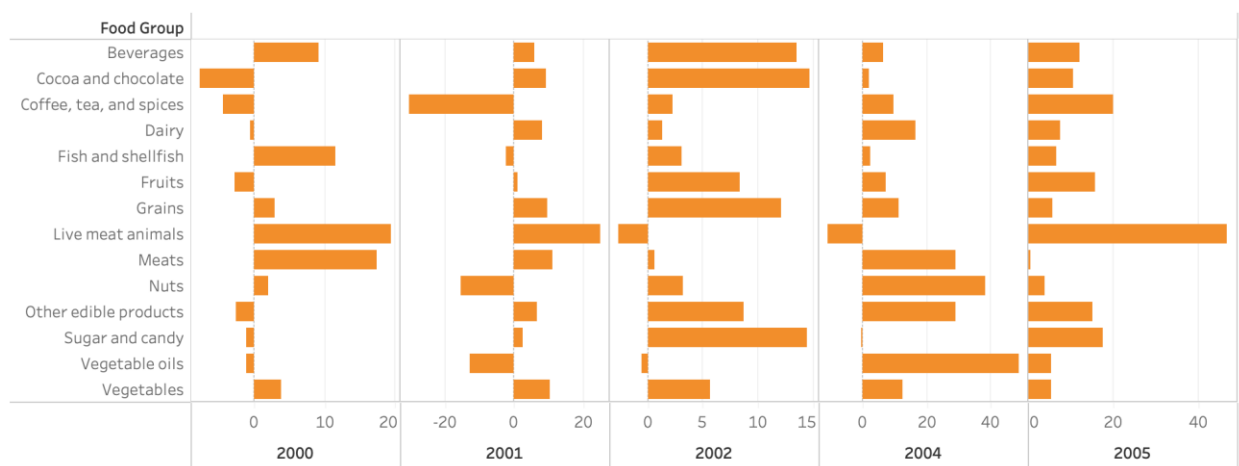
Investment analysis is used to help us better understand the risks and returns of potential investment opportunities and to help us make more informed decisions about which investments to make based on numerous factors. Investment analysis involves examining the financial, economic, and other relevant data associated with an investment, such as market trends, industry performance, company financial statements, and projected earnings.

By conducting investment analysis, investors can determine the likelihood of an investment generating a profit or a loss, as well as identify the optimal time frame for holding the investment. This information is crucially important for investors looking to make more level headed decisions about how to allocate their funds and manage their investment portfolios.

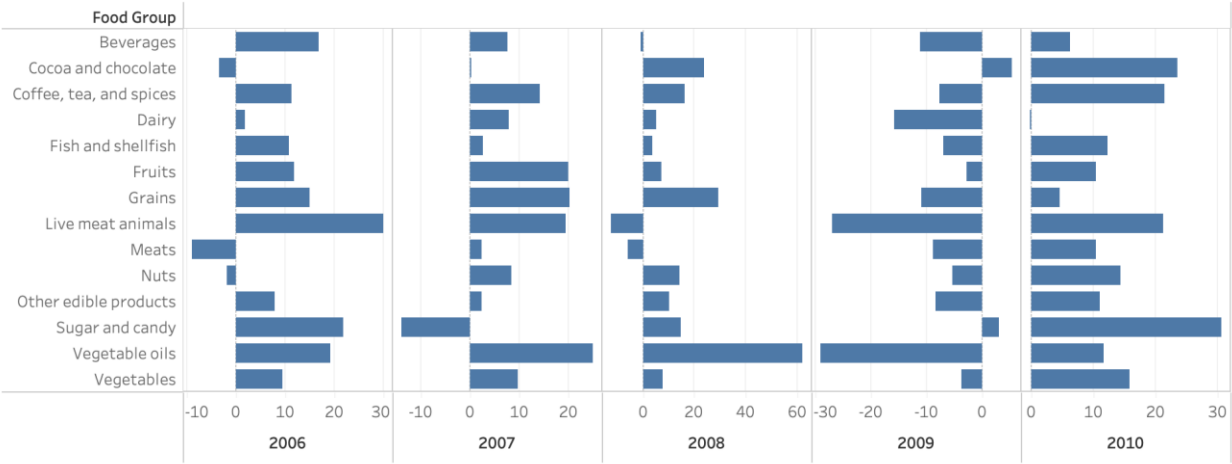
Investment analysis also helps investors understand the factors that can affect an investment's value, such as changes in interest rates, volatility in the stock market, and changes in consumer behavior/interests. This in-depth knowledge can allow investors to develop strategies for managing risk and maximizing returns, as well as identify potential niches for future investments.

With our data set, Tableau has been used to create a breakdown of the food groups by time frame. As seen below, the data ranges have been split into four different horizontal bar charts with different time frames (2000 - 2005), (2006 - 2010), (2011 - 2015) and (2016 - 2021).

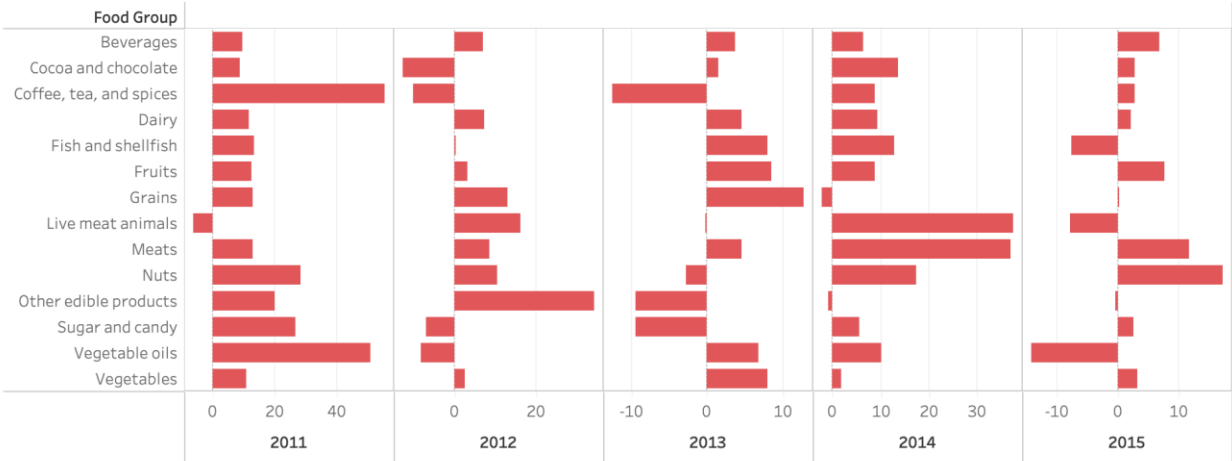
2000 - 2005



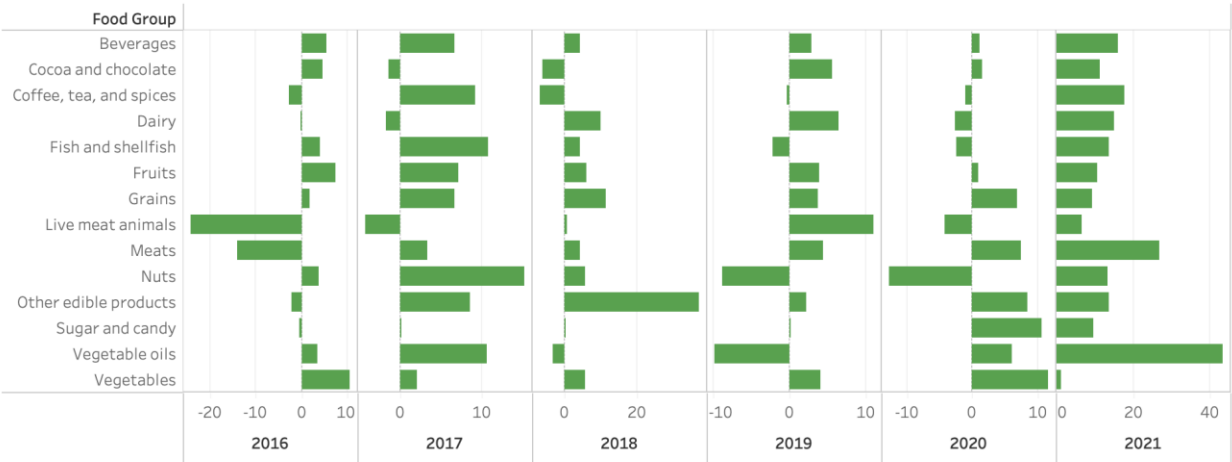
2006 - 2010



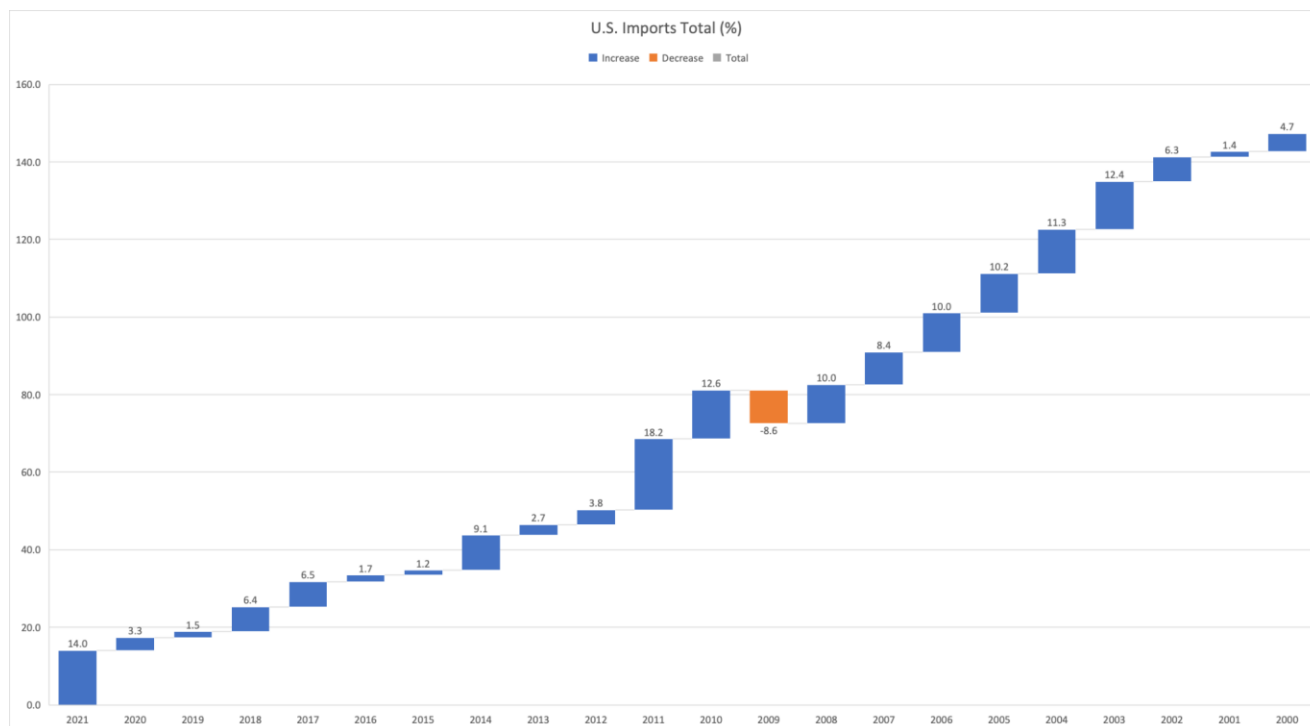
2011 - 2015



2016 - 2021



As we can see, this data is very detailed as it shows the percentage change for each individual food group, however it is fairly hard to read due to the amount of information given. For this reason, Excel has been used to create a waterfall graph and show the U.S. Total food imports as a whole. This allows a more indexed look into the market to help decide if the U.S. Import market is a viable investment opportunity.



This graph shows us that the U.S. Import Market provides fairly consistent returns year after year, with the exception of 2009 due to the reasons previously discussed. There is not a specific index fund or ETF that invests solely in U.S. food imports, as food imports can be a relatively small segment of the overall economy and are not typically tracked as a separate category in the financial markets.

However, there are some funds that invest in companies involved in the food industry, including those that may import food products. For example, the iShares U.S. Consumer Goods ETF invests in a range of consumer goods companies, including those involved in the food and beverage industry. Another example is the Invesco Dynamic Food & Beverage ETF, which invests in companies involved in the production, distribution, and retailing of food and beverage products.

Conclusion and Recommendations:

After working on the data set by using different analysis and tools we get an idea how the US is working on imports in the form of money, how much they are investing or from each country how much they are importing. With this data we get to know that the US is importing products from different countries all over the world. Here in each country who has more products, for example, India has more spices, Indonesia produces rice. By these countries the US is importing to match with the demand in the US. By importing from these countries by less value and selling in the US it is beneficial to the US. It is generating revenue and we can see that there is cash flow and by this the US government is also getting more tax on the sales and it is creating employment.

Recommendation:

Even it has advantages by importing products from other countries for lower prices. We have disadvantages by importing more products from different countries at larger values . The US will totally depend on those countries which is not good for the future. I recommend that they concentrate on self agriculture. In some places there is climate change. But we have different technology that helps us to grow plants in any climate. By doing all these it may not change in the form of imports. But at least we can see some changes for the future.

