

Country Cluster Visualization Based on Agricultural Imports: Unsupervised Learning Approach

R.S. Kamath^{1*}, P.G. Naik², S.S. Jamsandekar³

Abstract

The clustering algorithm used in this analysis makes it easier for policymakers to understand performance metrics. K-means clustering has been demonstrated to be a tool for analyzing countries' agricultural imports and a visualization tool for interpreting the results in this study. Authors have used agricultural imports from 190 nations from Knoema, a web-based open data platform. Cereals, meat, and coffee imports from 190 nations in 2016 are included in the dataset. Based on their imports of cereals, meat, and coffee, the nations were divided into three clusters of sizes 24, 13, and 159. According to the findings of this study, 159 nations fall into the cluster category, with average imports of 6.64 million tons of cereals, 0.72 million tons of meat, and 1.1 million tons of coffee, respectively. This shows the agrarian prosperity of these nations since the imports are less contrasted with the other two bunches. According to the findings, K-means has the potential to become the most widely used tool for country cluster analysis.

Keywords: K-means, cluster visualization, unsupervised learning, artificial intelligence, cluster designs

INTRODUCTION

Artificial Intelligence (AI) tools are speeding up data analysis processes in all sectors and making them even more effective. AI and machine learning often leverage historical data to forecast and anticipate future outcomes [1]. As the import of agricultural products depends on several factors such as international and domestic demand & supply situation, prices, quality concerns, and food security concerns no targets for imports are fixed. A broad study on machine learning techniques has convinced the digital transformation in agriculture [2]. The K-means Clustering Method is used to discuss the application of data mining to rice imports by primary country of origin in this study [3]. The research

findings can help ascertain the quantity of rice imported by the primary country of origin. Application of data mining methods illustrated with clustering customers of agriculture products [4]. The findings demonstrated that clustering with three groups can lengthen a customer's lifespan and increase their purchase value.

In this context, this paper visualizes 190 countries based on their imports of cereals, meat, and coffee by applying the K-means clustering approach. A comprehensive analysis has been offered by employing unsupervised learning to design prominent clusters among these countries. In this division, agricultural imports of 190 countries for the year 2016 are considered [5]. The authors aimed to extract latent insights from agricultural import datasets and generate clusters of countries grounded on their values.

*Author for Correspondence

R.S. Kamath
E-mail: rskamath@siberindia.edu.in

¹Associate Professor, Department of Computer Studies, Chhatrapati Shahu Institute of Business Education and Research (CSIBER), Kolhapur, Maharashtra, India

²Professor, Department of Computer Studies, Chhatrapati Shahu Institute of Business Education and Research (CSIBER), Kolhapur, Maharashtra, India

³Assistant Professor, Department of Computer Studies, Chhatrapati Shahu Institute of Business Education and Research (CSIBER), Kolhapur, Maharashtra, India

Received Date: November 17, 2023

Accepted Date: December 02, 2023

Published Date: January 17, 2024

Citation: R.S. Kamath, P.G. Naik, S.S. Jamsandekar. Country Cluster Visualization Based on Agricultural Imports: Unsupervised Learning Approach. Journal of Artificial Intelligence Research & Advances. 2024; 11(1): 1–7p.

AGRICULTURAL IMPORTS DATA

In this comprehensive study, we used agricultural imports from 190 countries from Knoema, a free-to-use web-based open data platform. The dataset includes cereals, meat, and coffee imports from 190 countries for the year 2016 (Knoema). These imports are by quantity in terms of tones [6].

Figure 1 provides an exploratory data analysis of the dataset designed for this study. Japan, China, and The United States of America are the top countries by cereals, meat, and coffee import quantity in the world respectively. As reported in Knoema, the world's total cereals, meat, and coffee import quantity was estimated at 435 million tons, 44.7 million tons, and 30.2 million tons in 2016 respectively.

K-means Clustering for Agricultural Imports

We have carried out unsupervised classification for the visualization of countries based on their agricultural imports. The K-means procedure determines an assortment of k groups utilizing a heuristic search beginning with a determination of k randomly picked groups every one of which before all else addresses a cluster mean [7]. Further, the cluster analysis depends on estimating the likeness between data points by finding the distance between each pair. The likeness is estimated concerning the mean worth of the information things in a group.

The present experiment is simulated in R, an open-source data mining platform. 190 countries selected for the study are clustered based on their quantity of cereals, meat, and coffee imports. Each record N_i is addressed as a multidimensional vector and is characterized as:

$$N_i = [\text{Cereal}_i, \text{Meat}_i, \text{Coffee}_i]$$

Where, $i = 1$ to 190

The K-means technique requires the researcher to determine the number of groups to be generated. R code for this process is displayed in Figure 2. We plotted the "quantity of groups against" the "inside clusters sum of squares", which is the parameter that should be limited during unsupervised learning. Figure 3 illustrates "within clusters sum of squares" for various quantities of groups. The plot reveals that the value of the within-cluster sum of squares is less for 3 clusters as compared to 4.

The K-means clustering technique divides the data items into k partitions in such a way that the resulting within-cluster similarity is more but the inter-cluster similarity is less. The R function *kmeans()* is applied to get clusters of countries [8]. The study is adjusted by altering the number of clusters while maintaining a constant number of iterations at 15. The results of this experiment are briefed in Table 1. Performance is assessed based on the "within-cluster sum of square errors" and "BetweenSS/TotalSS". Preferably, a clustering method that exhibits both inside cohesion and outside separation would have a BSS/TSS ratio approaching 1 [9].

Cereals		Coffee		Meat	
Min. :	26	Min. :	1	Min. :	115
1st Qu. :	136871	1st Qu. :	475	1st Qu. :	7979
Median :	591236	Median :	4596	Median :	41944
Mean :	2299421	Mean :	161756	Mean :	237693
3rd Qu. :	2102560	3rd Qu. :	81498	3rd Qu. :	160609
Max. :	23400454	Max. :	5746364	Max. :	3271819

Figure 1. Basic statistical summary of the dataset.

```
wss <- (nrow(import)-1)*sum(apply(import,2,var))
for (i in 2:10)
  wss[i] <- sum(kmeans(import, centers=i)$withinss)
plot(1:10, wss, type="b", xlab="Number of Clusters",
      ylab="Within groups sum of squares")
```

Figure 2. R code for determining the number of clusters.

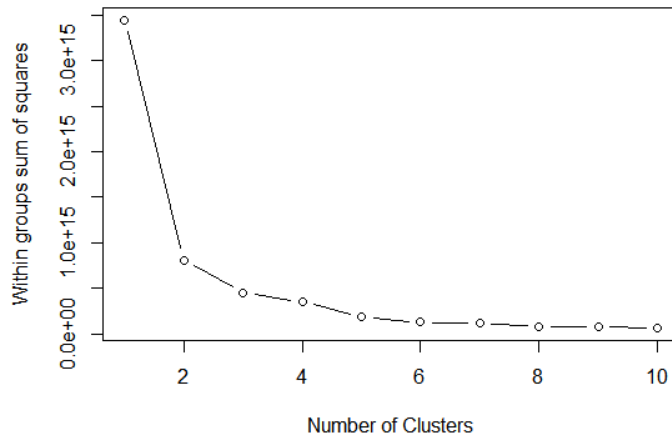


Figure 3. Number of clusters against within clusters sum of squares.

K-means clustering with 3 clusters of sizes 24, 13, 153

Cluster means:

	Cereals	Coffee	Meat
1	5666743.5	530449.79	494052.2
2	15541001.4	537434.62	1263993.7
3	646111.6	72001.05	110277.2

Clustering vector:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
3	3	3	3	2	1	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3
49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
1	3	3	3	3	2	3	3	3	1	3	1	3	3	3	3	3	3	3	3	3	2	3	2
73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
3	1	2	2	2	3	1	3	3	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144
3	3	3	3	3	3	3	3	3	1	3	3	2	1	3	3	3	3	3	3	3	3	3	3
145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168
3	3	3	3	3	1	1	1	1	3	2	3	3	3	3	3	2	3	2	3	2	3	1	3
169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190		
3	3	3	3	3	3	3	1	3	1	3	3	3	1	1	1	1	3	1	3	3	2		

within cluster sum of squares by cluster:

[1] 153764252636126 195977230412473 96704266960079
 (between_ss / total_ss = 87.0 %)

Available components:

[1] "cluster" "centers" "totss" "withinss" "tot.withinss" "betweenss"
 [7] "size" "iter" "ifault"

Figure 4. Cluster means and cluster vector for countries clusters.

RESULTS AND DISCUSSIONS

Considering cluster performance explained in Table 1 and Figure 3, it has been decided to design three clusters of 190 countries. The Table shows that for 3 clusters BSS/TSS ratio is 87% indicating a good fit [10]. However, the cluster size is also taken into consideration. The K-means algorithm is applied with three clusters.

The K-means clustering yielded three clusters with sizes 24, 13, and 153. Figure 4 displays the cluster means and clustering vectors obtained in the R environment. Within-cluster variation measures how much each event within a cluster differs from the others in the same cluster. [10]. Figures 5 and 6 show cluster plots for 1st and 2nd principal components and discriminant functions respectively. The plot reveals that cluster 3 is dense and some of the data points of cluster 1 are closer to cluster 3 [11, 12]. The matrix plot of countries' clusters based on agricultural imports is shown in Figure 7. The clusters and participating countries are tabulated in Table 2.

Table 1. Performance evaluation for accuracy of cluster designs.

No. of clusters	Cluster size	Between_SS/Total_SS
2	17, 173	76.4%
3	24, 13, 153	87%
4	12, 28, 139, 11	89.8%
5	3, 13, 34, 18, 122	94.6 %

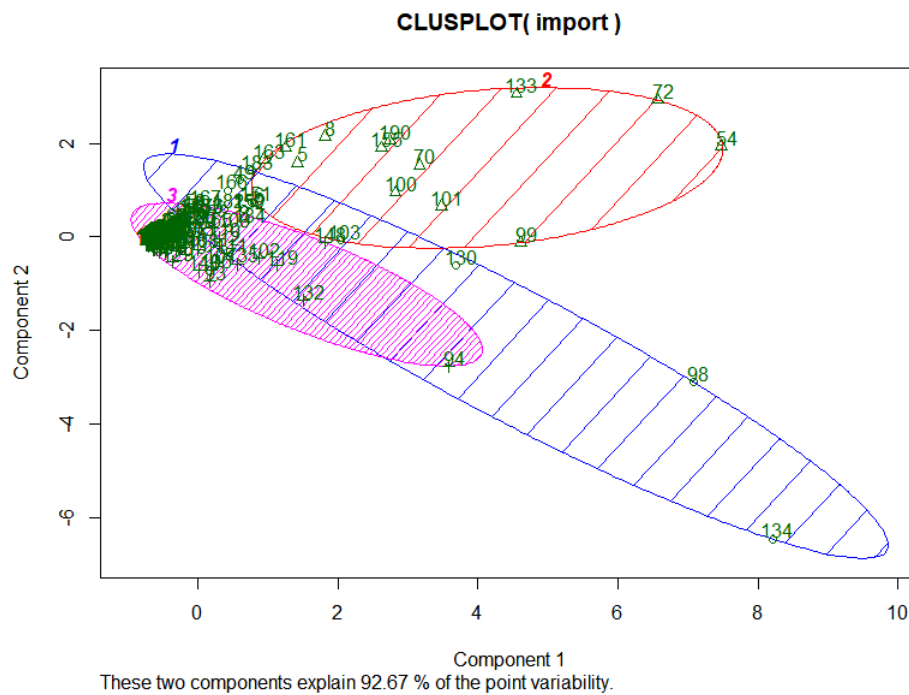


Figure 5. Cluster plot for 1st and 2nd principal components.

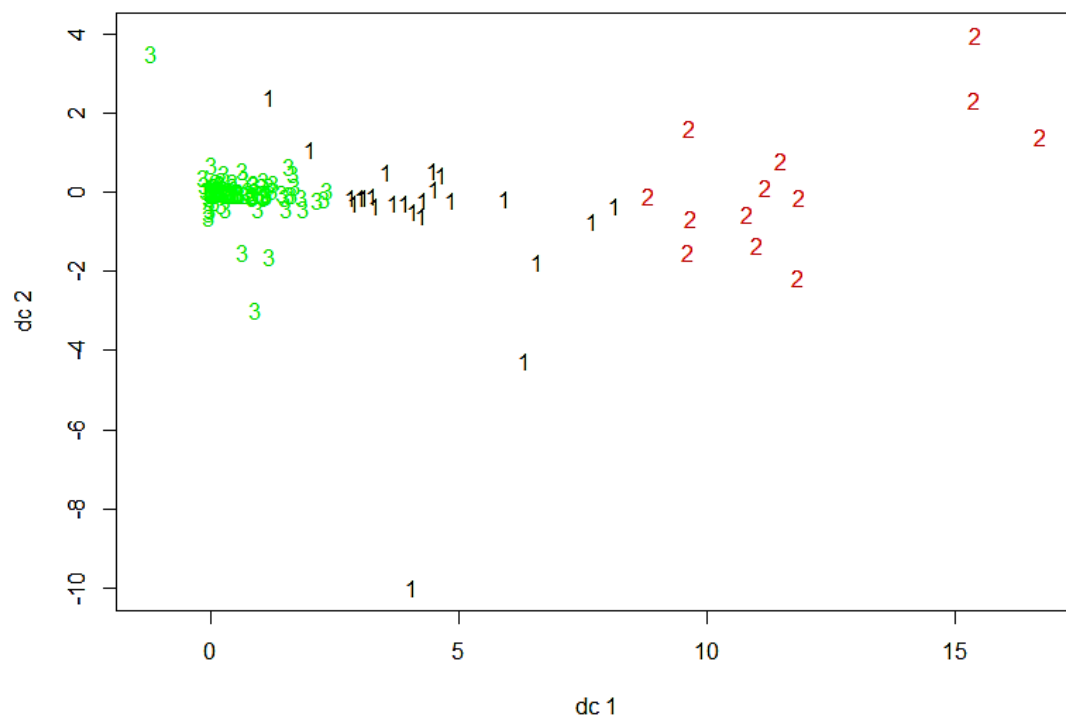


Figure 6. Cluster plot for 1st and 2nd discriminant functions.

Table 2. Clusters of countries based on agricultural imports.

Cluster 1	Cluster 2	Cluster 3		
South Africa	Algeria	Democratic Republic of the Congo	Armenia	Latvia
Tunisia	Egypt	Congo	Sri Lanka	Norway
Nigeria	Japan	Rwanda	Tajikistan	Russian Federation
Morocco	Republic of Korea	Ghana	India	Croatia
Bangladesh	China, mainland	Libya	Turkmenistan	Ukraine
Afghanistan	Italy	Guinea	Georgia	Republic of Moldova
Germany	Spain	Lesotho	Kyrgyzstan	The former Yugoslav Republic of Macedonia
Belgium	Netherlands	Angola	Uzbekistan	Albania
Portugal	Mexico	Central African Republic	Azerbaijan	Montenegro
United Kingdom	Viet Nam	Somalia	Democratic People's Republic of Korea	Bosnia and Herzegovina
United States of America	Indonesia	Burkina Faso	Maldives	Serbia
China, Taiwan Province of	Brazil	Gabon	Nepal	Iceland
Philippines	Saudi Arabia	Togo	Antigua and Barbuda	Finland
Malaysia		Sierra Leone	Barbados	Slovenia
Thailand		Liberia	Dominica	Canada
Colombia		Djibouti	Grenada	Oceania
Peru		Eritrea	Saint Kitts and Nevis	Micronesia
Turkey		Kenya	Saint Lucia	Kiribati
Yemen		Gambia	Saint Vincent and the Grenadines	Nauru
Iraq		Comoros	Cuba	Tonga
Iran (Islamic Republic of)		Cabo Verde	Guatemala	Tuvalu
United Arab Emirates		Côte d'Ivoire	Dominican Republic	Vanuatu
Israel		Cameroon	Jamaica	Samoa
Jordan		Burundi	Costa Rica	Australia
		Mali	Panama	New Zealand
		Chad	Honduras	Papua New Guinea
		Benin	Nicaragua	Solomon Islands
		Swaziland	Belize	Fiji
		Equatorial Guinea	El Salvador	China, Hong Kong SAR
		Zimbabwe	Trinidad and Tobago	China, Macao SAR
		Sudan	Bahamas	Singapore
		Malawi	Haiti	Lao People's Democratic Republic
		Mozambique	Switzerland	Myanmar
		Niger	France	Timor-Leste
		Guinea-Bissau	Luxembourg	Brunei Darussalam
		United Republic of Tanzania	Malta	Cambodia
		Senegal	Belarus	Venezuela (Bolivarian Republic of)
		Uganda	Poland	Argentina
		Zambia	Bulgaria	Chile
		Madagascar	Hungary	Ecuador

		Mauritius	Sweden	Paraguay
		Seychelles	Czechia	Uruguay
		Botswana	Denmark	Falkland Islands (Malvinas)
		Sao Tome and Principe	Romania	Bolivia (Plurinational State of Bolivia)
		Mauritania	Austria	Guyana
		Namibia	Greece	Suriname
		Kazakhstan	Slovakia	Cyprus
		Mongolia	Estonia	Bahrain
		Pakistan Syrian Arab Republic	Ireland	Kuwait
			Lithuania Lebanon	Oman
				Qatar
				Occupied Palestinian Territory

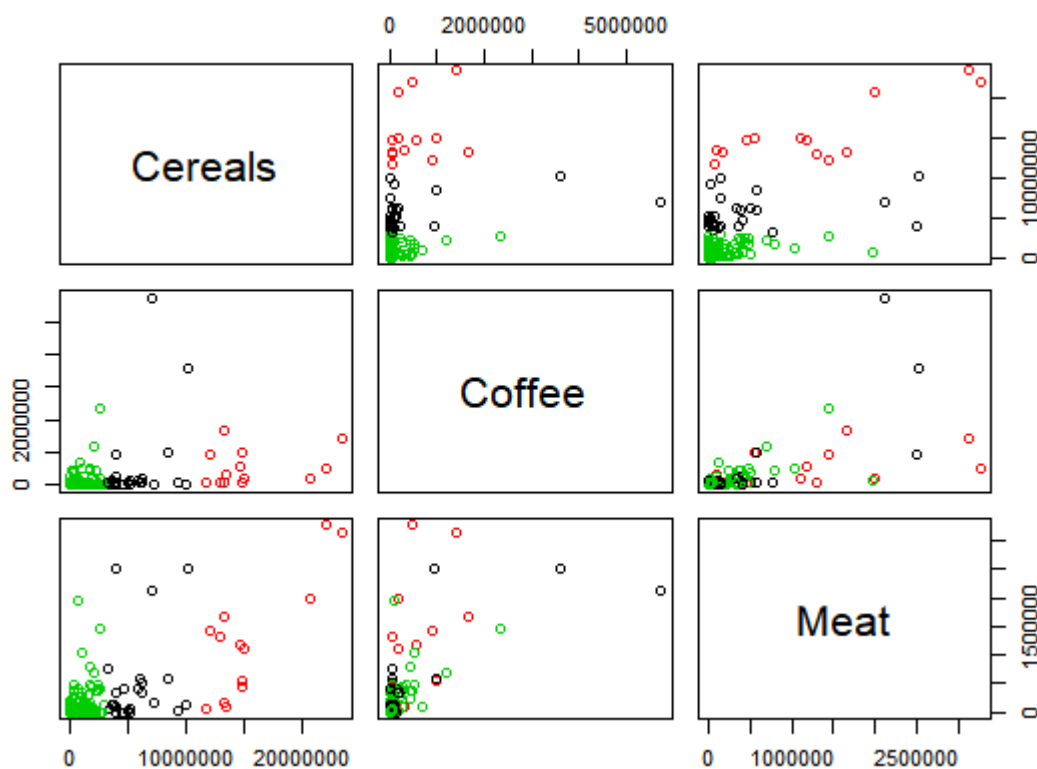


Figure 7. Matrix plot of country clusters.

Upon studying the cluster means, we can establish correlations between each group and the three classes of countries:

- Cluster 2 has the highest value for imports of cereals, meat, and coffee. There are 13 countries in this cluster.
- Cluster 1 is average for cereals and coffee imports and higher for meat imports. There are 24 countries in this cluster.
- The values for all three agricultural imports are lowest for cluster 3 where 159 countries fall under this category. This shows the agricultural well-being of these countries since the imports are less.

CONCLUSION

This research on employing unsupervised learning offers insights into performance metrics crucial for policymakers. This study has demonstrated the K-means clustering as a tool for analyzing

agricultural imports of countries together with visualization for interpreting the results. The countries were grouped into three clusters of sizes 24, 13, and 159 based on imports of cereals, meat, and coffee. This study concluded that 159 countries fall under the category of the cluster with mean imports of 6.64 Lakh tons, 0.72 Lakh tons, and 1.1 Lakh tons for cereals, meat, and coffee respectively. This shows the agricultural well-being of these countries since the imports are less compared to the rest of the two clusters. The findings indicate that K-means holds promise as a primary tool for conducting cluster analysis of countries grounded on their agricultural imports.

REFERENCES

- 1 Han J, Kamber M. Data Mining: Concepts and Techniques, 4th edition. Amsterdam: Elsevier; 2022.
- 2 Benos L, Tagarakis AC, Dolias G, Berruto R, Kateris D, Bochtis D. Machine learning in agriculture: A comprehensive updated review. *Sensors*. 2021;21:3758. DOI: 10.3390/s21113758.
- 3 Supriyatna A, Carolina I, Widiati W, Nuraeni C. Rice productivity analysis by province using K-means cluster algorithm. *IOP Conference Series: Materials Science and Engineering*. 2020;771:012025. DOI: 10.1088/1757-899X/771/1/012025.
- 4 Chen TC, Ibrahim Alazzawi FJ, Mavaluru D, Mahmudiono T, Enina Y, Chupradit S, et al. Application of data mining methods in grouping agricultural product customers. *Mathematical Problems in Engineering*. 2022;2022:1–9. DOI: 10.1155/2022/3942374.
- 5 Altuntas S, Selim S, Altuntas F. A hierarchical clustering based panel data approach: A case study of regional incentives. *International Journal of Information Management Data Insights*. 2022;2:100098. DOI: 10.1016/j.ijime.2022.100098.
- 6 Cereals imports quantity by country (2022). Knoema. Available from: <https://knoema.com/atlas/topics/Agriculture/Trade-Import-Quantity/Cereals-imports-quantity>. Accessed: 14 March, 2023.
- 7 Kamath RS, Jamsandekar SS, Kharade KG, Kamat RK. *Data Analytics in R: A Case Study Based Approach*. Mumbai, Maharashtra: Himalaya Publishing House; 2019.
- 8 Kamath RS, Kamat RK. Visualization of University Clusters based on NIRF and NAAC Scores: K-means Algorithm Approach. *Univ News*. 2019 Jan 21-27;57(03):3-15.
- 9 Jamsandekar SS, Kamath RS, Patil MB. Time series and cluster analysis of COVID-19 growth curve in India. *South Asian J Manag Res*. 2022;12:45–53.
- 10 Badadare VL, Jamsandekar SS, Kamath RS. Insights of exploratory data analysis (EDA) of COVID-19 Vaccine Administration in India. *South Asian J Manag Res*. 2022;12:36–44.
- 11 Coffee imports quantity by country (2022). Knoema. Available from: <https://knoema.com/atlas/topics/Agriculture/Trade-Import-Quantity/Coffee-imports-quantity>. Accessed: 14 March, 2023.
- 12 Meat imports quantity by country (2022). Knoema. Available from: <https://knoema.com/atlas/topics/Agriculture/Trade-Import-Quantity/Meat-imports-quantity>. Accessed: 14 March, 2023.