

COVID-19 effect on supply and demand of essential commodities using unsupervised learning method

P. Anitha (✉ anitha.palakshappa@gmail.com)

JSS Academy of Technical Education, Bengaluru-560060, Karnataka, India. <https://orcid.org/0000-0002-9904-8762>

Malini M. Patil

JSS Academy of Technical Education, Bengaluru-560060, Karnataka, India. <https://orcid.org/0000-0001-6959-7843>

Rekha B Venkatapur

KS Institute of Technology, Bengaluru-560060, Karnataka, India

Research Article

Keywords: Pandemic, Essential data, Forecasting, Prediction, Business Intelligence

Posted Date: December 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-110010/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Journal of The Institution of Engineers (India): Series B on June 10th, 2021. See the published version at <https://doi.org/10.1007/s40031-021-00594-6>.

COVID 19 effect on supply and demand of essential commodities using unsupervised learning method

Abstract: *The affliction caused by the Covid-19 Pandemic is diverse from other disasters seen so far. Supply chain industries are facing unique challenges in fulfilling the essential needs of the people. The objective of the paper is to analyse the supply and demand of essentials during pre-pandemic and post-pandemic lockdowns using machine learning algorithms. This helps for supply chain industries in forecasting and managing the supply and demand of essential stocks for the future. Data is analyzed using prediction algorithms to check the actual and predicted values. The clustering algorithm along with rolling mean is used for half-yearly data of 2019 and 2020 to identify the sales of different categories of essential commodities. This paper aims at applying intelligence in predicting various categories of sales by providing timely information for B2B Industries during the time of disasters.*

Keywords: *Pandemic, Essential data, Forecasting, Prediction, Business Intelligence.*

1. Introduction:

The pandemic is causing a high impact on the supply chain industries, which includes manufacturers, wholesalers, and retailers [2] all over the globe. Economically, affected countries are facing challenges related to the supply chain for transportation of essentials [9]. Covid-19 also affects the supply chain related to health care [5]. It causes suspension of retail trade, save for essential goods for sustainability (including medicines, food, and their supply chains) with financial, banking, and insurance services [4]. Industries are facing challenges in the supply chain for transportation of goods, especially essential grocery items during this COVID 19 and problem related to suppliers [5]. The challenging task faced by supply chain industries during a pandemic is predicting demand and supply, transportation issues, manpower issues, and government regulations. Managing these issues within and between the state has increased the attention of researchers towards the supply chain [1]. This type of disaster impacts mainly on customer behavior and preferences. Under this prevailing situation, customers are increasingly working out on what, where, and how the essential commodities are bought. Since the demand for essential commodities increases, industries are concentrating more on their supply chain for secure and immediate operations. At the same time, insight into the other categories of consumer needs also offers a preference on the consumer side.

A literature survey reveals that it is the consumer-driven business that needs to address from a supply chain perspective. Few facts to be engrossed for further analysis are summarized as follows.

1. Demand and supply: During this pandemic, companies are started facing huge demand for essential commodities which is not expected. This leads to a great challenge for the supply chain department. Also, it is difficult for Suppliers to arrange for such a huge demand. A contingency plan has been developed to take part in the supply of essential goods.
2. Manpower (labour issues): since lockdowns are unplanned, it created a serious issue on lack of manpower. So supply and demand depend on the manpower.
3. Maintaining safety: Another important challenge includes the safety of food items [3] and also the safety of people involved in transportation concerning SOP. It is important to check the safety while delivering the essentials and applications of disinfectants for surfaces and vehicles. Also thermal checks and sanitizers for people delivering the goods. Based on the service and policy environment Responsible Transportation is started with post-pandemic [8].
4. Government Regulations: It is important to know the reaction of the government rules and regulations which disturbs the supply chain, also to check whether alternative suppliers are available at a moment's notice.

To overcome the above issues, statutory bodies can inform the government and started receiving the e-passes for their transportation purpose. This leads to having better control over demand and supply of essentials.

The paper is organized as follows: Literature review is presented in section 2; Section 3 is about the methodology of the proposed work; Section 4 is about discussions on results and experiments followed with the conclusion in section 5.

2. Literature Survey:

In recent years, both national and global level supply chain risk management attracted the attention of researchers and practitioners [1]. Big data and machine learning approaches help in the detection of emerging risks, maintenance of relevant reports, and initiate suitable actions for a reformation of the supply chain [1]. Using analytics, supply chain issues like track and trace, route optimization, Green Logistics can be resolved [10]. During this pandemic, the supply chain has struggled for a steady flow of essential goods. So, the author discussed demand and supply challenges, technological challenges, and supply chain sustainability faced during COVID-19 [2].

The safety of food is another challenge in the field of the supply chain. The difficulties faced in each critical stage of the food supply chain, from farm to consumer has been explained and measures initiated to overcome these problems [3]. While the impact of COVID-19 is increasing, reduction measures are taken to reduce the risk across

the countries also increases [4]. Covid-19 disaster affects the supply chain related to health care, since the sudden rise in the demand for specific health care products [5]. Here healthcare equipment is considered as a product. K-Means is used to cluster the customer purchase based on their RFM values [6]. In future work, it is mentioned that K-Means can be used to cluster product wise sales for the given data [6].

Forecasting sales is another important segment of Business Intelligence [7]. Time series forecasting is used for validating the sales results obtained from the predictive machine learning models [7]. Because of the pandemic, transportation policies are reframed to solve the issues related to existing approaches [8]. Linear Regression is used to predict and compare the sales of a month [10]. Many research areas have been emerged in describing and solving the issues related to COVID-19. Few are supply chain, health care, economic, Information technology, sustainability, psychological problems, and many more [9].

3. Methodology of the work:

This section describes the comparison of data between pre and post-pandemic using the Prediction method and Clustering approach. The prediction of data on essential commodities is carried out based on their categories. There are 8 major categories in the dataset like Flours, Rice, Sugar, Grains, Pulses, Oil, Seasonal food, and Dried nuts. Each category consists of subcategories as shown in Fig 1. Only a few subcategories are mentioned in the diagram whereas the dataset includes other subcategories also.

The proposed work is implemented on the dataset collected from one of the B2B industry. The proposed approach includes 2 methods: Prediction/Forecasting and Data Clustering.

3.1 Prediction: Regression analysis helps to forecast the dependent variable based on one or more independent variables [10]. One of the important part of Business Intelligence in the current period is Sales prediction [7].

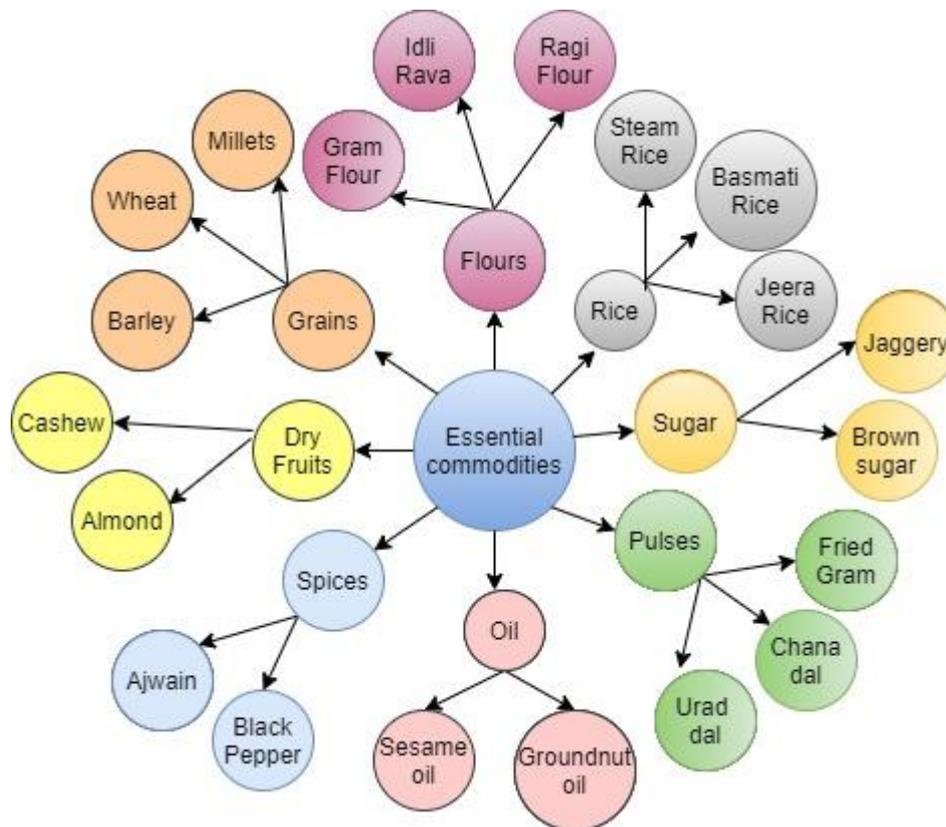


Fig 1: Categories and subcategories in Essential commodities

. In this work, value generated is used as a dependent variable whereas independent variables are considered as individual store number, month, category and quantity sold. The equation generated using dependent and independent variables is termed as a Regression model. Value of dependent variable changes based on the month wise execution.

3.2 Clustering: since the forecasted data is varying from pre and post covid after comparing actual and predicted values, k-means algorithm is used to analyse:

- a. what categories of data is varying?
- b. How much amount of different categories are varying in month wise?

K-Means: This helps us to analyse the sales of essential commodities month wise. Three clusters are formed as shown in the Fig 2, Lowest, Average and High sales. Out of 8 categories, K-Means is used to group the categories based on the 3 clusters formed for each month. Steps of the proposed work is as follows:

Step 1: Input of half yearly Dataset of 2019.

Step 2: Forecasting on the analysis of category and month wise separately.

Step 3: Input of half yearly Dataset of 2020.

Step 4: Comparison between predicted values of 2019 and 2020

Step 4: Major variation seen in the month of March to May, that is post covid prediction compared to pre pandemic.

Step 5: K-Means is applied for 2020 data to check the sales of various categories separately.

4. Experiment and Results:

Dataset includes 6 months (Jan to June) sales of essential commodities of the year 2019 and 2020. Table 1 describes the dataset used for the work. Initially Multiple Linear Regression is used to forecast the half yearly sales of 2019.

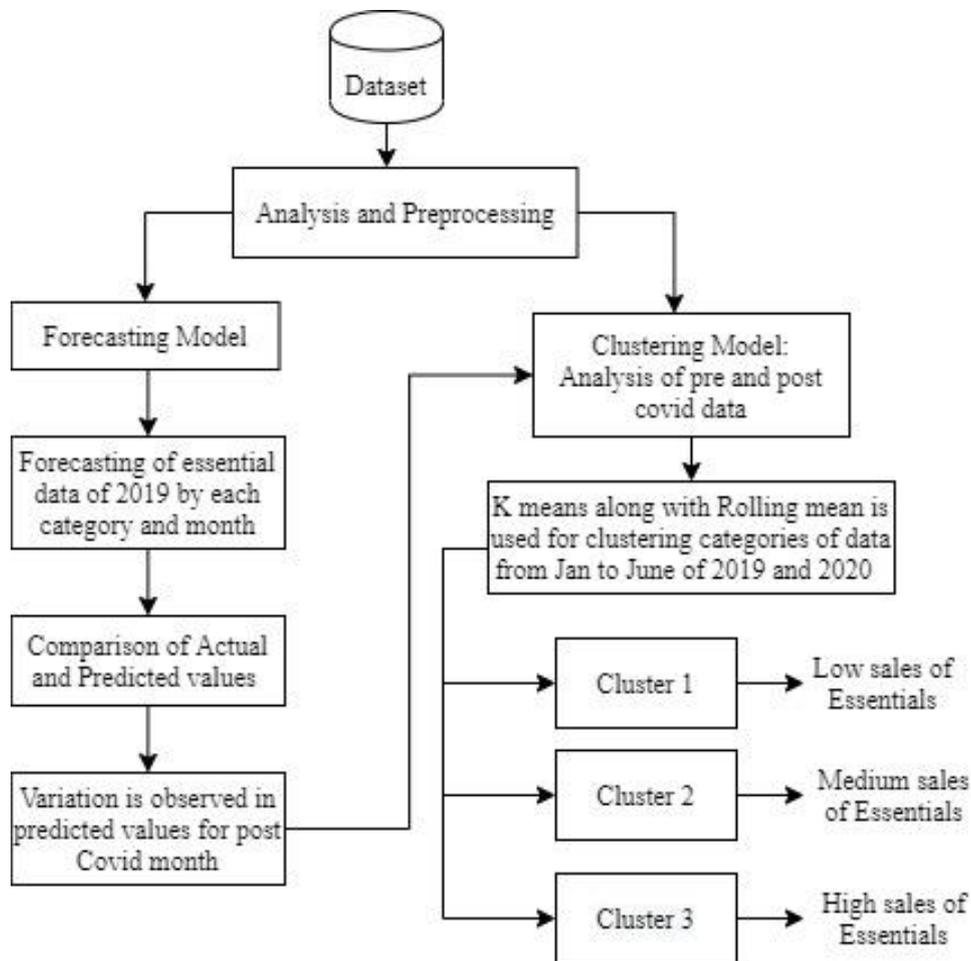


Fig 2: Proposed Approach

Table 1: Description of data

Sl No	Name of the Feature	Type of the Feature	Description of the feature
1	Store	Numeric	2-digit unique number assigned for each store.
2	Item_Number	Numeric	5 digit number assigned for each item
3	Quantity	Numeric	Article sold in terms of kilograms

4	Value_Generated	Numeric	Value generated from the sales of each article in a particular date from particular store.
5	Invoice_Date	Nominal	Date of sale of article
6	Month	Numeric	Represents Jan to June as 1 to 6
7	Category	Nominal	It represents the category, in which essential food items belongs to.

Fig 3 gives the prediction for 8 different categories for 6 months. It consists of number of instances, accuracy and root mean squared error(RMSE) values for each category of 2019 dataset. No entry in the table of Fig 3 represents no sales of the category in that month.

Predicted values are compared with the actual values of 2020 month. Because of the pandemic there is a variation in the month of March to May 2020 compared to predicted values of 2019. Fig 4 and 5 gives six-month percentage wise comparison of actual and predicted values of Category-Rice and Pulses.

Category	Jan			Feb			Mar			Apr			May			Jun		
	Instances	Accuracy	RMSE															
Rice	1139	92%	2.7	1160	95%	2.8	1087	94%	2.3	1057	94%	3.2	1085	94%	3.2	1028	93%	3.6
Pulses	1408	99.37	9.5	1481	99.37	1.1	1404	99.35	9.1	1346	99.3	9.3	1433	98.11	15.1	1433	98.11	15.1
Grains	95	94.45	5.4	109	98.5	1.7	168	96.55	2.6	196	92.42	2.9	130	97.19	2.5	82	94.78	2.5
Flours	203	97.97	3.8	251	97.51	3.6	256	96.3	5.2	237	97.65	3.6	160	98.83	3.7	129	98.22	7.4
Oil	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	12	50.41	47
Spices	4	-	-	78	90.71	7.2	8	-	-	141	85.38	9.7	175	85.71	42.2	159	87.8	29.01
Sugar	-	-	-	24	99.97	4.4	60	99.99	4.4	55	99.97	7.1	47	99.83	7.2	47	99.83	7.2
Dry Fruits	-	-	-	10	69.57	9.1	-	-	-	10	23.3	13.9	-	-	-	-	-	-

Fig 3: Category and Month wise prediction

According to the graph in Fig 4 and Fig 5, January and February month comparison of predicted and actual values are approximately equals to 99% whereas March to May values are varying by 64% for rice and 66% for pulses. Since actual sales value of post covid is more, the percentage in the comparison graph is decreasing. In June month, the graph is becoming normal as the comparison increases. Similarly remaining 6 categories of essentials are compared and analysed.

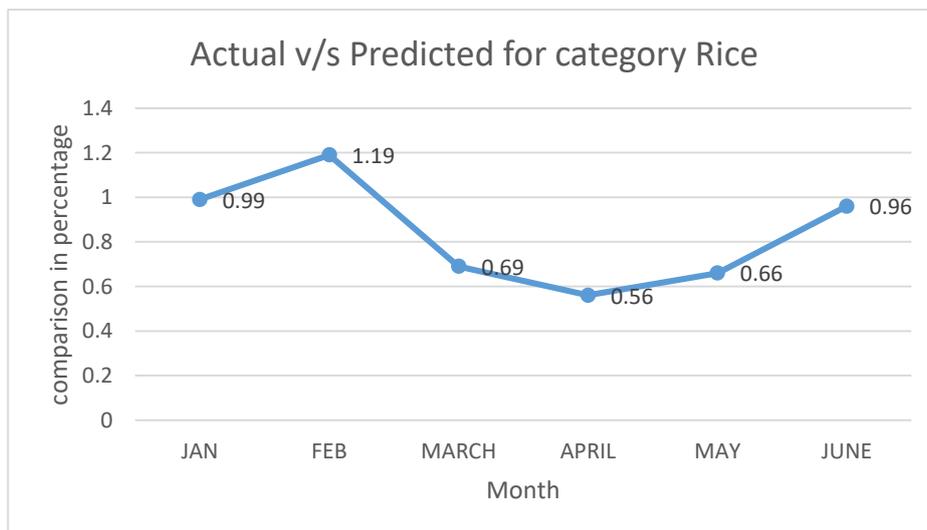


Fig 4: Comparison of Rice during pre and post covid.

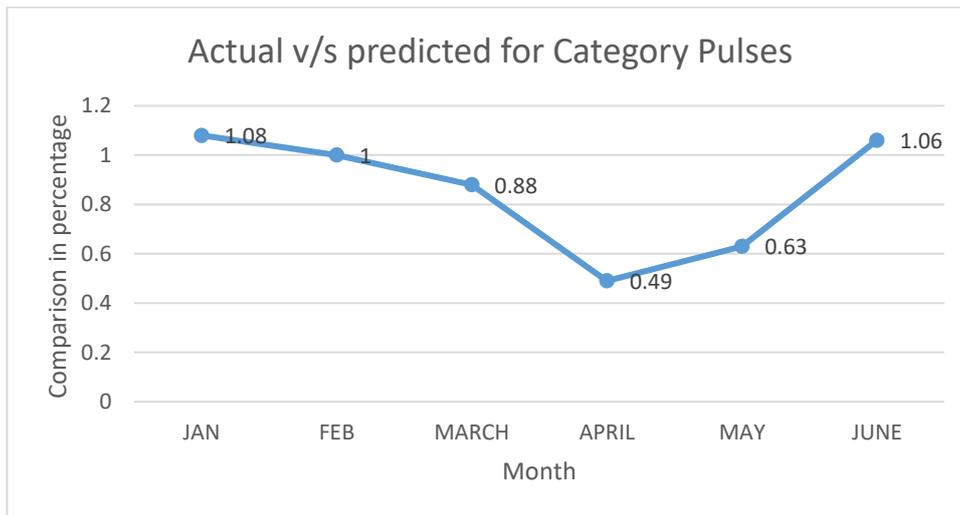


Fig 5: Comparison of Pulses during pre and post covid.

Clustering Algorithm: K-Means algorithm is selected along with rolling mean method. Compared to other clustering methods, K-Means chooses only 'k' as a single input parameter [6]. Rolling mean is used to analyse the dataset by taking average or mean value of dataset. Initially for calculation of current month sales, mean value of previous 2 month sales is used, which gives the sales value of current month. Mean value of each category is calculated separately. For example, to predict the march month sales, previous Jan and Feb data is used. Similarly, for April month calculation, Feb and March values are used.

To begin with a cluster, 3 groups or classes are formed based on the mean value calculations. The cluster centroids are chosen randomly. 3 clusters are classified as low, average and high sales of each category month wise. This helps the supply chain industry to analyse the category of essentials sold high, medium or low during this pandemic. Also they can compare with the previous year sales data of each category using the same algorithm. Fig 6 and Fig 7 shows the clustering for category Rice and Pulses for April 2020 dataset. Also it is calculated and analysed for other categories in the dataset. Figure 6 represents comparison of 2020 March sales of Categories, which is divided into 3 clusters based on their centroid values. According to the Fig 6, Purple (High Sales value), Red (Average sales value) and Green (Low sales value). Small circles in the figure gives the sales of each category. Small circles with purple colour is grouped into high sales cluster, red small circles to the average case and green small dots to Low sales cluster.

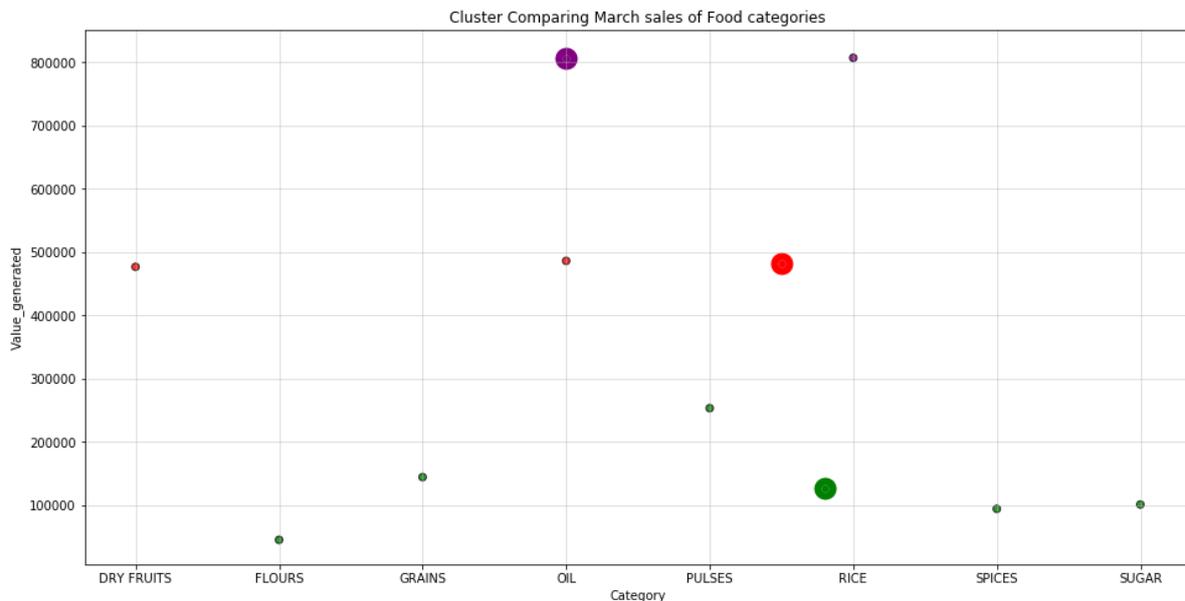


Fig 6: Clusters comparing March 2020 sales of different food categories.

Based on the Fig 6, comparative analysis of 2019 and 2020 sales value v/s category for the remaining months are implemented. The results obtained are noted separately for 2019 and 2020 for 8 categories as shown in the Table 2.

Table 2: Comparative analysis of month wise sales of various categories of essentials.

Comparison	March		April		May		June	
	2019	2020	2019	2020	2019	2020	2019	2020
Rice	2	3	2	3	1	3	1	1
Pulses	3	1	3	2	2	1	3	3
Grains	1	1	1	1	1	2	1	2
Flours	2	1	2	1	1	1	1	3
Oil	1	2	1	3	1	3	1	1
Sugar	1	1	1	1	1	1	1	2
Dry Fruits	3	2	3	2	3	2	2	1
Spices	1	1	1	1	1	1	1	3

*HIGH:3, AVG:2, LOW:1

5. Conclusion:

Prediction and clustering methods aims at the analysis of pre and post COVID data related to essential categories. Forecasting helps the Supply chain industries to improve their supply based on the percentage of demand as per the analysis done in the proposed work. This helps supply chain to analyse the issues regarding demand and supply during this type of disaster. The clustering method improves the analysis of essential categories for pre and post-COVID months. The proposed work and the obtained results provide the strong base for the business organizations for comparing the value generated before and during the time of the pandemic. The work carried out if found to be innovative in its own approach using the unsupervised learning approach.

References:

1. Baryannis, George, Samir Dani, and Grigoris Antoniou. "Predicting supply chain risks using machine learning: The trade-off between performance and interpretability." *Future Generation Computer Systems* 101 (2019): 993-1004.
2. Sharma, Amalesh, Anirban Adhikary, and Sourav Bikash Borah. "Covid-19's Impact on Supply Chain Decisions: Strategic Insights for NASDAQ 100 Firms using Twitter Data." *Journal of Business Research* (2020).
3. Rizou, Myrto, et al. "Safety of foods, food supply chain and environment within the COVID-19 pandemic." *Trends in Food Science & Technology* (2020).
4. Honerkamp, Yasine. "Initial impacts of global risk mitigation measures taken during the combatting of the COVID-19 pandemic [Summary]." (2020).
5. Govindan, Kannan, Hassan Mina, and Behrouz Alavi. "A decision support system for demand management in healthcare supply chains considering the epidemic outbreaks: A case study of coronavirus disease 2019 (COVID-19)." *Transportation Research Part E: Logistics and Transportation Review* (2020): 101967.
6. Anitha, Palaksha, and Malini M. Patil. "RFM model for customer purchase behavior using K-Means algorithm." *Journal of King Saud University-Computer and Information Sciences* (2019).
7. Pavlyshenko, Bohdan M. "Machine-learning models for sales time series forecasting." *Data* 4.1 (2019): 15.
8. Budd, L., & Ison, S. (2020). *Responsible Transport: A post-COVID agenda for transport policy and practice*. *Transportation Research Interdisciplinary Perspectives*, 6, 100151.
9. Haleem, A., Javaid, M., Vaishya, R., & Deshmukh, S. G. (2020). Areas of academic research with the impact of COVID-19. *The American Journal of Emergency Medicine*.
10. Anitha, P., and Malini M. Patil. "A review on data analytics for supply chain management: a case study." *International Journal of Information Engineering and Electronic Business* 11.5 (2018): 30.

Figures

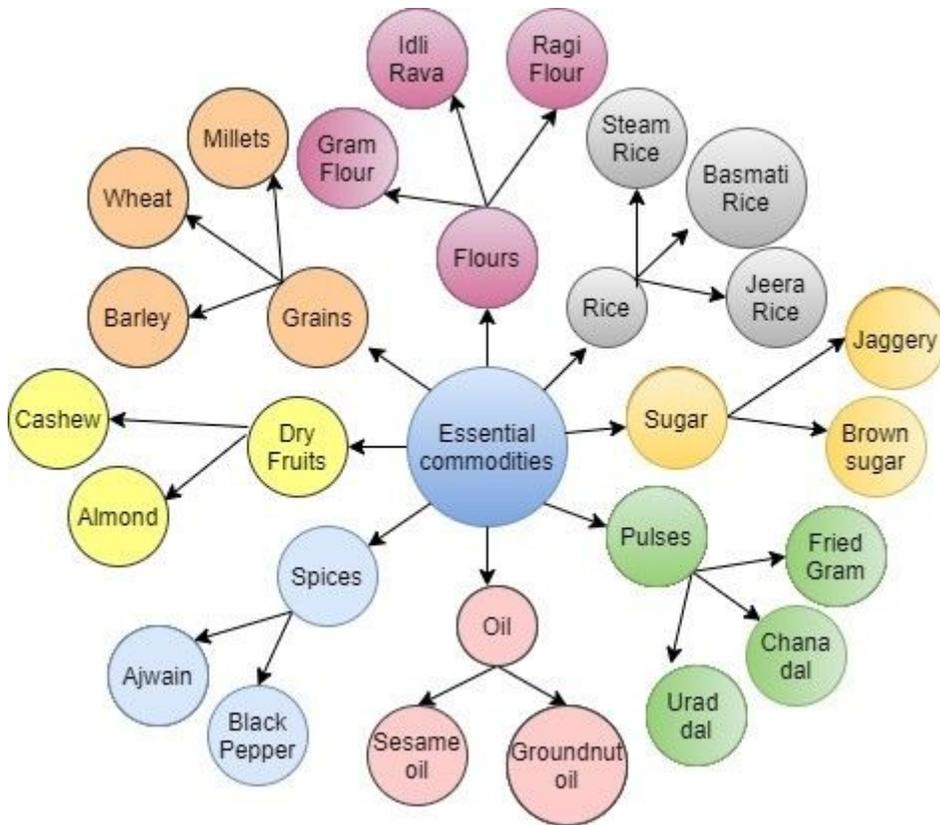


Figure 1

Categories and subcategories in Essential commodities

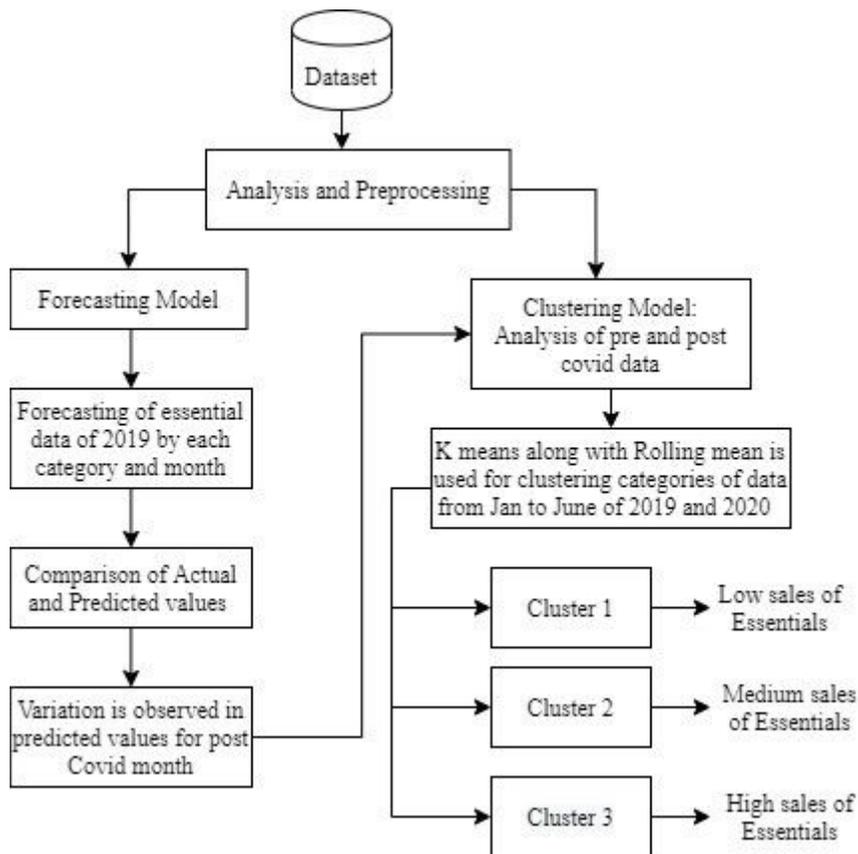


Figure 2

Proposed Approach

Category	Jan			Feb			Mar			Apr			May			Jun		
	Instances	Accuracy	RMSE															
Rice	1139	92%	2.7	1160	95%	2.8	1087	94%	2.3	1057	94%	3.2	1085	94%	3.2	1028	93%	3.6
Pulses	1408	99.37	9.5	1481	99.37	1.1	1404	99.35	9.1	1346	99.3	9.3	1433	98.11	15.1	1433	98.11	15.1
Grains	95	94.45	5.4	109	98.5	1.7	168	96.55	2.6	196	92.42	2.9	130	97.19	2.5	82	94.78	2.5
Flours	203	97.97	3.8	251	97.51	3.6	256	96.3	5.2	237	97.65	3.6	160	98.83	3.7	129	98.22	7.4
Oil	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	12	50.41	47
Spices	4	-	-	78	90.71	7.2	8	-	-	141	85.38	9.7	175	85.71	42.2	159	87.8	29.01
Sugar	-	-	-	24	99.97	4.4	60	99.99	4.4	55	99.97	7.1	47	99.83	7.2	47	99.83	7.2
Dry Fruits	-	-	-	10	69.57	9.1	-	-	-	10	23.3	13.9	-	-	-	-	-	-

Figure 3

Category and Month wise prediction

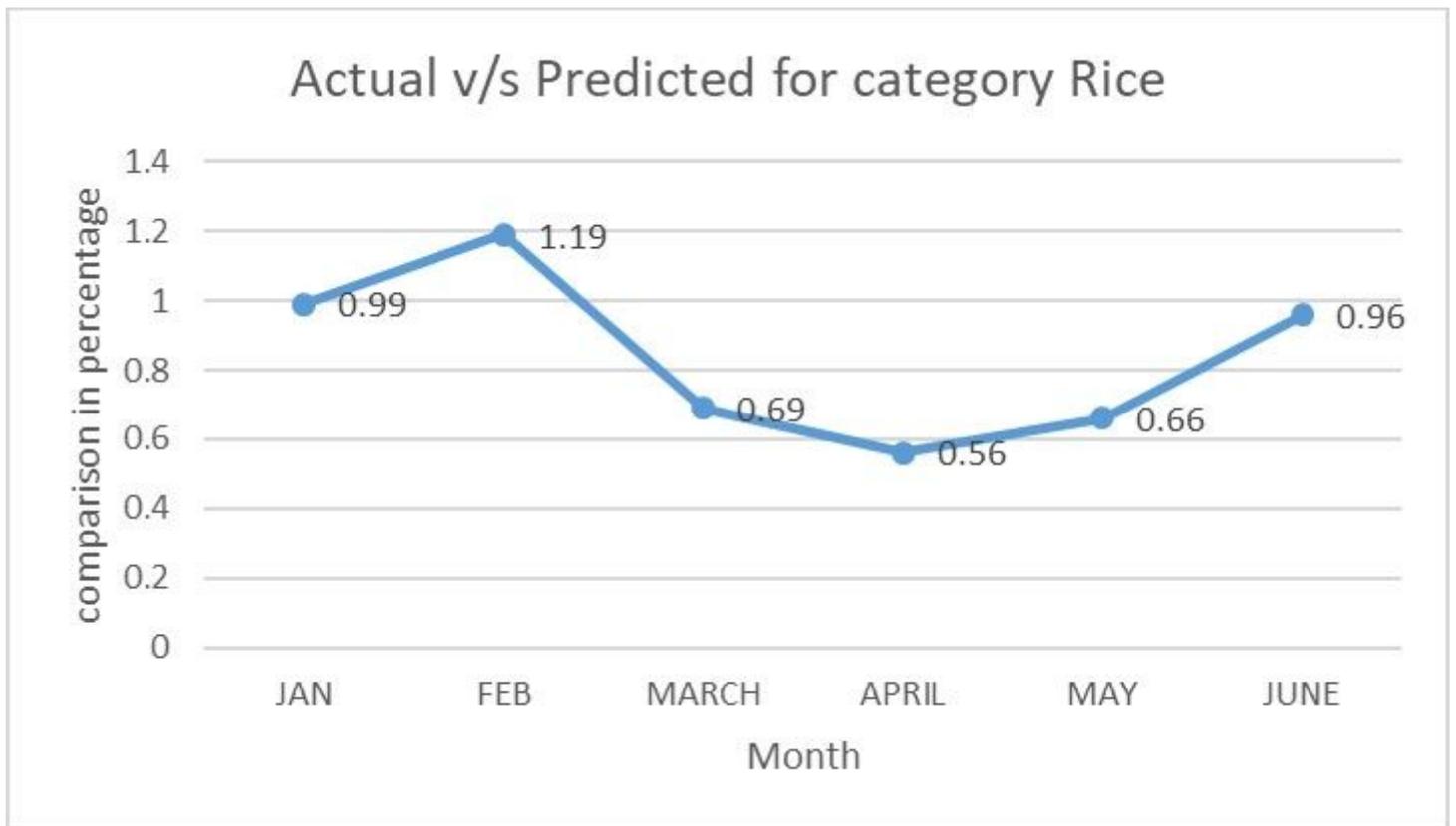


Figure 4

Comparison of Rice during pre and post covid.

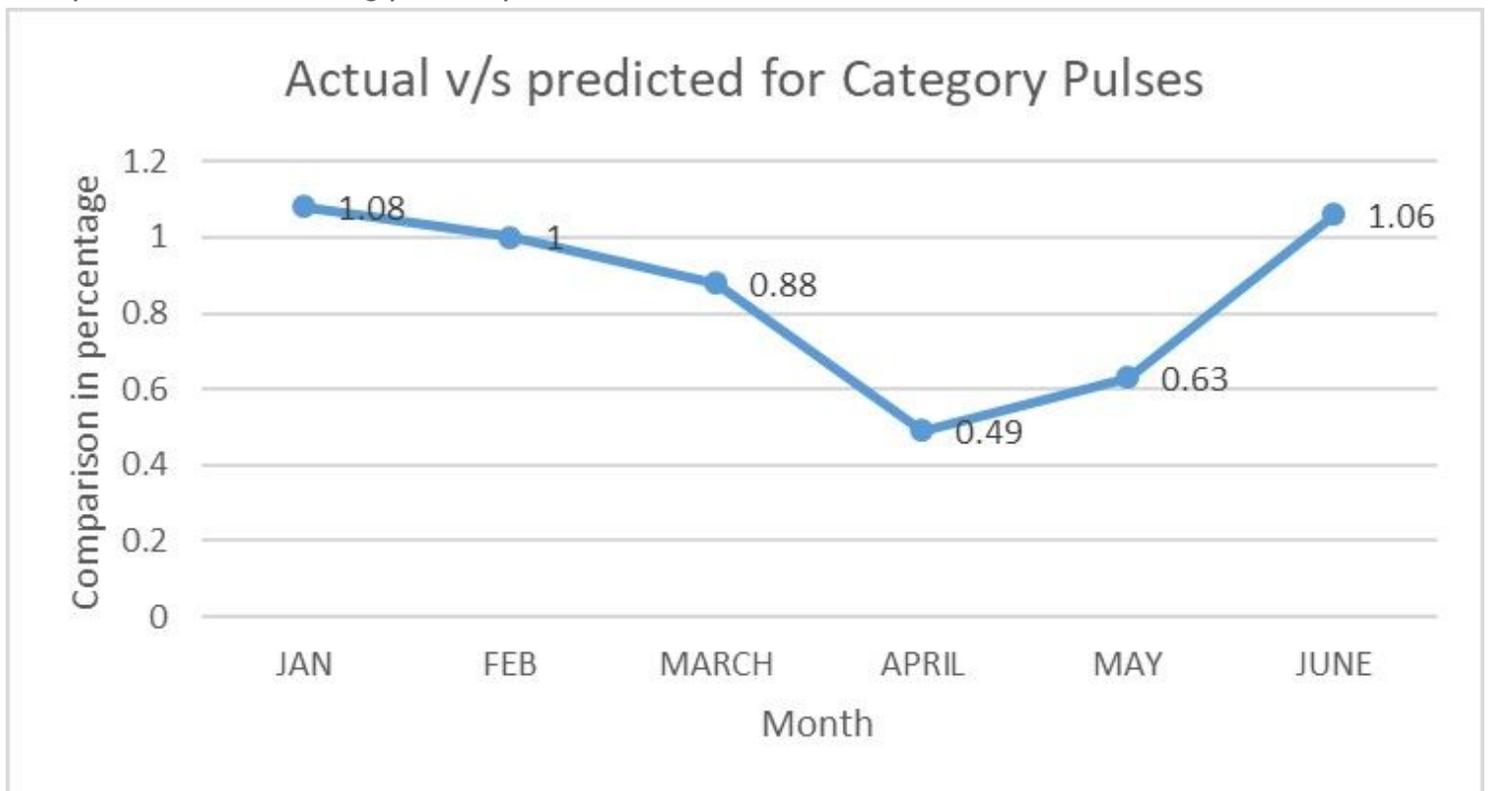


Figure 5

Comparison of Pulses during pre and post covid.

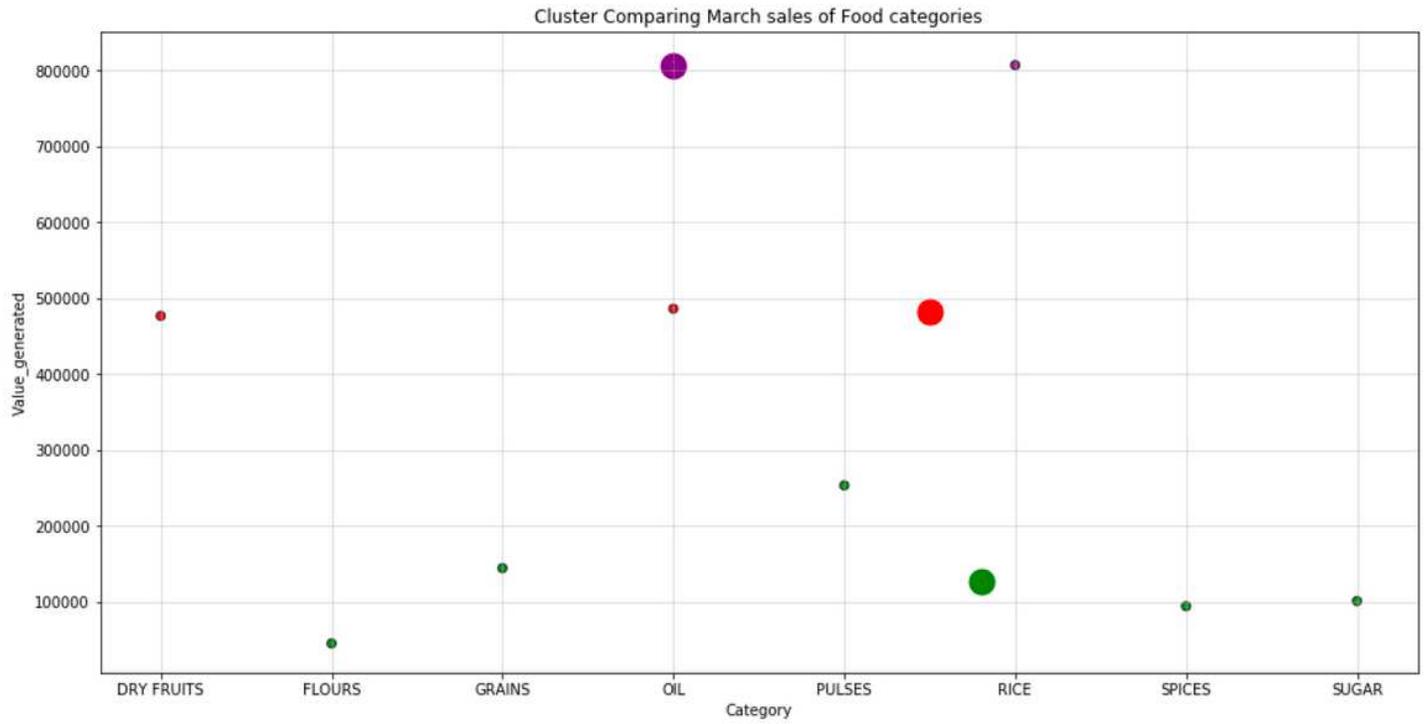


Figure 6

Clusters comparing March 2020 sales of different food categories.