DELIVERING LOCALLY SOURCED NUTRITIOUS FOOD TO INDIAN HOUSEHOLDS

by

Sanchita Das
B.Arch (Honors) and Minor in Economics

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ENGINEERING IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2021

Signature of Author: ______________________________

Department of Supply Chain Management
May 14, 2021

Certified by: ______________________________

Dr. Chris Mejia Argueta
Founder and Director, MIT Food and Retail Operations Lab
Director, MIT SCALE Network for Latin America & The Caribbean
Thesis Advisor

Certified by: ______________________________

Dr. Tatiana Sadalla Collese
Postdoctoral Associate
Thesis Co-Advisor

Accepted by: ______________________________

Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

DELIVERING LOCALLY SOURCED NUTRITIOUS FOOD TO INDIAN HOUSEHOLDS

by
Sanchita Das
Submitted to the Program in Supply Chain Management
on May 14, 2021 in Partial Fulfilment of the
Requirements for the Degree of Master of Engineering in Supply Chain Management

ABSTRACT

WHO reports that in the South Asian region, the number of undernourished has hardly decreased in the last decade. This situation calls for a concerted effort to combat malnutrition the world over. The effort must be grounded in nutrition and executed through a robust distribution mechanism to reach all segments of the society. In this thesis, we take a step in that direction by combining expertise from the supply chain and nutrition areas to address protein-energy malnutrition among poor households in India. While the country is the largest producer of pulses, milk, and other dairy products, and many food grains, Indian diets are traditionally low in protein intake, especially among the poor. Within our scope of the problem, we target the poorest households in India, which currently hold the Antyodaya Anna Yojana ration cards from the Government of India. We develop a framework to improve their diet diversity nutritionally. We propose matching the demand of food (as recommended by Indian Council of Medical Research for a balanced diet) with locally available, culturally preferred supply by designing 'customized food baskets' for different consumer clusters. We suggest distributing the proposed food baskets at scale to all target households via the government Public Distribution System mechanism operational in India. We use PCA and K-means clustering to segment the customers, create a food basket model inspired by the knapsack problem, and use a Mixed Integer Linear optimization program to solve the distribution problem. The key contribution of this thesis is a framework of basket assortment and distribution. The approach is generalizable and can be used on many different customer types and (public or private) distribution channels to match demand with supply of nutritious assortments and enable delivery at scale. We can serve 65 to 75% of recommended daily quantity of cereals and pulses to our target households via the proposed framework.

Thesis Advisor: Dr. Chris Mejia Argueta
Title: Founder and Director, MIT Food and Retail Operations Lab, the U.S.

Thesis Co-Advisor: Dr. Tatiana Sadalla Collese
Title: Postdoctoral Associate, Universidade de São Paulo, Brazil

ACKNOWLEDGMENTS

Throughout the work on this thesis, I was consistently supported by a few people. I am deeply grateful to my advisor Dr. Chris Mejia due to whose foresight we could scope the problem early on and get valuable help from my co-advisor Dr. Tatiana Collese, who is a nutrition expert. Dr Chris Mejia's Food and Retail Operations Lab fascinated me as an aspirant to MIT. The exposure by working with him in the last 10 months has been very gratifying and has fuelled my conviction for pursuing a research career in this area. The discussions between Dr Chris, Dr. Tatiana and myself have been very enlightening and I have received constructive feedback from time to time on scoping the big picture, attention to methodological detail, and access to experts from other areas. Thanks, are also due to the Center for Transportation Logistics at MIT for curating this unique program to address the learning needs of people with supply chain experience. The coursework helped complement my research and writing.

I would like to thank my family and friends without whose unflinching support, I could never pursue my calling. I am privileged to have a strong support system in them all the way.

I would also like to thank the University of Washington, Foster Business School for being very kind and supportive of my decision to pursue this unique opportunity at MIT. This was a very challenging year world over, and I am grateful for their consideration of my request to learn from experts here.

I am thankful to Professors Sarang Deo and Sripad Devalkar at the Indian School of Business, Hyderabad who patiently guided me through a research assistantship, which motivated me to pursue a career in research. Thanks, are also due to Flipkart, a Walmart Group company in India, where my work experiences helped grow a curiosity for the supply chain area. Finally, I am grateful to Professor Pulak Mishra, a faculty of Economics at the Indian Institute of Technology Kharagpur who thoroughly inspired me by his classes in Public Finance and Policy, and Econometrics. My interest in public systems area is largely due to his excellent teaching during my undergraduate education in India.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES IN APPENDIX

LIST OF FIGURES IN APPENDIX

1. INTRODUCTION

Lack of dietary diversity is a severe problem among low-income populations in the developing world because their diets are predominantly based on starchy (i.e., carbohydrate and calorie rich) staples. These diets often include few fresh fruits and vegetables and mainly comprise grains (Ruel, 2003). Such diets often fall short of meeting various nutritional needs required to achieve food security (in terms of balanced diets) in these vulnerable population groups. Mark, et al., (2016) suggest a high prevalence of nutrient inadequacies in the food supply in several countries in the South Asia region. These inadequacies were most severe among the low and lower-middle income countries, with Bangladesh, Sri Lanka and India being the three countries with the highest nutritional inadequacies.

India is the world's largest producer of milk, pulses, and millets, and the second largest producer of rice, wheat, sugarcane, groundnuts, vegetables, fruits, and cotton (FAO-a, 2020). However, the Indian population still suffers from food insecurity. As agreed upon at the 1996 World Food Summit (FAO, 1996) hosted by the Food and Agriculture Organization, food security is achieved 'when all people, at all times, have physical, social, and economic access to sufficient, safe, and nutritious food that meets their dietary needs and food preferences for an active and healthy life' (FAO-b, 2006).

To enable access to food grains required for sustenance of all segments of the society, the government in India operates a food security program called the Public Distribution System (PDS). The beneficiaries of this scheme are entitled to purchase cereals (primarily rice and wheat) at subsidized prices from government fair price shops. Given the affordable access, rice and wheat have become the choice of staples for consumption in most of the households that are beneficiaries of the PDS. However, rice and wheat have low protein content (Davis, et al., 2019). In contrast, coarse cereals, pulses, and legumes, which are also locally grown in a wide variety by different states of India, are better sources of proteins.

The proliferation of fine cereals in the PDS supply chain can be traced back to the advent of the Green Revolution in India in the 1960s. The Indian government promoted high-yield varieties of wheat and rice, leading to a reduction in the land on which (the more nutritious) coarse cereals were grown. The increase in high-yielding cereals at the expense of more indigenous varieties of cereals over the last 50 years has reduced the nutritional content of the cereal supply (DeFries, et al., 2016) which is the mainstay of Indian diets. Household consumption estimates[1] of nutritious cereals, which historically had been staples for the poor, declined by more than half between 2004-2005 and 2011-2012, while consumption of rice and wheat from the government's PDS more than doubled (DeFries, et al., 2016).

As per the National Sample Survey Organization's (NSSO) Consumption Expenditure survey, in 2011-2012, the share of energy intake contributed by cereals (i.e., rice, wheat, sorghum, pearl millet, finger millet, barley, small millets and their products) for rural Indians ranged approximately from 70% for the bottom'5% of the population (by monthly expenditure) to 42% for the top 5%. Animal sources, and fats fulfilled energy needs of the high-income groups. As can be seen, the choice of staples in low-income households is directly linked to affordability and to access (which is enabled by the PDS). With the 2013 National Food Security Act, the government intended to promote coarse cereals in the PDS (Rao, et al., 2018), but efforts are in their infancy.

The opportunity for each state to tap into this policy development is huge. Each Indian state grows a set of grains -- cereals (fine and coarse) and pulses -- traditionally suited to the local agro-ecological conditions. Therefore, local, short supply chains can greatly help in managing perishability and freshness and thus, in the deployment of high-performance food access models and distribution systems to world's communities. Local sourcing also presents opportunities of cost efficiency due to short distances that the food travels from the producer to reach the consumer's plate.

Studies in recent years have acknowledged the PDS as a useful mechanism for distribution of coarse cereals at scale in India because of its design that supports both producers (through price support) and consumers (through subsidized prices of grains) (Devalkar, et al., 2020). However, to the best of our knowledge, there is no systematic large-scale study to design a mechanism to include locally grown pulses and legumes in the PDS for increasing food diversity, improving nutrition for underserved households affordably at scale and reducing

[1] Consumption estimates are obtained by National Sample Survey Organization of India in the National Consumption Expenditure Survey, 68th Round. Details in the Section 3, Data. Sample households report an estimate of their consumption based on purchase in the last 30 days.

logistics costs. Our work proposes such alternative to improve dietary variety and quality in Indian households for improved nutrition outcomes. We draw conclusions from locally grown cereals and pulses, satisfying local taste preferences to build an assortment of items that is nutritionally balanced.

We build a methodology using mathematical modelling to understand the impact of supplementing Indian PDS with other nutritious millets. We use cluster analysis and an optimization model. Both techniques use input data that estimates household's purchase-consumption patterns and supply of grains from the public datasets like (ICMR, 2020), (Crop Production Statistics, 2019-20), and NSSO Consumption Expenditure Data,68th Round, 2011-12 (hereafter referred as, NSSO 2011-2012). The production dataset, which is our key input is for the year 2019-2020 and has a level of detail – 'per district' of Indian subcontinent. To model other relevant parameters such as the distance between districts, time taken for travel, and costs, we extract data from Google APIs, Food Corporation of India or we calibrate data to build estimated functions. The optimization formulation contains overall close to 80,000 constraints and on the order of 10 million decision variables, reason why we called it a large-scale distribution model.

Through this thesis, we intend to make the following contributions:

1. Determine a set of customized food assortments (i.e., food baskets) for distribution to different consumer clusters in India constrained by local production patterns and traditional taste preferences of consumers. For household diets, we particularly aim to improve the diversity of food group combinations purchased and enhance the protein content by adding pulses into the food basket.
2. Propose a distribution model for suggested grain baskets through PDS and alternative distribution channels using a Mixed Integer Linear Program (MILP) model.
3. Analyze scenarios/ strategies to check generalizability of results and applicability amidst diverse circumstances for India.

These steps will culminate into food security (addressing both hunger and nutrition) through a robust, localized supply chain. We find that all our target households across India can be segmented into six customer clusters with the principal factors of clustering being geography, and historical household purchase of a few key cereals and pulses like rice, wheat, jowar, masur, urd, arhar and gram. Our model suggests that the proposed food baskets can cater to around 100% of cereals and about 65% of pulses consumption recommended for healthy adults as proposed by the Indian Council of Medical Research (ICMR). Finally, our distribution model can achieve the required procurement of cereals and pulses by a localized, decentralized procurement, the details of which are elucidated in Section 5.

In the following sections, we describe the insights from Literature review in Section 2, the methodology and data in Section 3, results in section 4 and the managerial insights and discussion in Section 5.

2. LITERATURE REVIEW

This thesis relates and contributes to the literature in the following areas: i) Design of affordable, locally sourced nutritious food baskets for certain segments of Indian households; ii) Synergies between nutrition science and supply chain-based frameworks to combat food insecurity, and iii) Exploration of supply chain strategies for large-scale distribution of proposed food baskets.

In the following sections, we summarize our learnings from the literature. The first stream is about food systems design and its ‘nutrition’ dimension, especially in relation to the sustainable development goals. In the second stream, we deep dive into studies that present the local availability and consumption patterns of food groups in different parts of India. The third stream is about methods that help building consumer awareness of healthy food. Finally, the fourth stream addresses our solution approaches: i) a bin-packing or knapsack approach to formulate a basket of food groups for consumption in each major geographical cluster, and ii) a distribution model for large-scale distribution of the proposed baskets to households in the geographic clusters. We discuss various studies in the nutritional epidemiology and operations/supply chain management areas that address these problems and elucidate the novelty of our thesis in this context.

2.1 Food Systems Design: Context of Nutrition

Food is the strongest lever to optimize human health and environmental sustainability on Earth. However, food is currently threatening both people and the planet. An immense challenge facing humanity is to provide a growing world population with healthy diets from sustainable food systems. While global food production of calories has generally kept pace with population growth, more than 820 million people still lack sufficient food, and many more consume either low-quality diets or too much non-nutritious food (EAT, 2019).

World Health Organization, (2017) reports 462 million adults worldwide continue to be underweight, while 1.9 billion people suffer from overweight or obesity. Breaking down some facts, the same report documents that around 264 million women of reproductive age around the world are affected by iron deficiency-related anaemia, 155 million children under the age of 5 years around the world are stunted (i.e., low height for age) and that 41 million children under the age of 5 years worldwide suffer from overweight.

Nutrition and logistics are at the nucleus of efficient food ecosystems. When nutrition is properly linked to food supply and production systems, it usually triggers higher affordability, accessibility for nutritious food to end consumers. Current public health nutrition approaches should integrate food and nutrition systems with a wider focus on social, cultural, economic and environmental aspects and solutions (World Public Health Nutrition Association. The Giessed Declaration, 2005). Because human diet is complex, encompassing different factors such as biological, psychological, and social, that are intrinsically related to individuals. It depends on biological factors like age, gender; psychological factors like taste preferences and consumer choice; and social factors like cultural beliefs, affordability and ease of access. In addition, the nutrient content of a food varies with season, location of production, growing conditions, storage, processing, and cooking techniques, and many of these factors are unaccounted for in food composition tables. However, the content differs from nutrient to nutrient and individual to individual (Satija, 2015). Hence, measuring diet in free-living populations is challenging because individual diets are complex exposures with innumerable and sometimes, poorly characterized components that are consumed in varying amounts and combinations (Satija, 2015).

The overarching reality is that nutrition, public health, and supply chain experts have addressed the malnutrition challenges from their independent perspectives. For example, under the supply chain perspective, there is an impetus to target specific nutrient deficiencies by using optimization (Gazan et al., 2018), whereas this approach comes with challenges of making the proposed food plate palatable. Similarly, nutrition experts address the challenge of more practical ‘food groups’ instead of ‘nutrient’ assortment (Yan et al., 2020). However, scaling this approach so that it is accessible and affordable by the world population needs an operations/supply chain perspective.

A large body of work (EAT Lancet Commission on Food, 2019) has emerged on the environmental impacts of various diets (the so-called sustainable food), with most studies concluding that a diet rich in plant-based foods and with fewer animal source foods confers both improved health and environmental benefits. Overall, the body of literature indicates that such healthy diets[2] are “win-win” in that they are good for both people and the planet. However, there is still no global consensus on whether planetary healthy diets may be achieved for a global population of 10 billion people by 2050. By assessing the existing scientific evidence, the EAT Commission developed universal scientific targets into a common framework, so that planetary health diets (both healthy and environmentally sustainable) and the operating ground for food systems could be identified. This safe operating space is defined by scientific targets for intakes of specific ‘food groups’ (e.g., 100 to 300 g/day of fruits) to optimize human health and food production to ensure a stable Earth system (EAT Lancet Commission on Food, 2019).

These food security studies provide a framework/ guideline for the design of a healthy diet and connect nutrition to the food production, supply and transportation systems worldwide. We contribute to this stream of work by modelling baskets of food groups specific to Indian geographies based on the cultural preferences and local production patterns.

2.2 Local Production and Consumption Patterns in India

A large portion of Indian diets is made up of cereals and leguminous plants, which provide most of the required calories for the population. As per the NSSO 2011-2012 survey, cereals, starches, and roots comprised between 45% to 87% of calorie share of Indians across all geographical regions and income groups in India. The NSSO 2011-2012 survey presents self-reported data by sample households on monthly consumption of cereals (i.e., rice, wheat, jowar, bajra, ragi, barley, small millets, and their products) – both from open market and the PDS. The estimate of consumption that is reported in the survey by households is based on the survey respondents’ own estimate based on monthly purchase, stock, etc. We noted that rice and wheat comprise 91% of overall producer’s self-consumption (within the family) of cereals, and the PDS channel to acquire cereals comprised

[2] Planetary health refers to the “the health of human civilization and the state of the natural systems on which it depends”. This concept was put forth in 2015 by the Rockefeller Foundation-*Lancet* Commission on planetary health to transform the field of public health, which has traditionally focused on the health of human populations without considering natural systems. The EAT-*Lancet* Commission builds upon the concept of planetary health and puts forth the new term “planetary health diet” to highlight the critical role that diets play in linking human health and environmental sustainability and the need to integrate these often-separate agendas into a common global agenda for food system transformation to achieve the SDGs and Paris Agreement - (EAT, 2019).

only rice and wheat as of 2011. We infer that affordable access to rice and wheat via the PDS is a major contributor to the widespread consumption of these grains in India.

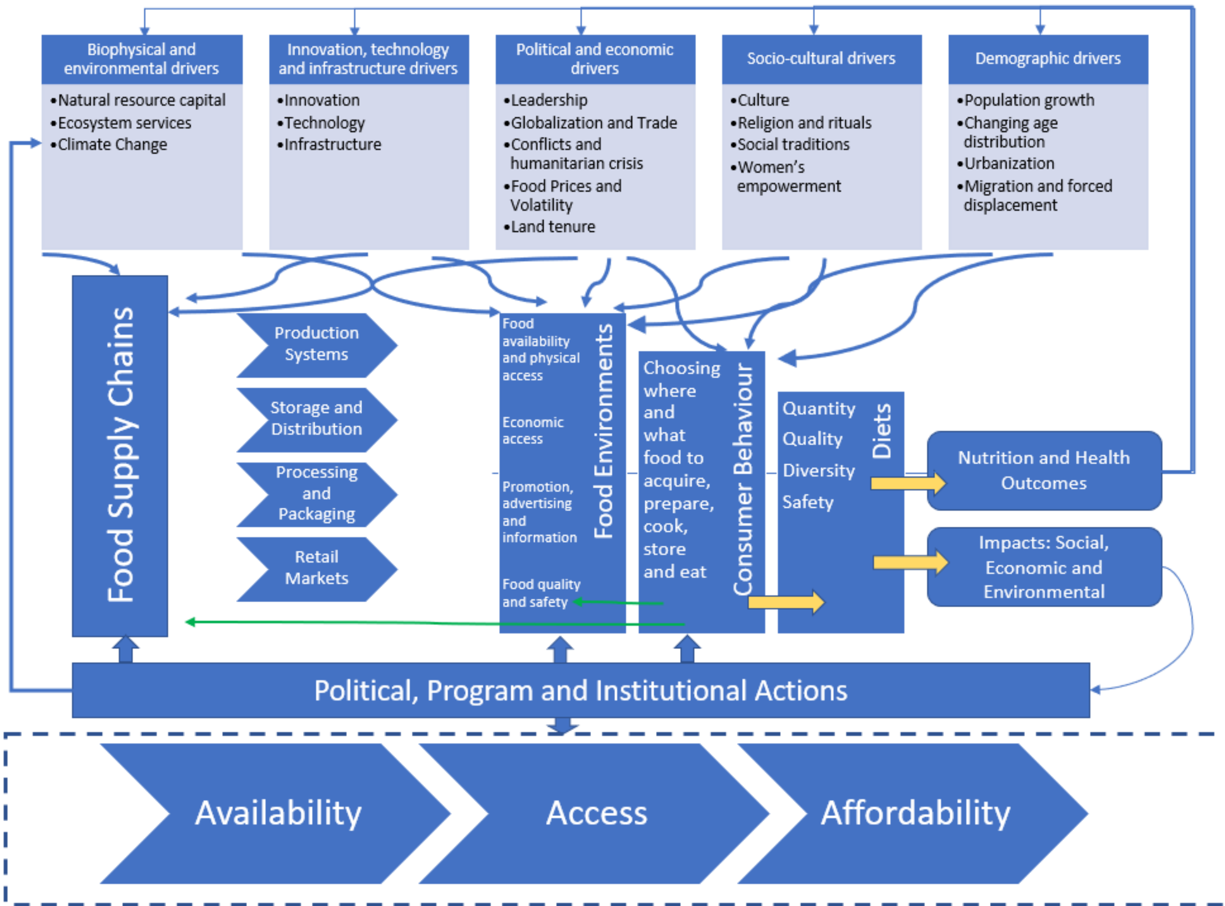
On the production side, the APY 2011-2012 survey indicates that rice and wheat comprised 82% of cereal production in the entire country as of 2011. However, each state also grows a set of other coarse cereals and pulses like jowar, bajra, ragi, barley, small millets that are suited to its agro-ecological conditions. These products are not distributed by the PDS channel despite their regional and cultural importance for each Indian state and region, where they are locally grown/sourced.

Furthermore, urban households in India consistently have higher dietary intake deficiencies than their rural counterparts at all income levels, except for Vitamin A (i.e., a micronutrient that helps maintaining a healthy muscular-skeletal structure, mucus membranes, skin and good eyesight). Fewer urban households in the highest income group are deficient compared to their rural counterparts (Rao, et al., 2018). In almost the entire North and West, rural diets show a higher diversity in cereal consumption. In the South, and particularly in the East, urban cereal consumption is slightly more diverse, because the rural poor areas rely more on cheap rice (Rao, et al., 2018).

2.3 Bringing a Supply Chain Perspective to Nutrition Problems

As can be seen in the studies of nutrition presented above, affordability and availability of food are important drivers of consumers’ diet preferences. The High-Level Panel of Experts on Food Security and Nutrition of the UN Committee on World Food Security - HLPE (2017) identified three core elements in food systems: food supply chains, food environments and consumer behavior (see Figure 1) (Yong-jun, 2019).

Figure 1 – Conceptual Framework of Food Systems for Diets and Nutrition *Source:* (Yong-jun, 2019)



Of these three core elements, the HLPE (HLPE, 2017) illustrated the central role of food environments in shaping consumer behavior and food choices and, finally, in determining diets and nutrition. Therefore, acting on food environments in the proper way can bring transformative changes across the whole food system for delivering healthy and sustainable diets. More specifically, the HLPE, 2017 identified three key elements of food environments on which to act to improve the diet and nutrition outcomes of our current food systems: (i) physical and economic access to food (i.e., proximity and affordability); (ii) promotion, advertising and information; (iii) food quality and safety. The HLPE, (2017) explored how to improve physical and economic access to healthy diets and ways to strengthen consumers' information and education on healthy diets. These two pathways are a direct reference to addressing the nutrition problems using a supply chain lens towards long-term sustainable food ecosystems.

Pollard & Rowley, (2010) conducted a systematic literature review of supply chain role in the fruits and vegetables availability, price and the impact of intervention schemes to increase nutrition awareness. Availability is a challenge in underserved communities and thus, prices for nutritious food are higher in small grocery stores than in large retailers (Deller, et al., 2015). A proper supply chain with the right partnership among stakeholders, can increase availability and lower prices for nutritious food (Sadler, et al., 2013). Another key leverage point that the supply chain can handle is advertising and promotion, there is little promotion of nutritious food and there is a higher promotion of "less nutritious" or "junk" foods that are sponsored by the big CPG manufacturers and distributors because these products are more profitable. The supply chain should aim to increase consumer awareness working together with the health sector to achieve a better informed and health-aware consumer Pollard & Rowley, (2010). Both lower-price and health information are expected to have a positive effect on the increase in healthy food consumption.

In the food supply chain stream in the Indian context, studies have explored the impact of cereal diversification on production sustainability and distribution costs. Devalkar et al., (2020) highlighted the potential of introducing locally available coarse cereals to reduce the government's cost of operating PDS without addressing the food variety required to gain more nutritional content. Rao et al., (2018) studied the impact of coarse cereals on improving the nutritional content within some cost-guardrails but without considering purpose of building local-based food ecosystems. Davis et al., (2019) detailed agro-ecological sustainability, which can be reclaimed by supplementing rice and wheat production with coarse cereals in many parts of India without deep diving into how the logistics and distribution models connect smallholder farmers to customers.

Most studies around interventions for awareness for food groups propose field experiments as a methodology to understand purchasing behaviors. Carroll and Samek, (2018) contributed with a methodology on how to conduct field experiments in low-income populations. They suggested forming partnerships with grocery stores and tailoring the study to community's characteristics with easy-to-follow instructions. They suggested a randomized block design for the experiment to reduce the variability within treatment conditions and better estimate treatment effects. In Latin America, studied the effects of price variation, food availability and building nutritional awareness on customers' decision-making by using an experimental design with stratified random sampling to understand causal inferences in Ecuador and Brazil using a small data set. Despite identifying that respondents were price sensitive and that nutritional awareness helps reinforcing the message, authors did not build any logistics strategy or food access model.

In our work, we build on the HLPE (2017)'s framework to match local production to consumption by designing a customized food basket for different geographies based on local cultural preferences and cereal availability in the Indian districts. In contrast to the existing studies, we intend to propose improved dietary diversity to boost food quality and security for the entire population. We then study ways to link farmers from all districts to households through the PDS, thus affecting the food environment to minimize transportation costs, reduce food waste and increase the availability of diverse cereals for the most vulnerable population segments.

2.4 Solution Approaches for the basket configuration and Distribution strategies

As highlighted earlier, our problem will solve two optimization problems. One formulation will address the best configuration of 'baskets' based on locally grown, available or sourced food groups for each geography. The second formulation will solve the distribution problem of delivering these customized assortments to the numerous households in the geography, which will be a standard Mixed Integer Linear Program (MILP).

In regard to the diet assortment problem, (Gazan R., Brouzes, Florent, Matthieu, & Anne, 2018) reviewed a set of studies that use mathematical diet optimization for understanding the relations between different dimensions of diet sustainability and how optimization can be a powerful method to define sustainable diets. Diet optimization aims to find the optimal selection and combination of foods for a population, or for an individual, which fulfils a set of constraints while minimizing or maximizing an objective function.

We reviewed a set of literature around the knapsack problem and bin packing problem (BPP) to model the customized assortment for each geography. Knapsack problem is a typical application of integer programming (IP). In knapsack problems, there is a container (the 'knapsack') with a fixed capacity (an integer) and a number of items. Each item has an associated weight (an integer) and an associated value or benefit (another integer). The problem consists of filling the knapsack without exceeding its capacity, while maximizing the overall value or benefit of its contents (IBM, 2020).

In the case of the bin packing problem, items of different volumes must be packed into a finite number of bins or containers each of a fixed given volume in a way that minimizes the number of bins used. In computational complexity theory, it is a combinatorial NP-hard problem and deciding if items will fit into a specified number of bins is NP-complete (Wikipedia, 2020).

As with all NP-hard problems, exact algorithms will suffice for only small instances of the problem. Large computationally intensive instances of the BPP can be solved by a combination of approximation techniques with an exact method, or by means of metaheuristics. Metaheuristics are more efficient in searching the solution space than the exact approaches by using different strategies even for large-scale instances; however, they do not guarantee finding optimal but high-quality solutions.

Santos, et al., (2018) reported literature on solution algorithms to solving the bin packing problem. Heuristics commonly applied to solve combinatorial optimization problems have been tested on the BPP. Authors introduced a new variant of the bin packing problem for ‘compatible categories’ (BPCC). Motivated by last mile deliveries to mom-and-pop stores (i.e., nanostores), this study analyses ‘groups’ of items for the same category. Authors show that VNS algorithms can solve the BPCC in very short CPU times. Our problem is closest to this literature.

Our problem set of choices is constrained by the type of foods available and we determine their quantities to boost nutrition via the accessibility to a larger variety of cereals and pulses than just rice and wheat through the PDS.

2.5 Gaps and contributions

There is a growing need for nutrition and supply chain experts to come together to combat different forms of malnutrition, through integrated research initiatives, policies, and programs. Efforts in the last decades have started to show some results in catering to ‘hunger’. Solving the food malnutrition problem will need a similar concerted effort at a massive scale. Our work is an initiative in this direction.

In summary, we propose an optimization model that will take as input the local crop production, and the cultural taste preferences of the population in a geographical cluster. The output of the model will be a basket of assorted ‘food groups’ that can be supplied to households to enable dietary diversity and quality. Through our analysis of literature, we found that the Public Distribution System (PDS) is the single largest food security program run by the Government of India, across all Indian districts. We learned that rice and wheat are the staple cereals distributed at scale through this program. Our proposal will diversify the consumption basket with other food groups that are locally available and are nutritious. We also present strategies to scale the distribution of the proposed food basket via the PDS and alternate channels like neighborhood retail stores and mid-day meal schemes, and thus enable their affordable access by all Indian households. This will also help reducing costs for the Government of India, boost cereal accessibility for everyone in the Indian districts while indirectly minimizing food waste.

3. METHODOLOGY AND DATA

3.1 Methodological Framework

We propose a model to match the existing production supply of cereals and pulses in India to address the minimum dietary requirements of the poorest of the poor households in the country. We obtain the minimum dietary requirements from the Indian Council of Medical Research (ICMR) food plate recommendations. The recommendations are in the form of a minimum daily consumption target for seven different food groups (ICMR, 2020) - 'fruits', 'vegetables', 'cereals and nutri-cereals', 'pulses, eggs and flesh food', 'nuts and seeds', 'fats and oils' and 'milk/curd'.

The scope of food groups proposed in our 'food basket' model is limited to 'cereals' and 'pulses'. This is because of two reasons. First, the aim is enhancing protein content in the diet and pulses are good protein source. Second, the long shelf life of these food groups in the Indian climate without much cold storage or nuanced handling requirements. The scope of the distribution strategy in our model is limited to the government channel. The model, however, is generalizable, and provides a framework for matching supply to nutritional requirements through any public or private channels. On the non-government channels front, we speak to the Akshaya Patra foundation for their mid-day meal scheme for school children and review literature on distribution of food grains through small, family-owned retailers or mom and pop stores (i.e., nanostores) to enhance availability of the grains for all population segments, especially in underserved communities. We realize a strong opportunity and interest from these non-government channel partners to serve as distribution centers for such 'food baskets' as well. However, due to the completeness of the publicly available household database served by the government Public Distribution System (PDS) channel, we scope our target audience and distribution channel to the poorest of poor households (Antyodaya Anna Yojana Households) currently served by the government Public Distribution System in India for this first study.

Figure 2 – Methodological Framework

Based on the aforementioned scope, we take the following approach as shown in Figure 2. First, we segment our target households (aggregated into districts) based on their historical consumption patterns considering their taste preferences into some ‘clusters’. Our target households are currently entitled by the Government of India to obtain a fixed monthly quantity of rice and wheat at subsidized prices from the fair price shops, which depend on the government, in their districts. We propose to redesign the composition of this ‘basket of items’ that the target households receive to incorporate other millets. So, keeping the total monthly entitlements (i.e.., total quantities) that each household receives fixed, we propose a new customized assortment (food basket) of cereals and pulses for each cluster.

These food grain baskets are constituted from locally grown pulses and cereals, have a history of consumption in these districts, and satisfy a large part of daily prescribed intake of cereals and pulses by the ICMR to enhance the protein content of the basket and constitute more ‘balanced diets.’ Finally, we design a distribution strategy to deliver the food baskets to the target households via the Public Distribution System operated by the Food Corporation of India (FCI) in each district cluster. In the following subsections, we elucidate our methodology.

3.1.1 Customer Segmentation by PCA and K-means Clustering

We characterize the cereals and pulses consumption patterns of households based on the National Sample Survey Dataset of Consumption Expenditure 68th Round described in section 3.2. For each district, over 31 characteristics of people living in the district including median household income, geographical location of the state, historical consumption of different cereals and pulses were available. We arrived at key features critical to characterizing taste preferences of cereals and pulses in each district, by performing a dimensionality reduction of this dataset using Principal Component Analysis (PCA). The details of this analysis are provided in Appendix A.1.

Based on these features, we perform k-means clustering to segment districts into certain clusters with similar taste preferences. Table 1 shows a sample cluster output template. We discuss results from the application of clustering in Section 4.

Table 1: Sample Clusters from k-means Clustering

Cluster Number	Vector of Districts *[Serial Number of Districts]*
1. Cluster 1	[1,3,7,9,10,11,...]
2. Cluster 2	-
3. Cluster 3	-

3.1.2. Food Basket Configuration

Once we obtain homogenous clusters of districts (i.e., small variability among the elements of the cluster), we design a customized food grain basket for each cluster. Each target household is entitled to 35 kg per household per month from the government (Entitlements under NFSA, 2020). We set this entitlement as the size of the monthly food baskets to be delivered to each target household. ICMR recommends 270g cereal intake per person per day, and 90 g pulses per person per day (ICMR, 2020). For each cluster, we design a basket that comprises cereals and pulses in proportion as prescribed by the ICMR as shown in Table 2.

Table 2: Food Basket Size based on daily targets proposed by the Indian Council of Medical Research for a balanced diet, 2020

	Food Type	Daily Target (ICMR)	Monthly Target	Share of Total
1.	Cereals	*0.27 kg*	*8.1 kg*	*75%*
2.	Pulses	*0.09 kg*	*2.7 kg*	*25%*
	Total	*0.36 kg*	*10.8 kg*	
Household Type	Basket Size	Cereals in Basket	Pulses in Basket	Beneficiary
AAY	**35 kg**	35 kg X 75% = 26kg	35 kg X 25% = 9 kg	Household

Table 3: Demand in each cluster

Cluster *Number*	Demand of *AAY baskets* *in baskets*	Demand of Cereals *kg per month*	Demand of Pulses *kg per month*
1. Cluster 1	600	=600 X 26	=600 X 9
2. Cluster 2	-	-	-
3. Cluster 3	-	-	-
4. Cluster 4	-	-	-

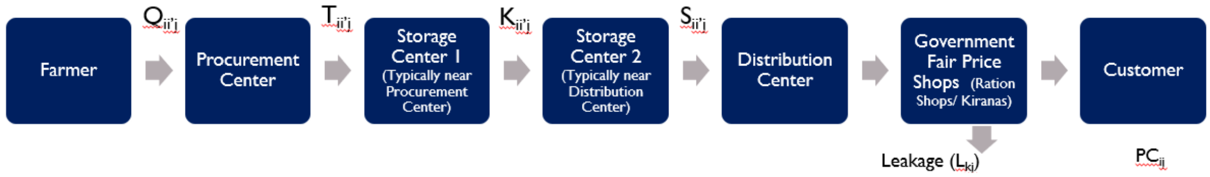
We work out the quantities of pulses and cereals per basket to be 9 kg and 26 kg, respectively as shown in Table 2. We combine the quantity of pulses and cereals per basket (or per household) in each cluster (see Table 3). We use the number of target household types in the district to obtain the total demand of cereals and pulses in each district.

We now model the food basket design as a simplified version of the knapsack/ bin packing problem. We have to select cereals and pulses for each cluster from multiple varieties grown in the country so as to maximize match to historical taste preference for the cluster, constrained by the weight/quantity that be filled into the basket (which is like a knapsack). The details of basket formulation are in Appendix A.2. The selection of food grains paves the way to determine our procurement strategy - the states/districts from which we should procure the grains.

3.1.3. Distribution of Baskets through Government Channel

The final step of our model is the procurement and distribution of grains in the food baskets to reach all households in each cluster district. As discussed earlier, we propose the government PDS channel operated by Food Corporation of India (FCI) to be the medium of distribution. We model a similar network as the FCI, as shown in Figure 3 (Devalkar, et al., 2020).

Figure 3: Schematic Map of the Public Distribution System, India



As shown, farmers sell their produce and grains to the government at procurement centers, the FCI stores the procured food grains in a storage center close to the procurement center. If the distribution center (DC) is close enough, then food grains are directly transported to the DC; otherwise, parked at a storage center near the DC before arriving at the DC. The food grains finally move from the DC to fair price shops from where the households come and purchase their monthly entitlements at a subsidized price determined by the government.

We, however, do not restrict the model to operate within the existing storage center footprint of the FCI. This is because of two reasons. First, relaxing the model over a large search space to select storage centers improves the alternatives of minimizing the cost of distribution. Second, cereals and pulses have minimum nuanced storage and handling requirements unlike crops that require cold storage (e.g., fruits and vegetables), and hence, if the model proposes new storage locations, it should be relatively feasible to rent/lease centers at alternative locations within a reasonable time frame.

Thus, storage centers to be opened in each state become decision variables/ constraints, while all the remaining decision variables and constraints are continuous. Overall, this turns out to be a Mixed Integer Linear Program with over 30,000 constraints. The key decision variables are quantities procured, stored, and transported between locations, and count of storage centers to operate in each location.

Our objective is to minimize the cost of assortment and distribution, subject to supply (i.e., production) and demand constraints (determined by ICMR, 2020 nutrition targets). We also ensure a high similarity of our proposed food baskets to the local taste preferences of households in the district to increase the food grain basket acceptability. We detail the model formulation in sub-section 3.3.

3.2 Data

Our models and analyses are based on the year of the primary input, production supply, that is 2019-2020. We use the following data sources for the model:

- **Demand:** We obtain the monthly entitlement of food grains for our target households from the Department of Food and Public Distribution website (Entitlements under NFSA, 2020). This entitlement is 35 kg of food grains per household for the poorest of poor families registered as Antyodaya Anna Yojana (AAY) ration cardholders with the government. We obtain the number of target households in each district from the NSSO Consumption Expenditure Data, 68th Round, 2011-2012. We combine the entitlement per household with the number of households to obtain the monthly demand in each district. The sample survey reports an estimate of 10.9 million AAY households in India spread in the 35 Indian states. Each state is further divided into districts, 625 districts all over India. Each district has many households. Some of these households are sampled in each district, and the sample size is combined with some multipliers to obtain an estimate of the total number of households per district.
- **Production Supply:** We use the area, production, and yield[3] dataset, available on the 'Crop Production Statistics Information System' webpage by the Directorate of Economics and Statistics (Crop Production Statistics, 2019-20). The report provides production estimates of crops in each of the APY survey districts for the year 2019-2020. For our analysis, we aggregate the reported seasonal production to obtain yearly production estimates for each APY survey district. We then obtain the average monthly production supply of each of the 24 grains from this aggregate statistic.

  From the production output, a part is consumed in the farmer's own household. From the remaining, a part is sold in the open market and the remainder is sold to the government. We assume that in all districts that produce at least 100 metric tonnes of a food grain, 60% is available for sale to/procurement by the government channel. This assumption is based on the ratios of self-consumption, government procurement and open market sales of rice and wheat observed (primary food grains in end-to-end circulation, that is, from farmer to end consumer through the PDS channel) historically.
- **Taste Preferences:** For our selected scope of cereals and pulses, we use the (NSSO Consumption Expenditure Data,68th Round, 2011-12) to obtain estimates of consumption for various cereals and pulses

---

[3] Production is reported in units of metric tonnes of crop output per season. Yield is reported in metric tonnes of crop output per unit area of cropland harvested.

by our target population in each district. Consumption estimates are combined with households count estimates in each district to obtain per household per month consumption estimate of cereals and pulses. We use this derived metric as an index to understand the district's relative taste preferences for different food grains. We use PCA and k-means clustering on this District—Geographic Region—Per Household Consumption Estimate (for each of 24 food grains) dataset to obtain district clusters with similar taste preferences.

- **Distance Between Districts:** We work with the Google API to collect the distance between district centroids. We have 625 districts and obtain a 625 X 625 distance matrix. Additional work will be needed to make these computations more granular and accurate between distribution centroids. For our analysis, we refrain from distances between facility locations for the lack of enough data.
- **Procurement Cost:** We get the Minimum Support Prices (MSPs) of food grains in 2019-2020 from the Department of Agriculture, Cooperation and Farmers Welfare, Government of India website (MSP of Foodgrains in India, 2021). We include the variable cost of operating a procurement center (7.5% of MSP) based on a report by the High-Level Committee (Kumar, 2015) that gave recommendations on reorienting the role and restructuring the FCI. The variable unit cost includes a trader commission equivalent to 2.5% MSP and the cost of packaging equivalent to 5% MSP.
- **Transportation Costs:** We use an estimate of rail transportation cost per unit distance from (Kumar, 2015). Based on an FCI operational spends report, we assume the unit cost (per tonne-km) of road transportation to double the rail transportation cost.
- **Storage Costs and Capacity:** We work with the average fixed cost of storage by adding the cost of power, fuel, electricity, rent of centers, and repair and maintenance of centers and dividing it by the number of storage centers. Each of these parameters is gleaned from the from FCI Annual Report (FCI, 2019-20). For handling cost at storage centers, we obtain an estimate from (Kumar, 2015). For the capacity that can move through storage centers, we use the total capacity and combine it with total count of storage centers. Each of these parameters is obtained from FCI Annual Report (FCI, 2019-20). In the

next section, we explain the results and managerial insights of our study.

3.3 MILP formulation

With regard to the objective function, we observe that the cost incurred by the government comprises three key components: Procurement Cost, Storage Cost and Transportation Cost. These costs can be defined as the following:

- Procurement Costs:
    - Cost of Purchase - Minimum Support Price X Quantity Procured
    - Cost of Travel – Per Unit Road Transportation X Distance Travelled
- Storage Costs:
    - Fixed Cost of Storage – Total cost of rent, electricity, employee salaries
    - Variable Cost of Storage – Handling charges at storage centers
- Transportation Costs:
    - Cost of movement of grains from Procurement to Storage to Distribution Center

Our model formulation is detailed below:

Sets

District i, i' $\in$ D = {625 districts of India}. Also, each district belongs to some state, k.

State k, k $\in$ S = {35 States and Union Territories of India}

Food grain j, j $\in$ F = {rice, wheat, jowar, bajra, ragi, maize, small millets, barley, arhar, gram, moong, masur, urd, peas, khesari, besan}

Cluster c, c $\in$ C = {Cluster 1, Cluster 2, Cluster 3, Cluster 4, Cluster 5, Cluster 6}

Parameters

All parameter values are for the year 2019-2020, unless mentioned otherwise.

MSP_j = Minimum Support Price for food grain j (Rs. per tonne)

P_{ij} = Average monthly quantity of food grain j produced in district i (tonnes)

D_{ij} , D_{cj} = Monthly demand of food group j in district i or cluster c (tonnes)

$tt_{ii'}$ = Physical driving distance between districts i and i' (km)

tc_f = Per unit km, per unit quantity cost of travelling from farmer location to procurement center

tc_road = Per unit km, per unit quantity cost of road movement

tc_rail = Per unit km, per unit quantity cost of rail movement

L_{kj} = Percentage of grain j in state k lost due to leakages in the PDS

flow cap = Average flow capacity of an FCI storage unit (tonnes/ storage center/ month)

f_p = Fixed cost of storage like rent, electricity, staff expenses (Rs./ storage center/ month)

f_v =Variable cost of storage such as loading and unloading (Rs. per tonne) (Same across movement legs because this is mainly loading/ unloading cost)

M = Large constant value computed as the sum of all capacities and demands

Decision Variables

$Q_{ii'j}$ = Monthly quantity of grain j available in district i' that is procured in district i

$T_{ii'j}$ = Monthly quantity of grain j procured in district i and stored in district i' in the first stage of storage

$K_{i'ij}$ = Monthly quantity of grain j transferred from first stage of storage in district i' to second stage of storage in district i

$S_{i'ij}$ = Monthly quantity of grain j from the second stage of storage in district i' distributed in district i

$PC_{i'j}$ = Monthly quantity of grain j purchased from the PDS by households in district i'

PC_{cj} = Monthly quantity of grain j purchased from the PDS by households in cluster c

Y_i = Number of first stage storage centers operating in district i

Y_{pi} = Number of second stage storage centers operating in district i

z_i = Sum of Y_i , Y_{pi}, i.e., total number of storages centers operating in district i

y_k = Binary variable to ensure a first stage storage unit is operational in state k if food grains are procured in any district i

Mathematical formulation

$$
\begin{aligned}
\text{P:} \quad min\ Cost = & \sum_{j}(MSP_j \sum_{i'} \sum_{i} Q_{i'ij}) + \\
& \sum_{i'}(fp * z_{i'} + \sum_{i} \sum_{j} fv * T_{ii'j}) + \sum_{i} \sum_{i'} \sum_{j} fv * K_{i'ij} + \\
& \sum_{i} \sum_{i'} \sum_{j} Q_{i'ij} * tc\_f * tt_{i'i} + \\
& \sum_{i} \sum_{i'} \sum_{j} T_{ii'j} * tc\_road * tt_{ii'} + \\
& \sum_{i} \sum_{i'} \sum_{j} K_{i'ij} * tc\_rail * tt_{i'i} + \\
& \sum_{i} \sum_{i'} \sum_{j} S_{ii'j} * tc\_road * tt_{ii'}
\end{aligned}
$$

s.t.

$$
\begin{aligned}
\sum_{i} Q_{i'ij} &\leqslant 60\% X P_{i'j}\ \forall\ i', j\_where P_{i'j} \geqslant 100 tonnes && (1) \\
\sum_{i'} Q_{i'ij} &= \sum_{i'} T_{ii'j}\ \forall\ i, j && (2) \\
\sum_{i} T_{ii'j} &= \sum_{i} K_{i'ij}\ \forall\ i', j && (3) \\
\sum_{i'} K_{i'ij} &= \sum_{i'} S_{ii'j}\ \forall\ i, j && (4) \\
\sum_{i \in k} \sum_{i'} \sum_{j} Q_{i'ij} &\leqslant M * (1 - y_k)\ \forall\ k && (5) \\
\sum_{i \in k} Y_i + M * y_k &\geqslant 1\ \forall\ k && (6) \\
\sum_{i} \sum_{j} T_{ii'j} &\leqslant flow\_cap * Y_{i'}\ \forall\ i' && (7) \\
\sum_{i'} \sum_{j} K_{i'ij} &\leqslant flow\_cap * Yp_i\ \forall\ i && (8) \\
z_i &\geqslant Y_i\ \forall\ i && (9) \\
z_i &\geqslant Yp_i\ \forall\ i && (10) \\
PC_{ij} &\geqslant D_{ij}\ \forall\ i'\_in\_c, j && (11) \\
\sum_{i' \in k} PC_{i'j} &\leqslant (1 - L_{kj}) * \sum_{i' \in k} \sum_{i} S_{ii'j}\ \forall\ k, j && (12) \\
Q_{i'ij}, T_{ii'j}, Tp_{i'ij}, S_{ii'j}, PC_{i'j} &\geqslant 0 \quad \forall\ i, i', j && (13) \\
Y_{i'}, Yp_i \in \mathbb{Z}^+ \quad \forall\ i\ \&\ i' & && (14) \\
y_k \in \{0, 1\} \forall\ k & && (15)
\end{aligned}
$$

The objective function aims to minimize the total cost of procurement, storage and transportation, formulated subject to the following constraints:

(1) is the supply constraint, it enforces the procurement of grains to be less than the available supply of grains for government purchase in each district;

(2) , (3) and (4) are flow balance constraints, they ensure total inflow of grains in a district is equal to the total outflow of grains from the district across the logistics facilities up to selling points;

(5) and (6) are linking constraints that ensure the facility is open if transportation occurs and vice-versa;

(7) and (8) are capacity constraints, they ensure that transportation of food grains into and from storage centers is within the capacity limits of the storage center

(9) and (10) enforce that the number of storage centers suggested in each stage are less than equal to the sum of storage centers opened;

(11) is the demand constraint, ensures purchase in each cluster to be greater or equal to the demand;

(12) ensures purchase is greater than equal to leakage;

(13), (14) and (15) refer to the domain constraints for the variables used.

We use the CPLEX ILOG studio for optimization with an Azure Virtual Machine with 64 GB RAM and 16GHz processor.

4. RESULTS

In this section, we first characterize the consumption estimates and taste preferences of pulses and cereals in India and note the gap from ICMR recommended nutritional targets. We then present the results of our proposed model. The results are organized as follows: i) Customer Segmentation using Principal Component Analysis (PCA) and K-means clustering, ii) Model baskets for each cluster, and iii) the distribution plan to deliver the baskets to all districts in the clusters, including sensitivity analyses.

4.1 Characterization of Historical Consumption of cereals and pulses

Table 4 shows the gap between consumption estimates and ICMR, 2020 targets for cereals and pulses in India. Clearly, India consumes less than recommended pulses. This reinforces the fact Indian diets are traditionally low in protein and rich in carbohydrates. We intend to modify the food basket of our target households to include more pulses in their daily diets.

Table 4: Average Deficiency at Country Level

(across all household types, *per household of a 4-member family per month*)

Quantity	2011-2012 Consumption Levels	ICMR Target	Gap
Cereals Total	51.6 kg	32.4 kg	19.2 kg excess
Pulses Total	3.7 kg	10.8 kg	7.1 kg short

We compute the correlation between taste preferences and geographical regions in Table 5. The figures represent the Spearman correlation coefficient between average household consumption of a food grain per month and a dummy binary variable representing each geographical region (for example, for North, the dummy variable is 1 for 'North' and 0 for all other regions). This analysis allows us to test the dependence of cereals and pulses in each Indian region. Given that rice and wheat are offered via PDS and also via open market (om), we split their analysis in the table. We note that among cereals, ragi has a strong positive correlation in the South. Bajra and Jowar have moderately strong positive correlation in the West. Rice and masoor (i.e., a pulses variety) have moderately strong positive correlation in the North Eastern region.

Table 5: Spearman Correlation between food grains and different Indian regions[4]
(Values in bold are different from 0 with significance level alpha=0.05)

Variables	North	South	East	West	Central	N. East
rice-PDS	**-0.252**	0.05	**0.275**	**-0.423**	**-0.13**	**0.401**
rice-om	**-0.293**	0.01	**0.258**	**-0.439**	**-0.128**	**0.523**
wheat-PDS	**0.374**	**-0.494**	0.07	**0.253**	**0.232**	**-0.412**
wheat-om	**0.249**	**-0.522**	**0.229**	**0.23**	**0.21**	**-0.418**
Jowar	**-0.184**	**0.232**	**-0.234**	**0.452**	0.006	**-0.234**
Bajra	-0.015	-0.046	**-0.157**	**0.57**	-0.075	**-0.257**
Maize	**0.135**	**-0.288**	0.046	0.071	0.02	0.021
Barley	0.025	0.021	-0.021	**0.147**	**-0.079**	**-0.103**
small millets	-0.072	-0.025	**-0.18**	**0.177**	0.03	**0.125**
Ragi	**-0.089**	**0.671**	**-0.266**	0.055	**-0.153**	**-0.207**
cereals total	**-0.1**	**-0.479**	**0.331**	**-0.179**	0.067	**0.315**
arha, tur	**-0.291**	**0.119**	**0.208**	-0.012	**0.286**	**-0.353**
gram: split	**0.124**	**-0.435**	**0.246**	-0.051	**0.204**	**-0.11**
gram: whole	0.064	**-0.303**	**0.208**	**-0.097**	**0.168**	-0.059
Moong	-0.068	**-0.298**	0.068	**0.086**	0.05	**0.169**
Masur	**-0.166**	**-0.322**	**0.322**	**-0.358**	0.062	**0.415**
Urd	-0.075	**0.103**	**0.093**	**-0.221**	**0.124**	-0.038
Peas	-0.019	**-0.222**	**0.313**	**-0.233**	-0.032	**0.122**
khesari	-0.077	**-0.098**	**0.23**	**-0.149**	0.06	-0.016
other pulses	**0.186**	**-0.216**	-0.059	**-0.083**	0.053	**0.161**
gram products	0.018	**0.122**	0.063	-0.044	**-0.107**	**-0.091**
Besan	0.05	**-0.185**	0.017	**0.231**	**0.233**	**-0.324**
pulses total	**0.212**	0.005	0.01	0.029	**0.09**	**-0.346**

Another interesting insight is that for all regions, rice and wheat have opposite signs of correlation with the region. This indicates that each region prefers only one of these two cereals (i.e., rice or wheat) as their staple.

[4] Major States in each region are listed as follows: **East** – West Bengal, Orissa, Jharkhand, Uttar Pradesh, Bihar; **West** – Rajasthan, Gujrat, Maharashtra; **South** – Karnataka, Tamil Nadu, Andhra Pradesh, Kerala; **Central** – Chhattisgarh, Madhya Pradesh; **North** – Jammu and Kashmir, Himachal Pradesh, Delhi, Haryana, Punjab, Uttarakhand; and, **N. East –** Assam, Manipur, Meghalaya, Mizoram, Nagaland, Tripura, Sikkim

The Eastern region is an exception to this insight. This is important for the design of food baskets, particularly, when looking for increasing accessibility to a larger diversity of grains. Today, all households are entitled to rice at around Indian Rupee (INR) 3 (~ US$0.042) per kg, wheat at INR 2 (~US$0.028) per kg across states, and a total quantity of 35 kg per month. In our basket proposal for each cluster, we can prioritize the preferred staple cereal (i.e., wheat or rice), and propose alternatives to the less preferred cereal in the form of pulses and other nutri-cereals.

We also compute the correlations between food grains to get preferences of grain combinations by households in Table 6. We see that (Rice and Masur) and (Wheat and besan or chickpea flour) are two (cereal, pulses) combinations that have moderately strong positive correlations. (Jowar, Bajra) are a cereal pair with a moderately strong positive relationship.

Table 6: Spearman Correlation among different food grains
(Values in bold are different from 0 with significance level alpha=0.05)

Variables	rice	wheat	jowar	bajra	ragi	arha	gram	moong	masur	urd	peas	besan
Rice	**1**	**-0.631**	**-0.273**	**-0.531**	-0.029	0.027	0.007	**0.125**	**0.52**	**0.119**	**0.168**	**-0.407**
Wheat	**-0.631**	**1**	-0.016	**0.325**	**-0.409**	0.062	**0.345**	**0.081**	-0.075	-0.077	**0.178**	**0.517**
Jowar	**-0.273**	-0.016	**1**	**0.444**	**0.341**	**0.263**	**-0.113**	-0.01	**-0.297**	**-0.107**	**-0.184**	**0.133**
Bajra	**-0.531**	**0.325**	**0.444**	**1**	0.044	0.051	-0.045	0.014	**-0.343**	**-0.186**	**-0.136**	**0.313**
Maize	**-0.117**	**0.252**	-0.006	**0.111**	**-0.241**	**-0.106**	**0.154**	**0.117**	0.068	-0.005	**0.09**	**0.08**
Barley	-0.068	-0.029	0.066	**0.135**	0.019	-0.067	-0.049	-0.028	-0.078	-0.01	**-0.095**	0.011
small millets	-0.002	**-0.084**	**0.179**	0.017	**0.22**	0.041	-0.007	0.026	0.005	0.014	0.019	-0.023
ragi	-0.029	**-0.409**	**0.341**	0.044	**1**	**0.159**	**-0.25**	**-0.236**	**-0.292**	**0.106**	**-0.178**	**-0.145**
arha, tur	0.027	0.062	**0.263**	0.051	**0.159**	**1**	**0.174**	0.039	0.016	**0.256**	**0.121**	**0.248**
gram: whole	0.007	**0.345**	**-0.113**	-0.045	**-0.25**	**0.174**	**1**	**0.267**	**0.284**	**0.147**	**0.344**	**0.324**
moong	**0.125**	**0.081**	-0.01	0.014	**-0.236**	0.039	**0.267**	**1**	**0.347**	0.035	0.069	**0.143**
masur	**0.52**	-0.075	**-0.297**	**-0.343**	**-0.292**	0.016	**0.284**	**0.347**	**1**	**0.164**	**0.31**	**-0.136**
urd	**0.119**	-0.077	**-0.107**	**-0.186**	**0.106**	**0.256**	**0.147**	0.035	**0.164**	**1**	**0.184**	**0.094**
peas	**0.168**	**0.178**	**-0.184**	**-0.136**	**-0.178**	**0.121**	**0.344**	0.069	**0.31**	**0.184**	**1**	**0.114**
besan	**-0.407**	**0.517**	**0.133**	**0.313**	**-0.145**	**0.248**	**0.324**	**0.143**	**-0.136**	**0.094**	**0.114**	**1**

4.2. Customer Segmentation

The correlation tables give important insights about consumer's taste preferences related to a) specific geographical regions and b) combinations of food grains historically opted for. Out of the 31 factors available in the data set, we identify the key variables that are critical for the segmentation of districts based on taste preferences through Principal Component Analysis (PCA). PCA performs a linear combination of some important variables to identify the 2 key orthogonal axes that can explain the data (thus, reducing complexity of handling 31 variables).

The factor circles in Figure 4a show that factors F1 and F2 from the PCA explain 32.5% of the variation in data. Figure 4b shows Factors F1 and F3 from the PCA explain 25.1% of the variation. The red lines in Figure 4 that are closer to the axes represent the variables that have maximum influence on these principal component axes.

Figure 4a: PCA representation for F1 and F2

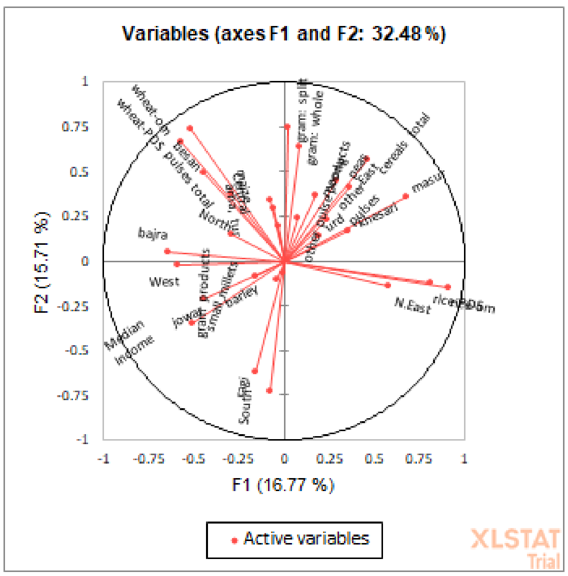
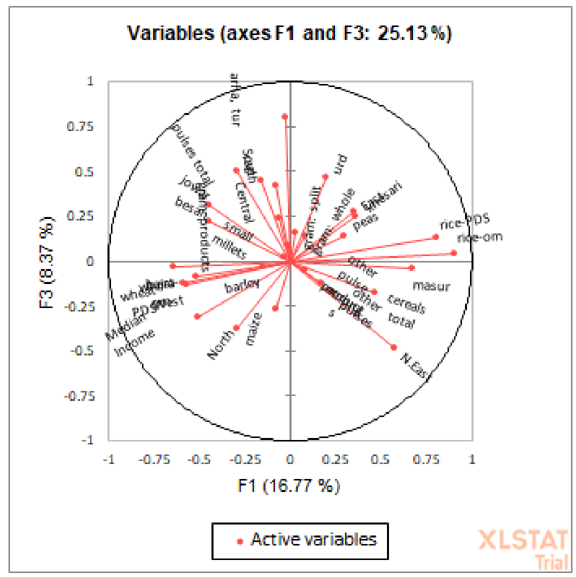
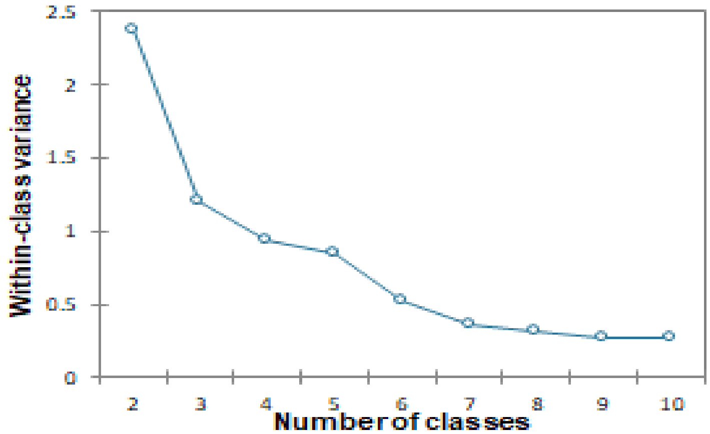


Figure 4b: PCA representation for F1 and F3



From the squared cosine table (in appendix A.1), we confirm the constituent factors that have maximum contributions to these principal components. The regions South, North East, Central and the household consumption estimates of rice-om, wheat-om, bajra, masur, split gram, arhar and urd were the most contributing factors. After identifying the key factors for segmentation from PCA, we perform a K-means clustering. We first plot the Within-Cluster variance against the number of clusters and obtain the ideal number of clusters as six. This value is established based on the elbow point shown in Figure 5. We note that beyond six clusters, the within-class variance does not drop steeply any further. Hence, the value of six is the elbow point, or the optimal number of clusters for our analysis.

Figure 5: Variation of Within Class Variance of Clusters with Number of Classes

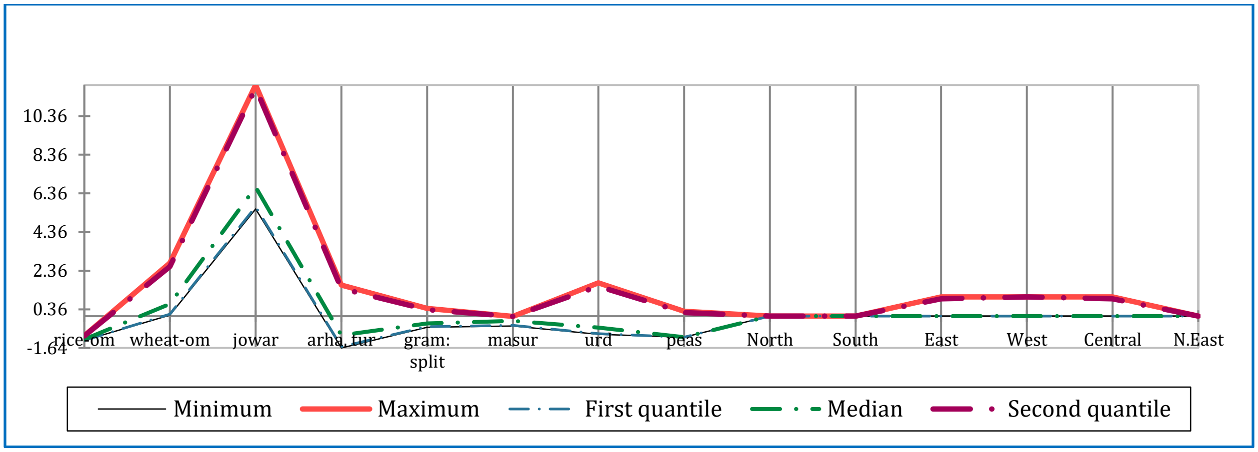


To compare the characteristics of different clusters, we draw parallel plots of the key variables for each cluster. Parallel plots in Figure 6 demonstrate the features of each cluster. The Y-axis in these plots denotes the standardized value of monthly per household consumption estimates. For example, a value of 2 for food grain 'jowar' represents an estimated consumption of twice the average. The 'first quantile' represents 2.5^{th} percentile observation, and 'second quintile' represents 97.5^{th} percentile observation. The median represents the 50^{th} percentile observation.

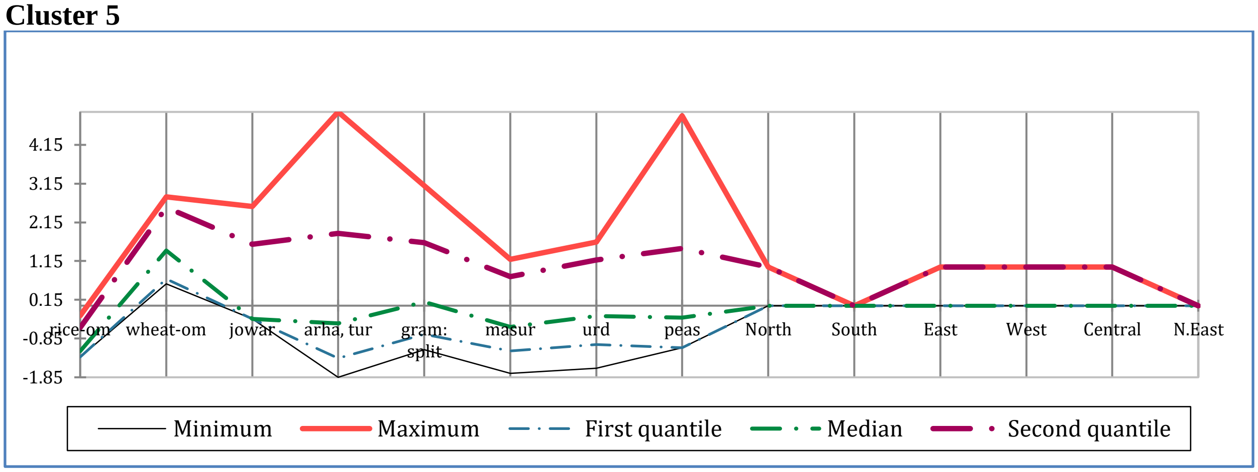
We observe that in Cluster 6, jowar is widely preferred cereal and urd the most opted pulses. Also districts in cluster 6 are spread across East, West and Central India. Similarly, in Cluster 5, wheat is the most preferred cereal; while arha, tur and peas are the most preferred pulses. Districts in that cluster are spread across except in South and North-East.

Figure 6: Parallel Plot of Cluster highlighting key features

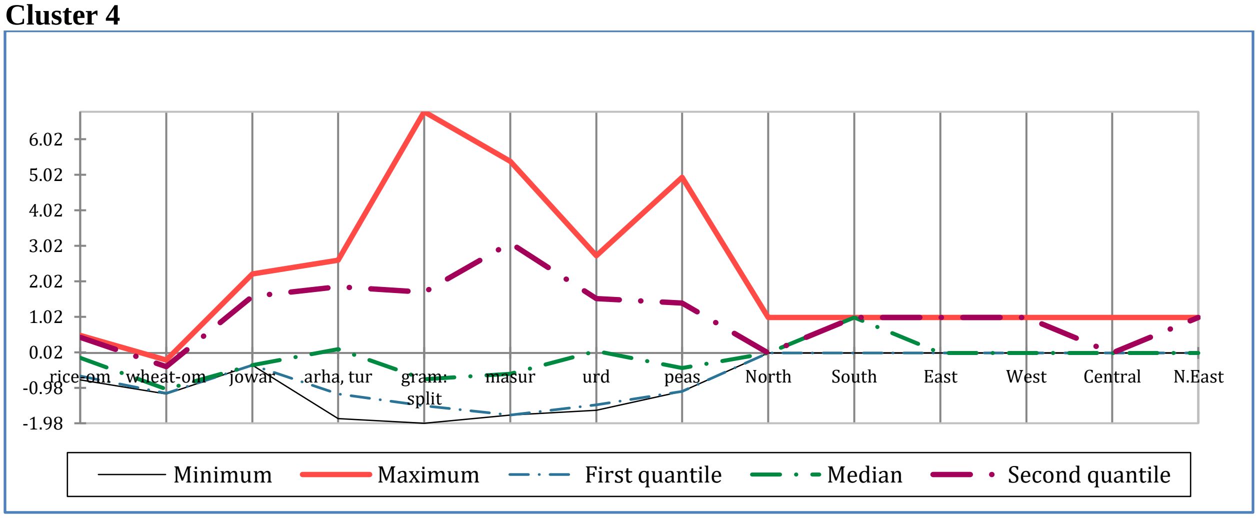
Cluster 6



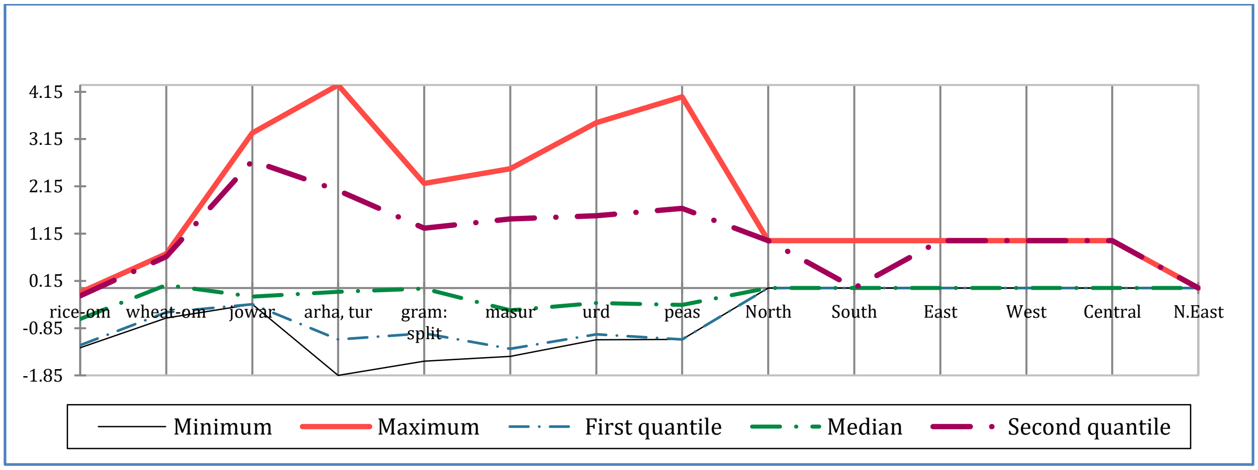
Cluster 5



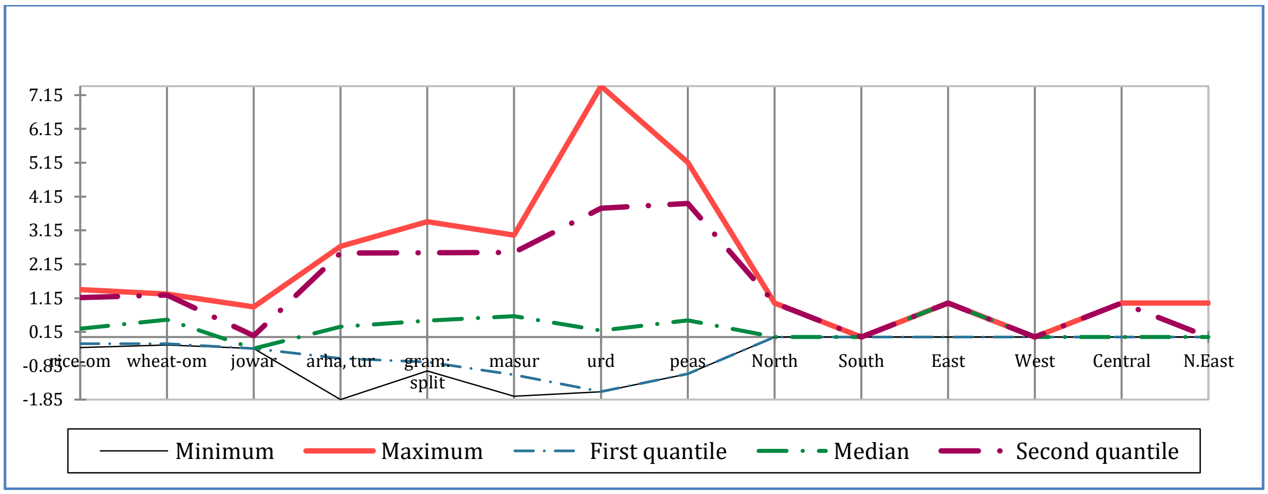
Cluster 4



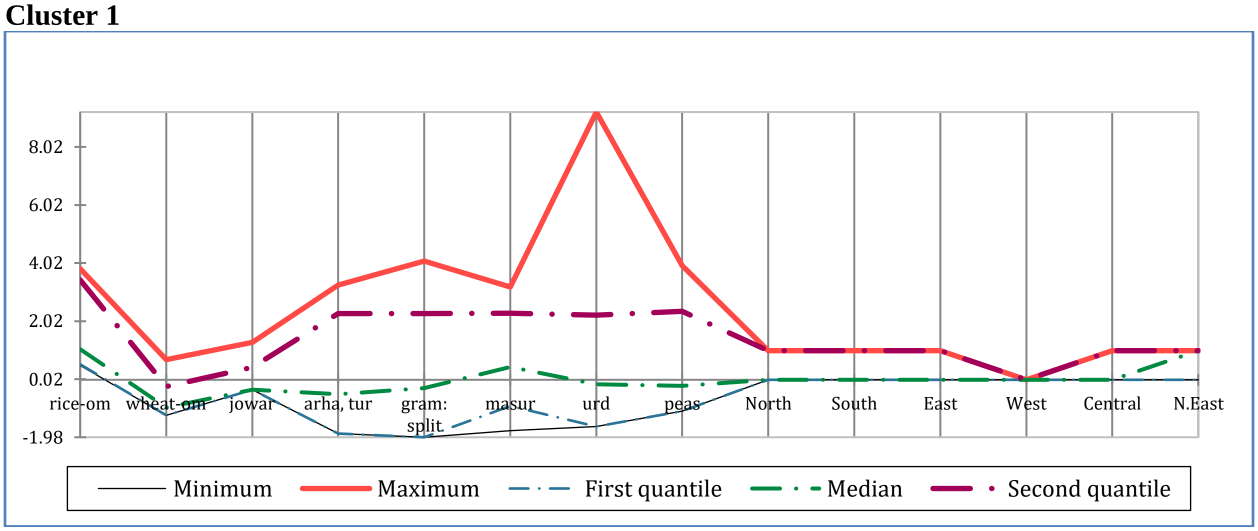
Cluster 3



Cluster 2



Cluster 1



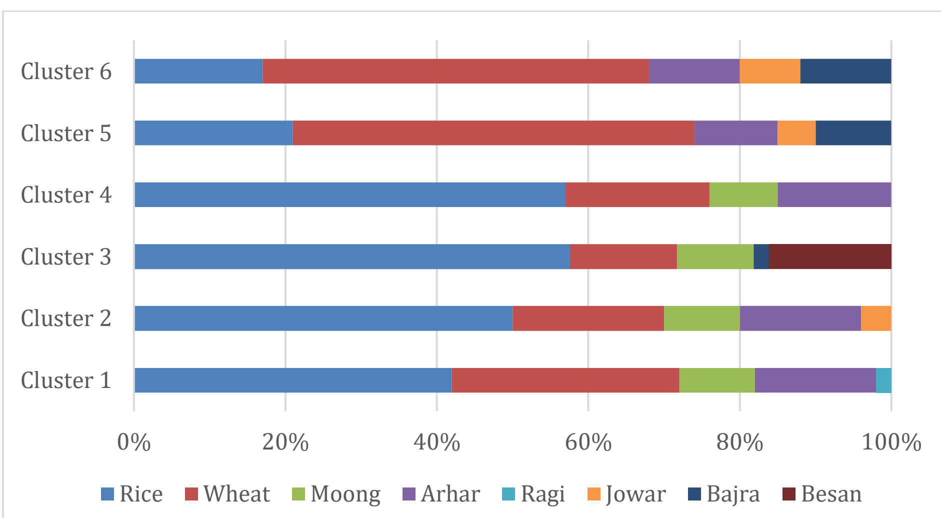
In clusters 1 and 2, rice is the key cereal preferred. As noted in the correlation matrix, clusters with rice preference have a lower preference for wheat. Masur seems to be a preference in Clusters 1 and 2. Interestingly, Western districts do not fall in this rice-preferring cluster. We build on these relative preferences in the following Section 4.3 to design customized grain assortments for each cluster.

We note that our clusters are reasonably balanced in terms of number of districts. Cluster 1 contains 150 districts, Cluster 2 contains 104 districts, Cluster 3 contains 105 districts, Cluster 4 contains 132 districts, Cluster 5 has 129 districts, Cluster 6 has 5 districts. Cluster 6 is exceptional in terms of high preference for jowar. In terms of geographical representations, in Cluster 1, 82% districts are in East and North East, in Cluster 2, 82% districts are in East, in Cluster 3, 82% districts are in North and West, in Cluster 4, 74% districts are in South. Cluster 5 consists a near equal representation from across geographies. Cluster 6 consists over half of the districts from West.

4.3 Model Baskets for each Cluster

For each of the determined clusters, we design a food basket by implementing an algorithm inspired by the 'Bin Packing' or 'Knapsack' problem. The following stacked bar charts show the composition of the basket for each cluster. Each household is eligible for a 35 kg food basket.

Figure 7: Food Basket Composition for each Cluster



As shown in Figure 7, each colour represents a particular food grain and the percentage represents a fraction of the household basket filled with each food grain type. As expected, rice and/or wheat are the staple food grains. However, their quantities are slightly reduced compared to the current values in PDS Operations. This part is substituted by locally grown and sourced pulses and coarse cereals that provide access to a larger variety of grain groups and food components. Clearly, in each cluster we propose a 20 to 35% basket composition by pulses and/or millets. It is interesting to note that a small intervention can help fulfil 65 to 75% of the ICMR recommended pulses intake.

4.4 MILP Results from Distribution Optimization Problem

We calculate an estimate of procurement, storage and distribution costs of the government by using the Mixed Integer Linear Program described in the Methodology Section 3.1.3. In Table 7, we report the aggregate cost statistics obtained from our model.

Table 7: Estimates of Overall Costs to be Incurred by Government Annually

Cost components	Cost (INR Crores)
Procurement Cost	47,284
Storage Cost	2887
Transportation Cost	937
Number of Storage Centers	845 (count, unitless)

The cost estimates shown in Table 7 serve as a reference benchmark. It is hard to compare these costs with overall PDS costs incurred by government because our target households belong to only one of the three groups (AAY, BPL, APL) that are PDS beneficiaries, currently.

4.4.1 Geographic Spread of Procurement

Figures 8a and 8b below shows the procurement strategy of India before and after our proposed analysis, respectively. We see that four states contribute to 65% of the procurement of rice and wheat in the status quo. With our analysis, we proposed sourcing equitably from all states in India. Figure 8b shows our proposed procurement strategy. The legend shows that all states contribute to 0 to 10% of the total procurement in the country. The two larger states of India, Madhya Pradesh and Uttar Pradesh, contribute to ~15% due to their sheer size of arable land area and given that they are the among most populated states in India.

Figure 8: Procurement

Figure 8a: Spatial Spread of current procurement

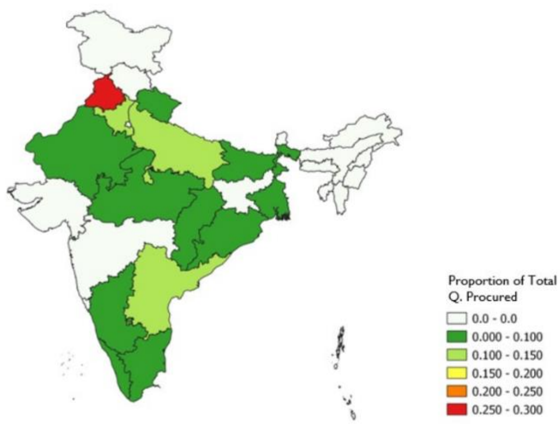
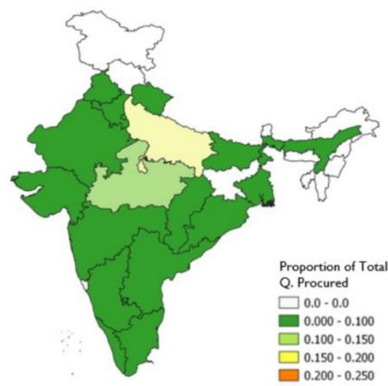


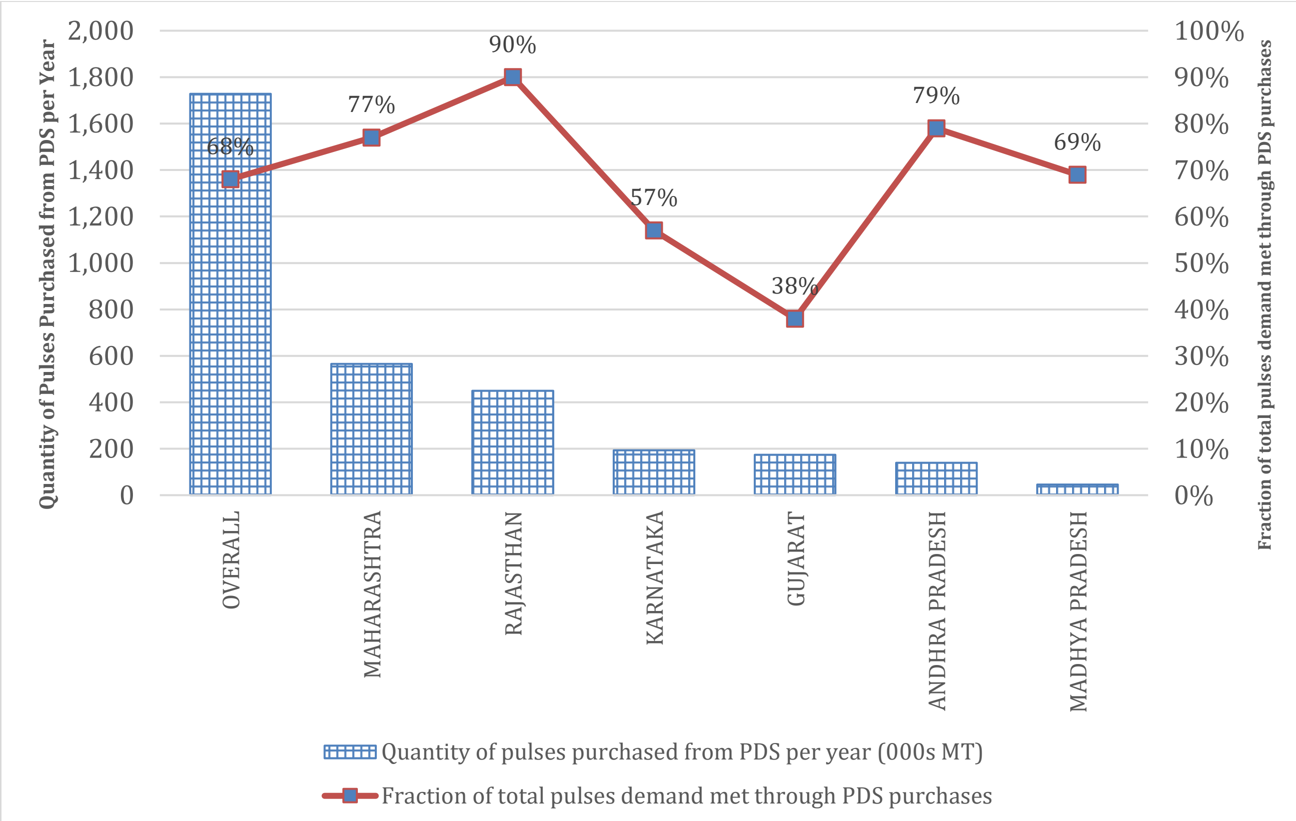
Figure 8b: Spatial Spread of proposed



4.4.2 Yearly Quantity of pulses distributed through PDS in the State

Our proposed food baskets are relevant to drive changes because they can help the target households in improving an affordable access to nutritious grains. Open market prices of pulses vary easily between INR 25 per kg and above. However, with our proposal, a large share of the recommended requirement of pulses can be catered via the PDS at a subsidized price. This means, target households will have to depend less on open market purchase of pulses for their daily requirements. Figure 9 shows the percentage of overall pulses purchase that can be addressed via the PDS by our proposed model. In the X – axis, we see the six states with the highest proposed purchase of pulses, and the quantities of their purchase of PDS pulses is depicted by the left Y - axis. Clearly, our model proposes that overall 68% of pulses, which are today accessed via the open market, can be purchased via the PDS according to the right Y – axis by our proposed intervention. This means that PDS channel can make more food grains available at subsidized prices across states, which may imply a migration from the open market to the PDS, and an increased consumption of these protein-rich food grains across poor households.

Figure 9: Yearly Purchase from Pulses that can be actualized through PDS



4.4.3 Scenario Analysis

We run a scenario where we put a threshold of 2 kg to the minimum quantity of food grain that can be added to a food basket. We note the proposed pulses/cereals in the food basket change for different clusters when we enforce the minimum threshold. Table 8 shows the quantities of food grains received by different districts in some states.

The percentages represent the fraction of districts out of total districts in each state that still receive cereals/pulses in the food basket after the minimum quantity threshold is enforced. The columns showing quantities in kg represent how much cereals/pulses do the districts receive in the food baskets in this scenario. We infer that even though there is a certain taste preference for diverse, more nutritious cereals and pulses in the states, it is relatively small in comparison to the staple cereals – rice and wheat, in many districts less than 2 kg per month per household. Hence, there is a need for increasing awareness of the benefits of these cereals/ pulses among the population through active campaigns, alongside improving their logistical access. Improving availability of food at affordable prices will be translated into increased nutrition levels only when people actively prefer these alternatives in their regular diets.

Table 8: Scenario Analysis – Minimum Quantity of Food grains in Basket

State *(Share of PDS Purchase of Millets and Pulses in Country)*	**Most Purchased Millets/ Pulses per State** *(Share of Millets/ Pulses in State PDS Purchase)*	**[2 kg - 3 kg)**	**[3 kg - 4 kg)**	**[4 kg - 5 kg)**	**≥ 5 kg**
Maharashtra (31%)	Jowar (21%)	6%	9%	12%	32%
	Arhar (10%)	3%	9%	0%	12%
Karnataka (29%)	Jowar (17%)	0%	3%	7%	34%
	Moong (12%)	3%	7%	0%	38%
Gujarat (17%)	Bajra (9%)	28%	12%	4%	16%
	Maize (8%)	8%	0%	8%	0%
Rajasthan (10%)	Bajra (6%)	16%	0%	9%	13%
	Besan (4%)	3%	9%	0%	0%
Andhra Pradesh (7%)	Ragi (7%)	0%	4%	0%	0%

5. TAKEAWAYS AND CONCLUSIONS

The key insights from our analyses are i) the decentralized, localized procurement of food grains, ii) access to a more nutritionally balanced supply of food basket at an affordable price. More generally, we propose a modelling framework for matching supply and demand of locally sourced nutritious food by customer segmentation and assembling a customized assortment or 'food basket' for each segment. We, then deliver the proposed food grain baskets to all districts (i.e., smaller geographical units within clusters) through a large-scale distribution network. In the following sub-sections, we conclude this investigation with the key insights on geographical procurement distribution, benchmarking nutrition, managerial insights and recommendations for future research.

5.1 Detailed conclusions

In this investigation, we build a methodology using mathematical and statistical modelling to understand the impact of supplementing with other millets, the traditional Indian public distribution system (PDS) that supplies rice and wheat to the Indian population. We use public data sources to consider household's purchase-consumption patterns and supply of grains for 625 districts in India. Also, we extract data from Google APIs, Food Corporation of India or we calibrate data to build estimated functions for relevant parameters such as the distance between districts, time taken for travel, and costs. We use cluster analysis to determine a group of customized food grain baskets) for distribution to different clusters in India constrained by local production patterns and traditional taste preferences of consumers.

We look for improving the diversity of food group combinations purchased and enhance the protein content by adding pulses into the food basket. In addition, we formulate a Mixed Integer Linear Program (MILP) model to distribute the grain baskets through the PDS minimizing government's total costs. The optimization model contains overall close to 80,000 constraints and on the order of 10 million decision variables. Finally, we performed some scenario analyses. Our model proposes that all target households can obtain ~100% of their cereal demands and ~60% of the pulses demand from the PDS channel at a subsidized rate of INR 2 per kg, against an open market rate of ~20 per kg. This can be a significant incentive for the poor households to increase the intake of pulses in their diets. Furthermore, this can be delivered via the PDS channel to gain economies of scale and take advantage of the penetration this social government-based program has all over India.

Moreover, our model includes important contextual factors, such as nutritional, social, cultural,

economic, agricultural, and food supply approaches, which are essential for its successful implementation (Ridgeway, et al., 2019).

5.2 Recommendations and managerial insights

One of the key contributions of this study, unlike traditional diet problems, is that the nutrition benchmarks at a 'food group' level are used to select an assortment of 'food items' that are local to the geography, also acknowledging the fact that people eat foods and not isolated nutrients (D.R. & Tapsell, 2007). It is critical to note that a complete diet at a food group level provides many nutrients that together enable their overall bio-absorption, which cannot be met via unpalatable food items high in nutrients individually, or supplements or fortified foods (*ICMR Food Group Recommendations for Adults in India*, 2020).

For example, we work with ICMR benchmark recommendations of 270g of cereals and 90g of pulses per person per day. However, the 'food items' that we propose for 'cereals and pulses' are different in each cluster and obtained by local demand, food preferences and supply matching. Also, in the case of non-availability of the particular cereal or lentil that we propose, it can be substituted by any other that is locally available. Also, from a supply chain perspective, 'food group' level recommendations are easier to scale across geographies. This is because, such recommendations are not too sensitive to the risk of unavailability of the exact food item proposed by model.

Our analysis can enable the government to build a sustainable food-ecosystem that can match supply with demand of nutritionally diverse food grains. The current public distribution system infrastructure is an added advantage to experiment such basket design and distribution. It can improve the affordability of end consumers, while being cost effective for the government due to economies of scale.

Recent news articles in India indicate that the government explored a temporary one-time distribution of pulses to many PDS beneficiary households during the COVID-19 pandemic induced food crisis for the poor. Our study can be a useful framework to the government for regularizing their initiative. One major challenge the government is reported to have faced was the location of milling facilities for pulses. As a result, in many cases, unmilled pulses were sent out from the storage facilities. It would be worthwhile to add another stage in the distribution network called 'milling facilities' to come up with their optimal count and locations to enable a large-scale distribution of edible pulses. A lot of such improvisations will happen based on first rollouts, and primary empirical evidence of customer behaviour and adoption of the initiative by the target households.

Overall, this is an active problem area open for cost effective and robust solutions. The societal impact of such an initiative is of even more importance in the COVID-19 pandemic because frequent lockdowns affect livelihoods of the poor. Ensuring food security through affordable access, and ready availability of food is very crucial in these times.

5.3 Limitations and future work

Some of the limitations of our analysis are that production dataset and estimates of taste preferences are based on secondary data sources. (NSSO Consumption Expenditure Data,68th Round, 2011-12) itself is self-reported. It would be worthwhile to collect primary data based on more accurate records. For the lack of data, we needed to calibrate and estimate some parameters like storage costs and rely on external sources. From the theoretical perspective, we did not address the problem from a tactical-operational perspective to model granularly several processes. Also, we did not follow stochastic or multi-criteria mathematical approaches, given that we wanted to build a base case.

Finally, many future research opportunities exist in this area. Our work is a step towards combating acute protein-energy malnutrition, especially in the developing countries through an interdisciplinary effort between supply chain management and nutrition science. It is increasingly crucial to make 'nutritious foods' available, affordable and accessible at scale through robust distribution mechanisms. One immediate extension of this study would be a design proposal for including perishable foods into the basket as well. That should need more careful consideration for either faster deliveries within the shorter shelf life of the product, or introduction of cold storage networks for increasing their shelf life. The target households can be expanded to increase more economic and social sections of the society. It will be interesting to assess if increasing the scope of target households warrants an increase in the production capacity of India.

An interesting extension of this work will be the design and implementation of this capability across non-government distribution channel partners. Moving to a different channel, might remove the incentive in the form of subsidy that the government is enabling today. It would be worthwhile to explore alternate mechanism designs for incentivizing consumption of nutritious baskets in non- government distribution channels. This work would need a better understanding of the open market.

References

APY Dataset. (2019-20). Retrieved May 9, 2021, from Crop Production Statistics Information System: https://aps.dac.gov.in/APY/Index.htm

Coverage and Entitlement Under NFSA. (2020). Retrieved May 8, 2021, from Government of India, Department of Food and Public Distribution: https://dfpd.gov.in/pds-caeunfsa.htm

Crespo, P., Ferreira, I., Naranjo, P., & Mejía-Argueta, C. (2020). *Increasing Consumer Awareness of Nutritious Food in Emerging Markets.* GCLOG Capstone Project Manuscript, MIT Center for Transportation Logistics.

D.R., J., & Tapsell, L. (2007). Food, not nutrients, is the fundamental unit in nutrition. *Nutrition Review*.

Davis, K., Chhatre, A., Rao, N., Singh, D., Ghosh-Jerath, S., Mridul, A., . . . DeFries, R. (2019). Assessing the Sustainability of Post Green Revolution Cereals in India. *Proceedings of the National Academy of Sciences*.

DeFries, R., Mondal, P., Singh, D., Agrawal, I., Fanzo, J., Remans, R., & and Wood, S. (2016). Synergies and trade-offs for Sustainable Agriculture: Nutritional Yields and Climate Resilience for Cereal Crops in Central India. *Global Food Security*, 11:44-53.

Deller, S., Canto, A., & Brown, L. (2015). Rural Poverty, Health and Food Access. *Regional Science, Policy and Practise*.

DES. (2011-12). *Department of Economics and Statistics, Government of India.* Retrieved from https://eands.dacnet.nic.in/

Devalkar, S., Deo, S., Chhatre, A., & Jha, B. (2020). Economic Impact of Introducing Coarse Grains in India's Public Distribution System. *Working Paper*.

EAT. (2019). *Food in The Anthropocene: the EAT-Lancet Commission on Healthy Diets From Sustainable Food Systems.* Lancet.

EAT Lancet Commission on Food, P. a. (2019). *The EAT-Lancet Commission on Food, Planet, Health*. Retrieved from The EAT-Lancet Commission on Food, Planet, Health: https://eatforum.org/eat-lancet-commission/

FAO, IFAD, UNICEF, WFP, & WHO. (2020). The State of Food Security and Nutrition in the World 2020. Transforming food systems for affordable healthy diets. *FAO.*

FAO-a. (2020). Retrieved from Food and Agricultural Organization of the United States: http://www.fao.org/india/fao-in-india/india-at-a-glance/en/

FAO-b. (2006). *Policy brief - food and agricultural organization of the United Nations.* . Retrieved from http://www.fao.org/fileadmin/templates/faoitaly/documents/pdf/pdf_ Food_Security_Cocept_Note.pdf

FCI. (2019-20). *Food Corporation of India Annual Report.* FCI.

Gazan R., Brouzes, M. C., Florent, V., Matthieu, M., & Anne, L. (2018). Mathematical Optimization to Explore Tomorrow's Sustainable Diets: A Narrative Review. *Oxford University Press on behalf of American Society of Nutrition*.

HLPE. (2017). *Nutrition and Food Systems. A report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security.* Rome: FAO.

IBM. (2020). *Knapsack Problems using CPLEX.* Retrieved from IBM CPLEX Help Document: https://www.ibm.com/docs/en/icos/20.1.0?topic=programming-knapsack-problems

ICMR. (2020, August). *ICMR Food Group Recommendations for Adults in India.* Retrieved May 7, 2021, from My_plate_for_the_day: https://www.nin.res.in/downloads/My_Plate_08-2020.pdf

K, H., Yan, B., Derek, H., & A, M. W. (2020). Affordability of the EAT Lancet Reference Diet: A Global Analysis. *Lancet Global Health*.

Kumar, S. (Ed.). (2015). *Report of the High Level Committee on Re-orienting the Role and Re-Structuring of Food Corporation of India.* External Committee. Retrieved from https://ageconsearch.umn.edu/record/230029/files/09-pg198-209-Documentation-01.pdf

Levi, R. P. (November, 2019). Optimal Interventions for Increasing Healthy Food Consumption Among Low Income Households. *MIT Sloan Research Paper No. 6053-19*, SSRN: https://ssrn.com/abstract=3486292.

Li C, G. A. (2018). Sociodemographics and access to organic and local food: A case study of New Orleans, Louisiana. *Cities*, pp. doi:10.1016/j.cities.2018.03.003, 79:141-150.

Lu C, B. D. (2008). Dietary intake and its contribution to longitudinal organophosphorus pesticide exposure in urban/suburban children. *Environmental Health Perspectives, 116, 537–542.)*.

Mark, H. E. (2016). Estimating dietary micronutrient supply and the prevalence of inadequate intakes from national food balance sheets in the South Asia region. *Asia Pacific Journal of Clinical Nutrition*.

Martello S., Pisinger D., & Vigo D. (2000). The Three Dimensional Bin Packing Problem. *Operations Research*.

MSP. (2021, January 8). Retrieved May 10, 2021, from Minimum Support Prices for Farmers: https://farmer.gov.in/mspstatements.aspx

National Sample Survey of India Dataset on Consumption Expenditure, 68th Round. (2011-12, August 2016 4). Retrieved from ICSSR: http://www.icssrdataservice.in/datarepository/index.php/catalog/91/study-description

Pollard, C., & Rowley, C. (2010). Working with Industry for the Promotion of Fruit and Vegetable Consumption. pp. 205-20. Retrieved May 2018

Rao, N., Min, J., DeFries, R., Suparna, G.-J., Hugo, V., & Jess, F. (2018). Healthy, affordable and climate friendly diets in India. *Global Environmental Change*.

Ridgeway, E., Baker, P., Woods, J., & M., L. (2019). Historical Developments and Paradigm Shifts in Public Health Nutrition Science, Guidance and Policy Actions: A Narrative View. *Nutrients*.

Ruel, M. (2003). Operationalizing dietary diversity: A review of measurement issues and research priorities. *The Journal of Nutrition*.

Sadler, R., Gilliland, J., & Arku, G. (2013). Community Development and Influence of New Food Retail Sources on the Price and Availability of Nutritious Food. *Journal of Urban Affairs*.

Santos, L. F. (2018). A Variable Neighborhood Search Algorithm for the Bin Packing Problem with Compatible Categories. *Expert Systems with Applications*.

Satija, A., Yu, E., Willett, W. C., & Hu, F. B. (2015). *Understanding Nutritional Epidemiology and Its Role in Policy*. American Society for Nutrition.

Wikipedia. (2020). *Bin packing problem*. Retrieved from Wikepedia: https://en.wikipedia.org/wiki/Bin_packing_problem#:~:text=In%20the%20bin%20packing%20problem,the%20number%20of%20bins%20used.&text=The%20decision%20problem%20(deciding%20if,bins)%20is%20NP%2Dcomplete.

World Health Organization, D. o. (2017). *Double duty actions for Nutrition.* Retrieved from Policy Brief.

World Public Health Nutrition Association. (2005). *The Giessen Declaration.* Public Health Nutrition. 8, 783–786.

Yong-jun, Q. (2019). Improved Food Environments for Healthy Diets and Enhanced Nutrition. *Science Direct*.

APPENDIX

Appendix A.1: Principal Component Analysis and K-means Clustering

For this analysis, we use the XLSTAT add-in for Microsoft Excel.

Data Preparation: Our consumption estimates data contains the following feature variables:

- District, State, Geographic Region (East, West North, South, Central, North East)
- For each district, per month, per household consumption estimates of following:
    - Pulses: arhar, gram, moong, masur, urd, peas, khesari, gram, besan
    - Cereals: rice (open market and PDS), wheat (open market and PDS), jowar, bajra, maize, ragi, barley, small millets

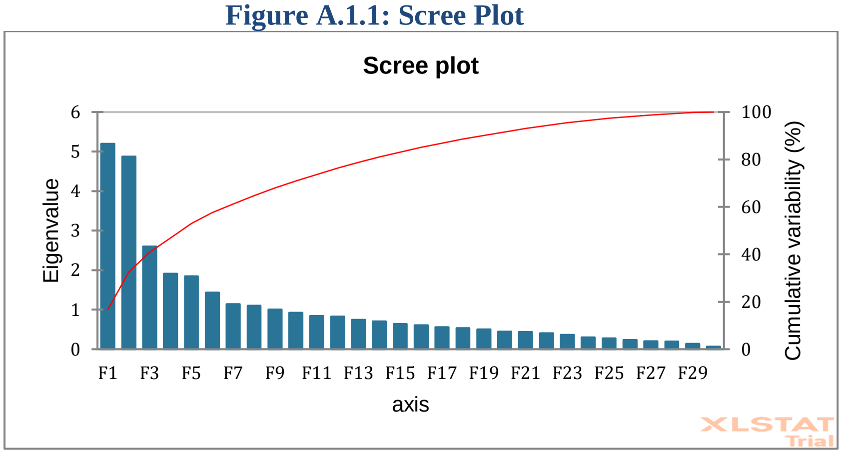
We represent each geographic region by a dummy variable and perform a dimensionality reduction to understand the key variables. Eigen values shown in Table A.1.1 represent the quality of projection from the n-dimensional space to lower number of dimensions. Each eigen value corresponds to a factor, and each factor to one dimension. A factor is a linear combination of the initial variables and all factors are uncorrelated. [5]

Table A.1.1: Eigenvalues from Principal Component Analysis

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Eigenvalue	5.198	4.871	2.594	1.907	1.847	1.433	1.136	1.104	1.001	0.920
Variability (%)	16.766	15.71	8.367	6.152	5.958	4.622	3.665	3.560	3.230	2.967
Cumulative %	16.766	32.48	40.85	46.99	52.957	57.58	61.24	64.80	68.03	71.00

Figure A.1.1 is a scree plot showing the cumulative variability in data explained by each incremental factor introduced.

Figure A.1.1: Scree Plot



[5] https://help.xlstat.com/s/article/principal-component-analysis-pca-in-excel?language=en_US

Figure A.1.2: Biplot (Axes F1 and F2: 32.48% explained variability)

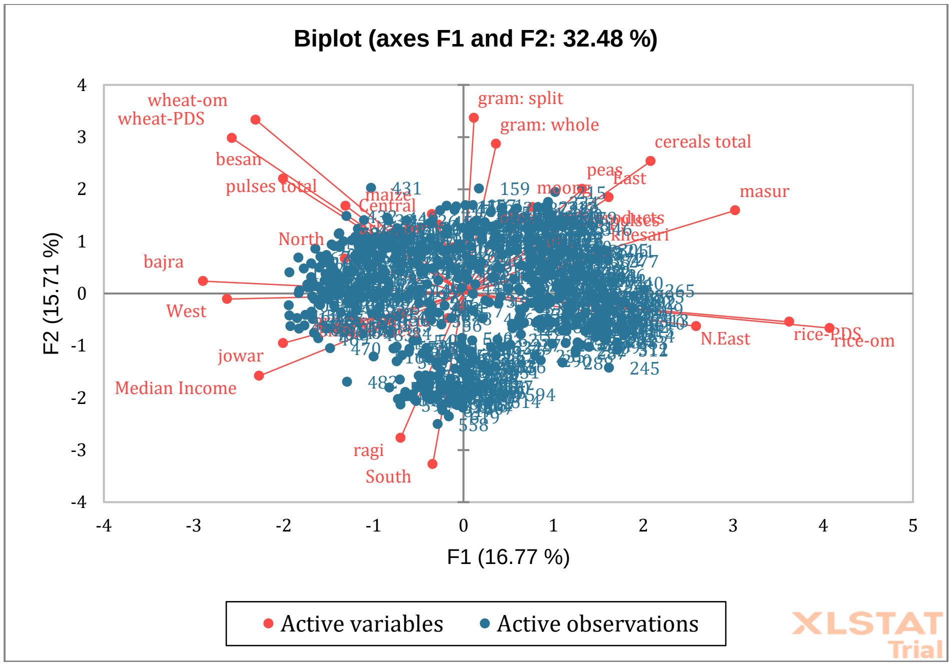
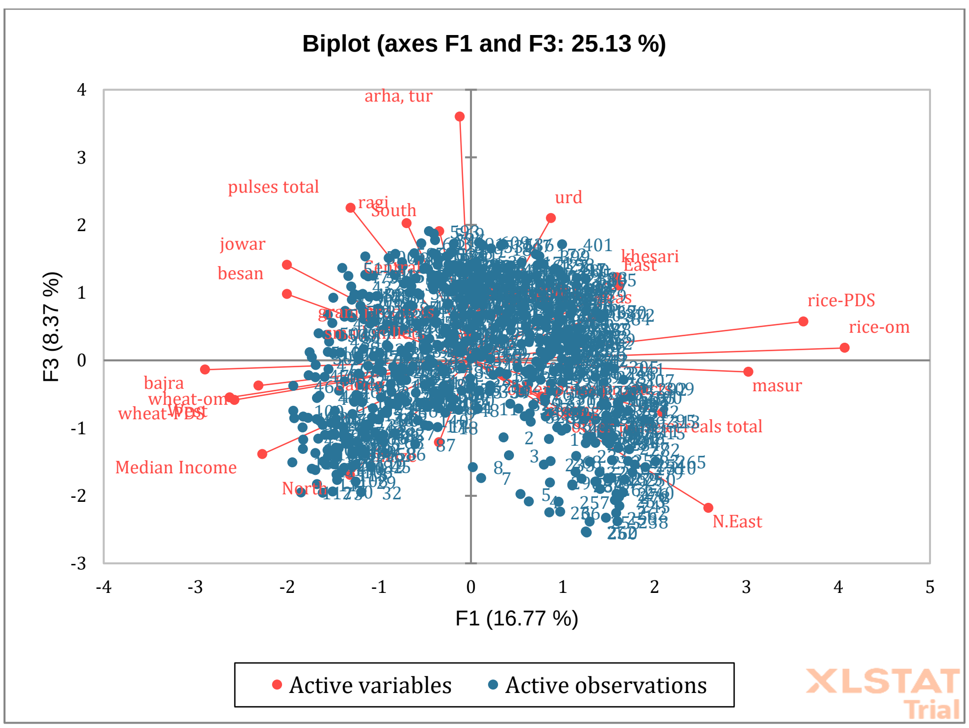


Figure A.1.3: Biplot (Axes F1 and F3: 25.13% explained variability)



Figures A.1.2 and A.1.3 represent the feature variables and their association with the factors. The feature variables closer to the 'factor axis' have a higher contribution to the factor. The feature variables whose axes make small angles to each other have high correlation.

Table A.1.2: Square cosines of variables

	F1	**F2**	F3	F4	F5
North	0.086	0.022	0.142	0.162	**0.336**
South	0.006	**0.530**	0.181	0.015	0.036
East	0.130	0.168	0.060	0.062	**0.182**
West	**0.343**	0.001	0.015	0.286	0.087
Central	0.004	**0.086**	0.056	0.000	0.001
N. East	**0.332**	0.020	0.236	0.095	0.011
Median Income	**0.256**	0.124	0.095	0.000	0.237
rice-PDS	**0.652**	0.015	0.016	0.006	0.001
rice-om	**0.823**	0.022	0.002	0.015	0.000
wheat-PDS	0.329	**0.440**	0.017	0.031	0.001
wheat-om	0.265	**0.550**	0.007	0.026	0.019
jowar	0.199	0.045	0.099	**0.287**	0.008
bajra	**0.416**	0.003	0.001	0.124	0.060
maize	0.006	**0.115**	0.073	0.040	0.013
barley	**0.024**	0.007	0.001	0.007	0.001
small millets	0.001	0.011	0.000	**0.230**	0.039
ragi	0.024	**0.380**	0.204	0.012	0.072
cereals total	0.216	**0.320**	0.030	0.014	0.020
arha, tur	0.001	0.039	**0.644**	0.022	0.003
gram: split	0.001	**0.563**	0.025	0.001	0.029
gram: whole	0.007	**0.408**	0.020	0.005	0.059
moong	0.029	0.136	0.014	**0.192**	0.002
masur	**0.453**	0.126	0.001	0.016	0.002
urd	0.038	0.019	**0.219**	0.011	0.079
peas	0.087	**0.199**	0.019	0.002	0.026
khesari	**0.126**	0.028	0.074	0.001	0.000
other pulses	0.055	0.053	0.030	0.093	**0.238**
gram products	0.000	0.004	0.008	0.005	**0.024**
besan	0.199	**0.240**	0.048	0.009	0.001
other pulse products	0.006	0.058	0.003	**0.137**	0.078
pulses total	0.085	0.140	**0.252**	0.000	0.181

Values in bold correspond for each variable to the factor for which the squared cosine is the largest

Table A.1.2 shows the fractional contribution of each feature variable to the factor axes. Based on these values, we select rice-om, wheat-om, jowar, arhar, gram:split, masur, urd, peas as the key drivers for clustering.

K-means clustering:

As discussed in Results section, we select k = 6, based on the elbow method (i.e., plot of within-cluster variance against number of clusters). For k=6, Table A.1.3 shows the within-class and between-class variances. Table A.1.4 shows the details of the clusters. The row 'Objects' represents the count of districts in each cluster. Clearly 5 clusters are balanced as the number of districts is reasonably close. The sixth cluster shows an exceptional preference for jowar and hence we keep it separate.

Table A.1.3: Variance decomposition for the optimal classification

	Absolute	Percent
Within-class	100.563	18.6%
Between-classes	441.402	81.4%
Total	541.965	100.0%

Table A.1.4: Cluster Characteristics

Class	**1**	**2**	**3**	**4**	**5**	**6**
Objects	150	104	105	132	129	5
Sum of weights	150	104	105	132	129	5
Within-class variance	173.271	74.071	82.177	53.964	85.904	547.626
Minimum distance to centroid	1.148	1.070	3.358	1.783	1.762	11.881
Average distance to centroid	10.568	7.768	8.536	6.764	8.265	19.254
Maximum distance to centroid	39.962	17.669	20.497	14.744	21.382	34.390

Appendix A.2: Basket Formulation for each Cluster

This formulation is inspired by the knapsack problem where the objective is to Maximize: Taste Preference Match (Value) for a Cluster constrained by: capacity of the basket

Sets:
Cereals C ={rice, wheat, jowar, bajra, ragi, maize, barley, small millets}
Pulses P ={arhar, moong, besan, gram, masur, urd, peas, khesari}

Parameters:
Total Weight of Cereals in Basket W_c = 26 kg
Total Weight of Pulses in Basket W_p = 9 kg
Value (Taste Preference) of grains in each Cluster:

- Standardized Value of Average monthly consumption per household (lies between 0 and 1). This was considered as a proxy for taste preference in the model. So, we populate the following table:

Table A.2.1: Grain Cluster Value Matrix

Grain\Cluster	c_1	c_2	c_3	c_4	c_5	c_6
rice						
wheat						
jowar						
bajra						
ragi						
maize						
barley						
small millets						
arhar						
moong						
besan						
gram						
Masur						
urd						
peas						
khesari						

Each food grain has a weight and a value.

- The weight is its quantity that goes into the basket,
- The value is the taste preference match score as given in Table A.2.1, Grain Cluster Matrix.

We feed the inputs into CPLEX ILOG prebuilt knapsack-module. The output is the maximum value (or taste preference) that can be assigned to a cluster subject to weight (basket size) constraints. It is of the form:

Cluster i → [Cereal 1…j, Pulses 1…k], for i <- (1 to 6)