CHANGING SHOPPING HABITS IN SUPERMARKET DURING COVID-19

by
EDA HELİN GÜNDEŞ

Submitted to the Sabancı Graduate Business School
in partial fulfilment of
the requirements for the degree of Master of Science in Business Analytics

Sabancı University
June 2021

CHANGING SHOPPING HABITS IN SUPERMARKET DURING COVID-19

Approved by:

Prof. FUSUN ULENGIN ....
(Thesis Supervisor)

Assoc. Prof. AYSE KOCABIYIKOGLU ....

Prof. YUSUF ILKER TOPCU ....

Date of Approval: June 30, 2021

ABSTRACT

CHANGING SHOPPING HABITS IN SUPERMARKET DURING COVID-19

EDA HELIN GUNDES

BUSINESS ANALYTICS M.Sc. THESIS, MAY 2021

Thesis Supervisor: Prof. FUSUN ULENGIN

Keywords: COVID-19, consumer behavior , consumption , supply-chain, online-transaction

On December 31, 2019, the world was shaken by the news of a new virus observed in Wuhan city, Hubei, China. Covid-19 affected all the industries but mostly the retail industry. Stock-pilling and panic buying behaviors were seen in supermarkets. Most of the retailers could not satisfied the basic demands with their supply plans. Therefore, the main aim of this research is to find out the changing habits and consumer behaviors due to the effect of COVID-19. The analysis was conducted using the CRISP-DM model steps. The data set was revealed from one of the large retailers in Turkey. Initially, cluster analysis was conducted to group the products according to their similarity in terms of shopping behaviors. Subsequently, stepwise regression, best subset selection and Lasso models were applied to each cluster in order to analyze the relation between consumer shopping behavior and Covid cases. The best regression model for each cluster is selected using 10-fold cross-validation as well as Cp and adjusted R2. These analyses were applied both to the physical and online market data to highlight the shift from physical to online market. Eventually, this research found out that there is a significant shift from physical to online market during Covid-19. Many of the product groups have seen a significant increase in sales compared to 2019 with the effect of panic-buying. The shifts in sales levels are statistically related with the Covid-19 case numbers and curfews.

ÖZET

COVID-19 SÜRECİNDE SÜPERMARKET ALIŞVERİŞİ ALIŞKANLIKLARININ DEĞİŞİMİ

EDA HELİN GÜNDEŞ

İŞ ANALİTİĞİ YÜKSEK LİSANS TEZİ, MAYIS 2021

Tez Danışmanı: Prof. Dr. FÜSUN ÜLENGİN

Anahtar Kelimeler: COVID-19, tüketici davranışları , tüketim, tedarik zinciri, çevrimiçi satınalma

31 Aralık 2019'da Dünya, Çin'in Hubei eyaletine bağlı Wuhan kentinde gözlemlenen yeni bir virüs haberiyle sarsıldı. Covid-19, perakende sektörü başta olmak üzere tüm sektörleri etkiledi. Süpermarketlerde stok biriktirme ve panik satın alma davranışları görüldü. Perakendecilerin çoğu, tedarik planlarıyla temel talepleri karşılayamadı. Bu araştırmanın temel amacı, COVID-19'un etkisiyle değişen alışkanlıkları ve tüketici davranışlarını ortaya çıkarmaktır. Analizler, CRISP-DM model adımları kullanılarak yapılmıştır. Veri seti, Türkiye'deki büyük perakendecilerden birinden sağlanmıştır. Öncelikle, ürünleri alışveriş davranışları açısından benzerliklerine göre gruplandırmak için küme analizi yapılmıştır. Daha sonra tüketici alışveriş davranışı ile Covid vakaları arasındaki ilişkiyi analiz etmek için her bir kümeye Stepwise regresyon, En iyi alt küme seçimi ve Lasso modelleri uygulanmıştır. Her küme için en iyi regresyon modeli, Cp ve Adjusted R2'nin yanı sıra 10 kat çapraz doğrulama kullanılarak seçilmiştir. Bu analizler, fiziksel pazardan çevrimiçi pazara geçişi vurgulamak için hem fiziksel hem de çevrimiçi pazar verilerine uygulanmıştır. Sonuç olarak, bu araştırma, Covid-19 sırasında fiziksel pazardan çevrimiçi pazara önemli bir geçiş olduğunu , panik satın almanın etkisiyle birçok ürün grubunda 2019'a kıyasla satışlarında önemli bir artış görüldüğünü, ve satış rakamlarındaki değişimlerin istatistiksel olarak Covid-19 vaka sayıları ve sokağa çıkma yasakları ile ilişkili olduğunu ortaya çıkarmıştır.

ACKNOWLEDGEMENTS

First of all, I would like to express my thanks to my beloved thesis advisor and mentor Prof. Füsun Ülengin for her help and support not just for this thesis and also throughout my master education. Without her guidance and help I would not be able to find my way and understand what I want for the rest of my life.

I would like to express my special thanks of gratitude to Prof. Burç Ülengin for his guidance and support in completing this research.

I also want to thank all of my professors that I had a chance to take class from throughout my master education. Prof. Abdullah Daşçı, Assoc. Prof. Burak Gökgür and Assoc. Prof. Ayşe Kocabıyıkoğlu thanks for all of your guidance and precious classes that you've given to me.

I would like to thank Prof. Emel Aktaş and Prof. İlker Topçu for letting me a part of their team. It was a pleasure to work with you and learn from you.

Lastly, I would like to thank to my mother Zeynep Begüm Gündeş, my father Mustafa Gündeş, my little brother Altuğ Baran Gündeş and all of my friends. They always supported me throughout this journey. I am really grateful to them.

This thesis is dedicated to my beloved family, professors and friends who supported me in all parts of this journey no matter what.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATONS

1. INTRODUCTION

On December 31, 2019, the world was shaken by the news of a new virus observed in Wuhan city, Hubei, China. The Chinese government warned the World Health Organization (WHO) that the reason for the spreading cases was unknown. The International Committee on Taxonomy of Viruses(ICTV) named this disease Severe Acute Respiratory Syndrome Coronavirus-2(SARS-COV-2). After a while, the name was changed to “COVID-19" by WHO, making the official name COVID-19. As of 12 of June 2021, the total number of confirmed cases is 174,918,667 million and confirmed deaths are 3,782,490 million. (Who, 2021) COVID-19 has changed all parts of life. People have learned to be more careful about self-hygiene, and masks have become an important part of our lives, even starting to be used as accessories with different colors and designs. Countries however used different rules and regulations such as curfews and mask requirement to deal with pandemic in a most efficient way for its citizens. National economies have also been affected since the first news from China. As China is one of the most important countries in production, all other countries including Turkey were affected by the fact that it became unreliable in terms of shipping products. During this period, the market fluctuated in every field. This fluctuation and changes in the marketplace showed that supply chains that were once expected to be profitable and reliable, now need to be resilient and sustainable. The ones who offer these standards had a competitive advantage, among others. The companies have seen possible risks that they will face if they could not operate its supply chain during crisis. (Saenz, Stephan, Terino, Bysong & Gnamm, 2021) Therefore, companies started to focus on making their supply chain distribution more resilient than before. Labor intensive industries were the most affected from pandemics as they were most exposed to these types of shocks in general. A research conducted by McKinsey shows that, companies can adopt different resilience strategies under 3 main headlines; namely minimizing exposure to shocks, strengthening the risk management and transparency, and improving financial and operational capacity.(Lund et al., 2020) Figure 1.1 shows the tactics that they provided to other companies to successfully overcome their current problems.

Figure 1.1 Mckinsey suggestions of resilience strategies(Lund et al. (2020) pg.74)

In addition to the supply side, the demand side is also affected in all industries. Retail industry is one of the most affected ones during the pandemic. Changing rules and regulations, especially curfews affected the demands. The limitations that governments applied led people to panic-buying all around the world. After most of the companies preferred to work home-office, people started to spend more time in their home which led them to cook at home to spend time instead of eating outside or ordering food. The markets were mostly unready for this panic-buying; most of the shelves were empty for a couple of days due to supply problems with the producers. At the beginning of March 2020, when compared to March 2019, grocery spending extremely increased after public awareness and media effect showed itself. People started to be stockpiling and panic-buying before the first lockdown and started to be normalize at the end of march.(Hall, Prayag, Fieger & Dyason, 2020). People mostly preferred to buy legumes and pasta to stock at home in case of curfews. Lots of people started to bake their own bread at home as the lock downs started to happen all weekends. Most of the people have chosen online shopping during this period instead of going outside. Most of the markets had an infrastructure for online shopping but with the increasing demand a supply and logistic problem occurred both in online and physical stores. Many of the markets could not met the demands and had to give customers 3-4 days deliver time for their orders. These 3-4 days was even higher for non-consumable products. Logistic firms had a very high level of demand from the firms all over the country and with the limited number of employees the logistic process started to be delayed for all citizens all over the country.

Turkey started to become ready for the upcoming threat after the news from China spread all around the world. The first case was seen in March 11, 2020, in Turkey. Subsequently, lots of measures had been started to apply in several parts of the country. On January 10, 2020, the Ministry of Health decided to establish a committee that is called 'Coronavirus Scientific Board' to limit the spread of the virus. Five days later, the government closed all primary and high schools by 1 week, the universities by 3 weeks until a further notice. Most of the public places were shut down. Such as parks, pubs, restaurants, sport centers, shopping centers etc. At the end of March 2020, the total number of cases was 13.531. As of April, the limitations for going outside had started on weekdays and it was restricted for all age groups on weekends. Additionally, government closed 31 big city's borders to limit the spreading. At the end of April 2020, the total number of cases was 120.204. Compared to the March it was 8 times greater even with all these measures. At the beginning of the May 2020, the age group of 65+ started to go outside just 1 day and for 4 hours. Shopping centers were opened again but many of the stores did not opened it or opened for limited hours. Normalization period has been planned for the beginning of June 2020. With the normalization, all the social and physical limitations planned as to be stopped and the life will be continuing as it was before just to be careful about self-hygiene, social distance, and compulsory use of mask both inside and outside. At the end of May 2020, the total number of cases was 163.942. The enormous increase from March to April was decreased. After the normalization period begin at 1st of June 2020, Turkish government almost did not apply any limitations until the 20th of November 2020. At the end of the June, the total number of cases reached to 199.906. After the normalization period started, even with the masks and social distance the total number of cases increased to 35.964. At the end of July, the total number of cases increased to 230.837. The cases in August and September were 270.133 and 318.663 respectively. Even with social distancing and masks the number of cases increased almost 30.000 each month for the first wave. (Wikipedia, 2021)

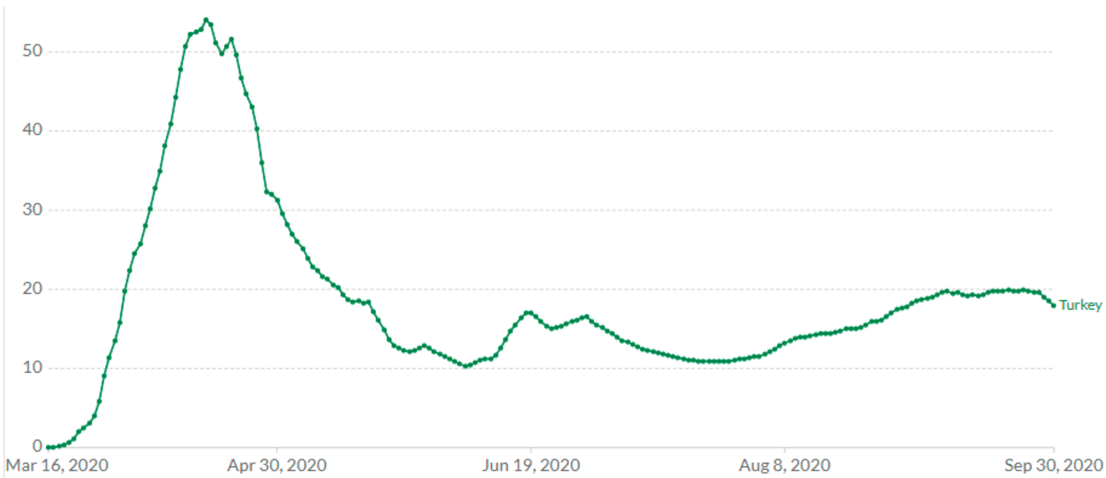


Figure 1.2 March 16 - September 30 2020 Number of New Cases in Turkey(Owid, 2020)

On October 2nd 2020, large public gatherings prohibited in all cities until December 1st. At the end of October, the total number of cases were 375.367. On November 20th, government started to apply curfew for age older than 65 and under 20 again. Grocery stores and pharmacies continued to work within limited capacity but besides them all businesses and worship places stopped their indoor activities. At the end of November 2020, the number of total cases were 500.865. On December 10, seven-day averages of deaths, cases and hospitalization reached its peak from the beginning. As can be seen from the graph, there is an enormous increase at the case numbers at 25th of November. The reason for this peak is that Turkish government started to publish the real numbers of cases. The government stated that before that date they were only stating the numbers of people who are showing symptoms. At the last day of year, the total number of cases were 2.208.652 till the Covid-19 first seen in Turkey. (Wikipedia, 2021)

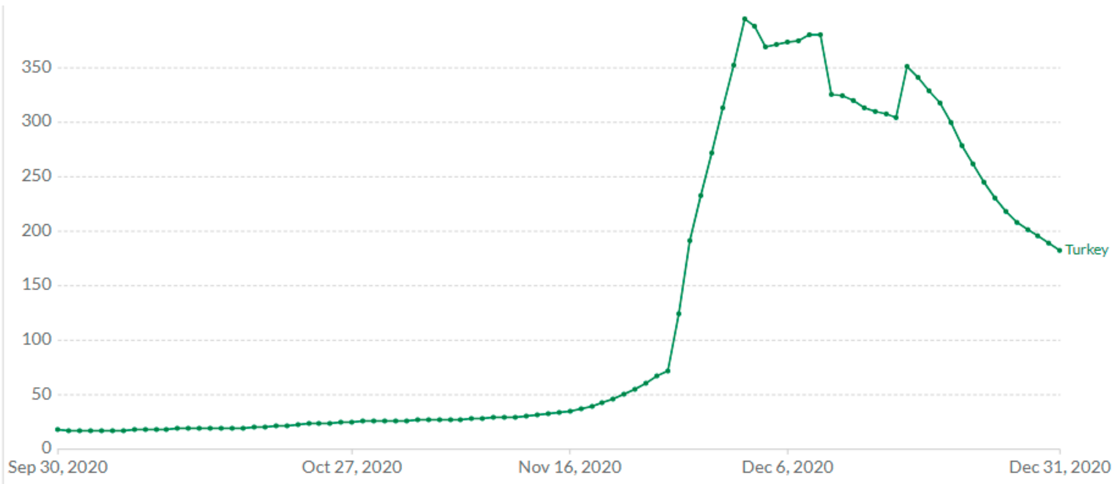


Figure 1.3 September 30 - December 31 2020 Number of New Cases in Turkey(Owid, 2020)

The aim of this research is to understand the effect of Covid-19 on consumer behavior for retail industry. To understand the supply chain disruptions and differences compared to 2019, the answer of following questions are investigated.

1-Which product groups experienced a significant shopping behavior change compared to 2019?

2-Is there any difference between online shopping and physical market sales compared to 2019?

3- How has the Covid-19 affected customer buying habits?

For this manner, the analysis is conducted using one of the Turkey's largest retailer's 2019-2020 sales data. The retailer provides service with a total of 2,370 stores and approximately 45,000 employees in 81 provinces in Turkey. There are over 2000 variety of product groups and brands in its stores. Besides having food and food related groups, it also provides glassware, textile, toys and electronics. Retailer's turnover was 22.9 billion TL in 2019. In 2020, this figure increased to 28.8 billion TL. There are two different datasets. One of them includes information about physical market sales. The other dataset contains information about online market sales. Besides having a platform difference all columns, product groups and dates are the same. The rest of the study organized as follows. At the first part of this research, the dataset was cleaned from insignificant and irrelevant data points. Thirty main product groups were selected for further analyses. The changes of sales levels between 2019 to 2020 were calculated for each month. Cluster analysis were applied to understand whether these thirty product groups were showing similar shopping patterns. In the second part of the research, the main aim was to understand if there is a Covid effect on the shopping behaviors or not. For this purpose, the research used Stepwise Regression Analysis, the Best Subset Model and Lasso. The efficiency of these regression models was evaluated and compared using 10-fold cross-validation as well as considering their Cp and Adjusted R2 values. A new dataset was created with clusters that been made in first part. Covid case numbers, cluster sales levels and 3 most significant dates were considered to explain and create these models.

The first section gives an explanation for Cluster Analysis that have been conducted and the products groups showing similar shopping patterns are revealed. How both physical market and online market clusters acted throughout the year were provided under this section.

Second part mainly focuses on physical market results. All the models that have been applied, which are Stepwise Regression, Lasso and Best Subset Model, for each cluster and the analysis were provided under this section.

Third section includes the same structure and methods as second section has but for online market. For the last section, the comparison between physical and online market results were provided. The research concludes with a Conclusion and Further Suggestions part.

2. Literature Review

Although the pandemic started in 2019 and is still continuing in 2021, there are limited sources and research that points out the effect of Covid-19 on consumer behavior and shopping habits. To understand the changing shopping habits and consumer behavior, initially the disruption that occurred in the supply chain due to pandemic should be investigated. Since the supply chain starts with producers and ends with customers, any kind of disruption in the process affects the final step which is customer. Covid-19 showed all industries their weaknesses of inability to react on time to large-scale disruptions. Resilience gained importance and lots of industries are still trying to be more resilient to unexpected situations as they are facing right now. Resilience as a meaning, is the ability of resisting to disruptions and recovering the performance. Supply chains that are mostly affected from lack of resilience, had seen in life sciences, health care and food industries (Simchi-Levi & Simchi-Levi, 2020). Most of the companies needed to implement new sustainability strategies faster than they had expected. Whether there is still uncertainty and concerns, there is transitions into more sustainable supply chains. Sustainability strategies like buying local or building community trust contributes to company's own supply chain resilience. It became important to overcome this crisis and its risk with sustainability opportunity and be more resilient. Risk responses and crisis management techniques of companies to reduce the risks and be more resilient led them to transform themselves by using sustainability (Sarkis, 2021). Supply-chains were divided into two groups in this crisis. Some of them faced with an extreme demand which they are not able to supply and the others faced with an extreme decrease both in demand and supply which led them their productions to stop. Many of the companies dealt with the danger of bankruptcy and did not get any governmental support in this period (Ivanov, 2020). Food industry is one of the most affected industries from this sudden increase in demand. Most of the consumer goods were not easily reachable as before during this crisis. Demand for food still continues to increase even people have sufficient supplies because of the extreme financial conditions. If supply chains became inefficient and continues to be disrupted shortages will be expected for food industry (Sarkis, Cohen, Dewick & Schröder, 2020).

Shopping habits and consumers started to change in March 2020. Retailing industry were affected from increasing online orders because of restrictions even stores remained open in this period. As the cases increased, most of the people were scared to shop in physical market conditions and this resulted an extreme increase in online orders and retailers started to focus on their online services more (Park, Brumberg & Yonezawa, 2020). In many of the natural disasters and crisis, panic-buying, stock-pilling and hoarding behaviors has been seen all over the world. After the pandemic started, world started to see examples of this behavior from people in UK, Italy and Australia. Most of the people stock-pilled food, medicine and toilet papers and emptied the shelves of local stores (Chen, Rajabifard, Sabri, Potts & Laylavi, 2020). Same behavior also seen in US. Stock-pilling and panic-buying made stores to out-of-stock for toilet papers and sanitizers continued with grains, pasta, canned products and lots of other different food groups (Park et al., 2020). Stock-pilling occurs as an individual response to scarcity derived from stress,fear and panic environment. People started to buy or order more than they need and stock them in case of any circumstances (Micalizzi, Zambrotta & Bernstein, 2021). Commodity theory claims that scarcity may explain the behavior of stock-pilling (Brock, 1968). Prospect theory also suggest that this behavior is connected with risk aversion . According to prospect theory, risk aversion motivates this behavior if food sources seen risky by consumer, although the possibility of scarcity is very low (Tversky & Kahneman, 1992). As a result, this behavior should not be considered as completely irrational since it is human nature. On the contrary, stock-pilling as a social behavior caused negative effects on supply chains by disrupting them and creating shortages for others (Micalizzi et al., 2021). It has serious negative effects on the economy and society. Supply-chains are open to any disruptions during disasters. Depending to preparedness, these disruptions may last for several cycles. Along with the disruptions that disasters caused, stock-pilling creates more complex environment for inventory management. As mentioned before, panic-buying led people to empty shelves. When people could not meet their demand before restocking, they are eager to look for substitute products. This behavior affects the whole structure and creates double-sided problems for both the retailers and consumers. At the beginning of the pandemic several examples of this behavior seen in UK and Australia. When people did not find toilet papers, they bought baby nappies, kitchen towels even though these products are not substitute products (Chen et al., 2020). The ones who needed baby nappies did not found them in many stores because of irrational panic buying. Despite having disruptions on supply chains, panic buying also affects the social life by creating a chaos environment and scarcity for others. Since the elderly, disabled and working people cannot shop in regular times, they are much more affected in this panic-buying chaos. When reaching a product becomes hard, the value of a product also starts to increase. Another negative effect of stock-pilling is it creates a competition environment that leads prices to increase (Chen et al., 2020). Empty shelves, problems in restocking and extreme crowds in local stores led people to shop online. Most of the retailers were not ready for increasing demand in online market resulted as 3-4 days delays on deliveries. Late deliveries in online markets, empty shelves in local stores, increasing prices in competitors led people to spend much more money in this period. The consumers who could not found the product that they are looking for had to go to other more distant stores (Chen et al., 2020).

After a year from Covid started, even though people not stock-pilling right now, the effects of stock-pilling still show itself. Many of the stores are still trying to deal with disruptions that they had faced. Supply-chains are looking for permanent solutions for upcoming threats. Governments working on preventing panic-buying behavior in case of possible disasters. Online-market options increasing day by day. Many of the current retailers started to launch their own apps which allows people to order easily. They provide more product options compared to their physical stores by establishing warehouses dedicated solely on online orders.

3. Data set, Methodology and Applications

This project focuses on real case study based on a big retailer in Turkey during Covid-19 . It covers the transactions experienced on the product portfolio during March 2020 until December 2020. To extract knowledge from a mass data to solve real life problems requires a systematical process. The Cross Industry Standard Process for Data Mining (CRISP-DM) is one of these processes that providing a structure when solving data analytical problems. The CRISP-DM model can be seen in Figure 3.1 . The model consists of 6 steps, namely; Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment. Overall model with iterative steps is mainly used for data exploration. (Shearer, 2000)

Figure 3.1 CRISP-DM Model (Shearer, 2000)

The process starts with Business Understanding. It is important to understand the problem that will be solved. Business problems might be vague and unclear to solve usually. Understanding and naming the main problem is an important process to design a solution. To reach the optimal solution, it is important to go through these steps several times as needed. In this research, these iterations also done several times until understanding the real problem that will be solved. There were several options or different methodologies to be focused on. This research mainly focused on 'How consumer behavior changed in retail industry during Covid-19 process' by looking at physical market and online market sales data of a big retailer in Turkey for the period March 2020 to December 2020. This topic was one of the major problems that businesses encountered.

Data Understanding is important step for to understand if the dataset is suitable for the problem that will be solved. There might be limitations on data that may prevent to find a solution to the problem. If there are any limitations the process can iterate to the first step again to find out a new problem that can be solved with the dataset. It is important to understand the structure of the problem and the data to be able to work with data mining techniques afterwards. This step can iterate between first step several times until finding the optimal solution and problem. Before starting this research, several meetings were done with the retailer company. Data warehouse and reporting manager and also one of the data science researcher were guide this research about what they need, their current problem and the data that they are collecting every day from each store. After discussing several different problems and research topics, the research topic was decided to be Covid effect on changing consumer habits. If the dataset they are collecting was not sufficient enough to do this research, the topic must be changed. The analyst of this research and data science researcher of the retailing company decided on which information will be needed in this research. Before they send the dataset, the data analysts of the retailing company explained everything that the dataset consists. All of the columns and single elements under those columns explained through several meetings. As was mentioned before, without understanding the data it would be very challenging to solve business problems. It is important to work and communicate with the businesses for this step.

Data preparation step moves along with data understanding step. To understand the dataset better some manipulations and cleaning can be done. This step involves conversions of the dataset, removing missing or unnecessary values, normalization and scaling. Making the dataset ready to analyze depending on the models that will be used is a very important step for reaching the solution. In this research, the data preparation step done by several times according to model that used.

Modelling step is where the data analysis techniques are applied. Depending on the model that will be used, this step might made analysts to make some changes and conversions on the dataset. This research used three different datasets according to models that have been applied. First, the research used the original dataset that the retailer provided and cleaned it to create the second dataset. After the cleaning process 30 main group were decided for further analysis. Second dataset derived from the first set for cluster analysis. It contains percentage change in sales from January 2019 to December 2020 for each group and market format on monthly basis. Third dataset were created to run regression analysis. It was done after the cluster analysis and consists sales level information of clusters, along with Covid-19 case numbers and 3 dummy dates.

Evaluation of the models is an important step which gave researchers a confidence to move along with the research that they are conducting if the results are valid and fitted for the main problem. It is important that the results are satisfying the business problem. Even if the model provides accurate results, evaluation in the business context is equally important. If there is any inconsistency between the results and the main problem, the process can start over to reach an optimal solution. In this research, all the models that have been applied and their results are evaluated statistically as well by considering the main business problem. Luckily, the models that had been applied gives an understanding of the business problem the retailer investigated is experiencing the consumers' changing shopping habits during Covid-19 period. Based on empirical findings, it is possible to see that consumer behavior shifted during Covid-19 process and number of Covid cases are one of the reasons of this change

The last step is Deployment. In this step the results of data analysis put into real life. Implementing a business model or information system is deployment examples. From this research, results or models can be implement into real life problems and help other retailers and researchers to take an example the result (Provost & Fawcett, 2017).

3.1 The Dataset

The data set that had been used and examined throughout this research was retrieved from a large retail company in Turkey. There are over 5 million rows of data for each month from 2019 to 2020. From the data set that have been used in this paper, it is possible to see the daily transactions throughout the country with sales units and prices. To understand the shopping habits before and after Covid, the shopping data of both 2019 and 2020 are analyzed together. As unit prices changed a lot during these two years, the paper focused on the quantities sold for to understand the habit change regardless of price. To get rid of the effects of different measurements on products such as kg, gram , liter and ml, the paper used the quantity changes between years of investigated month. The dataset mainly consist of 18 columns. Those are as follows:

Table 3.1 Variables of the retailer data

Period :	Gives an information about which month and year (Ex:201901)
Date:	Shows specific dates for each month
City Code:	The government stated city codes from 01-81
City Name:	All city names in Turkey
Format Code:	Specified format code of large retailer
Format Name:	Specified format name of large retailer
Degree Code:	Specified degree code of large retailer. Much detailed version of format code
Degree Name:	Specified degree name of large retailer. Much detailed version of format name
Main Group Number:	Specified group numbers of large retailer
Main Group Name:	Most general grouped version of single products
Sub Group Number:	Specified sub group numbers of large retailer
Sub Group Name:	Grouped version of main classes
Main Class Number:	Specified main class number of large retailer
Main Class Name:	More generalized classed version of sub class products
Sub Class Number:	Specified class number of a single product
Sub Class Name:	Contains each product name in details
Sales Price:	Sales price for a product
Quantities Sold:	The quantities sold for a product

For data understanding step, 9 meetings are conducted with Data warehouse& Reporting Manager and Data Science Researcher of the retailing company through July 2020 to December 2020. It was important to match the dataset to the business problem. This period mainly passed by evaluating the alternative problems and deciding which one is most important among them along with evaluating the dataset that they are collecting throughout the years. (Details and examples of the data set can be found in the appendix section.)

3.2 Data Preparation

In the initial phase of the analysis, the data set were cleaned and unnecessary information were eliminated. As CRISP process suggested, it is important to prepare the data according to business problem. In that manner, the research only considered the main consumption products that can show the consumer behavior. Besides that product groups all other insignificant groups were cleared from the dataset. Such as, expenses that company made for their stores or boxes, bottles and shelves that currently using in stores. For cleaning process, basic R commands were used (James, Witten, Hastie & Tibshirani, 2013). Main group items were consisting Turkish characters which R could not process. After deciding which main groups to be eliminated, their related code numbers took into consideration and rows that are consisting unnecessary information cleaned according to these code numbers. As the main aim was to investigate the effect of Covid-19's effect on consumer behavior the main groups were decreased to 30 main group after the cleaning process. At the end of the cleaning process the main groups were as follows;

Table 3.2 30 Main Groups after cleaning process

Alcoholic Beverages	Non-Alcoholic Beverages	Fish&Seafood	Chips&Snacks	Cleaning	Frozen Goods
Cookies,Chocolate&Candy	Electronics	Imported Fruit	Meat-Deli	House Care	Takeouts
Paper&Baby	Meat	Beauty	Seasonal Products	Cheese	Textile
Canned,Beverages&Breakfast	Poultry	Toys,Pet Care&Media	Dairy Products	Tobacco	Eggs
Grains, Pasta&Sides	Packaged Meat	Produce :Vegetables&Fruits	Bread&Bakery	Olive Oil&Butters	Glassware

The same main group cleaning process is applied to online-market dataset. However, as the government regulations do not allow to sell Alcoholic beverages and Tobacco products through online channels, there are total of 28 main groups in online version of this dataset. The subsequent analyses were done based on these 30 for physical and 28 for online main group elements.

3.3 Methods

In order to decide on the methods that will be used in this research, the business problem that retailers went through Covid-19 process was considered. In the 9 meetings held, it was determined that the research questions of the company managers were as follows:

1-Which product groups had similar purchases during the Covid -19 period?

2-How has Covid-19 affected customer buying habits?

3-Are there any shifts from physical market to online market?

In order to answer these research questions, initially cluster analysis is conducted to group the products showing similar shopping behavior. Subsequently, different regression methods are used and their efficiency are compared.

Cluster Analysis: To understand which product groups acted together in Covid-19

process. Which product groups are selling together and behaving in the same manner.
Stepwise Regression Analysis: To see if the clusters that found in previous method are affected by the Covid-19 case numbers from March to end of December. This method allowing to understand that which product groups affected the most from case numbers during this period. Also it is important to see the difference between online market and physical market conditions.
Lasso: Similarly to Stepwise Regression, the research also focused on Lasso regression aiming to find a better model that explains the Covid effect.
Best Subset Model Selection: Even if the two methodologies above gave accurate results, the evaluation process of a business problems needs a satisfying model that explains the business problem. Best subset selection also done to make sure the research shows a best an optimal solution for the business problem.
To calculate the efficiency of the models, 10-fold cross validation was applied and the results were evaluated considering Cp, Adjusted R squared and Cross-Validation test error rate.

3.3.1 Cluster Analysis

In order to find the product groups showing similar purchase behavior a cluster analysis is conducted. Hierarchical clustering was used in this part of the research. Contrary to K-means, Hierarchical clustering does not limit the number of groups and grouping them according to their similarity. (James et al., 2013) The result of hierarchical clustering can be seen through a graph called dendrogram. Dendrograms are tree-shaped diagrams that consisting of nodes. The nodes consist of initial data and it groups upwards to the top. The arrows between these nodes represents the distance between individual nodes. This distance is very important when deciding on clusters. Hierarchical clustering aims to group individual data points that have the least distance between them. Dendrogram mainly gives the best result at the earlier fusions that have been investigated. Observations that fused close to the top will be show less similarity. For that manner, a threshold should be set according to fusions. Hierarchical Clustering does not tell where to cut the tree or how many clusters there are. The dendrogram should be cut at a given height to partition the data (Kassambra, 2013). Limiting the threshold to 10, gave the best groups for this research. Neither too similar nor too distant groups were created according to their distance. Also, the number of clusters was not either too many to handle nor too less to be insignificant.

A different data set was created to run the clustering algorithm. For this purpose, the difference in sales levels from 2019 to 2020 for each 6 different market formats that large retailer has and total sales levels were considered. As stated above there were 30 main groups that have been investigated during this process. In order to prepare the data for clustering, a pivot table analysis is conducted. The pivot tables consist of 30 main groups as rows and 6 different store types as well as their sum as columns and filled the numbers with sales numbers for the sale levels for 24 months from January 2019 to December 2020 were created

In order to find the change in the consumption of the products , to find out the change compared to 2019, the percentage change for each month were calculated.(Example of this tables provided in Appendix section). Then, to run the clustering analysis all 12 month changes consisting 30 groups and 6 different stores and total number of sales were merged as single table. Since the main aim is to understand which product groups acted together and consumed more or less during Covid process, Hierarchical Clustering were applied by looking at their similarities. In this case the similarity is the percentage change in sales numbers compared to 2019.

However, this dataset generation methodology could only be applied to the physical market. The reason of not applying this method to online market is that online market structure is not the same with physical market. The large retailer was provided the orders only from its large scaled markets in 2019. There are no data points for small sized markets for online orders in 2019. However, when the orders started to increase in 2020 after Covid-19 cases started to increase as well, they started to send orders to small-sized stores and there are data points for that store formats in 2020. Since this research consider the change in sales compared to 2019, when there is no data point in small-sized formats, the analysis shows insignificant change results. Taking only the total change in months for online-market creates 12 columns to 30 rows dataset which gives poor result of clustering analysis as a result.

Since, the aim is comparing the physical market to online market and see the Covid effect, for further analysis this research used physical market clusters to understand the online market. However, using the same clusters that were revealed from physical market, even a clustering analysis did not work on online market, gave the similar result for online market too. It can be seen that the same cluster groups also behaved the same way in online market dataset. As a result, similar trends observed in online market gave a possibility to compare physical and online market for same clusters for further parts of this research.

3.3.2 Stepwise Regression Analysis

After running the cluster analysis, different regression models were applied to understand the Covid effect on clusters much more efficiently. The 3 questions that have been asked for this analysis were as follows;
-> Is there a relationship between covid case numbers and sales levels of a specific cluster?
-> How strong is the relationship between covid case numbers and sales levels of a specific cluster?
->Which cluster is affected the most in 2020 from Covid-19?

Initially, Stepwise regression which is the combination of backward and forward model is used. Stepwise starts as a blank model with no variable in it. When it goes forward it adds variables that fits the model most. P-values can be increase when new variables entered the model until a threshold point. After reached at its threshold that variable will be removed from the model. This process continues forward and backwards, adding and removing until all variables have significantly lower p-values. For this research, stepwise model had been used and evaluated. (James et al., 2013)
A new dataset were created for to run stepwise regression. The data set for regression analysis were created as follows;

Table 3.3 Variables that have been used in regression analysis

Date:	Starting from March 2020 at the end of December 2020
LnCaseNum:	Natural logarithm of case numbers
LnCluster(A):	Natural logarithm of total sales numbers for each cluster
LnCluster(A_1):	Natural logarithm of total sales numbers 1 day before the original date
LnCluster(A_2):	Natural logarithm of total sales numbers 2 day before the original date
LnCluster(A_3):	Natural logarithm of total sales numbers 3 day before the original date
LnCluster(A_4):	Natural logarithm of total sales numbers 4 day before the original date
LnCluster(A_5):	Natural logarithm of total sales numbers 5 day before the original date
LnCluster(A_6):	Natural logarithm of total sales numbers 6 day before the original date
LnCluster(A_7):	Natural logarithm of total sales numbers 7 day before the original date
DUM24_05:	Dummy variable for the 24 of May because there is a significant decrease of sales for every cluster
DUM01_06:	Dummy variable for 1st of June because there is a significant increase on that day
DUM12_12	Dummy variable for 12th of Decembers because there is a significant increase on that day.

Stepwise Regression Analysis is made by using R-studio. The package called 'olsrr' was used to conduct stepwise regression (Hebbali, 2020). This package is useful when conducting Ordinary Least Squares regression models, variable selection and model fit assessments. Logarithmic transformation is applied for all of the variables. Case numbers and Cluster data were not in normal distribution. In some dates, mainly because of the restrictions that had been applied to supermarkets, sales numbers were too low or too high that affecting the overall performance of the dataset. Also, the case numbers showed a highly skewed performance. To overcome this issue, all of the variables related with case numbers and cluster sales numbers normalized. This process applied for all clusters in both physical and online market dataset. Dummy variables were created for 24-25-26th of May, 1st of June and 12th of December. These dummy dates were added after noticing that there is a significant change on those days in sales numbers. These 7 variables have the same values with LnCluster(A) but follows the values with 1-7 days later. These were added to capture the dynamic relationship between case numbers and sales numbers change for each day. (Details about the dataset prepared for regression can be found in the Appendix section)

3.3.3 Lasso

The predictive accuracy of stepwise regression is analyzed by comparing its results with Lasso and best fit model. In general, Lasso regression provides a better prediction method and an alternative model for research. The logic of the Lasso is, it shrinks all the coefficient to zero and even exactly zero when the alpha gets higher. Since Lasso also use variable selection, it gives better and simpler model for interpretation better than both linear and ridge regression (James et al., 2013).(Details related to Lasso equation can be found in the appendix)

3.3.4 Best Subset Model Selection

Best subset model selection allows to calculate all possible models. For this research it is important to find the best fitting model that explains the relationship between Covid cases and sales shifts. Best subset model selection fits separate least square regressions for all possible combination of the predictors. In this research, its 11 predictors to explain cluster sales numbers. It starts creating models with each predictor individually and continues with two, three, four... It creates 2^p models and the aim is finding the best subset model that explains the best among others (James et al., 2013). (The detailed results of the best subset selection model are given in the appendix)

3.3.5 Comparison of the regression models

In order to analyze the efficiency of the regression models cross-validation analysis were applied. This approach randomly divided the dataset into approximately equal 10 groups. The first fold set aside as validation set and the model implemented on the other 9 groups. This process continued 10 times by making one different group validation set. The Mean Squared Error were calculated at each step for validation set. Again, this process repeated 10 times until all MSE values for each validation set found. The average of these MSE values provides Cross Validation Test Error value. With that result, it can be easily understood that if the model is good or not. The lower the test error rate, the better model fits. (James et al., 2013) Additionally, the regression model's results are also compared according to the Cp, Adjusted R2 values along with the 10-fold cross-validation test error rate. The best model were provided for each cluster in both market type.

4. Results & Evaluation

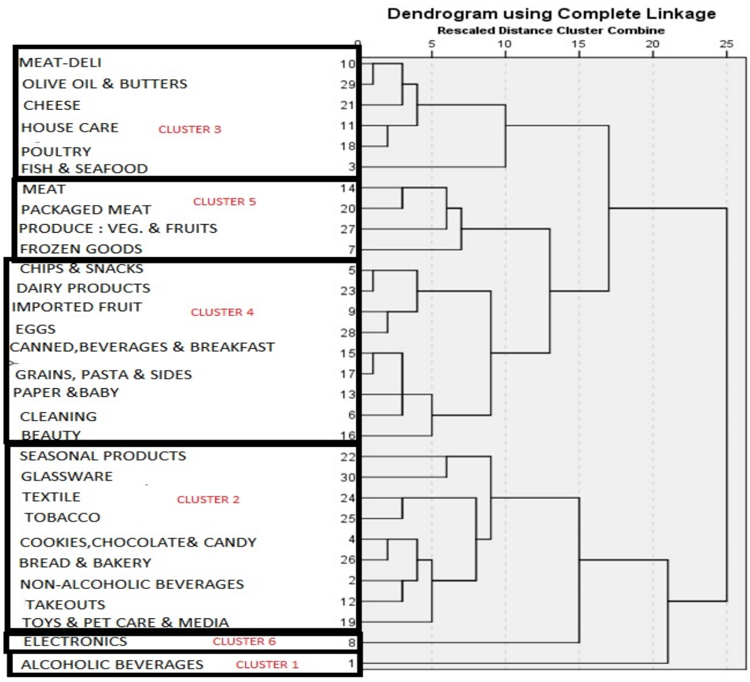
Empirical findings are given in four parts. Initially, Cluster Analysis results which have been obtained using SPSS were provided The resulting clusters are given "names" according to their general characteristics. Afterwards, all of the clusters were added in graph format for both physical and online market to show the relationship and trend among variables in clusters. The second part provides regression analysis results conducted with different models , namely stepwise regression, Lasso and best-subset, and the prediction accuracy of them are compared in order to select the best model for each cluster for the physical market. The third part gives the same analysis for online market. The last section provides information about the comparisons between physical and online market. For each cluster, their best models compared according to their Cp, Adjusted R squared and CV test error values.

4.1 Cluster Analysis

Hierarchical Clustering results that are provided below is conducted using by SPSS (IBM, 2016). One of the problems with Hierarchical Clustering is that the researcher had to define a threshold point, a height, to define the clusters (Kassambra, 2013). The threshold point was set at 10 for this research. Neither too similar nor too distant groups were created according to their distance. Also, the number of clusters was not either too many to handle nor too less to be insignificant. If the threshold point were set at 5 it was going to create 14 different set of groups mostly with ungrouped single items. To decrease the number of groups to an optimal point, the threshold was set at 10. With this threshold the total number of clusters is 6 where 2 of them has single item. Also, the Data Warehouse and Reporting Center Manager of the large retailer stated that they were mostly explain 'Alcoholic Beverages' and 'Electronics' as single items on their research. The Data Warehouse & Reporting

Center Manager and Data Science Researcher that have been discussed with believed that the purchasing behavior of the alcoholic beverages show differences with respect to the size of the town. In fact, they believed that this behaviour is different in big town when compared to small towns. For the small towns in Turkey, this retailer mostly is the only place that sells Alcoholic Beverages. The research conducted in this thesis also supported their claim. These two items acting differently from other items.

Figure 4.1 Hierarchical Clustering Result based on Physical Market



The cluster results can be found below. The clusters are named based on their general qualifications.

Table 4.1 Cluster Results

ALCOHOLIC BEVERAGES	DAILY NEEDS	PROTEIN-BASED	BASIC CONSUMPTION GOODS	COLD CHAIN PRODUCTS	ELECTRONICS
ALCOHOLIC BEVERAGES	SEASONAL PRODUCT	MEAT-DELI	CHIPS /& SNACKS	MEAT	ELECTRONICS
	GLASSWARE	OLIVE-OIL /& BUTTERS	DAIRY PRODUCTS	PACKAGED MEAT	
	TEXTILE	CHEESE	IMPORTED FRUIT	PRODUCE : VEGETABLES /& FRUIT	
	TOBACCO	HOUSECARE	EGGS	FROZEN GOODS	
	COOKIES,CHOCOLATE & CANDY	POULTRY	CANNED,BEVERAGES/& BREAKFAST		
	BREAD & BAKERY	FISH & SEAFOOD	GRAINS,PASTA/& SIDES		
	NON-ALCOHOLIC BEVERAGES		PAPER /& BABY		
	TAKEOUTS		CLEANING		
	TOYS, PET CARE & MEDIA		BEAUTY		
1	**9**	**6**	**9**	**4**	**1**

4.1.1 Alcoholic Beverages

According to research results, the Alcoholic Beverages saw a significant decrease March 2020. It was the month that first Covid case seen and people were started to be getting anxious and serious about this disease. With that period most of the doctors and scientist were warning people about self-care and hazardous effect of alcoholic beverages and tobacco on human body. Following months after the first shock the consumption increased to its regular levels and with the home restrictions on May 2020 it reached its peak for the first wave of Covid. After the normalization period started in June 2020 the sales mostly stayed at the normal levels. However, we can see that after November, the sales enormously increased at the second wave period of Turkey. There is no information related to online market for this product group because it is forbidden to sell Alcoholic Beverages by government

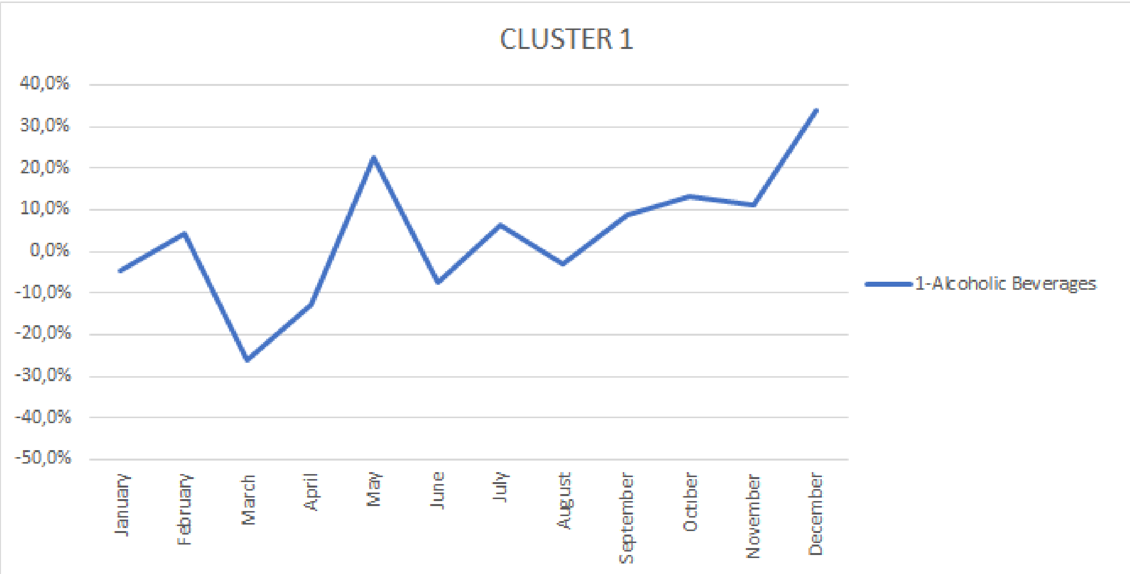


Figure 4.2 Alcoholic Beverages January-December trends on physical market

.

4.1.2 Daily Needs

Cluster 2 elements shows the same behavior from January to December 2020. Whether there are differences at their level of increase, when one of them increases the other elements has an increase for the next month. The same rule applies for online market too. From the physical market side, there is a significant decrease in April for all elements in this cluster. This may be due to the fact that people started to bake their own goods at home much often compared to 2019 because of the restrictions. Since July 2020, it can be seen that these levels did not turn into normal spending. After the normalization period it can be seen that people started to spend more. Whether it is still lower compared to 2019, we can say that they started to bake less and buy more. In July 2020, Seasonal Products have a significant increase. Normally, this is an expected increase. However, even with the Covid cases increasing day by day, people again preferred to buy pool supplies, sea toys or products related with vacation. At the time of second wave, between October to December, it can be seen that the numbers stay at the higher levels compare to first shock. This suggest that people getting used to live with the idea of this virus.

When we look at the online market trends, it can be seen that decrease in April 2020 for physical market is not the case for online market. This suggest that people preferred to buy mostly from online store instead of physical markets. It can be seen that there is a significant increase in Bread &Bakery and Cookies, Chocolate & Candy in April. At the beginning of the year 2020 we can see that there is an increase in preference of online market. When we look at the end of the year, it can be seen that whether it decreases compared to 1st wave, people still prefers to shop online.

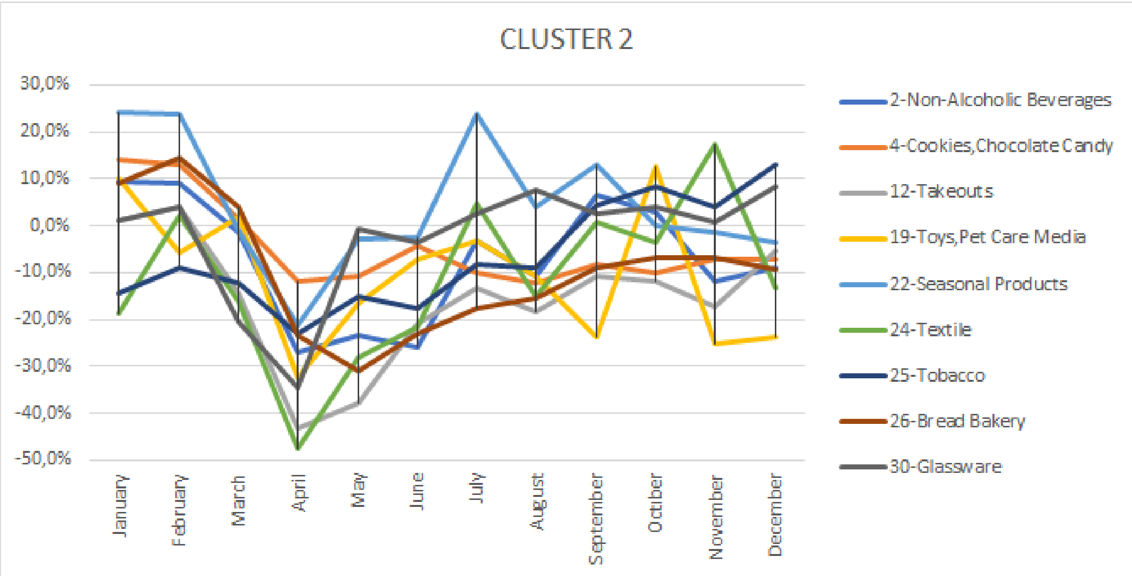


Figure 4.3 Daily Needs January-December trends on physical market

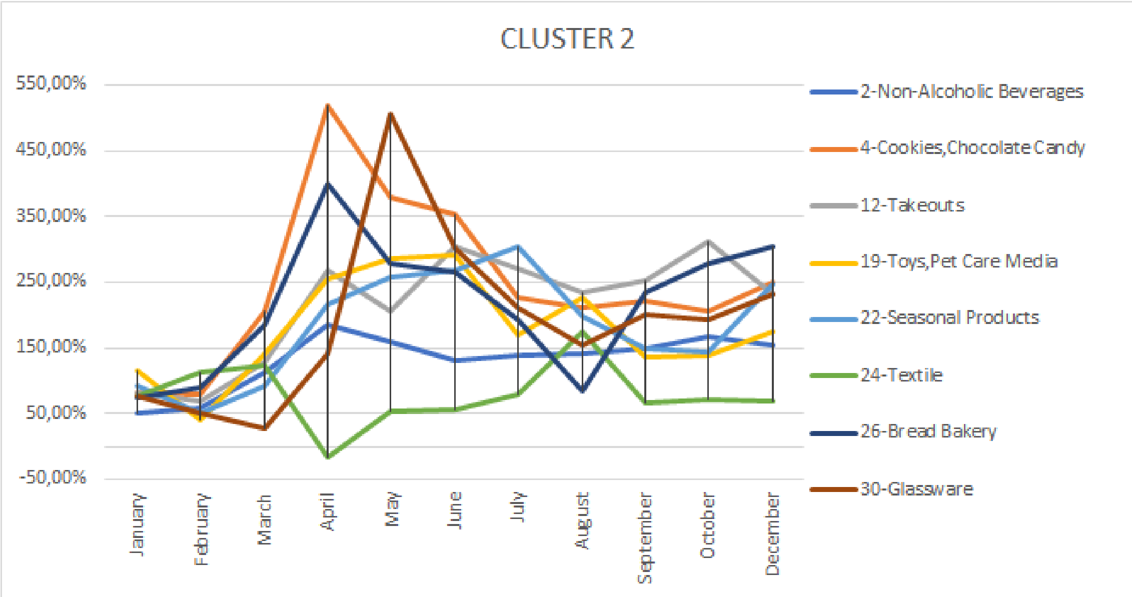


Figure 4.4 Daily Needs January-December trends on online market

4.1.3 Protein-Based Products

The third cluster also shows similar trends from January to December. For physical market there is an enormous change in House-care products in March 2020. House care products also contains elements similar to Cleaning. With the first cases started to show up in Turkey, people started to be stock-pilling of house-care products on their home in case of emergency. It can be seen that the other elements stayed mostly at the same levels till the end of the year. The most changed in this period is House care products.

For the online side, again we see similar behavior on these 6 groups. At the beginning of Covid period, there is an increase similar to other clusters. Compared to physical market it can be seen that people mostly preferred to shop online. Instead of house-care fish & seafood category sees an extreme increase with respect to others.

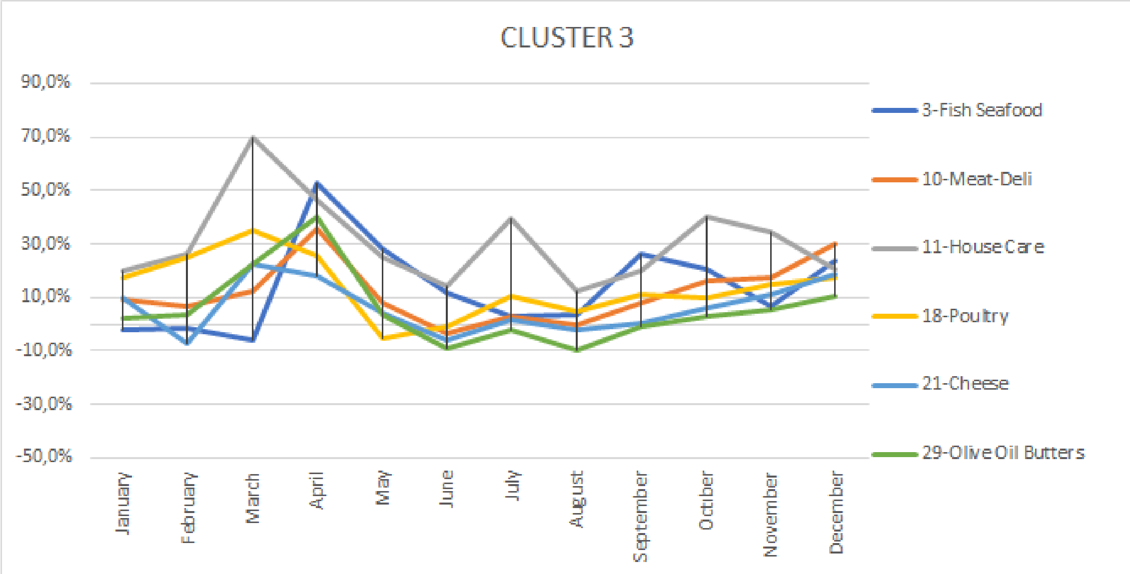


Figure 4.5 Protein-Based Products January-December trends on physical market

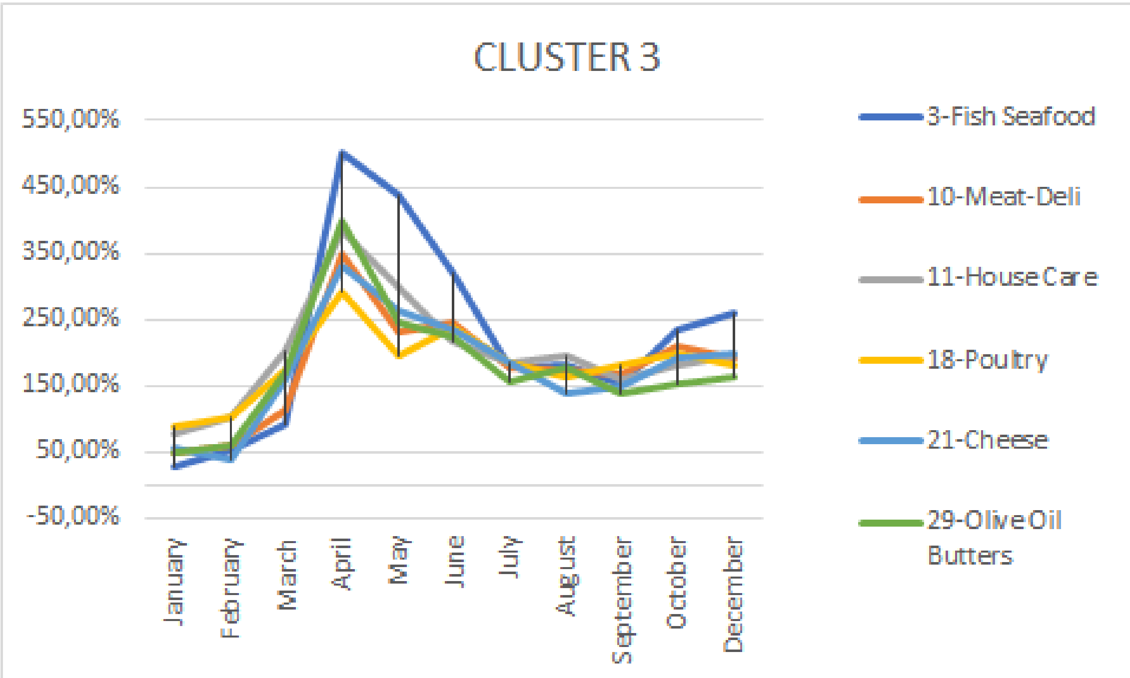


Figure 4.6 Protein-Based Products January-December trends on online market

4.1.4 Basic Consumption Goods

With the effect of the first cases, we can see that the most affected elements are "Grains,Pasta and sides","Paper & Baby products" and "Cleaning supplies". These 3 categories also were the most affected categories throughout the world. People started to stock-pilling of these categories in case of emergency and sanitary purposes. Most of the markets were out-of-stock in this period for these groups of products. People could not find diapers or toilet papers for weeks. Markets on the other hand, could not met the needs and filled their shelves as a result of this sudden shock.

For the online market, compared to physical we see that most preferred groups are not the most popular ones in the 1st wave period. Instead of "Grains,Pasta and sides","Paper & Baby products" and "Cleaning supplies", Chips & Snacks has a sudden increase. Interestingly, Grains, Pasta & sides sees an extreme decrease in August.

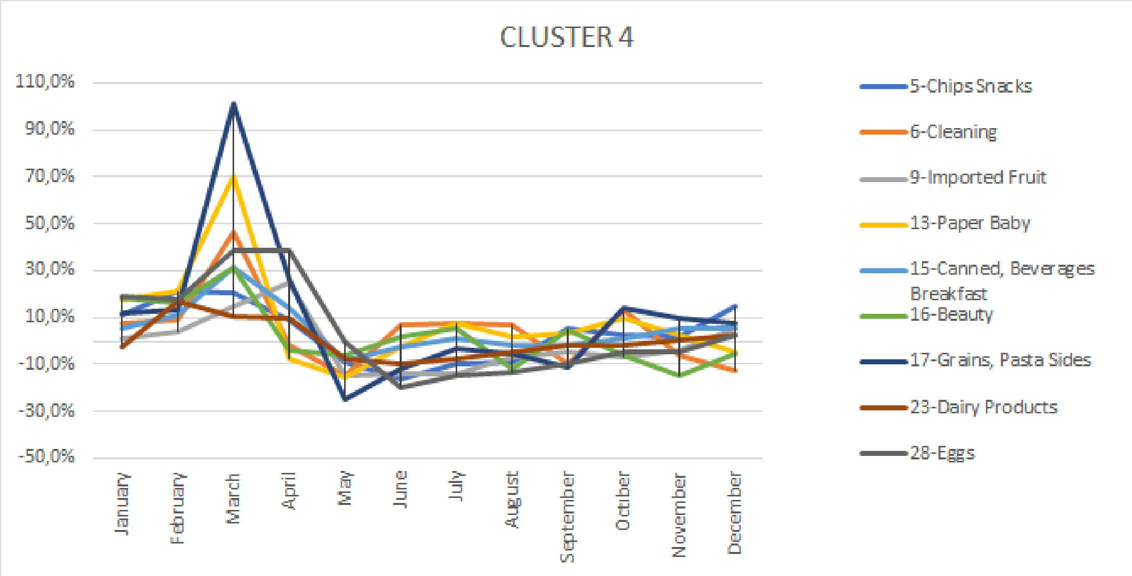


Figure 4.7 Basic Consumption Goods January-December trends on physical market

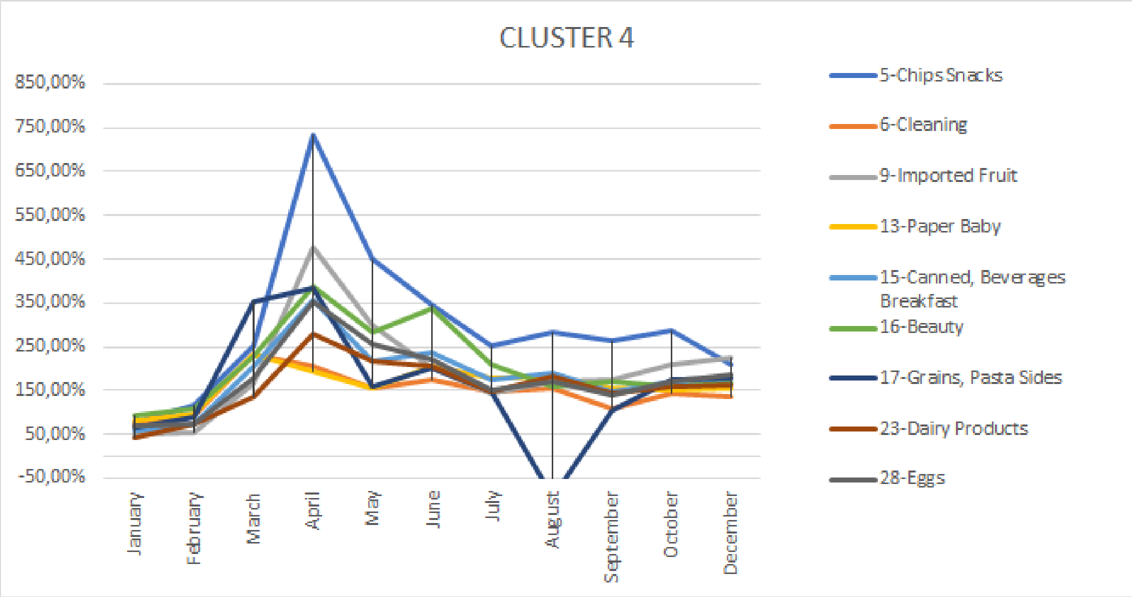


Figure 4.8 Basic Consumption Goods January-December trends on online market

4.1.5 Cold-Chain Products

This cluster shows that, after the first shock till normalization period(June 2020-September 2020) this group of products stayed at the normal levels. However, with the beginning of normalization period the sales were decreased. The reason behind this decrease is that people started to go and eat outside in this period. So, started to cook less in their home and with this way spend less on these products. When the second wave started, we see that there is again significant demand on these products. Since the number of cases had increased a lot and winter came, people started to spend much more time at home and cook more often.

For online market, 4 groups are acting in a similar way to the physical market. Again, the first shock of Covid can be seen from the graph.

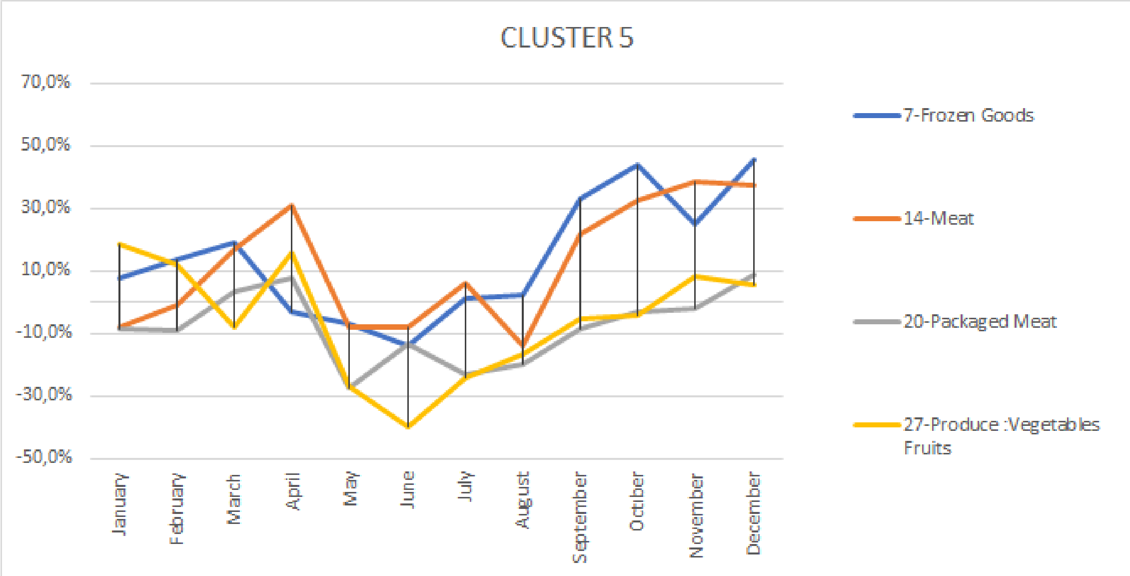


Figure 4.9 Cold-Chain Products January-December trends on physical market

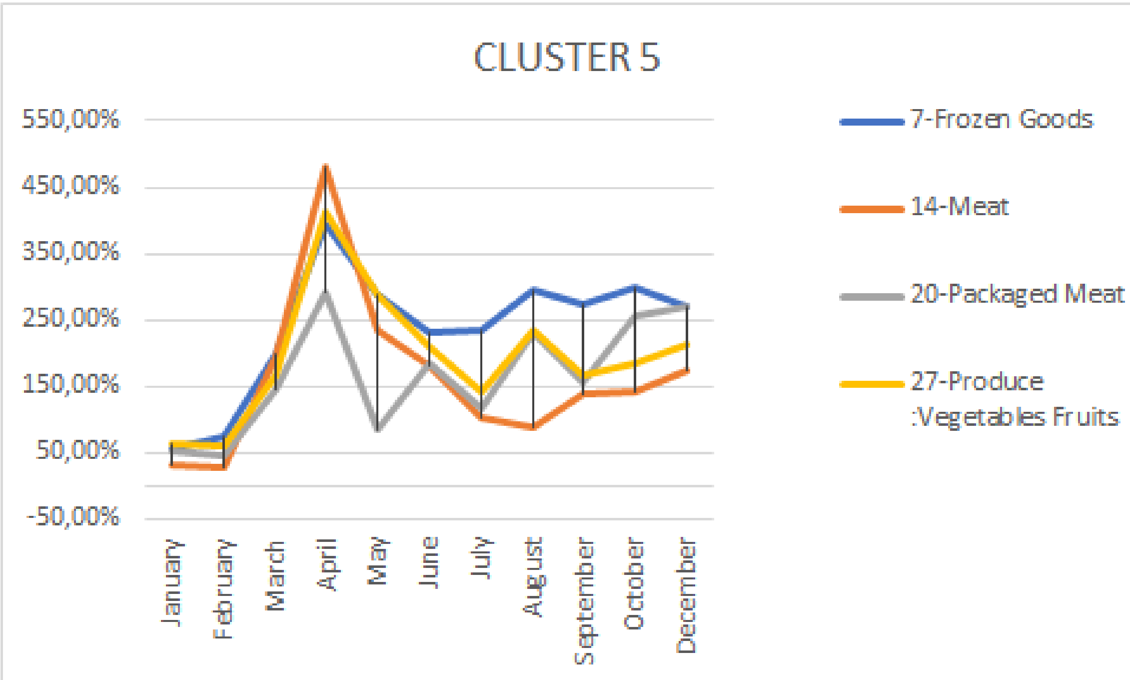


Figure 4.10 Cold-Chain Products January-December trends on online market

4.1.6 Electronics

There is a limited access to these products. The variety of the electronics equipment mostly differs for different format sizes of the retailer. It can be seen that there is a significant decrease in this product type. Compared to first three months of 2020, the sales of electronics could not reach its normal levels till the end of the year. However, there is a significant increase in June 2020. Before June 2020, people were restricted at their home. This led most of the people to renovate their home or change the electronic products that they own for better ones in between April and May 2020. With the beginning of the normalization period in June 2020, this increase may be the result of this need of change. However, since the retailer mostly focused on food and beverages and there are electronic shops that giving more options, even with that demand large retailer could not reached its normal levels that we see in January and February.

For the online market, despite not having all products in all market types, the application provides all electronic products that they are selling. As stated in previous part, large retailer started to send their orders from small-sized markets of their own after Covid period started. This method gave an opportunity to the customers to reach different electronics much more easily instead of finding them in physical stores. It can be seen that there is again a significant increase in June for electronics. Compared to physical market we can see that online market ,even for electronics, preferred much more in this period.

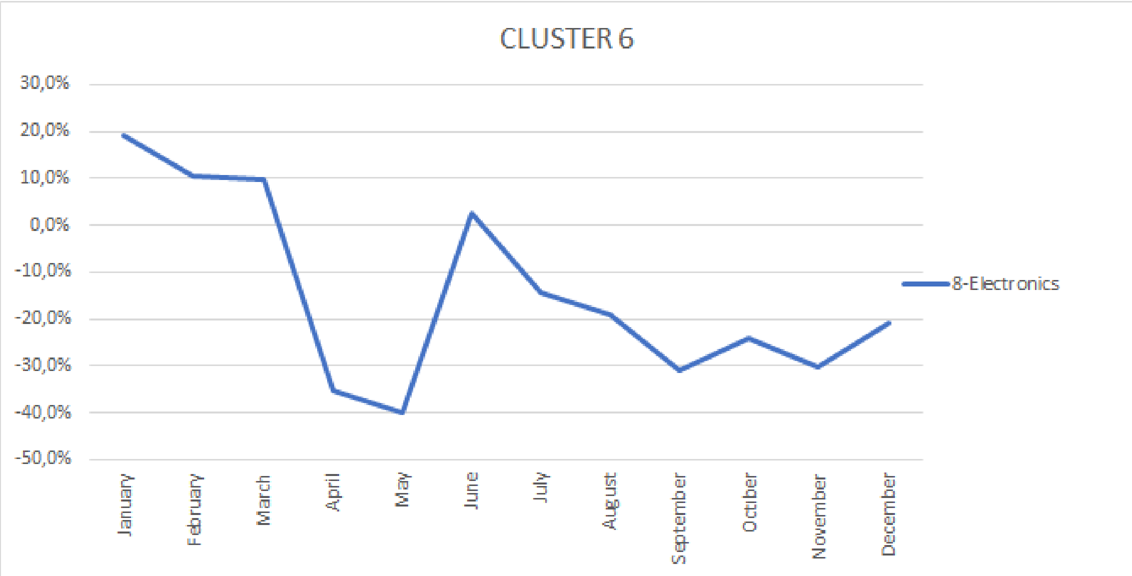


Figure 4.11 Electronics January-December trends on physical market

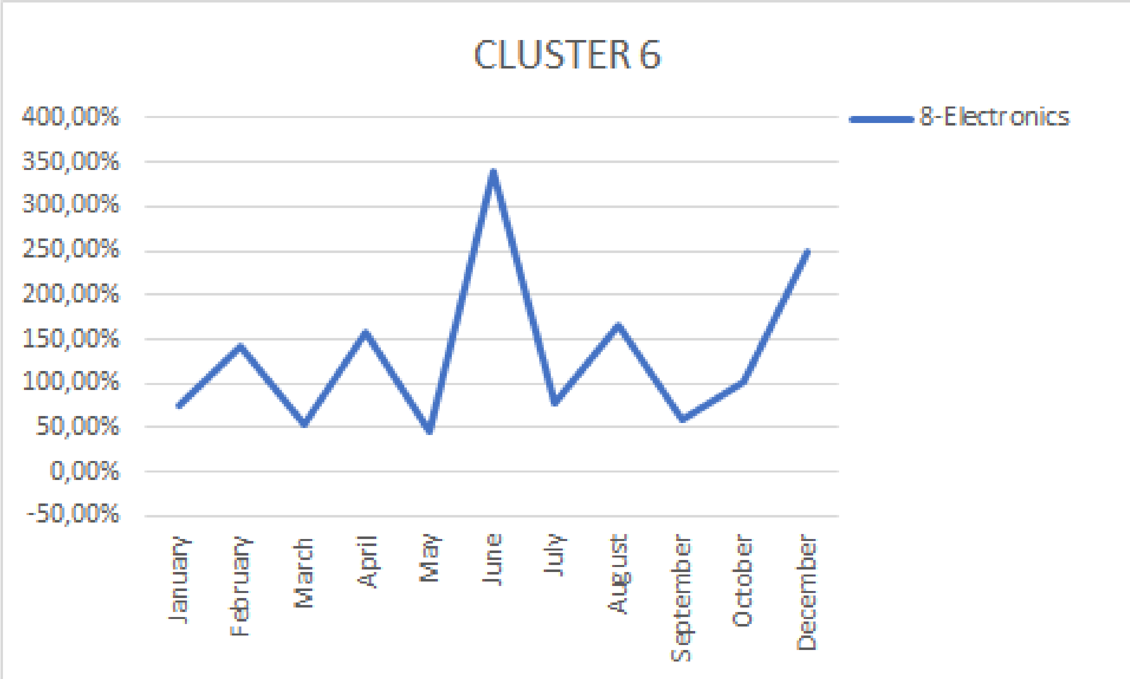


Figure 4.12 Electronics January-December trends on online market

4.2 Physical Market Results

Physical Market Results sectioned into 6 parts. Below parts will provide results of three regression models that had been applied for each cluster separately. These clusters were examined with three different models. Stepwise regression, Lasso regression and Best Subset Model Selection. These three models were evaluated by 3 performance metrics which are R-squared, Cp and Cross-Validation Test error. R squared value shows how close the data to the fitted regression line. The higher the r-squared, the model better explains the variability of the data around its mean. Mallow's Cp should be close to the number of predictors plus constant to be considered as unbiased and a good estimation on coefficients. When comparing these three models, test error rate mainly took into consideration. The final model coefficients were provided under 'Comparison between Physical and Online Market' section for each cluster.

4.2.1 Alcoholic Beverages

Table 4.2 shows the results of these three models on Alcoholic Beverages cluster. As can be seen from the table, the model that the lowest test error rate is Best Subset Selection. R-squared value is greater than the other models which suggests a better explanation of the dataset. Cp value of best subset regression suggest that the model uses all 11 predictors and the constant. As a result , CV Test Error value is smaller and R-squared, Cp values are higher than the other models which is statistically significant to claim that the best model is Best Subset Selection.

Table 4.2 Alcoholic Beverages Model Results

Alcoholic Beverages	**R-Squared**	**Cp**	**Cv Test Error**
Stepwise	0.825	11	2.317
Lasso	0.824	10	2.108
Best Subset	0.826	12	**0.944**

4.2.2 Daily Needs

Table 4.3 shows the results of these three models on Daily Needs cluster. As can be seen from the table, the model that the lowest test error rate is Best Subset Selection. R-squared value is greater than the other models which suggests a better explanation of the dataset. Cp value of 8.75 considered as 9 predictors. As a result, Cp, R-squared and CV test error rate gives a statistically significant and better result than the other models. Best Subset Model was selected for the final and significant model among others.

Table 4.3 Daily Needs Model Results

Daily Needs	**R-Squared**	**Cp**	**Cv Test Error**
Stepwise	0.728	9	0.138
Lasso	0.713	8	0.135
Best Subset	0.735	8.75	0.100

4.2.3 Protein-Based Products

Table 4.4 shows the results of these three models on Protein-Based Products cluster. As can be seen from the table, the model that the lowest test error rate is Best Subset Selection. R-squared value is greater than the other models which suggests a better explanation of the dataset. Cp value is giving the best result in Stepwise Regression. However, Stepwise regression results as high test error rate compared to Best Subset Selection. As a result, Cp, R-squared and CV test error rate gives a statistically significant and better result than the other models. Best Subset Model was selected for the final and significant model among others.

Table 4.4 Protein-Based Products Model Results

Protein-Based Products	**R-Squared**	**Cp**	**Cv Test Error**
Stepwise	0.772	10	0.215
Lasso	0.738	6	0.190
Best Subset	0.773	8.31	0.142

4.2.4 Basic Consumption Goods

Table 4.5 shows the results of these three models on Protein-Based Products cluster. As can be seen from the table, the model that the lowest test error rate is Best Subset Selection. R-squared value is greater than the other models which suggests a better explanation of the dataset. Cp value is giving the best result in Stepwise Regression. However, Stepwise regression results as high test error rate compared to Best Subset Selection. As a result, Cp, R-squared and CV test error rate gives a statistically significant and better result than the other models. Best Subset Model was selected for the final and significant model among others.

Table 4.5 Basic Consumption Goods Model Results

Basic Consumption Goods	**R-Squared**	**Cp**	**Cv Test Error**
Stepwise	0.753	10	0.184
Lasso	0.726	7	0.167
Best Subset	0.753	8.24	0.128

4.2.5 Cold-Chain Products

Table 4.6 shows the results of these three models on Protein-Based Products cluster. As can be seen from the table, the model that the lowest test error rate is Best Subset Selection. R-squared value is greater than the other models which suggests a better explanation of the dataset. Cp value is giving the best result in Stepwise Regression. However, Stepwise regression results as high test error rate compared to Best Subset Selection. As a result, Cp, R-squared and CV test error rate gives a statistically significant and better result than the other models. Best Subset Model was selected for the final and significant model among others.

Table 4.6 Cold Chain Products Model Results

Cold-Chain Products	**R-Squared**	**Cp**	**Cv Test Error**
Stepwise	0.764	10	0.207
Lasso	0.729	7	0.179
Best Subset	0.764	8.27	0.135

4.2.6 Electronics

Table 4.7 shows the results of these three models on Electronics cluster. As can be seen from the table, the model that the lowest test error rate is Lasso. However, R-squared value is also slightly lower than the other models. However, when three models compared T values of Lasso resulted statistically significant than the others. As a result, Lasso model was chosen as final model because having significantly lower value of CV test error rate compared to other models.

Table 4.7 Electronics Model Results

Electronics	**R-Squared**	**Cp**	**Cv Test Error**
Stepwise	0.666	7	0.199
Lasso	0.664	7	0.162
Best Subset	0.673	12	0.165

4.3 Online Shopping Results

Online Shopping Results sectioned into 5 parts same as physical market results. Since Online Market dataset did not have any information for Alcoholic Beverages cluster, that cluster had to be eliminated for this part of this research. Below parts will again provide results of three regression models that had been applied for each cluster separately. The final model coefficients were provided under 'Comparison between Physical and Online Market' section for each cluster.

4.3.1 Daily Needs

For this cluster, Best Subset Model has the highest R-squared value among other models. Cp value is exactly 12 in Best Subset as wanted. There is a significant difference in CV test error rate among three models. Best Subset model provides the lowest Cv test error rate among others. When considering the lowest CV test error and highest R-squared and Cp value , the final model for this cluster is Best Subset Model.

Table 4.8 Daily Needs Model Results

Daily Needs	**R-Squared**	**Cp**	**Cv Test Error**
Stepwise	0.886	10	0.495
Lasso	0.883	7	0.264
Best Subset	0.886	12.00	0.171

4.3.2 Protein-Based Products

For Protein-Based products, the CV test error of the Best Subset Model is the lowest one. Cp value also statistically the best value. Whether R-squared value is the highest in Stepwise Regression, stepwise regression resulted as the highest test error rate compared to other models. The best result for this cluster achieved by using Best Subset Model.

Table 4.9 Protein-Based Products Model Results

Protein-Based Products	**R-Squared**	**Cp**	**Cv Test Error**
Stepwise	0.886	10	0.440
Lasso	0.844	7	0.251
Best Subset	0.867	12	0.188

4.3.3 Basic Consumption Goods

Basic consumption goods again have the lowest CV test error rate in Best Subset Model. Additionally, R-squared value is higher compared to Lasso. Cp value is 10.04 which is closer to 12. The results are similar with Stepwise regression. Since the main consideration when deciding the final model for this research is CV test error rate, the final model decided as Best Subset Model.

Table 4.10 Basic Consumption Goods Model Results

Basic Consumption Goods	**R-Squared**	**Cp**	**Cv Test Error**
Stepwise	0.870	10	0.608
Lasso	0.776	7	0.358
Best Subset	0.870	10.04	0.210

4.3.4 Cold-Chain Products

Cold-Chain Products again has the lowest CV test error rate in Best Subset Model. Additionally, R-squared value is higher than the other models. Cp value is 10.30 which is closer to 12. The results are similar with Stepwise regression but stepwise regression resulted as a very high test error rate. The final model decided as Best Subset Model.

Table 4.11 Cold-Chain Products Model Results

Cold-Chain Products	R-Squared	Cp	Cv Test Error
Stepwise	0.880	10	0.410
Lasso	0.865	7	0.225
Best Subset	0.8815	10.30	0.177

4.3.5 Electronics

Compared to Physical Market Electronics result, online market dataset gave the best result in Best Subset Model by having a highest value of R-squared, significant Cp value and lowest CV test error rate. Comparing R-squared, Cp and most importantly Cv test error rate with other models, Best Subset Model provides a much more statistically significant result.

Table 4.12 Electronics Model Results

Electronics	R-Squared	Cp	Cv Test Error
Stepwise	0.656	8	0.489
Lasso	0.654	10	0.540
Best Subset	0.659	10.00	0.409

4.4 Comparison between Physical and Online Market

This section compares the physical and online market results for each cluster with their final model that had been selected. Related coefficients of the final models, T-values and P-values will provide an insight about Covid-19 effect on sales for each cluster. As comparing physical and online market models the research aimed to see the shift between physical to online market during Covid-19 process. 4 main predictors took into consideration for comparison. Namely; LnCasenum(Natural Logarithm of number of cases between March 2020 to December 2020), DUM24_5(24th of May-Holiday-significant change on sales levels) , DUM01_06(1st of June 2020-Beginning of the normalization period and significant change on sales levels) and lastly DUM12_12 (significant change on sales levels). The Comparison part sectioned into 6 parts according to clusters.

4.4.1 Alcoholic Beverages

It can be seen that Case numbers has no specific effect on sales of Alcoholic Beverages. The coefficient of LnCasenum statistically low. However, when looking at the T-value and P-value of Case numbers, the effect of case numbers on sales levels cannot be denied. On the other hand, all three dummy dates have a significant negative effect on sales numbers. It can be said that during these days the sales levels were statistically low than usual. T-values and P-values also proving that claim. Since there is no data for online market in this cluster. The results were only provided for physical market.

Table 4.13 Alcoholic Beverages-Physical Market Final Model Result

Alcoholic Beverages	**Predictors**	**Coefficients**	**T values**	**P values**
Physical	(Intercept)	3.94	5.11	0.00
	LnCasenum	**-0.08**	**-2.37**	**0.01**
	Lnclust1_1	0.36	10.55	0.00
	Lnclust1_2	-0.24	-6.43	0.00
	Lnclust1_3	0.05	1.42	0.15
	Lnclust1_4	0.07	1.82	0.06
	Lnclust1_5	-0.23	-5.62	0.00
	Lnclust1_6	0.23	5.33	0.00
	Lnclust1_7	0.48	13.02	0.00
	DUM24_05	**-10.03**	**-16.78**	**0.00**
	DUM12_12	**-12.51**	**-13.01**	**0.00**
	DUM01_06	**9.36**	**8.64**	**0.00**

4.4.2 Daily Needs

For physical market, it can be seen that Case numbers has no specific effect on sales of Daily Needs. The coefficient of LnCasenum statistically low. However, when looking at the T-value and P-value of Case numbers, the effect of case numbers on sales levels cannot be denied. On the other hand, DUM24_05 had a negative effect on sales levels. It can be said that during these days the sales levels were statistically low than usual. On the contrary, DUM01_06 has a positive effect on sales levels.T-values and P-values of these two dates also proving that claim.
On the online market, it can be seen that Case numbers has positive effect compared to physical market on sales of Daily Needs. The coefficient of LnCasenum statistically low. However, when looking at the T-value and P-value of Case numbers, the effect of case numbers on sales levels cannot be denied. On the other hand, DUM24_05 had a significant negative effect on sales levels. It can be said that during these days the sales levels were statistically low than usual. On the contrary, DUM01_06 has a significant positive effect on sales levels.T-values and P-values of these two dates also proving that claim.
Online market results proving the claim of a shift from physical to online market for this cluster by suggesting a higher value of coefficients, T-values and better P-values.

Table 4.14 Daily Needs-Physical Market Final Model Result

Daily Needs	**Predictors**	**Coefficients**	**T values**	**P values**
Physical	(Intercept)	4.22000	4.505	0.00
	LnCasenum	**-0.01629**	**-1.442**	**0.15**
	Lnclust2_1	0.29661	7.023	0.00
	Lnclust2_2	-0.18145	-4.044	0.00
	Lnclust2_3	0.03032	0.660	0.50
	Lnclust2_4	0.08434	1.822	0.06 .
	Lnclust2_5	-0.13263	-2.813	0.00
	Lnclust2_6	0.14605	3.125	0.00
	Lnclust2_7	0.48271	12.090	0.00
	DUM24_5	**-2.87985**	**-14.239**	**0.00**
	DUM12_12	**-0.18367**	**-0.580**	**0.56**
	DUM1_6	**2.51723**	**7.141**	**0.00**

Table 4.15 Daily Needs-Online Market Final Model Result

Daily Needs	Predictors	Coefficients	T values	P values
Online	(Intercept)	8.28224	15.766	0.00
	LnCasenum	**0.07459**	**4.714**	**0.00**
	Lnclust2_1	0.22889	6.395	0.00
	Lnclust2_2	-0.37918	-10.280	0.00
	Lnclust2_3	0.17105	4.375	0.00
	Lnclust2_4	0.01915	0.484	0.62
	Lnclust2_5	-0.16515	-4.177	0.00
	Lnclust2_6	0.13807	3.719	0.00
	Lnclust2_7	0.25181	7.871	0.00
	DUM24_5	**-9.71935**	**-28.824**	**0.00**
	DUM12_12	**0.37904**	**0.928**	**0.35**
	DUM1_6	**4.42341**	**8.404**	**0.00**

4.4.3 Protein-Based Products

For physical market, it can be seen that case numbers have no specific effect on sales of Protein-Based Products. The coefficient of LnCasenum statistically low. However, when looking at the T-value and P-value of Case numbers, the effect of case numbers on sales levels cannot be denied. On the other hand, DUM24_05 had a negative effect on sales levels. It can be said that during these days the sales levels were statistically low than usual. On the contrary, DUM01_06 has a positive effect on sales levels.T-values and P-values of these two dates also proving that claim.

On the online market, it can be seen that case numbers have similar but positive effect compared to physical market on sales of Protein-Based Products. The coefficient of LnCasenum statistically low. However, when looking at the T-value and P-value of Case numbers, the effect of case numbers on sales levels cannot be denied. DUM24_05 had a significant negative effect on sales levels. It can be said that during these days the sales levels were statistically low than usual. DUM01_06 has a significant positive effect on sales levels.T-values and P-values of these two dates also proving that claim.

Online market results proving the claim of a shift from physical to online market for this cluster by suggesting a higher value of coefficients, T-values and better P-values.

Table 4.16 Protein-Based Products-Physical Market Final Model Result

Protein-Based Products	Predictors	Coefficients	T values	P values
Physical	(Intercept)	6.974998	8.298	0.00
	LnCasenum	**-0.030281**	**-2.280**	**0.02**
	Lnclust3_1	0.259949	6.491	0.00
	Lnclust3_2	-0.185242	-4.418	0.00
	Lnclust3_3	-0.001608	-0.038	0.96
	Lnclust3_4	0.112936	2.615	0.00
	Lnclust3_5	-0.203428	-4.520	0.00
	Lnclust3_6	0.164681	3.597	0.00
	Lnclust3_7	0.342701	9.359	0.00
	DUM24_5	**-5.055043**	**-19.748**	**0.00**
	DUM12_12	**-0.211549**	**-0.561**	**0.57**
	DUM1_6	**3.206452**	**6.998**	**0.00**

Table 4.17 Protein-Based Products-Online Market Final Model Result

Protein-Based Products	Predictors	Coefficients	T values	P values
Online	(Intercept)	6.88227	13.166	0.00
	LnCasenum	**0.05259**	**3.180**	**0.00**
	Lnclust3_1	0.26290	7.099	0.00
	Lnclust3_2	-0.33695	-8.758	0.00
	Lnclust3_3	0.11268	2.797	0.00
	Lnclust3_4	0.05814	1.427	0.15
	Lnclust3_5	-0.19793	-4.805	0.00
	Lnclust3_6	0.15940	4.074	0.00
	Lnclust3_7	0.28866	8.712	0.00
	DUM24_5	**-8.75795**	**-25.863**	**0.00**
	DUM12_12	**0.31999**	**0.746**	**0.45**
	DUM1_6	**5.05109**	**9.312**	**0.00**

4.4.4 Basic Consumption Goods

For physical market, it can be seen that case numbers have no specific effect on sales of Basic Consumption Goods. The coefficient of LnCasenum statistically low. However, when looking at the T-value and P-value of Case numbers, the effect of

case numbers on sales levels cannot be denied. On the other hand, DUM24_05 had a negative effect on sales levels. It can be said that during these days the sales levels were statistically low than usual. On the contrary, DUM01_06 has a positive effect on sales levels.T-values and P-values of these two dates also proving that claim.
On the online market, it can be seen that case numbers have same but positive effect compared to physical market on sales of Basic Consumption Goods. The coefficient of LnCasenum statistically low. However, when looking at the T-value and P-value of Case numbers, the effect of case numbers on sales levels cannot be denied. DUM24_05 had a significant negative effect on sales levels. It can be said that during these days the sales levels were statistically low than usual. DUM01_06 has a significant positive effect on sales levels.T-values and P-values of these two dates also proving that claim.
Online market results proving the claim of a shift from physical to online market for this cluster by suggesting a higher value of coefficients, T-values and better P-values.

Table 4.18 Basic Consumption Goods-Physical Market Final Model Result

Basic Consumption Goods	**Predictors**	**Coefficients**	**T values**	**P values**
Physical	(Intercept)	7.111703	7.254	0.00
	LnCasenum	**-0.038233**	**-2.976**	**0.00**
	Lnclust4_1	0.266468	6.466	0.00
	Lnclust4_2	-0.184126	-4.268	0.00
	Lnclust4_3	-0.006281	-0.143	0.88
	Lnclust4_4	0.102171	2.294	0.02
	Lnclust4_5	-0.183306	-3.990	0.00
	Lnclust4_6	0.157594	3.403	0.00
	Lnclust4_7	0.383103	10.068	0.00
	DUM24_5	**-4.200313**	**-17.624**	**0.00**
	DUM12_12	**-0.168354**	**-0.470**	**0.63**
	DUM1_6	**2.994115**	**7.069**	**0.00**

Table 4.19 Basic Consumption Goods-Online Market Final Model Result

Basic Consumption Goods	**Predictors**	**Coefficients**	**T values**	**P values**
Online	(Intercept)	8.52205	14.637	0.00
	LnCasenum	**0.03354**	**1.999**	**0.00**
	Lnclust4_1	0.25291	6.687	0.00
	Lnclust4_2	-0.39198	-9.964	0.00
	Lnclust4_3	0.18274	4.374	0.00
	Lnclust4_4	0.00865	0.205	0.83
	Lnclust4_5	-0.15094	-3.588	0.00
	Lnclust4_6	0.13106	3.322	0.00
	Lnclust4_7	0.26859	7.857	0.00
	DUM24_5	**-9.89079**	**-26.478**	**0.00**
	DUM12_12	**0.35507**	**0.781**	**0.43**
	DUM1_6	**4.82857**	**8.332**	**0.00**

4.4.5 Cold-Chain Products

For physical market, it can be seen that case numbers have no specific effect on sales of Cold-Chain Products. The coefficient of LnCasenum statistically low. However, when looking at the T-value and P-value of Case numbers, the effect of case numbers on sales levels cannot be denied. On the other hand, DUM24_05 had a negative effect on sales levels. It can be said that during these days the sales levels were statistically low than usual. On the contrary, DUM01_06 has a positive effect on sales levels.T-values and P-values of these two dates also proving that claim.
On the online market, it can be seen that case numbers have some more higher and positive effect compared to physical market on sales of Cold-Chain Products. The coefficient of LnCasenum statistically low. However, when looking at the T-value and P-value of Case numbers, the effect of case numbers on sales levels cannot be denied. DUM24_05 had a significant negative effect on sales levels. It can be said that during these days the sales levels were statistically low than usual. DUM01_06 has a significant positive effect on sales levels.T-values and P-values of these two dates also proving that claim.
Online market results proving the claim of a shift from physical to online market for this cluster by suggesting a higher value of coefficients, T-values and better P-values.

Table 4.20 Cold-Chain Products-Physical Market Final Model Result

Cold-Chain Products	Predictors	Coefficients	T values	P values
Physical	(Intercept)	6.591133	7.413	0.00
	LnCasenum	**-0.030444**	**-2.338**	**0.02**
	Lnclust5_1	0.269870	6.781	0.00
	Lnclust5_2	-0.183991	-4.398	0.00
	Lnclust5_3	-0.004747	-0.111	0.91
	Lnclust5_4	0.114220	2.647	0.00
	Lnclust5_5	-0.204178	-4.550	0.00
	Lnclust5_6	0.163787	3.600	0.00
	Lnclust5_7	0.371445	10.094	0.00
	DUM24_5	**-4.709145**	**-19.175**	**0.00**
	DUM12_12	**-0.189945**	**-0.515**	**0.60**
	DUM1_6	**3.399490**	**7.662**	**0.00**

Table 4.21 Cold-Chain Products-Online Market Final Model Result

Cold-Chain Products	Predictors	Coefficients	T values	P values
Online	(Intercept)	7.81728	14.878	0.00
	LnCasenum	**0.06015**	**3.726**	**0.00**
	Lnclust5_1	0.23830	6.833	0.00
	Lnclust5_2	-0.32138	-8.954	0.00
	Lnclust5_3	0.09473	2.528	0.00
	Lnclust5_4	0.05906	1.553	0.12
	Lnclust5_5	-0.21260	-5.489	0.00
	Lnclust5_6	0.16577	4.482	0.00
	Lnclust5_7	0.25960	8.457	0.00
	DUM24_5	**-9.40371**	**-28.451**	**0.00**
	DUM12_12	**0.22916**	**0.549**	**0.58**
	DUM1_6	**5.11184**	**9.507**	**0.00**

4.4.6 Electronics

For physical market, it can be seen that case numbers are not add to the Lasso model. This suggest that case numbers had no significant effect on sales levels so it did not added to the final model. On the other hand, DUM24_05 had a negative effect on sales levels. It can be said that during these days the sales levels were statistically low than usual. On the contrary, DUM01_06 has a positive effect on sales levels.T-values and P-values of these two dates also proving that claim.

On the online market, it can be seen that case numbers have some effect on sales of Electronics. This believed that online market providing much more electronics product options. The coefficient of LnCasenum statistically low. However, when looking at the T-value and P-value of Case numbers, the effect of case numbers on sales levels cannot be denied. DUM24_05 had a significant negative effect on sales levels. It can be said that during these days the sales levels were statistically low than usual. DUM01_06 has a significant positive effect on sales levels.T-values and P-values of these two dates also proving that claim.

Online market results proving the claim of a shift from physical to online market for this cluster by suggesting a higher value of coefficients, T-values and better P-values.

Table 4.22 Electronics-Physical Market Final Model Result

Electronics	Predictors	Coefficients	T values	P values
Physical	(Intercept)	4.19186	7.357	0.00
	Lnclust6_1	0.36515	7.712	0.00
	Lnclust6_2	-0.15418	-3.687	0.00
	Lnclust6_6	0.06268	1.469	0.14
	Lnclust6_7	0.30210	6.873	0.00
	DUM24_5	**-3.62236**	**-13.624**	**0.00**
	DUM1_6	**2.06296**	**4.599**	**0.00**

Table 4.23 Electronics-Online Market Final Model Result

Electronics	Predictors	Coefficients	T values	P values
Online	(Intercept)	1.527323	4.071	0.00
	LnCasenum	**0.087545**	**3.378**	**0.00**
	Lnclust6_1	0.433464	8.686	0.00
	Lnclust6_2	-0.183326	-3.417	0.00
	Lnclust6_3	0.002154	0.040	0.96
	Lnclust6_4	0.011608	0.214	0.83
	Lnclust6_5	-0.083583	-1.514	0.13
	Lnclust6_6	0.144355	2.612	0.00
	Lnclust6_7	0.321085	6.943	0.00
	DUM24_5	**-4.160791**	**-10.224**	**0.00**
	DUM12_12	**0.589596**	**0.926**	**0.35**
	DUM1_6	**4.066505**	**5.759**	**0.00**

5. Conclusion and Further Suggestions

Initially, this research aimed to find an answer to following questions.
Which product groups experienced a significant shopping behavior change compared to 2019?
With the cluster analysis had been done, it can be clearly seen that all of the clusters had experienced a significant change compared to 2019. Most demanded product groups in both physical and online market were Basic Consumption Goods(which are grains& pasta and sides, cleaning and paper products). Online market sales increased for all product groups compared to 2019. Even after the curfews started, physical market sales were still significantly higher than 2019 for all product groups. Is there any difference between online shopping and physical market sales compared to 2019?
All of the clusters both in physical and online market seen a sudden increase in sales at the beginning of the Covid-19 period. This increase suggesting stock-pilling and panic-buying behavior in both shopping types. At the end of 2020, it can be seen that these panic behaviors turning into normal buying behaviors since the sales numbers starts to be decrease after months passed. Online-market still has a significantly higher sales ratio. That is believed that people are starting to be getting used to shop from online as online market provides more variety of products and easy. How has the Covid-19 affected customer buying habits?
As with the regression analysis were done on this research, it can be seen that Covid-19 case numbers had an effect on shopping habits. Also, there is a significant effect of curfews on sales numbers. The dummy dates that had been created after noticing a significant change on 25th of May, 1st of June and 12th of December showed that these dates affected the sales numbers a lot. Limitations, curfews and case numbers made people to shift from physical market to online market. People started to be preferred to shop online rather than physical market as the case numbers and curfews increased.

To answer the questions above, this research used CRISP-DM process. The dataset was provided from one of the largest retailers in Turkey and the topic were decided to solve one of the business problems that they are encountering during this process. Data analysis techniques were applied such as cleaning dataset, clustering the product groups, creating models and calculating the efficiency of the models to answer this business problem. Implementing a business model or information system from this research can be beneficial for businesses to overcome their similar problems related with changing shopping habits during Covid-19. With the methods or results derived from this research, companies can evaluate their own customers, act according to result and overcome their problems much faster. Deployment is an important step for companies to overcome their business problems by implementing the results that had been found like in this research. This research will also be a useful guideline for other countries which are still dealing with the changing consumer behaviors and its responses on demand. Since the dataset is only limited to one retailer, adding more retailers to the models that had been implemented will definitely change the results. Whether the consumer behavior mostly changed similar all around the world, adding much more information on the methods that had been applied will give much more statistically significant results. So, this topic will be open to any further examinations to understand how the consumer behavior affected from COVID-19 process.

BIBLIOGRAPHY

Brock, T. C. (1968). Implications of commodity theory for value change. *Physchological foundations of attitudes*, 243–275.

Chen, Y., Rajabifard, A., Sabri, S., Potts, K., & Laylavi, F. (2020). A discussion of irrational stockpiling behaviour during crisis. *Journal of Safety Science and Resilience*.

Hall, M., Prayag, G., Fieger, P., & Dyason, D. (2020). Beyond panic buying: consumption displacement and covid-19. *Journal of Service Management*, *32*(1), 113–128.

Hebbali, A. (2020). Olsrr: Tools for building ols regression models, https://olsrr.rsquaredacademy.com/index.html.

IBM (2016). Hierarchical cluster analysis. *IBM Documentation*, https://www.ibm.com/docs/en/spss–statistics/24.0.0?topic=option–hierarchical–cluster–analysis.

Ivanov, D. (2020). Viable supply chain model: integrating agility, resilience and sustainability perspectives—lessons from and thinking beyond the covid-19 pandemic. *Annals of Operations Research*.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. New York: Springer.

Kassambra, A. (2013). *Practical Guide To Cluster Analysis in R*. Sebastopol: STHDA.

Lund, S., Manyika, J., Woetzel, J., Barribal, E., Krishnan, M., Alicke, K., Birshnan, M., George, K., Smit, S., Swan, D., & Hutzler, K. (2020). Risk, resilience, and rebalancing in global value chains. *McKinsey Global Institute*.

Micalizzi, L., Zambrotta, N., & Bernstein, M. (2021). Stockpiling in the time of covid-19. *British Journal of Health Psychology*.

Owid (2020). Covid-19 data explorer. *Our world in data*, https://ourworldindata.org/explorers/coronavirus–data–explorer.

Park, K., Brumberg, A., & Yonezawa, K. (2020). The covid-19 shopper: Shopping habits during covid-19. *Department of Applied Economics and Management Cornell University Extension Bulletin*.

Provost, F. & Fawcett, T. (2017). *Data Science for Business*. O'Reilly.

Saenz, H., Stephan, J., Terino, J., Bysong, T., & Gnamm, J. (2021). How to trace a path to resilient, sustainable supply chains. *Bain Company*.

Sarkis, J. (2021). Supply chain sustainability: learning from the covid-19 pandemic. *International Journal of Operations Production Management*, *41*(1), 63–73.

Sarkis, J., Cohen, M., Dewick, P., & Schröder, P. (2020). A brave new world: Lessons from the covid-19 pandemic for transitioning to sustainable supply and production. *Resources, Conservation Recycling*.

Shearer, C. (2000). The crisp-dm model:the new blueprint for data mining. *Journal of Data Warehousing*, *5*(4), 13–22.

Simchi-Levi, D. & Simchi-Levi, E. (2020). Building resilient supply chains. *Harvard Business Review*.

Tversky, A. & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.

Who (2021). Coronavirus disease(covid-19) pandemic. *World Health Organization*, https://www.who.int/emergencies/diseases/novel–coronavirus–2019.

Wikipedia (2021). Covid-19 pandemic in turkey. *Wikipedia Free Encyclopedia*, https://en.wikipedia.org/wiki/Timeline_of_the_COVID–19_pandemic_in_Turkey.

APPENDIX A

Table A.1 Example from the dataset

PERIOD	DATE	CITY CODE	CITY NAME	MAIN GROUP NAME	SUB GROUP NAME	MAIN CLASS NAME	SUB CLASS NAME	SALES PRICE	QTY SOLD
201901	01.01.2019	01	ADANA	CLEANING	LAUNDRY DETERGENT	FABRIC SOFTENER	CONCENTRATED	53.30	3
201902	06.02.2019	06	ANKARA	GRAINS, PASTA&SIDES	SALT & SPICE	SPICE	CUMIN	11.12	2
201911	23.11.2019	34	İSTANBUL	GLASSWARE	HOUSE DECORATION	DECORATIVE PRODUCTS	FRAME	15.90	1
202004	08.04.2020	35	İZMİR	NON-ALC BEV.	CARBONATED BEVERAGE	ENERGY DRINKS	ENERGY DRINKS	18.95	1
202012	19.12.2020	09	AYDIN	PAPER & BABY	CLEANING PAPERS	TOILET PAPERS	PREMIUM TOILET PAPERS	71.60	3

Notes :NON-ALC. BEV is an acronym for Non-Alcoholic beverages. There are total of 18 columns in this dataset, which also mentioned in Dataset,Methods and Applications part. However, the company that provided this dataset wanted their company codes and format names to be confidential. That is why those columns were not added to the table above.

Table A.2 Example from the pivot table

	Jan_Hipermarket	January_1	January_Lux	January_Fast	January_2	January_3	January_Total
1-Alcoholic Beverages	-10,10%	-5,51%	6,09%	1,36%	-6,46%	-7,19%	-4,43%
2-Non-Alcoholic Beverages	-7,17%	14,28%	1,51%	23,86%	10,39%	5,48%	9,59%
3-Fish&Seafood	-9,39%	3,02%	6,94%	20,00%	-2,86%	1,40%	-2,10%
4-Cookies,Chocolate&Candy	8,21%	13,57%	7,98%	21,52%	17,37%	9,63%	14,25%
5-Chips&Snacks	1,66%	10,98%	7,70%	16,40%	15,28%	7,12%	10,98%
6-Cleaning	6,89%	11,38%	2,80%	12,14%	11,21%	-3,96%	7,24%
7-Frozen Goods	-0,52%	10,13%	4,80%	11,78%	11,73%	3,55%	7,96%
8-Electronics	21,31%	36,57%	-9,46%	43,20%	35,49%	3,82%	19,13%
9-Imported Fruit	-5,01%	0,28%	12,84%	4,77%	1,95%	-0,12%	1,17%
10-Meat-Deli	-2,54%	9,09%	9,22%	14,65%	12,55%	6,44%	8,93%
11-House Care	10,57%	26,37%	12,02%	19,22%	24,73%	12,70%	19,64%
12-Takeouts	-15,94%	17,21%	18,36%	24,07%	9,93%	-2,20%	0,96%
13-Paper&Baby	10,52%	14,99%	6,61%	34,14%	22,28%	12,75%	17,79%
14-Meat	-11,69%	-9,42%	-5,68%	-11,75%	-5,29%	-6,90%	-7,65%
15-Canned, Beverages&Breakfast	-3,66%	6,35%	2,40%	9,84%	9,94%	1,55%	5,70%
16-Beauty	21,32%	18,00%	1,58%	29,98%	17,11%	15,21%	18,29%
17-Grains, Pasta&Sides	0,99%	14,66%	15,64%	18,97%	14,64%	7,24%	12,13%
18-Poultry	11,60%	16,13%	5,47%	27,80%	19,94%	14,03%	17,43%
19-Toys,Pet Care&Media	7,38%	7,67%	2,36%	10,51%	11,60%	14,16%	10,14%
20-Packaged Meat	-24,42%	-6,25%	5,32%	5,18%	-8,21%	-18,13%	-8,42%
21-Cheese	4,34%	12,04%	5,37%	15,14%	12,96%	5,51%	10,12%
22-Seasonal Products	17,30%	25,91%	20,34%	26,56%	27,84%	21,38%	24,14%
23-Dairy Products	-10,55%	-2,26%	2,83%	5,82%	-1,03%	-7,28%	-2,52%
24-Textile	-16,31%	-23,84%	-30,19%	-37,40%	-14,68%	-12,98%	-18,65%
25-Tobacco	-17,38%	-16,38%	-10,32%	-11,28%	-14,46%	-16,86%	-14,34%
26-Bread&Bakery	0,58%	9,08%	7,79%	22,08%	12,46%	1,78%	9,13%
27-Produce :Vegetables&Fruits	15,56%	20,66%	2,75%	46,82%	19,24%	10,53%	18,65%
28-Eggs	11,04%	22,94%	3,18%	30,67%	22,72%	10,12%	19,13%
29-Olive Oil&Butters	-8,50%	4,82%	5,51%	11,78%	4,30%	-2,74%	2,50%
30-Glassware	-7,47%	11,29%	10,46%	24,95%	0,09%	-5,32%	1,08%

Table A.3 Example from Regression Dataset

	LnCasenum	LnClust3	LnClust3_1	LnClust3_2	LnClust3_3	LnClust3_4	LnClust3_5	LnClust3_6	LnClust3_7	DUM24_5	DUM12_12	DUM1_6
1.03.2020	0	13								0	0	0
2.03.2020	0	13	13							0	0	0
3.03.2020	0	13	13	13						0	0	0
4.03.2020	0	13	13	13	13					0	0	0
5.03.2020	0	13	13	13	13	13				0	0	0
6.03.2020	0	13	13	13	13	13	13			0	0	0
7.03.2020	0	13	13	13	13	13	13	13		0	0	0
8.03.2020	0	13	13	13	13	13	13	13	13	0	0	0
9.03.2020	0	13	13	13	13	13	13	13	13	0	0	0
10.03.2020	0	13	13	13	13	13	13	13	13	0	0	0
11.03.2020	1	14	13	13	13	13	13	13	13	0	0	0
12.03.2020	0	13	14	13	13	13	13	13	13	0	0	0
13.03.2020	2	14	13	14	13	13	13	13	13	0	0	0
14.03.2020	0	14	14	13	14	13	13	13	13	0	0	0
15.03.2020	1	13	14	14	13	14	13	13	13	0	0	0
16.03.2020	3	14	13	14	14	13	14	13	13	0	0	0
17.03.2020	3	14	14	13	14	14	13	14	13	0	0	0
18.03.2020	4	14	14	14	13	14	14	13	14	0	0	0
19.03.2020	5	14	14	14	14	13	14	14	13	0	0	0
20.03.2020	5	14	14	14	14	14	13	14	14	0	0	0
21.03.2020	6	14	14	14	14	14	14	13	14	0	0	0
22.03.2020	6	13	14	14	14	14	14	14	13	0	0	0
23.03.2020	6	13	13	14	14	14	14	14	14	0	0	0
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
22.05.2020	7	14	14	14	13	13	12	12	14	0	0	0
23.05.2020	7	13	14	14	14	13	13	12	12	0	0	0
24.05.2020	7	7	13	14	14	14	13	13	12	1	0	0
25.05.2020	7	7	7	13	14	14	14	13	13	1	0	0
26.05.2020	7	8	7	7	13	14	14	14	13	1	0	0
27.05.2020	7	14	8	7	7	13	14	14	14	0	0	0
28.05.2020	7	13	14	8	7	7	13	14	14	0	0	0
29.05.2020	7	14	13	14	8	7	7	13	14	0	0	0
30.05.2020	7	13	14	13	14	8	7	7	13	0	0	0
31.05.2020	7	12	13	14	13	14	8	7	7	0	0	0
1.06.2020	7	13	12	13	14	13	14	8	7	0	0	1
2.06.2020	7	13	13	12	13	14	13	14	8	0	0	0
3.06.2020	7	13	13	13	12	13	14	13	14	0	0	0
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
10.12.2020	10	13	13	13	13	12	12	14	13	0	0	0
11.12.2020	10	14	13	13	13	13	12	12	14	0	0	0
12.12.2020	10	13	14	13	13	13	13	12	12	0	1	0
13.12.2020	10	12	13	14	13	13	13	13	12	0	0	0
14.12.2020	10	13	12	13	14	13	13	13	13	0	0	0
15.12.2020	10	13	13	12	13	14	13	13	13	0	0	0
16.12.2020	10	13	13	13	12	13	14	13	13	0	0	0
17.12.2020	10	13	13	13	13	12	13	14	13	0	0	0
18.12.2020	10	14	13	13	13	13	12	13	14	0	0	0
19.12.2020	10	13	14	13	13	13	13	12	13	0	0	0
20.12.2020	10	13	13	14	13	13	13	13	12	0	0	0
21.12.2020	10	13	13	13	14	13	13	13	13	0	0	0
22.12.2020	10	13	13	13	13	14	13	13	13	0	0	0
23.12.2020	10	13	13	13	13	13	14	13	13	0	0	0
24.12.2020	10	13	13	13	13	13	13	14	13	0	0	0
25.12.2020	10	14	13	13	13	13	13	13	14	0	0	0
26.12.2020	10	13	14	13	13	13	13	13	13	0	0	0
27.12.2020	10	13	13	14	13	13	13	13	13	0	0	0
28.12.2020	10	14	13	13	14	13	13	13	13	0	0	0
29.12.2020	10	14	14	13	13	14	13	13	13	0	0	0
30.12.2020	10	14	14	14	13	13	14	13	13	0	0	0
31.12.2020	10	14	14	14	14	13	13	14	13	0	0	0

STEPWISE RESULTS

```
Final Model Output
------------------

                        Model Summary
--------------------------------------------------------------
R                         0.909       RMSE              0.955
R-Squared                 0.826       Coef. Var         7.810
Adj. R-Squared            0.819       MSE               0.912
Pred R-Squared             -Inf       MAE               0.418
--------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                                ANOVA
-----------------------------------------------------------------------
                Sum of
               Squares        DF    Mean Square       F         Sig.
-----------------------------------------------------------------------
Regression    1242.333        10        124.233    136.292     0.0000
Residual       262.520       288          0.912
Total         1504.853       298
-----------------------------------------------------------------------

                                    Parameter Estimates
------------------------------------------------------------------------------------------------
      model        Beta    Std. Error    Std. Beta        t        Sig       lower       upper
------------------------------------------------------------------------------------------------
(Intercept)       4.099         0.767                   5.346    0.000       2.590       5.608
    DUM24_5     -10.009         0.599       -0.445    -16.712    0.000     -11.187      -8.830
 LnClust1_7       0.491         0.037        0.437     13.204    0.000       0.418       0.564
   DUM12_12     -12.489         0.963       -0.321    -12.968    0.000     -14.384     -10.593
 LnClust1_1       0.353         0.033        0.352     10.579    0.000       0.287       0.418
     DUM1_6       9.520         1.079        0.245      8.820    0.000       7.396      11.645
 LnClust1_2      -0.210         0.030       -0.210     -7.043    0.000      -0.269      -0.151
 LnClust1_5      -0.252         0.039       -0.239     -6.400    0.000      -0.330      -0.175
 LnClust1_6       0.234         0.043        0.209      5.438    0.000       0.150       0.319
 LnClust1_4       0.108         0.031        0.108      3.434    0.001       0.046       0.170
 LnCasenum       -0.087         0.036       -0.063     -2.443    0.015      -0.157      -0.017
------------------------------------------------------------------------------------------------
```

Figure A.1 Alcoholic Beverages Physical Market Regression Result

```
Residuals:
     Min       1Q   Median       3Q      Max
-2.16770 -0.05375  0.02160  0.07254  1.29277

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.17061    0.87735   5.893 1.05e-08 ***
LnClust2_1   0.27887    0.04145   6.727 9.22e-11 ***
LnClust2_2  -0.14846    0.03653  -4.064 6.23e-05 ***
LnClust2_5  -0.07114    0.03807  -1.869   0.0627 .
LnClust2_6   0.11559    0.04530   2.552   0.0112 *
LnClust2_7   0.48858    0.04015  12.169  < 2e-16 ***
LnCasenum   -0.01926    0.01130  -1.704   0.0894 .
DUM24_5     -2.84101    0.20315 -13.985  < 2e-16 ***
DUM1_6       2.43109    0.35076   6.931 2.71e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3145 on 290 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:  0.7285, Adjusted R-squared:  0.721
F-statistic: 97.24 on 8 and 290 DF,  p-value: < 2.2e-16
```

Figure A.2 Daily Needs Physical Market Regression Result

```
Final Model Output
------------------

                        Model Summary
-------------------------------------------------------------
R                        0.879       RMSE              0.371
R-Squared                0.773       Coef. Var         2.820
Adj. R-Squared           0.766       MSE               0.137
Pred R-Squared            -Inf       MAE               0.208
-------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                                ANOVA
---------------------------------------------------------------------
              Sum of
              Squares        DF    Mean Square       F         Sig.
---------------------------------------------------------------------
Regression    135.074         9         15.008    109.205    0.0000
Residual       39.718       289          0.137
Total         174.792       298
---------------------------------------------------------------------

                                   Parameter Estimates
-----------------------------------------------------------------------------------------
      model        Beta    Std. Error    Std. Beta        t         Sig     lower    upper
-----------------------------------------------------------------------------------------
(Intercept)      6.970        0.833                   8.372     0.000    5.332    8.609
   DUM24_5      -5.061        0.253       -0.660    -20.005     0.000   -5.559   -4.563
LnClust3_7       0.344        0.036        0.341      9.432     0.000    0.272    0.415
    DUM1_6       3.219        0.456        0.243      7.058     0.000    2.322    4.117
LnClust3_1       0.259        0.040        0.258      6.525     0.000    0.181    0.337
LnClust3_2      -0.186        0.035       -0.185     -5.301     0.000   -0.255   -0.117
 LnCasenum      -0.031        0.013       -0.066     -2.359     0.019   -0.057   -0.005
LnClust3_5      -0.205        0.045       -0.204     -4.587     0.000   -0.293   -0.117
LnClust3_6       0.166        0.045        0.166      3.666     0.000    0.077    0.256
LnClust3_4       0.113        0.037        0.112      3.059     0.002    0.040    0.185
```

Figure A.3 Protein-Based Products Physical Market Regression Result

```
Final Model Output
------------------

                        Model Summary
-------------------------------------------------------------
R                        0.868       RMSE              0.352
R-Squared                0.753       Coef. Var         2.407
Adj. R-Squared           0.745       MSE               0.124
Pred R-Squared            -Inf       MAE               0.192
-------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                               ANOVA
--------------------------------------------------------------------
              Sum of
              Squares        DF    Mean Square       F        Sig.
--------------------------------------------------------------------
Regression    109.269         9         12.141    97.975    0.0000
Residual       35.813       289          0.124
Total         145.081       298
--------------------------------------------------------------------

                                   Parameter Estimates
-----------------------------------------------------------------------------------------
      model        Beta    Std. Error    Std. Beta        t         Sig     lower    upper
-----------------------------------------------------------------------------------------
(Intercept)      7.093        0.970                   7.313     0.000    5.184    9.002
   DUM24_5      -4.207        0.236       -0.602    -17.796     0.000   -4.672   -3.742
LnClust4_7       0.384        0.038        0.381     10.137     0.000    0.309    0.458
    DUM1_6       3.000        0.421        0.249      7.132     0.000    2.172    3.829
LnClust4_1       0.266        0.041        0.265      6.551     0.000    0.186    0.346
LnClust4_2      -0.187        0.036       -0.186     -5.221     0.000   -0.258   -0.117
 LnCasenum      -0.039        0.013       -0.091     -3.045     0.003   -0.064   -0.014
LnClust4_5      -0.184        0.045       -0.183     -4.062     0.000   -0.273   -0.095
LnClust4_6       0.159        0.046        0.158      3.461     0.001    0.069    0.250
LnClust4_4       0.099        0.038        0.099      2.641     0.009    0.025    0.173
-----------------------------------------------------------------------------------------
```

Figure A.4 Basic Consumption Goods Physical Market Regression Result

```
Final Model Output
------------------

                        Model Summary
---------------------------------------------------------------
R                       0.874       RMSE              0.363
R-Squared               0.765       Coef. Var         2.715
Adj. R-Squared          0.757       MSE               0.132
Pred R-Squared           -Inf       MAE               0.214
---------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                              ANOVA
-----------------------------------------------------------------------
               Sum of
              Squares        DF    Mean Square       F         Sig.
-----------------------------------------------------------------------
Regression    123.535         9         13.726     104.37    0.0000
Residual       38.007       289          0.132
Total         161.542       298
-----------------------------------------------------------------------

                                 Parameter Estimates
-------------------------------------------------------------------------------------------
      model      Beta    Std. Error    Std. Beta        t        Sig      lower     upper
-------------------------------------------------------------------------------------------
(Intercept)     6.563         0.878                   7.477     0.000     4.835     8.290
    DUM24_5    -4.716         0.243       -0.639    -19.407     0.000    -5.194    -4.238
LnClust5_7      0.372         0.037        0.371     10.165     0.000     0.300     0.444
     DUM1_6     3.411         0.441        0.268      7.730     0.000     2.543     4.280
LnClust5_1      0.269         0.039        0.269      6.848     0.000     0.192     0.347
LnClust5_2     -0.186         0.035       -0.186     -5.306     0.000    -0.255    -0.117
 LnCasenum     -0.031         0.013       -0.069     -2.407     0.017    -0.057    -0.006
LnClust5_5     -0.205         0.044       -0.204     -4.617     0.000    -0.292    -0.118
LnClust5_6      0.166         0.045        0.165      3.678     0.000     0.077     0.255
LnClust5_4      0.112         0.037        0.112      3.054     0.002     0.040     0.185
-------------------------------------------------------------------------------------------
```

Figure A.5 Cold-Chain Products Physical Market Regression Result

```
Residuals:
     Min       1Q   Median       3Q      Max
-2.01436 -0.14539 -0.02899  0.14513  1.29176

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.564837   0.540742   8.442 1.48e-15 ***
LnClust6_1   0.362777   0.047499   7.638 3.18e-13 ***
LnClust6_2  -0.156320   0.041945  -3.727 0.000233 ***
LnClust6_7   0.333916   0.038250   8.730  < 2e-16 ***
LnCasenum   -0.003422   0.014198  -0.241 0.809697
DUM24_5     -3.628497   0.267086 -13.586  < 2e-16 ***
DUM1_6       1.925429   0.441527   4.361 1.80e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4002 on 292 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:  0.6661, Adjusted R-squared:  0.6593
F-statistic:  97.1 on 6 and 292 DF,  p-value: < 2.2e-16
```

Figure A.6 Electronics Physical Market Regression Result

```
Final Model Output
------------------

                        Model Summary
-------------------------------------------------------------
R                       0.941       RMSE              0.405
R-Squared               0.886       Coef. Var         3.404
Adj. R-Squared          0.882       MSE               0.164
Pred R-Squared           -Inf       MAE               0.224
-------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                                ANOVA
---------------------------------------------------------------------
               Sum of
              Squares        DF    Mean Square       F         Sig.
---------------------------------------------------------------------
Regression    367.229         9         40.803    248.758    0.0000
Residual       47.404       289          0.164
Total         414.633       298
---------------------------------------------------------------------

                                   Parameter Estimates
------------------------------------------------------------------------------------------
      model       Beta    Std. Error    Std. Beta       t        Sig       lower     upper
------------------------------------------------------------------------------------------
(Intercept)      8.334         0.504                 16.525    0.000       7.342     9.327
    DUM24_5     -9.744         0.331       -0.825   -29.413    0.000     -10.396    -9.092
LnClust2_7       0.250        0.032         0.250     7.879   0.000       0.188     0.313
     DUM1_6      4.367         0.513        0.214     8.506    0.000       3.356     5.377
LnClust2_2      -0.382        0.036        -0.382   -10.579   0.000      -0.454    -0.311
LnClust2_1       0.227        0.035         0.227     6.416   0.000       0.158     0.297
 LnCasenum       0.077        0.016         0.106     4.895   0.000       0.046     0.107
LnClust2_3       0.183        0.030         0.183     6.039   0.000       0.123     0.243
LnClust2_5      -0.152        0.030        -0.152    -5.016   0.000      -0.212    -0.093
LnClust2_6       0.134        0.036         0.134     3.695   0.000       0.062     0.205
------------------------------------------------------------------------------------------
```

Figure A.7 Daily Needs Online Market Regression Result

```
Final Model Output
------------------

                        Model Summary
-------------------------------------------------------------
R                       0.930       RMSE              0.426
R-Squared               0.866       Coef. Var         3.867
Adj. R-Squared          0.862       MSE               0.182
Pred R-Squared           -Inf       MAE               0.244
-------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                                ANOVA
---------------------------------------------------------------------
               Sum of
              Squares        DF    Mean Square       F         Sig.
---------------------------------------------------------------------
Regression    338.437         9         37.604    207.135    0.0000
Residual       52.466       289          0.182
Total         390.903       298
---------------------------------------------------------------------

                                   Parameter Estimates
-----------------------------------------------------------------------------------------
      model       Beta    Std. Error    Std. Beta       t        Sig      lower     upper
-----------------------------------------------------------------------------------------
(Intercept)      7.068         0.504                 14.035    0.000      6.077     8.059
    DUM24_5     -8.829         0.335       -0.770   -26.374    0.000     -9.488    -8.170
LnClust3_7       0.283        0.033         0.282     8.594   0.000      0.218     0.348
     DUM1_6      4.888         0.531        0.247     9.201    0.000      3.843     5.934
LnClust3_2      -0.345        0.038        -0.345    -9.077   0.000     -0.420    -0.270
LnClust3_1       0.256        0.037         0.256     6.982   0.000      0.184     0.329
 LnCasenum       0.056        0.016         0.080     3.416   0.001      0.024     0.088
LnClust3_3       0.147        0.032         0.147     4.569   0.000      0.084     0.210
LnClust3_5      -0.161        0.032        -0.161    -4.956   0.000     -0.225    -0.097
LnClust3_6       0.148        0.038         0.148     3.853   0.000      0.072     0.224
-----------------------------------------------------------------------------------------
```

Figure A.8 Protein-Based Products Online Market Regression Result

Final Model Output

Model Summary

R	0.933	RMSE	0.450
R-Squared	0.870	Coef. Var	3.621
Adj. R-Squared	0.866	MSE	0.202
Pred R-Squared	-Inf	MAE	0.255

RMSE: Root Mean Square Error
MSE: Mean Square Error
MAE: Mean Absolute Error

ANOVA

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	391.351	9	43.483	214.781	0.0000
Residual	58.509	289	0.202		
Total	449.860	298			

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	8.534	0.562		15.181	0.000	7.427	9.640
DUM24_5	-9.898	0.368	-0.804	-26.921	0.000	-10.622	-9.175
LnClust4_7	0.268	0.034	0.267	7.887	0.000	0.201	0.335
DUM1_6	4.804	0.567	0.226	8.466	0.000	3.687	5.921
LnClust4_2	-0.394	0.038	-0.393	-10.283	0.000	-0.469	-0.318
LnClust4_1	0.253	0.038	0.253	6.733	0.000	0.179	0.327
LnClust4_3	0.188	0.032	0.188	5.914	0.000	0.126	0.251
LnClust4_5	-0.145	0.032	-0.144	-4.535	0.000	-0.207	-0.082
LnClust4_6	0.129	0.038	0.128	3.362	0.001	0.053	0.204
LnCasenum	0.035	0.017	0.047	2.100	0.037	0.002	0.068

Figure A.9 Basic Consumption Goods Online Market Regression Result

Final Model Output

Model Summary

R	0.938	RMSE	0.415
R-Squared	0.880	Coef. Var	3.634
Adj. R-Squared	0.877	MSE	0.172
Pred R-Squared	-Inf	MAE	0.226

RMSE: Root Mean Square Error
MSE: Mean Square Error
MAE: Mean Absolute Error

ANOVA

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	366.294	9	40.699	236.319	0.0000
Residual	49.772	289	0.172		
Total	416.066	298			

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	8.051	0.503		16.020	0.000	7.062	9.040
DUM24_5	-9.492	0.326	-0.802	-29.133	0.000	-10.133	-8.850
LnClust5_7	0.252	0.030	0.253	8.306	0.000	0.192	0.312
DUM1_6	4.925	0.525	0.241	9.378	0.000	3.891	5.958
LnClust5_2	-0.327	0.036	-0.328	-9.167	0.000	-0.398	-0.257
LnClust5_1	0.229	0.034	0.229	6.662	0.000	0.161	0.296
LnCasenum	0.064	0.016	0.088	3.988	0.000	0.032	0.095
LnClust5_5	-0.176	0.031	-0.177	-5.669	0.000	-0.237	-0.115
LnClust5_6	0.156	0.036	0.156	4.265	0.000	0.084	0.227
LnClust5_3	0.128	0.031	0.128	4.190	0.000	0.068	0.188

Figure A.10 Cold-Chain Products Online Market Regression Result

```
Final Model Output
------------------

                          Model Summary
-------------------------------------------------------------------
R                          0.810       RMSE                 0.631
R-Squared                  0.656       Coef. Var           10.431
Adj. R-Squared             0.647       MSE                  0.399
Pred R-Squared              -Inf       MAE                  0.419
-------------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                                  ANOVA
--------------------------------------------------------------------------
               Sum of
              Squares          DF     Mean Square       F          Sig.
--------------------------------------------------------------------------
Regression    220.831           7          31.547     79.143     0.0000
Residual      115.996         291           0.399
Total         336.827         298
--------------------------------------------------------------------------

                                     Parameter Estimates
-------------------------------------------------------------------------------------------------
      model         Beta    Std. Error    Std. Beta        t         Sig       lower      upper
-------------------------------------------------------------------------------------------------
(Intercept)        1.373        0.343                    4.005     0.000      0.698      2.047
LnClust6_1        0.437        0.049        0.437        8.904     0.000      0.341      0.534
    DUM24_5       -4.154        0.405       -0.390     -10.248     0.000     -4.952     -3.356
LnClust6_7        0.332        0.046        0.328        7.302     0.000      0.243      0.422
     DUM1_6        3.822        0.689        0.208        5.546     0.000      2.465      5.178
LnClust6_2       -0.182        0.044       -0.181       -4.151     0.000     -0.268     -0.096
 LnCasenum        0.084        0.025        0.129        3.335     0.001      0.034      0.133
LnClust6_6        0.089        0.044        0.088        2.011     0.045      0.002      0.176
-------------------------------------------------------------------------------------------------
```

Figure A.11 Electronics Online Market Regression Result

LASSO RESULTS

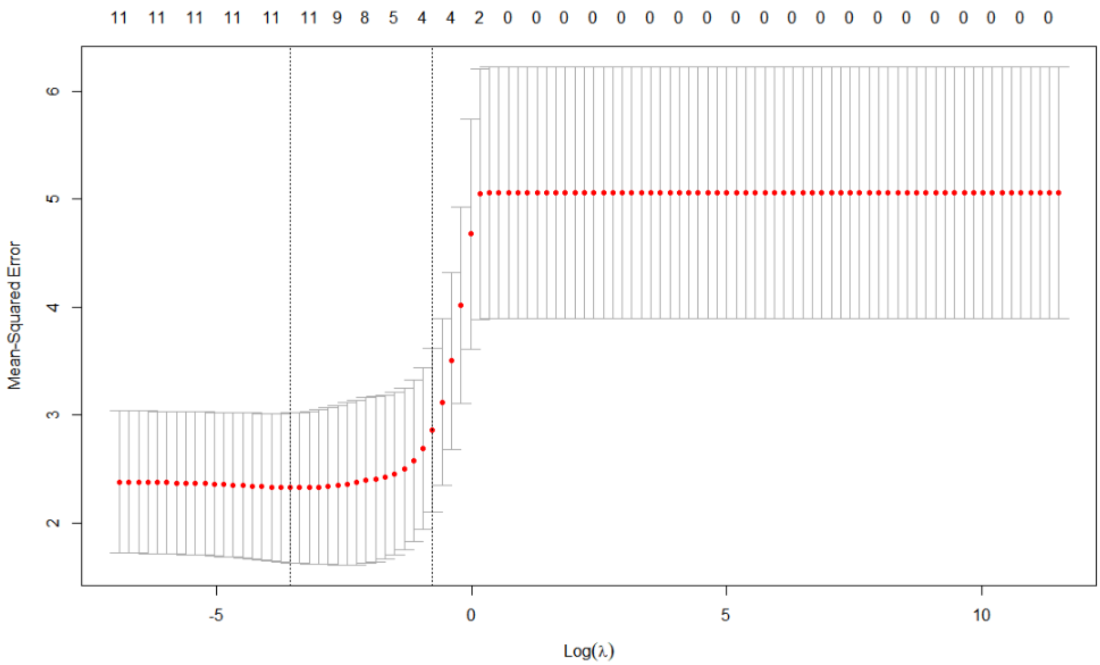


Figure A.12 Alcoholic Beverages Physical Market Lasso Result

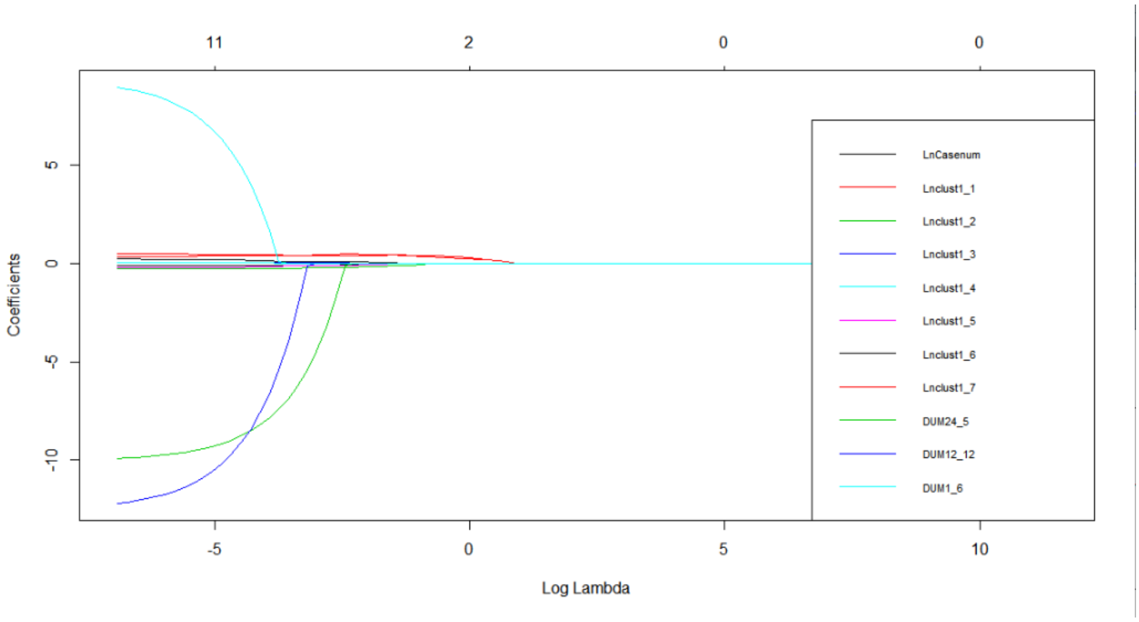


Figure A.13 Alcoholic Beverages Physical Market Lasso Result

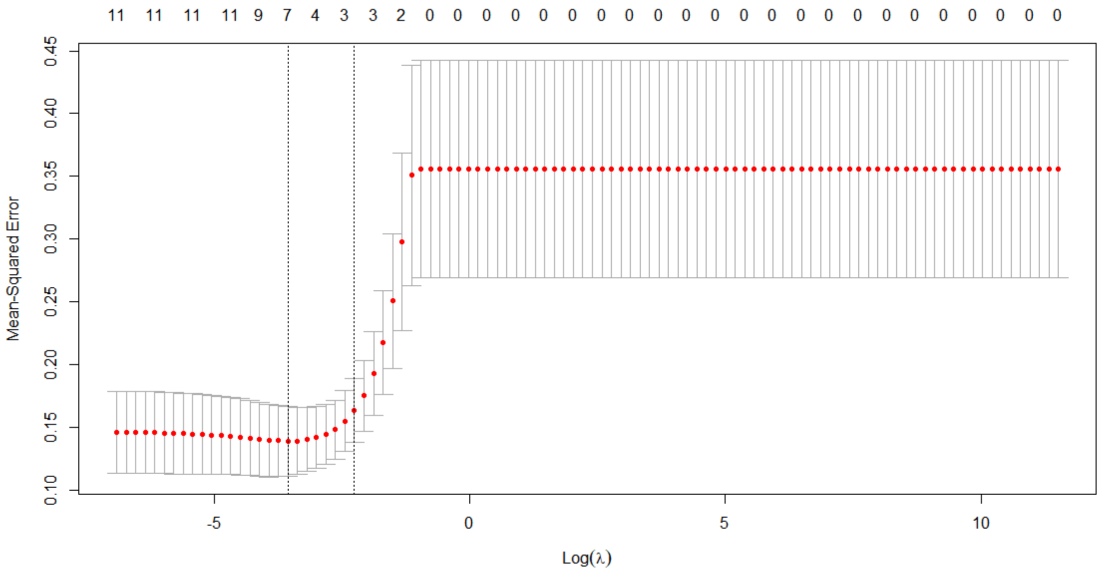


Figure A.14 Daily Needs Physical Market Mean Squared Error Result

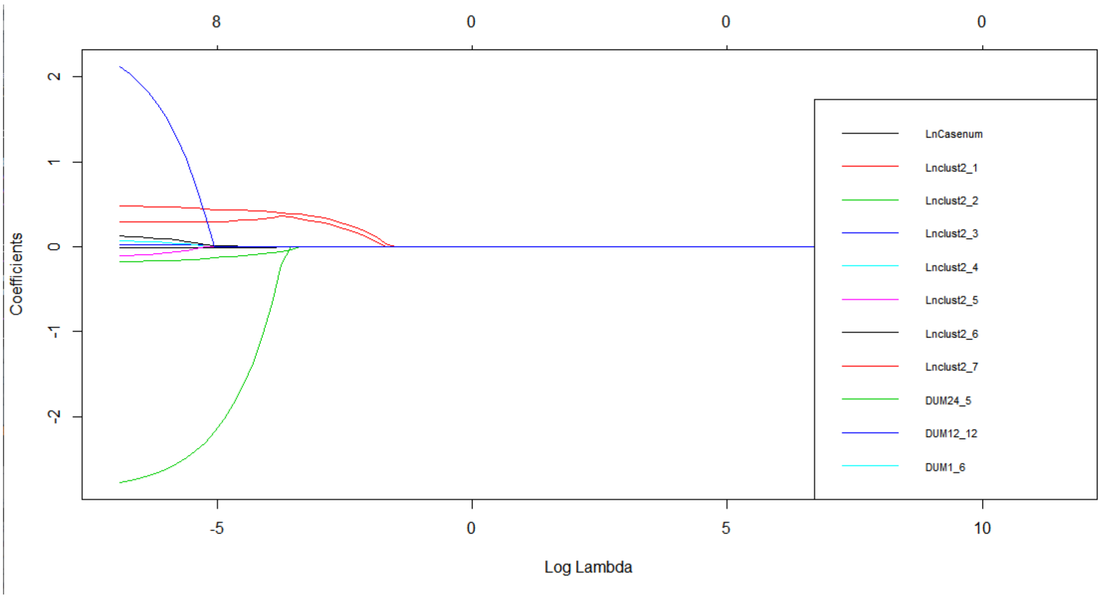


Figure A.15 Daily Needs Physical Market Lasso Result

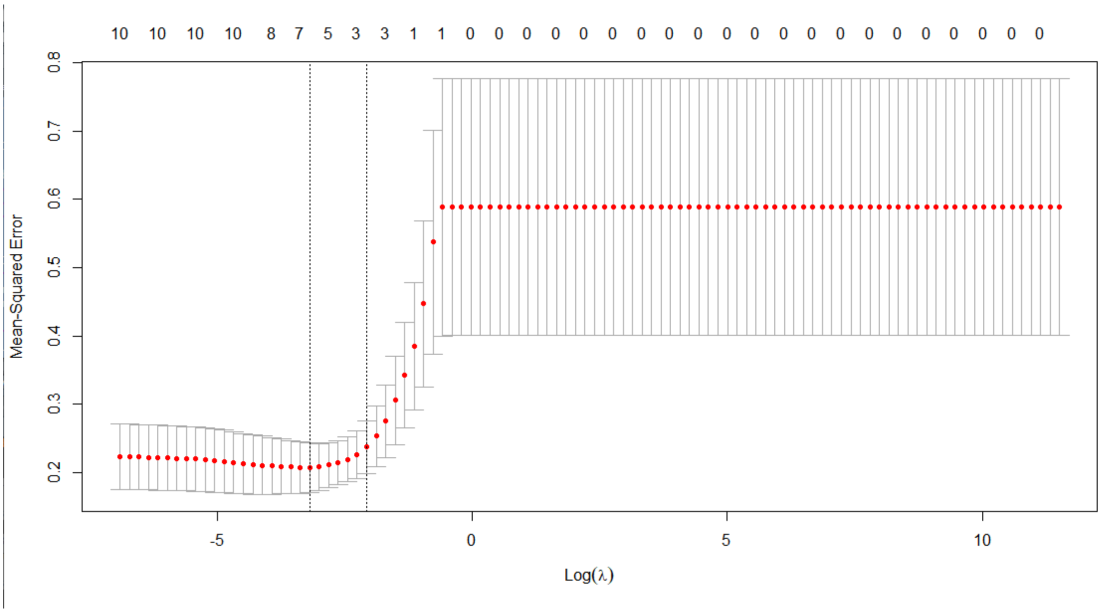


Figure A.16 Protein-based products Physical Market Mean Squared Error

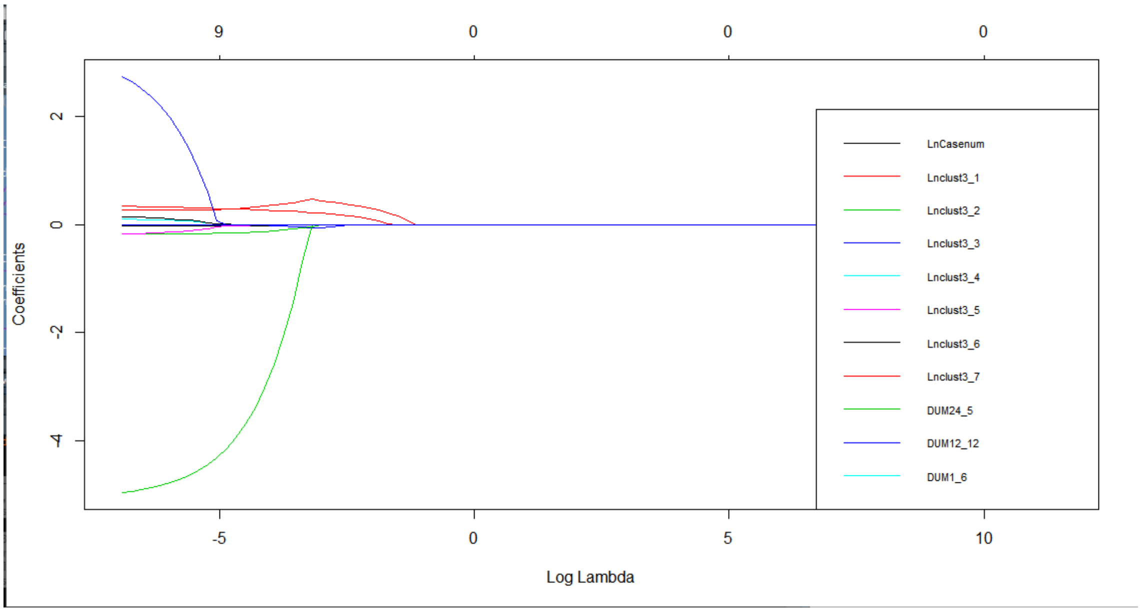


Figure A.17 Protein-based products Physical Market Lasso Result

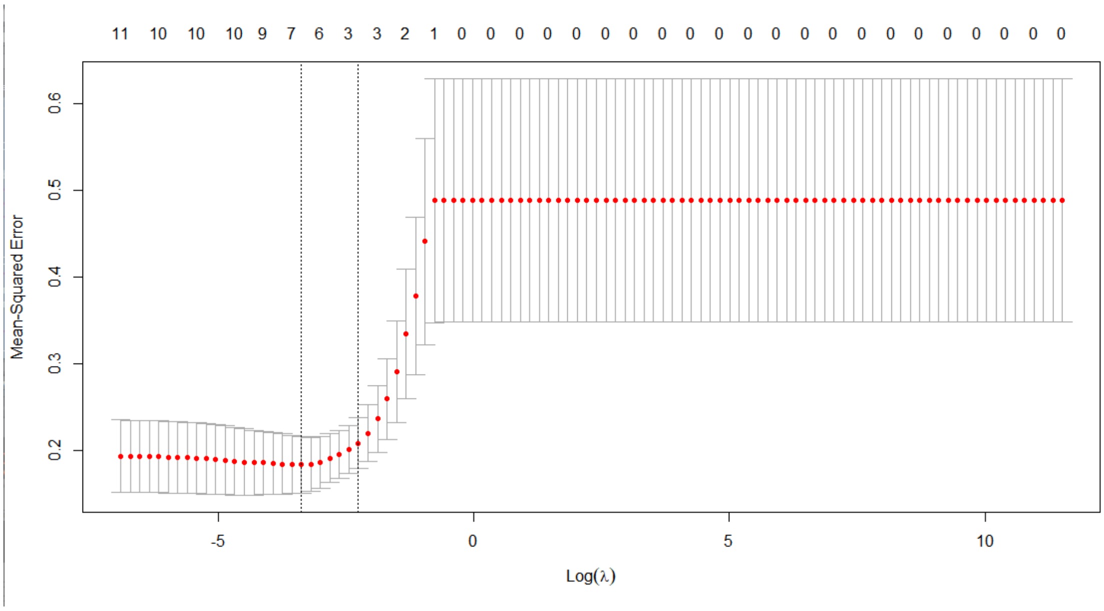


Figure A.18 Basic Consumption Goods Physical Market Mean Squared Error

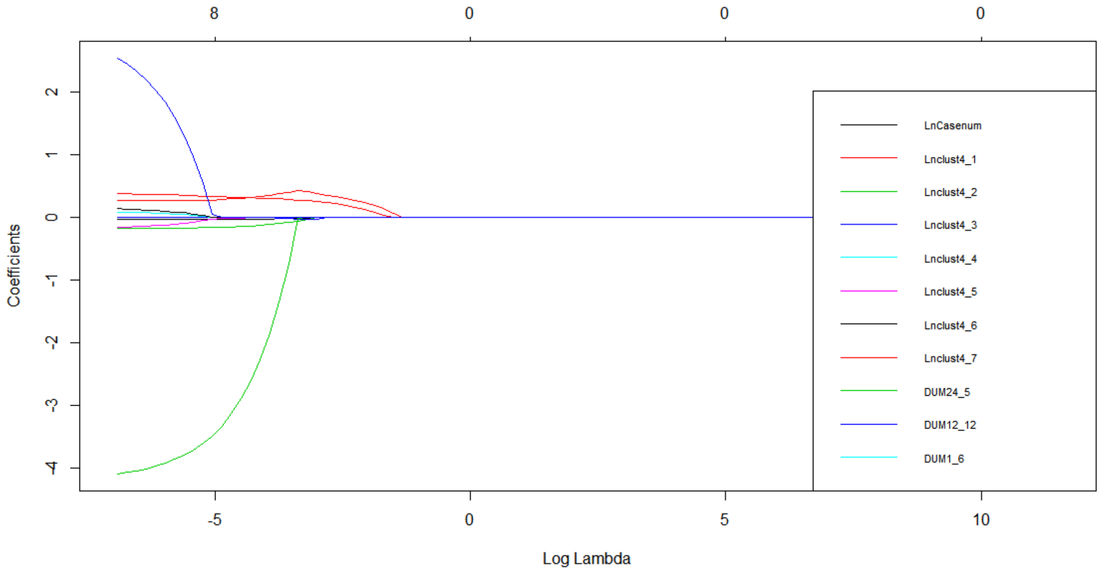


Figure A.19 Basic Consumption Goods Physical Market Lasso Result

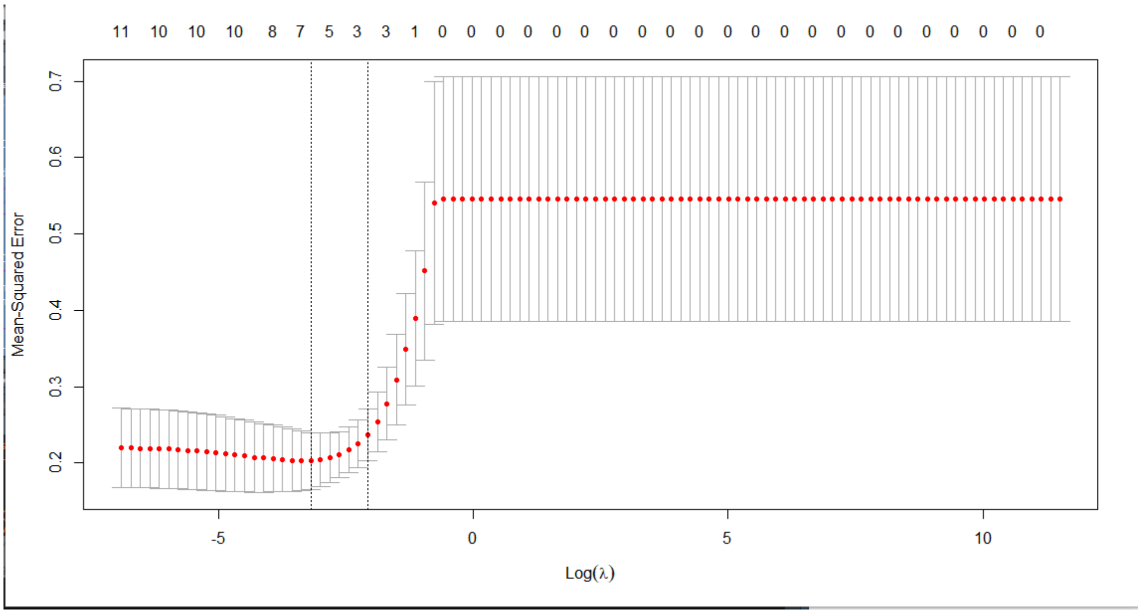


Figure A.20 Cold Chain products Physical Market Mean Squared Error

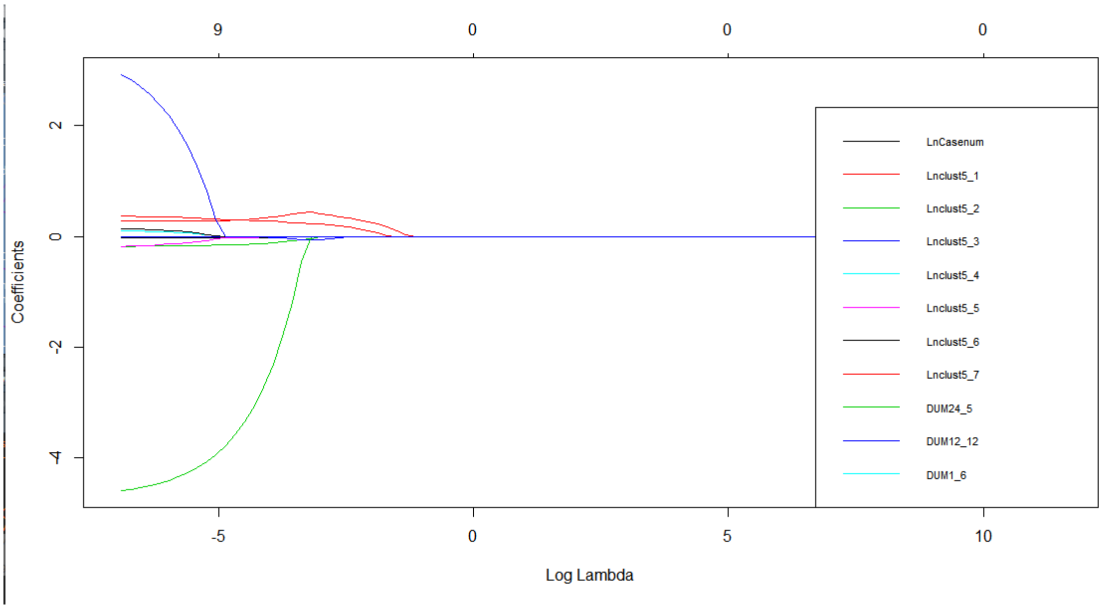


Figure A.21 Cold Chain products Physical Market Lasso Result

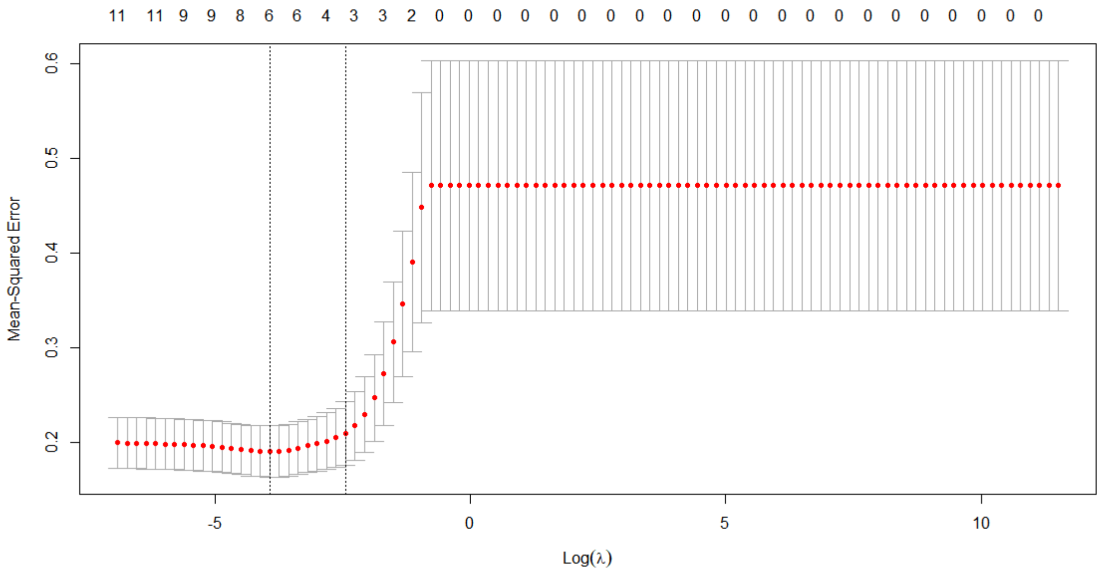


Figure A.22 Electronics Physical Market Mean Squared Error

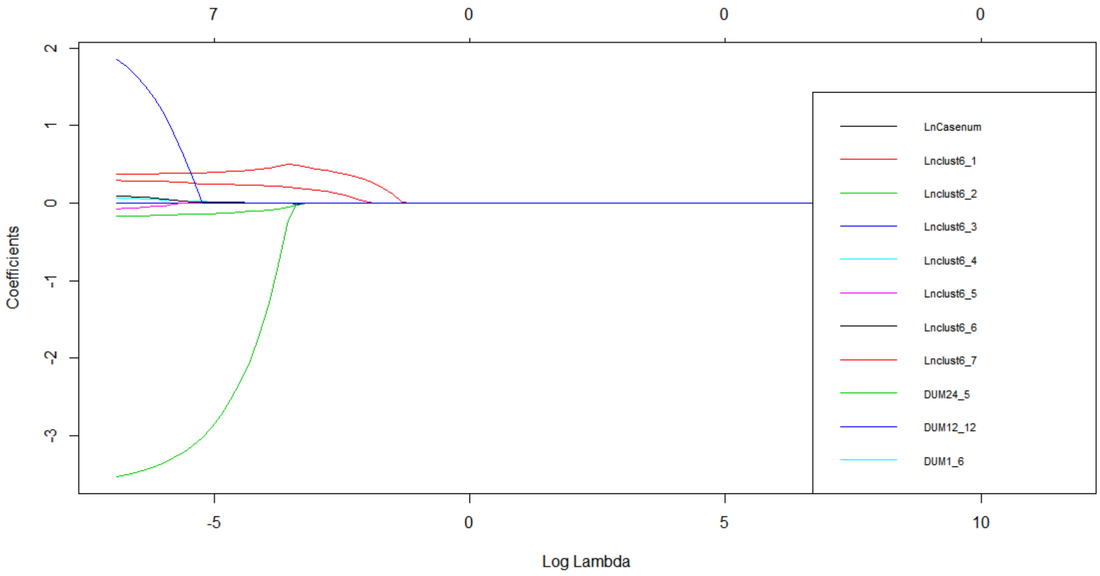


Figure A.23 Electronics Physical Market Lasso Result

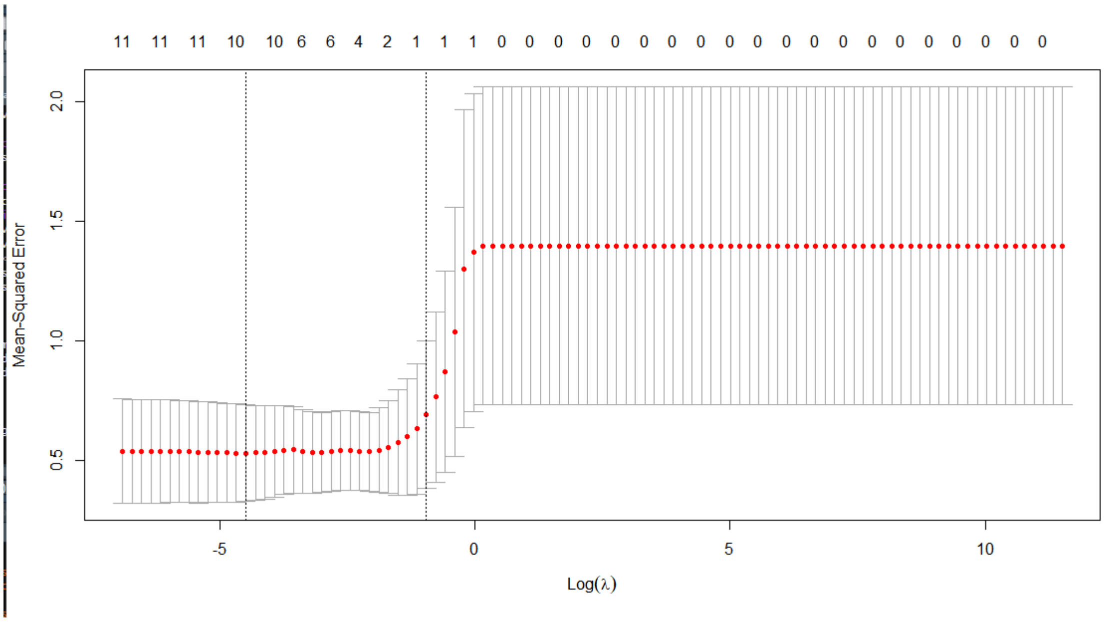


Figure A.24 Daily Needs Online Market Mean Squared Error Result

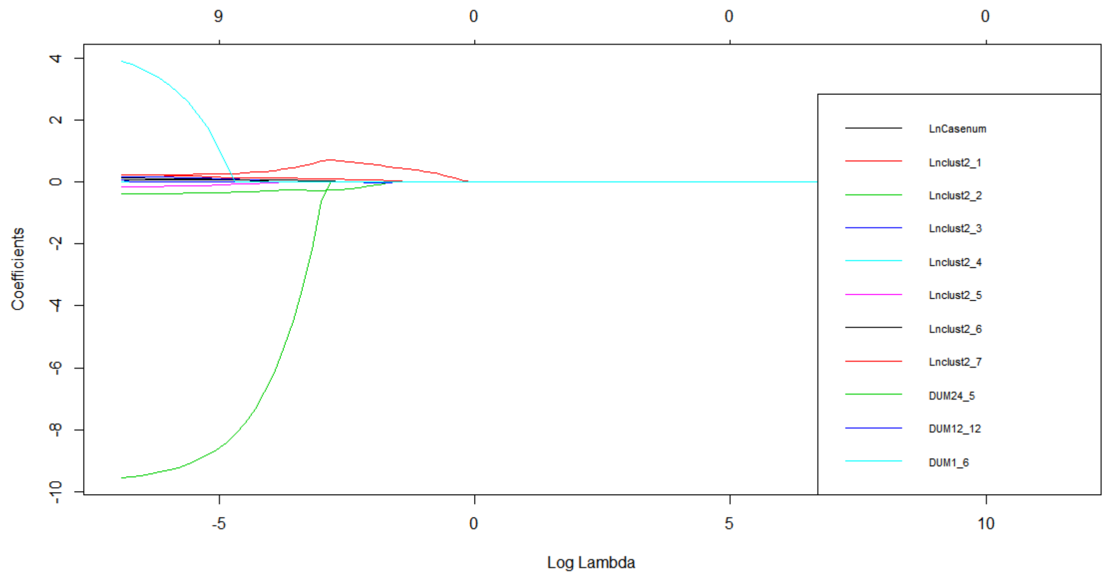


Figure A.25 Daily Needs Online Market Lasso Result

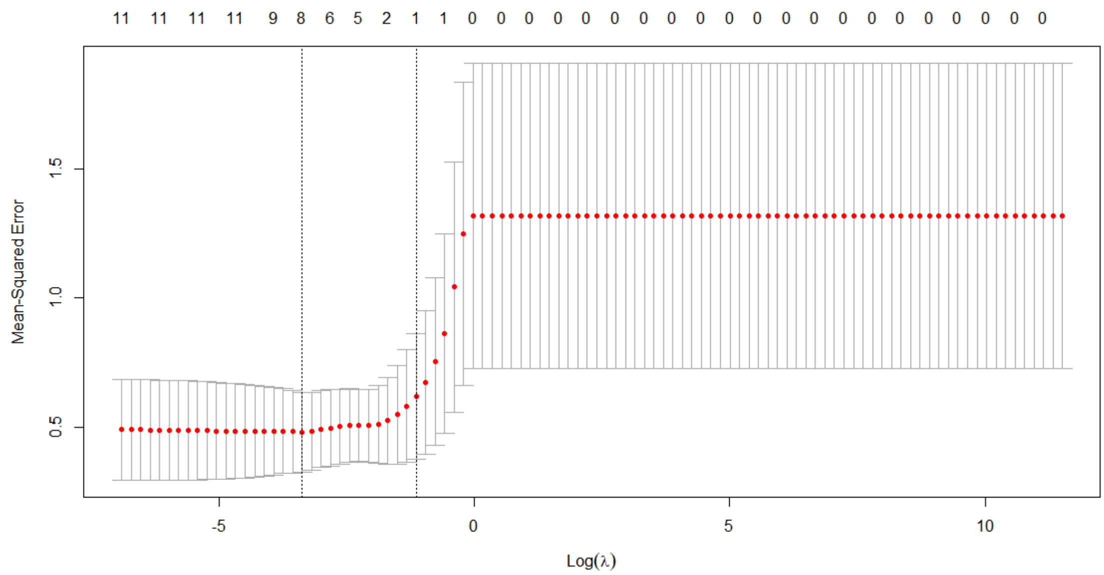


Figure A.26 Protein-based products Online Market Mean Squared Error

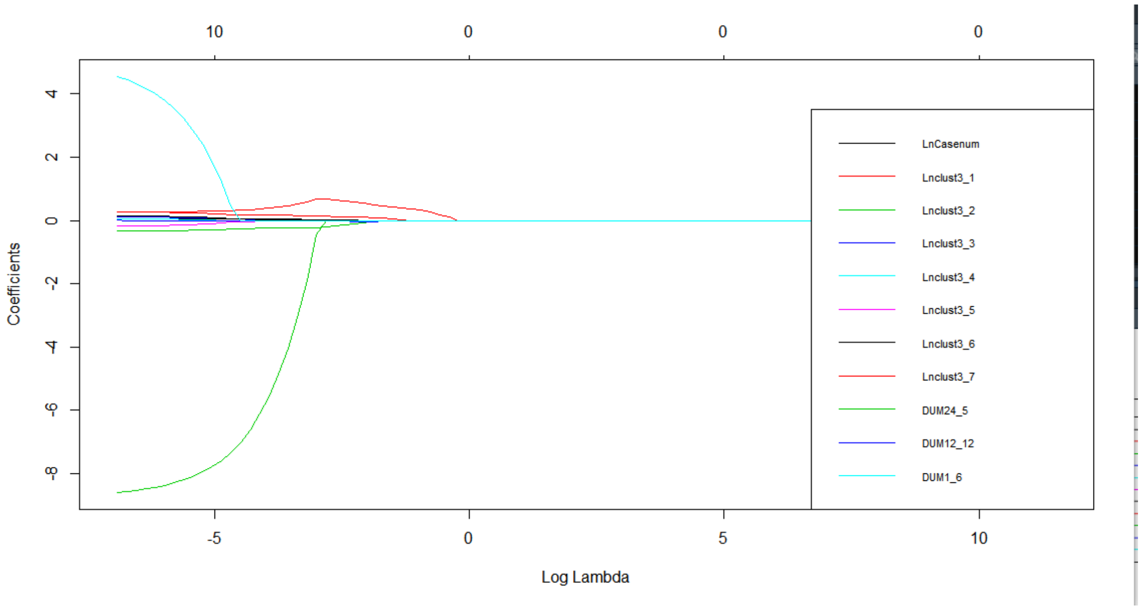


Figure A.27 Protein-based products Online Market Lasso Result

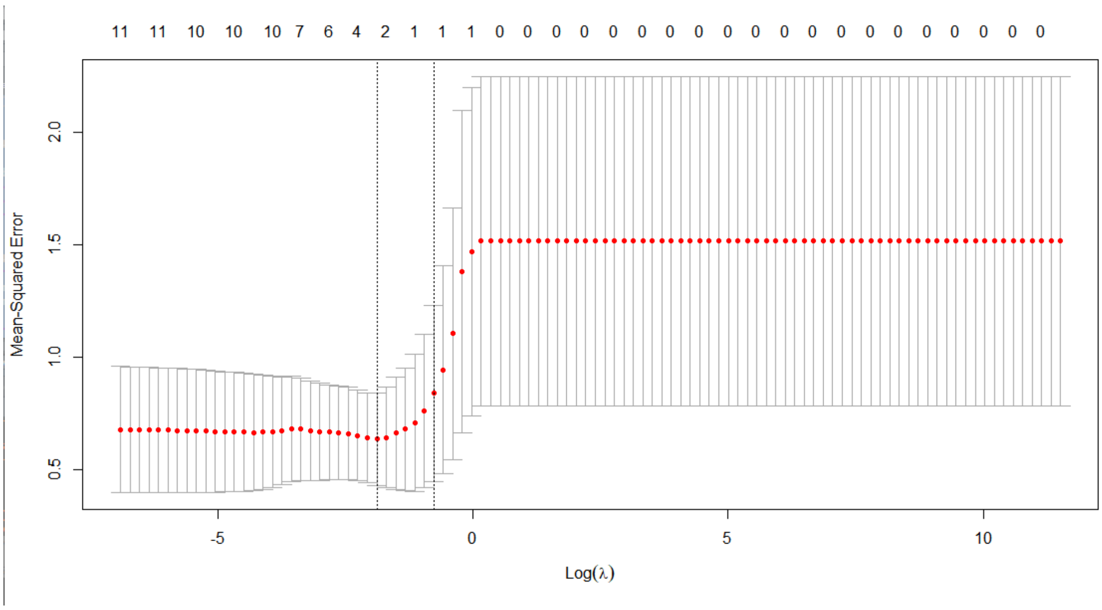


Figure A.28 Basic Consumption Goods Online Market Mean Squared Error

Figure A.29 Basic Consumption Goods Online Market Lasso Result

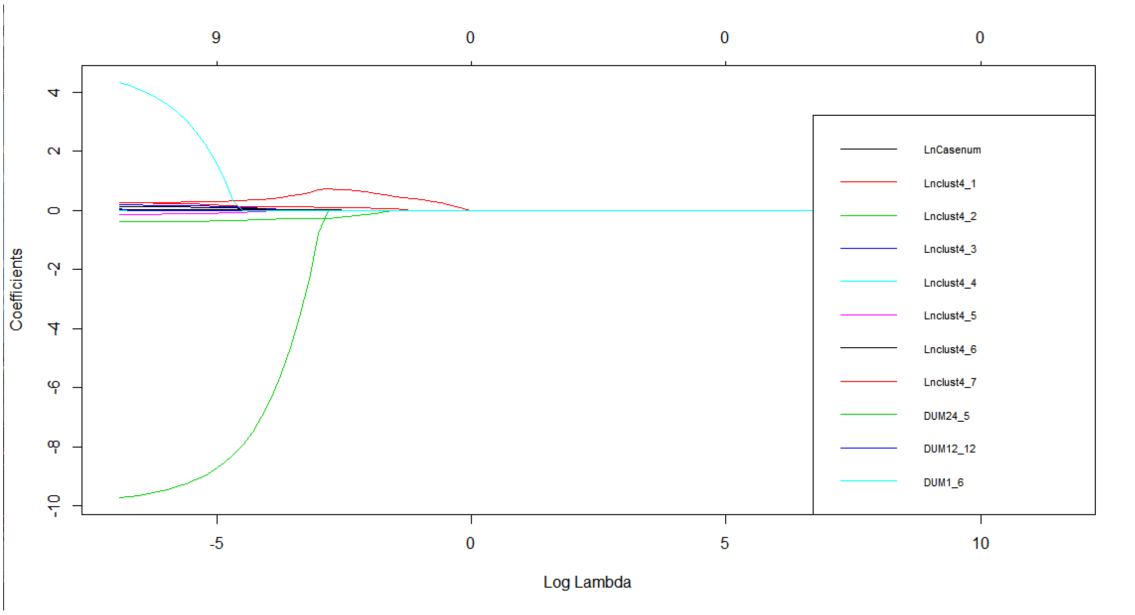


Figure A.30 Basic Consumption Goods Online Market Lasso Result

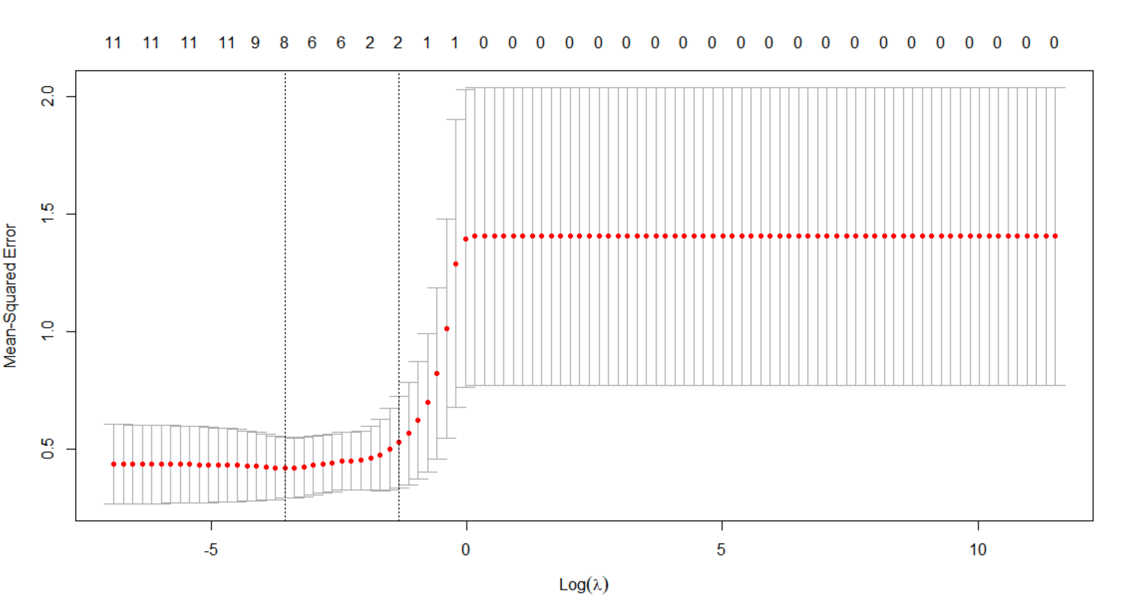


Figure A.31 Cold Chain products Online Market Mean Squared Error

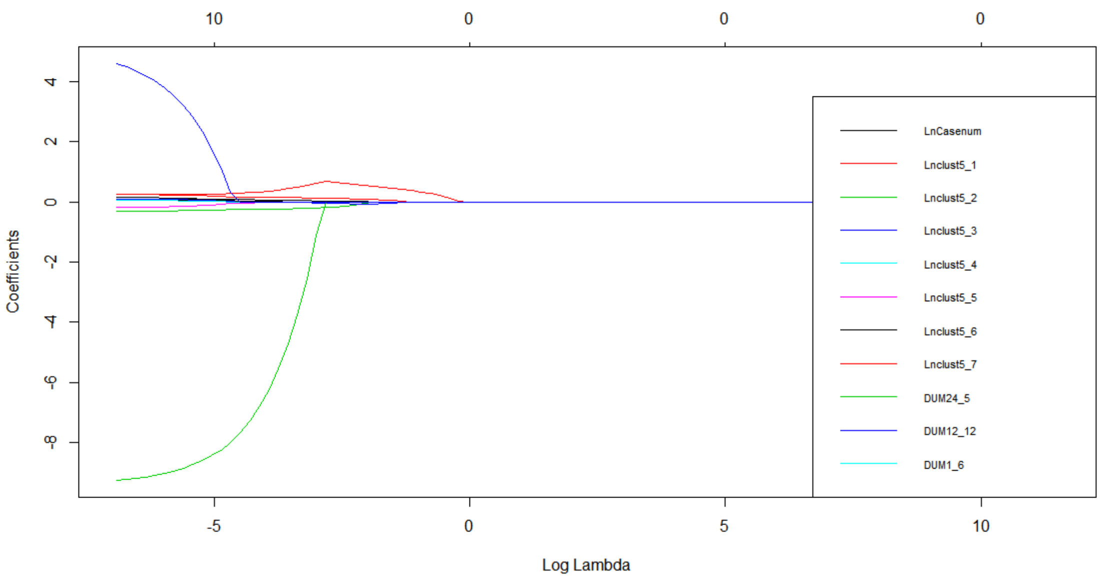


Figure A.32 Cold Chain products Online Market Lasso Result

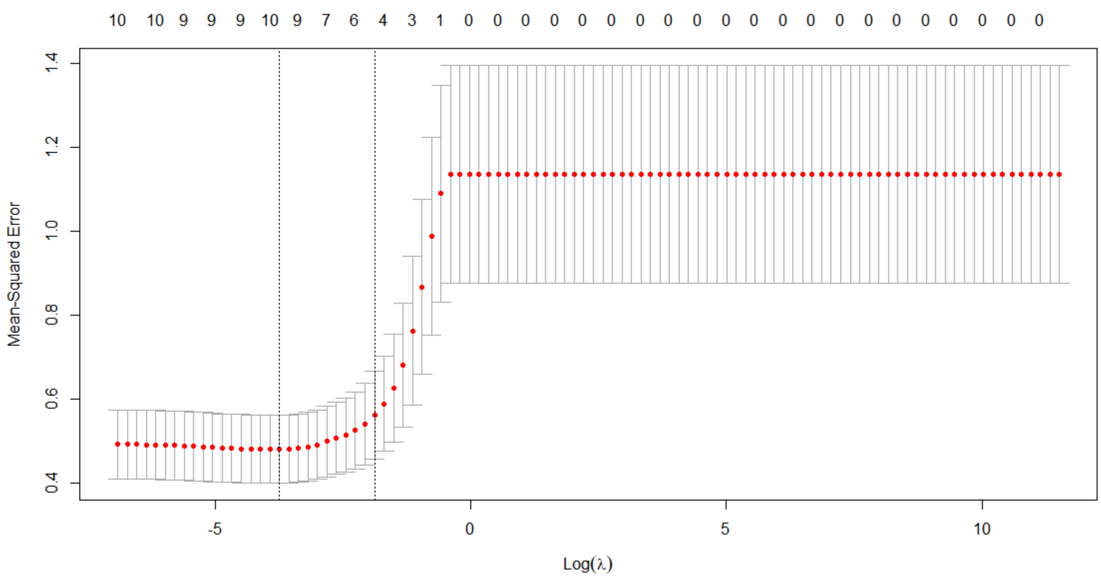


Figure A.33 Electronics Online Market Mean Squared Error

Figure A.34 Electronics Online Market Lasso Result

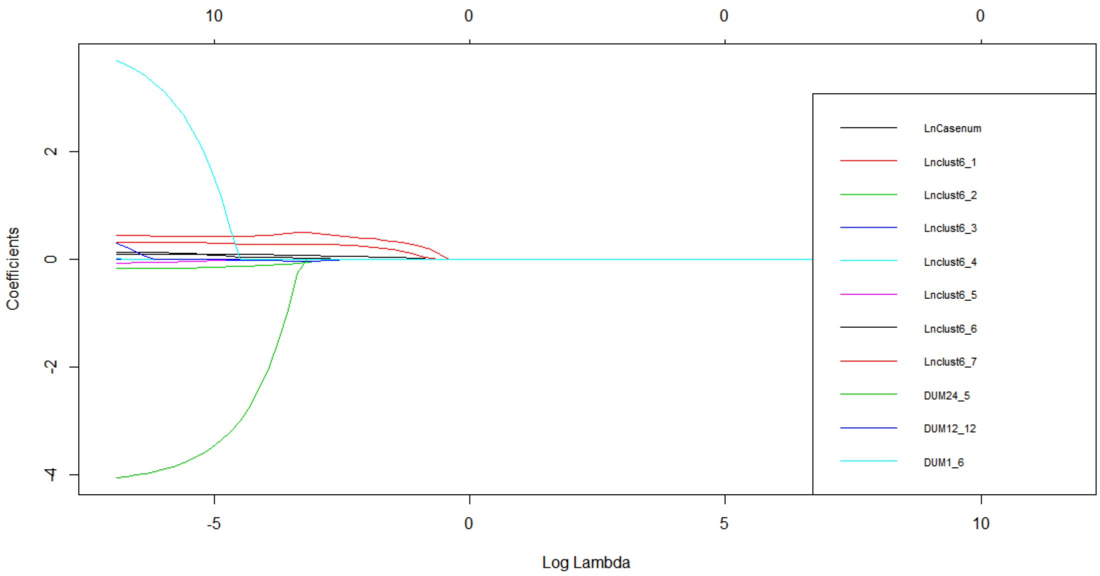


Figure A.35 Electronics Online Market Lasso Result

BEST SUBSET MODEL RESULTS

```
ols_step_best_subset(model)
                                   Best Subsets Regression
---------------------------------------------------------------------------------------------------------------
odel Index    Predictors
---------------------------------------------------------------------------------------------------------------
    1         Lnclust1_7
    2         Lnclust1_7 DUM24_5
    3         Lnclust1_7 DUM24_5 DUM12_12
    4         Lnclust1_1 Lnclust1_7 DUM24_5 DUM12_12
    5         Lnclust1_1 Lnclust1_7 DUM24_5 DUM12_12 DUM1_6
    6         Lnclust1_1 Lnclust1_2 Lnclust1_7 DUM24_5 DUM12_12 DUM1_6
    7         Lnclust1_1 Lnclust1_2 Lnclust1_5 Lnclust1_7 DUM24_5 DUM12_12 DUM1_6
    8         Lnclust1_1 Lnclust1_2 Lnclust1_5 Lnclust1_6 Lnclust1_7 DUM24_5 DUM12_12 DUM1_6
    9         Lnclust1_1 Lnclust1_2 Lnclust1_4 Lnclust1_5 Lnclust1_6 Lnclust1_7 DUM24_5 DUM12_12 DUM1_6
   10         LnCasenum Lnclust1_1 Lnclust1_2 Lnclust1_4 Lnclust1_5 Lnclust1_6 Lnclust1_7 DUM24_5 DUM12_12 DUM1_6
   11         LnCasenum Lnclust1_1 Lnclust1_2 Lnclust1_3 Lnclust1_4 Lnclust1_5 Lnclust1_6 Lnclust1_7 DUM24_5 DUM12_12 DUM1_6
---------------------------------------------------------------------------------------------------------------

                                                   Subsets Regression Summary
-----------------------------------------------------------------------------------------------------------------------
                       Adj.        Pred
odel     R-Square    R-Square    R-Square     C(p)        AIC         SBIC        SBC         MSEP       FPE       HSP       APC
-----------------------------------------------------------------------------------------------------------------------
 1        0.3036      0.3013      0.2433    858.7401   1229.5098    376.9234   1240.6111   1054.9872   3.5520    0.0119    0.7057
 2        0.5864      0.5836      0.5284    392.1990   1075.7143    223.1719   1090.5161    628.6742   2.1237    0.0071    0.4220
 3        0.6891      0.6859        -Inf    224.1416    992.4193    140.1854   1010.9216    474.2531   1.6073    0.0054    0.3194
 4        0.7366      0.7330        -Inf    147.3489    944.7854     92.9144    966.9881    403.0855   1.3706    0.0046    0.2723
 5        0.7733      0.7695        -Inf     88.5189    901.8956     50.8421    927.7987    348.0807   1.1875    0.0040    0.2359
 6        0.7943      0.7901        -Inf     55.7443    874.8398     24.5694    904.4433    316.9323   1.0847    0.0036    0.2155
 7        0.8001      0.7952        -Inf     48.2642    868.4032     18.3007    901.7072    309.1772   1.0616    0.0036    0.2109
 8        0.8133      0.8082        -Inf     28.2876    849.8784      0.7523    886.8829    289.6647   0.9979    0.0034    0.1983
 9        0.8219      0.8164        -Inf     16.0149    837.7520    -10.5034    878.4569    277.2569   0.9582    0.0032    0.1904
10        0.8256      0.8195        -Inf     12.0263    833.6200    -14.1484    878.0253    272.5752   0.9451    0.0032    0.1878
11        0.8268      0.8201        -Inf     12.0000    833.5164    -14.0088    881.6222    271.6107   0.9447    0.0032    0.1877
-----------------------------------------------------------------------------------------------------------------------
IC: Akaike Information Criteria
SBIC: Sawa's Bayesian Information Criteria
SBC: Schwarz Bayesian Criteria
MSEP: Estimated error of prediction, assuming multivariate normality
FPE: Final Prediction Error
```

Figure A.36 Alcoholic Beverages Physical Market Best Subset Model Result

Best Subsets Regression

Model Index	Predictors
1	DUM24_5
2	Lnclust2_7 DUM24_5
3	Lnclust2_7 DUM24_5 DUM1_6
4	Lnclust2_1 Lnclust2_7 DUM24_5 DUM1_6
5	Lnclust2_1 Lnclust2_2 Lnclust2_7 DUM24_5 DUM1_6
6	Lnclust2_1 Lnclust2_2 Lnclust2_6 Lnclust2_7 DUM24_5 DUM1_6
7	Lnclust2_1 Lnclust2_2 Lnclust2_3 Lnclust2_6 Lnclust2_7 DUM24_5 DUM1_6
8	Lnclust2_1 Lnclust2_2 Lnclust2_4 Lnclust2_5 Lnclust2_6 Lnclust2_7 DUM24_5 DUM1_6
9	LnCasenum Lnclust2_1 Lnclust2_2 Lnclust2_4 Lnclust2_5 Lnclust2_6 Lnclust2_7 DUM24_5 DUM1_6
10	LnCasenum Lnclust2_1 Lnclust2_2 Lnclust2_3 Lnclust2_4 Lnclust2_5 Lnclust2_6 Lnclust2_7 DUM24_5 DUM1_6
11	LnCasenum Lnclust2_1 Lnclust2_2 Lnclust2_3 Lnclust2_4 Lnclust2_5 Lnclust2_6 Lnclust2_7 DUM24_5 DUM12_12 DUM1_6

Subsets Regression Summary

Model	R-Square	Adj. R-Square	Pred R-Square	C(p)	AIC	SBIC	SBC	MSEP	FPE	HSP	APC
1	0.3660	0.3639	0.3603	393.7108	407.1665	-444.2626	418.2678	67.4220	0.2270	8e-04	0.6425
2	0.6371	0.6346	0.6041	101.2296	242.3579	-607.7334	257.1597	38.7243	0.1308	4e-04	0.3703
3	0.6780	0.6747	-Inf	58.8256	208.6278	-641.1020	227.1300	34.4795	0.1169	4e-04	0.3308
4	0.7050	0.7009	-Inf	31.4957	184.4457	-664.7593	206.6483	31.6965	0.1078	4e-04	0.3051
5	0.7194	0.7146	-Inf	17.8610	171.4918	-677.2649	197.3949	30.2535	0.1032	3e-04	0.2921
6	0.7228	0.7171	-Inf	16.1202	169.8002	-678.8237	199.4037	29.9850	0.1026	3e-04	0.2905
7	0.7259	0.7193	-Inf	14.7197	168.4044	-680.0521	201.7084	29.7486	0.1021	3e-04	0.2891
8	0.7330	0.7256	-Inf	9.0817	162.6334	-685.3403	199.6379	29.0857	0.1002	3e-04	0.2836
9	0.7351	0.7269	-Inf	8.7549	162.2255	-685.5212	202.9304	28.9526	0.1001	3e-04	0.2832
10	0.7355	0.7263	-Inf	10.3366	163.7905	-683.8451	208.1958	29.0112	0.1006	3e-04	0.2847
11	0.7358	0.7257	-Inf	12.0000	165.4400	-682.0852	213.5458	29.0785	0.1011	3e-04	0.2863

AIC: Akaike Information Criteria
SBIC: Sawa's Bayesian Information Criteria
SBC: Schwarz Bayesian Criteria
MSEP: Estimated error of prediction, assuming multivariate normality
FPE: Final Prediction Error

Figure A.37 Daily Needs Physical Market Best Subset Model Result

ols_step_best_subset(model3)

Best Subsets Regression

odel Index	Predictors
1	DUM24_5
2	Lnclust3_7 DUM24_5
3	Lnclust3_7 DUM24_5 DUM1_6
4	Lnclust3_1 Lnclust3_7 DUM24_5 DUM1_6
5	Lnclust3_1 Lnclust3_2 Lnclust3_7 DUM24_5 DUM1_6
6	LnCasenum Lnclust3_1 Lnclust3_2 Lnclust3_7 DUM24_5 DUM1_6
7	Lnclust3_1 Lnclust3_2 Lnclust3_5 Lnclust3_6 Lnclust3_7 DUM24_5 DUM1_6
8	Lnclust3_1 Lnclust3_2 Lnclust3_4 Lnclust3_5 Lnclust3_6 Lnclust3_7 DUM24_5 DUM1_6
9	LnCasenum Lnclust3_1 Lnclust3_2 Lnclust3_4 Lnclust3_5 Lnclust3_6 Lnclust3_7 DUM24_5 DUM1_6
10	LnCasenum Lnclust3_1 Lnclust3_2 Lnclust3_4 Lnclust3_5 Lnclust3_6 Lnclust3_7 DUM24_5 DUM12_12 DUM1_6
11	LnCasenum Lnclust3_1 Lnclust3_2 Lnclust3_3 Lnclust3_4 Lnclust3_5 Lnclust3_6 Lnclust3_7 DUM24_5 DUM12_12 DUM1_6

Subsets Regression Summary

odel	R-Square	Adj. R-Square	Pred R-Square	C(p)	AIC	SBIC	SBC	MSEP	FPE	HSP	APC
1	0.5807	0.5793	0.5775	235.1898	434.1341	-416.5156	445.2355	73.7857	0.2484	8e-04	0.4250
2	0.6899	0.6878	0.6572	99.0455	345.8788	-504.1831	360.6806	54.7453	0.1849	6e-04	0.3163
3	0.7173	0.7144	-Inf	66.4921	320.2918	-529.5960	338.7940	50.0901	0.1698	6e-04	0.2904
4	0.7334	0.7297	-Inf	48.1288	304.7559	-544.9259	326.9586	47.3981	0.1612	5e-04	0.2757
5	0.7504	0.7462	-Inf	28.5850	287.0108	-562.1506	312.9139	44.5211	0.1519	5e-04	0.2598
6	0.7550	0.7500	-Inf	24.7678	283.4479	-565.5664	313.0514	43.8506	0.1501	5e-04	0.2567
7	0.7607	0.7550	-Inf	19.5219	278.3708	-570.3410	311.6748	42.9726	0.1476	5e-04	0.2524
8	0.7684	0.7620	-Inf	11.8481	270.6535	-577.4945	307.6579	41.7424	0.1438	5e-04	0.2460
9	0.7728	0.7657	-Inf	8.3160	266.9511	-580.7642	307.6560	41.0960	0.1420	5e-04	0.2430
10	0.7730	0.7651	-Inf	10.0014	268.6235	-578.9855	313.0289	41.1941	0.1428	5e-04	0.2443
11	0.7730	0.7643	-Inf	12.0000	270.6221	-576.9032	318.7278	41.3379	0.1438	5e-04	0.2460

IC: Akaike Information Criteria
SBIC: Sawa's Bayesian Information Criteria
SBC: Schwarz Bayesian Criteria
MSEP: Estimated error of prediction, assuming multivariate normality
FPE: Final Prediction Error

Figure A.38 Protein-Based Products Physical Market Best Subset Model Result

```
> ols_step_best_subset(model4)
                                          Best Subsets Regression
--------------------------------------------------------------------------------------------------------------------------
Model Index    Predictors
--------------------------------------------------------------------------------------------------------------------------
     1         DUM24_5
     2         Lnclust4_7 DUM24_5
     3         Lnclust4_7 DUM24_5 DUM1_6
     4         Lnclust4_1 Lnclust4_7 DUM24_5 DUM1_6
     5         Lnclust4_1 Lnclust4_2 Lnclust4_7 DUM24_5 DUM1_6
     6         LnCasenum Lnclust4_1 Lnclust4_2 Lnclust4_7 DUM24_5 DUM1_6
     7         LnCasenum Lnclust4_1 Lnclust4_2 Lnclust4_5 Lnclust4_7 DUM24_5 DUM1_6
     8         LnCasenum Lnclust4_1 Lnclust4_2 Lnclust4_5 Lnclust4_6 Lnclust4_7 DUM24_5 DUM1_6
     9         LnCasenum Lnclust4_1 Lnclust4_2 Lnclust4_4 Lnclust4_5 Lnclust4_6 Lnclust4_7 DUM24_5 DUM1_6
    10         LnCasenum Lnclust4_1 Lnclust4_2 Lnclust4_4 Lnclust4_5 Lnclust4_6 Lnclust4_7 DUM24_5 DUM12_12 DUM1_6
    11         LnCasenum Lnclust4_1 Lnclust4_2 Lnclust4_3 Lnclust4_4 Lnclust4_5 Lnclust4_6 Lnclust4_7 DUM24_5 DUM12_12 DUM1_6
--------------------------------------------------------------------------------------------------------------------------

                                                   Subsets Regression Summary
-------------------------------------------------------------------------------------------------------------------------------
                       Adj.        Pred
Model    R-Square    R-Square    R-Square     C(p)        AIC         SBIC         SBC        MSEP       FPE       HSP       APC
-------------------------------------------------------------------------------------------------------------------------------
  1        0.5067      0.5051      0.5028    278.9949    427.0016    -423.8990    438.1030    72.0464    0.2426    8e-04    0.4999
  2        0.6566      0.6543      0.6221    106.6276    320.7375    -529.4253    335.5393    50.3303    0.1700    6e-04    0.3504
  3        0.6906      0.6875        -Inf     69.0301    291.5382    -558.4006    310.0405    45.4975    0.1542    5e-04    0.3178
  4        0.7108      0.7068        -Inf     47.5610    273.3831    -576.2830    295.5858    42.6768    0.1451    5e-04    0.2991
  5        0.7286      0.7240        -Inf     28.7654    256.3130    -592.8550    282.2161    40.1770    0.1371    5e-04    0.2825
  6        0.7375      0.7321        -Inf     20.4847    248.4161    -600.4072    278.0196    39.0025    0.1335    4e-04    0.2751
  7        0.7401      0.7339        -Inf     19.4249    247.4061    -601.3006    280.7101    38.7450    0.1330    4e-04    0.2742
  8        0.7472      0.7402        -Inf     13.1766    241.1380    -607.0927    278.1424    37.8187    0.1303    4e-04    0.2685
  9        0.7532      0.7455        -Inf      8.2417    236.0050    -611.7049    276.7099    37.0554    0.1281    4e-04    0.2639
 10        0.7533      0.7448        -Inf     10.0204    237.7746    -609.8359    282.1799    37.1558    0.1288    4e-04    0.2655
 11        0.7534      0.7439        -Inf     12.0000    239.7533    -607.7720    287.8591    37.2831    0.1297    4e-04    0.2673
-------------------------------------------------------------------------------------------------------------------------------
AIC: Akaike Information Criteria
 SBIC: Sawa's Bayesian Information Criteria
 SBC: Schwarz Bayesian Criteria
 MSEP: Estimated error of prediction, assuming multivariate normality
```

Figure A.39 Basic Consumption Goods Physical Market Best Subset Model Result

```
                                          Best Subsets Regression
---------------------------------------------------------------------------------------------------------------------
Model Index    Predictors
---------------------------------------------------------------------------------------------------------------------
     1         DUM24_5
     2         Lnclust5_7 DUM24_5
     3         Lnclust5_7 DUM24_5 DUM1_6
     4         Lnclust5_1 Lnclust5_7 DUM24_5 DUM1_6
     5         Lnclust5_1 Lnclust5_2 Lnclust5_7 DUM24_5 DUM1_6
     6         LnCasenum Lnclust5_1 Lnclust5_2 Lnclust5_7 DUM24_5 DUM1_6
     7         Lnclust5_1 Lnclust5_2 Lnclust5_5 Lnclust5_6 Lnclust5_7 DUM24_5 DUM1_6
     8         Lnclust5_1 Lnclust5_2 Lnclust5_4 Lnclust5_5 Lnclust5_6 Lnclust5_7 DUM24_5 DUM1_6
     9         LnCasenum Lnclust5_1 Lnclust5_2 Lnclust5_4 Lnclust5_5 Lnclust5_6 Lnclust5_7 DUM24_5 DUM1_6
    10         LnCasenum Lnclust5_1 Lnclust5_2 Lnclust5_4 Lnclust5_5 Lnclust5_6 Lnclust5_7 DUM24_5 DUM12_12 DUM1_6
    11         LnCasenum Lnclust5_1 Lnclust5_2 Lnclust5_3 Lnclust5_4 Lnclust5_5 Lnclust5_6 Lnclust5_7 DUM24_5 DUM12_12 DUM1_6
---------------------------------------------------------------------------------------------------------------------

                                                   Subsets Regression Summary
-------------------------------------------------------------------------------------------------------------------------------
                       Adj.        Pred
Model    R-Square    R-Square    R-Square     C(p)        AIC         SBIC         SBC        MSEP       FPE       HSP       APC
-------------------------------------------------------------------------------------------------------------------------------
  1        0.5400      0.5384      0.5365    266.6768    438.2625    -412.5709    449.3638    74.8115    0.2519    8e-04    0.4662
  2        0.6680      0.6658       0.631    112.3563    342.7406    -507.4966    357.5423    54.1737    0.1830    6e-04    0.3387
  3        0.7037      0.7007        -Inf     70.8171    310.7659    -539.2086    329.2681    48.5194    0.1644    6e-04    0.3044
  4        0.7237      0.7199        -Inf     48.3745    291.8523    -557.8363    314.0549    45.3961    0.1544    5e-04    0.2857
  5        0.7410      0.7366        -Inf     29.2322    274.5020    -574.6831    300.4051    42.6970    0.1457    5e-04    0.2696
  6        0.7461      0.7409        -Inf     25.0242    270.5738    -578.4518    300.1774    42.0026    0.1438    5e-04    0.2661
  7        0.7519      0.7459        -Inf     19.9756    265.6974    -583.0381    299.0014    41.1892    0.1414    5e-04    0.2618
  8        0.7600      0.7534        -Inf     12.0382    257.7260    -590.4339    294.7304    39.9761    0.1377    5e-04    0.2549
  9        0.7647      0.7574        -Inf      8.2770    253.7890    -593.9234    294.4939    39.3262    0.1359    5e-04    0.2516
 10        0.7649      0.7568        -Inf     10.0124    255.5135    -592.0964    299.9188    39.4269    0.1367    5e-04    0.2530
 11        0.7649      0.7559        -Inf     12.0000    257.5006    -590.0247    305.6063    39.5630    0.1376    5e-04    0.2547
-------------------------------------------------------------------------------------------------------------------------------
AIC: Akaike Information Criteria
 SBIC: Sawa's Bayesian Information Criteria
 SBC: Schwarz Bayesian Criteria
 MSEP: Estimated error of prediction, assuming multivariate normality
 FPE: Final Prediction Error
 HSP: Hocking's Sp
 APC: Amemiya Prediction Criteria
```

Figure A.40 Cold-Chain Products Physical Market Best Subset Model Result

```
> ols_step_best_subset(model6)
                                              Best Subsets Regression
---------------------------------------------------------------------------------------------------------------------
Model Index    Predictors
---------------------------------------------------------------------------------------------------------------------
     1         DUM24_5
     2         Lnclust6_7 DUM24_5
     3         Lnclust6_1 Lnclust6_7 DUM24_5
     4         Lnclust6_1 Lnclust6_7 DUM24_5 DUM1_6
     5         Lnclust6_1 Lnclust6_2 Lnclust6_7 DUM24_5 DUM1_6
     6         Lnclust6_1 Lnclust6_2 Lnclust6_6 Lnclust6_7 DUM24_5 DUM1_6
     7         Lnclust6_1 Lnclust6_2 Lnclust6_5 Lnclust6_6 Lnclust6_7 DUM24_5 DUM1_6
     8         Lnclust6_1 Lnclust6_2 Lnclust6_4 Lnclust6_5 Lnclust6_6 Lnclust6_7 DUM24_5 DUM1_6
     9         Lnclust6_1 Lnclust6_2 Lnclust6_4 Lnclust6_5 Lnclust6_6 Lnclust6_7 DUM24_5 DUM12_12 DUM1_6
    10         Lnclust6_1 Lnclust6_2 Lnclust6_3 Lnclust6_4 Lnclust6_5 Lnclust6_6 Lnclust6_7 DUM24_5 DUM12_12 DUM1_6
    11         LnCasenum Lnclust6_1 Lnclust6_2 Lnclust6_3 Lnclust6_4 Lnclust6_5 Lnclust6_6 Lnclust6_7 DUM24_5 DUM12_12 DUM1_6
---------------------------------------------------------------------------------------------------------------------

                                              Subsets Regression Summary
-------------------------------------------------------------------------------------------------------------------------------
                     Adj.        Pred
Model    R-Square    R-Square    R-Square     C(p)        AIC         SBIC          SBC         MSEP       FPE       HSP       APC
-------------------------------------------------------------------------------------------------------------------------------
  1        0.4778      0.4761      0.4738    164.1335    433.4939    -416.6697    444.5953    73.6279    0.2479    8e-04    0.5292
  2        0.5781      0.5753      0.5536     77.9563    371.7302    -478.0342    386.5319    59.6892    0.2016    7e-04    0.4304
  3        0.6286      0.6248      0.5868     35.5817    335.6363    -513.5789    354.1385    52.7278    0.1787    6e-04    0.3815
  4        0.6502      0.6455        -Inf     18.5492    319.6829    -529.1219    341.8855    49.8244    0.1694    6e-04    0.3617
  5        0.6661      0.6604        -Inf      6.6091    307.8135    -540.4921    333.7166    47.7289    0.1628    5e-04    0.3476
  6        0.6685      0.6617        -Inf      6.4541    307.6109    -540.5533    337.2144    47.5414    0.1627    5e-04    0.3474
  7        0.6698      0.6619        -Inf      7.3328    308.4582    -539.5911    341.7622    47.5218    0.1632    5e-04    0.3484
  8        0.6734      0.6644        -Inf      6.1514    307.1638    -540.6221    344.1682    47.1637    0.1625    5e-04    0.3469
  9        0.6735      0.6634        -Inf      8.0387    309.0465    -538.6488    349.7513    47.3089    0.1635    5e-04    0.3490
 10        0.6736      0.6622        -Inf     10.0141    311.0209    -536.5891    355.4262    47.4697    0.1646    6e-04    0.3514
 11        0.6736      0.6611         NaN     12.0000    313.0061    -534.5191    361.1119    47.6333    0.1657    6e-04    0.3537
-------------------------------------------------------------------------------------------------------------------------------
AIC: Akaike Information Criteria
 SBIC: Sawa's Bayesian Information Criteria
 SBC: Schwarz Bayesian Criteria
 MSEP: Estimated error of prediction, assuming multivariate normality
```

Figure A.41 Electronics Physical Market Best Subset Model Result

```
                                              Best Subsets Regression
---------------------------------------------------------------------------------------------------------------
Model Index    Predictors
---------------------------------------------------------------------------------------------------------------
     1         DUM24_5
     2         Lnclust2_7 DUM24_5
     3         Lnclust2_7 DUM24_5 DUM1_6
     4         Lnclust2_2 Lnclust2_7 DUM24_5 DUM1_6
     5         Lnclust2_1 Lnclust2_2 Lnclust2_7 DUM24_5 DUM1_6
     6         LnCasenum Lnclust2_1 Lnclust2_2 Lnclust2_7 DUM24_5 DUM1_6
     7         LnCasenum Lnclust2_1 Lnclust2_2 Lnclust2_3 Lnclust2_7 DUM24_5 DUM1_6
     8         LnCasenum Lnclust2_1 Lnclust2_2 Lnclust2_3 Lnclust2_5 Lnclust2_7 DUM24_5 DUM1_6
     9         LnCasenum Lnclust2_1 Lnclust2_2 Lnclust2_3 Lnclust2_5 Lnclust2_6 Lnclust2_7 DUM24_5 DUM1_6
    10         LnCasenum Lnclust2_1 Lnclust2_2 Lnclust2_3 Lnclust2_5 Lnclust2_6 Lnclust2_7 DUM24_5 DUM12_12 DUM1_6
    11         LnCasenum Lnclust2_1 Lnclust2_2 Lnclust2_3 Lnclust2_4 Lnclust2_5 Lnclust2_6 Lnclust2_7 DUM24_5 DUM12_12 DUM1_6
---------------------------------------------------------------------------------------------------------------

                                              Subsets Regression Summary
--------------------------------------------------------------------------------------------------------------------------------
                     Adj.        Pred
Model    R-Square    R-Square    R-Square     C(p)        AIC         SBIC          SBC         MSEP        FPE       HSP       APC
--------------------------------------------------------------------------------------------------------------------------------
  1        0.7497      0.7489      0.6916    335.7012    538.1241    -313.0580    549.2254    104.4760    0.3518    0.0012    0.2537
  2        0.8004      0.7990      0.7182    210.0760    472.5229    -378.7654    487.3247     83.6173    0.2825     9e-04    0.2037
  3        0.8290      0.8273        -Inf    139.8765    428.2019    -422.9592    446.7041     71.8605    0.2435     8e-04    0.1756
  4        0.8433      0.8412        -Inf    105.7866    404.0466    -447.0227    426.2492     66.0662    0.2246     8e-04    0.1620
  5        0.8571      0.8547        -Inf     73.0372    378.4972    -472.1198    404.4003     60.4573    0.2062     7e-04    0.1487
  6        0.8678      0.8651        -Inf     48.1129    357.2572    -492.7227    386.8607     56.1284    0.1921     6e-04    0.1385
  7        0.8757      0.8727        -Inf     30.2392    340.8644    -508.3923    374.1684     52.9618    0.1819     6e-04    0.1311
  8        0.8803      0.8770        -Inf     20.7095    331.6512    -517.0358    368.6556     51.1889    0.1763     6e-04    0.1272
  9        0.8857      0.8821        -Inf      9.0964    319.8465    -527.9246    360.5514     49.0490    0.1695     6e-04    0.1222
 10        0.8860      0.8821        -Inf     10.2346    320.9508    -526.6767    365.3561     49.0726    0.1701     6e-04    0.1227
 11        0.8861      0.8817        -Inf     12.0000    322.7064    -524.8188    370.8122     49.2040    0.1711     6e-04    0.1234
--------------------------------------------------------------------------------------------------------------------------------
AIC: Akaike Information Criteria
 SBIC: Sawa's Bayesian Information Criteria
 SBC: Schwarz Bayesian Criteria
 MSEP: Estimated error of prediction, assuming multivariate normality
```

Figure A.42 Daily Needs Online Market Best Subset Model Result

```
> ols_step_best_subset(model3)
                                                  Best Subsets Regression
-----------------------------------------------------------------------------------------------------------------------
Model Index    Predictors
-----------------------------------------------------------------------------------------------------------------------
     1         DUM24_5
     2         Lnclust3_7 DUM24_5
     3         Lnclust3_7 DUM24_5 DUM1_6
     4         Lnclust3_2 Lnclust3_7 DUM24_5 DUM1_6
     5         Lnclust3_1 Lnclust3_2 Lnclust3_7 DUM24_5 DUM1_6
     6         LnCasenum Lnclust3_1 Lnclust3_2 Lnclust3_7 DUM24_5 DUM1_6
     7         LnCasenum Lnclust3_1 Lnclust3_2 Lnclust3_3 Lnclust3_7 DUM24_5 DUM1_6
     8         Lnclust3_1 Lnclust3_2 Lnclust3_3 Lnclust3_5 Lnclust3_6 Lnclust3_7 DUM24_5 DUM1_6
     9         LnCasenum Lnclust3_1 Lnclust3_2 Lnclust3_3 Lnclust3_5 Lnclust3_6 Lnclust3_7 DUM24_5 DUM1_6
    10         LnCasenum Lnclust3_1 Lnclust3_2 Lnclust3_3 Lnclust3_4 Lnclust3_5 Lnclust3_6 Lnclust3_7 DUM24_5 DUM1_6
    11         LnCasenum Lnclust3_1 Lnclust3_2 Lnclust3_3 Lnclust3_4 Lnclust3_5 Lnclust3_6 Lnclust3_7 DUM24_5 DUM12_12 DUM1_6
-----------------------------------------------------------------------------------------------------------------------

                                                   Subsets Regression Summary
------------------------------------------------------------------------------------------------------------------------------
                       Adj.        Pred
Model    R-Square    R-Square    R-Square     C(p)        AIC         SBIC         SBC         MSEP       FPE       HSP       APC
------------------------------------------------------------------------------------------------------------------------------
  1        0.7127      0.7117      0.6717    324.9936    561.7892    -289.3431    572.8906    113.0811    0.3807    0.0013    0.2912
  2        0.7778      0.7763      0.7084    186.3787    486.8810    -364.1851    501.6827     87.7306    0.2964    0.0010    0.2267
  3        0.8131      0.8112        -Inf    112.2527    437.1756    -413.5515    455.6779     74.0499    0.2510     8e-04    0.1920
  4        0.8238      0.8214        -Inf     91.1316    421.5209    -429.2297    443.7236     70.0424    0.2382     8e-04    0.1822
  5        0.8444      0.8418        -Inf     48.6369    386.2991    -463.5598    412.2022     62.0556    0.2117     7e-04    0.1619
  6        0.8502      0.8472        -Inf     38.1175    376.9330    -472.6484    406.5365     59.9462    0.2052     7e-04    0.1569
  7        0.8542      0.8507        -Inf     31.5677    370.9147    -478.4072    404.2187     58.5613    0.2011     7e-04    0.1538
  8        0.8604      0.8565        -Inf     20.2871    360.0172    -488.6448    397.0216     56.2830    0.1939     7e-04    0.1483
  9        0.8658      0.8616        -Inf     10.5957    350.1834    -497.6945    390.8883     54.2868    0.1876     6e-04    0.1435
 10        0.8667      0.8621        -Inf     10.5568    350.0708    -497.5821    394.4762     54.0924    0.1875     6e-04    0.1435
 11        0.8670      0.8619        -Inf     12.0000    351.4914    -496.0339    399.5971     54.1764    0.1884     6e-04    0.1441
------------------------------------------------------------------------------------------------------------------------------
AIC: Akaike Information Criteria
 SBIC: Sawa's Bayesian Information Criteria
 SBC: Schwarz Bayesian Criteria
```

Figure A.43 Protein-Based Products Online Market Best Subset Model Result

```
> ols_step_best_subset(model4)
                                                  Best Subsets Regression
-----------------------------------------------------------------------------------------------------------------------
Model Index    Predictors
-----------------------------------------------------------------------------------------------------------------------
     1         DUM24_5
     2         Lnclust4_7 DUM24_5
     3         Lnclust4_7 DUM24_5 DUM1_6
     4         Lnclust4_2 Lnclust4_7 DUM24_5 DUM1_6
     5         Lnclust4_1 Lnclust4_2 Lnclust4_7 DUM24_5 DUM1_6
     6         Lnclust4_1 Lnclust4_2 Lnclust4_3 Lnclust4_7 DUM24_5 DUM1_6
     7         Lnclust4_1 Lnclust4_2 Lnclust4_3 Lnclust4_5 Lnclust4_7 DUM24_5 DUM1_6
     8         Lnclust4_1 Lnclust4_2 Lnclust4_3 Lnclust4_5 Lnclust4_6 Lnclust4_7 DUM24_5 DUM1_6
     9         LnCasenum Lnclust4_1 Lnclust4_2 Lnclust4_3 Lnclust4_5 Lnclust4_6 Lnclust4_7 DUM24_5 DUM1_6
    10         LnCasenum Lnclust4_1 Lnclust4_2 Lnclust4_3 Lnclust4_5 Lnclust4_6 Lnclust4_7 DUM24_5 DUM12_12 DUM1_6
    11         LnCasenum Lnclust4_1 Lnclust4_2 Lnclust4_3 Lnclust4_4 Lnclust4_5 Lnclust4_6 Lnclust4_7 DUM24_5 DUM12_12 DUM1_6
-----------------------------------------------------------------------------------------------------------------------

                                                   Subsets Regression Summary
------------------------------------------------------------------------------------------------------------------------------
                       Adj.        Pred
Model    R-Square    R-Square    R-Square     C(p)        AIC         SBIC         SBC         MSEP       FPE       HSP       APC
------------------------------------------------------------------------------------------------------------------------------
  1        0.7376      0.7368      0.6722    285.2464    576.5881    -274.3458    587.6894    118.8188    0.4000    0.0013    0.2659
  2        0.7867      0.7853      0.6955    178.7557    516.6976    -334.2930    531.4994     96.9303    0.3274    0.0011    0.2176
  3        0.8166      0.8148        -Inf    114.5407    473.4768    -377.2881    491.9791     83.6087    0.2834    0.0010    0.1883
  4        0.8343      0.8321        -Inf     77.3998    445.1295    -405.3029    467.3322     75.7970    0.2577     9e-04    0.1713
  5        0.8490      0.8464        -Inf     47.0677    419.5074    -430.2995    445.4105     69.3451    0.2366     8e-04    0.1572
  6        0.8592      0.8563        -Inf     26.3680    400.4674    -448.6171    430.0710     64.8553    0.2220     7e-04    0.1475
  7        0.8628      0.8595        -Inf     20.3865    394.7029    -454.0541    428.0069     63.4107    0.2177     7e-04    0.1447
  8        0.8680      0.8643        -Inf     11.0428    385.3089    -462.7887    422.3133     61.2510    0.2110     7e-04    0.1402
  9        0.8699      0.8659        -Inf      8.6527    382.7801    -464.9593    423.4850     60.5397    0.2092     7e-04    0.1391
 10        0.8702      0.8657        -Inf     10.0421    384.1448    -463.4674    428.5501     60.6217    0.2102     7e-04    0.1397
 11        0.8702      0.8653         NaN     12.0000    386.1009    -461.4244    434.2067     60.8247    0.2116     7e-04    0.1406
------------------------------------------------------------------------------------------------------------------------------
AIC: Akaike Information Criteria
 SBIC: Sawa's Bayesian Information Criteria
 SBC: Schwarz Bayesian Criteria
 MSEP: Estimated error of prediction, assuming multivariate normality
```

Figure A.44 Basic Consumption Goods Online Market Best Subset Model Result

```
> ols_step_best_subset(model5)
                                              Best Subsets Regression
------------------------------------------------------------------------------------------------------------------
Model Index    Predictors
------------------------------------------------------------------------------------------------------------------
     1         DUM24_5
     2         Lnclust5_7 DUM24_5
     3         Lnclust5_7 DUM24_5 DUM1_6
     4         Lnclust5_2 Lnclust5_7 DUM24_5 DUM1_6
     5         Lnclust5_1 Lnclust5_2 Lnclust5_7 DUM24_5 DUM1_6
     6         LnCasenum Lnclust5_1 Lnclust5_2 Lnclust5_7 DUM24_5 DUM1_6
     7         LnCasenum Lnclust5_1 Lnclust5_2 Lnclust5_5 Lnclust5_7 DUM24_5 DUM1_6
     8         Lnclust5_1 Lnclust5_2 Lnclust5_3 Lnclust5_5 Lnclust5_6 Lnclust5_7 DUM24_5 DUM1_6
     9         LnCasenum Lnclust5_1 Lnclust5_2 Lnclust5_3 Lnclust5_5 Lnclust5_6 Lnclust5_7 DUM24_5 DUM1_6
    10         LnCasenum Lnclust5_1 Lnclust5_2 Lnclust5_3 Lnclust5_4 Lnclust5_5 Lnclust5_6 Lnclust5_7 DUM24_5 DUM1_6
    11         LnCasenum Lnclust5_1 Lnclust5_2 Lnclust5_3 Lnclust5_4 Lnclust5_5 Lnclust5_6 Lnclust5_7 DUM24_5 DUM12_12 DUM1_6
------------------------------------------------------------------------------------------------------------------

                                                Subsets Regression Summary
-------------------------------------------------------------------------------------------------------------------------
                       Adj.        Pred
Model    R-Square    R-Square    R-Square     C(p)        AIC        SBIC         SBC         MSEP      FPE       HSP       APC
-------------------------------------------------------------------------------------------------------------------------
  1       0.7454      0.7446      0.7132    321.5541    544.2457    -306.8703    555.3470    106.6370    0.3590    0.0012    0.2580
  2       0.7979      0.7965      0.7403    196.5337    477.2686    -373.8949    492.0704     84.9551    0.2870    0.0010    0.2062
  3       0.8286      0.8269        -Inf    124.0521    429.9188    -420.9997    448.4210     72.2743    0.2449     8e-04    0.1760
  4       0.8404      0.8382        -Inf     97.5039    410.6115    -440.2803    432.8141     67.5328    0.2296     8e-04    0.1650
  5       0.8586      0.8562        -Inf     55.3479    376.3385    -473.7383    402.2416     60.0224    0.2048     7e-04    0.1471
  6       0.8652      0.8624        -Inf     41.5561    364.2179    -485.5029    393.8215     57.4504    0.1966     7e-04    0.1413
  7       0.8689      0.8658        -Inf     34.4316    357.7445    -491.7165    391.0485     56.0378    0.1924     6e-04    0.1383
  8       0.8738      0.8703        -Inf     24.6537    348.4394    -500.4786    385.4438     54.1453    0.1865     6e-04    0.1340
  9       0.8804      0.8766        -Inf     10.7101    334.4214    -513.4646    375.1263     51.4991    0.1780     6e-04    0.1279
 10       0.8814      0.8773        -Inf     10.3012    333.9249    -513.7079    378.3302     51.2489    0.1777     6e-04    0.1277
 11       0.8815      0.8770        -Inf     12.0000    335.6112    -511.9141    383.7170     51.3741    0.1787     6e-04    0.1284
-------------------------------------------------------------------------------------------------------------------------
AIC: Akaike Information Criteria
 SBIC: Sawa's Bayesian Information Criteria
 SBC: Schwarz Bayesian Criteria
```

Figure A.45 Cold-Chain Products Online Market Best Subset Model Result

```
> ols_step_best_subset(model6)
                                              Best Subsets Regression
-------------------------------------------------------------------------------------------------------------------
Model Index    Predictors
-------------------------------------------------------------------------------------------------------------------
     1         Lnclust6_1
     2         Lnclust6_7 DUM24_5
     3         Lnclust6_1 Lnclust6_7 DUM24_5
     4         Lnclust6_1 Lnclust6_7 DUM24_5 DUM1_6
     5         Lnclust6_1 Lnclust6_2 Lnclust6_7 DUM24_5 DUM1_6
     6         LnCasenum Lnclust6_1 Lnclust6_2 Lnclust6_7 DUM24_5 DUM1_6
     7         LnCasenum Lnclust6_1 Lnclust6_2 Lnclust6_6 Lnclust6_7 DUM24_5 DUM1_6
     8         LnCasenum Lnclust6_1 Lnclust6_2 Lnclust6_5 Lnclust6_6 Lnclust6_7 DUM24_5 DUM1_6
     9         LnCasenum Lnclust6_1 Lnclust6_2 Lnclust6_5 Lnclust6_6 Lnclust6_7 DUM24_5 DUM12_12 DUM1_6
    10         LnCasenum Lnclust6_1 Lnclust6_2 Lnclust6_4 Lnclust6_5 Lnclust6_6 Lnclust6_7 DUM24_5 DUM12_12 DUM1_6
    11         LnCasenum Lnclust6_1 Lnclust6_2 Lnclust6_3 Lnclust6_4 Lnclust6_5 Lnclust6_6 Lnclust6_7 DUM24_5 DUM12_12 DUM1_6
-------------------------------------------------------------------------------------------------------------------

                                                Subsets Regression Summary
--------------------------------------------------------------------------------------------------------------------------
                       Adj.        Pred
Model    R-Square    R-Square    R-Square     C(p)        AIC        SBIC         SBC         MSEP      FPE       HSP       APC
--------------------------------------------------------------------------------------------------------------------------
 1        0.3513      0.3491      0.3242    252.4205    760.7551     -89.9972    771.8564    219.9813    0.7406    0.0025    0.6575
 2        0.4977      0.4943      0.4773    130.8885    686.2868    -164.1798    701.0886    170.9173    0.5774    0.0019    0.5125
 3        0.5838      0.5796      0.5505     60.1719    632.0166    -217.7413    650.5189    142.0787    0.4815    0.0016    0.4274
 4        0.6178      0.6126        -Inf     33.5413    608.5883    -240.6775    630.7910    130.9406    0.4452    0.0015    0.3952
 5        0.6338      0.6276        -Inf     21.9994    597.7637    -251.1520    623.6668    125.8726    0.4294    0.0014    0.3812
 6        0.6508      0.6437        -Inf      9.6386    585.5343    -262.7842    615.1378    120.4351    0.4122    0.0014    0.3659
 7        0.6556      0.6473        -Inf      7.5990    583.4065    -264.6578    616.7105    119.1935    0.4093    0.0014    0.3633
 8        0.6588      0.6494        -Inf      6.9339    582.6517    -265.1847    619.6561    118.5090    0.4082    0.0014    0.3624
 9        0.6598      0.6492        -Inf      8.0773    583.7609    -263.9372    624.4658    118.5667    0.4098    0.0014    0.3637
10        0.6599      0.6481        -Inf     10.0016    585.6820    -261.9270    630.0873    118.9485    0.4124    0.0014    0.3661
11        0.6599      0.6469        -Inf     12.0000    587.6804    -259.8449    635.7861    119.3637    0.4152    0.0014    0.3686
--------------------------------------------------------------------------------------------------------------------------
AIC: Akaike Information Criteria
 SBIC: Sawa's Bayesian Information Criteria
 SBC: Schwarz Bayesian Criteria
 MSEP: Estimated error of prediction, assuming multivariate normality
```

Figure A.46 Electronics Online Market Best Subset Model Result

FINAL MODEL RESULTS

```
   Min      1Q  Median      3Q     Max
-7.2574 -0.0682  0.0478  0.1339  5.7498

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.94932    0.77248   5.113 5.82e-07 ***
LnCasenum    -0.08421    0.03550  -2.372   0.0183 *
Lnclust1_1    0.36697    0.03475  10.559  < 2e-16 ***
Lnclust1_2   -0.24331    0.03784  -6.431 5.29e-10 ***
Lnclust1_3    0.05558    0.03905   1.423   0.1557
Lnclust1_4    0.07283    0.04001   1.820   0.0698 .
Lnclust1_5   -0.23366    0.04153  -5.627 4.37e-08 ***
Lnclust1_6    0.23016    0.04313   5.336 1.93e-07 ***
Lnclust1_7    0.48577    0.03728  13.029  < 2e-16 ***
DUM24_5     -10.03756    0.59815 -16.781  < 2e-16 ***
DUM12_12    -12.51643    0.96155 -13.017  < 2e-16 ***
DUM1_6        9.36646    1.08293   8.649 3.75e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure A.47 Alcoholic Beverages Physical Market Final Model Result

```
Residuals:
     Min       1Q   Median       3Q      Max
-2.15930 -0.05216  0.01137  0.06492  1.21257

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.22000    0.93676   4.505 9.67e-06 ***
LnCasenum   -0.01629    0.01129  -1.442  0.15035
Lnclust2_1   0.29661    0.04223   7.023 1.57e-11 ***
Lnclust2_2  -0.18145    0.04487  -4.044 6.75e-05 ***
Lnclust2_3   0.03032    0.04593   0.660  0.50972
Lnclust2_4   0.08434    0.04630   1.822  0.06956 .
Lnclust2_5  -0.13263    0.04714  -2.813  0.00524 **
Lnclust2_6   0.14605    0.04674   3.125  0.00196 **
Lnclust2_7   0.48271    0.03993  12.090  < 2e-16 ***
DUM24_5     -2.87985    0.20224 -14.239  < 2e-16 ***
DUM12_12    -0.18367    0.31659  -0.580  0.56228
DUM1_6       2.51723    0.35251   7.141 7.64e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure A.48 Daily Needs Physical Market Final Model Result

```
Residuals:
     Min       1Q   Median       3Q      Max
-2.20907 -0.08616 -0.00522  0.11327  1.56468

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.974998   0.840587   8.298 4.19e-15 ***
LnCasenum   -0.030281   0.013279  -2.280  0.02331 *
Lnclust3_1   0.259949   0.040047   6.491 3.73e-10 ***
Lnclust3_2  -0.185242   0.041929  -4.418 1.41e-05 ***
Lnclust3_3  -0.001608   0.042705  -0.038  0.96999
Lnclust3_4   0.112936   0.043186   2.615  0.00939 **
Lnclust3_5  -0.203428   0.045007  -4.520 9.06e-06 ***
Lnclust3_6   0.164681   0.045789   3.597  0.00038 ***
Lnclust3_7   0.342701   0.036618   9.359  < 2e-16 ***
DUM24_5     -5.055043   0.255976 -19.748  < 2e-16 ***
DUM12_12    -0.211549   0.376994  -0.561  0.57514
DUM1_6       3.206452   0.458191   6.998 1.84e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure A.49 Protein-Based Products Physical Market Final Model Result

```
Call:
lm(formula = Lnclust4 ~ ., data = denemeeC4)

Residuals:
     Min       1Q   Median       3Q      Max
-2.10484 -0.06375  0.00014  0.09591  1.37823

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.111703   0.980449   7.254 3.80e-12 ***
LnCasenum   -0.038233   0.012847  -2.976 0.003168 **
Lnclust4_1   0.266468   0.041211   6.466 4.32e-10 ***
Lnclust4_2  -0.184126   0.043141  -4.268 2.68e-05 ***
Lnclust4_3  -0.006281   0.043976  -0.143 0.886520
Lnclust4_4   0.102171   0.044541   2.294 0.022520 *
Lnclust4_5  -0.183306   0.045945  -3.990 8.40e-05 ***
Lnclust4_6   0.157594   0.046314   3.403 0.000762 ***
Lnclust4_7   0.383103   0.038052  10.068  < 2e-16 ***
DUM24_5     -4.200313   0.238334 -17.624  < 2e-16 ***
DUM12_12    -0.168354   0.358350  -0.470 0.638852
DUM1_6       2.994115   0.423527   7.069 1.19e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure A.50 Basic Consumption Goods Physical Market Final Model Result

```
Call:
lm(formula = Lnclust5 ~ ., data = denemeeC5)

Residuals:
     Min       1Q   Median       3Q      Max
-2.16971 -0.12371  0.01222  0.12324  1.64196

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.591133   0.889168   7.413 1.40e-12 ***
LnCasenum   -0.030444   0.013020  -2.338 0.020061 *
Lnclust5_1   0.269870   0.039800   6.781 6.81e-11 ***
Lnclust5_2  -0.183991   0.041836  -4.398 1.54e-05 ***
Lnclust5_3  -0.004747   0.042643  -0.111 0.911446
Lnclust5_4   0.114220   0.043143   2.647 0.008558 **
Lnclust5_5  -0.204178   0.044875  -4.550 7.93e-06 ***
Lnclust5_6   0.163787   0.045491   3.600 0.000374 ***
Lnclust5_7   0.371445   0.036800  10.094  < 2e-16 ***
DUM24_5     -4.709145   0.245583 -19.175  < 2e-16 ***
DUM12_12    -0.189945   0.369071  -0.515 0.607188
DUM1_6       3.399490   0.443677   7.662 2.83e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure A.51 Cold-Chain Products Physical Market Final Model Result

```
     Min       1Q   Median       3Q      Max
-2.00948 -0.14588 -0.02361  0.14844  1.28093

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.19186    0.56974   7.357 1.91e-12 ***
Lnclust6_1   0.36515    0.04735   7.712 1.97e-13 ***
Lnclust6_2  -0.15418    0.04181  -3.687  0.00027 ***
Lnclust6_6   0.06268    0.04266   1.469  0.14281
Lnclust6_7   0.30210    0.04396   6.873 3.81e-11 ***
DUM24_5     -3.62236    0.26588 -13.624  < 2e-16 ***
DUM1_6       2.06296    0.44853   4.599 6.32e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure A.52 Electronics Physical Market Final Model Result

```
Residuals:
     Min       1Q   Median       3Q      Max 
-2.10460 -0.07340  0.02916  0.15997  1.59464 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  8.28224    0.52533  15.766  < 2e-16 ***
LnCasenum    0.07459    0.01582   4.714 3.80e-06 ***
Lnclust2_1   0.22889    0.03579   6.395 6.49e-10 ***
Lnclust2_2  -0.37918    0.03689 -10.280  < 2e-16 ***
Lnclust2_3   0.17105    0.03910   4.375 1.70e-05 ***
Lnclust2_4   0.01915    0.03953   0.484 0.628477    
Lnclust2_5  -0.16515    0.03954  -4.177 3.92e-05 ***
Lnclust2_6   0.13807    0.03713   3.719 0.000241 ***
Lnclust2_7   0.25181    0.03199   7.871 7.24e-14 ***
DUM24_5     -9.71935    0.33719 -28.824  < 2e-16 ***
DUM12_12     0.37904    0.40856   0.928 0.354327    
DUM1_6       4.42341    0.52633   8.404 2.03e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure A.53 Daily Needs Online Market Final Model Result

```
Residuals:
     Min       1Q   Median       3Q      Max 
-2.26339 -0.10705  0.00018  0.17078  1.59308 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.88227    0.52273  13.166  < 2e-16 ***
LnCasenum    0.05259    0.01654   3.180  0.00163 ** 
Lnclust3_1   0.26290    0.03703   7.099 9.87e-12 ***
Lnclust3_2  -0.33695    0.03847  -8.758  < 2e-16 ***
Lnclust3_3   0.11268    0.04028   2.797  0.00550 ** 
Lnclust3_4   0.05814    0.04074   1.427  0.15467    
Lnclust3_5  -0.19793    0.04119  -4.805 2.50e-06 ***
Lnclust3_6   0.15940    0.03912   4.074 5.98e-05 ***
Lnclust3_7   0.28866    0.03313   8.712 2.42e-16 ***
DUM24_5     -8.75795    0.33863 -25.863  < 2e-16 ***
DUM12_12     0.31999    0.42885   0.746  0.45618    
DUM1_6       5.05109    0.54243   9.312  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure A.54 Protein-Based Products Online Market Final Model Result

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.52205    0.58223  14.637  < 2e-16 ***
LnCasenum     0.03354    0.01678   1.999 0.046541 *
Lnclust4_1    0.25291    0.03782   6.687 1.19e-10 ***
Lnclust4_2   -0.39198    0.03934  -9.964  < 2e-16 ***
Lnclust4_3    0.18274    0.04178   4.374 1.71e-05 ***
Lnclust4_4    0.00865    0.04214   0.205 0.837523
Lnclust4_5   -0.15094    0.04207  -3.588 0.000391 ***
Lnclust4_6    0.13106    0.03945   3.322 0.001010 **
Lnclust4_7    0.26859    0.03418   7.857 7.92e-14 ***
DUM24_5      -9.89079    0.37355 -26.478  < 2e-16 ***
DUM12_12      0.35507    0.45436   0.781 0.435174
DUM1_6        4.82857    0.57952   8.332 3.32e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure A.55 Basic Consumption Goods Online Market Final Model Result

```
Residuals:
     Min       1Q   Median       3Q      Max
-2.28393 -0.07520  0.01693  0.13591  1.47175

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.81728    0.52541  14.878  < 2e-16 ***
LnCasenum    0.06015    0.01614   3.726 0.000235 ***
Lnclust5_1   0.23830    0.03487   6.833 4.98e-11 ***
Lnclust5_2  -0.32138    0.03589  -8.954  < 2e-16 ***
Lnclust5_3   0.09473    0.03747   2.528 0.012005 *
Lnclust5_4   0.05906    0.03803   1.553 0.121513
Lnclust5_5  -0.21260    0.03873  -5.489 8.91e-08 ***
Lnclust5_6   0.16577    0.03698   4.482 1.07e-05 ***
Lnclust5_7   0.25960    0.03069   8.457 1.41e-15 ***
DUM24_5     -9.40371    0.33052 -28.451  < 2e-16 ***
DUM12_12     0.22916    0.41752   0.549 0.583533
DUM1_6       5.11184    0.53767   9.507  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure A.56 Cold-Chain Products Online Market Final Model Result

```
Residuals:
     Min       1Q   Median       3Q      Max 
-2.78324 -0.27371 -0.00594  0.25664  2.48145 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.527323   0.375197   4.071 6.07e-05 ***
LnCasenum    0.087545   0.025920   3.378 0.000832 ***
Lnclust6_1   0.433464   0.049904   8.686 2.90e-16 ***
Lnclust6_2  -0.183326   0.053650  -3.417 0.000725 ***
Lnclust6_3   0.002154   0.054126   0.040 0.968285    
Lnclust6_4   0.011608   0.054172   0.214 0.830479    
Lnclust6_5  -0.083583   0.055200  -1.514 0.131079    
Lnclust6_6   0.144355   0.055259   2.612 0.009466 ** 
Lnclust6_7   0.321085   0.046245   6.943 2.56e-11 ***
DUM24_5     -4.160791   0.406980 -10.224  < 2e-16 ***
DUM12_12     0.589596   0.636405   0.926 0.354991    
DUM1_6       4.066505   0.706130   5.759 2.18e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure A.57 Electronics Online Market Final Model Result