

Article

# An Unsupervised Machine Learning-Based Framework for Transferring Local Factories into Supply Chain Networks

Mohd Fahmi Bin Mad Ali, Mohd Khairol Anuar Bin Mohd Ariffin \* , Faizal Bin Mustapha and Eris Elianddy Bin Supeni 

Department of Mechanical and Manufacturing Engineering, Universiti Putra Malaysia, Serdang 43400, Malaysia; madali.fahmi@gmail.com (M.F.B.M.A.); faizalms@upm.edu.my (F.B.M.); eris@upm.edu.my (E.E.B.S.)

\* Correspondence: Khairol@upm.edu.my

**Abstract:** Transferring a local manufacturing company to a national-wide supply chain network with wholesalers and retailers is a significant problem in manufacturing systems. In this research, a hybrid PCA-K-means is used to transfer a local chocolate manufacturing firm near Kuala Lumpur into a national-wide supply chain. For this purpose, the appropriate locations of the wholesaler's center points were found according to the geographical and population features of the markets in Malaysia. To this end, four wholesalers on the left island of Malaysia are recognized, which were located in the north area, right area, middle area, and south area. Similarly, two wholesalers were identified on the right island, which were in Sarawak and WP Labuan. In order to evaluate the performance of the proposed method, its outcomes are compared with other unsupervised-learning methods such as the WARD and CLINK methods. The outcomes indicated that K-means could successfully determine the best locations for the wholesalers in the supply chain network with a higher score (0.812).

**Keywords:** food supply chain; food distribution; design supply chain; unsupervised machine learning



**Citation:** Mad Ali, M.F.B.; Ariffin, M.K.A.B.M.; Mustapha, F.B.; Supeni, E.E.B. An Unsupervised Machine Learning-Based Framework for Transferring Local Factories into Supply Chain Networks. *Mathematics* **2021**, *9*, 3114. <https://doi.org/10.3390/math9233114>

Academic Editors: Andrea Prati, Carlos A. Iglesias, Luis Javier García Villalba and Vincent A. Ciciarello

Received: 28 September 2021  
Accepted: 11 November 2021  
Published: 3 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The food industry plays a significant role in the economy of Malaysia. A total of 35.9% of the Malaysian economy in 2020 belongs to this industry sector (Figure 1) [1]. Besides, a well-known trade Center in Europe reports that the Food and Beverage sector in Malaysia is growing on an annual average of 7.6% and is the source of €22.12 billion of Malaysian GDP in 2018 [2].

The question of how to expand a local business into a national and international one is an excellent issue to be addressed. There are many reasons to expand a local factory, but perhaps, the main reason for expanding a business is that the larger company will be more reliable, safer, and earn more profits than smaller ones. Many factors must be taken into consideration while expanding a local business. Each of the factors can help grow a newly designed business or play an adverse role.

Besides, ignoring the investments that cause growth for any local industry may cause bankruptcy. Figure 2 indicates the number of bankrupt companies in Malaysia annually [3].

The Statista website listed the reasons for the bankruptcy of Malaysian companies (Figure 3).

Therefore, transforming a local food industry into a supply chain in Malaysia is a significant problem to be addressed.

The main questions are: what are the most important factors in designing and scheduling while transferring a local food company to a supply chain? Are there any effective personal, geographical, urban factors that can increase the performance of the newly designed supply chain?

How should a local manufacturing system be designed according to the identified factors?

Food products must be distributed to the market according to some factors. Such factors may be varied depending on individual-social characteristics in different countries,

such as population, age, gender, underlying diseases, distance from target market to the source of supply (geographical location) and similar factors (which will be addressed in this study). It seems that such factors can directly impact how food must be distributed to destination markets that require it.

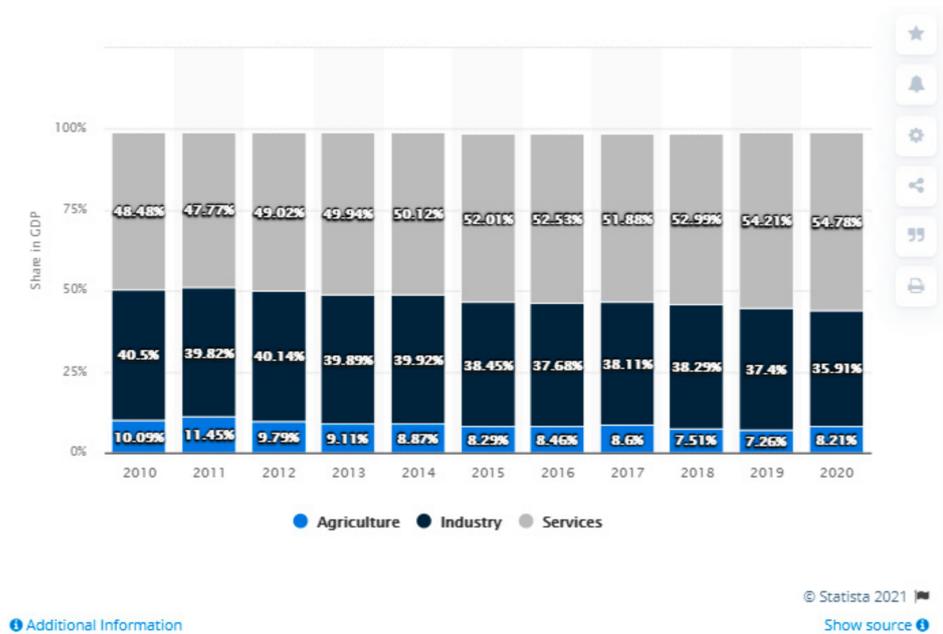


Figure 1. Malaysian economy by sector.

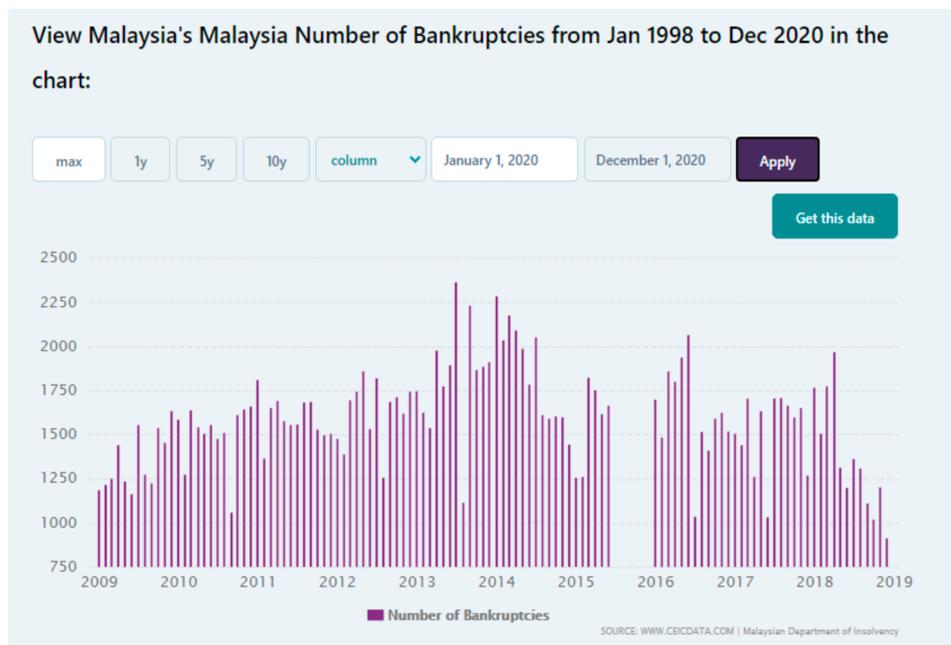
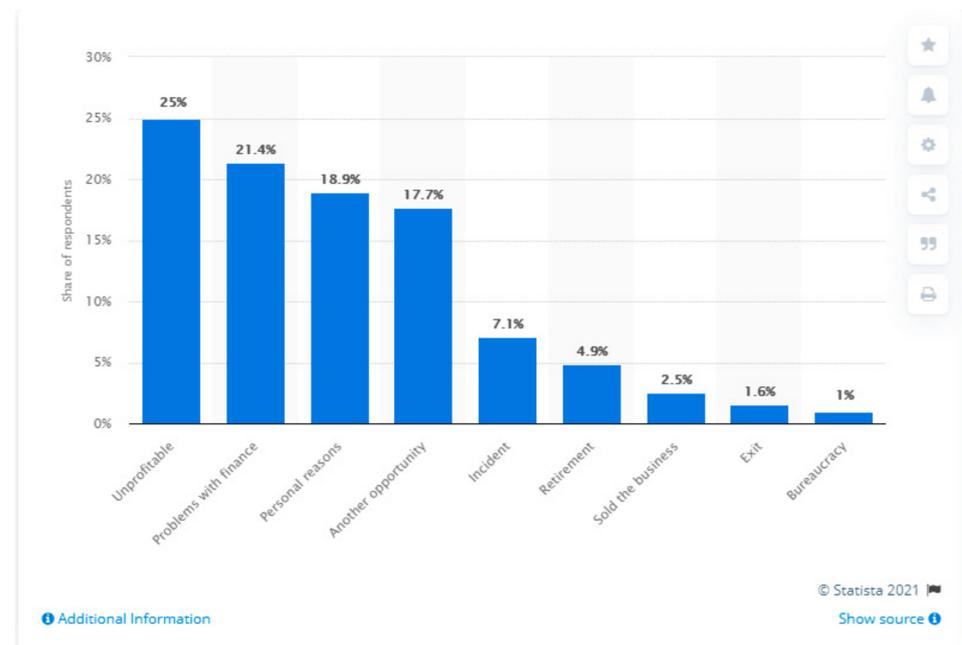


Figure 2. Number of bankrupt companies in Malaysia.



**Figure 3.** Reasons for Malaysian entrepreneurs to discontinue their businesses as of June 2017.

This research aims to propose a machine-learning platform for computer systems that can determine the product clusters according to the end-users features and identify the best locations for the wholesalers' center points based on the geographical and population features. Moreover, the proposed method must provide the best product distribution for the designed supply chain to maximize profit and minimize product transferring costs.

Therefore, proposing a method that can provide a pattern to recognize the best amount of food distribution in Malaysia is crucial and can help develop the business more successfully, specifically during the early stages after transformation.

We identify adequate food consumption and distribution factors in food supply chains and propose a machine learning-based algorithm to predict the best food delivery schedule using effective features in food supply chains.

Designing new supply chains based on real features of the consumers, geographical and population data can increase the chance of success when transferring from a local business into a national one. Failing to distribute and transfer enough products in the food sector can cause losing the market share and impose adverse effects on the supply chain.

As mentioned before, the markets can be clustered according to product consumption and product distribution features.

Therefore, proposing a method that can recognize the pattern of food distribution in a country according to the features of the society is vital and can enhance the chance of success in the new environment.

The outcomes of this research will be helpful for the top managers of a local food factory to transfer their local business to a national or international supply chain according to the accurate data of customers, geographical, and population features.

Some benefits of using the proposing machine learning method will be:

- To determine the most important features that influence the product consumption pattern;
- To identify the most important features that influence the product distribution (design new supply chain);
- To use machine learning for transferring a local business to a national or international business for the first time;
- To determine the best product distribution in Malaysia according to accurate data gathered from the society.

## 2. Literature Review

In this section, 145 research studies are reviewed. The aim of this section is twofold. Firstly, to find out whether the topic of this research has been done before and; secondly, to discuss the promising methods that can be used for this research. At the end of the section, a statistical analysis will be conducted to select the best method and discuss the gaps.

During the last two decades, supply chain management has been taking into consideration by scientists due to its high advantage.

### 2.1. Supply Chain Management

In simple words, a classic supply chain consists of a central factory, several wholesalers and several retailers. Supply chains can forecast and fulfil various markets' demands more effectively using a hierarchical series of processes. With the growth of the market needs during the last decades, new technologies have emerged to speed up the fulfilment of customers' needs and, thus, many new methods for modeling and scheduling of supply chains from different points of view are proposed. In continuation, several important references related to this research's problem statement are opted and discussed.

A critical problem in supply chains is the issue of product transportation. Product transportation can play a crucial role in satisfying the needs of a market on time and, at the same time, it can decrease (or increase) the system costs through the transportation cost.

### 2.2. Designing Supply Chain Networks

The main aim of this research is to transfer a local food factory to a national supply chain. For this purpose, several related research studies are opted and reviewed. Fornasiero et al., (2015) [4] argued that using historical data of a supply chain can help customize it via a discrete-event simulation. Şen (2008) [5] focused on the advantages of positive communications between manufacturers and retailers in the supply chain's performance. Cultural Issues must be considered important factors in designing supply chains (Montagna, 2015) [6] and (Delgado & Albuquerque, 2015) [7]. Iannone et al., (2015) [8] focused on correlations between factors in integrating the retailer's network. By carrying out statistical analysis on the information of 132 Italian factories producing fashion goods, Macchion et al., (2015) [9] identified three different branches of factories in which different ways were found to organize production and distribution network with specific competitive preferences. Zilberman et al., (2019) [10] proposed a framework to outline the important factors in innovative supply chains from various points of view, including product, technology, or system.

Many authors have identified the focus of the strategy of geographical diversity in product distribution. Caniato et al., (2014) [11] investigated the issue of designing a comprehensive network while integrating the new products and international retailing was considered.

J.-M. Chen & Chang (2013) [12] focused on product type and sales time in supply chains and proposed an analytical decision framework to solve it. Some interactions exist between physical processes, information flows, and the management in supply chains (Mehrjoo & Pasek, 2014) [13]. Therefore, Zhou et al., (2015) [14] used an optimal strategy style to show its importance during the scheduling process of supply chains. S. Yang et al., (2017) [15] proposed a method to increase retailers' profit by utilizing a supply chain design where perishable food products exist. Soolaki & Arkat (2018) [16] proposed a method for designing new supply chains using the mathematical programming method. To solve their model, they used a hybrid genetic ant lion optimization algorithm. Delgoshaei et al., (2014) [17] proposed a simulated annealing algorithm for maximizing the profit during construction of supply chains. Allaoui et al., (2018) [18] proposed a 2-stage framework where in the first stage, the best partners were selected using a hybrid AHP and Ordered Weighted Averaging methods, and the outcomes are then used in a mathematical model, as the second stage, to find out the best designs for the supply chain network.

Cohen & Lee (2020) [19] outlined strategies and methods for designing effective global supply chains aligned with the government's policy. Mogale et al., (2020) [20] developed a mathematical model to utilize the wheat supply chain in India. In order to solve their complex model, they used a variant of the particle swarm optimization method. Singh et al., (2021) [21] focused on the adverse impact of COVID-19 on the elements of supply chains. They proposed a public distribution network to simulate the three main drawbacks in food supply chains. The outcomes of this section are outlined as below:

1. It is important to know the product types according to the client's cultural and urban factors;
2. While designing a new supply chain, it is important to design new products;
3. Designing a new supply chain network must be aligned with the country's policy or economic growth;
4. Proposing appropriate multi-stage frameworks for designing a new supply chain network has been successfully used by scientists.

### 2.3. Market Demand Forecasting

Market demands is another critical issue that scientists investigate. Perhaps the reason for such importance is that the main aim of supply chain management is to fulfil the customer's demands in various markets. To continue, several significant recent research studies, which are primarily aligned with the problem statement of the research, will be investigated.

Ni & Fan (2011) [22] proposed models for forecasting the short-term and long-term customer demands. Lo et al., (2012) [23] focused on the advantages of using environmental management systems on performance and economic results that can be obtained accordingly. Delgoshaei et al., (2016) [24] stated that the dynamic conditions can impact adversely on the scheduling process. Dye & Hsieh (2012) [25] provided an inventory model with a variable rate of deterioration and a slight downgrade that considered the amount of capital involved in the product conservation technology to measure the maximum profit. Basu & Nair (2014) [26] presented a multi-period inventory control formula in 2013 in a dynamic random programming model.

### 2.4. Perishable Items in Supply Chain Planning

Food products are mostly perishable. Scheduling a supply chain while perishable items are available is more important, must be done accurately as perishable products are sensitive to time and can be wasted if they are not distributed correctly or on time. Scientists have investigated perishable items during the last decade. Ouyang et al., (2006) [27] presented an inventory model minimizing annual inventory costs of a supply chain while perishable items were available. For this purpose, they focused on determining repository replenishment optimal policies. Hsu et al., (2010) [28] addressed the inventory control problem in supply chains using new technologies to improve the maintenance of corrosive items. Widyadana & Wee (2011) [29] focused on the role of production capacity and machine failure in developing an inventory control model for perishable items in production systems. Musa & Sani (2012) [30] developed a mathematical programming method for managing the inventory system while perishable items were available in a system. Mishra et al., (2013) [31] developed a model for minimizing costs where order time, maintenance costs, and rate of corruption were taken into account.

### 2.5. Transportation and Logistics in Supply Chains

Transferring a local factory to a national or global supply chain can impose intolerable costs on the supply chain owners. Thus, in continuation, several critical references will be reviewed. Mula et al., (2010) [32] reviewed mathematical models that were successfully developed for production and transportation planning supply chains. Qin et al., (2011) [33] addressed product transmission problem using a selective control system. MacCarthy & Jayarathne (2013) [34] showed that the distribution of components in apparel retailers

would affect supply chain performance; thus, an important consistency among the retailers and distribution chain must be considered. Delgoshaei, Ariffin et al., (2016) [35] reviewed various internal and external transferring methods in manufacturing systems and outlined the major drawbacks that were usually emerged while scheduling manufacturing systems. Paciarotti & Torregiani (2020) [36] focused on the role of logistics in increasing the sustainability of short food supply chains. Their research outlined that optimizing the location of supply chain nodes, improving the distribution route, and restructuring the supply chains are important factors in improving food supply chains' sustainability.

By reviewing the references in this section, the following points are outlined:

1. It is important to find the best locations of supply chain nodes (wholesaler's centers);
2. It is essential to recognize the possible ways of transporting products according to geographical factors and vehicles availability.

### 2.6. Sustainability in Supply Chains

Many research studies have also been carried out to consider environmental issues (Macchion, Moretto, Caniato, Caridi, Danese, Spina et al., 2017) [37] and to increase the sustainability of supply chains. The reason for reviewing the sustainable designing and scheduling of supply chains is their effects on designing supply chain networks.

Levner & Proth (2005) [38] focused on the role of global policies in managing the ecological factors that protect oceans from industrial pollution and overexploitation. Levner (2007) [39] used risk analysis methods for sustainable wastewater management in supply chains using an economic mathematical model where economic, technological, and social constraints were considered. Nagurney & Yu (2012) [40] addressed a new method for scheduling multi-product supply chains in the fashion sector where environmental issues were considered. L. Yang & Ng (2014) [41] developed a strategic model for capacity-restricted multi-product supply chains where market demands were uncertain. Lion et al., (2016) [42] focused on the sustainability of drivers and practices within the Italian fashion industry. They proposed a taxonomy of these approaches by adopting the supplier perspective, a novelty in the sustainability literature. Macchion et al., (2018) [43] focused on the role of strategic approaches in increasing the level of sustainability in supply chains. Meanwhile, Moretto et al., (2018) [44] drew a five-step roadmap for achieving higher sustainability levels in supply chains.

It is concluded that the new supply chain must be designed in a way that:

1. It is aligned with the policy of the supply chain owners;
2. It is essential to identify the factors that can increase sustainability while designing a supply chain network;
3. Minimize the product transportation that could reduce fuel consumption and air pollution accordingly.

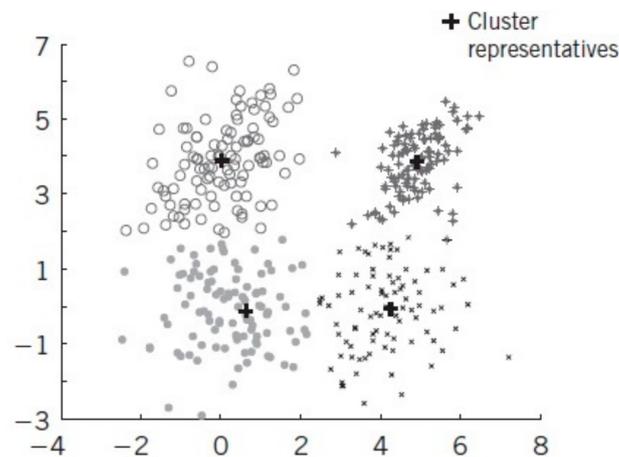
### 2.7. Machine Learning Methods

Machine learning methods have been widely used during the last decade, specifically by the emerging industry 4.0. Machine learning algorithms can be used for two primary purposes: pattern recognition and clustering. Despite supervised learning, the unsupervised machine learning methods do not use the label of a dataset for classifying its members. Unsupervised machine learning algorithms are mainly used while the labels for data are not available or unknown.

Clustering is one of the most critical applications of unsupervised machine learning algorithms. In clustering, an attempt will identify the relations between objects and then dedicate them to the most related cluster accordingly. Clustering techniques can be divided into two main methods: agglomerative clustering and divisive clustering, where agglomerative objects will be grouped to gather step by step until a unique cluster emerges (Theodoridis et al., 2010) [45]. Despite, in divisive clustering, a big group of objects will be taking part into several subgroups, and then, each of the subgroups can be divided into

other smaller subgroups as well. This process will be repeated until all objects, which are mostly related, are grouped in a cluster.

Figure 4 shows a graphical view of objectives that are clustered using unsupervised machine learning algorithms.



**Figure 4.** Graphical view of cluster simulating.

Depending on the clustering algorithm, different functions for grouping objects may be used. In partitioning methods, patterns of objects will be created to set similar objects with less distance (or any other cost function) into the K partition. K-mean, K-medoids and C-mean are among the partitioning methods that scientists mostly use.

Generally, in a K-means algorithm, the aim is to find objects with less distance from a center point. As evident, considering different numbers of K points will result in different partitions. Therefore, one crucial step is to find out how many partitions must be taken into account. Chitta & Narasimha Murty (2010) [46] proposed step K-mean algorithm and tried to find the correlations between the number and size of the partitions.

It is worth knowing that calculating the distance between an object with a K center point can be done using the Euclidean distance method; however, other important measures such as Manhattan are frequently used where applicable. For example, a K-mean harmonic algorithm is proposed by B. Zhang et al., (1999) [47], where a harmonic distance between an object and kth center point is calculated using the following formula:

$$\text{KHM}(X, C) = \sum_{i=1}^N \frac{k}{\sum_{j=1}^k \frac{1}{|(X_i - C_j)|^p}} \quad (1)$$

In the above formula, K is the number of partitions, C shows the partition number and N represents the number of available objects.

Their method is used by Ünler & Güngör (2009) [48] for making partitions objects based on the degree of membership. Kaufman and Rousseeuw (2009) [49] stated that in K-medoids, a set of medoids that will form partitions are used as the benchmark points to calculate the cost (distance) function with every single object and, then, the object will be grouped with a partition with lowest distance function value. Won & Chang Lee (2004) [50] proposed two versions of the P-median problem (PMP) method to maximize the sum of the similarities between machines in a cell forming problem. Then, Ashayeri et al., (2005) [51] proposed a PMP-based heuristic method for partitioning machine locations. Won & Currie (2006) [52] used a variant of PMP that calculated similarity coefficients using machine-component index matrix (MICM). Goldengorin et al., (2012) [53] addressed a rapid machine location method based on PMP that minimizes the dissimilarities between centroid and machine locations. Krushinsky & Goldengorin (2012) [54] mentioned that the information of the MCIM matrix is not sufficient for providing good layouts. In their

research, a new straightforward formulation and the alternative formulation were used for minimizing dissimilarities. In some cases, meta-heuristic algorithms were successfully used as a part of a method for portioning or clustering objects. For example, Paydar & Saidi-Mehrabad (2013) [55] proposed a genetic algorithm and variable neighborhood search method to maximize grouping efficacy.

Izakian & Abraham (2011) [56] presented a combination of fuzzy C-mean and Particle Swarm Optimization, where a threshold value is used for determining partitions.

In Complete Linkage Clustering (CLINK), each object is considered a cluster at the beginning. Smaller clusters will join if their distance similarity coefficient is small enough. Angra et al., (2008) [57] proposed 2 algorithms for the cell forming problem where in the first algorithm, the objects could be clustered and machines could also be grouped based on their processing times. Afterward, a second algorithm was applied to calculate the total processing time of activities in each cell. Oliveira et al., (2009) [58] used a spectral clustering algorithm for minimizing the inter-cell movements in a cell forming problem where the clusters are formed based on the cell-size constraint.

The literature review outcomes show that machine-learning algorithms have been widely used for various problems in the supply chain sector and thus are a promising way. Besides, it was found that the unsupervised machine learning methods have not been used for food consumption pattern recognition and product distribution.

#### Hybrid of Metaheuristics and Clustering Methods

Unsupervised machine learning algorithms have been widely used as a part of hybrid meta-heuristic algorithms. In most cases, they are used for clustering or partitioning a set of data. Banerjee & Das (2012) [59] addressed a 2-phase genetic algorithm for generating adaptive clusters in a manufacturing layout and identified bottleneck machines accordingly. Kao & Li (2008) [60] used the ant colony optimization method in a clustering algorithm where the ants were used for object recognition in an agglomerative way. F. Yang et al., (2009) [61] used hybrid particle swarm methods called KHM and PSOKHM to overcome local optimum traps during clustering objects. Later, Nouri et al., (2010) [62] used Bacteria Forging Algorithm to address the material transferring problem in manufacturing systems.

Chattopadhyay et al., (2012) [63] argued that choosing the optimum size of the SOM is a big problem while using it. Therefore, they offered a new method that used the average distortion values during the training process in SOM as a criterion to determine the correct size of SOM. Kuo et al., (2006) [64] proposed a fuzzy set in a variant of ART to produce family problems that can improve learning procedures and result in better patterns. Özdemir et al., (2007) [65] focused on preventing unnecessary clusters in the clustering process using the ART method called the proliferation problem. M.-S. Yang & Yang (2008) [66] used a new method for improving the learning process in ART by modifying vigilance parameters and training vectors. Pandian & Mahapatra (2009) [67] used a new version of ART that could obtain operation sequences and time as inputs.

In this section, statistical analysis will be done to compare the opted research studies. According to our findings, Production Scheduling, Technology in Use and Material Transferring are the most investigated methods for minimizing the product completion time in the opted research studies (Figure 5).

While machine learning methods are investigated, it is found that in more than 62% of the studied cases supervised machine learning methods have been used and, in 3% of the cases, unsupervised machine learning methods (Figure 6).

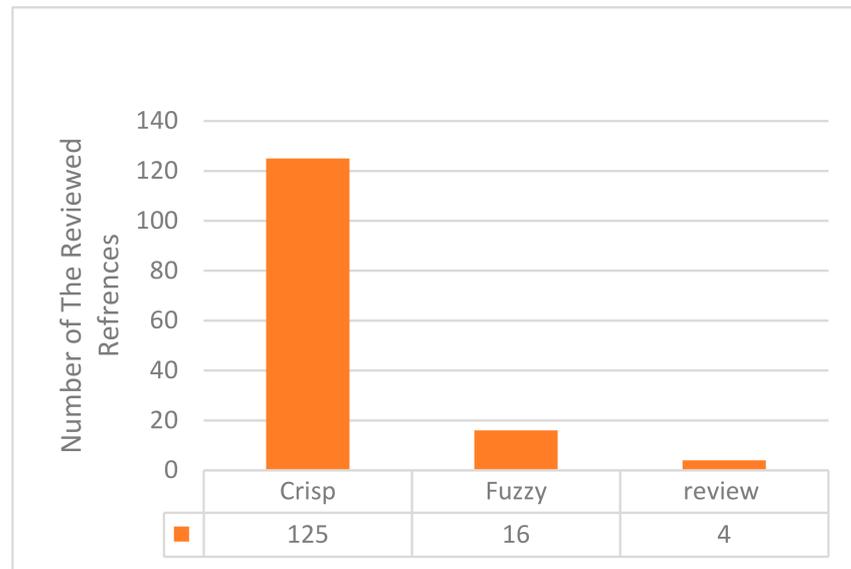


Figure 5. Frequency of the reviewed research studies according to the categories.

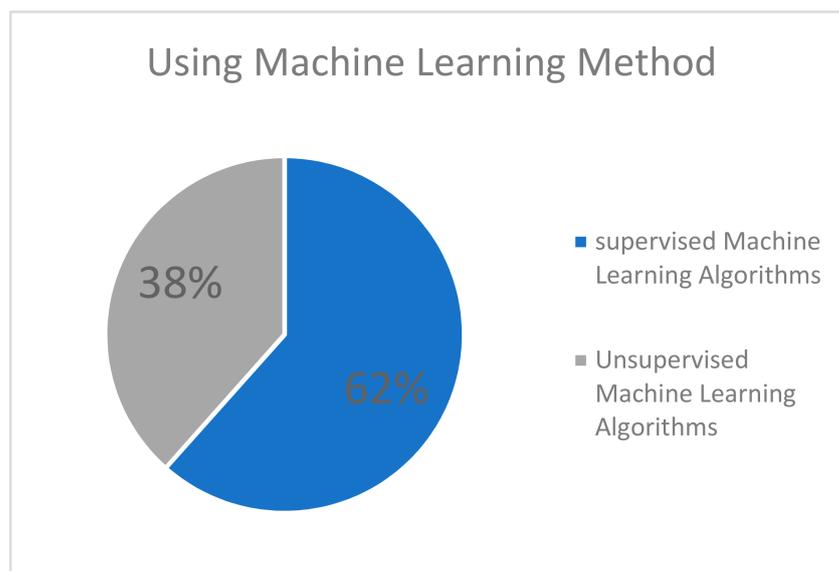


Figure 6. Comparing frequency of using machine learning methods.

As far as found by reviewing the papers, the following findings were achieved:

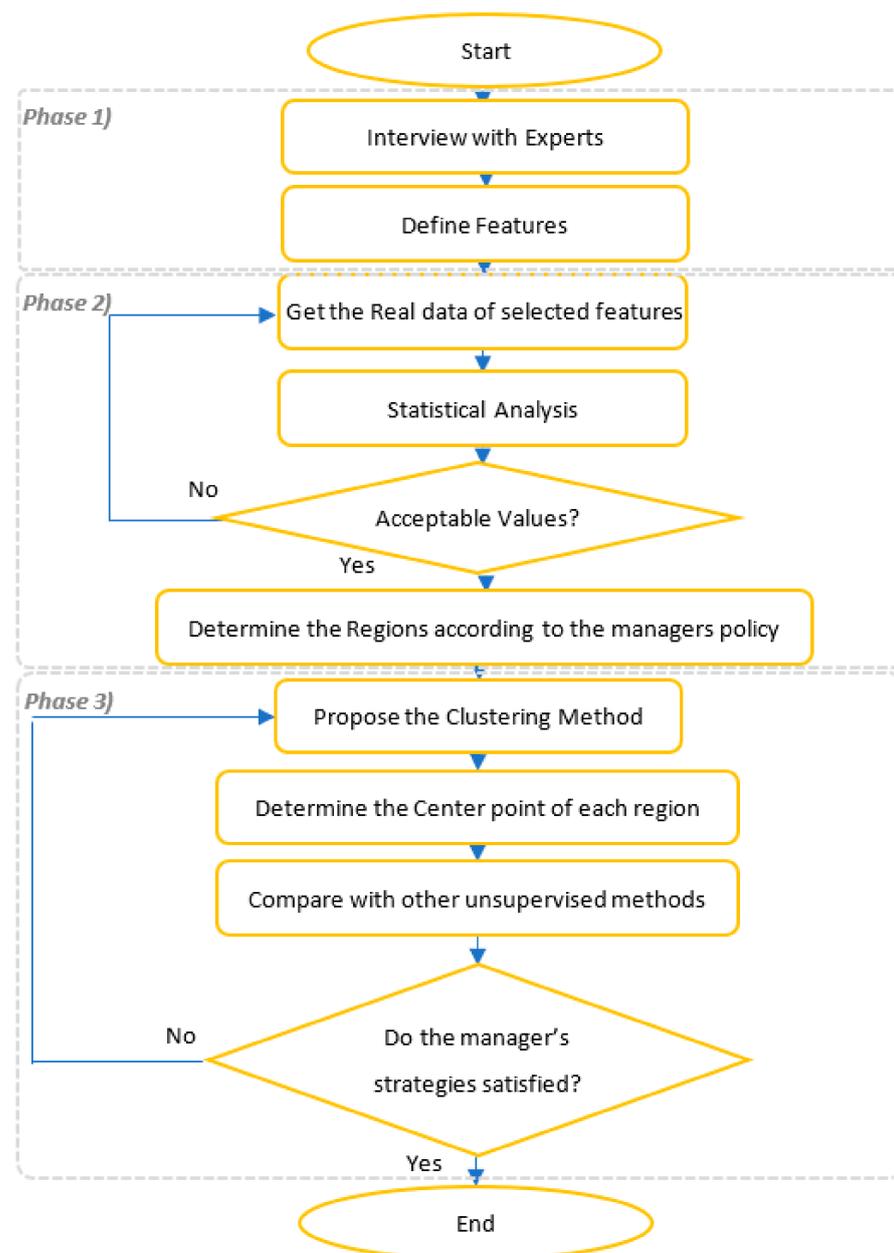
1. According to the literature review, it is essential to determine the critical factors for designing a new supply chain network;
2. In designing a new supply chain, determining the product types is crucial and must be considered;
3. When designing the new supply chain, determining the best locations for the supply chain nodes (wholesaler’s center points) is very important and can decrease transportation time and cost;
4. Machine learning algorithms have been widely used for various engineering problems, including supply chains; however, they were not used for transforming a local factory into a national supply chain network.

### 3. Research Methodology

In this section, an unsupervised machine-learning algorithm (K-means) will be proposed according to the data of the local chocolate factory located near Kuala Lumpur to turn into a national wide supply chain. This section aims to determine the appropriate places for the wholesalers' center points for designing the supply chain network according to the population and destination features. The K-means algorithm outcomes will be compared with a number of the most frequently used machine-learning algorithms.

#### 3.1. Flowchart of the Proposed Method

Figure 7 shows the flowchart of the research methodology in more detail.



**Figure 7.** Flowchart of the research methodology.

According to the research flowchart, in the next section, the effective features that can influence the wholesaler's center points in designing national-wide supply chain will be identified first by interview with top managers of the company (phase 1). Then, the real

data of the features will be extracted and prepared for the next phase (phase 2). Afterward, an unsupervised machine learning method will be applied (phase 3) to cluster the centers and find the best location for the wholesaler's center points according to the selected features. The performance of the outcomes of the method will then be evaluated by using some metrics.

### 3.2. Choosing an Appropriate Machine Learning Algorithm for This Research

In this section, the reasons for choosing an unsupervised machine learning algorithm will be explained:

1. The data used in this research do not have any label as the scope of this research is to enter new markets by establishing new wholesalers and schedule product distribution in markets;
2. Unsupervised machine learning algorithm can find the center of clusters (centroids) based on the data. Such centroids can be considered as the center point of wholesalers in a supply chain;
3. Unsupervised methods can effectively find the various pattern of data without having a specific label.

Table 1 compares the attributes of the machine learning methods. Therefore, since data do not have a label in this research, unsupervised machine learning methods must be used (Table 1).

**Table 1.** Comparing machine learning methods.

Machine Learning Type	Label	Complexity	Training	Accuracy
Supervised Learning	✓	Low	Using labeled data	High
Unsupervised Learning	✗	High	Using Data Information	Medium
Reinforcement Learning	✓	Medium	Using Actions (Reward, Punishment)	High

Considering the aim of clustering to find the best center point for the wholesalers, K-means seems to match better with the method as they can find the best center points of clusters using a series of data.

However, in Section 4, more unsupervised machine learning algorithms will be used to evaluate their performance.

### 3.3. K-Means Algorithm

As mentioned in the previous section, choosing an appropriate machine learning algorithm is entirely dependent on the nature of a problem. Therefore, according to the findings of the unsupervised machine learning algorithms, in this research, we focused on the K-means algorithm as the most fitted unsupervised ML algorithms for finding the best center points in our research.

K-means algorithm, which is also known as Lloyd's Algorithm, is a promising way to cluster the objects in a dataset and is frequently used by many scientists in various computer science and engineering problems.

The nature of the algorithm is based on clustering objects into K clusters, where each object is placed in a cluster with the nearest mean. In the K-means algorithm, the distance between objects is measured. The aim is to assign objects to clusters, to achieve the minimum squared Euclidean distance. The most crucial method to calculate the distances between an object and the center (or centroid) of a cluster is the Euclidean Distance method. However, depending on the nature of the distances, other distance methods, such as Euclidean, Manhattan and Cosine are used.

The steps of the K-mean algorithm are as follows:

- Step-1: Determine the number of clusters (K);  
 Step-2: Specify the K- points (or centroids);

- Step-3: Calculate the distance of each object to the closest centroid. It will form the pre-determined clusters (K);
- Step-4: Determine the centroid of each cluster by calculating the variance values;
- Step-5: Go back to Step-3 and re-calculate the distance of each object considering the new closest centroid of each cluster;
- Step-6: If a new reassignment is possible, then go to Step-4.

Figure 8 shows the outcomes of a K-mean algorithm for a problem. As seen, the K-mean algorithm could effectively assign objects into the most related cluster.

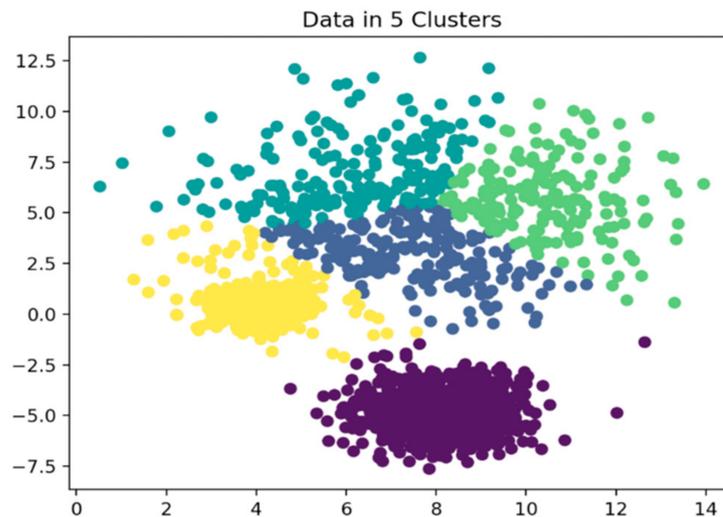


Figure 8. Using K-means algorithm in clustering data into 5 clusters.

### 3.4. Develop a Machine Learning Algorithm

This section will show the necessary steps to develop unsupervised machine learning algorithms used in phases 2 and 4 of the framework (Figure 9).

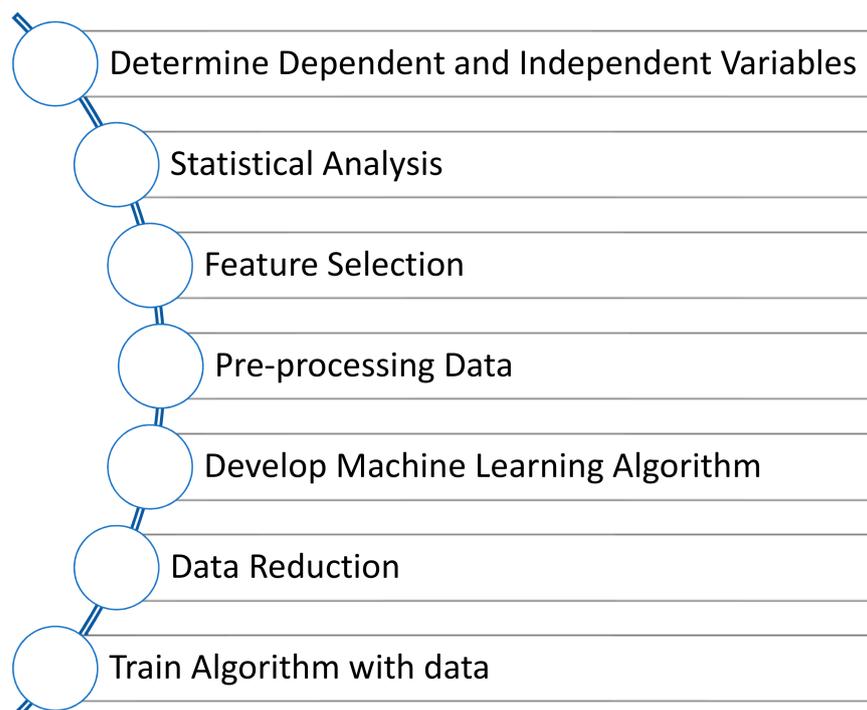


Figure 9. Steps of developing the machine learning algorithm.

In the following, the necessary steps for developing a machine learning method will be illustrated.

#### 3.4.1. Python

In this research, Python was used for coding the proposed machine learning algorithm.

- Python is one of the most frequently used software with a powerful engine for calculating mathematical equations and models;
- Python has many applications. Jupyter is a handy platform for coding the machine learning algorithms, as each line of the scripts can be executed separately and the results can be seen.

#### 3.4.2. Libraries

While using Python, choosing correct libraries is crucial, as they contain many formulas and commands that can be used to develop and evaluate the algorithms' performance.

Below, libraries that must be used for developing the machine learning algorithm in the next section will be explained briefly:

- Numpy

Numpy is a Python library for generating and working with homogeneous multi-dimensional arrays. It is also used for applying basic mathematical formulas. These arrays are tables of elements (usually numbers) of the same type and are indexed by a few positive integers. In NumPy, dimensions are known as axes. The number of axes is called the rank.

- Pandas

Pandas is the second library that will be used in this research. The main aim of using Pandas is to develop the data frameworks. With pandas, it is possible to import the data with different file types such as CSV and XLSX.

Pandas is also an excellent library to work with matrixes and perform various functions, such as adding a row, deleting a column, multiplying two matrixes, etc.

- Matplotlib

Matplotlib is used for drawing various plots, including histogram, box-chart, bar chart and scatter chart.

- Seaborn

Seaborn is a library that contains powerful formulas for statistical analysis.

- Scipy

Scipy is widely used for various purposes. However, in this research, Scipy will be applied for calculating the correlations between factors and optimization purposes.

- Scikit-learn

The scikit-learn is an essential library for this research. This library will be used for applying the supervised and unsupervised machine learning methods.

Python must be installed on a computer before importing libraries into it.

#### 3.4.3. Pre-Processing Data

Pre-processing data will be used to pre-process data such as missing data, duplicate data, concatenation, scaling data and indicating the outliers.

#### 3.5. Processing Machine

The computer features have a direct impact on the processing speed. For this purpose, it is necessary to introduce computer information. In this research, a personal laptop with the following characteristics is used:

- CPU: Intel Core i7;
- VGA: 4 GB;
- RAM: 8 GB;

- Windows: 10 Enterprise.

**4. Results and Discussion**

*4.1. Identify and Analyze the Information of the Geographical, Transportation System and Clients for the Destinations (Phase 3)*

The survey outcomes showed that the following factors could be considered independent variables that can affect food delivery.

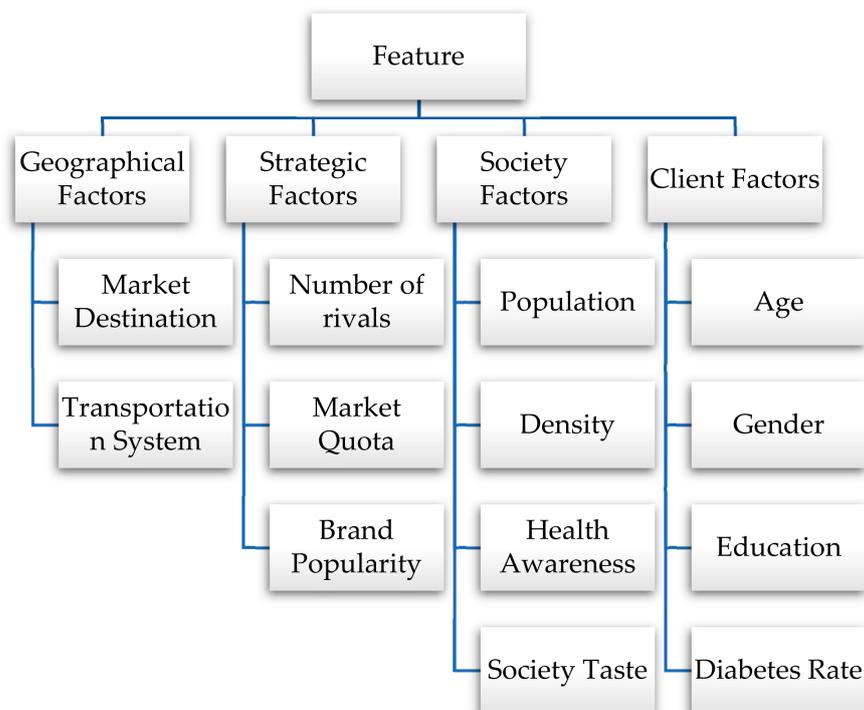
- Market Destination;
- The population of the market;
- The density of the Market;
- Transportation System;
- Market Quota in the destination;
- Number of rivals in the destination;
- Brand Popularity;
- Age;
- Gender;
- Education;
- Health Awareness;
- Society Taste;
- Diabetes rate of the destination population.

The main question in the first phase of the proposed framework that can be asked to clarify the research methodology is the following:

What would be the best priority for food distribution in an area such as a city or county considering the features of it (independent variables)?

It is a crucial question in the product distribution of the food industry sector. This section aims to find the best center point for the wholesalers and retailers accordingly.

According to our findings, the features can be classified into four main clusters, as shown in Figure 10:



**Figure 10.** Independent factors that can influence the food delivery in an area.

This research assumes that the company wants to find the best location for establishing the wholesalers. As a result, it is assumed that there is no statistical evidence for consumption rate (as the dependent variable) in the markets throughout the studied cities. Therefore, the data will not have any labels as the consumption rate. It means that for this section, unsupervised machine learning algorithms must be used to determine the wholesaler center points. According to the section above, the independent variables can now be represented in Table 2.

**Table 2.** The independent variables of the research.

NO.	Type	Cluster	Variable
1	Dependent	-	Consumption Rate (unknown)
2	Independent	Geographical	Market Destination
3			Transportation System
4			Market Quota in the destination
5		Strategic	Number of rivals in the destination
6		Brand Popularity	
7		Age	
8		Personal	Gender
9			Education
10			Diabetes Rate
11		Society	Population
12			Density
13			Health Awareness
14			Society Taste

This section is essential to show the correlations between the research questions, objectives, and variables.

#### 4.2. An Unsupervised Machine Learning Algorithm to Determine the Best Center Points for Wholesalers Location (Phase 4)

The first part is to find where to locate the central manufacturing companies in the country. The aim is to find the best locations that could minimize product transportation costs. In this regard, two main features must be considered: population and distance. Those states with more population will request more product demand. Besides, distant locations need more transportation costs. Therefore, in the first step, using two K-means algorithms prepared in terms of states and their populations and distances, the best locations for the manufacturing factories will be determined. Then, using a hybrid PCA and K-means method, all features are used to estimate the best location for the main factories.

#### 4.3. Libraries

Libraries used for this section are shown as follows:

- From sklearn import linear\_model;
- From sklearn import cluster, datasets;
- From sklearn.datasets import load\_digits;
- From sklearn.cluster import KMeans;
- Import numpy as np;
- Import pandas as pd;
- Import scipy.cluster.hierarchy as shc;

- Import matplotlib.pyplot as plt;
- From `__future__` import division;
- From matplotlib import colors as mcolors;
- Import matplotlib.gridspec as gridspec;
- Import itertools;
- From scipy import interpolate;
- From scipy.spatial import ConvexHull.

#### 4.4. Dataset

In this section, the information of the Malaysian population based on states and federal territories will be used as data. The information of population is extracted from the Statistica website [68]. In addition, the information of distances between capital cities of the states is extracted from the distance calculator website [69]. Since the main factory is located in Kuala Lumpur, then, all distances are measured with it. Then, data will be prepared to be used in Python (Table 3).

**Table 3.** Data of population and distance of capital cities of the states and federal territories of Malaysia.

Row	State	Capital	Distance	Population
1	Selangor	Shah Alam	23.58	6,569,500
2	Sabah	Kota Kinabalu	1621.98	3,907,500
3	Johor	Johor Bahru	295.61	3,776,600
4	Sarawak	Kuching	974.49	2,828,700
5	Perak	Ipoh	175.51	2,518,600
6	Kedah	Alor Setar	362	2,193,900
7	Kelantan	Kota Bharu	334.61	1,904,900
8	Penang	George Town	293.45	1,783,600
9	W.P. Kuala Lumpur	Kuala Lumpur	0	1,773,900
10	Pahang	Kuantan	194.57	1,682,200
11	Terengganu	Kuala Terengganu	287.5	1,259,000
12	Negari Sembilan	Seremban	54.51	1,135,900
13	Melaka	Malacca City	121.36	936,900
14	Perlis	Kangar	402.38	255,000
15	W.P. Labuan	W.P. Labuan	1518.56	99,600
Total Population				32,625,800

Since the country has two sides (left and right islands) and each side has a significant population size, each side must be considered as a group; otherwise, the best center point for the wholesaler will be estimated in the middle of the sea (Figure 11).

#### 4.5. Evaluating the Features

In this section, the important features must be determined before choosing features to be considered in the learning process. For this purpose, we used the Shapiro method for ranking the features. The outcomes showed that both features are important and must be considered in the learning process (Figure 12).

$$\text{visualizer} = \text{Rank1D}(\text{algorithm} = \text{'shapiro'}) \quad (2)$$

The results showed that both destination (feature 1) and population (feature 2) are important enough to be considered in the machine learning algorithm.



Figure 11. The map of Malaysia with state names [70].

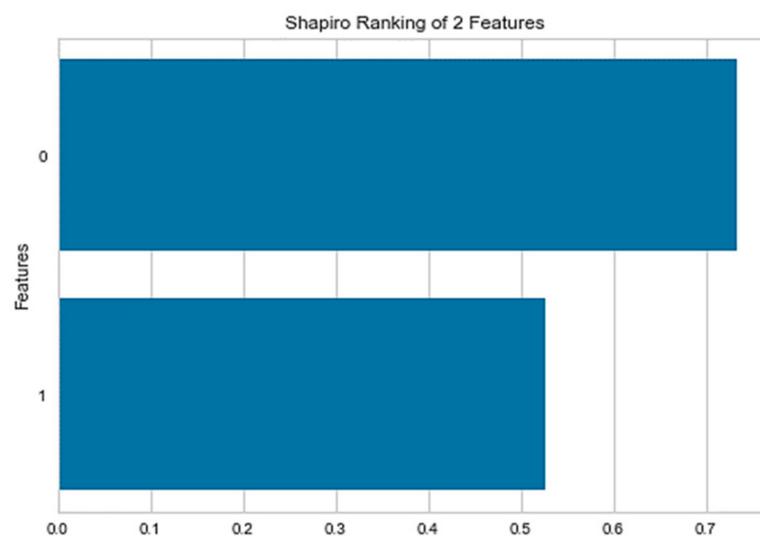


Figure 12. Results of the Shapiro method for the supervised machine learning method.

#### 4.6. Strategies of the Top Managers

Before designing the supply chain, the strategies of the top managers of the chocolate factory are considered as followings:

- (1) The supply chain will be designed according to the top managers' strategies. Due to the specific shape of Malaysia (having separated islands), the right island must be considered separately; otherwise, the centroid of the wholesaler for the right island will be located in the middle of the ocean. Therefore, the data of Malaysia will be considered accordingly.
- (2) To find the wholesalers' best locations using the clustering algorithm more precisely, the dataset will be divided into four main areas according to the geographical distribution of population, which are north, middle, right and south areas of the left island and also the right island. Such grouping will help prevent wrong clustering. To clarify, suppose the two largest cities are located in the north and south of the country. If they are considered together, then the clustering method will consider the two states as a separate cluster and then consider the centroid of these two states as the center point for the wholesaler. Moreover, the small cities around these two cities will be allocated

to another cluster. Such a wrong strategy will cause increasing transportation costs and time.

- (3) The chocolates will be transferred to Sarawak and then transferred to the destinations.
- (4) Johor must be considered a separate database due to its geographical location; otherwise, if the Johor and Melaka are clustered together, the products will be transferred to a center near the Johor and returned to Melaka!
- (5) To draw the clusters, each 1 million people will be considered one black point in the scatter charts.

Then, the dataset will be divided into 4 main groups (Table 4).

**Table 4.** Sorted version of the population of states and federal territories of Malaysia.

Row	State	Capital	Distance	Population	Geographical Location
1	Perak	Ipoh	175.51	2,518,600	North
2	Kedah	Alor Setar	362	2,193,900	North
3	Kelantan	Kota Bharu	334.61	1,904,900	North
4	Penang	George Town	293.45	1,783,600	Right
5	Terengganu	Kuala Terengganu	287.5	1,259,000	Right
6	Perlis	Kangar	402.38	255,000	North
7	Selangor	Shah Alam	23.58	6,569,500	Mid
8	W.P. Kuala Lumpur	Kuala Lumpur	0	1,773,900	Mid
9	Pahang	Kuantan	194.57	1,682,200	Mid
10	Negari Sembilan	Seremban	54.51	1,135,900	Mid
11	Melaka	Malacca City	121.36	936,900	South
12	Johor	Johor Bahru	295.61	3,776,600	South
13	Sabah	Kota Kinabalu	1621.98	3,907,500	Right Island
14	Sarawak	Kuching	974.49	2,828,700	Right Island
15	W.P. Labuan	W.P. Labuan	1518.56	99,600	Right Island
Total Population				32,625,800	

#### 4.7. Preparing the Dataset

In this section, to draw the scatter chart more precisely, each point in the scatter chart represents 1 million people. Note that a point is also considered for the cities with a population of less than 1 million.

Then, the dataset for each zone is prepared as shown in Tables 5–9:

**Table 5.** The population of states and federal territories of northern areas of the left island of Malaysia (North-DB).

Row	State	Capital	Distance	Population	Geographical Location	Ratio
1	Perak	Ipoh	169	2,516,489	north	3
2	Perak	Ipoh	166	2,688,505	north	3
3	Perak	Ipoh	195	2,532,088	north	3
4	Kedah	Alor Setar	354	2,045,540	north	2
5	Kedah	Alor Setar	362	2,021,907	north	2
6	Kelantan	Kota Bharu	318	1,996,528	north	2
7	Kelantan	Kota Bharu	324	1,904,900	north	2
8	Penang	George Town	283	1,643,600	north	2
9	Penang	George Town	298	1,783,600	north	2
10	Perlis	Kangar	402	255,000	north	1

**Table 6.** The population of states and federal territories of middle areas of the left island of Malaysia (MID-DB).

Row	State	Capital	Distance	Population	Geographical Location	Ratio
1	Selangor	Shah Alam	19	6,672,400	mid	7
2	Selangor	Shah Alam	14	6,537,984	mid	7
3	Selangor	Shah Alam	32	6,740,606	mid	7
4	Selangor	Shah Alam	32	6,456,529	mid	7
5	Selangor	Shah Alam	25	6,507,105	mid	7
6	Selangor	Shah Alam	16	6,412,373	mid	7
7	Selangor	Shah Alam	30	6,671,739	mid	7
8	W.P. Kuala Lumpur	Kuala Lumpur	0	1,763,907	mid	2
9	W.P. Kuala Lumpur	Kuala Lumpur	3	1,754,979	mid	2
12	Negari Sembilan	Seremban	54	1,135,900	mid	1
13	Melaka	Malacca City	121	936,900	mid	1

**Table 7.** The population of states and federal territories of right areas of the left island of Malaysia (Right-DB).

Row	State	Capital	Distance	Population	Geographical Location	Ratio
1	Pahang	Kuantan	194	1,765,494	Right	2
2	Pahang	Kuantan	186	1,839,706	Right	2
3	Terengganu	Kuala Terengganu	287	1,259,000	Right	1

**Table 8.** The population of states and federal territories of southern areas of the left island of Malaysia (South-DB).

Row	State	Capital	Distance	Population	Geographical Location	Ratio
1	Johor	Johor Bahru	295	37,586,504	south	4
2	Johor	Johor Bahru	273	36,678,300	south	4
3	Johor	Johor Bahru	292	36,525,857	south	4
4	Johor	Johor Bahru	277	39,771,087	south	4

**Table 9.** The population of states and federal territories of right island of Malaysia (areas of right island, Malaysia-DB).

Row	State	Capital	Distance	Population	Geographical Location	Ratio
1	Sabah	Kota Kinabalu	1688	3,753,368	right island	4
2	Sabah	Kota Kinabalu	1609	3,703,829	right island	4
3	Sabah	Kota Kinabalu	1583	3,804,832	right island	4
4	Sabah	Kota Kinabalu	1592	3,824,530	right island	4
5	Sarawak	Kuching	1025	2,787,745	right island	3
6	Sarawak	Kuching	1081	2,639,873	right island	3
7	Sarawak	Kuching	1120	2,855,566	right island	3
8	W.P. Labuan	W.P. Labuan	1518	108,891	right island	1

It should be noted that each dataset will be run separately in the machine learning algorithm.

Therefore, the data entered into the model will be prepared according to Equation (3). The results are shown in Tables 10–14.

$$X = \text{AreaInfo.iloc}[:, [3, 4]].\text{values} \quad (3)$$

**Table 10.** Selected features of the population and distance (areas of left island Malaysia-North DB).

---

```

array([[169, 2,516,489],
       [166, 2,688,505],
       [195, 2,532,088],
       [354, 2,045,540],
       [362, 2,021,907],
       [318, 1,996,528],
       [324, 1,904,900],
       [283, 1,643,600],
       [298, 1,783,600],
       [402, 255,000]], dtype = int64)

```

---

**Table 11.** Selected features of the population and distance (areas of left island Malaysia-Mid DB).

---

```

array([[19, 6672400],
       [14, 6,537,984],
       [32, 6,740,606],
       [32, 6,456,529],
       [25, 6,507,105],
       [16, 6,412,373],
       [30, 6,671,739],
       [0, 1,763,907],
       [3, 1,754,979],
       [54, 1,135,900],
       [121, 936,900]], dtype = int64)

```

---

**Table 12.** Selected features of the population and distance (areas of left island Malaysia-Right DB).

---

```

array([[194, 1,765,494],
       [186, 1,839,706],
       [287, 1,259,000]], dtype = int64)

```

---

**Table 13.** Selected features of the population and distance (areas of left island Malaysia-South DB).

---

```

array([[295, 37,586,504],
       [273, 36,678,300],
       [292, 36,525,857],
       [277, 39,771,087]], dtype = int64)

```

---

**Table 14.** Selected features of the population and distance (areas of right island Malaysia-DB).

---

```

array([[1705, 3,818,840],
       [1548, 3,746,718],
       [1622, 4,019,239],
       [1608, 3,714,913],
       [1145, 2,880,068],
       [963, 2,963,976],
       [1145, 2,858,991],
       [1518, 88,875]], dtype = int64)

```

---

#### 4.8. Choosing the Appropriate Machine Learning Algorithm

In this section, a number of methods will be applied to find the most suitable method for determining the center points in the next section. For this purpose, the outcomes of K-means, WARD, SLINK and CLINK will be compared (Table 15). The results indicate that the K-means method could provide a better silhouette score while three clusters are considered (0.812).

**Table 15.** Results of comparing the K-means, WARD and CLINK algorithms.

Method	Silhouette Score	calinski_harabasz Score	Number of Clusters (K)
K-means	0.812	32.693	3
WARD	0.781	263.751	4
CLINK	0.781	263.751	4

4.9. Machine Learning Algorithm Settings and Results

In this section, four K-means algorithms for datasets will be applied:

$$\text{model} = \text{KMeans}(\text{n\_clusters} = 1, \text{random\_state} = 0) \tag{4}$$

Results of performing K-means for the datasets show that the K-means algorithm could successfully solve all datasets (Table 16):

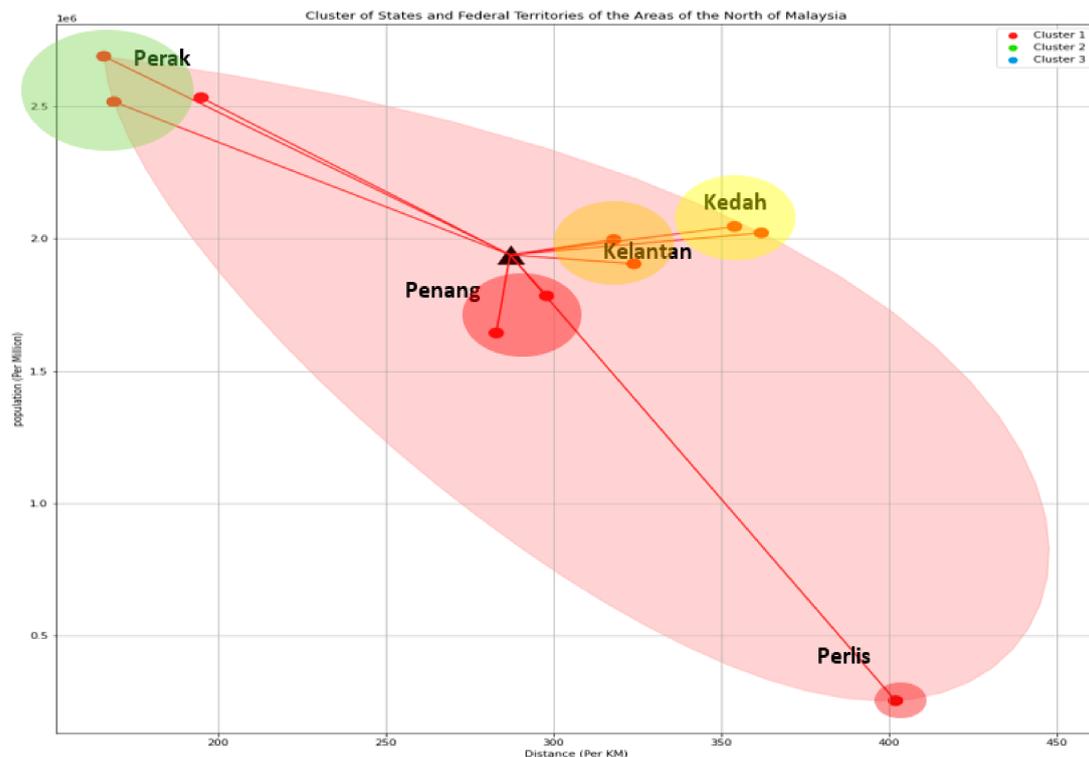
**Table 16.** Cluster centers of the different regions.

Dataset	Cluster Center Point	Iterations	Ratio
North Region-Left Island DB	[287.1, 1,938,815]	2	6.753
Mid Region-Left Island DB	[50.0, 4,874,349]	2	97.487
Right Region t-Left Island DB	[223.5, 1,622,000]	2	7.257
South Region-Left Island DB	[284.2, 3,764,043]	2	13.244
Right Island DB	[141, 3,013,952]	2	21.376

The ratio of determined cluster centers in terms of population and distance can be calculated based on the following formula:

$$\text{Ratio} = \text{Population of Cluster Center Point (per 1000 person)} / \text{Distance of Cluster Center Point (km)} \tag{5}$$

In the following, the center points for each area (centroids) will be shown in Figures 13–17:



**Figure 13.** Cluster and centroid of states and federal territories of the northern region of the left island of Malaysia.

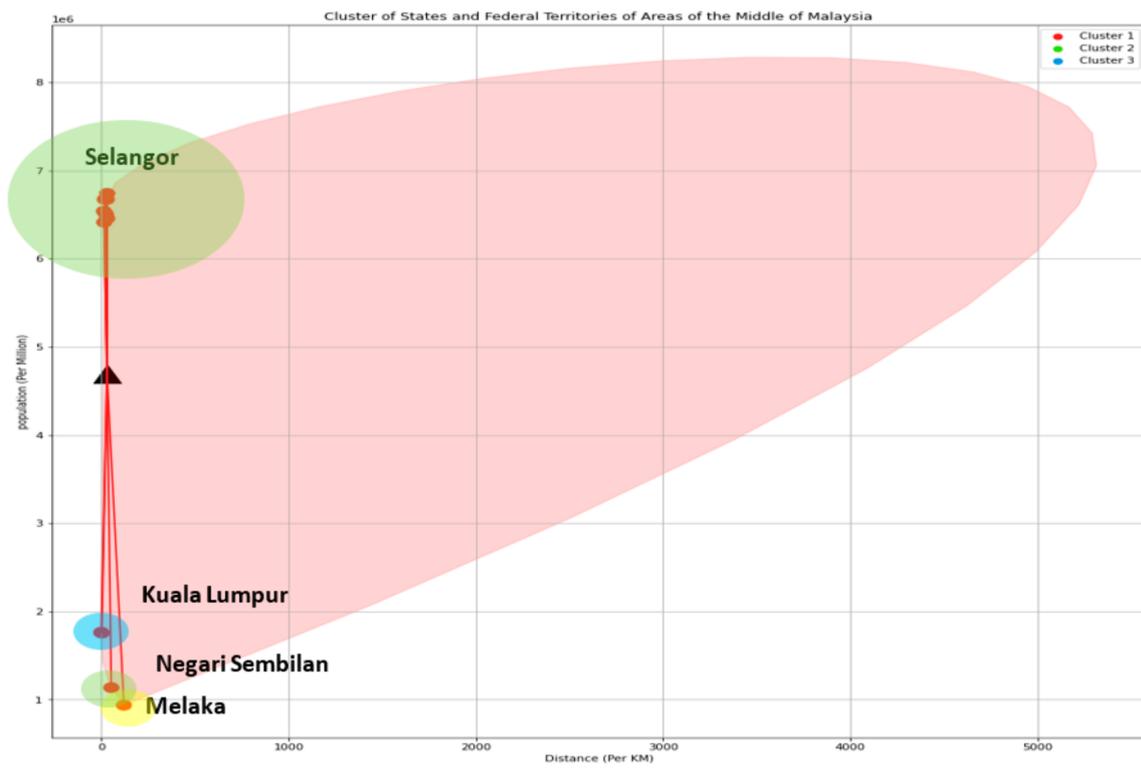


Figure 14. Cluster and centroid of states and federal territories of the middle region of the left island of Malaysia.

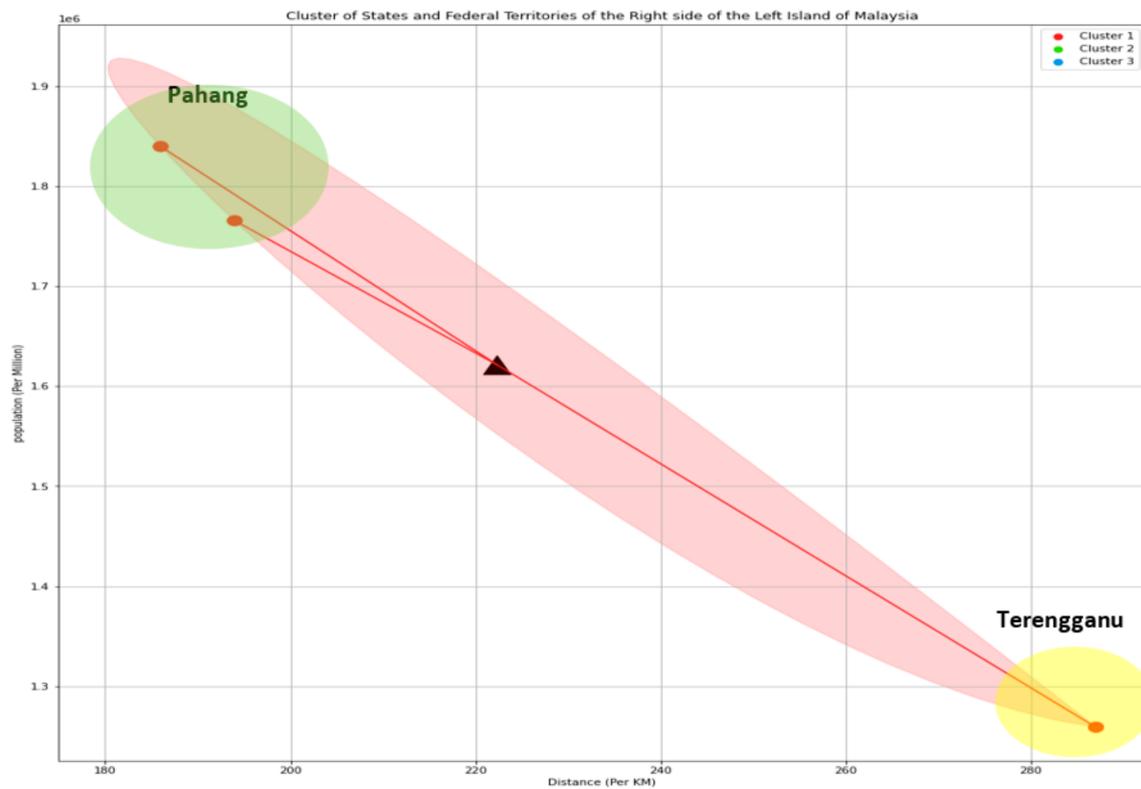


Figure 15. Cluster and centroid of states and federal territories of the right region of the left island of Malaysia.

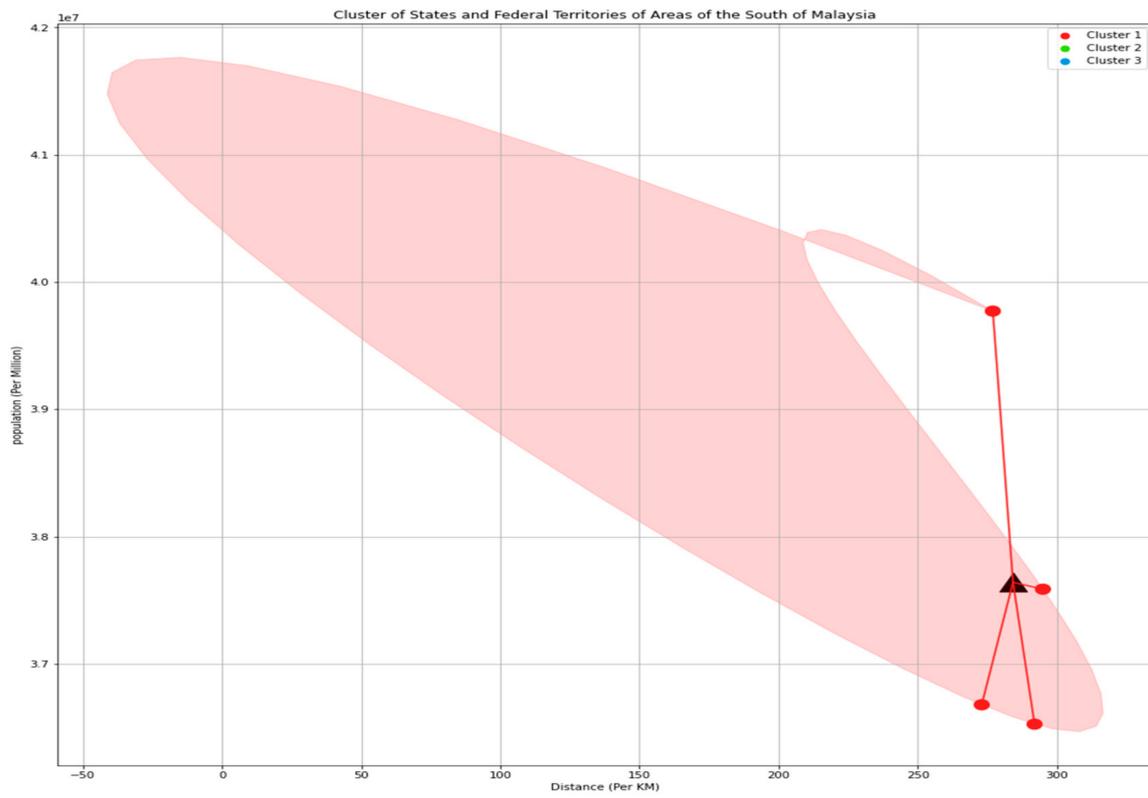


Figure 16. Cluster and centroid of states and federal territories of the southern region of the left island of Malaysia.

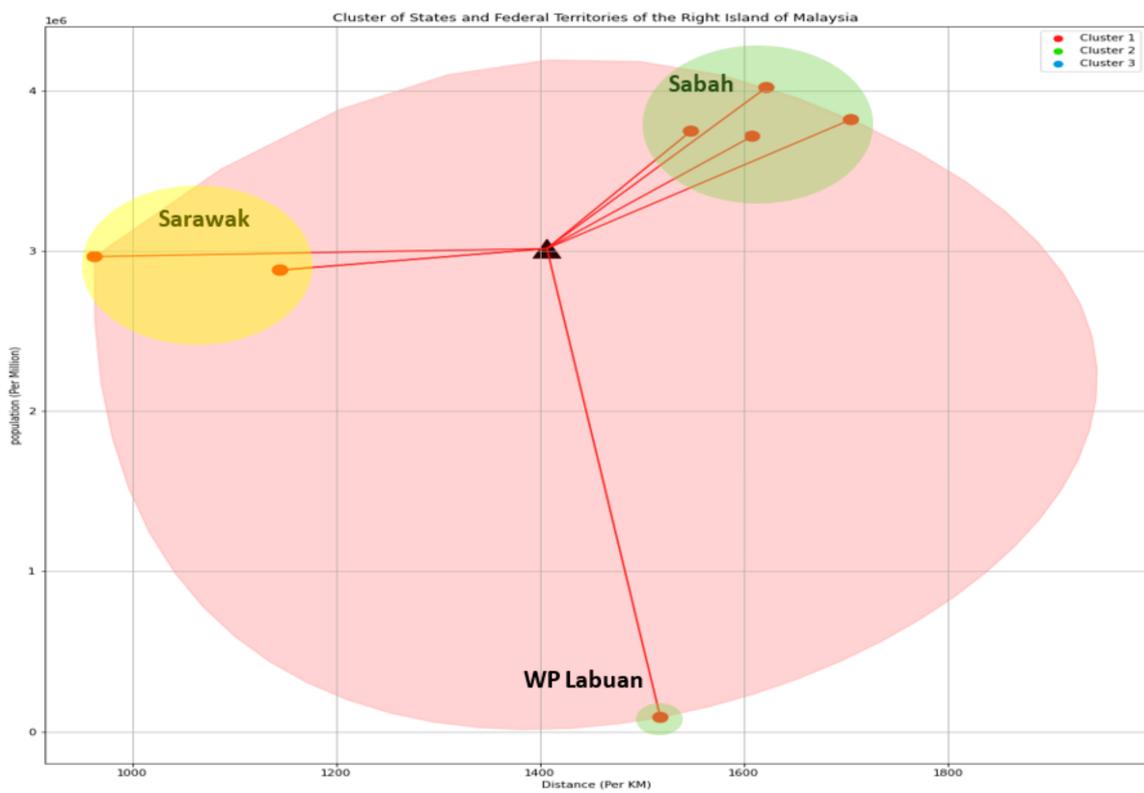


Figure 17. Cluster and centroid of states and federal territories of the right island of Malaysia.

The weighted 3D views of the K-means algorithm are represented in Figures 18–22. In these figures, the weights of cities in terms of population are shown by the size of the points where states with a higher number of populations will show bigger points.

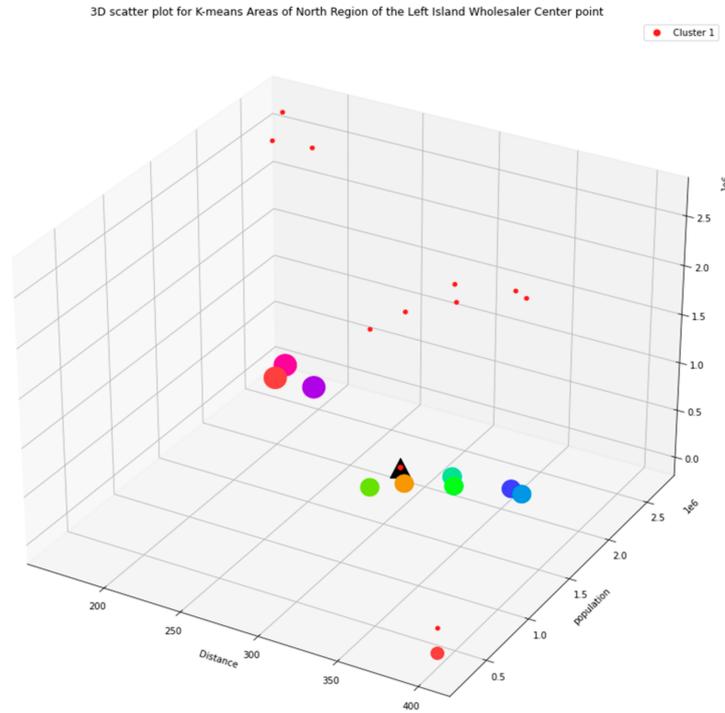


Figure 18. Weighted 3D view of the cluster and centroid of states and federal territories of the northern region of the left island of Malaysia.

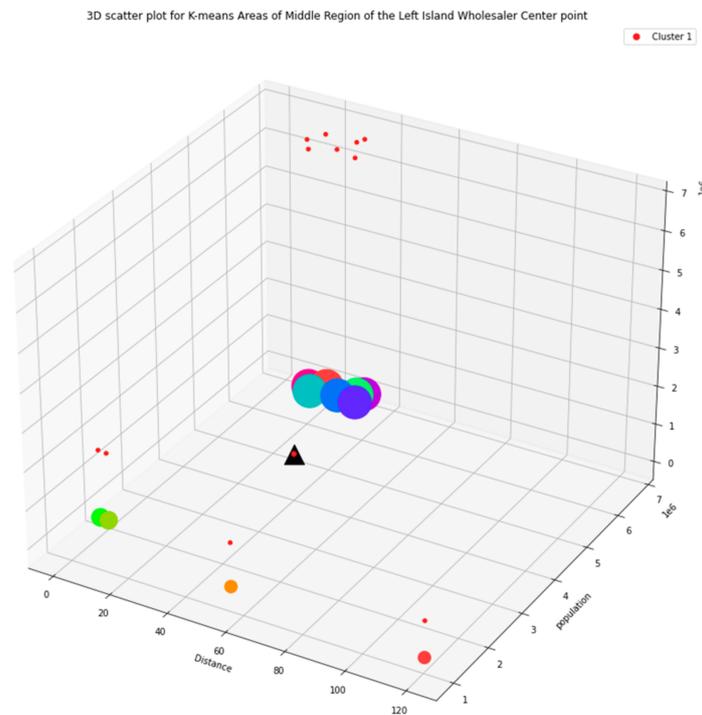
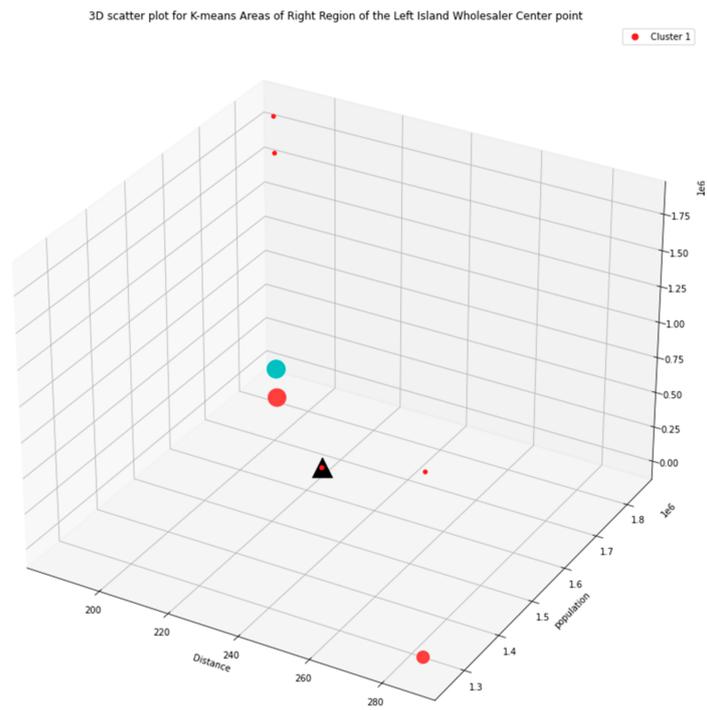
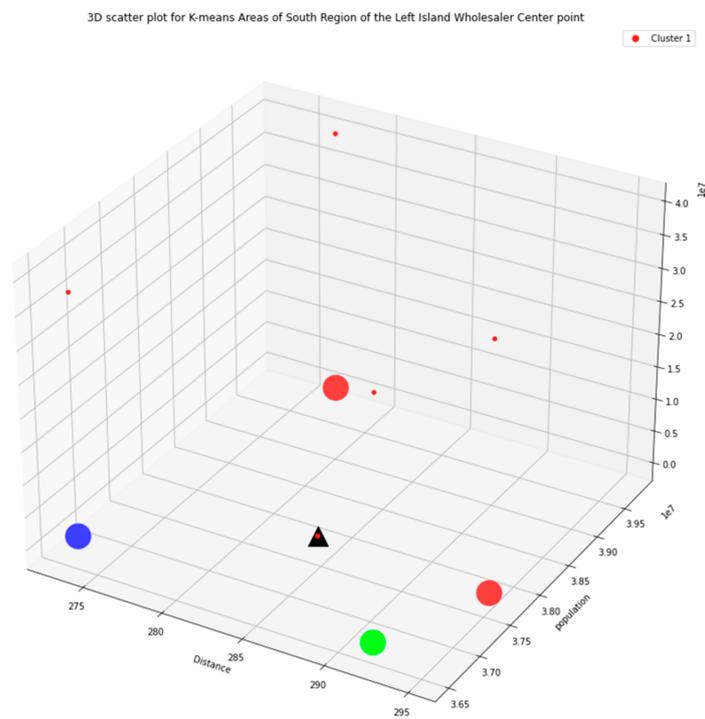


Figure 19. Weighted 3D view of the cluster and centroid of states and federal territories of the middle region of the left island of Malaysia.



**Figure 20.** Weighted 3D view of the cluster and centroid of states and federal territories of the right region of the left island of Malaysia.



**Figure 21.** Weighted 3D view of the cluster and centroid of states and federal territories of the southern region of the left island of Malaysia.

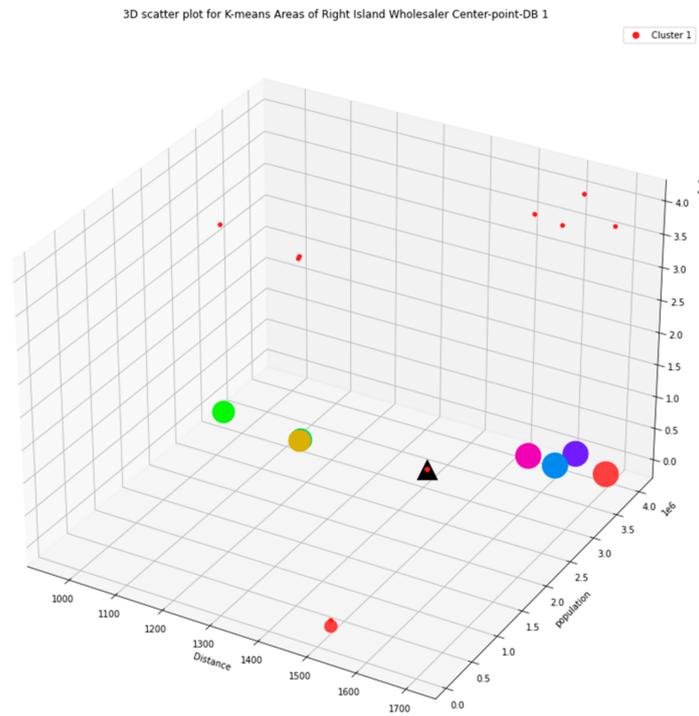


Figure 22. Weighted 3D view of the cluster and centroid of states and federal territories of the right island of Malaysia.

In the following, the main factory and the wholesaler center points will be shown on the map of Malaysia to clarify the designed supply chain (Figures 23 and 24).



Figure 23. The designed supply chain including main factory and wholesaler center points for each region in the left island of Malaysia [70].



Figure 24. The designed supply chain including main factory and wholesaler center points for each region in the right island of Malaysia [70].

The suggested locations by the proposed method are in accordance with the strategies of the company’s top manager that were mentioned in Section 4.6 in terms of number of the wholesalers (6 points) and population/distance ratio calculated by Equation (5) ([6.753, 97.487, 7.257, 13.244, 21.376]).

The proposed algorithm can determine the best location of the center points according to the geographical and population features. The algorithm can be combined to a cloud-based application to take the real time data of Malaysia for future updates accordingly. As the next step (future studies), designing the supply chain and a transportation planning method can be proposed for producing and distributing the products to the wholesalers to minimize the transportation cost according to the designed supply chain.

#### 4.10. Measuring Performance of the Proposed Method

In this section, the performance of the proposed method will be evaluated using silhouette and calinski\_harabasz scores. The higher the values for both scores the better the performance of the proposed method (Figures 25 and 26).

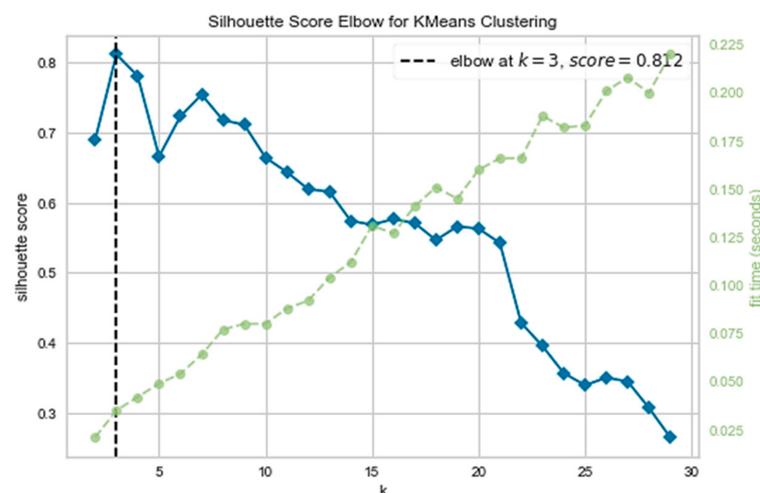


Figure 25. Results of silhouette score for the proposed method (k = 3).

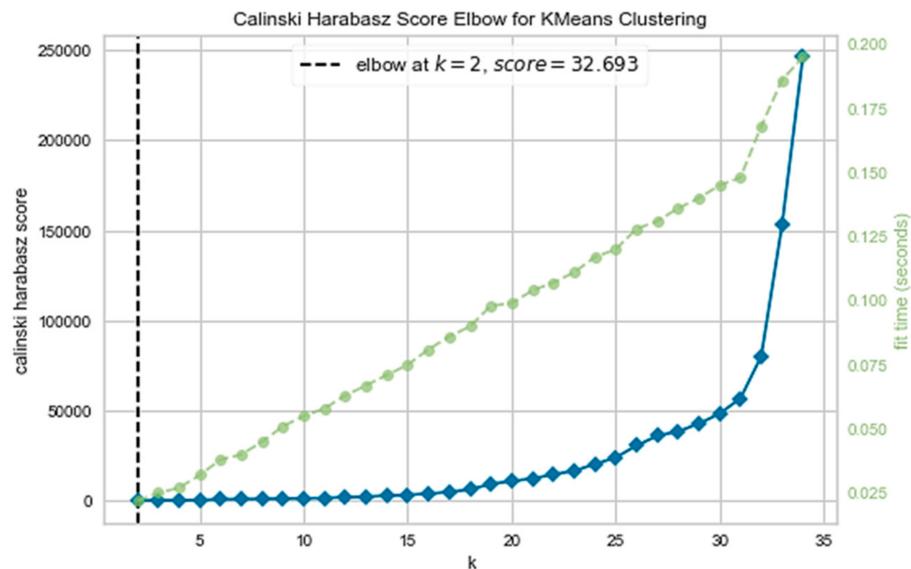


Figure 26. Results of silhouette score for the proposed method ( $k = 2$ ).

## 5. Conclusions

In this research, an unsupervised machine learning algorithm is proposed for transforming a local factory into a national supply chain in Malaysia. For this purpose, a hybrid PCA and K-means is used to figure out the wholesalers' center-points according to the geographical and population of the markets in Malaysia. The method is applied for transferring a local chocolate manufacturing company near Kuala Lumpur into a national-wide supply chain according to the top managers' strategies. To this end, and according to the regions recognized in objective two, four wholesalers on the left island of Malaysia are recognized in the north, right, central and south areas. Similarly, two wholesalers were identified on the right island, which were in Sarawak and WP Labuan. The outcomes indicate that the machine learning algorithms can be successfully used to design supply chain networks by considering real features, such as geographical, population, and supply chain features.

The outcomes indicated that machine learning algorithms can be considered as an appropriate way to determine the supply chain nodes considering the population and distance with high score (0.812). Such an approach will help top managers to minimize the transportation cost.

Further expansion of this research is recommended by considering the internal factors of the local factory (such as infrastructures, human resources, machinery, and available transportation system) while designing the supply chain. In addition, the method could be combined with a cloud-based application to reflect the real population regional data during production planning.

**Author Contributions:** Conceptualization, M.K.A.B.M.A. and E.E.B.S.; Methodology, M.F.B.M.A., M.K.A.B.M.A. and F.B.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Authors would like to thank anonymous reviewers and the editor for their positive comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Statista. Available online: <https://www.statista.com/statistics/318732/share-of-economic-sectors-in-the-gdp-in-malaysia/> (accessed on 20 February 2021).
2. Husin, M.M.; Kamarudin, S.; Rizal, A.M. Food and beverage industry competitiveness and halal logistics: Perspective from small and medium enterprises in Malaysia. *Asian J. Islam. Manag.* **2021**, *3*, 1–10. [[CrossRef](#)]
3. Ceicdata. Available online: <https://www.ceicdata.com/en/malaysia/number-of-bankruptcies/number-of-bankruptcies> (accessed on 21 February 2021).
4. Fornasiero, R.; Macchion, L.; Vinelli, A. Supply chain configuration towards customization: A comparison between small and large series production. *IFAC-PapersOnLine* **2015**, *48*, 1428–1433. [[CrossRef](#)]
5. Şen, A. The US fashion industry: A supply chain review. *Int. J. Prod. Econ.* **2008**, *114*, 571–593. [[CrossRef](#)]
6. Montagna, G. Multi-dimensional Consumers: Fashion and Human Factors. *Procedia Manuf.* **2015**, *3*, 6550–6556. [[CrossRef](#)]
7. Delgado, M.J.; Albuquerque, M.H.F. The Contribution of Regional Costume in Fashion. *Procedia Manuf.* **2015**, *3*, 6380–6387. [[CrossRef](#)]
8. Iannone, R.; Martino, G.; Miranda, S.; Riemma, S. Modeling Fashion Retail Supply Chain through Causal Loop Diagram. *IFAC-PapersOnLine* **2015**, *48*, 1290–1295. [[CrossRef](#)]
9. Macchion, L.; Moretto, A.; Caniato, F.; Caridi, M.; Danese, P.; Vinelli, A. Production and supply network strategies within the fashion industry. *Int. J. Prod. Econ.* **2015**, *163*, 173–188. [[CrossRef](#)]
10. Zilberman, D.; Lu, L.; Reardon, T. Innovation-induced food supply chain design. *Food Policy* **2019**, *83*, 289–297. [[CrossRef](#)]
11. Caniato, F.; Caridi, M.; Moretto, A.; Sianesi, A.; Spina, G. Integrating international fashion retail into new product development. *Int. J. Prod. Econ.* **2014**, *147*, 294–306. [[CrossRef](#)]
12. Chen, J.-M.; Chang, C.-I. Dynamic pricing for new and remanufactured products in a closed-loop supply chain. *Int. J. Prod. Econ.* **2013**, *146*, 153–160. [[CrossRef](#)]
13. Mehrjoo, M.; Pasek, Z. Impact of Product Variety on Supply Chain in Fast Fashion Apparel Industry. *Procedia CIRP* **2014**, *17*, 296–301. [[CrossRef](#)]
14. Zhou, E.; Zhang, J.; Gou, Q.; Liang, L. A two period pricing model for new fashion style launching strategy. *Int. J. Prod. Econ.* **2015**, *160*, 144–156. [[CrossRef](#)]
15. Yang, S.; Xiao, Y.; Kuo, Y.-H. The Supply Chain Design for Perishable Food with Stochastic Demand. *Sustainability* **2017**, *9*, 1195. [[CrossRef](#)]
16. Soolaki, M.; Arkat, J. Incorporating dynamic cellular manufacturing into strategic supply chain design. *Int. J. Adv. Manuf. Technol.* **2018**, *95*, 2429–2447. [[CrossRef](#)]
17. Delgoshaei, A.; Ariffin, M.K.; Baharudin, B.H.T.B.; Leman, Z. A backward approach for maximizing net present value of multi-mode pre-emptive resource-constrained project scheduling problem with discounted cash flows using simulated annealing algorithm. *Int. J. Ind. Eng. Manag.* **2014**, *5*, 151–158.
18. Allaoui, H.; Guo, Y.; Choudhary, A.; Bloemhof, J. Sustainable agro-food supply chain design using two-stage hybrid multi-objective decision-making approach. *Comput. Oper. Res.* **2018**, *89*, 369–384. [[CrossRef](#)]
19. Cohen, M.A.; Lee, H.L. Designing the Right Global Supply Chain Network. *Manuf. Serv. Oper. Manag.* **2020**, *22*, 15–24. [[CrossRef](#)]
20. Mogale, D.; Kumar, S.K.; Tiwari, M.K. Green food supply chain design considering risk and post-harvest losses: A case study. *Ann. Oper. Res.* **2020**, *295*, 257–284. [[CrossRef](#)]
21. Singh, S.; Kumar, R.; Panchal, R.; Tiwari, M.K. Impact of COVID-19 on logistics systems and disruptions in food supply chain. *Int. J. Prod. Res.* **2021**, *59*, 1993–2008. [[CrossRef](#)]
22. Ni, Y.; Fan, F. A two-stage dynamic sales forecasting model for the fashion retail. *Expert Syst. Appl.* **2011**, *38*, 1529–1536. [[CrossRef](#)]
23. Lo, C.K.Y.; Yeung, A.C.L.; Cheng, E. The impact of environmental management systems on financial performance in fashion and textiles industries. *Int. J. Prod. Econ.* **2012**, *135*, 561–567. [[CrossRef](#)]
24. Delgoshaei, A.; Ariffin, M.K.A.M.; Baharudin, B.T.H.T. Pre-emptive resource-constrained multimode project scheduling using genetic algorithm: A dynamic forward approach. *J. Ind. Eng. Manag.* **2016**, *9*, 732–785. [[CrossRef](#)]
25. Dye, C.-Y.; Hsieh, T.-P. An optimal replenishment policy for deteriorating items with effective investment in preservation technology. *Eur. J. Oper. Res.* **2012**, *218*, 106–112. [[CrossRef](#)]
26. Basu, P.; Nair, S.K. A decision support system for mean-variance analysis in multi-period inventory control. *Decis. Support Syst.* **2014**, *57*, 285–295. [[CrossRef](#)]
27. Ouyang, L.-Y.; Wu, K.-S.; Yang, C.-T. A study on an inventory model for non-instantaneous deteriorating items with permissible delay in payments. *Comput. Ind. Eng.* **2006**, *51*, 637–651. [[CrossRef](#)]
28. Hsu, P.; Wee, H.; Teng, H. Preservation technology investment for deteriorating inventory. *Int. J. Prod. Econ.* **2010**, *124*, 388–394. [[CrossRef](#)]
29. Widyadana, G.A.; Wee, H.M. Optimal deteriorating items production inventory models with random machine breakdown and stochastic repair time. *Appl. Math. Model.* **2011**, *35*, 3495–3508. [[CrossRef](#)]
30. Musa, A.; Sani, B. Inventory ordering policies of delayed deteriorating items under permissible delay in payments. *Int. J. Prod. Econ.* **2012**, *136*, 75–83. [[CrossRef](#)]
31. Mishra, V.K.; Singh, L.S.; Kumar, R. An inventory model for deteriorating items with time-dependent demand and time-varying holding cost under partial backlogging. *J. Ind. Eng. Int.* **2013**, *9*, 4. [[CrossRef](#)]

32. Mula, J.; Peidro, D.; Poler, R. The effectiveness of a fuzzy mathematical programming approach for supply chain production planning with fuzzy demand. *Int. J. Prod. Econ.* **2010**, *128*, 136–143. [[CrossRef](#)]
33. Qin, Z.; Bai, M.; Ralescu, D. A fuzzy control system with application to production planning problems. *Inf. Sci.* **2011**, *181*, 1018–1027. [[CrossRef](#)]
34. MacCarthy, B.L.; Jayarathne, P. Supply network structures in the international clothing industry: Differences across retailer types. *Int. J. Oper. Prod. Manag.* **2013**, *33*, 858–886. [[CrossRef](#)]
35. Delgoshaei, A.; Ariffin, M.K.A.M.; Leman, Z.; Bin Baharudin, B.T.H.T.; Gomes, C. Review of evolution of cellular manufacturing system's approaches: Material transferring models. *Int. J. Precis. Eng. Manuf.* **2016**, *17*, 131–149. [[CrossRef](#)]
36. Paciarotti, C.; Torregiani, F. The logistics of the short food supply chain: A literature review. *Sustain. Prod. Consum.* **2021**, *26*, 428–442. [[CrossRef](#)]
37. Macchion, L.; Moretto, A.; Caniato, F.; Caridi, M.; Danese, P.; Spina, G.; Vinelli, A. Improving innovation performance through environmental practices in the fashion industry: The moderating effect of internationalisation and the influence of collaboration. *Prod. Plan. Control* **2017**, *28*, 190–201. [[CrossRef](#)]
38. Levner, E.; Proth, J.-M. Strategic Management of Ecological Systems: A Supply Chain Perspective. In *Strategic Management of Marine Ecosystems*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 95–107.
39. Levner, E. Risk/Cost Analysis of Sustainable Management of Wastewater for Irrigation: Supply Chain Approach. In *Wastewater Reuse—Risk Assessment, Decision-Making and Environmental Security*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 33–42.
40. Nagurney, A.; Yu, M. Sustainable fashion supply chain management under oligopolistic competition and brand differentiation. *Int. J. Prod. Econ.* **2012**, *135*, 532–540. [[CrossRef](#)]
41. Yang, L.; Ng, C.T. Flexible capacity strategy with multiple market periods under demand uncertainty and investment constraint. *Eur. J. Oper. Res.* **2014**, *236*, 511–521. [[CrossRef](#)]
42. Lion, A.; Macchion, L.; Danese, P.; Vinelli, A. Sustainability approaches within the fashion industry: The supplier perspective. *Supply Chain Forum Int. J.* **2016**, *17*, 95–108. [[CrossRef](#)]
43. Macchion, L.; Da Giau, A.; Caniato, F.; Caridi, M.; Danese, P.; Rinaldi, R.; Vinelli, A. Strategic approaches to sustainability in fashion supply chain management. *Prod. Plan. Control* **2018**, *29*, 9–28. [[CrossRef](#)]
44. Moretto, A.; Macchion, L.; Lion, A.; Caniato, F.; Danese, P.; Vinelli, A. Designing a roadmap towards a sustainable supply chain: A focus on the fashion industry. *J. Clean. Prod.* **2018**, *193*, 169–184. [[CrossRef](#)]
45. Theodoridis, S.; Pikrakis, A.; Koutroumbas, K.; Cavouras, D. *Introduction to Pattern Recognition: A Matlab Approach*; Academic Press: Cambridge, MA, USA, 2010.
46. Chitta, R.; Murty, M.N. Two-level k-means clustering algorithm for k- $\tau$  relationship establishment and linear-time classification. *Pattern Recognit.* **2010**, *43*, 796–804. [[CrossRef](#)]
47. Zhang, B.; Hsu, M.; Dayal, U. *K-Harmonic Means—A Data Clustering Algorithm*; Hewlett-Packard Labs Technical Report HPL-1999-124; Hewlett-Packard Company: Palo Alto, CA, USA, 1999.
48. Ünler, A.; Güngör, Z. Applying K-harmonic means clustering to the part-machine classification problem. *Expert Syst. Appl.* **2009**, *36*, 1179–1194. [[CrossRef](#)]
49. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
50. Won, Y.; Chang Lee, K. Modified p-median approach for efficient GT cell formation. *Comput. Ind. Eng.* **2004**, *46*, 495–510. [[CrossRef](#)]
51. Ashayeri, J.; Heuts, R.; Tammel, B. A modified simple heuristic for the p-median problem, with facilities design applications. *Robot. Comput.-Integr. Manuf.* **2005**, *21*, 451–464. [[CrossRef](#)]
52. Won, Y.; Currie, K.R. An effective p-median model considering production factors in machine cell/part family formation. *J. Manuf. Syst.* **2006**, *25*, 58–64. [[CrossRef](#)]
53. Goldengorin, B.; Krushinsky, D.; Slomp, J. Flexible PMP Approach for Large-Size Cell Formation. *Oper. Res.* **2012**, *60*, 1157–1166. [[CrossRef](#)]
54. Krushinsky, D.; Goldengorin, B. An exact model for cell formation in group technology. *Comput. Manag. Sci.* **2012**, *9*, 323–338. [[CrossRef](#)]
55. Paydar, M.M.; Saidi-Mehrabad, M. A hybrid genetic-variable neighborhood search algorithm for the cell formation problem based on grouping efficacy. *Comput. Oper. Res.* **2013**, *40*, 980–990. [[CrossRef](#)]
56. Izakian, H.; Abraham, A. Fuzzy C-means and fuzzy swarm for fuzzy clustering problem. *Expert Syst. Appl.* **2011**, *38*, 1835–1838. [[CrossRef](#)]
57. Angra, S.; Sehgal, R.; Noori, Z.S. Cellular manufacturing—A time-based analysis to the layout problem. *Int. J. Prod. Econ.* **2008**, *112*, 427–438. [[CrossRef](#)]
58. Oliveira, S.; Ribeiro, J.; Seok, S. A spectral clustering algorithm for manufacturing cell formation. *Comput. Ind. Eng.* **2009**, *57*, 1008–1014. [[CrossRef](#)]
59. Banerjee, I.; Das, P. Group technology based adaptive cell formation using predator-prey genetic algorithm. *Appl. Soft Comput.* **2012**, *12*, 559–572. [[CrossRef](#)]
60. Kao, Y.; Li, Y.L. Ant colony recognition systems for part clustering problems. *Int. J. Prod. Res.* **2008**, *46*, 4237–4258. [[CrossRef](#)]

61. Yang, F.; Sun, T.; Zhang, C. An efficient hybrid data clustering method based on K-harmonic means and Particle Swarm Optimization. *Expert Syst. Appl.* **2009**, *36*, 9847–9852. [[CrossRef](#)]
62. Nouri, H.; Tang, S.; Tuah, B.H.; Anuar, M. BASE: A bacteria foraging algorithm for cell formation with sequence data. *J. Manuf. Syst.* **2010**, *29*, 102–110. [[CrossRef](#)]
63. Chattopadhyay, M.; Dan, P.K.; Mazumdar, S. Application of visual clustering properties of self organizing map in machine-part cell formation. *Appl. Soft Comput.* **2012**, *12*, 600–610. [[CrossRef](#)]
64. Kuo, R.; Su, Y.; Chiu, C.; Chen, K.-Y.; Tien, F. Part family formation through fuzzy ART2 neural network. *Decis. Support Syst.* **2006**, *42*, 89–103. [[CrossRef](#)]
65. Özdemir, R.G.; Gençyılmaz, G.; Aktin, T. The modified fuzzy art and a two-stage clustering approach to cell design. *Inf. Sci.* **2007**, *177*, 5219–5236. [[CrossRef](#)]
66. Yang, M.-S.; Yang, J.-H. Machine-part cell formation in group technology using a modified ART1 method. *Eur. J. Oper. Res.* **2008**, *188*, 140–152. [[CrossRef](#)]
67. Pandian, R.S.; Mahapatra, S. Manufacturing cell formation with production data using neural networks. *Comput. Ind. Eng.* **2009**, *56*, 1340–1347. [[CrossRef](#)]
68. Statistica. Available online: <https://www.statista.com/statistics/1040670/malaysia-population-distribution-by-state/> (accessed on 15 May 2021).
69. Distance Calculator. Available online: <https://www.distancecalculator.net/> (accessed on 15 May 2021).
70. Pinterest. Available online: <https://www.pinterest.com/pin/468022586283801336/> (accessed on 15 May 2021).