

University of Wales Trinity Saint David

DATA HANDLING AND BUSINESS INTELLIGENCE

-LCBB5000-

~~Student name~~

~~Student number~~

Lecturer: Md. Nazmul Hasan

Part 1

The first part of the report analyses the data of a superstore to determine the decline in sales and profit over the years. For this part, the analysis has been carried out on Microsoft Excel for pre-processing the data and visualizing it for a more in-depth explanation.

Significance of data in business

Data offers the structure for the info to be imparted with others. Hence, according to Cracana et al. (2019), data handling is vital in ensuring data integrity (in this case, the business data) because it addresses concerns associated with safety, maintenance, and confidentiality of the business data. Appropriate planning for data handling can also provide upshots in effective and economical storage, recovery, and data removal (Cracana et al., 2019). Data is additionally of utmost significance in the business as data analytics can assist a business in forecasting the customers' behavior, enhance decision-making, market trends, and ascertain the return of investment (ROI) of its marketing endeavours. According to Desai (2019), the data need to be relevant when analyzing it for business purposes to comprehend the users, the market reach, and the competition, as the business results and performance depend on the analysis of data entirely.

The analysis of data was done on MS Excel, utilizing suitable summary statistics, including the computation of standard statistics, i.e., mean, median, mode, range, and standard deviation. The findings were also shown by the visual representation through graphs and charts to further explain the findings. With the help of data visualization, the segments of data that require attention can be identified easily (Aparicio and Costa, 2015). Nowadays, accounting efficiency is based on a fundamental spreadsheet to various enterprise resource planning software (for example, SAP). This software helps accountants in interpreting the result (Lee et al., 2018).

Determining Sales and Profit for four years:

	2009	2010	2011	2012
Sales	4169325	3551642	3455452	3739182
Profit	421514	371072	384288	345106

(Table 1: Sales Vs Profit 2009-12)

The Superstore data shows the fluctuation in the sales and profit since 2009. The sales margin decreases from 2009 to 2010 by 14.81% and further decreases in 2011 by 2.7%. However, the sales data shows an increase in 2012 that rose to 3739181, which was reported at 3455452 at the end of 2011. This is the decrease in quantity purchased by customers in the affected years, as shown in Table 2. Sales decrease is where the firm's sales are being diminished, contrasting with the typical deals due to factors the organization might have missed to carry out. So, by this, the organization can even go on a misfortune when time goes and decrease proceeds (Thrun, 2018). All organisations worldwide exist to sell something and without deals, all organisations would go into bankruptcy. That's why it's consistently essential to discover the real reasons why clients are not accepting your items or administrations.



(Fig 1: Profit Vs Sales Superstore)

From the sales and profits of Superstore presented in Fig 1, it can be said that both sales and profits have been decreased significantly. Both sales and profits are directly proportional, particularly for the Superstore; therefore, since 2009, it can be said that the management of

Superstore has not been able to strike a better balance between their profit and sales which also putting the negative impact on their overall revenues and allowing their competitors to get a competitive lead over them. From figure 1, it can also be articulated that there is a decrease in profit from 2009 to 2010 and increases in 2011. However, the sales revenue decreased in 2011, but the profit margin is higher in 2011 than in 2010.

The reason behind it is the discount rate between 1 to 9% but reaches 10% in 2010. The more discount rate decreases the portion of profit in sales revenue. This condition and uncertain situation that Superstore is facing have also been discussed in Poinar et al., (2020). However, the core focus was on evaluating the impact of sales on profits. From the analysis, it was found that the problem of the declining profit rates in the modern economies comes to the front and becomes effective considering the stoppage in monetary development and the stagnation and emergencies inclinations. These inclinations have been examined for the most part inside Keynesian, post-Keynesian. With benefit rates decay being just one of the numerous reasons of the precariousness and crisis-prone nature of the private enterprise, Marxist ways of thinking, different clarifications including underutilization issue, capitalism's anarchic nature and absence of preparation, over the top financialization, among others (Trofimov, 2018).

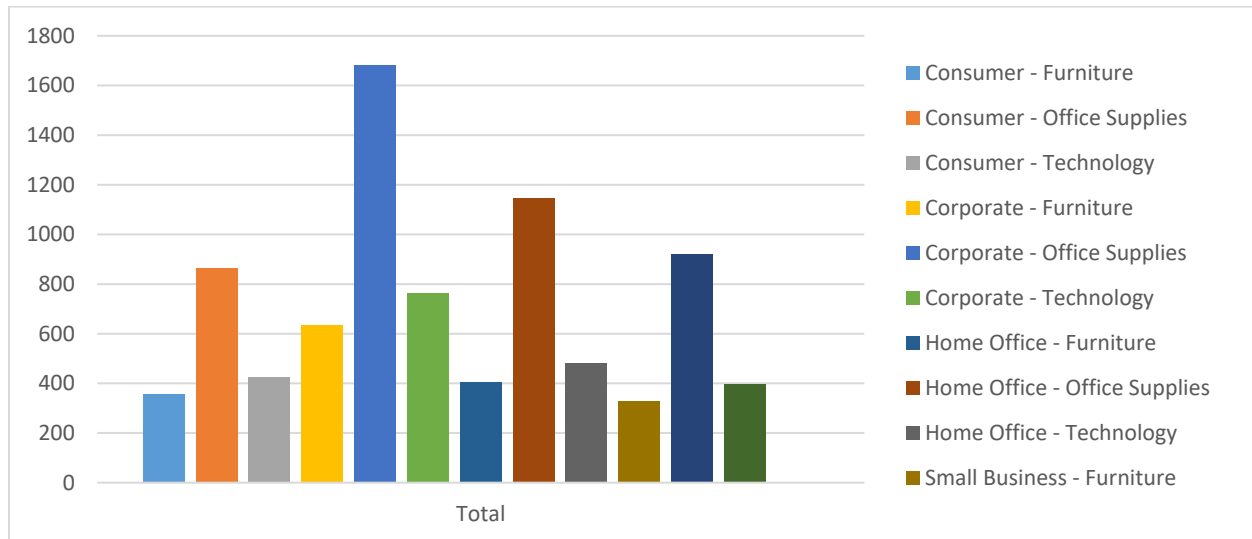
Order Quantity

This section defines the orders and how they have been received over the years. Therefore, this section aims to define the number of orders that the management of Superstore has received during the said frame of time.

2009	2010	2011	2012
54040	54155	51891	54589

(Table 2: Order Quantity)

From table 2, it can be illustrated that during 2009, the numbers of orders booked for the delivery were 54040, which increased in 2010 to 54155 and then decreased to 51891 when they were calculated at the end of 2011. However, it again increased in 2012, which were found to be recorded at 54589 highest compared to the previous three years. This shows that when the sales and profits were decreased, the biggest reason was the decrease in order booking, the years when the highest orders were booked, in those years, the management of Superstore was able to get sufficient level of sales and revenues. However, the decrease in order booking has also negatively impacted the sales and revenues.

Analysis through Pivot Table for Profit:

(Pivot Table 1)

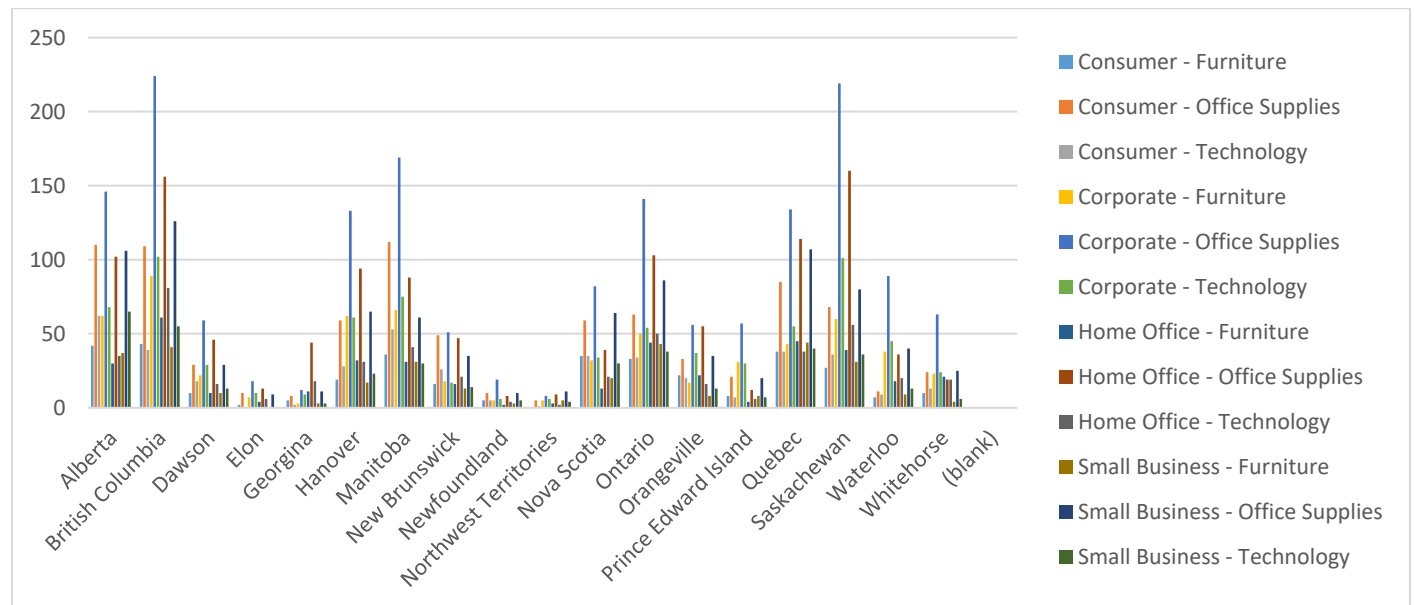
The Pivot mentioned above Table 1 interprets the count of sales and profit segmented according to different customer segments and product categories of the Superstore. With the help of the Pivot table, it is easy to understand the consumption of products in different regions and provinces. The customers of the Superstore are segmented into corporate, home office, small business, and consumer. However, the profit for consumer furniture and corporate technology is relatively high compared to the profit of corporate furniture, consumer technology, and corporate office supplies. Moreover, this pivot table demonstrates the comparison among the profits of customer segments. For example, the profit of home office technology is 400; on the other hand, corporate technology shows approximately 760.

From the analysis of the results, it can be interpreted that right now, if the data is calculated for the last four years, such as 2009, 2010, 2011, and 2012, the Superstore has been able to target consumers who like to purchase products and services for their homes, corporate companies that avail the products for the office use, and home office for the employees who work from home. From the analysis made in the Pivot Table 1, it can be said that furniture that the consumers buy is something that is purchased by most of the customers during the said time frame. After this, office supplies stand at second and home office stands at third. Therefore, to ensure that Superstore can compete with the competitors, the management must ensure that they are constantly changing these products according to the changing or varying needs and

preferences of their customers so that their customers can feel the sense of loyalty and value and customize the products with their vital feedback.

However, small business furniture and home furniture are the two products that were not sold as they were supposed to be; therefore, these two areas need to be improved to ensure the quality of services and products.

Analysis through Pivot Table for Order Quantity:



(Pivot Table 2)

The Pivot Table 2 defines the quantity ordered by the consumers in the different segments and according to different regions and provinces. For example, the consumers in Ontario and the West region are above 400. However, if we choose the same scenario for the Yukon region, the order quantity for corporate office supplies is 120. The pivot table 2 might be used to interpret the order quantity of different products, but in general, it shows that corporate office supplies demand higher. On the other hand, consumer office supplies are second in demand. For the Atlantic region, the demand for consumer office supplies is at 140.

On the other hand, the demand for consumer office supplies for the Yukon region is between 40 and 60 but stood second in demand in the same region. Moreover, if we interpret the province segment, corporate and consumer office supplies stood the same. If we link quantity ordered and profit, it is seen that the higher the order quantity, the higher will be the profit. For example, in Pivot Table 2, corporate office supplies are highly demanded. Therefore, the margin for this product and segment in all the regions and provinces interprets the highest profit level.

Part 2

Data Analysis through Excel and SPSS:

This part of the analysis shows the descriptive statistics. Descriptive statistics are used to define the data described in terms of numbers representing data extracted for research. Moreover, it is helpful to easily understand which part statistic measure is feasible for the research (Rodrigues et al., 2017). Descriptive statistics involve two parts one is the central tendency, and the other is dispersion. Our focus is on the central tendency. This part of the research segments the individual in two parts,

- Customers of Smile Clinic who eat rice
- Customer of Smile Clinic who do not eat rice.

According to the descriptive statistics data, ten males and thirty females don't eat rice. However, on the same side, there are forty males and twenty females who eat rice.

<i>Customer</i>	<i>Those who eat rice</i>	<i>Those who do not eat rice</i>
<i>Male</i>	<i>40</i>	<i>10</i>
<i>Female</i>	<i>20</i>	<i>30</i>

(Table 3: Those who eat/not eat Rice)

Rice					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	40	36.7	40.0	40.0
	1.00	60	55.0	60.0	100.0
	Total	100	91.7	100.0	
Missing	System	9	8.3		
Total		109	100.0		

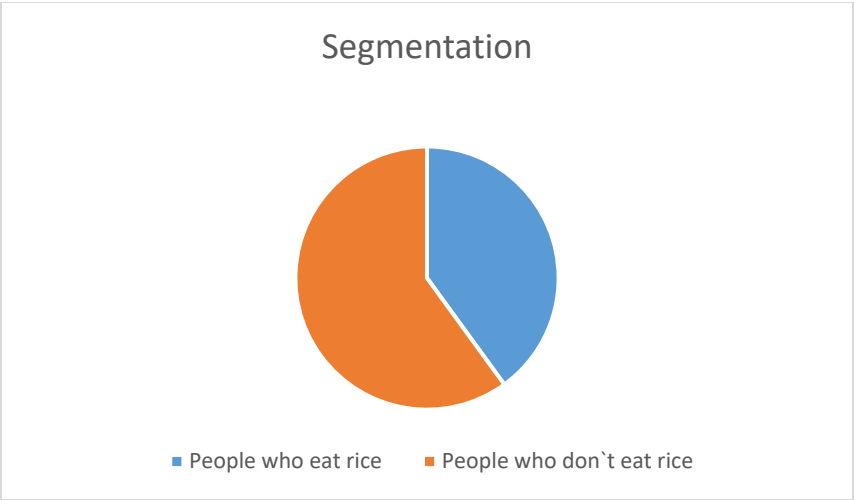


Fig 2: Segmentation Description through Pie chart

We have divided the data into two segments. Furthermore, we plot Pie chart that easily interprets the segmentation of data samples through percentage slices. With the help of the pie chart, we interpret that 60% of the data comprises on the customer of Smile Clinic who eat rice and 40% participation represents the customer of Smile Clinic who don't eat rice. This shows that people who eat rice daily take care of their health in maintaining the vital minerals required to be significant for being healthy.

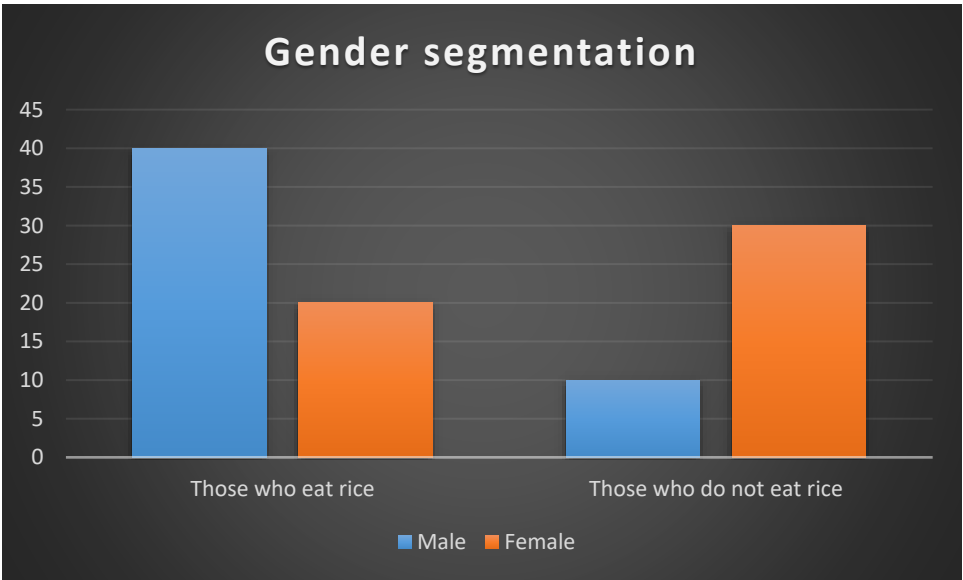


Fig 3: Gender Segmentation

Fig 4 defines the segmentation of the sample mentioned in the pie chart into male and female customers who eat rice and male and female customers who don't eat rice. From this, it can be interpreted that most of the males who participated in the proposed research tend to eat rice daily. At the same time, only 20% of the females, half of the male participants tend to eat rice daily.

Statistical Analysis through SPSS:



		Age	Rice
N	Valid	100	100
	Missing	9	9
Mean		20.3500	.6000
Median		19.0000	1.0000

Table 4: SPSS representation of Data

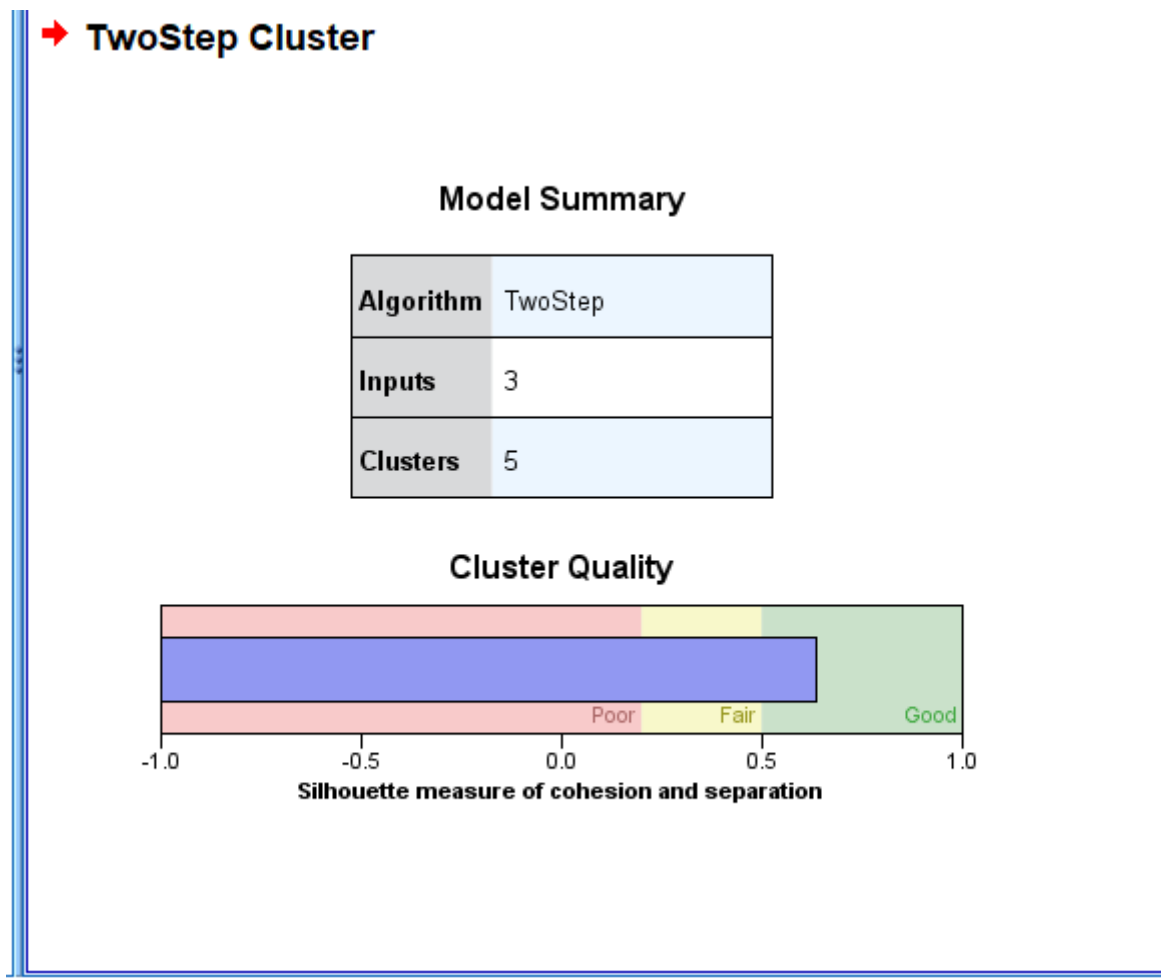
The SPSS represents the mean and median of 100 customers of Smile Clinic. According to the SPSS calculation, about 20.35 aged individuals (males and females) eat rice that lies under 60% and the mid-value of the data set is 19. The data interprets the same result as in Excel. But the main difference between both the software is that Excel is, in general, a spreadsheet with multiple functions. However, SPSS is proficient in statistics. It is a complete software that is relatively used for statistical analysis. By using SPSS, Smile Clinic gets more appropriate results. According to Segun Olabode and Akeem Bakare (2020), SPSS works on a point and click interface that helps the user pull down the menus and provide the appropriate command to perform.

a) Measures of cohesion and separation:

Clustering is different from machine learning methods because it focuses on the intrinsic structure, divisions, and data specifications. In this way, it is easy to compare the data in one cluster with the data of another cluster. However, the data and its validation and correct analysis

provide the basis for interpreting the data. The clustering analysis provides an analysis scale that is between -1 and +1. The +1 means that data in one cluster matches the data in another cluster.

Moreover, the clustering of data is essential because large number of data is incorporated in the sample that is different in features (Wang et al., 2017). Therefore, we calculate the mean coefficient value under silhouette measure to measure the cluster quality for one cluster to compare it. The clustering quality between 0.5 and 1 represents the good cluster quality, below 0.5 but above 0.2 represents fair cluster quality; however, below 0.2 shows the poor-quality cluster that means the data in one cluster does not match the data in another cluster. From the analysis made in fig 5, it can be said that the cluster quality falls within 0.5 to 1, which ensures that the articulated data for the proposed research is reliable and valid.



(Fig 5: Cluster Quality)

Fig.6 Cluster sizes

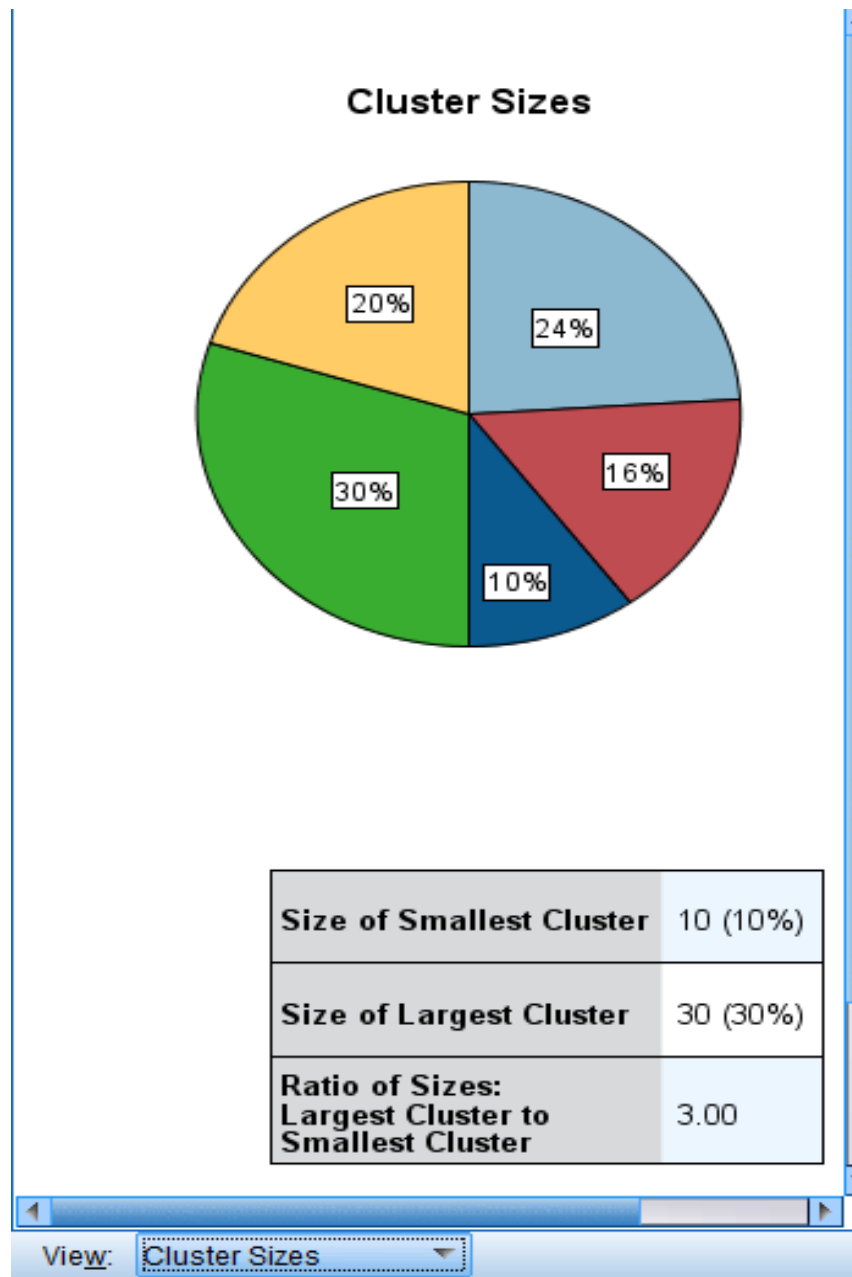
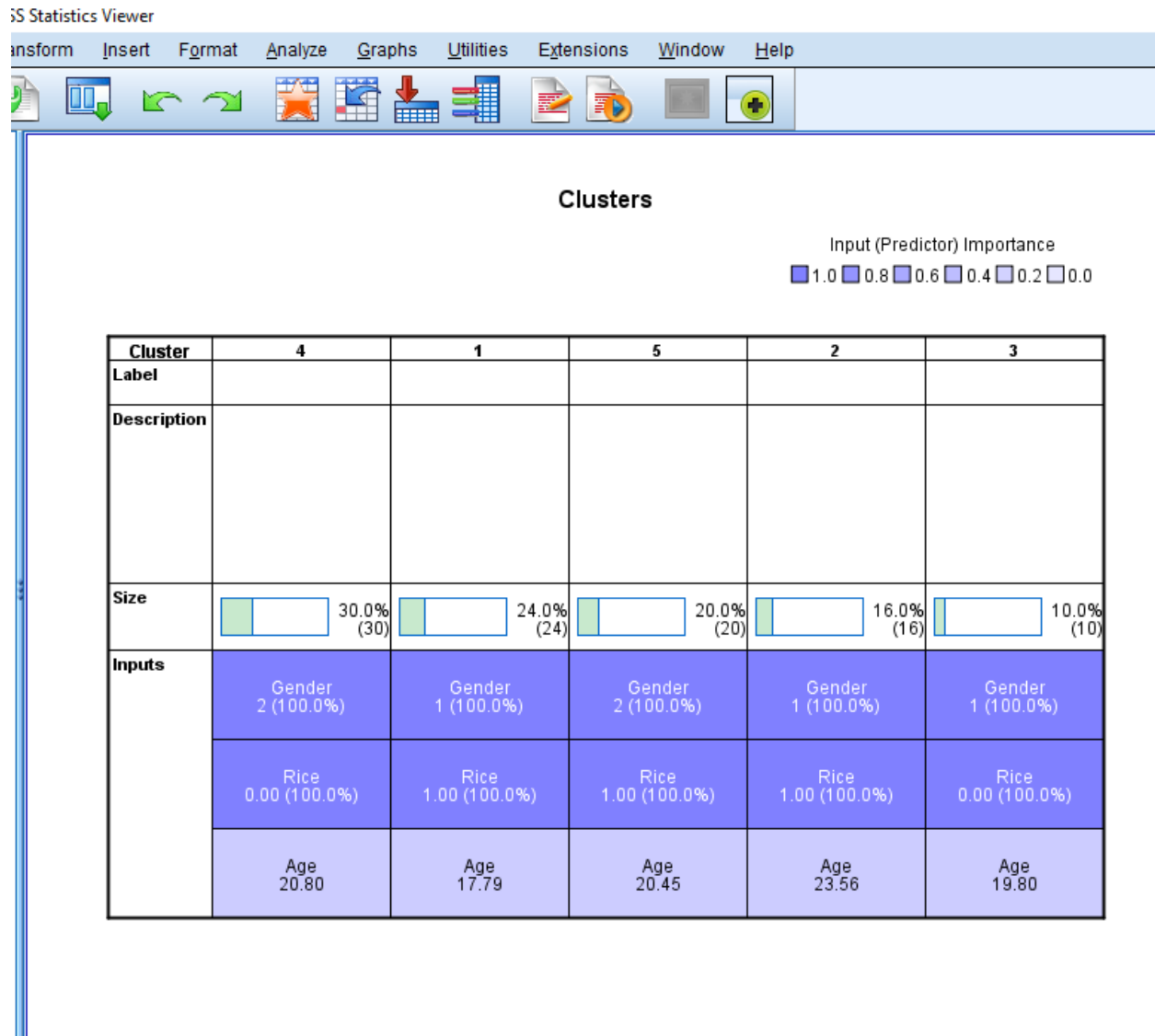


Fig. 7 Clusters



To get a deeper knowledge of the customers in fig 6 and 7 we divided the cluster analysis into gender, average age and eating/not eating categories and discovered that the male customers that are eating/not eating rice are younger than the female customers that are eating/not eating rice.

b) Methods of Data Mining and its Examples:

To extract the data from a wide range of populations, it is important to understand the relationship and the data pattern. For this purpose, the data mining process is an interesting and useful thing for researchers and data analysts. However, accurate data mining is important because it helps achieve the estimated profit margin, sustainability of performance, mitigate risk factors in obtaining goals and timely detection of errors and fraud. Various techniques help businesses to extract data for research. Knowledge Discovery of Databases (KDD) is a process that draws out relevant information for research purposes. Data mining is an important part of KDD (Alasadi, et al. 2017). Data mining techniques involve Data cleaning, tracking patterns, association, clustering, regression, prediction, and sequential pattern.

1. Tracking pattern

A tracking Pattern is a basic data mining pattern useful in identifying and tracking patterns in data. Monitoring and identifying the data is important to evaluate the future outcome of the business (Hooman et al., 2017). For example, Covid-19 negatively impacts the stock market. Therefore, the stock investment requires an in-depth analysis of past trends and patterns and track them while investing in the stock market.

2. Clustering and Regression

Data mining technique such as clustering, and regression is an important process model. This technique is used to analyze the data and interpret the support of SPSS in access to the model. For example, assigning tasks and setting objects is done in clustering to compare the data segments. On the other hand, regression is the statistical tool that relates the dependent variable in response to the other variable that is independent. Essentially, regression takes a mathematical dataset and fosters a numerical recipe that fits the information. For example, one research is made on Kish Island to gauge the growth of the experimental dataset and extract data geographically under this data mining technique (Hooman Fetanat et al., 2017).

c) SPSS vs Excel

Trading review information from Alchemy to SPSSs restrictive. SAV design makes the way toward pulling, controlling, and dissecting information perfect and simple. Utilizing the SAV design, SPSS consequently sets up and imports the assigned variable names, variable sorts, titles, and worth names, making the interaction a lot simpler on analysts. Conclusively, make sure

to utilize SPSS when researchers need an adaptable, adjustable approach to get very granular on even the most unpredictable informational collections (Arkkelin, 2014). This allows the scientists to do what they want to specialize in, recognize patterns, create prescient models, and reach educated inferences. Moreover, SPSS is an essential tool that depends on statistical data analysis. The following differences show the comparison between SPSS and Excel.

Source: (Analytica, 2019)	SPSS	Excel
Field	It incorporates the entire innovative region, which is considered as the superset of Data Science.	It is a subset of organization capacity where the data is made by applying various advancements and techniques.
Objectives	It is used for information control strategies to get substantial outcomes.	It is used to save the data and cautiously inspect it.
Usage	It rehearses for super quick and progressed gadgets like supercomputers and significantly more. Spss CSV file can contain large data sets along with the proper coding	It is supporting and managing a lot of customer information. However, excel files cannot contain coding-related information, making it difficult to interpret the results.

References

- Analytica, S. (2019) SPSS vs Excel: Which One is The Best Tool for Statistics, Statanalytica.
- Aparicio, M. and Costa, C.J., 2015. Data visualization. Communication design quarterly
- Arkkelin, D., 2014. Using SPSS to understand research and data analysis.
- BHAT, MA, IMPORTANCE OF SPSS FOR SOCIAL SCIENCES RESEARCH. PRAGYA, p.32.
- Cracana, A.I., Osman, H.D.M., Al-Shehab, S.M. and Ciocîrlan, V., 2019. Ethical data handling.
- Desai, R., 2019. Why is DATA important for your business? Available at: <https://towardsdatascience.com/how-important-is-data-for-your-business-c15a35c6935e>
- Divisi, D., Di Leonardo, G., Zaccagna, G. and Crisci, R., 2017. Basic statistics with Microsoft Excel: a review. Journal of thoracic disease, 9(6), p.1734.
- Hooman Fetanat, Leila Mortazavi, Far Narsis Zarshenas, Ciência e Natura, 2017, The Application of Data Mining Techniques in Agricultural Science, v. 37 Part 2 2015, p. 108–116
- Lee, L., Kerler, W. and Ivancevich, D., 2018. Beyond Excel: Software tools and the accounting curriculum. AIS Educator Journal, 13(1), pp.44-61.
- Poinar, G. and Ellenberger, S., 2020. Burmese amber fossils, mining, sales, and profits. Geoconservation Research, 3(1), pp.12-16.
- review, 3(1), pp.7-11.
- Rodrigues, C.F.D.S., Lima, F.J.C.D. and Barbosa, F.T., 2017. Importance of using basic statistics adequately in clinical research☆. Revista brasileira de anestesiologia, 67, pp.619-625.
- Segun Olabode and Akeem Bakare, 2020. A COMPARISON OF STATISTICAL SOFTWARE PACKAGES: A PRACTICAL APPLICATION OF MICROSOFT EXCEL AND IBM SPSS STATISTICS.
- Thrun, M.C., 2018. Approaches to Cluster Analysis. In Projection-Based Clustering through Self-Organization and Swarm Intelligence (pp. 21-31). Springer Vieweg, Wiesbaden.
- Trofimov, I.D., 2018. The secular decline in profit rates: time series analysis of a classical hypothesis.
- Wang, F., Franco-Penya, H.H., Kelleher, J.D., Pugh, J. and Ross, R., 2017, July. An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity. In International Conference on Machine Learning and Data Mining in Pattern Recognition (pp. 291-305). Springer, Cham.