BUSINESS INTELLIGENCE APPLICATIONS ON HOUSEHOLD EXPENDITURE SURVEY OF 2016 AND 2019
(Aplikasi Kecerdasan Perniagaan Mengenai Kajian Perbelanjaan Isi Rumah Tahun 2016 dan 2019)

SERAH LATHAN A/L SATHIAMOORTHI, SITI NOOR ASYIKIN BINTI MOHD RAZALI*,
MUHAMMAD AMMAR SHAFI & ROSSHAIRY ABD RAHMAN

ABSTRACT

According to The Malaysia Reserve, following the implementation of the new RM1,500 minimum wage, prices of basic items in the country will likely rise by 50 percent to 60 percent beginning in June. This will affect the manufacturing cost. Due to increased manufacturing costs and inflation, beverage firms have been under a great deal of pressure to boost their pricing. This will impact the Malaysian household expenditure. The spending of consumer will increase drastically. Therefore, a proper study must be conducted on the expenditures among Malaysian people to understand the trend and current statistics of various races, income, occupation, and education level. In this research, descriptive analysis, tree decomposition method, and scatter plot clustering has been applied to examine household expenditure. Microsoft Power Business Intelligence (Power BI) has been applied as a tool to visualize the statistic, pattern or trends that could be explored through an attractive dashboard. The dashboard represents the monthly expenditure averages by state, marital status, strata, and gender. In addition, average total income and expenditures by occupation, ethnicity, highest level of education, and industry also presented that could be explored by user in detail. Result shows that the income and expenditures in Malaysia have been increased from 2016 to 2019. According to the ethnicity, Chinese concur the highest monthly expenses in both 2016 and 2019 compared to other races. Meanwhile based on industry, activities of household as employers has the highest expenses and occupation as a manager is also represents the highest income as well as highest monthly expenses. Furthermore, based on education level, Malaysian people who studied until degree and above has highest average income and expenses monthly. Future works could be done by analyzing factors that affect the increases of expenses among Malaysian people.

Keywords: analyze; descriptive; business intelligence; Power BI

ABSTRAK

Perbelanjaan pengguna telah meningkat secara mendadak. Oleh itu, kajian yang sewajarnya perlu dilakukan terhadap perbelanjaan dalam kalangan rakyat Malaysia untuk memahami trend dan statistik semasa pelbagai kaum, pendapatan, pekerjaan dan tahap pendidikan. Dalam penyelidikan ini, analisis deskriptif, kaedah penguraian pokok, dan pengelompokan plot taburan telah digunakan untuk mengkaji perbelanjaan isi rumah. Microsoft Power Business Intelligence (Power BI) telah digunakan sebagai alat untuk menggambarkan statistik, corak atau trend yang boleh diterokai melalui papan pemuka yang menarik. Papan pemuka mewakili purata perbelanjaan bulanan mengikut negeri, status perkahwinan, strata dan jantina. Di samping itu, purata jumlah pendapatan dan perbelanjaan mengikut pekerjaan, etnik, tahap pendidikan tertinggi dan industri juga dibentangkan yang boleh diterokai oleh pengguna secara terperinci. Keputusan menunjukkan bahawa pendapatan dan perbelanjaan di Malaysia telah meningkat dari 2016 hingga 2019. Mengikut etnik, Cina bersetuju dengan perbelanjaan bulanan tertinggi pada 2016 dan 2019 berbanding kaum lain. Manakala berdasarkan industri, aktiviti isi rumah sebagai majikan mempunyai perbelanjaan tertinggi dan pekerjaan sebagai pengurus juga mewakili pendapatan tertinggi serta perbelanjaan bulanan tertinggi. Tambahan

pula, berdasarkan tahap pendidikan, rakyat Malaysia yang belajar sehingga ijazah dan ke atas mempunyai purata pendapatan dan perbelanjaan bulanan tertinggi. Kerja masa hadapan boleh dilakukan dengan menganalisis faktor-faktor yang mempengaruhi peningkatan perbelanjaan dalam kalangan rakyat Malaysia.

Kata kunci: analisis; deskriptif; kecerdasan perniagaan; Power BI

1. Introduction

Descriptive analysis is defined as the form of data analysis that aids in describing, illustrating, or summarizing data points in a constructive manner so that patterns might develop that satisfy all data conditions (Rawat 2021). In this research, descriptive analysis is the primary analytic technique. The purpose of descriptive research is to describe a phenomenon and its characteristics. In contrast, a descriptive method emphasizes reporting the events that have occurred and statistically assessing the data (Nassaji 2015). In industrial, governmental, and research settings, descriptive analysis is a versatile source of product information because it delivers objective, comprehensive, and relevant data (Kemp *et al.* 2018). During descriptive analysis, the mean or average of the data set was computed since it serves as the foundation for all future analyses. It aids in determining the general trend and is simple to calculate (Zhang & Yang 2017). Diagnostic analytics is a subfield of analytics aimed at answering the question "Why did this occur?" Using diagnostic analytics, businesses may get insight into the underlying reasons of observed data trends (Holliday 2021). The diagnostic analysis delves further into the sources of the outcomes discovered by descriptive analytics (Gibson 2021).

There is certain research have been carried out using descriptive analysis on household expenditure. In one of the research by Punt, analysed the impact of different household income levels on household spending habits and how they change among regions and ethnic groups and in her analysis used 1995 IES data (Punt 2003). In another research by Ahmad, mentioned that in a descriptive analysis aside from household income, other characteristics such as race and location of homes also influence household consumption patterns (Ahmad & Fatima 2011). Cingona used panel data on families to figure out the short-term and long-term income elasticity of different types of spending (Cingano 2014).

Classification, clustering, logistic regression, decision tree, and time series are top predictive analytics models (Gupta 2020). In a research study, Luque et al. (2017) had done categorization systems exploited household income, consumption, and spending data. Okori and Obua have done research on predicting food insecurity using decision tree and random forest classification methods (Okori & Obua 2011). Etaati, mentioned that cluster analysis organizes items so that those in the same group called a cluster are more similar than those in other groups (Etaati 2017). Dogan, intends to analyze household catastrophic health expenditures using the clustering technique (Dogan *et al.* 2019).

Another study by Niu and friends were done using the random forest algorithm to measure poverty in urban areas (Niu *et al.* 2020). Koc and Yalcin, mentioned the classification model's training method differentiates between negative and positive events (Koc & Yalcin 2021). Boobier, proposed that clustering is a means of putting objects together because they are like each other (Boobier 2018). Cluster analysis groups households based on budget allocation similarities by maximizing within-group similarities and between-group disparities (Johnson & Wichern 2007). Fan and friends said this cluster analysis results in multiple clusters of households, each with a different food spending trend (Fan *et al.* 2007).

According to The Malaysia Reserve, following the implementation of the new RM1,500 minimum wage, prices of basic items in the country will likely rise by 50 percent to 60 percent beginning in June. This will affect the manufacturing cost. Due to increased manufacturing costs and inflation, beverage firms have been under a great deal of pressure to boost their pricing. This will impact the Malaysian household expenditure. The spending of consumer will increase drastically. Therefore, there must be a proper study regarding the expenditure of Malaysian people so that they can manage their expenditure accordingly. In this study descriptive and diagnostic analysis has been carried out to analyze the expenditure of Malaysians. For this study, the household expenditure data has been taken from Malaysia Department of Statistics (DOSM). It is one of the data suppliers on the Malaysian Open Data Portal. Proper research using business intelligence tools should be carried out in household expenditure. Using AI features of Power BI software will permit individuals to readily use data to uncover hidden, actionable insights that assist them in making better decisions and spending less. In this study, the first objective is to visualize the household expenditure survey in Malaysia by using descriptive analysis. The second objective is to investigate the Malaysian household expenditure using tree decomposition method. The third objective is to demonstrate the Malaysian household expenditure pattern using scatter plot clustering method.

2. Methodology

Major focus of this research is to employ descriptive and diagnostic analysis in household expenditure surveys using AI visualization. Prior to data analysis, data preparation is conducted. The descriptive analysis phase examines data or content to determine what has occurred or is occurring in the data's context. This is performed by descriptively identifying distinct expenditure patterns and relationships from the data. The data has been undergone data cleaning process by clearing all the missing datasets before starting the analysis. This process has been done in Microsoft Excel by find and clear method. The variable in the dataset is in numerical form, which needs to be changed to a categorical variable. This has been done using Microsoft Excel, where the description of the variable is already given in the DOSM dataset blocks.

2.1. *Descriptive analysis*

First, for descriptive analysis summary statistic has been generated. In summary statistics mean, mode, median, standard deviation, standard error, kurtosis, and skewness are labelled in Table 1 and Table 2. Then dashboard has been created using Power BI because it is the most fundamental form of data analysis and is required for predictive analysis. During descriptive analysis, the mean, of the data set has been calculated since it provides the foundation for all further analyses. It aids in determining the general trend and is simple to compute. The descriptive analysis then has been visualized in dashboard using Power BI which known as diagnostic analysis. The visualization is for household expenditure survey, both 2016 and 2019.

2.2. *Tree decomposition method*

In this study, expenditure has been classified according to the ethnic. In expenditure data there is Malay, Chinese, Indians, and others. So, the expenditure data has been classified among these 4 ethnic group in Malaysia. This classification comprises Average monthly total expenses, Ethnicity, Highest level of education, and Industry. Classification is created for two

distinct years, 2016 and 2019. Although there are several classification algorithms is existed, for this census data the decomposition tree method in Power BI has been used.

2.3. *Scatter plot*

In this study, scatter plot clustering has been used in Power BI tool. It is used to categorize household expenditure into low income and low expenses, middle income and middle expenses, and high income and high expense. During the analysis in Power BI its compulsory to include a column of the data in the 'Details' field of the visual because clustering is not available if do not include the data in detail. In this analysis the industries have been listed as details.

3. Results and Discussion

Figure 1 represents, the dashboard the analysis of household expenditure survey of 2016 and 2019. The analysis is consisting of average total monthly expenditure by states, marital status, strata, and gender. Apart from that, the analysis consist of average total income and expenditure by occupation, ethnic, highest certificate, and industry.

The average monthly expenses of living in urban areas rose from RM 3989 to RM4,483 from 2016 to 2019, an increase of 4.12 percent. The expenses of living in rural areas increased from RM2,720 in 2016 to RM3,060 in 2019 at an average annual rate of 4.16 percent. The average total monthly expenses of W.P Putrajaya has increased from RM 6615 to RM 7770 which shows an increase in 4.95 percent. W.P Putrajaya is the state with the highest cost of living in Malaysia. Meanwhile, the average total monthly expenses of Sabah have increased from RM 2862 to RM 3301 which shows an increase in 4.43 percent. Sabah has the least living cost in Malaysia. In both 2016 and 2019, the Chinese mean monthly income was greater than other ethnic groups, at RM 8860 and RM 10,130 per month, respectively. This was followed by the Indians with a monthly mean income of RM8,010 and the Bumiputera at RM7,080 in 2019. The Chinese also reported the highest median income in 2016 and 2019, with RM 7700 and RM 6800, respectively, while the Indians and Bumiputera reported RM5,981 and RM5,500, respectively. Chinese play the major role in expenses in Malaysia. Since they have most of the income, they tend to spend much.

From 2016 to 2019, monthly expenditures for Chinese, Indians, Bumiputera, and others all increased by 4.92%, 2.87%, 4.51%, and 3.51%, respectively, based on ethnicity. It has been recorded that the average income has increased by 4.17 % for Chinese, 3.03% for Indians, 4.31% for Bumiputera, and 4.65% for those of other ethnicities. For marital status the increase in expenses has been recorded for all. Especially married people who has highest increase in expenses around 5.07 percent and separated people who has lowest increase in their expenses around 3.22 percent. From 2016 to 2019 the Malaysian people who have a degree has an increase in expenses around 4.61 percent. On the other side, the people who do not have any certificate has an increase in expenses about 4.38 percent.
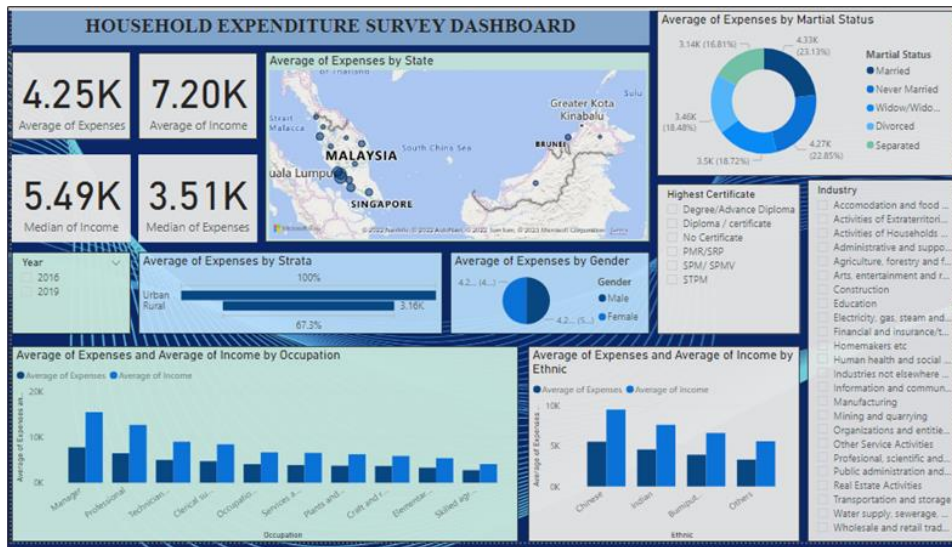
Figure 1: Household expenditure survey dashboard for 2016 and 2019

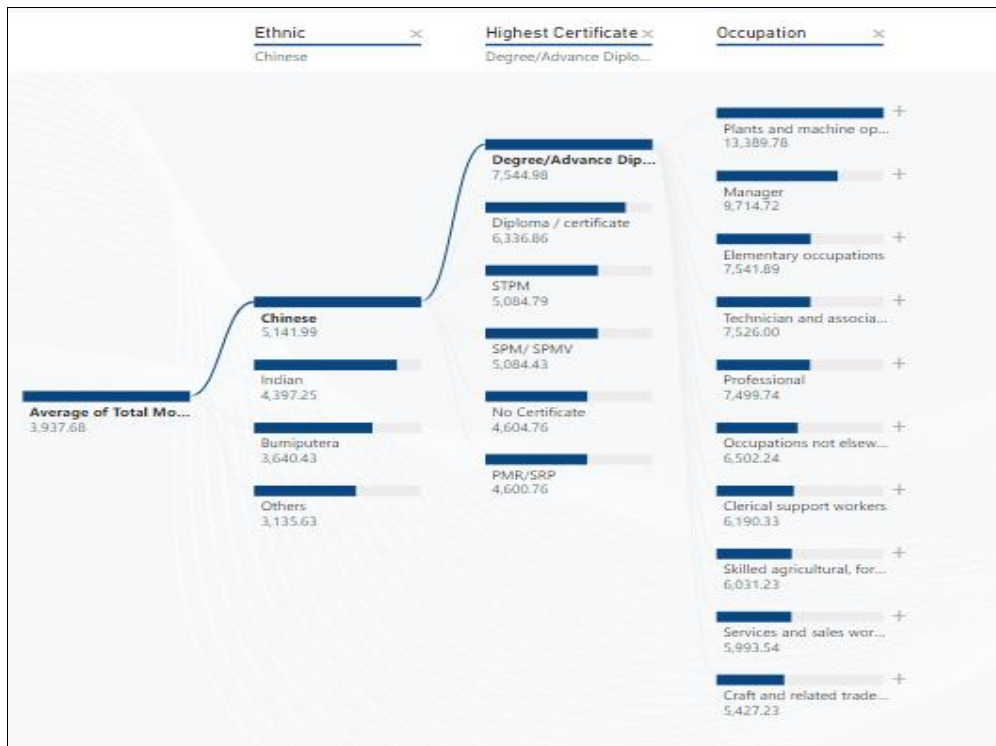### 3.1. *Tree decomposition method*

Figure 2: Tree decomposition method for expenses of Chinese 2016

Figure 3: Tree decomposition method for expenses of Chinese 2019

The Malaysian people who have degree has the most expenditures. According to the Figure 2 and Figure 3, the Chinese have 31.5 percent of the total expenses in Malaysia in 2016 and 32.05 percent in 2019. Although ethnic Bumiputera recorded the highest percentage of population with 68.6 per cent, followed by Chinese (23.4 %), Indians (7.0 %) and Others (1.0 %) but the percentage of Chinese and Indians of spending is higher than Bumiputera. Among Chinese, who got degree has higher expenses than others. The percentage of average expenses of Chinese who got degree is 23.16 percent. The lowest is the Chinese people who studied until PMR or SRP with 13.70 percent.

From the Figure 2 and Figure 3, from a total of RM 355 044.49, the expenditures of degree individuals are around RM 68562.2 in 2019. This represents around 19.31 percent of total average expenses. When further segmented by highest education level in degree, the data indicate that plant and machine operators and assemblers have the largest expenditures compared to other occupations. It represents around 17.67 percent of total expenses in the degree column. The lowest expenditure level in degree column is craft and related trades workers who has the total of 7.1 percent of total expenses in the degree column. In 2019, the data indicate that managers have the largest expenditures compared to other occupations with mean expenses around RM 11842.21. Although the plant and machine operators and assemblers have high expenses in 2016 but managers have become highest expenditure occupation in 2019. This happened because the income of plant and machine operators and assemblers' income has been reduced and managers income has increased.

Figure 4: Tree decomposition method for expenses of Indian 2016

Figure 5: Tree decomposition method for expenses of Indian 2019

Based on Figure 4 and Figure 5, Indian people get second place in spending although they are just 7% in overall population. In 2016 the percentage of average expenses of Indians who got degree is 25.62 percent. The lowest is the Indian people who studied until PMR or SRP with 12.69 percent. From a total of RM29,465.15, the expenditures of degree individuals are around RM 7330.59 per month. When the data are broken down by the highest level of education in a degree, they show that Managers spend the most money compared to other jobs. It's about 18.40 percent of all the expenses in the degree column. This is because a big population of the Indians has been working as managers in industries and getting paid high.

In 2019, the percentage that Indian people spend in Malaysia is 25.97 percent of all expenses. From a total of RM 260 843.29, the expenditures of degree individuals are around RM 52779.29. When further segmented in degree, the data indicate that managers have the largest expenditures compared to other occupations. Compared to 2016 the percentage has been decreased slightly. This is because the number of Indian people working as manager has decreased in 2019 compared to 2016. This effect the income and later effect the expenses as well.

Figure 6: Tree decomposition method for expenses of Bumiputera 2016

Figure 7: Tree decomposition method for expenses of Bumiputera 2019

Based on Figure 6 and Figure 7, Bumiputera has third highest expenses level than compared to all the other ethnicities. Among Bumiputera who got degree has higher expenses than others. The mean expenses of Bumiputera who got degree is RM 6060.90 in 2016 and RM 6840.92 in 2019. The lowest is the Bumiputera people who studied until PMR or SRP with 3276.09 in 2016 while RM 3724.93 in 2019. From a total of RM 260 843.29, degree-holding individuals spend around RM 52 779.29. This represents approximately 20.23 percent of average total expenses in 2016. When the data are further split by degree, managers had the highest expenses relative to other occupations. It accounts for approximately 15.67 percent of the total expenses in the degree column.

In 2019, from the total of RM 239 294.20, degree individuals spend around RM 55 474.25. This represents around 23.18 percent of average overall expenses. The results reveal that managers again have the highest expenses relative to other occupations when further split by highest education level in degree. It accounts for approximately 16.74 percent of the total expenses in the degree column. The expenses of managers of both years are higher than other occupations because the average income recorded for managers are higher than other occupations.
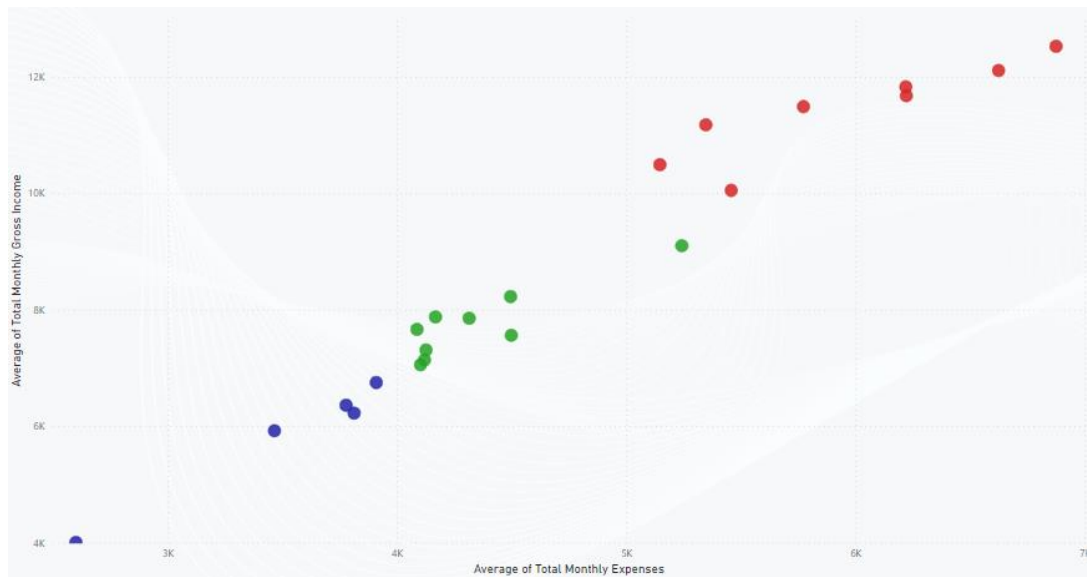
Figure 8a: The analysis of average of total monthly expense by industry in 2016
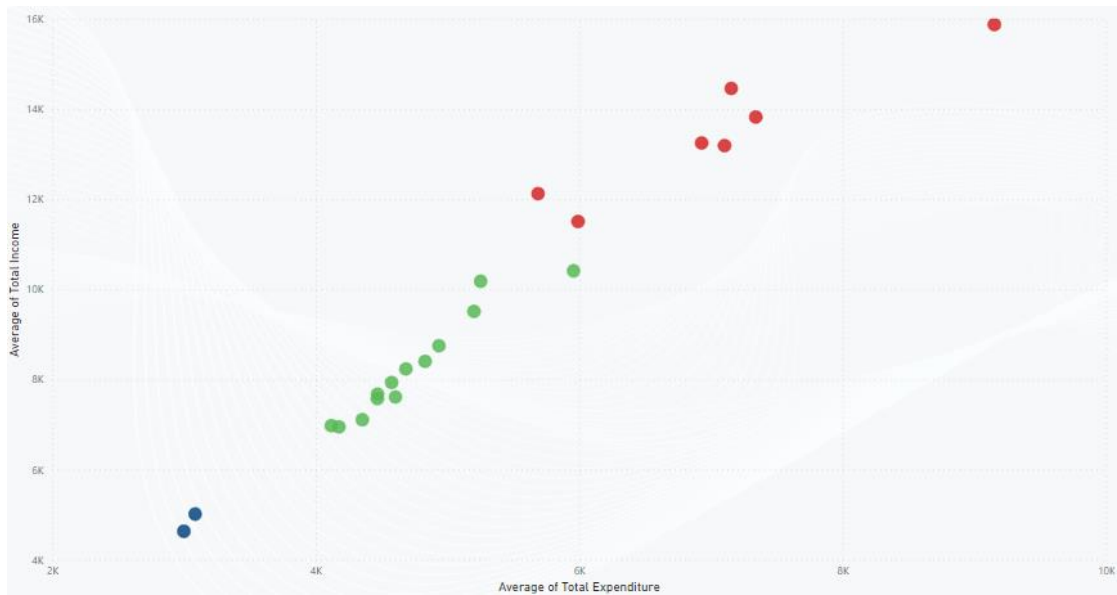

Figure 8b: The analysis of average of total monthly expense by industry in 2019

Based on Figure 8a and Figure 8b, there are three group have been analysed in this scatter plot clustering. The clustering analysis performed in this research is based on the industries in 2016 and 2019. This method will aid in the identification of similar data groups within a dataset. Group 1 has been labelled as low income and low expenses group. Group 2 has been listed as middle income and middle expenses group. The third group has been labelled as high income and high expense. In 2016 there are five industries has been listed in group 1. There are 9 industries has been listed in group 2. There are 8 industries has been in listed in group 3.

Group 1 consists of industries with incomes and expenses below RM 7,000 and RM 4,000, respectively. The income of group 2 ranges from RM 9000 to RM 7000, whilst its expenses range from RM 4000 to RM 6000. Group 3 comprises the industries with an income between RM 10,000 and RM 13,000 and expenditures between RM 5,000 and RM 7,000.

4. Tables

Table 1: Summary statistics of total monthly income of 2016 and 2019

Total Monthly Gross Income 2016		Total Monthly Gross Income 2019	
Mean	6680.07209	Mean	7690.401865
Standard Error	49.43711772	Standard Error	51.56435679
Median	5090	Median	5920.920085
Mode	1075	Mode	7544.166667
Standard Deviation	5963.477198	Standard Deviation	6594.192202
Sample Variance	35563060.29	Sample Variance	43483370.8
Kurtosis	66.19237014	Kurtosis	77.25914543
Skewness	5.615777335	Skewness	5.694563178
Range	147846.75	Range	173828.5685
Minimum	338.5	Minimum	318.2127075
Maximum	148185.25	Maximum	174146.7813
Sum	91032104.98	Sum	116348928.1
Count	14551	Count	16354

Table 2: Summary statistics of total monthly expenses of 2016 and 2019

Total Monthly Expenses 2016		Total Monthly Expenses 2019	
Mean	3720.603767	Mean	4220.829442
Standard Error	23.56196052	Standard Error	24.90101498
Median	3028.580078	Median	3730.210083
Mode	4166.990234	Mode	3000
Standard Deviation	2842.221003	Standard Deviation	3184.410492
Sample Variance	8078220.229	Sample Variance	10140470.18
Kurtosis	62.61360654	Kurtosis	51.52408411
Skewness	5.440261536	Skewness	4.822647204
Range	56203.0603	Range	72554.42093
Minimum	377.0100098	Minimum	460.2900085
Maximum	56580.07031	Maximum	73014.71094
Sum	54138505.42	Sum	69027444.7
Count	14551	Count	16354

In 2019 there are only two industries has been listed in group 1 compared to 2016. There are 13 industries has been listed in group 2. There are eight industries has been in listed in group 3. Group 1 is with mean incomes and mean expenses of less than RM 5,000 and RM 4,000, respectively. Group 2 has a range of RM 6000 to RM 10000 in mean income, and RM 4000 to RM 6000 in mean expenses. Group 3 includes industries with mean income between RM 11,000 and RM 16,000 and expenses between RM 5,000 and RM 9,000.

The analysis shows that the industries in group 1 has reduced from 5 to 2 industries. The lowest income and expenses in both year is agriculture, forestry and fishing industry which shows an increase by 13.61 percent and 15.34 percent in both income and expenses respectively. The second group increase from 9 to 13 industries and the industry which has the highest mean expenses and mean income has changed from activities of extraterritorial to arts, entertainment, and recreation. The third group has faces reduction from 8 industries to 7 industries. The highest mean expenses and mean income for both year is activities of

households as employers which shows and increase by 26.82 percent and 33.11 percent in both income and expenses.

Table 3: Average of monthly expense in 2016 for group 1

Industry	Average of Total Expenses (RM)	Average of Total Income (RM)
Group 1: Low Income and Low Expenses		
Accommodation & Food Service	3777.64	6359.43
Administrative & Support Service	3465.17	5921.42
Agriculture, Forestry & Fishing	2599.21	4001.91
Construction	3909.26	6749.65
Homemakers etc	3812.56	6224.32

Table 4: Average of monthly expense in 2016 for group 2

Industry	Average of Total Expenses (RM)	Average of Total Income (RM)
Group 2: Middle Income and Middle Expenses		
Activities of Extraterritorial	5242.04	9099.79
Electric Gas Steam & Air Conditioning	4167.68	7876.53
Human Health & Social Work Activities	4495.11	8225.84
Manufacturing	4119.89	7138.66
Other Service Activities	4498.35	7560.96
Public Administration & Defence	4314.62	7853.64
Transportation & Storage	4102.13	7054.25
Water Supply; Sewerage, Waste	4087.04	7662.49
Wholesale & Retail Trade	4126.47	7308.31

Table 5: Average of monthly expense in 2016 for group 3

Industry	Average of Total Expenses (RM)	Average of Total Income (RM)
Group 3: High Income and High Expense		
Activities of Households as Employers;	6874.55	12525.19
Arts, Entertainment & Recreation	5457.56	10051.03
Education	5146.55	10491.80
Financial & Insurance/Takaful Activities	5772.63	11489.25
Information & Communication	6221.53	11676.08
Mining & Quarrying	5346.89	11176.63
Professional, Scientific & Technical	6218.97	11828.74
Real Estate Activities	6623.83	12110.06

The mean income and expenses of group 1 has been decreased from 2016 to 2019. This is because the number of people working in group 1 industries has moved to other industries. This can be clearly seen from the number of people working in each industry. At the same time the mean income and expenses of group 2 and group 3 has been increased. From this can find out that the mean and median income group in Malaysia has been increased from 2016 to 2019. Most of the Malaysians preferred to be work in activities of households as employers'

industry since the mean income has been increased drastically. From this can understand that most of the Malaysian people preferred to work or do business from house which led the mean income of activities of household as employers to increase. Agriculture, forestry, and fishing might be the least option Malaysian people to choose as an industry to work in. This is because the mean income and mean expenses decreases from 2016 to 2019.

Table 6: Average of monthly expense in 2019 for group 1

Industry	Average of Total Expenditure (RM)	Average of Total Income (RM)
Group 1: Low Income and Low Expenses		
Agriculture, forestry, and fishing	2998.17	4632.62
Organizations and entities outside the territory	3084.06	5015.15

Table 7: Average of monthly expense in 2019 for group 2

Industry	Average of Total Expenditure (RM)	Average of Total Income (RM)
Group 2: Middle Income and Middle Expenses		
Accommodation and food services	4468.99	7677.60
Administrative and support activities	4117.67	6978.48
Arts, entertainment, and recreation	5956.47	10415.08
Construction	4467.29	7577.383
Electricity, gas, steam, and air conditioning	5251.33	10184.37
Human health and social work activities	5200.64	9515.379
Industries not elsewhere classified	4353.27	7109.79
Manufacturing	4829.70	8406.20
Other service activities	4604.47	7615.64
Public administration and defence	4934.68	8751.57
Transportation and storage	4575.43	7936.81
Water supply, sewerage, waste management	4175.36	6950.75
Wholesale and retail trade	4683.93	8237.47

Table 8: Average of monthly expense in 2019 for group 3

Industry	Average of Total Expenditure (RM)	Average of Total Income (RM)
Group 3: High Income and High Expense		
Activities of household as employers	9150.98	15885.31
Education	5990.74	11511.64
Financial and insurance/takaful activities	7339.96	13830.72
Information and communication	6929.02	13254.47
Mining and quarrying	5687.04	12129.47
Professional, scientific, and technical activities	7102.90	13196.05
Real estate activities	7153.86	14465.62

5. Conclusion

From the dashboard the datasets of the household expenditure survey in Malaysia have been visualized by using diagnostic analysis in Power BI tool. The data have been categorized using the decomposition tree method, and expenditures have been categorized based on

ethnicity. This method examined the 2016 and 2019 expenditures and earnings of Malaysians by industry. The scatter plot clustering technique primarily divides Malaysian income and expenses by sector. There are 22 distinct industries, each with unique expenses and income levels. The initiatives have been clustered based on their income and expenditures. There are many factors that can be affect the income and expenses of Malaysian people. From this analysis can be concluded that the income and expenses in Malaysia have been increased from 2016 to 2019. According to the ethnicity, Chinese concur the expenses in 2016 and 2019. According to industry, activities of household as employers has highest expenses and according to occupation, managers are with high income and expenses level. Based on education level, Malaysian people who studied until degree and above has highest mean income and expenses level.

The limitation of the study is the dataset is a census dataset, which further complicated the matter by being large data. As a result of the size of the data, it will take longer for every software to process. To obtain the exact result, no data can be lost during the transformation process. As a recommendation, in future study researchers may come up with variety of different analysis to analyses the data. This is because in this study there is only 2 analyses has been carried out.

Acknowledgments

References

Ahmad Z. & Fatima A. 2011. Prediction of household expenditure on the basis of household characteristics. In *Proc. ICCS-11*, pp. 351–367.

Boobier T. 2018. *Advanced Analytics and AI: Impact, Implementation, and the Future of Work*. 1st Ed. United Kingdom: John Wiley & Sons, Ltd.

Cingano F. 2014. Trends in income inequality and its impact on economic growth. *OECD Social, Employment, and Migration Working Papers*, No. 163. Paris: OECD Publishing.

Dogan O., Kaya G., Kaya A. & Beyhan H. 2019. Catastrophic household expenditure for healthcare in Turkey: Clustering analysis of categorical data. *Data* **4**(3): 112.

Etaati L. 2017. *Advance Analytics with Power BI and R*. Auckland: *RADACAD Systems Limited*.

Fan J.X., Brown B.B., Kowaleski-Jones L., Smith K.R. & Zick C.D. 2007. Household food expenditure patterns: A cluster analysis. *Monthly Labor Review* **130**(4): 38–51.

Gibson P. 2021. Types of data analysis. https://chartio.com/learn/data-analytics/types-of-data-analysis/ (9 July 2022).

Gupta S. What is predictive analytics? Learn 10 essential predictive analytics techniques. https://www.springboard.com/blog/data-analytics/predictive-analytics-techniques/ (7 August 2020).

Holliday M. What is diagnostic analytics? How it works and examples. https://www.netsuite.com/portal/resource/articles/data-warehouse/diagnostic-analytics.shtml (9 December 2021).

Johnson R.A. & Wichern D.W. 2007. *Applied Multivariate Statistical Analysis*. 6th Ed. Upper Saddle River, New Jersey: Pearson Prentice Hall.

Kemp S.E., Hort J. & Hollowood T. 2018. *Descriptive Analysis in Sensory Evaluation*. 1st Ed. New York: John Wiley & Sons, Ltd.

Koc P. & Yalcin C. 2021. Future of deep learning for cancer diagnosis. In Kose U. & Alzubi J. (eds.). *Deep Learning for Cancer Diagnosis*: 227-238. Singapore: Springer.

Nassaji H. 2015. Qualitative and descriptive research: Data type versus data analysis. *Language Teaching Research* **19**(2): 129–132.

Niu T., Chen Y. & Yuan Y. 2020. Measuring urban poverty using multi-source data and a random forest algorithm: A case study in Guangzhou. *Sustainable Cities and Society* **54**: 102014.

Okori W. & Obua J. 2011. Machine learning classification technique for famine prediction. In *Proceedings of the*

World Congress on Engineering 2011 (WCE 2011), pp. 991–996.

Punt C. 2003. Household expenditure patterns in South Africa – 1995. *PROVIDE Project Background Paper 2003: 2*. Elsenburg: PROVIDE Project.

Rawat A.S. An overview of descriptive analysis. https://www.analyticssteps.com/blogs/overview-descriptive-analysis (31 March 2021).

Luque J.P., Young H. & Miller C. 2017. *The Subprime Crisis: Lessons for Business Students*. Singapore: World Scientific Publishing.

Zhang T. & Yang B. 2017. Box–cox transformation in big data. *Technometrics* **59**(2): 189–201.

Department of Mathematic and Statistic
Faculty of Applied Science and Technology
Universiti Tun Hussein Onn Malaysia (UTHM)
84600 Hub Pendidikan Pagoh
Johor, MALAYSIA
E-mail: serahlathan7@gmail.com, asyikinr@uthm.edu.my[*]

Department of Technology and Management
Faculty of Technology Management and Business
Universiti Tun Hussein Onn Malaysia (UTHM)
86400 Batu Pahat
Johor, MALAYSIA
E-mail: ammar@uthm.edu.my

School of Quantitative Sciences
Universiti Utara Malaysia (UUM)
06010 Sintok
Kedah, MALAYSIA
E-mail: shairy@uum.edu.my

[*]Corresponding author