

L4 - Data Analysis and Visualisation ACCB4002

by Tyler Williamson

Submission date: 09-Jan-2024 05:41PM (UTC+0000)

Submission ID: 221229664

File name: 45148_Tyler_Williamson_L4_-_Data_Analysis_and_Visualisation_ACCB4002_547400_1040007248.docx (206.58K)

Word count: 2541

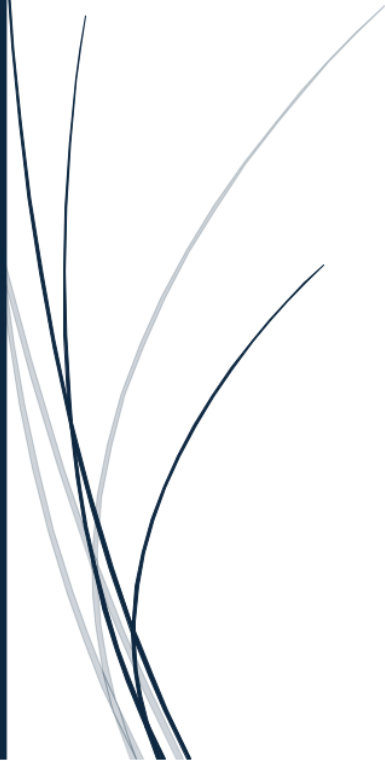
Character count: 14783

A thick dark blue vertical bar runs down the left side of the page. A blue arrow points to the right from this bar, containing the date.

05/01/2024

L4 – Data Analysis and Visualisation (ACCB4002)

Statistical Analysis of Retail Store Performance

Several thin, curved lines in shades of blue and grey originate from the bottom of the vertical bar and sweep upwards and to the right.

Tyler Williamson

Student No. 2237644

Table of Contents

1. Abstract.....	2
2. Introduction	2
3. Descriptive Statistics.....	2
3.1. Q1	2
3.1.1. A.....	2
3.1.2. B.....	3
3.1.3. C.....	4
3.2. Q2	5
3.2.1. A.....	5
3.2.2. B.....	5
3.2.3. C.....	5
4. Inferential Statistics.....	6
4.1. Q3	6
4.1.1. A.....	6
4.1.2. B.....	6
4.1.3. C.....	6
4.2. Q4	7
4.2.1. A.....	7
4.2.2. B.....	7
4.2.3. C.....	8
4.3. Q5	8
4.3.1. A.....	8
4.3.2. B.....	8
5. Summary	9
6. References	9
Figure 1 - Q1A Mean, Median and Mode.....	3
Figure 2 - Q1B Range, Variance and Standard Deviation	4
Figure 3 - Q1C Calculation of Outliers	4
Figure 4 - 75th Percentile Calculations	5
Figure 5 - Product category performing better than the 75th Percentile	5
Figure 6 - Correlation Coefficient.....	6
Figure 7 - Two-Tailed Distribution T-Test.....	7
Figure 8 - Mean and Standard Deviation of Rural and Urban Stores	7
Figure 9 - T-Statistic Level	7
Figure 10 - Stacked Bar Chart for Product Category Count Per Location	8
Figure 11 - Scatter Plot Chart	9

1. Abstract

This study investigates the dataset for a company's stores and the impact of various variables on the company's overall performance and efficiency, exploring effective ways for improvement. Utilising statistical analysis methods, including mean, median, ranges, correlation coefficient, 75th percentile, standard deviations, variance, interquartile, t-tests, t-statistics levels, degrees of freedom, p-values, and hypotheses. The investigation draws the conclusion that urban stores outperformed rural stores. The findings suggest potential improvements that could be achieved by influencing decisions to invest further in urban stores and making adjustments to rural stores, such as inventory and staffing levels. Furthermore, targeted marketing efforts may attract more clients to rural stores, subsequently improving the revenue generated by these stores.

2. Introduction

This report delves into the statistical analysis of a company's store dataset, aiming to uncover crucial insights that inform strategic decisions. The focus is on optimizing store performance for overall company improvements. This investigation employs a comprehensive set of statistical tools, including mean, median, mode, correlation coefficients, 75th percentiles, and t-tests etcetera. Visualisations, such as stacked bar charts and scatter plots, present the data in an accessible manner. The significance of this research is to provide quality recommendations for enhancing the efficiency of all stores. The aim is to guide the company's strategic decisions positively, ensuring further growth.

3. Descriptive Statistics

3.1. Q1

3.1.1. A

Identifying the central tendency of the sales distribution is reflected in the calculations for mean and median [1]. As shown in Figure 1 below, the mean and median for each store are closely

related and show similarities in comparable figures, indicating that the sales distribution is quite equal across all stores [2]. The store with the most differentiated figure is Store 4, which shares the product category Electronics with Store 1. The difference between these two stores is that one is in a rural area, and the other is in an urban location, which could impact the sales distribution of daily sales, resulting in a slightly larger difference between the mean and median than the rest of the stores.

Q1A	Store	Mean
	1	19.7143
	2	15
	3	13
	4	13.1429
	5	23.8333
	Store	Median
	1	20
	2	15
	3	13
	4	12
	5	23.5
	Store	Mode
	1	20
	2	#N/A
	3	#N/A
	4	#N/A
	5	#N/A

Figure 1 - Q1A Mean, Median and Mode

3.1.2. B

The variance figures of this dataset as illustrated within Figure 2, is also represented by the standard deviation figures, which is just the square root of the variance calculation. A higher variance or standard deviation figure indicates that the stores sales are more relative to the calculated mean [3]. The lower the variance or standard deviation figure is shows that the variability of the data is more restricted and in a closer range. Store 3 is the identified as having the lowest variance of this dataset and means that the stores performance is stable because the variance is similar to the calculated mean. The range highlights the similarity of performance between stores 2 and 3 where the daily sales are close in terms of variability and has a similarity to the value of the calculated mean meaning that there isn't much variance unlike store 4, which has the most variability by having a higher figure in the variance calculations and shows as being more distant in relation to the mean figure. This means that we can see stores 2 and 3 as having quite steady income in daily sales in comparison to other stores where daily sales may have differences in figures and potentially higher and lower figures quite often through the sales period.

Q1B	Store	MIN	MAX	Range
	1	12	28	16
	2	10	20	10
	3	8	18	10
	4	5	22	17
	5	18	30	12
	Store	Variance		
	1	30.2381		
	2	17		
	3	14		
	4	38.8095		
	5	21.7667		
	Store	Standard Deviation		
	1	5.49892		
	2	4.12311		
	3	3.74166		
	4	6.22973		
	5	4.66548		



Figure 2 - Q1B Range, Variance and Standard Deviation

3.1.3. C

Investigation of outliers in this dataset involves calculating interquartile values for each store and identifying the lower and upper boundaries of their respective figures [4]. This allows for the analysis of any potential outliers for each store. As demonstrated in Figure 3 and using the original dataset, analysis shows there are no potential outliers within this dataset as the data for the daily sales falls inside boundaries established calculating the boundaries using the interquartile, first quartile and third quartile values. The impact of outliers on analysis can significantly influence business decisions and can provide insights into the strengths and weaknesses of the dataset and how the business performs in relation to other stores. Identifying outliers can be crucial in making improvements to the core structure of the business and its operations. Evaluation of the stores and the differentiated data between them can tell a story on its own accord. For example, this data set can show the differences between rural and urban located stores in the same product category and that this possibly may have influence on the overall daily sales. Such evaluation could lead to a business owner making informed decisions such as scaling back operations in the rural stores and investing more into the urban stores or give consideration to adjusting the way the store is operating in the rural location versus the urban store, which is identified as having a more stable performance.

Q1C				
Identify any outliers using: IQR (Interquartile calculations)				
	Store	Quartile 1	Quartile 3 (Q3)	Interquartile (IQR)
	1	16.5	22.5	6
	2	12	18	6
	3	10.5	15.5	5
	4	9	17.5	8.5
	5	20.5	27.25	6.75
Identifying outliers				
	Store	IQR	Lower Bound	Upper Bound
	1	6	7.5	31.5
	2	6	3	27
	3	5	3	23
	4	8.5	-3.75	30.25
	5	6.75	10.375	37.375



Figure 3 - Q1C Calculation of Outliers

3.2. Q2

3.2.1. A

Calculating 75th Percentile figures for a dataset, while analysing it can assist in assessing store performances by locating the stores with higher daily sales values [5]. For instance, as shown in Figure 4, the 75th percentile for Store 5 is the highest, while Store 3 is the lowest. This insight provides an understanding of how stores are performing in terms of daily sales amounts. It enables the business to clearly identify areas where sales are stronger or weaker and determine which sales contribute the most to the overall store performance in daily sales.

Q2A	Store	75th Percentile
	1	22.5
	2	18
	3	15.5
	4	17.5
	5	27.25

Figure 4 - 75th Percentile Calculations

3.2.2. B

Figure 5 illustrates the stores, product categories, and the 75th percentile for each. Through the calculation of the 75th Percentile of daily sales and analysis of the original dataset, the findings revealed that Stores 1, 2 and 5 have outperformed the 75th Percentile in their respective product categories.

Q2B	Store	Product Category	75th Percentile	
	1	Electronic	22.5	ANSWER
	2	Apparel	18	Store 1
	3	Home Goods	15.5	Store 2
	4	Electronic	17.5	Store 5
	5	Home Goods	27.25	

Figure 5 - Product category performing better than the 75th Percentile

3.2.3. C

Findings from the analysis of the dataset and the 75th Percentile of each store for each product category can contribute to the strategic evaluation of the overall company sales. By strategically reviewing the product categories that are in highest demand and identifying the stores that perform the best in sales for these items, the analysis can influence the company's overall sales strategy. These insights assist in guidance as to how and where the business could make adjustments to potentially enhance performance and, consequently, generate more revenue for the company.

4. Inferential Statistics

4.1. Q3

4.1.1. A

As demonstrated in Figure 6, the correlation coefficient results representing the relationship between daily sales and average customer ratings for each store show there is no correlation between the two figures. The correlation for Stores 1 and 5 reveals no variability, leading to calculations attempting to divide by zero and resulting in the output #DIV/0!. Store 2, on the other hand, displays a positive linear correlation between daily sales and average customer ratings, while Store 3 shows no correlated relationship between these two values. The correlation calculation for Store 4 indicates a very weak or insignificant relationship between the two values, in essence nearing zero in result.

Q3A	Store	Correlation Coefficient
	1	#DIV/0!
	2	0.6
	3	0
	4	-8.79967E-17
	5	#DIV/0!

Figure 6 - Correlation Coefficient

4.1.2. B

As previously mentioned, the relationship between the stores is reflected in the correlation coefficient calculations [6]. Store 1 and 5 exhibits no apparent relationship between daily sales and average customer ratings. Store 2 demonstrates a positive linear relationship between the two variables. Store 3 is identified as having no correlated relationship for daily sales and average customer ratings. Store 4 shows a very weak correlation, which could be considered insignificant in determining a relationship between the variables. Store 2 reflects a positive relationship where customer satisfaction increases daily sales, therefore generating more income. In contrast, other stores that present no or weak correlated relationships do not show a significant influence of customer satisfaction on daily sales. This implies that this statistical analysis method may not be the most effective for identifying areas where the company could make improvements to its operations.

4.1.3. C

Providing recommendations to the company based on the findings from the correlation scores proves challenging due to the absence of negative and positive linear relationships that could be used to assess the performance of different stores against one another. However, upon analysis, Store 2 is the only store selling the product category of apparel. Therefore, the positive linear relationship gives an assumption that the increasing demands, coupled with the absence of other stores selling this product is influential to this positive correlation. The impact of the other stores displaying weak or no correlated relationships may be triggered by two out of the four stores each selling the same product (two for electronics and two for home goods). These four stores have varying locations, customer ratings, and daily sale averages in comparison to one another, resulting in a weak or no correlation relationship. This complexity

makes it difficult to provide logical and meaningful recommendations on how best to improve the company using this particular methodology of calculations.

4.2. Q4

4.2.1. A

As illustrated in Figure 7, the Two-Tailed Distribution T-Test is formulated to identify which hypothesis the daily sales performed such as Null Hypothesis and Alternative Hypothesis [7]. Highlighting the stores sales that are the best in either rural or urban located stores with no expectation of which direction this will go in the first instance. The calculation used to identify these hypotheses is the array of the daily sales for each group of stores in either rural or the urban.

T-Test	Values
Paired	#N/A
Equal Variance	0.002235315
Un-Equal Variance	0.001875848

Figure 7 - Two-Tailed Distribution T-Test



4.2.2. B

The findings from the T-Test conducted, as represented in Figure 7, strongly supports the evidence for an alternative hypothesis against the null hypothesis [8]. The reasoning behind this decision is reflected in the low p-value, indicating a significant difference in daily sales between urban and rurally located stores. The t-statistic level, as displayed in Figure 9, further supports the evidence against the null hypothesis and is much higher than the p-value mentioned in the previous paragraph. This demonstrates a significant difference between the two groups. The analysis of the dataset for the t-statistic level involved investigating the probability and degree of freedoms sample size of 31 across the two groups. This calculation is performed using Excel.

Location	Mean	Standard Deviation
Urban	19.77777778	5.63279638
Rural	13.07692308	4.827191228

Figure 8 - Mean and Standard Deviation of Rural and Urban Stores

T-Statistic	Values
Paired	#N/A
Equal Variance	3.332808883
Un-Equal Variance	3.399067828

Figure 9 - T-Statistic Levels

4.2.3. C

The investigation on this data set using the T-Test, T-Statistic, degrees of freedom, the p-value, and the hypotheses strongly suggest that urban stores are performing significantly better than the rurally located stores. The results provide solid evidence that the location of the stores has a substantial impact on d  10 les. Based on this insight the company can make varying conscientious decisions on their strategic actions. The business could capture the opportunity to invest further into the successfully performing urban stores. Additionally, the company could optimize inventory management by lowering the levels in rural stores and increasing them in urban stores, which may enhance the overall efficiency of the higher performing stores. Staff relocation to urban stores could be considered to further improve the stores operations. Furthermore, targeted marketing for rural stores could be implemented to encourage clients to purchase more products and, consequently, generate more sales. The investigation results have the potential to positively influence the company's performance and daily sales by allowing the company to make strategic decisions effectively.

4.3. Q5

4.3.1. A

As illustrated in Figure 10, the stacked bar chart represents the data visualisation of the Product categories distributed across both locations [9]. The visualisation shows that the variance of product categories between urban and rural stores is nearly equal for two of the product categories. However, urban stores have an additional product category, namely Apparel. Consequently, rural stores have roughly two-thirds of the product categories available in urban stores.

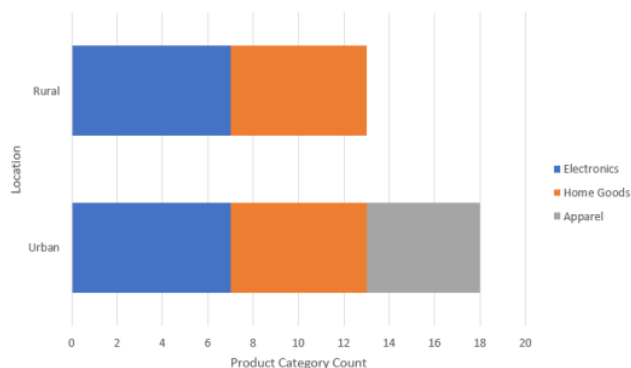


Figure 10 - Stacked Bar Chart for Product Category Count Per Location

4.3.2. B

The Scatter Plot chart as displayed in Figure 11, shows that the average customer ratings across all product categories falls within the range of 3.5 and 4.5 ratings [10]. Additionally, the

chart illustrates a diverse range of daily sales amounts, spanning approximately 15 from the lowest to the highest daily sales figure across all three product categories.

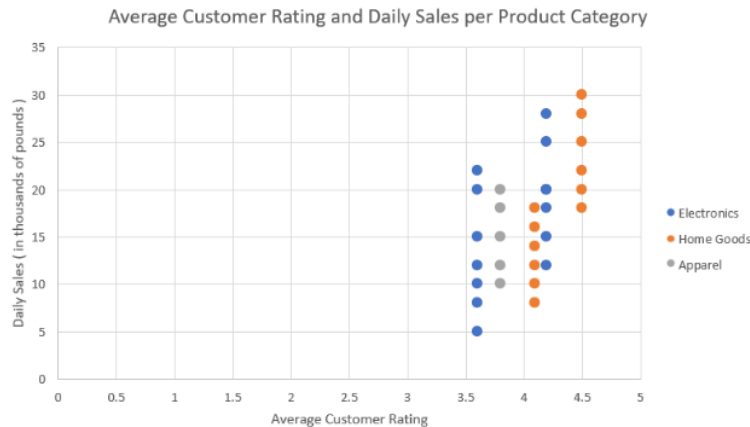


Figure 11 - Scatter Plot Chart

5. Summary

In conclusion, this report and its investigations have provided the company with recommendations for strategic decisions to improve the store's performance efficiency based on insights obtained from the analysis of daily sales, average customer ratings and locations. This investigation has been achieved through calculations such as mean, median, mode, standard deviations, variance, ranges, and the 75th percentile. Additionally, the analysis incorporated investigations into correlation relationships using correlation coefficient calculations in Excel, exploring the linear relationships between the correlations. Furthermore, interquartile calculations have been implemented to extract information related to potential outliers in the dataset by analysing the lower and upper boundaries and identifying figures that fall outside of these boundaries. This report also explored sales distributions using t-tests, t-statistics levels, degrees of freedom, p-values, and hypotheses. In addition, it included the visual representation of some data through stacked bar and scatter plot charts.

6. References

- [1] J. H. Kranzler, *Statistics for the terrified, sixth edition*, 6th ed. Ed. Rowman & Littlefield Publishers, 2017. [Online]. Available: <http://portal.igpublish.com/iglibrary/search/ROWMANB0017370.html>
- [2] J. H. Kranzler, *Statistics for the terrified, sixth edition*, 6th ed. Ed. Rowman & Littlefield Publishers, 2017. [Online]. Available: <http://portal.igpublish.com/iglibrary/search/ROWMANB0017370.html>

- [3] A. Hayes. "What Is Variance in Statistics? Definition, Formula, and Example." <https://www.investopedia.com/>. <https://www.investopedia.com/terms/v/variance.asp> (accessed Dec 21,).
- [4] W3Schools. "Statistics - Interquartile Range." <https://www.w3schools.com/>. https://www.w3schools.com/statistics/statistics_interquartile_range.php (accessed Dec 24,).
- [5] W3Schools. "Statistics - Quartiles and Percentiles." <https://www.w3schools.com/>. https://www.w3schools.com/statistics/statistics_quartiles_and_percentiles.php (accessed Dec 21,).
- [6] OpenStax. "12.5: Testing the Significance of the Correlation Coefficient." <https://stats.libretexts.org/>. [https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Introductory_Statistics_\(OpenStax\)/12%3A_Linear_Regression_and_Correlation/12.05%3A_Testing_the_Significance_of_the_Correlation_Coefficient](https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Introductory_Statistics_(OpenStax)/12%3A_Linear_Regression_and_Correlation/12.05%3A_Testing_the_Significance_of_the_Correlation_Coefficient) (accessed Dec 28,).
- [7] J. H. Krantzler, *Statistics for the terrified, sixth edition*, 6th ed. Ed. Rowman & Littlefield Publishers, 2017.
- [8] A. Hayes. "T-Test: What It Is With Multiple Formulas and When To Use Them." <https://www.investopedia.com/>. <https://www.investopedia.com/terms/t/t-test.asp> (accessed Dec 28,).
- [9] Microsoft. "Create a chart from start to finish." <https://support.microsoft.com/>. <https://support.microsoft.com/en-us/office/create-a-chart-from-start-to-finish-0baf399e-dd61-4e18-8a73-b3fd5d5680c2> (accessed Jan 3,).
- [10] Microsoft. "Present your data in a scatter chart or a line chart." <https://support.microsoft.com/>. <https://support.microsoft.com/en-us/topic/present-your-data-in-a-scatter-chart-or-a-line-chart-4570a80f-599a-4d6b-a155-104a9018b86e> (accessed Jan 3,).

L4 - Data Analysis and Visualisation ACCB4002

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

54/100

Question 1	15
Question 2	16
Question 3	8
Question 4	4
Question 5	6
Writing Skills	5

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5



Comment 1

well done



Comment 2

Outliers are datasets that are not in this quartile

PAGE 6



Comment 3

well analysed



Comment 4

More specific details could have been included



Comment 5

few strategies like

'Capitalize on the popularity of high-selling products.' or 'promoting and expanding the product range in the upper price tier.' could be included

PAGE 7



Comment 6

DIV/0 is an error in values that are input .

We get a positive sign of the correlation coefficient (0.56)and it's strength is moderate.



Comment 7

Continuous monitoring, adaptation, and responsiveness to customer feedback will be key to sustaining and enhancing both customer satisfaction and sales performance.

PAGE 8



Comment 8

but we focus on one tailed test



Comment 9

only one will be choosen

PAGE 9



Comment 10

Tailor sales and marketing strategies based on the specific characteristics and needs of each location.

