

Assignment 2

by Kelsey Priestley

Submission date: 12-Jan-2024 12:28AM (UTC+0000)

Submission ID: 221653690

File name: 50356_Kelsey_Priestley_Assignment_2_547400_1282436061.docx (66.89K)

Word count: 734

Character count: 3518

ASSIGNMENT 2

Kelsey Priestley

UWTSD Data Analysis and Visualisation

Data Set

Store	Product Category	Daily Sales (in thousands of pounds)	Average Customer Rating (on a scale of 1 to 5)	Location
1	Electronics	12.00	4.2	Urban
1	Electronics	15.00	4.2	Urban
1	Electronics	18.00	4.2	Urban
1	Electronics	20.00	4.2	Urban
1	Electronics	20.00	4.2	Urban
1	Electronics	25.00	4.2	Urban
1	Electronics	28.00	4.2	Urban
2	Apparel	10.00	3.8	Urban
2	Apparel	12.00	3.8	Urban
2	Apparel	15.00	3.8	Urban
2	Apparel	18.00	3.8	Urban
2	Apparel	20.00	3.8	Urban
3	Home Goods	8.00	4.1	Rural
3	Home Goods	10.00	4.1	Rural
3	Home Goods	12.00	4.1	Rural
3	Home Goods	14.00	4.1	Rural
3	Home Goods	16.00	4.1	Rural
3	Home Goods	18.00	4.1	Rural
4	Electronics	5.00	3.6	Rural
4	Electronics	8.00	3.6	Rural
4	Electronics	10.00	3.6	Rural
4	Electronics	12.00	3.6	Rural
4	Electronics	15.00	3.6	Rural
4	Electronics	20.00	3.6	Rural
4	Electronics	22.00	3.6	Rural
5	Home Goods	18.00	4.5	Urban
5	Home Goods	20.00	4.5	Urban
5	Home Goods	22.00	4.5	Urban
5	Home Goods	25.00	4.5	Urban
5	Home Goods	28.00	4.5	Urban
5	Home Goods	30.00	4.5	Urban

This is the data set sorted by store in ascending order.

Section A

Mean, Mode, Median, Range, Variance and Standard Deviation

Calculating the mean, median and mode can offer a broad view and help analyse the distribution and central tendency of different products and stores. For example, the mean which is the average will provide an overall sales pattern. But a mean that is inconsistent overtime indicates irregularities. The mode the value that appears the most in a data set. The mode indicated the most common daily sale figure and can further help identify patterns that follow a certain trend. The mode is not as susceptible to outliers in a data set as the mean is. The median is the middle value of a data set. The median might provide a better value than the mode when it comes to sales in the extreme highs or lows.

The range, variance and standard deviation will help identify volatility, compare performance, and help predict future sales. The range is the highest value and lowest value taken away from each other. The highest value for store 1 was 28 and the lowest value was 12, this makes the range for store one 16. A larger range can show more variety in sales, while a smaller range shows consistency.

The variance is the average of the squared differences of the mean. Variance can provide how individual sales differentiate from the average sales. Standard deviation is the square root of the variance. Standard deviation is like variance in the sense that it highlights values around the mean.

Here is the mean, median and mode for the retail store data set.

Mean		Median		Mode	
Store 1	16.43	Store 1	20	Store 1	20
Store 2	15	Store 2	15	Store 2	No Mode
Store 3	13	Store 3	13	Store 3	No Mode
Store 4	13.14	Store 4	12	Store 4	No Mode
Store 5	23.83	Store 5	23.5	Store 5	No Mode

There is an outlier with the three methods, these being there is only a mode for one of the stores. Having no mode is not necessarily a negative toward sales of a store but it could infer a lack of central tendency.

There are no outliers in the mean calculations, but it could be argued that thirteen and fifteen are the outliers as they are the minority as the other numbers have additional decimal values. However, as the numbers are not drastically different, it should not skew any results that the retail stores want to gather.

All the median values have no discrepancies with the main outlier being 23.5. However, the difference is not major.

Here is the range, variance, and standard deviation for the retail store data set.

Variance		Standard Deviation		Range	
Store 1	25.92	Store 1	5.09	Store 1	16.00
Store 2	13.60	Store 2	3.69	Store 2	10.00
Store 3	11.67	Store 3	3.42	Store 3	10.00
Store 4	33.27	Store 4	5.77	Store 4	17.00
Store 5	18.14	Store 5	4.26	Store 5	12.00

There does not appear to be any discrepancies with the calculations provided.

75th Percentile

By using the 75th percentile, it will help stores determine if they are performing higher than other stores. It also serves as a goal or success rate for a store. Stores with sales that reach higher than the 75th percentile are performing above average and will help plan in their marketing. The 75th percentile was used on the 5 stores in the data sheet.

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

As seen above, certain stores hit higher than others when calculating the 75th percentile. Store 2 is significantly lower than store 5 with it being a 9.25 difference. For store 4, since 17.5 falls between the values of 15 and 20, Store 4 does reach the 75th percentile. This means that 75% of the daily sales values for Store 4 are below 17.5. Store 1 also reaches the 75th percentile as 75% of sales fall below 22.5.

Stores can use this information to identify if they reach certain thresholds that they set. The 75th percentile is a very helpful benchmark to reach as it shows how sales perform. For marketing it can show what specific area is performing well. For example, if electronics is performing well, they can focus on marketing in that area by creating adverts for an audience who buy electronics.

Section B

Correlation Coefficient

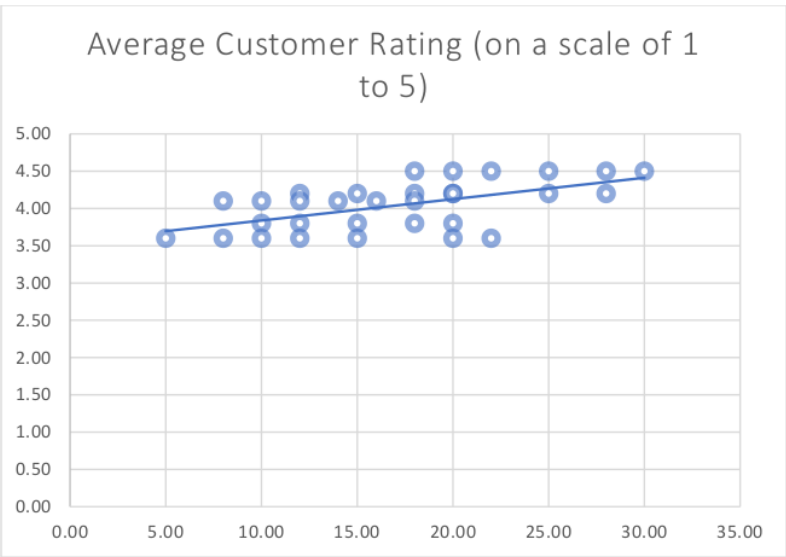
The correlation coefficient is value that measures the strength of a linear relationship between 2 different variables. For this dataset, the 2 values were the average customer ratings and the daily sales for each store.

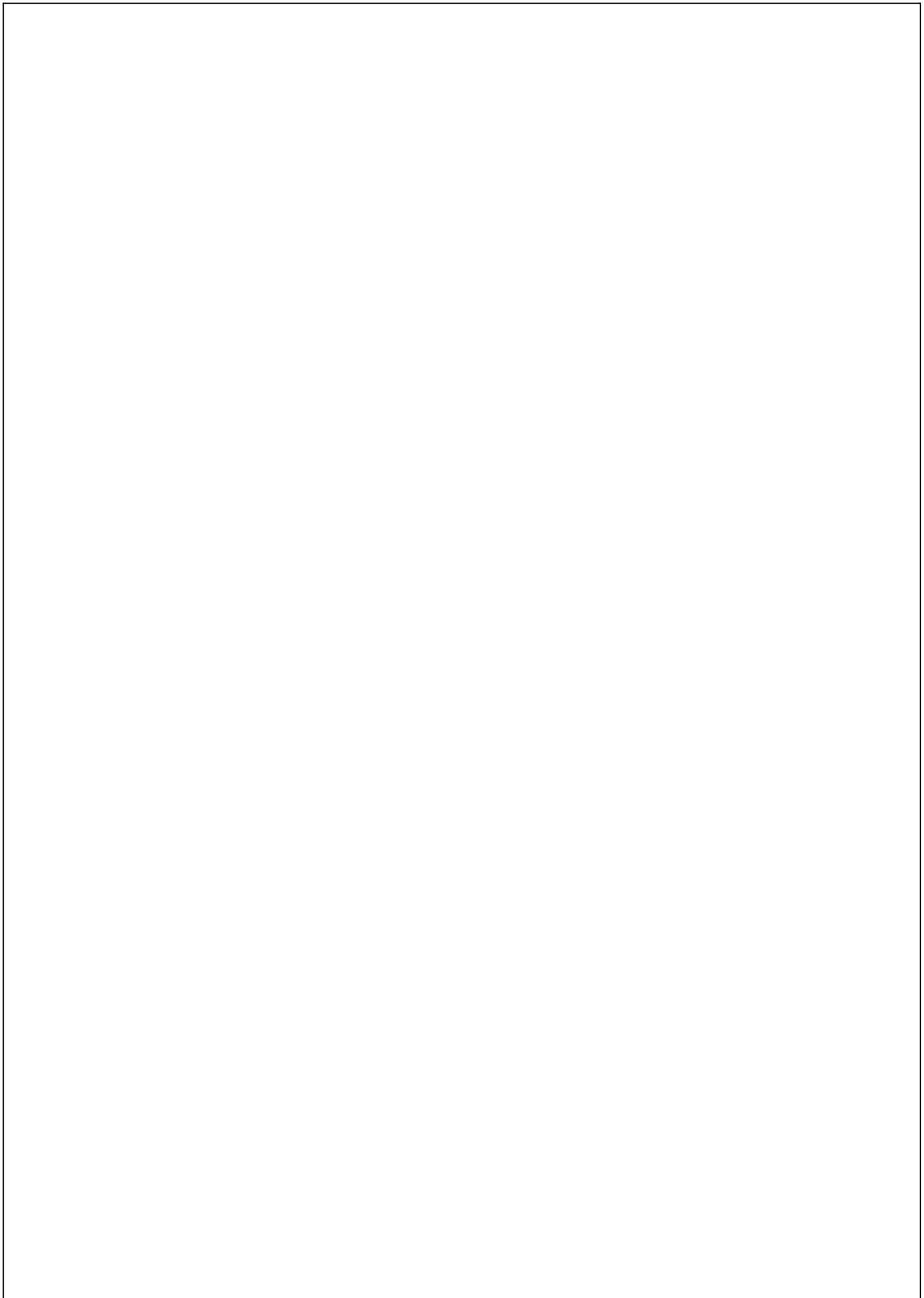
Correlation Coefficient		
Store 1	Positive	0.9894
Store 2	Positive	0.9999
Store 3	Positive	0.9999
Store 4	Perfect	1
Store 5	Positive	0.9945

Most of the correlation coefficient values were either strongly positive or perfect with very miniscule variations. Store 4 has a perfect relationship between the two values with it being a 1. All these relationships are strong linear relationships. As one increases so does the other.

Graphs

This is a scatter plot with the average customer rating on the Y-axis and the daily sales (in thousands) on the X-axis.





Assignment 2

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

43/100

Question 1	19
Question 2	15
Question 3	5
Question 4	0
Question 5	0
Writing Skills	4

PAGE 1

PAGE 2



Comment 1

clear numbering shpould have been done

PAGE 3



Comment 2

well represented



Comment 3

should be calculated for sample and not population dataset



Comment 4

well presented but how did you find the 75th percentile?

PAGE 4



Comment 5

you can include stratigies like 'Explore partnerships or collaborations to enhance product visibility'.

