

Title: Kwikie-Mart Sales Data Analysis

List of Contents:

- Executive Summary**
- Pivot Tables**
- Bar and Line Graphs**
- IF statements**
- VLOOKUP**
- WEKA**
- Clustering**
- Linear regression**
- Decision Trees**
- Association**
- Regression**

By Bayden Bromley

Business Management - Year 2

1800628

Trinity Saint Davids University

Executive Summary:

Our business has been approached by Kwik-e-mart office and supplies to analysis their data sales in order to determine weak areas that need attention. This will help their business grow and function more fluidly as well as making sales forecasting more transparent. With the information at hand our business can assist your business in putting in place mitigating strategies such as end of life promotion methods to ensure their business isn't supporting any products that are not profitable, like 'Dogs' in reference to the Boston Matrix. We want to provide insight and boost sales by using qualitative data to analyse the data and then identify areas where the profitability is low and work on ways to fix the problems through various techniques.

Pivot table:

This is the raw data that Kwikie-Mart sent our business. This information set is very hard to overview and understand, it was over 50 pages long and had all types of data. We want to narrow down Pivot table allows a graph to show.

[illegible]

The screenshot displays the Microsoft Excel interface. A PivotTable is set up in the range A3:M10, with 'Region' as the row label, 'Category' as the column label, and 'Total' as the values. The PivotTable data is as follows:

Region	Category	Total
Central	Funhouse	287,249.6
	Office Supplies	827,177.9
	Technology	1,067,143.2
Central Total		2,181,570.7
East	Funhouse	344,445.6
	Office Supplies	813,447.9
	Technology	4,362,110.4
East Total		5,520,003.9
North	Funhouse	877,128.4
	Office Supplies	1,948,390.9
	Technology	3,059,121.4
North Total		5,884,640.7
West	Funhouse	1,104,933.6
	Office Supplies	1,240,849
	Technology	4,963,447.6
West Total		7,309,230.2
Grand Total		20,977,444.5

On the right side, the 'PivotTable Fields' task pane is visible, showing a list of fields: Region, Category, Funhouse, Office Supplies, Technology, and Total. The 'Region' field is currently selected in the 'Rows' list.

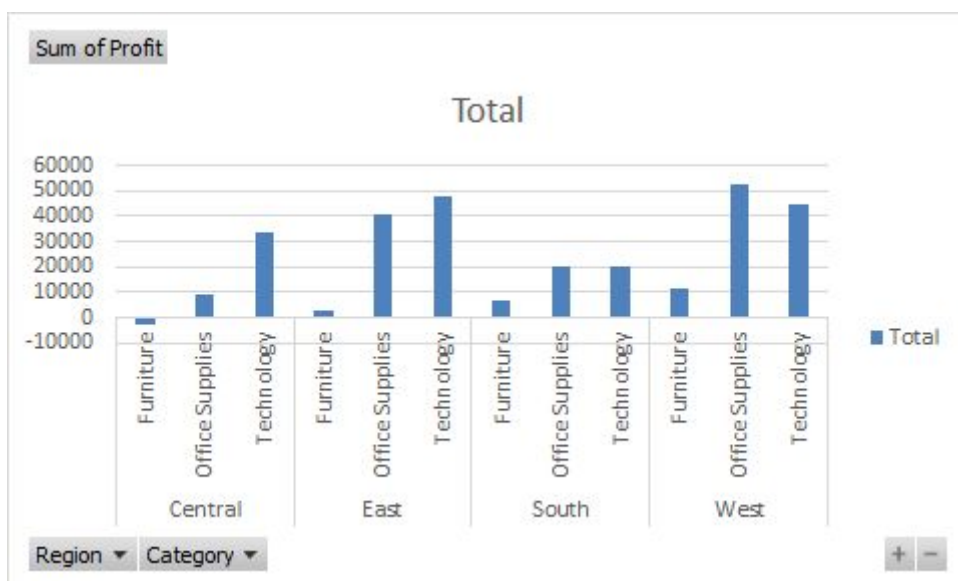
In this section of Excel you can pick the data that you want your pivot table to show. For this example I selected; region, profit and categories.

This allows the data to display these categories based off of the large data that we were given at the start by Kwikiee-mart office and supplies. Using this data you can see the businesses locations and which area is most profitable. For example this pivot table shows us that the West is the most profitable area, making '44303.64 off technology alone. This gives us the ability to see the profitability of the different store locations as well as the grand total at the bottom. We identified using this data a profitability problem with the business. The Central Region is making a loss of -2871.0494 on furniture. This can then guide the decisions of the management accordingly.

	A	B	C
1			
2			
3	Sum of Prof		
4	Region	Category	Total
5	Central	Furniture	-2871.0494
6		Office Supplies	8879.9799
7		Technology	33697.432
8	Central Total		39706.3625
9	East	Furniture	3046.1658
10		Office Supplies	41014.5791
11		Technology	47462.0351
12	East Total		91522.78
13	South	Furniture	6771.2061
14		Office Supplies	19986.3928
15		Technology	19991.8314
16	South Total		46749.4303
17	West	Furniture	11504.9503
18		Office Supplies	52609.849
19		Technology	44303.6496
20	West Total		108418.4489
21	Grand Total		286397.0217
22			

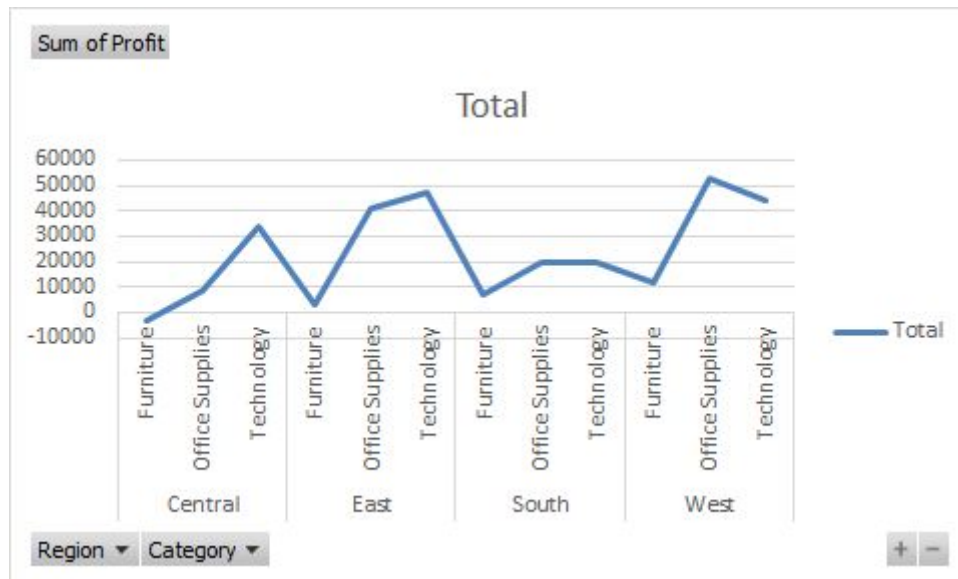
Adversely the west is looking very strong, winning in most categories for profitability, therefore the management should consider opening up another store in this region or expanding on the store they already have. They can also see that they may need to stop selling furniture or do some market research in the Central location as its pulling their profits down. However it's not only the Central region that is struggling for profits with furniture. The Easts profits are considerable lower in Furniture than the South and West. At only 3046.1658 in the East compared to 11504.9503 in the West and 6771.2061 in the South. The store in the West is head and shoulders above the rest of the locations, with an

overall profit doubling the overall profit in any other region. This is a very clear sign to management that their is a lot of demand in the West for their business. Opening a second store there would be a very smart move based on these numbers. The main benefit of this quantitative data is that it gives you insight into your business's financial stance and highlights weak areas that the business can then take decisions on. It's a very powerful tool for your business to utilise.



Excel has many functions, including visualizing data in a graphic form. This bar and line graph helps visualize the successful locations in

reference to their profitability. Technology and Office Supplies are clearly doing very well in the West as well as in the East. Furniture across the board is low in



profitability, so this could highlight to the business that this is not a very smart area to expand on, resulting in them potentially making the decision to stop selling furniture all together and

use that revenue to further invest in the areas that are succeeding. The East and West are the most profitable and successful locations, management should take note of this and visit the locations to see what they are doing that makes them succeed so well. This can then be mitigated in the other less successful locations. Or they may find that each store is running in a similar fashion, the location is what is making the West and East so successful. In this instance they could make the decision to open more stores in successful locations and may close the store in the South or Central as they are the sites with the weakest profitability. There are many courses of action the business can choose to take. We are just data analysts so we inform the business of what the data means and put it into a perspective that is easier to understand. The business then takes this information on board and makes their own executive decisions on the required course of action.

If Statement:

The IF statement performs the function of displaying information in a more organised format as well as to check if the condition is true. The IF statement is a simple bit of programming; =IF(X3U2>0, "profit", "loss"). If you enter this code and select the profit region you're data will display itself down the row. It shows the business which product is profitable or in loss.

The screenshot shows an Excel spreadsheet with the following data in columns T to Y:

Discount	Profit	W	X	Y
0	41.9136	profit		
0	219.582	profit		
0	6.8714	profit		
0.45	-383.05	loss		
0.2	2.5164	profit		
0	14.1694	profit		
0	1.9566	profit		
0.2	90.7152	profit		
0.2	5.7825	profit		
0	34.47	profit		
0.2	85.3092	profit		
0.2	68.3568	profit		
0.2	5.4432	profit		
0.2	132.5922	profit		
0.8	-123.858	loss		
0.8	-3.816	loss		
0	13.3176	profit		
0	9.99	profit		
0	2.4824	profit		
0.2	16.011	profit		
0.2	7.594	profit		
0	5.0996	profit		
0	15.6884	profit		
0.3	-1.0196	loss		
0	240.2649	profit		
0.2	4.2224	profit		
0	1.741	profit		
0.5	-1665.0522	loss		
0.7	-7.0532	loss		
0.2	15.525	profit		
0.2	1.1016	profit		
0.2	9.7092	profit		
0.7	5.215	loss		
0.2	3.546	profit		

Each product is labeled on one side and at the end the IF statement details which is a profitable product and which is in a loss. Assisting business managers in making decisions on their inventory.

Vlookup:

Vlookup is a method in excel that allows you to search for a value in the same row as the index number position. It acts like a search tool. For example, you have to

input a code into Excel. The code relates to how many data sets you want to include as well as where you want excel to display the data. For Kwiee-mart office and supplies the vlookup code is: =VLOOKUP(Z3,N1:U9995,8,FALSE).

N1 refers to the product ID.

U9995, refers to the last input at the bottom of this table. There are over 9000 columns so this just means that all the data sets have been selected.

The use of this is that it allows me to select any data set and see which products are in profit and which are making a loss. You can apply Vlookup to show a variety of different types of data. For example in this picture to the left, the data relates to

The screenshot shows an Excel spreadsheet with the following data in columns K to P:

State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name
Kentucky	42420	South	FUR-BO-1000	Furniture	Bookcases	Bush Somerset Co
Kentucky	42420	South	FUR-CH-1000	Furniture	Chairs	Hon Dekane Fabric
California	90036	West	OFF-LA-1000	Office Supplies	Labels	Self-Adhesive Add
Florida	33311	South	FUR-TA-1000	Furniture	Tables	Bretford CR4500 S
Florida	33311	South	OFF-ST-1000	Office Supplies	Storage	Eldon Fold 'N Roll
California	90032	West	FUR-FU-1000	Furniture	Furnishings	Eldon Expressions
California	90032	West	TEC-PH-1000	Technology	Phones	Mitel 5320 IP Pho
California	90032	West	OFF-BI-1000	Office Supplies	Binders	DXL Angle-View B
California	90032	West	OFF-AP-1000	Office Supplies	Appliances	Belkin F5C206VTE
California	90032	West	FUR-FU-1000	Furniture	Tables	Chromcraft Rectar
California	90032	West	TEC-PH-1000	Technology	Phones	Konftel 250 Corfe
North Carolin	28027	South	OFF-PA-1000	Office Supplies	Paper	Xerox 1067
Washington	98103	West	OFF-BI-1000	Office Supplies	Binders	Fellowes PR200 PI
Texas	76106	Central	OFF-AP-1000	Office Supplies	Appliances	Holmes Replacem
Texas	76106	Central	OFF-BI-1000	Office Supplies	Binders	Storex DuraTech R
Wisconsin	53711	Central	OFF-ST-1000	Office Supplies	Storage	Stur-D-Stor Shelvi
Utah	84084	West	OFF-ST-1000	Office Supplies	Storage	Fellowes Super St
California	94109	West	OFF-AR-1000	Office Supplies	Art	Newell 341
California	94109	West	TEC-PH-1000	Technology	Phones	Cisco SPA 5015 IP
California	94109	West	OFF-BI-1000	Office Supplies	Binders	Wilson Jones Han
Nebraska	68025	Central	OFF-AR-1000	Office Supplies	Art	Newell 318
Nebraska	68025	Central	OFF-AP-1000	Office Supplies	Appliances	Acco Six-Outlet P
Pennsylvania	19140	East	FUR-CH-1000	Furniture	Chairs	Global Deluxe Sta
Utah	84057	West	FUR-TA-1000	Furniture	Tables	Bretford CR4500 S
California	90049	West	OFF-BI-1000	Office Supplies	Binders	Wilson Jones Acti
California	90049	West	TEC-AC-1000	Technology	Accessories	Imation 8GB Mini
Pennsylvania	19140	East	FUR-BO-1000	Furniture	Bookcases	Riverside Palas R
Pennsylvania	19140	East	OFF-BI-1000	Office Supplies	Binders	Avery Recycled Fl
Pennsylvania	19140	East	FUR-FU-1000	Furniture	Furnishings	Howard Miller 13
Pennsylvania	19140	East	OFF-EN-1000	Office Supplies	Envelopes	Poly String Tie Em
Pennsylvania	19140	East	OFF-AR-1000	Office Supplies	Art	BOSTON Model 11
Pennsylvania	19140	East	OFF-BI-1000	Office Supplies	Binders	Acco Pressboard C
Pennsylvania	19140	East	OFF-AR-1000	Office Supplies	Art	Lumber Crayons

N4, so I get the data from that set. Whereas in the picture below I have selected to get data from the set N8. So you can select different data sets and find out the profitability of a single item that the store is selling. Giving insight into the benefit of the different products that the store sells as well as highlighting products that are making a loss.

Vlookup doesn't show as much data as some may want, it's also harder to use than other data providers on excel like pivot tables. However it is more niche as it can show the profitability of each product where as pivot tables show the profit overall. In the marketing mix the Vlookup helps with the product aspect a lot. It's a powerful tool to use, potentially saving the business from continuing to invest in non profitable products.

State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name
Kentucky	42420	South	FUR-BO-1000	Furniture	Bookcases	Bush Somerset Coll.
Kentucky	42420	South	FUR-CH-1000	Furniture	Chairs	Hon Deluxe Fabric U.
California	90036	West	OFF-LA-1000	Office Supplies	Labels	Self-Adhesive Adre
Florida	33311	South	FUR-TA-1000	Furniture	Tables	Bretford CR4500 Ser
Florida	33311	South	OFF-ST-1000	Office Supplies	Storage	Eldon Fold 'N Roll C

Sales	Quantity	Discount	Profit	product ID
261.96	2	0	41.9136	
731.94	3	0	219.582	
14.62	2	0	6.8714	
957.5775	5	0.45	-383.031	
22.368	2	0.2	2.5164	
48.86	7	0	14.1694	
7.28	4	0	1.9656	
907.152	6	0.2	90.7152	
18.504	3	0.2	5.7825	
114.9	5	0	34.47	
1706.184	9	0.2	85.3092	
911.424	4	0.2	68.3568	
15.552	3	0.2	5.4432	
407.976	3	0.2	132.5922	
68.81	5	0.8	-123.858	
2.544	3	0.8	-3.816	
665.88	6	0	13.3176	
55.5	2	0	9.99	
8.56	2	0	2.4824	
213.48	3	0.2	16.011	
22.72	4	0.2	7.384	
19.46	7	0	5.0596	
60.34	7	0	15.6884	
71.772	2	0.3	-1.0196	
1044.63	3	0	240.7649	
11.648	2	0.2	4.2224	
90.57	3	0	11.7741	
3083.43	7	0.5	-1665.0522	
9.618	2	0.7	-7.0532	
124.2	3	0.2	15.525	
3.264	2	0.2	1.1016	
86.304	6	0.2	9.7092	
6.858	6	0.7	-5.715	
15.76	2	0.2	3.546	

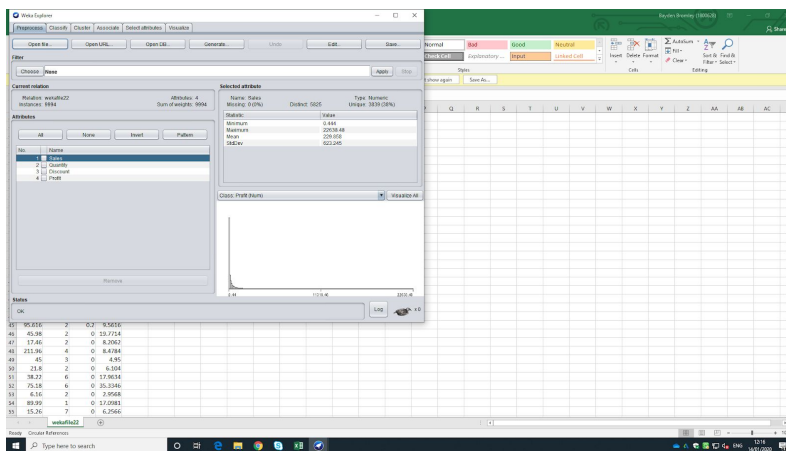
Z6 refers to where I want excel to put the data. N2 is the start of the data and U9995 is where the data ends. You then input 8, false

to finish the code. Excel will then put the Vlookup into your Z6 column that you selected at the start. Giving you the number that you were looking for based on all the datasets you selected between N2 and U9995.

WEKA:

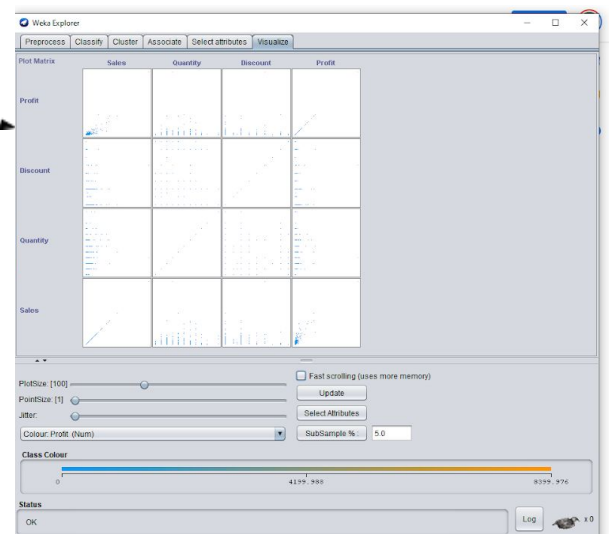
To input the data from Excel to WEKA you need to convert the file to a CSV file and also remove all /'s and commas in order for WEKA to be able to process the information easily.

Region	Product ID	Category	Sub-Category	Product Name
South	FUR-BO-1000	Furniture	Bookcases	Bush Somerset Collection Bookcase
South	FUR-CH-1000	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs Rounded Back
West	OFF-LA-1000	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters by Universal
South	FUR-TA-1000	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table
South	OFF-ST-1000	Office Supplies	Storage	Eldon Fold N Roll Cart System
West	FUR-FU-1000	Furniture	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood
West	OFF-AR-1000	Office Supplies	Art	Newell 322
West	TEC-PH-1000	Technology		
West	OFF-BI-1000	Office Supplies		
West	OFF-AP-1000	Office Supplies		
West	FUR-TA-1000	Furniture		
West	TEC-PH-1000	Technology		
South	OFF-PA-1000	Office Supplies		
West	OFF-BI-1000	Office Supplies		
Central	OFF-AP-1000	Office Supplies		
Central	OFF-BI-1000	Office Supplies		
Central	OFF-ST-1000	Office Supplies	Storage	Stur-D Stor Shelving, Vertical 5-Shelf: 72"H x 36"W x 18 1/2"D
West	OFF-ST-1000	Office Supplies	Storage	Fellowes Super StorDrawer
West	OFF-AR-1000	Office Supplies	Art	Newell 341

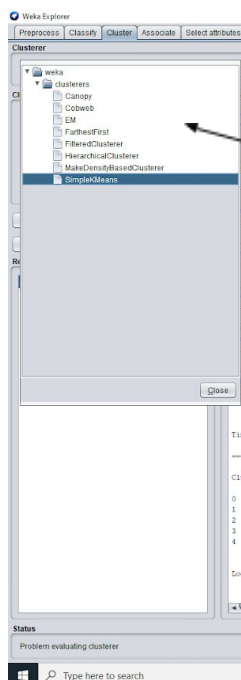


Once the data is in the correct format for WEKA to accept it, it will display this page. From here you can click on the subheadings to see the clustered data as well as the associate and classify data. There's also a section that allows you to visualize the profit, discount, quantity and sales for

the input you selected. The disadvantage of this is that the data isn't very clear or easy to understand. However using this software we can produce decision trees and other forms of data that are much easier to understand and are very useful.



Clustering:

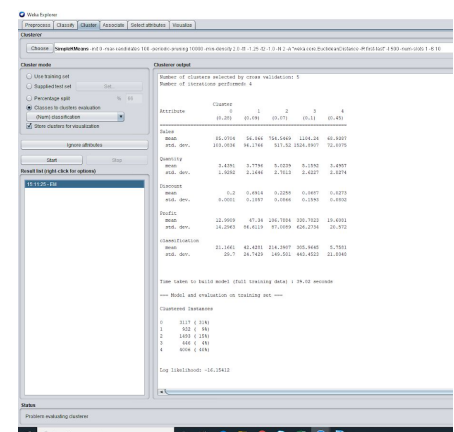


Clustering is a data mining method that uses an algorithm to find elements of similarity in the dataset. WEKA supports;

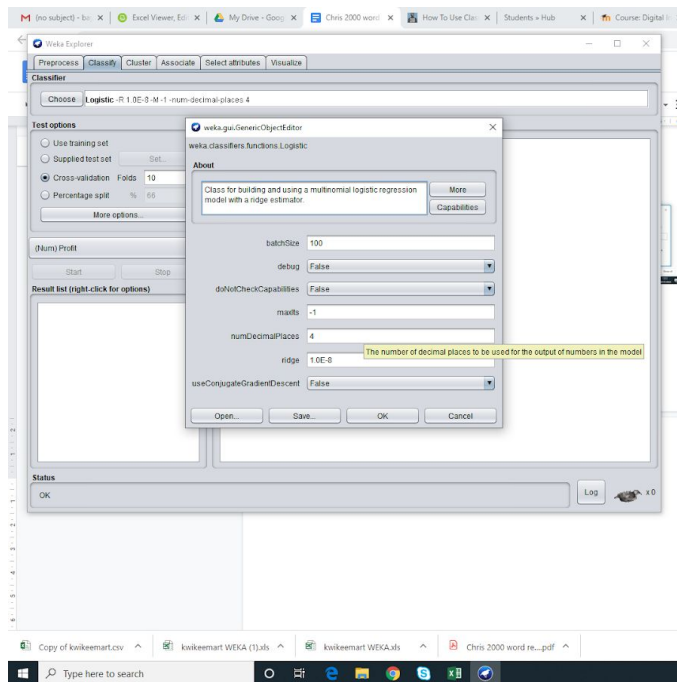
FilterClusterer, HierarchicalClusterer and SimpleKMeans, these are all examples of different types of clustering. The significance of clustering is that it puts data elements in their similar groups by dividing the data into subclasses. It is a fantastic tool for identifying patterns in the data.

Clustering is commonly used in statistical data analysis as it produces insights into the data by seeing what group certain products or profits the data points fall into.

As you can see the data has been split into relevant sections based on the input I selected.

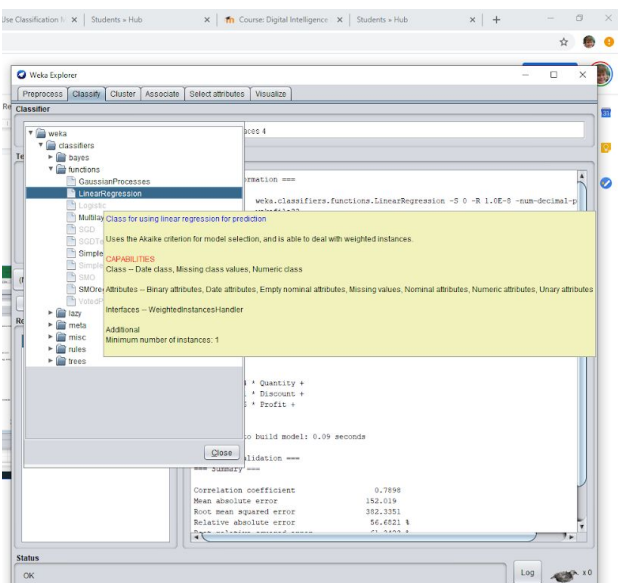
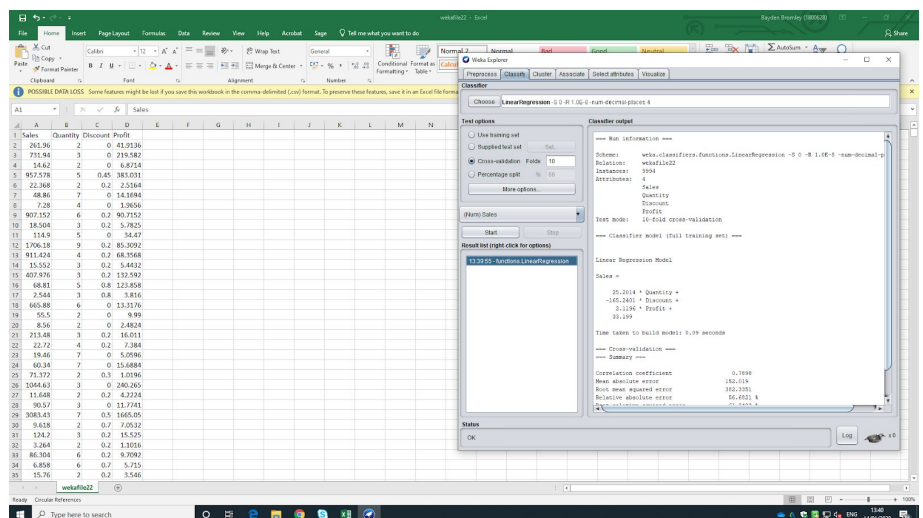


Logistic / linear regression:



Linear regression is a predictive analysis tool that is recognized worldwide. It estimates coefficients for a line that best matched the training data. The benefits of it is that its very easy to train and produces information that's very useful to businesses as well as easy to understand. Linear regression is used in finding relationships between two variables.

Its very commonly used in statistical analysis to determine the linear relationship between a dependent variable and one or more independent variables.



Linear regression is useful for determining relationships between the two continuous variables; the independent and the dependant variables. The predictor is the independent variable and the dependent is the response variable. It's interesting for businesses to identify patterns in their data using this method.

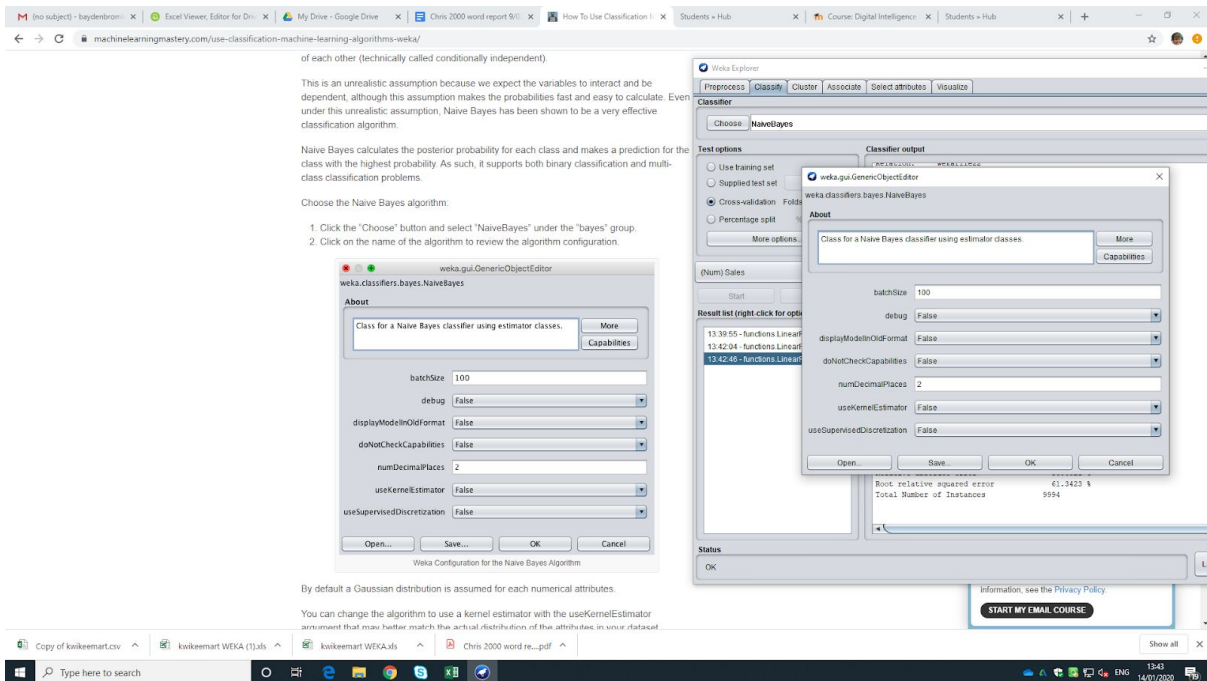
of each other (technically called conditionally independent).

This is an unrealistic assumption because we expect the variables to interact and be dependent, although this assumption makes the probabilities fast and easy to calculate. Even under this unrealistic assumption, Naive Bayes has been shown to be a very effective classification algorithm.

Naive Bayes calculates the posterior probability for each class and makes a prediction for the class with the highest probability. As such, it supports both binary classification and multi-class classification problems.

Choose the Naive Bayes algorithm:

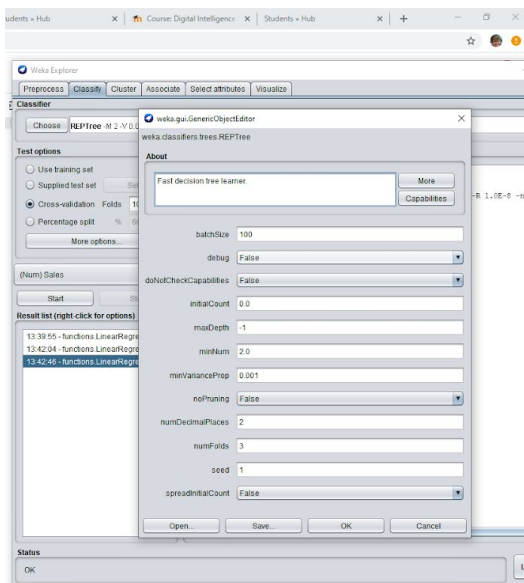
1. Click the "Choose" button and select "NaiveBayes" under the "bayes" group.
2. Click on the name of the algorithm to review the algorithm configuration.



By default a Gaussian distribution is assumed for each numerical attributes.

You can change the algorithm to use a kernel estimator with the `useKernelEstimator` argument that may better match the actual distribution of the attributes in your dataset.

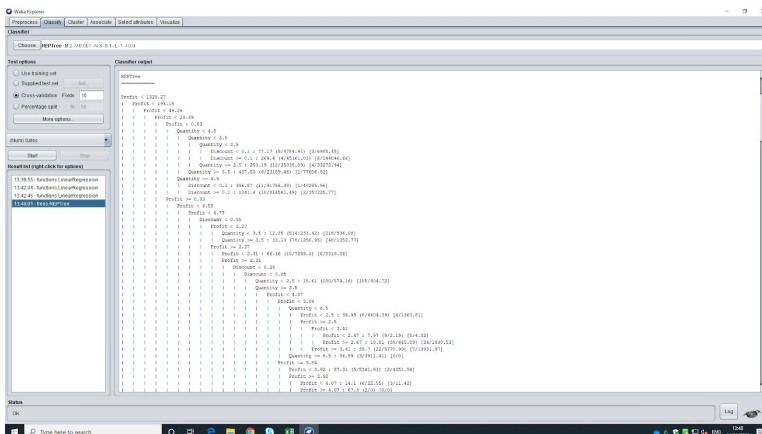
Classifications: Decision trees



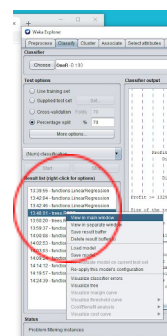
Decision trees are very easy to produce in WEKA and also provide a lot of information. The main downside is how cluttered the information can present itself, but that all depends on how big your data set is.

Firstly you have to select the REGtree option in the top bar then you have to ensure all the values are the same as they are here.

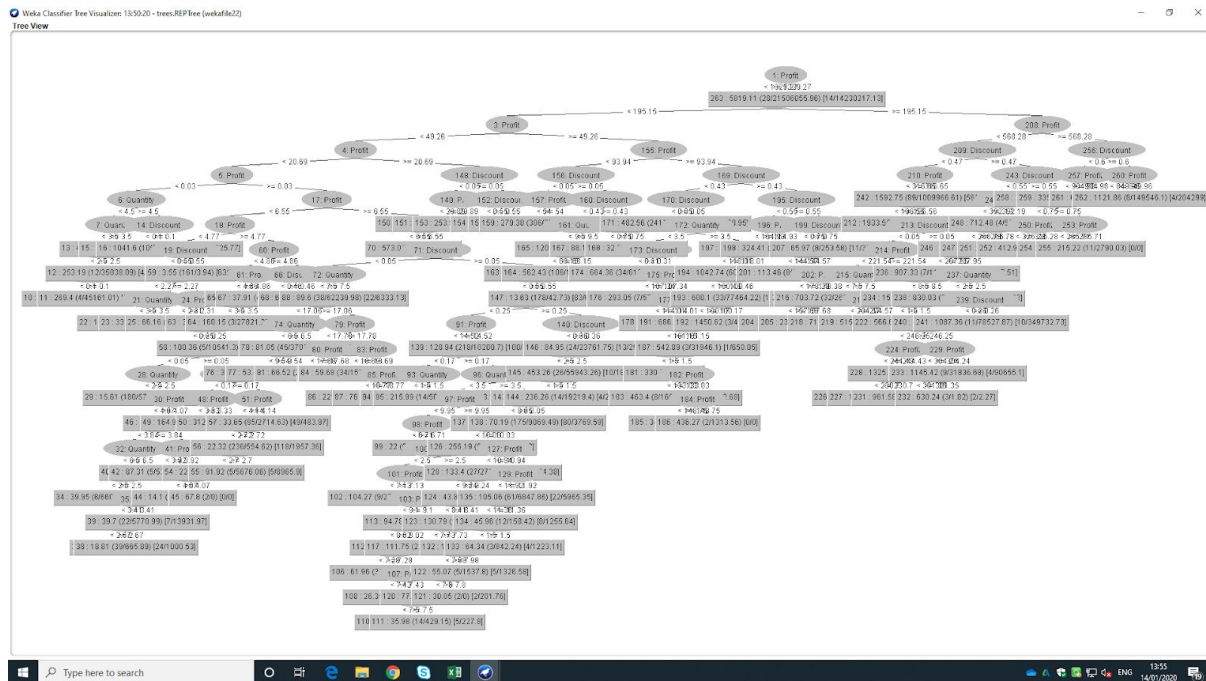
Once this is done you simply press start. This then gives you a form of cluttered data like this.



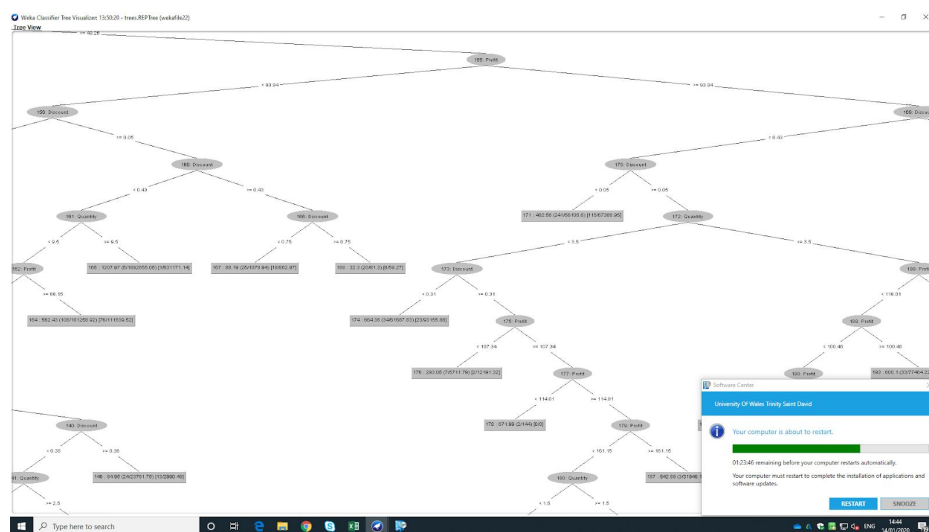
To make it into a decision tree you simply right click



on it and press 'visualize tree.'



The data shown is very cluttered which makes it hard for Kwikie-mart office and supplies to analysis but you can zoom in to make the data clearer.



It shows patterns between the data by putting them into their relevant categories. Abbreviating the method by showing predicted potential outcomes for each segment of the tree really helps businesses visualize their decision making. By identifying the patterns the problems are simplified for the business.

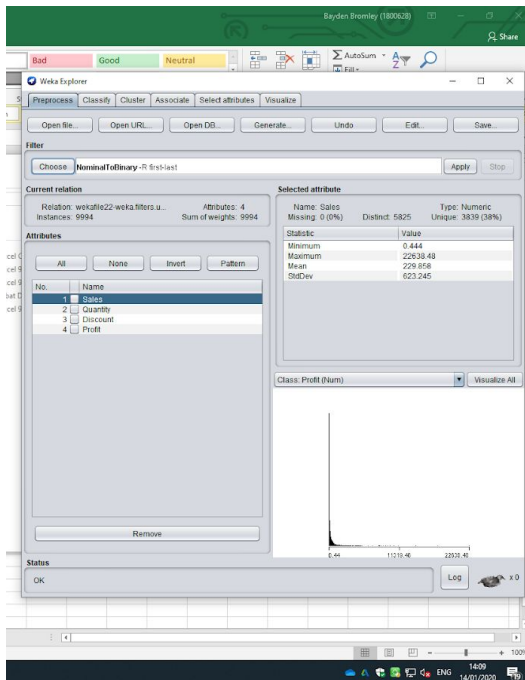
Regression

Regression analysis is a method of highlighting which variables have an impact on the topic of interest. It allows you to understand which factors matter the most as well as how the factors influence each other and which factors can be ignored.

In order for regression to work you must go to the filter tab and select 'unsupervised' and then 'attribute' the option of nominal to binary will be available. This is to change the class file to numeric.

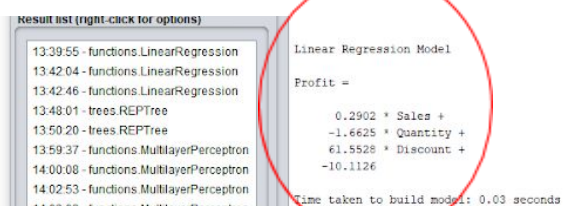
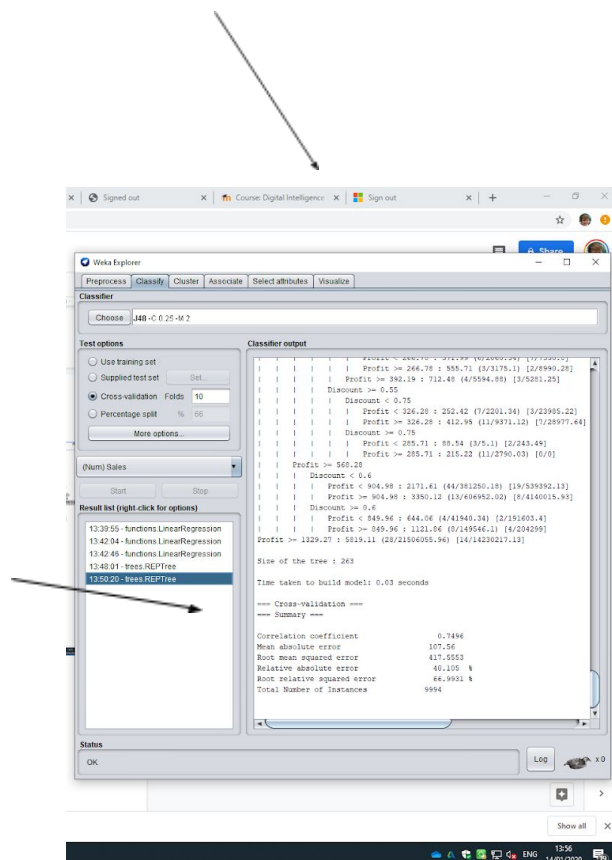
Once this is completed you will be able to select and use linear regression for your data set.

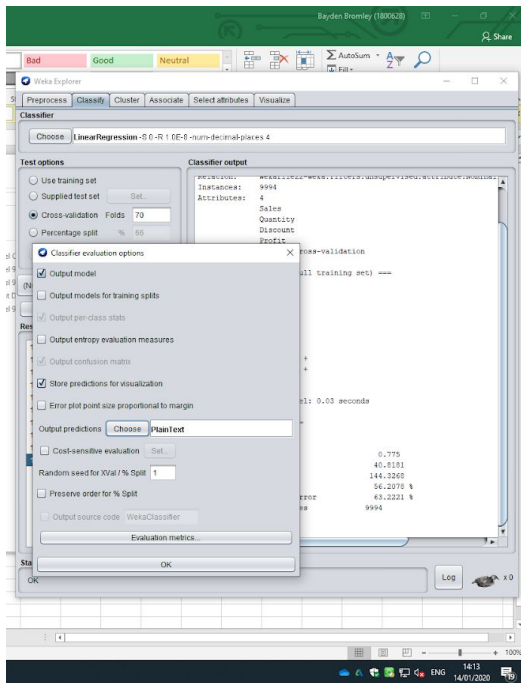
When you press start you will get data that looks like this:



This is our correlation coefficient:

This is our regression line:



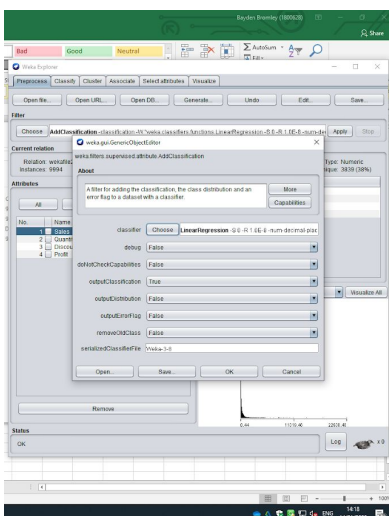
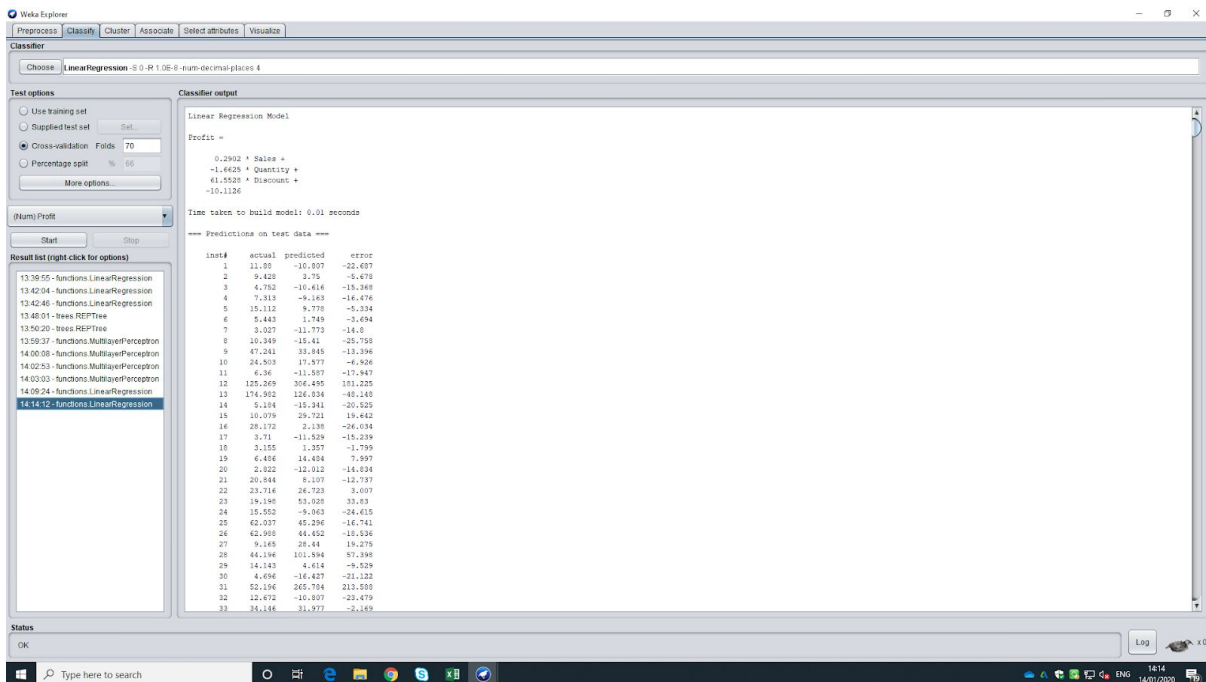


Another way to view the data in regression is if you change 'null' to 'plain text'

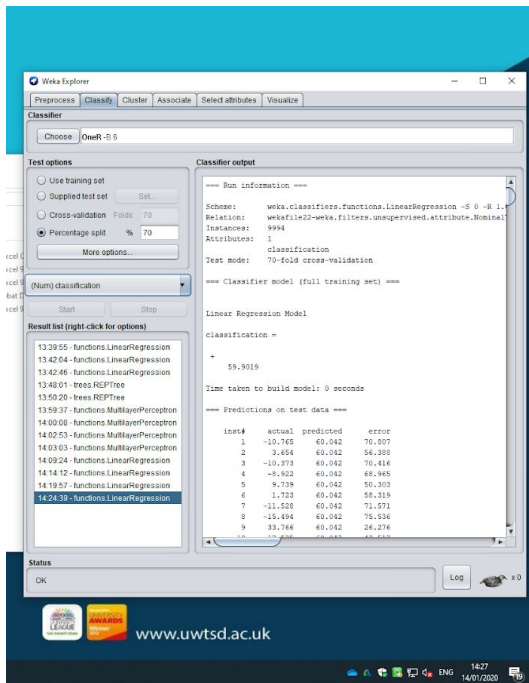
This will present the data in a long list. Allowing the business to view the actual and predicted data as well as the error data.

Regression is used to find the causal relationship between variables. The importance of it for small businesses is its ability to examine the relationships between two or more variables of interest.

So having multiple ways of presenting the data helps.

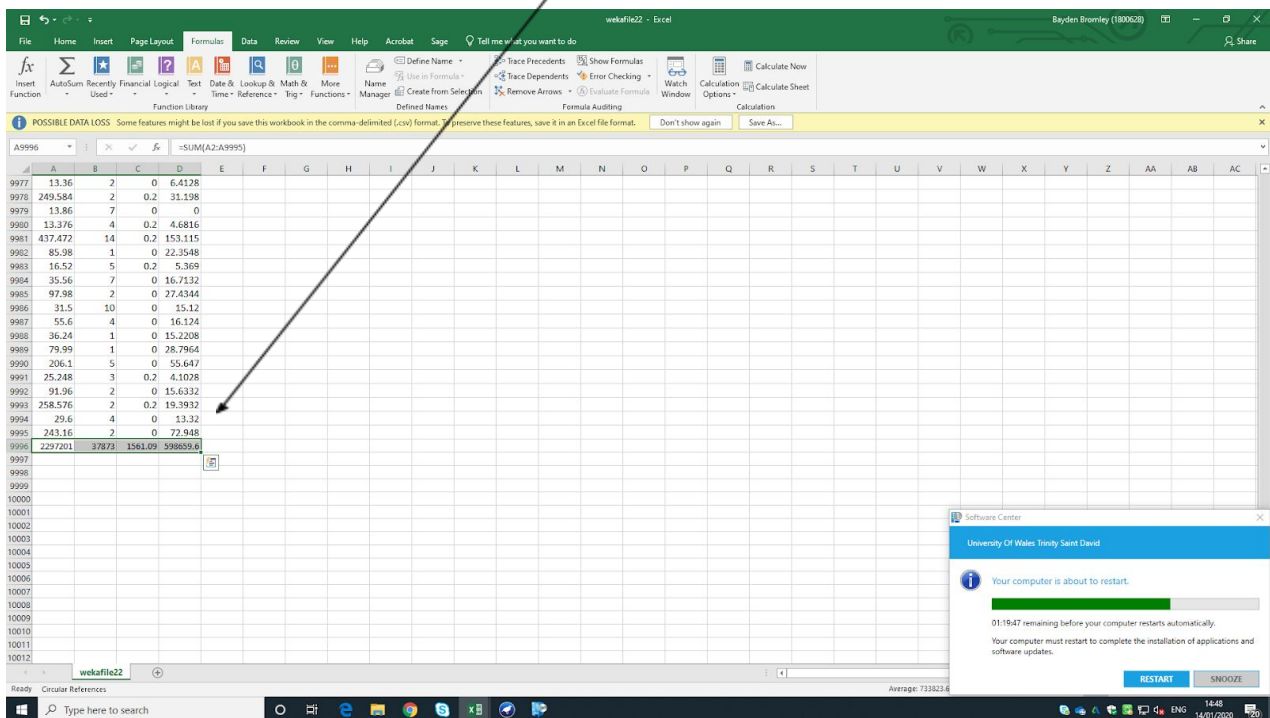


A simple regression model can help Kwikiee-mart office and supplies find the relationship between profit and quantity bought or revenue and time of the year. There are many types of data that the business can input into a regression model to help identify the trends and relationships between the data in order to try and boost the businesses market share and competitiveness and well as profitability.



Now finally using the regression line information we were given earlier. The business now just inputs the missing sum information to get the final profits for 2018.

Result list (right-click for options)	
13:39:55 - functions LinearRegression	Linear Regression Model
13:42:04 - functions LinearRegression	
13:42:46 - functions LinearRegression	Profit =
13:48:01 - trees REPTree	0.2902 * Sales +
13:50:20 - trees REPTree	-1.6625 * Quantity +
13:59:37 - functions MultilayerPerceptron	61.5528 * Discount +
14:00:08 - functions MultilayerPerceptron	-10.1126
14:02:53 - functions MultilayerPerceptron	
14:03:03 - functions MultilayerPerceptron	Time taken to build model: 0.03 seconds



With all this information at their disposal Kwikie-mart office and supplies can put in place strategies to boost their businesses product range as well as profitability. This data will give them insight into the products to use and products to discard. It will make their business more efficient by not supporting DOGs or Question Mark products in relation to Boston Matrix but to support products that are positively impacting the profitability of their business. This in turn will hopefully give them a better understanding of areas which need attention, which is what Kwikie-mart office

and supplies initially stated they wanted our help with.

Reference:

<https://www.phocassoftware.com/business-intelligence-blog/benefits-of-sales-analytics-software> -13/01/2020

<https://www.tutor2u.net/business/reference/boston-matrix-and-product-portfolios> 13/01/2020

<https://www.phocassoftware.com/business-intelligence-blog/benefits-of-sales-analytics-software> 13/01/2020

<https://www.springboard.com/blog/what-is-data-analytics/> 13/01/2020

<https://www.techopedia.com/definition/26418/data-analytics> 13/01/2020

<https://www.simplypsychology.org/qualitative-quantitative.html> 15/01/2020

<https://www.businessinsider.com.au/category/data-analytics> 15/01/2020

<https://www.businessinsider.com.au/category/big-data> 15/01/2020

<https://machinelearningmastery.com/use-classification-machine-learning-algorithms-weka/> 16/01/2020

<https://www.youtube.com/watch?v=kFKxUQDnLV8> 16/01/2020

<https://machinelearningmastery.com/use-classification-machine-learning-algorithms-weka/> 16/01/2020