

Assessment Component 2 – 75%

Written Report (3000 words)

You are a marketing assistant, responsible for interpreting data from your company's website. You use traditional software packages such as Excel for day-to-day analysis and more specialised data for mining the data.

Part 1 – 50%

Using the Superstore data set provided, determine the reasons for the store's decline in sales/profits over the years. Critically evaluate the strengths of using Excel for pre-processing the data, analysing the data and visualising the data. You will also need to demonstrate how you can do this practically with the use of Excel functions such as : IF, LOOKUP, PIVOT TABLES , charts and graphs. (1500 words)

Part 2 – 50%

You are an intern at an Audi car dealership has been tasked to analyse customer data to see if it can be used for competitive advantage. The dealership has a copy of Microsoft Excel and has just downloaded a free copy of the open source Weka data mining software. The company has used Microsoft Excel before but not Weka.

You need to produce a report giving a **critical evaluation** of why a small business such as this Audi dealership could use data mining software such as Weka for competitive advantage.

You will need to discuss the advantages/disadvantages of Weka over Excel. This should be a mix of theoretical argument as well as practical argument using the csv and arff files for the dataset.

Explain the most common data mining methods that can be used in business with real world examples.

Using the **audidealershipdata1.arff** provided on Moodle in conjunction with Weka give a specific example of using the J48 algorithm. Show your workings and screenshots and explain your results.

Using the **audidealershipdata2.arff** provided on Moodle in conjunction with Weka give a specific example of clustering. Show your workings with screenshots and explain you results.

The datasets are also provided in csv format: **audidealershipdata1.csv**
audidealershipdata2.csv

Consider what additional columns could be included in the audidealership1.csv file to improve the data gathered. Explain why the columns should be included and how the data when analysed will allow the business to become competitive. Further explore what types of functions could be used to analyse the data and generate graphical data and summary data and show evidence of using these functions using Excel.

A note on the datasets:

audidealershipdata.arff/csv

The dealership is starting a promotional campaign, whereby it is trying to push a two-year extended warranty to its past customers. **(An extended warranty can sometimes be a form of insurance policy that covers you for repair costs after the manufacturer's or retailer's guarantee has expired).** The dealership has done this before and has gathered 1,500 data points from past sales of extended warranties. The attributes in the data set are:

- Income bracket [0=£0-£15k, 1=£16k-£20k, 2=£21k-£30k, 3=£31k-£40k, 4=£41k-£50k, 5=£51k-£75k, 6=£76k-£250k, 7=£251k+]
- Year/month first Audi bought
- Year/month most recent Audi bought
- Whether they responded to the extended warranty offer in the past

audidealershipdata2.arff/csv

The dealership has kept track of how people walk through the dealership and the showroom, what cars they look at, and how often they ultimately make purchases. They are hoping to mine this data by finding patterns in the data and by using clusters to determine if certain behaviours in their customers emerge. There are 100 rows of data in this sample, and each column describes the steps that the customers reached in their BMW experience, with a column having a 1 (they made it to this step or looked at this car), or 0 (they didn't reach this step).