Data Mining Assignment 1

Identify a problem from your own experience that you think would be amenable to data mining. For that problem describe:

1. What the data is.  
2. What type of benefit you might hope to get from data mining.  
3. What type of data mining (classification, clustering, etc.) you think would be relevant.  
4. Name one type of data mining that you think would not be relevant, and describe briefly why not.  
For each, illustrate with an example, e.g., if you think clustering is relevant, describe what you think a likely cluster might contain and what the real-world meaning would be.

Write one to two pages of 11 point single-spaced typeset text - you aren't writing a paper, but it isn't short answer either.

1. What the data is.

Data are characteristics or information, usually numerical, that are collected through observation.Data is a collection of facts, such as numbers, words, measurements, observations or just descriptions of things. In a more technical sense, data are a set of values of qualitative or quantitative variables about one or more persons or objects, while a datum is a single value of a single variable. Data can be qualitative or quantitative.

* **Qualitative data** is descriptive information (it *describes* something)
* **Quantitative data** is numerical information (numbers)

**Quantitative data** can be:

* **Discrete data** can only take certain values (like whole numbers)
* **Continuous data** can take any value (within a range)

1. What type of benefit you might hope to get from data mining.

Data mining is mostly used by many of the big gaints in the information technology sector and also some small industries by making use of their own techniques. Some of the popular domains are,

* Market Analysis and Managemeent
* Corporate Analysis & Risk Management
* Fraud Detection

Data mining technology is something that helps one person in their decision making and that decision making is a process wherein which all the factors of mining is involved precisely.The benefits of data mining include:-

* Data mining technique helps companies to get knowledge-based information.
* Data mining helps organizations to make the profitable adjustments in operation and production.
* The data mining is a cost-effective and efficient solution compared to other statistical data applications.
* Data mining helps with the decision-making process.
* Facilitates automated prediction of trends and behaviors as well as automated discovery of hidden patterns.
* It can be implemented in new systems as well as existing platforms
* It is the speedy process which makes it easy for the users to analyze huge amount of data in less time.

1. What type of data mining (classification, clustering, etc.) you think would be relevant.

Classification analysis is used to retrieve important and relevant information about data, and metadata. This data mining method helps to classify data in different classes.Clustering analysis is a data mining technique to identify data that are like each other. This process helps to understand the differences and similarities between the data..Clustering would be relevant a lot more than classification, because of the fact that most of the data in the real life is unknown and we don't know about the classification of the training data before hand. Thus clustering can help a lot in the real life.

4. Name one type of data mining that you think would not be relevant, and describe briefly why not.  
For each, illustrate with an example, e.g., if you think clustering is relevant, describe what you think a likely cluster might contain and what the real-world meaning would be.