**NAME**

**COLLEGE NUMBER**

**TITLE**

**Design and Application of a Machine Learning System for a Practical Problem:**

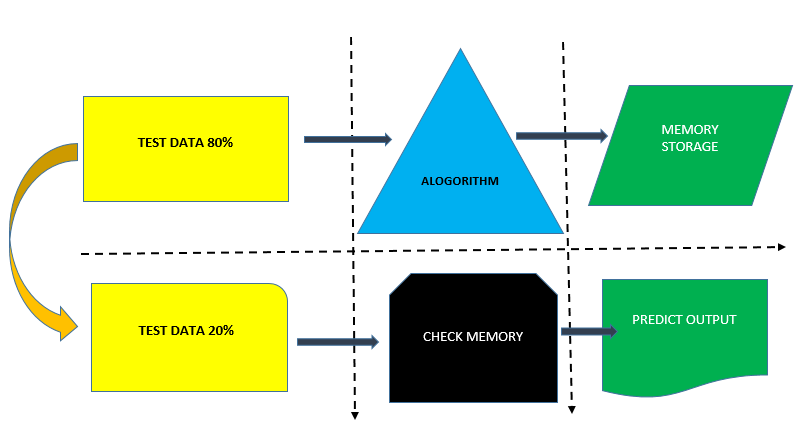
**A case study of the hotel industry openings in new locations**

**Introduction**

Data is becoming the core central performance of every organisation and application systems. Developers and engineers are consistently coming up with algorithms and software products that are tailored to solving the existing organisation problems. One of the existing problems that corporations are fighting to solve today are data related. Most corporations are producing a lot of data, in fact so much that they can’t consume within the allocated time and periods. So this is where data security comes in. Data security takes into considerations all the factors needed to ensure that the propose system, in both theory and practise is adaptable to the existing risking from both its internal surface and the external alienation.

However, as part of a major operative carried out by most banks, of the most of the factors that are concerned with drawing inferences from the datasets that are acquired by the organisations. Several departments like Human resource, supply chain and procurement, Information technology and marketing acquire a lot of data that the company can use in ensuring that the organisation is steered into the right direction. At this point, the role of data engineers and scientists come into play. Since management consistently requires regular updates and insights onto the existing departmental interactions within the organisation, data engineers have gone ahead to develop platforms and software applications that basically support this role and idea, they have done this by developing Business intelligent applications that robustly consume data, apply algorithms to it and then beautifully apply presentational graphs and linear visuals on the analysis to help the organisations make informed decisions. On top of the added functionalities provided by most analytical tools and software, the need for predictive analytics has also risen and companies want to know how much they should expect in future.

For instance, in case of a risk, how much is enough risk and to what extent should the risk go. The materiality of the risk entirely depends on how well the company has predicted the future reliance of the current state of affairs within the company given the current dataset and the future datasets. Predictive analytics relies in a core concept in the A/B rule of train and test. For 80% of the data supplied should be training purpose and the other 20% for the purposes of testing. During the training phase, the algorithm learns based on the dataset supplied, the learning curve helps the algorithm to identify what would be the most expected ideal scenario of on input. Consider the situation below:



From the above illustration, it can be observed that the 80% data supplied into the algorithm is specifically split for training and then the resultant learns are store in the computer memory. Later, when another test is supplied into the memory, it checks from the same memory and then reads the exact associated rule and the prints the required output from the table.

**Problem statement**

The existing hotel company needs a new machine learning model and algorithm to identify these best way forward to help it know if the next hotel branch opened in another part of town is going to profitable or not. For this particular exercise, in order to achieve this, the identification of the dataset existing and the machine learning algorithms amiable is going to help quantify, match and predict the next chain of hotel store profitability based on the output of the dataset.

**Objectives and aims**

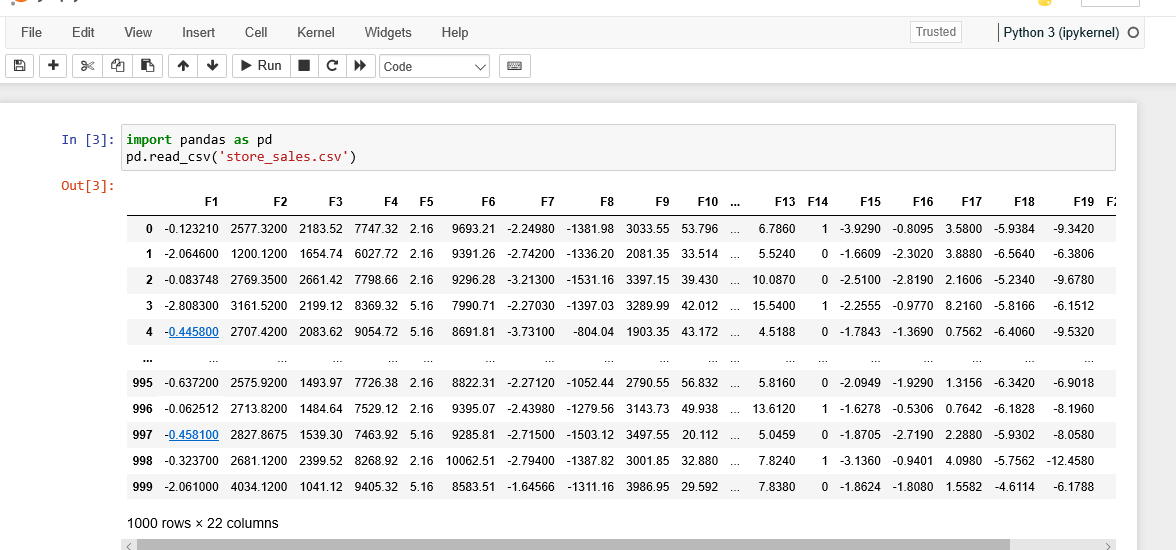
* To learn to identify machine learning techniques appropriate for a particular practical problem
* To undertake a comparative evaluation of several machine learning procedures when applied to the speciﬁc problem
* To predict the possibility of a new hotel branch being profitable in a new town

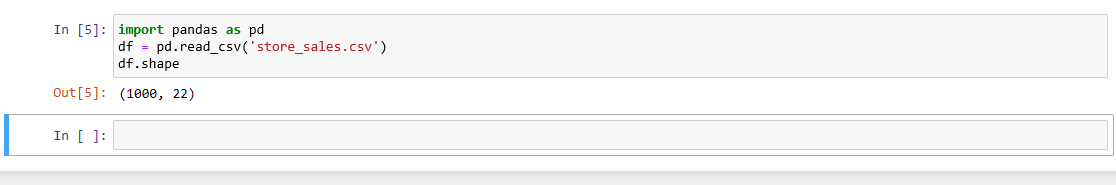
**Research methodology and design**

The analysis of this dataset is based on the python program and the Python program Jupyter Notebook. All code run on Python Jupyter Notebook web and interpreted. The stores dataset contains a list of 1000 records of store locations and their potential profits in the different locations that they are operating.

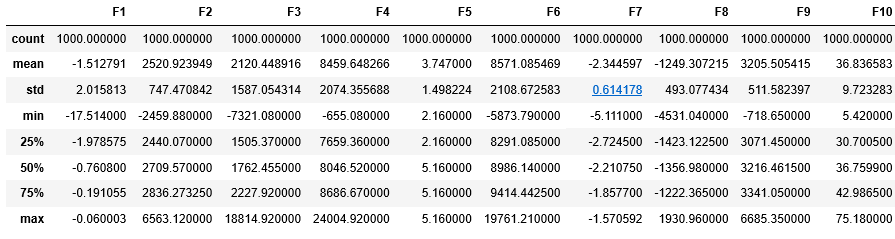
**Analysis:**

A preview of the dataset on Jupyter looks as follows:

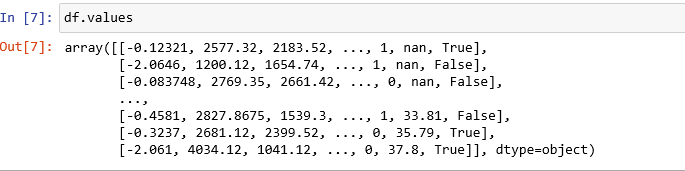


****

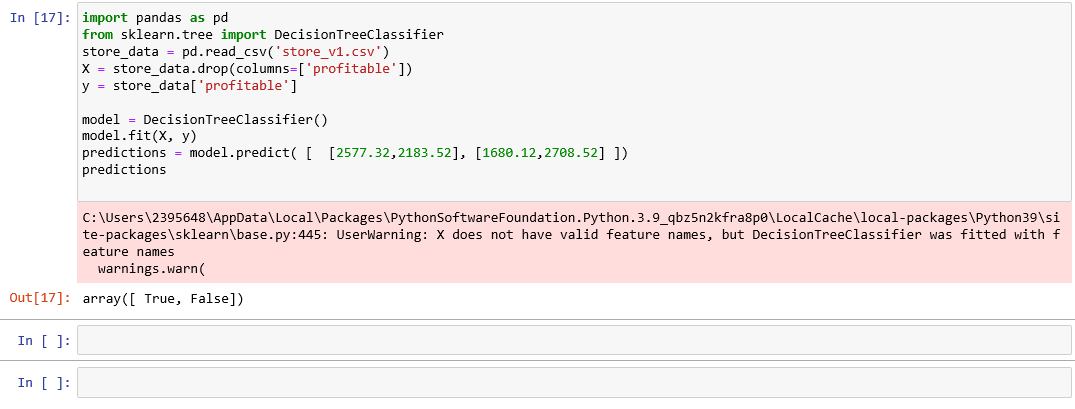
Using the def (), it was possible to show the records in the dataset and how many columns are part of the dataset. In this case, it was a total of 1000 records and 22 columns of data.



A summary of the column looked like below



For final machine model learning, the model was split by taking samples of this dataset that included the population of the area, the sales per shop and the status of the shop i.e. TRUE for profitable and FALSE for not profitable. The resultant dataset was put on test



The predictions from the above show that with a sample input of 2577 population and a sale of 2183 daily revenue, the shop in this place would be making profits whereas the reverse is true when the population is 1680 and the sale is 2708 . Statistically, this two variables re not directly related to one another and are subjects to other external forces.