**NAME**

**COLLEGE NUMBER**

**Abstract**

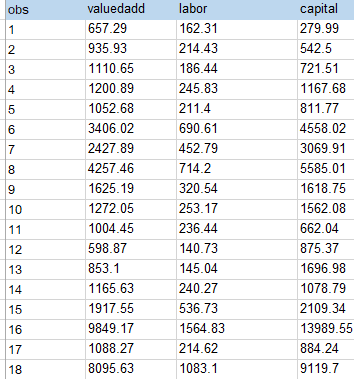
The application of data analyses systems span across nearly almost all industries today. Data science is the gold of the future, individuals and organisation are taking the advantages derived from the use and exploitation of the principle of the data science. Finanace, Healthcare and the education insdtries are currenly investing heavily as they tap into the full benefitis and capabilities of the dats scince industry. Engineers qwith a atechical bacjground have also gine ahead to develop software tools and programs that are focused on making the whole analysis process faster by developing algosthmic tools and software programs that automatically drive statistical inferences at the click of a button. Further, functions like sales and marketing, health sectos and the financial depwrtemnts are now employing the use of big data to help predict the next trend or point that their frecast might lland in. Examples here include future diseases discocveries and explorations, the next direction that the stock market might take and whether or not to invest on the particular shares and bonds, given past and current market prevailing factors. To achieve a given level of ecxpetced output and predictipm, the choice has to be made on the correct dataset, the model and the calassfication method to apply.

**Introduction**

One final objective of the data engineer/data scineties is to finaly be abe to produce analysis thata speaks to anyone. The data analysis process should contain a clear scenario of where the objectibe of the snalayis is coming fomr, the problem its trying to solcve and the future of such analysis. Managers and top management consistently rely on the analysis of these data to make useful and meaningful derivations. One way to achieving this is by using R programmimg language wihich is a sttsiasticak tool as well as adata viauslalisation pack. The R studio environment gives the user the full potenatial to grab the analytical caoabilities of this tool. R as a programming lanugae on te hand has benne used for by induuvia =ls and coiorrations ovr the years ti dkiver businessneeds and meet statsistics analayis that area rmeaningful. Addiotiona,another siginificance of the language is that its higly supported, easy to learn and can be has a number of packages that suppor quick analysis and visualosations.

**Dataset description**

There are two datsets provided for this partuclur project, all the datasets are delivered in comm saeparated files. A quick sneep in to he the frist dtaastet tahts is a sprodcution datset revelas the following contents:



Whereas the snip into the second dataset reveals the following:

