**NAME**

**COLEGE NUMBER**

**Abstract**

The application of Business intelligence information systems have been used across different across organisational functions. Companies and organisation re spending more and financial resources I trying to either acquire data, protect it or analyse the same information. Its vital thigh to consider the three stages of data even as when devolve deeper into the discussion surrounding the whole story of BIs systems. The BI process takes into account four different stages of managing the data process:

1. Data acquisition
2. Data storage
3. Data manipulations and analysis
4. Data processing and forecasting

Prior to the inception of the business application programming interdicted by Watsons COBOL, the programming languages developed during this era of time did not focus fully onto some of the various applications and possibilities poised by the BI systems, in reality though, the shifts was rather focused in the different possibilities that originated from the inventors that IBM was produced at this time to help organisation and corporations achieve a given level of business solution and survive mechanisms to its customers.

Later on such advancements would later be taken over by data scientist who would were experienced in the various s fields and saw the need to make good use of the knowledge and experience that they had gained working in this industry. Moving forward this era saw the development of software programs that focused on the development and production of tools that fitted the industry, some of these tools included FOTRAN DNA SCALA. These tools are still in use to date but their inception is a story that can only be derived from the pioneers and engineers who saw the need to invent something that individuals, corporates and other engineers could adopt into their daily work programs.

Further, statistical problems were discovered and engineers from the various fields continued to develop and implement other tools that could support other researchers and students from this field. It is at this point that other statistical analysis tools like RStudio, IBM studio, Matlab, SATA and STRATA applications. These tools have consistently been used over the years by data scientist to generate very amazing insights to the target audience and forma big part of preferred tools among some of the data modelling platforms whose applications have spanned across Finance, healthcare, education and marketing.

Another component of the data process is the data storage mechanisms. In order to understand this at depths, its crucial to also understand the length upon which how organisation go to keep and manage their data, keen attention however has to be drawn there types and elements of consuming data. Data node also known as data entry points act as the funnel upon which incoming data is derived and channelled into the correct data silo. Data silos are considered as the large stores upon which data and information be stored for future need. An example of data silo application is in healthcare where data is stored as it comes from the field and stored in different compartments within the server. This data can later be accessed by data engineers and parts of it extracted by data engineer, monitored, extracted and trained, based on an 80/20 rule where parts of this information I used to predict disease patterns , trends and monitors the current infections and treatments in the healthcare industry.

**Drate the data silo adnd wareheouse infogarpha here:**

There are three types of datasets that are considered when storing information within the database environment:

* Structured databases
* Unstructured databases
* Independent data systems

Structured database systems basically store data in organised rows and columns. Within the database are tables that define the different entities that are supposed to be considered as choice for the various data points originally connected to the database system? Each table is identified by a unique table name, sometimes a schema prefix comes into before the table name. As a master of security though, table names are sometimes not called by their specific table’s names, but rather a certain random gibberish name is set just in case hackers get access to or system becomes composed for whatever the reason is valuable ta that time.

Each table within the database is identified by unique primary key that references that particular row in the table. Another key unique identifier in the table is the foreign key, foreign key are used to join the different elements of the tables together without any compromise to the existing previous table connected to it. As we shall see later in this discussion; by using the primary keys and the secondary keys associated with other tales in the columns, it’s easier to reference the tables within the same dataset and join the tables together. Structured database systems have for long time been considered as the primary source of database type among developers and data analysts. They include examples of Postgres, DB2 and Oracle. The language of the database is considered as structured query language (SQL). As shall be seen in the later discussions, SQL is able to:

* Get dataset information
* Read from tables
* Update tables
* Insert into tables
* Delete tables
* Join tables together to get information

SQL procedure can also be written within server applications to help grab the required statistics on the database that they are reading form. This information can in turn be read and presented on the Business Intelligence program. Unstructured database systems on the other hand do exactly the opposite of what that structured database systems are supposed to do. The data in these systems do not have any organised rows or columns but instead the data there in comes on heterogeneity, that is, information cutting across the database can be referenced just with singles id and the result set can pull almost any required records under that particular ID. Not so many developers and data scientists are up to depth and ware of this kind of database so its use in the industry is also quite new and low. With the introduction of big data technology, this kind of database suites the heterogeneous kind and nature that the structure of big data offers, in later and further studies. Unstructured data types can accommodate almost all types of data variables including characters, strings, images, documents and videos in no organised columns or tables.

The other database type has been introduced by the inception of web 3.0. This kind of technology is almost the latest in the industry. Its use and inception cuts across the financial sector. It’s the kind of database that supports the block chain technology. Its major features is that data is not stored on any major applications or server, but rather the data is stored in smart contacts. Web 3.0 views each and every transaction on the etherium as a single block of separated and independent for every block of transaction identified by a particular ID. So, transactions can be interpreted by the other users in the same type of transaction. This makes the smart contracts running across the etherium network hard to guess and infiltrate since very high encryption algorithms have been applied in the network and there is in single database entity sitting somewhere on the particular network to crack. It’s also highly considered that the web 3.0 is going to take over as the next phase of data handling.

Once data has been captured and stored in the database, the next phase in the development of the Business application systems to determine which type of analysis to employ. These analyses range from general statistics on to which measures of central tendencies, specific counts, averages and the deviations. Then other factors to be considered in place include the kind of models to employ on the BI system. These may be include:

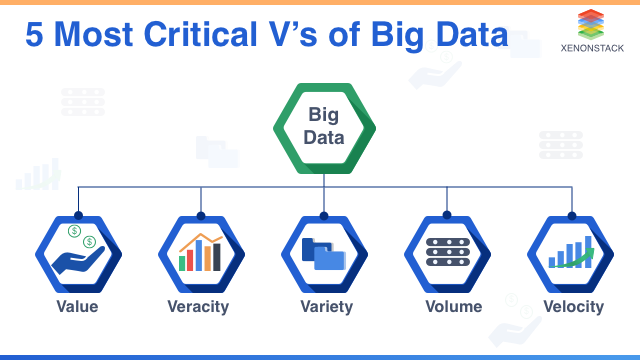
* ARIMA
* KNN
* Random forests
* Linear regressions
* Logistics regressions

The significance of each model shall be specific to the kind of organisational problem at hand. Their effectiveness may also be slightly restructured to the kind and nature set being put in place. Forecasting is another key proponents in the data analysis stage. Top management would always want to know what to expect next, where to expect it and from whom. What impact can such decisions have on the company performance and how should the company prepare in case the market changes in any way. A quick instance here is how the finance industry uses forecasting to predict the market behaviours so that accurate information can be made on before trading the next stock

BI application systems cannot be complete without t visualisations and presentations. Actually, the whole reason behind the rigorous analyses that is done on these datasets is other the company’s management can get the correct interpretation of the data being prepared. Some of the tools that have been previously used for visualisations include Microsoft excel, PowerBi, Rstusio SAS studio. In this study, the researcher focused on the production of interactive charts using the analyses on Microsoft Excel and SAS studio. As shall be seen in the last chapter of these analyses.

**Big data**

Organisational data points and nodes produce massive information that is fact moving,is in varous forms ,high inconta and gas mot elemnts of factual statisctis. Thye arecalled the 5vs of big data. They infomahe excat nature of big data and how its shuldlld be handled.



This is the kidnd of scenein in which comoanues andoraganisatins have foud themselves in today, they produce more tdata tahatn what ttehte can evargelyl consume. These data originate re formthe vatou data apnts that the organisation has put forward. They incude email ysystems, social media appliactions. Other deciosn support systems such as koknwodlefe meanamgement sysmets, iformatn oortala, wesbiyte ffedbac forsums amd the clll logs are consided as some of the maho jyrsoucres of big data.

**Introducing Business information systems**

The develomeny fo yhe BI system si fogansed itn o threemahot pahes:

1. Definitin of the requirements
2. Desingin gof he sporortopy
3. Development if the application
4. Testug anf inommeyatuin
5. Hand oberv and meentance

Since the BI is a kind of a soatware programm, its evekipment nust slaos fowwthe princeeplnadesigned ad define sd by IEEE.



The fisrst oahse during this process is to define the sfteare [secifucataina dna dinstructins that aren nededd toecsceuted yhe plan. The reqyurementsincude the finacia costs to besinsvcurred , the user exepernec sand skikks thata re nended, like he programmers and data a egineers, spgrama tsters and may be tye ysvala installalation or perchase of specific severs witjin the company. Another requremenbt that is uakkuya of chcoce when it cime to sosfwtare prgigrams ais to wther to consdre aoutsorucug or developing the aoplicatuij fromi inside the house, the choice within hwoch depends on management desvins and comarson sof the finacies aneeded to achove the same.

The next step is to design the prorotyype of theapplication. Usually the desgnand prortyep fo the BI system will give the inversors , management and the develipers a rough estimate and visual imoresson of twhat the final application is most likely to look at. As this is done software bigs are identified and patched befre the release of the appliocation. The prortyoyype can also be sued as sa tool fof fianacial marketing to poetentaol investors an finacuers.

The development do the BI will require he help of the shoetaee engineers t code the aplocatio, eoether from scarthnor by getting a acutom sysyem which can then be edited and replatetd across other systems. The role of data enhneers atthis levela snd tage if thev develomentv is to help identify ow the data is being cibfreted fromits spurce, write the exact code and querries that will get ethedofetare and cpde from its location to where its needs to be. The BI slos relies much o charts and visual impressions, the dofetare programmers with the help of the data analysts define these stages. Once the devekpment has been done, the next phase is to test the applovation across other cross fiunctionla departments and check if its meets the ruwored output.Sample dta can be thrown inot it from the warehouse. Under diffrenet nnevronments and suscicmustnacesm this data can eba measured to see if te works and meets the required output. Once the gre light has bee achieved, its now time to move on to the jandover part.The technical depattment stllbhas to define as to ehtaher this data nd ofnromation needs to be stores in the serber someher eor should just be used localy withihn agiven IP address nektwork, the choices of whoch deepend and telly in the kind of insftarctiure alresuces atharereadily availlal to tie the irahnisaio in question.

**Significance of BI systems**

The value of Bi systsems iseviedbet and can be qaunatified:

Foremoest, time is reduced drig reporting. Nlike other caes wherewe ave data scmeyes scsrouing through fles and files fo excel trying to find thataine r report and datsoirce to analyse, BI systems a uatmaticaklygenerate reoorts and taa that can ist be clcked adnd dowolaoed in form of pdfsm,excel documents of just normal world docunsmunets

Finacial risks are also educed in the process of mlmnentaion the BI.Uuslay reveue leaks coour beacue three could be missing oints in reporting some aspects pf the applications missing pr not clearly captured by the business analysts.

Reuvctin of errors and missing onformation. Once eberthing is sautomated, its not easier to identify or come across any misisig information from the application. The reporting is real and airomated.

Decsuionsare made fastre. Tatactsics and nubers fromg the Bi are reported in real time, wat this means is that the time magetsr ad deceion makesrs take to act oom data is nw reduced and information action now happnd fatster. This helps the company to e decionsfacts and implement any cnges earlier idnetfied and disucseed.

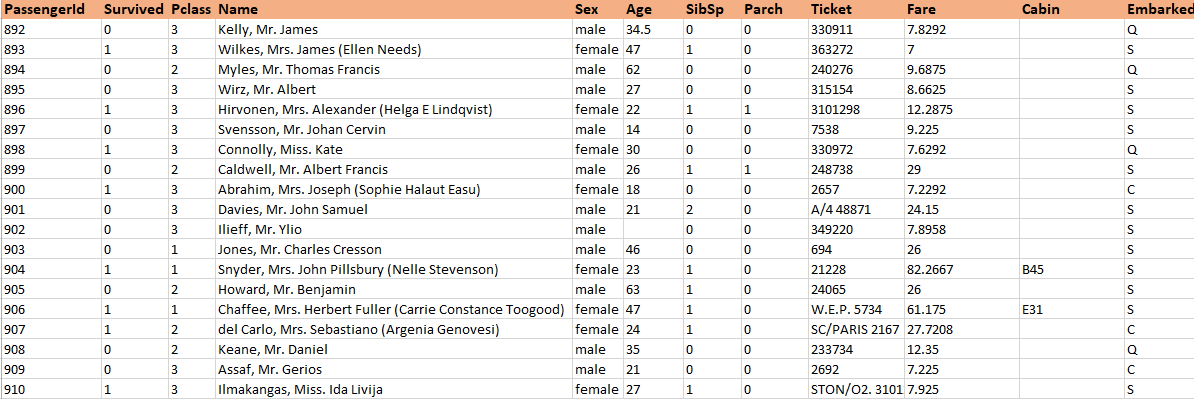
**Data analysis**

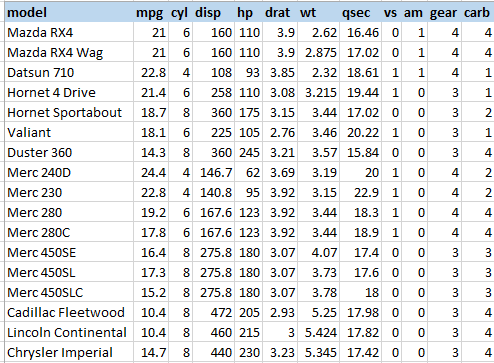
The last step in this statge is the analysis of the datatset. To do this, two sepate dataset ate conosredd:

1. The titanic dataset obtained from Kaggle website
2. The mtvars dataetobtained from the kaglele website

**Data datacsrption**

A highlight of the two data sets reveal thefoloonig imfomtsyomnncmncering thevariables ibtaine dtherein:





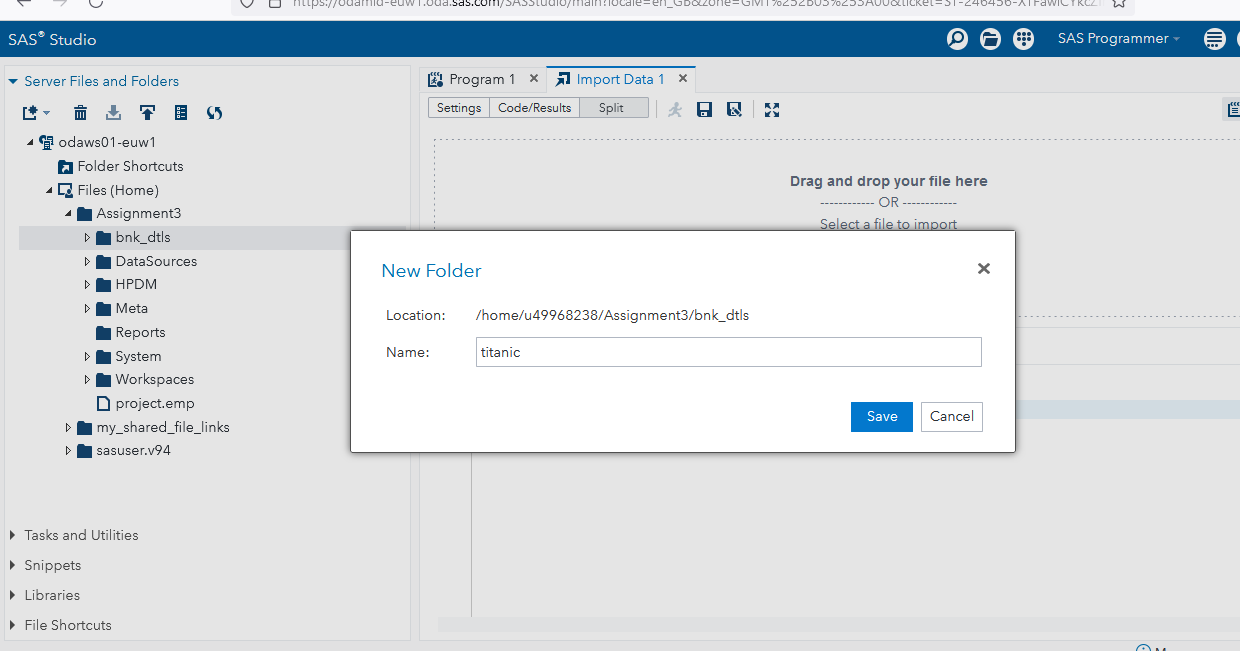
The two datasets shows vaables that will b emsuerd afains ineach other. The taaitanic dataet is acaollecton of lla thise who were onbairdc during the night of 1912 when the titanic sank with its coovpunats, most of the data here can be use nalayse the summary of e couupants ta the stime the acctident hapmed. There are a toyal of 418 records of data ta the time of this nalaysis. The seond datastas is a list of the fanus mtcars dataset, this datset is a averu good stadrad for measuting the re;ationhip between datasets, on how speed of a vehicles is facted by other datsets in the domain such as dinstance and the capacity of the engine.

**Aims and objectives**

1. Get a suammry of the mtcars dataset
2. Get a summaru of the titINC dtastet
3. Derermine the rlrationhsipn between fators dsoplayed in the totnci dataset
4. Dtermine the relathop btewen the factors in the mtvcars dataset

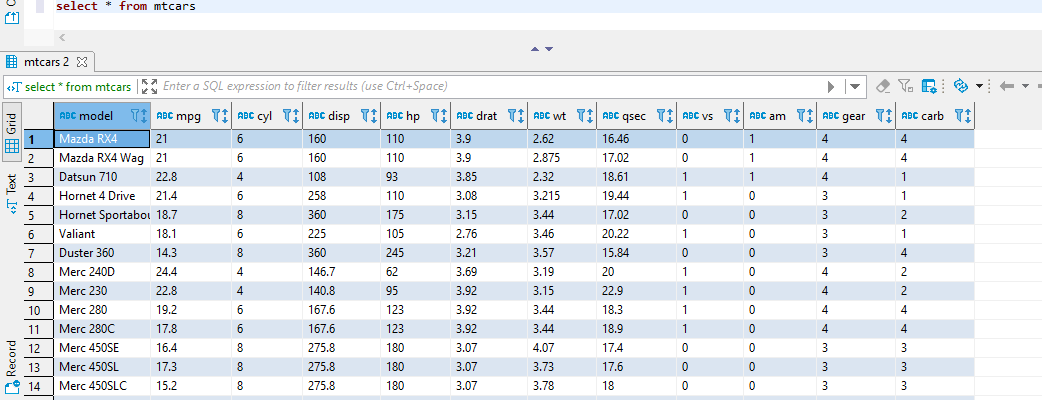
**SAS studio analyses**

The first process is to load the dataset into the SASstudio environment as below:

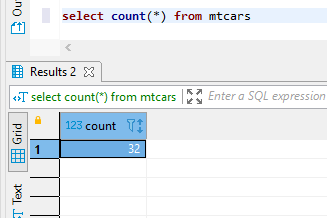


**Data analysis with SQL**

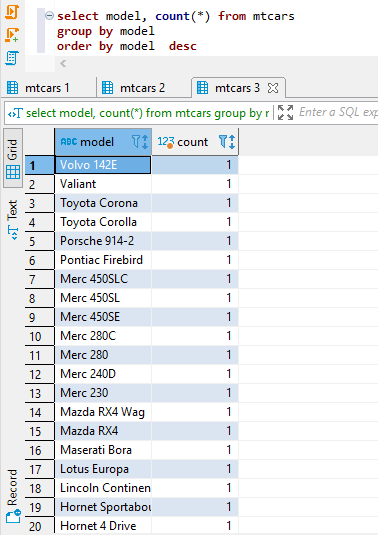
Selecting all records from the mtcars from h the dataset:



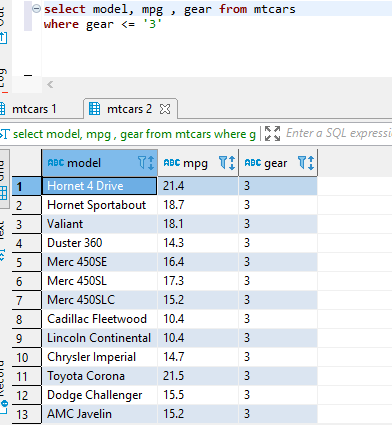
Selecting the count f records within the mtarcars tables



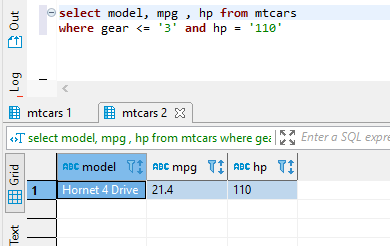
Group the cars by model



Seletcing model, mpg, gear where gear is less than 3

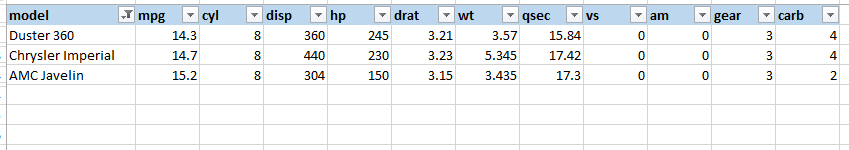


Selecting car model. Mpg, horsepower where horsepower is equal to 110

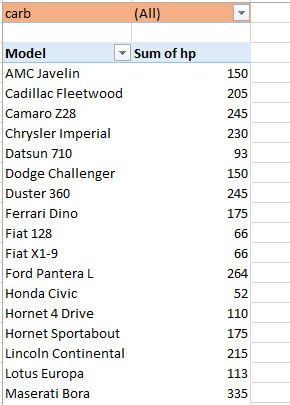


**Analysis on the Microsft Ecxcel**

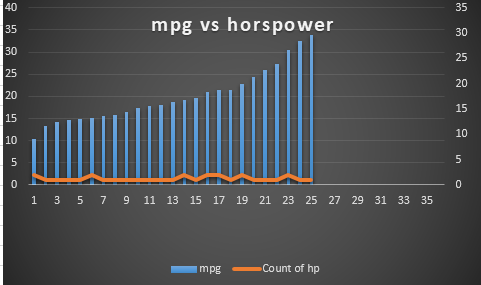
Filterimg datsasets for AMC, hcrystler and Duster 360. The relust returns three items frim the list:



Appluying pivit tables on the dataset and filetrng by carbs

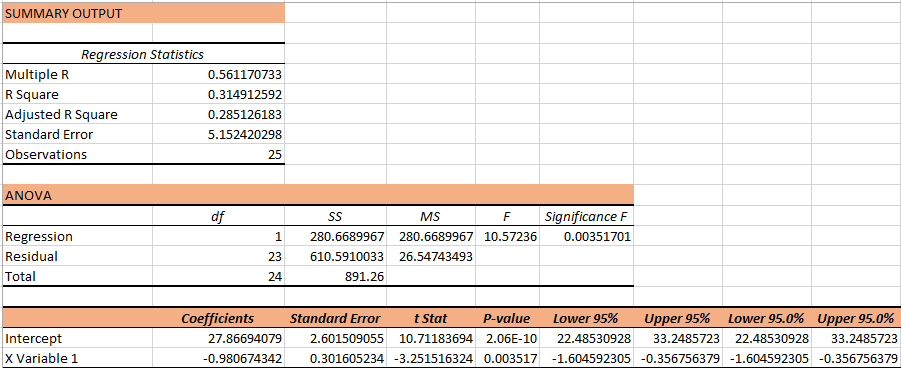


**Plotting a scatter graph and chart based on the dataset**

****

The graph shows the raleatiohip ntwen the miles per galling of the car and the different hiurspier, in reality thugh, this is nt just enough to determine if there is a relationhip between the two items since an increase in the mpg does not in turn show that there was an increase oor devrase in the hoisrse power. So at this point we cannot still establish any relationhips between the two datasets.

Summary of the datset between the mpg and the cylinder capacities



If we pay attention to the regression value, the result given above is 1. This is very key in understanding thereatinhip btween the two variables, the understanding is that the two varaibles are closely related and as one moves in the positive direction, so does the other.

**Summary and conclusion**

BI systems are effective in managing data system espacualy where large amnuts of data are involved. Dta asilos are large stotage devise for randomky andfrequently incoming data. This data can be stored and used for future use. Dta cana be stired ether in stsructred or sbtertycerd adtabses systems. Strcrured database systems are stored in tables and columns. Web 3.0 does not store any kind of data in any way, instead data is managed by smart contracts. The BI systems can benenfit an prfanixation in many ways that include the ability to patch revenue leaks, to report acciately and timely and finaly to reduce reporting time on data. Its advidables fpr business organisation s to adopt BI systems hiven its ability to trsnafrm the tenteprisr and scake them to another level of deceiosn support prgrammig. However, this whole process requires cost, lbour and skill managements in place.

**References**