**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.

**Introduction**

**Draw the data silo and warehouse in graph 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:

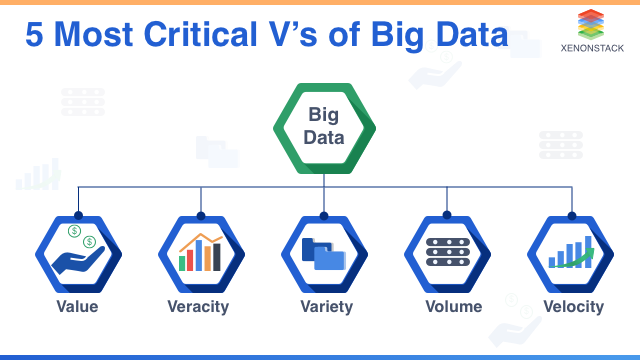
* ARIMA
* KNN
* Random forests
* Linear regressions
* Logistics regressions

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 various forms, high in content and has more elements of factual statistics. They are called the 5vs of big data. They inform the exact nature of big data and how it should be handled.



This is the kind of scenario in which companies and organisations have found themselves in today, they produce more data than what they can averagely consume. These data originate from the various data points that the organisation has put forward. They include email systems, social media applications. Other decision support systems such as knowledge management systems, information portals, website feedback forums and the call logs are considered as some of the major resources of big data.

**Introducing Business information systems**

The development of the BI system is organised into three major phases:

1. Definition of the requirements
2. Designing of the prototypes
3. Development of the application
4. Testing and implementation
5. Hand over and maintenance

Since the BI is a kind of a software program, its development must also follow the principle as designed ad define by IEEE.



The first stance during this process is to define the software specifications and instructions that are needed to execute the plan. The requirements include the financial costs to be incurred, the user experience sand skills that are needed, like he programmers and data a engineers, programmers, a testers and may be thee system installation or purchase of specific severs within the company. Another requirement that is key of choice when it comes to software programs as to whether to consider outsourcing or developing the applications from inside the house, the choice within which depends on management designs and comparison of the finances needed to achieve the same.

The next step is to design the prototype of the application. Usually the designed prototype of the BI system will give the investors, management and the developers a rough estimate and visual impression of what the final application is most likely to look at. As this is done software bugs are identified and patched before the release of the application. The prototype can also be used as tool for financial marketing to potential investors a financiers.

The development do the BI will require he help of the software engineers to code the application, either from scratch or by getting a custom system which can then be edited and replicated across other systems. The role of data engineers at this level and stage of the development is to help identify which of the data is being taken from its source, write the exact code and queries that will get application and code from its location to where its needs to be. The BI also relies much on charts and visual impressions, the efforts of which relies on works of programmers and with the help of the data analysts to help define these stages.

Once the development has been done, the next phase is to test the application across other cross functional departments and check if its meets the required output. Sample data can be thrown into it from the warehouse. Under different environments this data can be measured to see if the works and meets the required output. Once the green light has been achieved, it’s now time to move on to the handover part. The technical department still has to define as to whether this data information needs to be stores in the server somewhere where it should just be used locally within given IP address network, the choices of which depend and tally in the kind of infrastructural resources and readily available to tie the organization in question.

**Significance of BI systems**

The value of Bi systems are evident and can be quantified:

Foremost, time is reduced during reporting. Unlike other cases where we have data located in some files and with the need to convert them to excel, BI systems a automatically generate reports and that can be clicked and downloaded in form of pdf, excel documents of just normal world documents

Financial risks are also reduced in the process of implementation the BI. Usually revenue leaks occur because three could be missing points in reporting some aspects of the applications missing or not clearly captured by the business analysts.

Reduction of errors and missing information. Once everything is automated, it’s not easier to identify or come across any missing information from the application. The reporting is real and automated.

Discussions made faster. Statisticians and members from the Bi are reported in real time, what this means is that the time managers take to implement decision makers take to act zoom data is now reduced and information action now happen faster. This helps the company to reach at implementing any changes earlier identified and discussed.

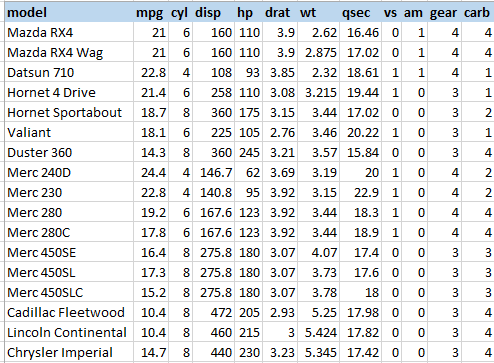
**Data analysis**

The last step in this stage is the analysis of the dataset. To do this, two separate dataset are considered:

1. The titanic dataset obtained from Kaggle website
2. The mtvars data obtained from the kaggle website

**Data description**

A highlight of the two data sets reveal the following information concerning the variables obtained therein:



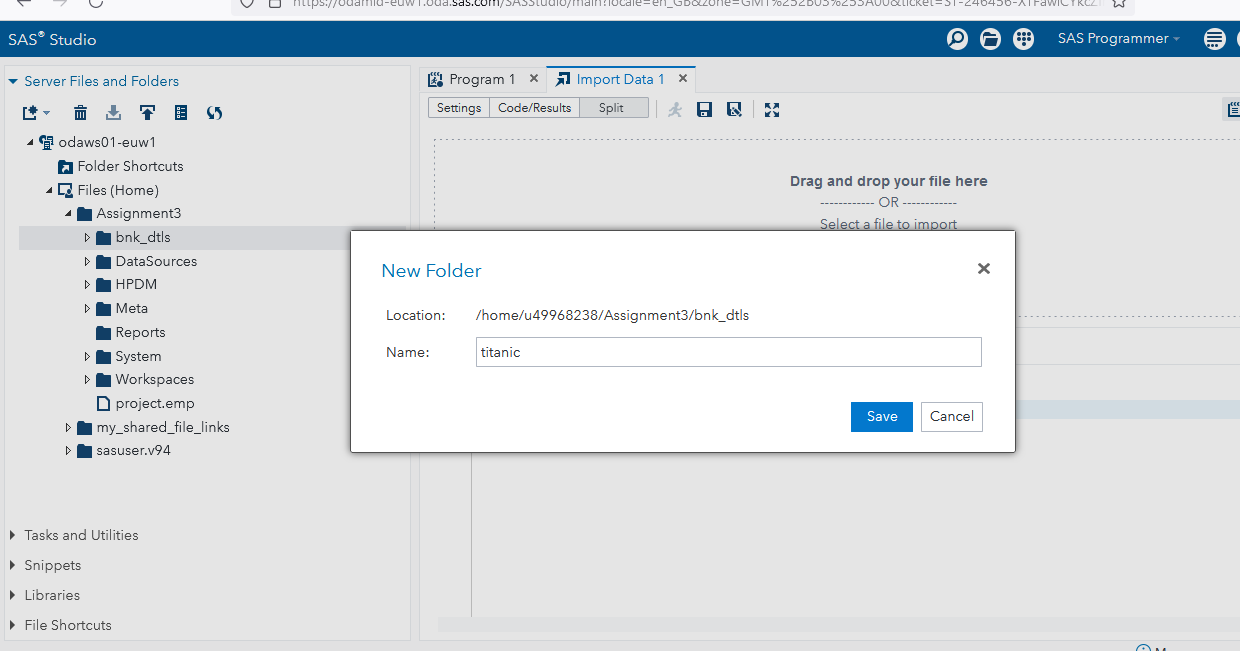
There are a total of 418 records of data at the time of this analysis.

**Aims and objectives**

1. Get a summary of the mtcars dataset
2. Get a summary of the titanic dataset
3. Determine the relationship between factors displayed in the titanic dataset
4. Determine the relationship between the factors in the mtvcars dataset

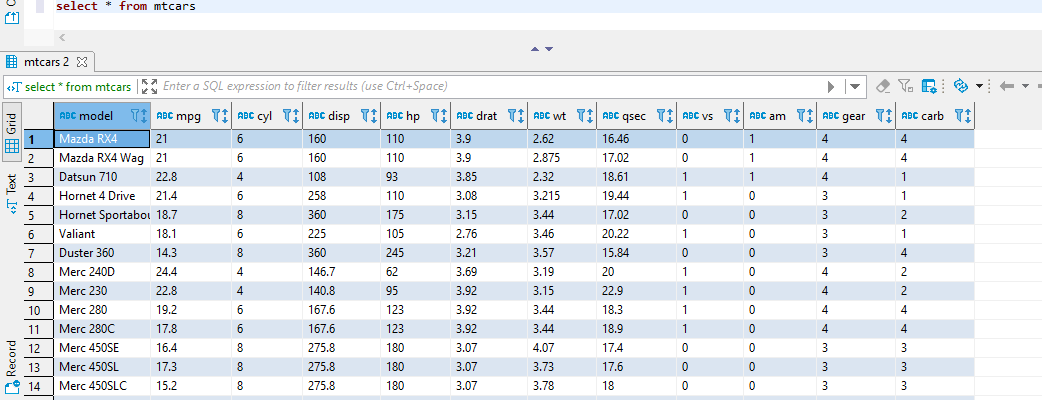
**SAS studio analyses**

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

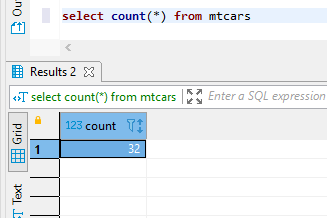


**Data analysis with SQL**

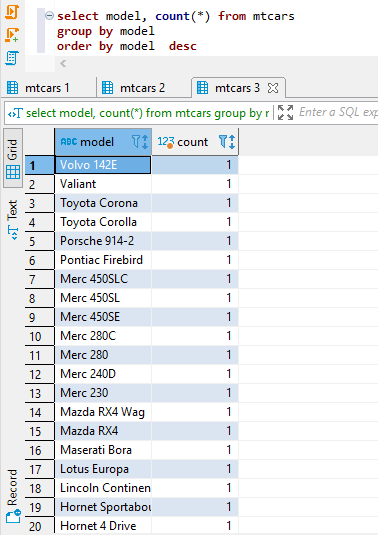
Selecting all records from the mtcars from h the dataset:



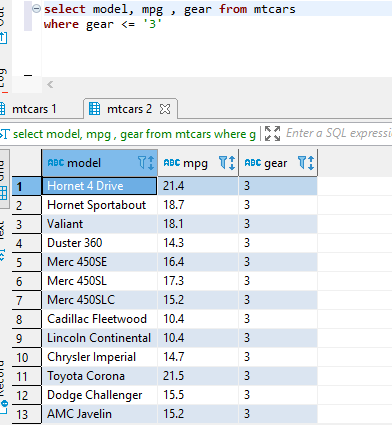
Selecting the count f records within the mtarcars tables



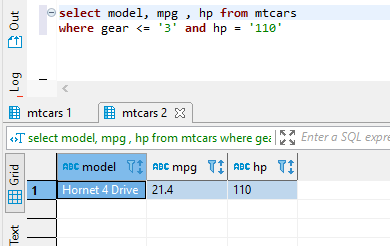
Group the cars by model



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

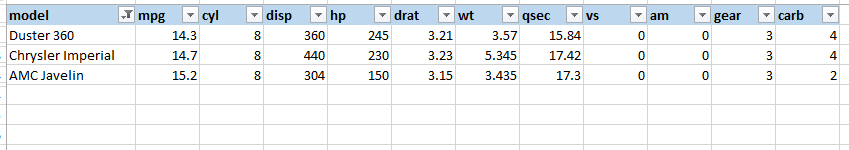


Selecting car model. Mpg, horsepower where horsepower is equal to 110

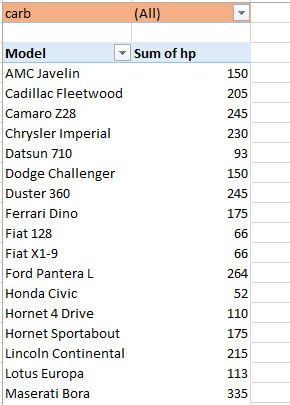


**Analysis on the Microsoft Excel**

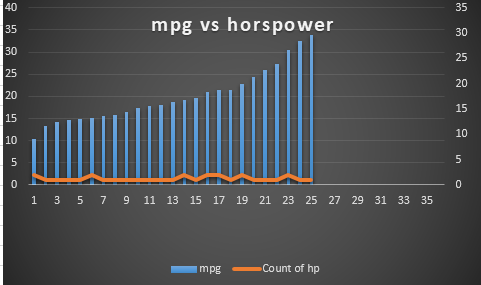
Filtering datasets for AMC, Chrysler and Duster 360. The result returns three items from the list:



Applying pivot tables on the dataset and filtering by carbs

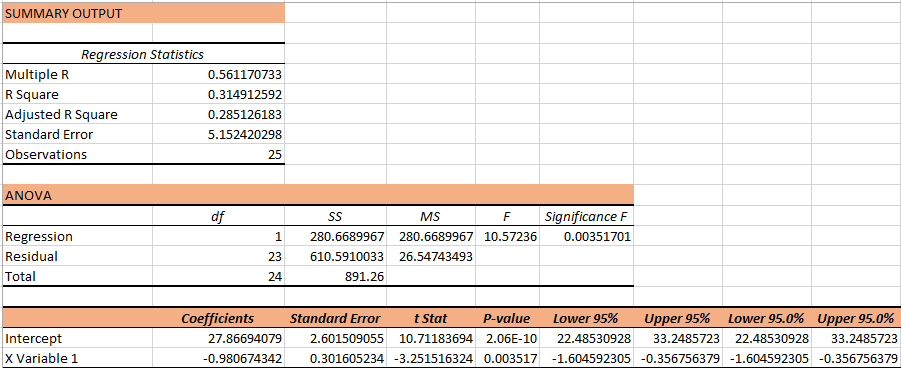


**Plotting a scatter graph and chart based on the dataset**

****

The graph shows the relationship between the miles per gallon of the car and the different horse powers, in reality though, this is not just enough to determine if there is a relationship between the two items since an increase in the mpg does not in turn show that there was an increase or decrease in the horse power. So at this point we cannot still establish any relationships between the two datasets.

Summary of the dataset 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 the relationship between the two variables, the understanding is that the two variables 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 especially where large amounts of data are involved. Data silos are large storage devise for randomly and frequently incoming data. This data can be stored and used for future use. Data can be stored ether in structured or unstructured d databases systems. Structured 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 benefit an organisation in many ways that include the ability to patch revenue leaks, to report accurately and timely and finally to reduce reporting time on data. It’s advisable for business organisation s to adopt BI systems given its ability to transform the enterprise and take them to another level of decisions support programming. However, this whole process requires cost, labour and skill managements in place.

**References**