**NAME**

**COLEGE NUMBER**

**Abstract**

The application of Bsuines ontellgence information systems havesoabn across diffenrt crss orgnisatunal functions. Comaonies and orgnaisatia re spending more more and finanacil resiurces I trying to either r acquire data, protect it or analyse the same information. Its vital thigh to cnsdre the three stages of data ecven as weh devolve deepr intonth discus susrtounfdnug the whoeile syery of BIs systems. The Bi percss takes incto account for difetstages of manbging the data prpcess:

1. Data acaquisuiion
2. Data storage
3. Data manioulatins and anaykyso
4. Data peesbtstains and firesctsing

Opuo to the inceptin of the business appliicatio prgrammig intirdiced by Watsons COBOL, the prgaeimg lnauafes gecevloped atthis era of time did no t focus fully ontosome of the varous applicatons and ppsblities poised by the BI systems, in reality thoug, the shofts was rather focused in the dffent posbbibities that originated from the sogetares that IBM was peodcug ta this time to help iragnisatina sna dcouorations o achieve a given kevel of business solution and servive sprviosn toits customers. Later on such advancements would lete rbat atake over by datascincetis who worled in te aviusous fields and sw the need to make good use of theknowldege and eixeprinece that the had gained working in this industry.Monng firward this era saw the development of dofwtare prgrams that focused on ten deveopement and prictuon of tools that fittend the industry,Some of these tools included FROTRAN DNA SCALA. These tools are stillmin use to date but their inception is a sory that can only be dreivedf orm the opuoneers and engineers who saw the need to incent smehing that indiuvauals, ciprtate and other negineers could adopt into theor dailywork prgrams.

Further, statsitsical problesm were dicvered and engoneers from the varous fileds continued to evelpe and implement other tools tools that could suoppppt other reaercgers and stsudents from this field. It is at thid posint that other sttatsicla analyais tools like Rstuio, IBM studio, Matlab, SATA and STRATA applicatuions. These tools have consistly been userd over the years bby data sceinetis stogenrate very amaaing insights to the ragtet audience and form amogitly of some of the data modelling aplatforms wose appliactions have spanned across Finace, healthcare, education and makekinting.

Another coponnet of the data process is the data storage mechanisms. In order to understand this at depnths, its ccrusicial to also undeyrsrand the length ipn whycg how organisatuon go to keep and imainat their data, keem attnyoon haowever ahs to bedwarn there tyes and eleents of cosnumbg data. Data node also known as data enetryb points act as the funnel upon which incomemnibg data iss derved amd channled into the correct data silo. Data silos are considered as the large stotres upon which fata and informatiojnmayb be stored for future need. Annexample of adata silo us I hekath care where data is stored as it comes ffomr the feld and stired in duffrentc compartements witinthe serevr. This datacan laetr be accessed by adta senginenrs and porto sof its ectsracted by data engineer, mornintired, ecsracted and trained, based on an 80/20 ule where perts of this imformation I sused tn predict isease pattrerns , tendsd and monitors the cutrrent infections and rtreatments in the healthcare industry.

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

There are three types of datasbes tahata are consired when storing information within the daatsbase environment:

* Stsructured adtasbess
* Unstrtctured databasesl
* Independt data sysemsts

Structured database systems basically stored data in organised rows and columns.Within the datasebase are tables that define the doffenrt enetieteis athat are supporseed to be cosnisdered as choice for the dferevtv data points oroginaly coennecetd to the adtasbe system. Each atabes isaodentied by a aunique table name, sometimes a schema prefix comes int before the table name. As a amstetr of sevurty though, table names are somentiems not called by tjeir specific tabeles nu=ames, but rather a cerstain tandom gibberish name is sed just in case hakjers gegt access to or asysmet becomes comporsed for whatetver thereason is vauaobe ta that time.

Each table within the databse is identgied bya nique primary key that refenecs thatpartculura row in the tabl.Another key unique identiefier in the table is the foreigh key , foreign key are used to join the different elments of the tables together without any comspsormise to the existining perevius tablec connected to it. As we shall see later in this discussion; by using the primary keys and the socodary keys associated with other tales in the coluns , its easier to ferenece the tables within the same dataset and join the tables together. Stsuctied databse systems have for long time been cosmidred as the primary source of database type amang developers and data analysts. They include examples of Pstfress DB2 and Oracle. The language of the database is considerds as strcyrured qurry language (SQL). As shallbe seen in the laters discusions, SQL is able to:

* Get datasbe information
* Read from tables
* Upadted tables
* Insert into tables
* Delet tables
* Join tables together to get information

SQL procedure can also be written withhinaserevr aopplocations tyo help grab the required staticstics on the databse that the ar erading form. This information can int turn be read an d prwsenetd on the Business Intellnce program.Unsrtcured databse systems con the other hand doe cstaly the opposite ifethat stsruterdd database ystems are supposed to do. The data in these systems do not have anay organised rows or clumns but instead the adta there in comes on heteoginty, that is =, infkation cuttignaros the databse can be referneces just with singles id and the result sset can pull almost ansy wrequierd recird 8under that artrculuar ID. Not so mamny deleveoperss and data scinetstis are up to depth and ware of this kind of database so its use in the instdry is also quite new and low. With the introduction of big data technology, this kind of databse suites the herteorgenus kind and nuatrue of thab the strcture of big data offers, in later and furher studie. Ustrcyrued data types camn accaomodate alost all types of data fraibels including cahartecst, strings, images, socuments and videoe in no trgnaised columns or tables.

The ither stabse type has been intoeudrfc by the inception of web 3.0. This kinsdo fo technology is lamost the lastsst in the industry. Its use and icnceropttion cuts across the finaccial sector. It’s the kind od databse that supports theblcokanin technology, its mahotr featre is that data is not stored on an any major appliacttion or server, but rather the data is stored in smart contac. Web 3.0 views each and every trsnaction on the etheriemu as a sinfleg eoblock of separetd and indedeopende for every block of tarsacyion identified by a particular ID. So, tarnsactioncan be interpceted by the orther util ine type of tdrsnaction is free. This makes the smart cintact running across the thereium network hard to guess and ifniltart since very high encryption algrthsm have been aolied in the netweork and there is ni single databse senntity sittin gokewhere on the ofe partculiatr network to crack. Its lso highly cnisdred that the web 3.0 is going to take over as the next phase of data handling.

Once data has been caturpred nd stored in the datasbese, the next pahes ein the deveopmentof the Buisnes applicatioj system sis tondtermie whuch tyepe of anakysies to emloy. These analyses range from genera;;ststatsitsics on measures of centera tendecinesc,specgifc counts, avaeraes and the deaviations. Then other factots to be socnisdrdd in olace incude the kind of meodels to employon the BI system. Thesemay be include:

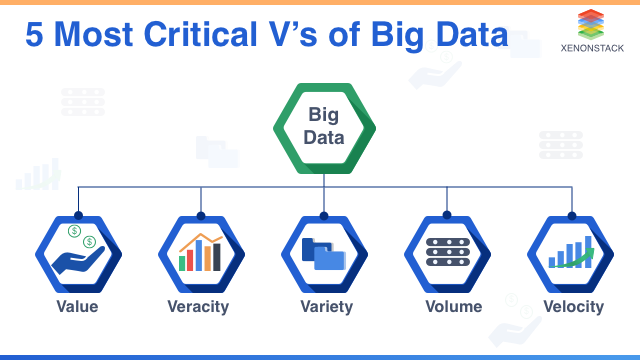
* ARIMA
* KNN
* Random forests
* Linar regessions
* Logstics regressions

The digjinifaceof each model shall be spcifvi to the kindo f ogransnaizatio probel at hand. Their efectives may also be slightly restucredto the kind and nature set being put in place. Foreactiing is another key proornent n the data analysis stage.Top maanegentb would always want to know what to expect next, whete toe expect it and from whom. What ipac acan such e[eectaton gave on the company paeomave and how hsouls the company prepare in case the market changes in any way.a quicik instance here is how the finance industry useses focrescatng to predict the maret behaious so that accaourate and informatinedd ecionsmamiauthiritues can be made onbefore trading thenext stock

Bi application systems cannot b coote without t vsiualaisations and presnatations. Actualy, th hwe reason nbehind the rigourous analyses that is dien on thesedatasts is otha the comapnaies nmanagement can get the correct interepteation of the data being prepared. Some of the tools tha have been presiously used for visulaisatyions include Micsroft exevl, PowrrBi, Rstusio SAS studio.In this stdy,the reasearxher focused on the production of oof aimteractvie charts usung the analyses on Microsft Excel and SAS studio. As shall be seen in the lst chapter of these analyses.

**Big data**

Organisational data points and nodes prodice massive information that cisfact 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.

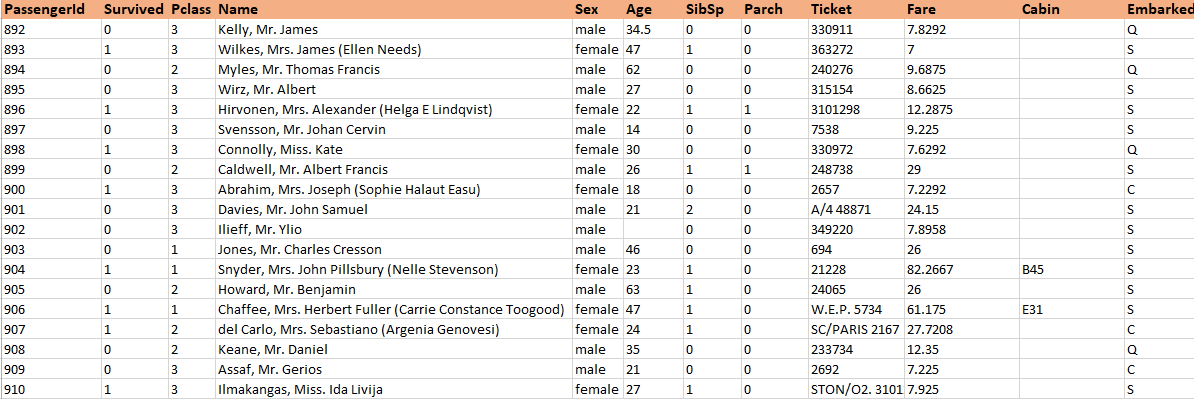
**Data analysis**

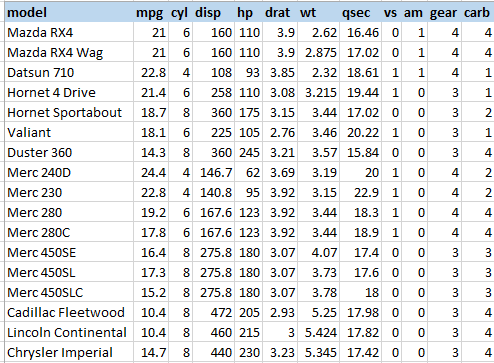
The last step in this statge is the analysis of the datatset. To do this, two sepate dataset ate conosredd:

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:





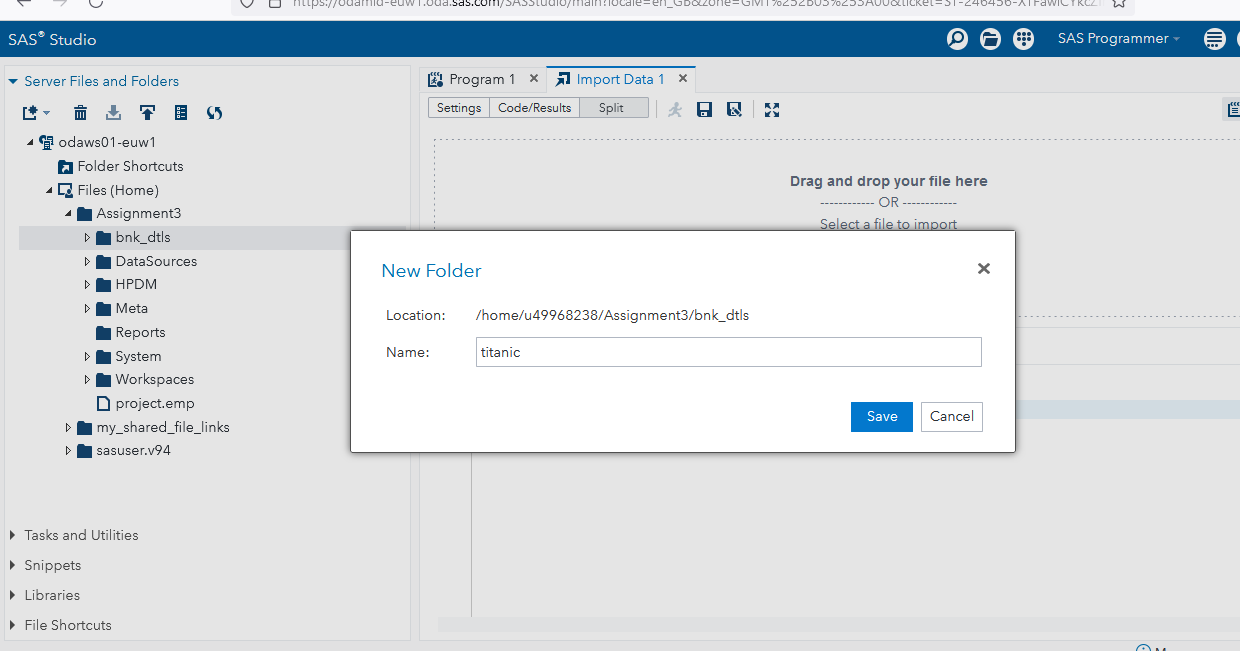
The two datasets shows vaables that will b emsuerd afains ineach other. The taaitanic dataet is acaollecton of lla thise who were onbairdc during the night of 1912 when the titanic sank with its coovpunats, most of the data here can be use nalayse the summary of e couupants ta the stime the acctident 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

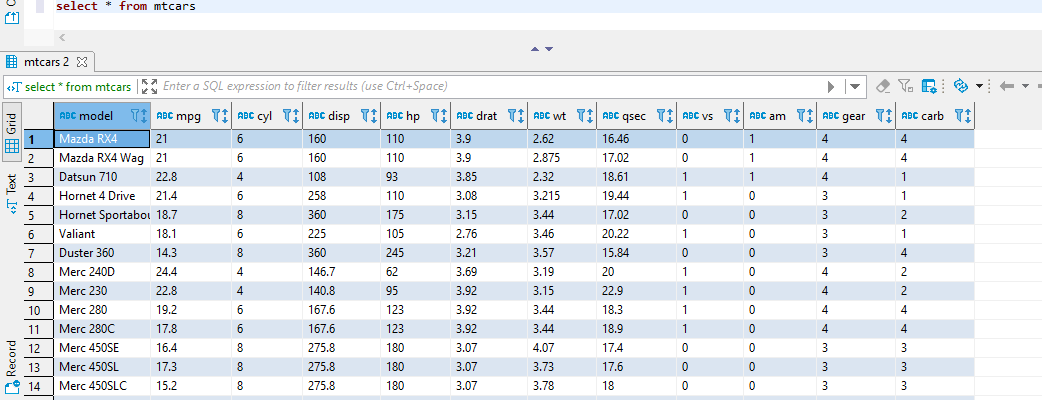
**SAS studio analyses**

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

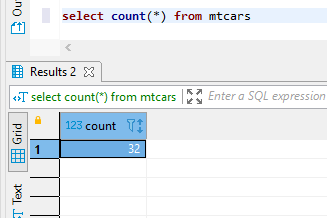


**Data analysis with SQL**

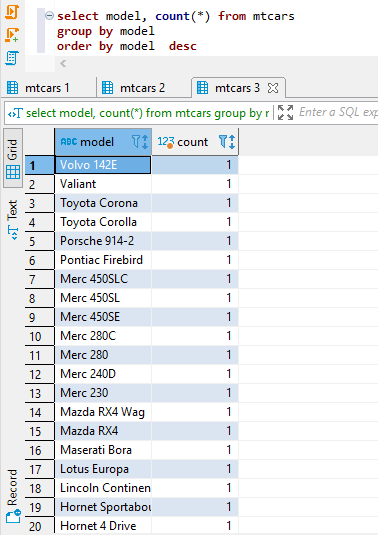
Selecting all records from the mtcars from h the dataset:



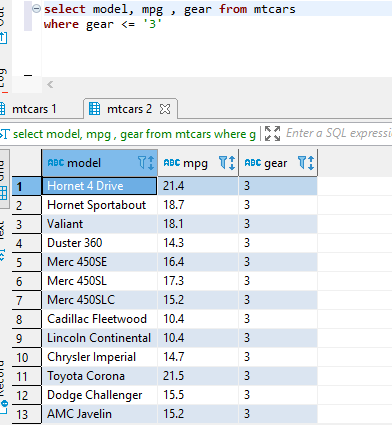
Selecting the count f records within the mtarcars tables



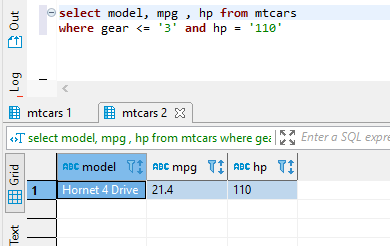
Group the cars by model



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

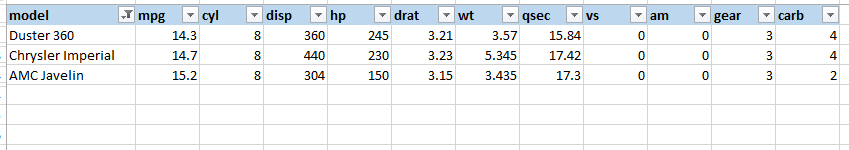


Selecting car model. Mpg, horsepower where horsepower is equal to 110

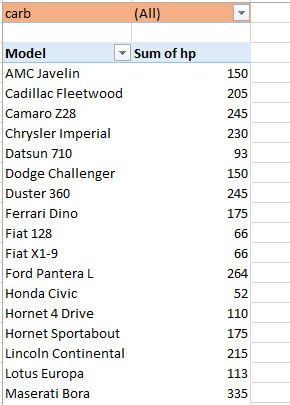


**Analysis on the Microsft Ecxcel**

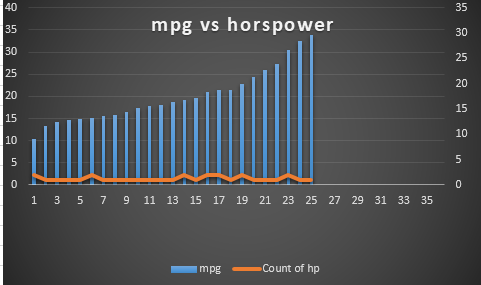
Filterimg datsasets for AMC, hcrystler and Duster 360. The relust returns three items frim the list:



Appluying pivit tables on the dataset and filetrng by carbs

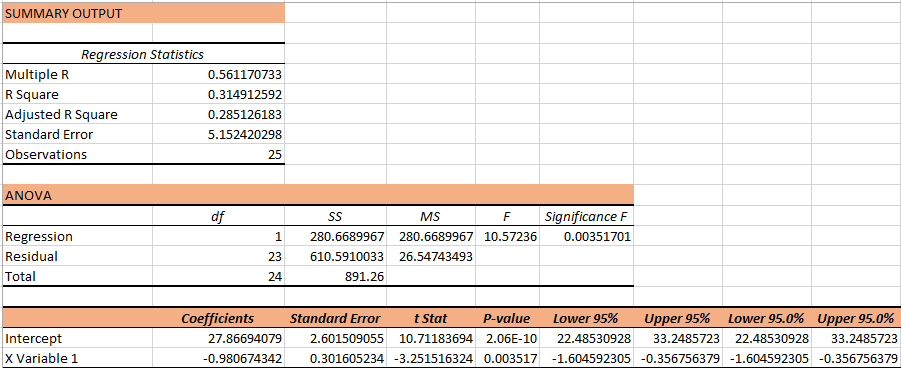


**Plotting a scatter graph and chart based on the dataset**

****

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.