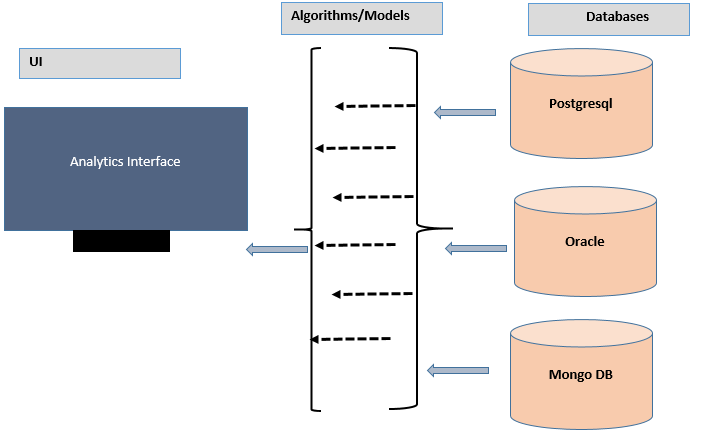
**NAME**

**COLLEGE NUMBER**

**Introduction**

**B**usiness information systems can be defined as a set of integrated set of web tools that consume data, analyse them and present the insights to an end user. Before proceeding here is the outline of the BI architecture:



The simple illustration of the BI architecture illustrates three items:

**The user interface:** Is the screen view upon which the user sees and can make reports upon which all submitted queries and reports. The user interface is the final product upon which the user/management can make decision based on the visible analytics as we shall see in the next step.

**The Algorithm and API/Modules:** The part of the interface takes the data from the data stores and applies models, algorithms and functions consume it to produce the required output. Some of the required and usual statistical models look for, counts, means, maximum, minimum, low, averages ad percentile values. Others include limits, top counts and latest records. Further after analysis, the presenting of this data in form of graphs, charts and scatter plots. Advanced methods in the process of integrating with the BI also includes predictive analytics where certain statistical c models like LDA, QDA, KNN, Arima and Linear regression methods are applied. This is so since organisations would always want to analyse and know the future of the data they are holding. For instance, if its sales, the department would like to predict when and how they are going to be selling in the near future and how this matters in their current business processes as an organisation.

**Databases:**

There are two types of databases that exist:

* Structured
* Unstructured Databases

Structured databases store data in in organised rows and columns in what we define as tables. The items in these tables are identified by primary keys as and foreign keys. Let’s look at a sample customer table below:

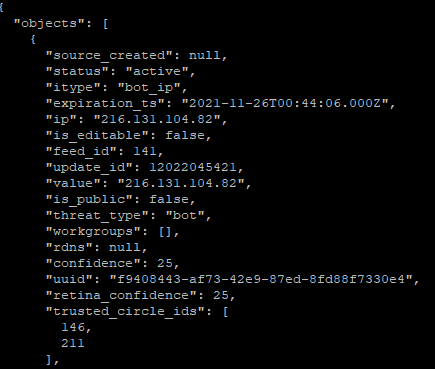
|  |  |  |
| --- | --- | --- |
|  |  |  |
| id | name |  |
| A54649 | Jeremy lans |  |
| REF763 | Rebbeca Ohms |  |
| YYE735 | Annet kieler |  |
|  |  |  |
|  |  |  |

The structured database above shows the unique records for a customer details that are uniquely identified by a key. On the other than, we can have another table for sales that show the sales details for the customer above.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sales id | FOR KEY | Prod\_name | amount | Sale\_date |
| 9655 | A54649 | Jeremy lans | 6457 | 2021-11-08 |
| 5444 | REF763 | Rebbeca Ohms | 896974 | 2021-10-12 |
| 7864 | YYE735 | Annet kieler | 745338 | 2021-09-05 |
|  |  |  |  |  |
|  |  |  |  |  |

The two tables identified above are linked by their respective primary keys, a primary key could be a foreign key in another table. This element of identifying records across structured datasets in the BI is essential because it will be crucial in joining tables and sharing the resultant outputs on the user interface during analytics.

**Unstructured databases**: on the other hand not organise data in any specific tables, columns or rows. Instead the data is stored in unstructured rows and columns and can take any data type available in as source. Below is an example of an unstructured data array that takes any data type.



**Literature Review**

In an article written by Negash and Gray (2008), prior to the year 2008, the term Business Intelligence majorly constituted cross functional analyses of dataset cutting across the different department as a decision support tool. The terms data warehouse and data mart were originally confined to the sources of large transactional data upon which analyses would be drawn for management.

These data sources provided real time analyses, queries and forecasting of events. According to Lee & Park, there exists a very thin line between measurement of data and strategic planning in an organization. That systematic data analysis and sound application of the web 2.0 there is need for companies to significantly employ the use of data analysis tools in order to measure data performance, predict and even leader others very well. This study finalized that the efficient use of business information systems, could also be used whenever the need arose to access knowledge managing information systems.

Moss and Atre (2003) state that that the benefits of having a Business intelligence applications exceed its limitations by far. In however much it’s not easier to quantify the benefits of such an application, the benefits have always been narrowed down and the user can always draw from the benefits financially and be able to measure business impacts in quantifiable numbers. For instance, of a company was losing millions because it was not able to tire done sales of a given product within a region to certain customers, then it’s easier to manage this with a BI application, where no data is left unanalysed and put on the table for decision making. Something that so also notable about BI systems is the element of the third eye.

Before BI systems were invented in the early 1980s, running SQL queries and commands on single server databases, would result in errors in numbers and statistics that was not easy for anyone else to notice except the data engineers or peer data developer. This problem was however solved by having efficient queries run on a single server and code optimization check for failures; this originated from the fact that managers and fellow BI users could observe numbers on the BI screens and still be able to critique them, bringing the elements of the third eye.

Loshin (2012) writes on some of the limitations and shortcomings of a typical BI application and in his papers states that, some organizations have experienced resistance to some of these applications, making it difficult for the organisation to adopt such systems, on the pretext t that the application shall render them jobless and unable to don their jobs effectively as required. However, focusing in a the financial sectors of the organization, companies have been more focused than ever before or make sense form their numbers. Managers need to make sense from sales numbers, to know who is buying where are they buying from, how frequently do they buy, any related products they are buying and at what cots do they buy, when it’s the period of the day, week or month that they purchase. Its worthwhile noting that such question scan be answered In 4 -6 lines of a code running inside the server, the same reason why again the BI is a must for the banks, SMES and other financial departments.

**Part B**

**Data analytics and reporting**

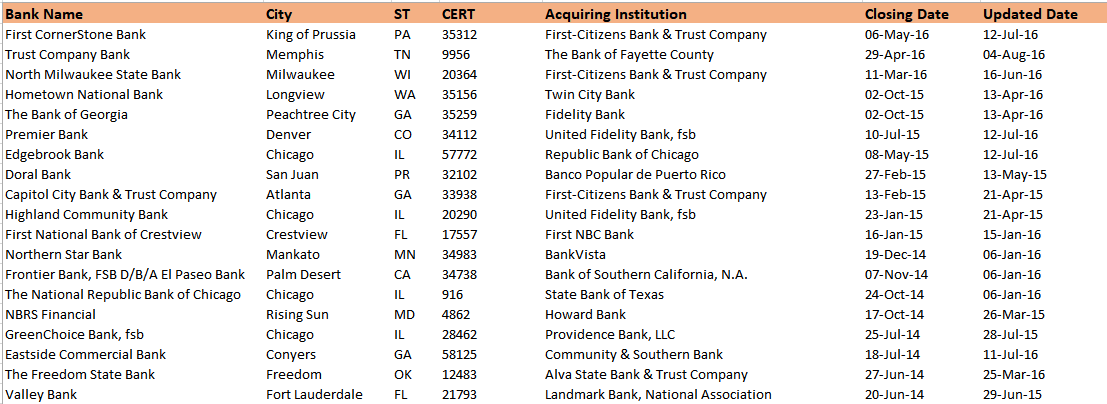
**Data description**

For this particular analyst, the researcher shall employ the financial datasets obtained at data.world <https://data.world/drew-atx/failedbankdata> this is a dataset containing a list of banks with failed banks and the companies that acquired them and when such an acquisition was done, the opening and closing of the respective banks.

Moreover, the dataset consists of the bank name, the bank state location, the state short code, the acquiring institution, the certificate of the bank, the closing date and the updated date.

**Data wrangling**

The next step in this dataset is to try and clean it by removing any spaces, unrequired characters and values that are in misplaced columns of the datasets. Occasionally, data wrangling would involve renaming columns to match the required column attributes. A quick sneak into the dataset revealed the following:



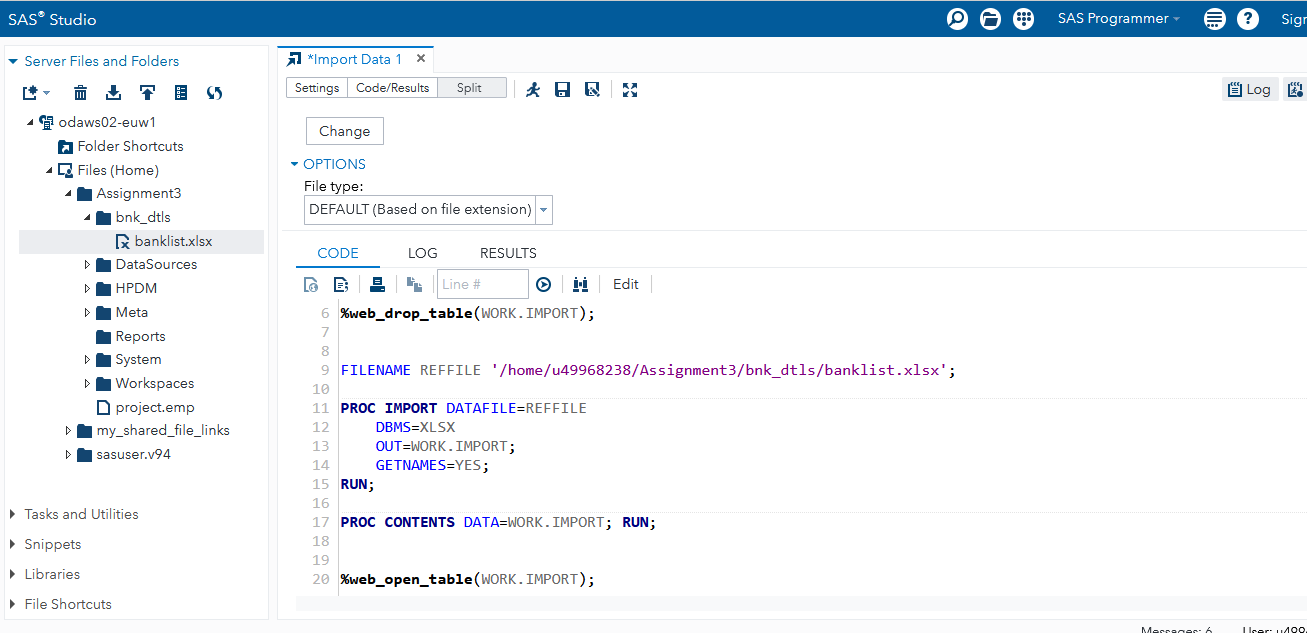
**Data analysis on SAS**

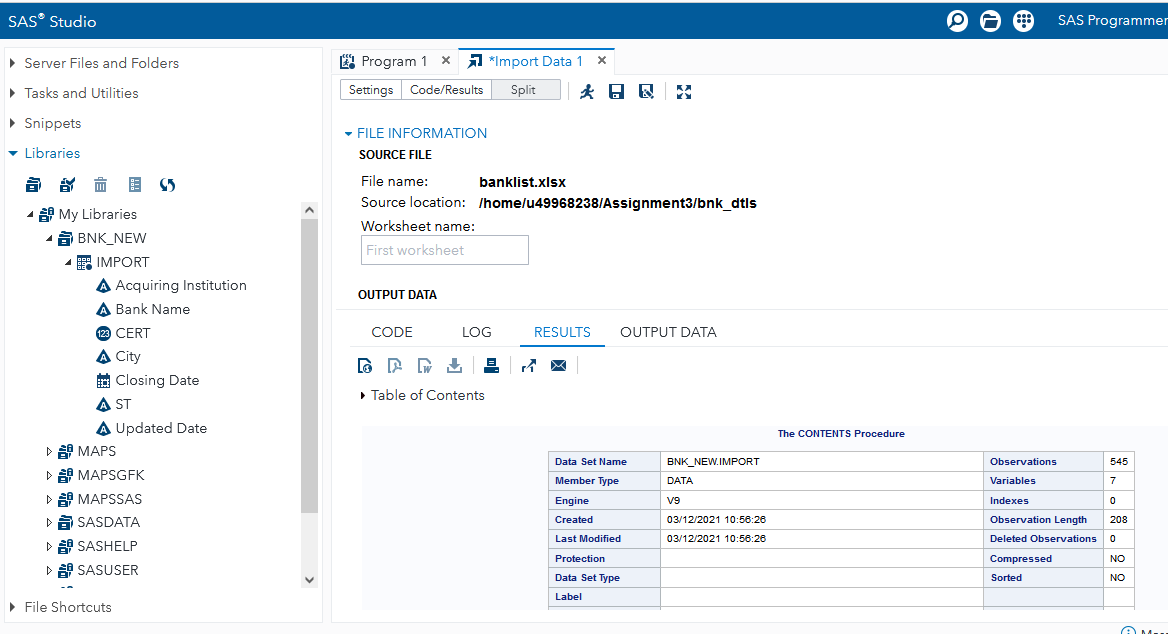
Prior to proceeding with data analysis on SAS objectives based on the above dataset shall be established:

* To determine the frequency of occurrence of the failed banks
* To determine the trend of failing banks
* To determine which cities had the most failures
* To determine duration upon which most banks were failing
* To determine If there is a correlation between the failing banks and their years
* To forecast any potential failures predisposed to these banks in the near future
* To predict which state is most likely to have a failing bank in the near future

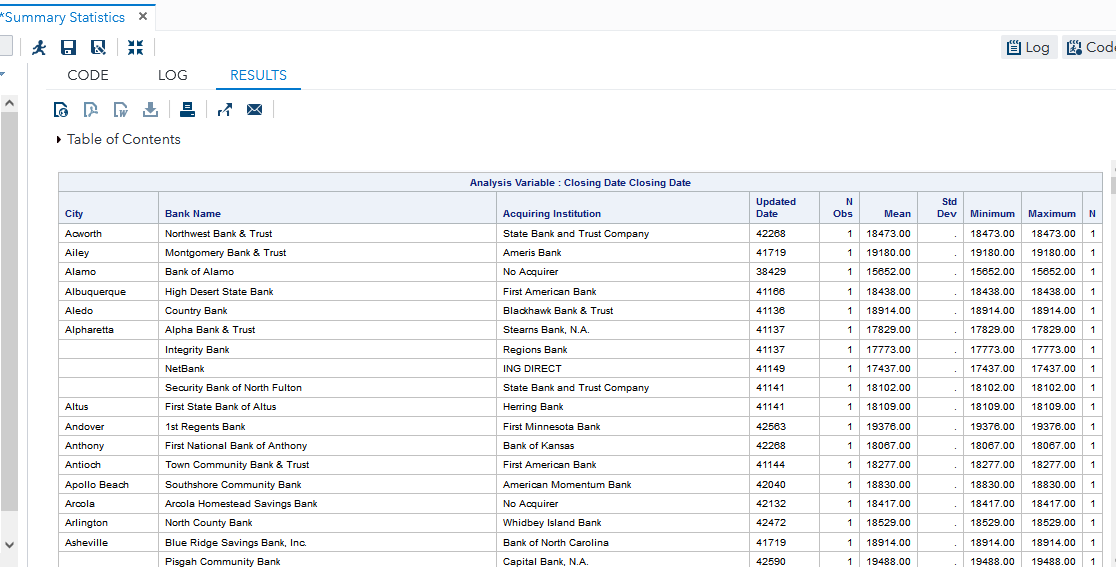
**Analysis on SAS studio**

The first step is to load the dataset into the SAS studio importing it as Excel and then creating a server directory for the dataset. In this project, the researcher created a directory called the BNK details. From which the dataset is then convened into the .bat, a format that SAS studio can read and understand. This has been done as per below;

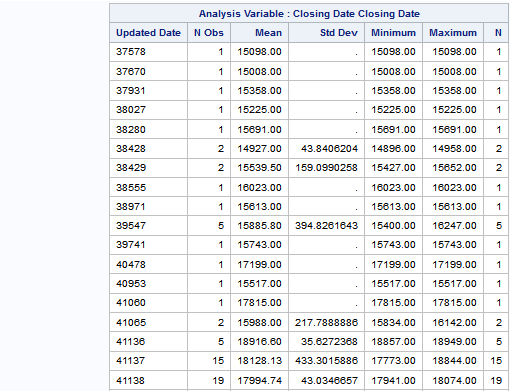




Once this has been done, then we can take a quick peek into our dataset and the Studio environment reveals the following dataset variables associated with the failed bank details. From this pane, we can choose which variables that we want to measure alongside each other, depending on the expected output of the data we want.



**Summary statistics**



Usually a summary statistics will give the overall ramp up of the variables in measurement. This measurement is based on the measures of central tendencies such as mean, median, average. This helps the researcher to know exactly where the concentration of his/her data is, what averages are cutting through the variables and what to consider as the minimum or maximum values in the dataset. The above summary is based on the means of opening and closing dates of these banks. Their standard deviation, which basically describes how far the date ranges are from the expected means.

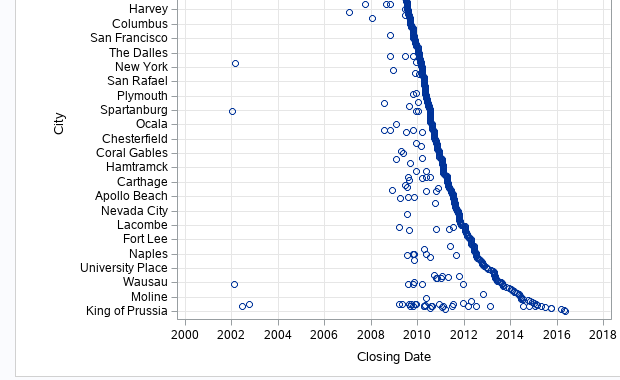
Prior to moving in with correlational analysis or logistics regression models, these summaries shall give a data scientist what is most likely expected of the models that they are about to predict so that the necessary hypotheses can be made and evaluated. After the models have been actioned, another comparison check can be made to identify if the hypotheses met the aims and objectives of the researcher and how they relate to the summary statistics derived above. This kind of exploratory data analysis is bests fit for linearly regressing data and gives much deeper insights to the researcher. For instance, in the dataset described above, a linear regression model can be applied to data that’s changing overtime. Suppose we consider the cities that were having their banks being closed overtime, we could most likely come up with a trend and then by sung this trend, we come up with the forecasts similarly related to this information. Since trends are based on timelines, one hypotheses the researcher could make based on the failed banks dataset is that;

* There are higher taxes and legal requirements in certain cities that make banks not to perfume well in search cities
* That there are social crimes such as burglary and robberies in particular cities that make the banks to close so often ,
* That the earning power and the gross income of individuals and cooperation’s in some of these cities is so low that the individuals cannot afford to save nor do business with these banks hence their slow o early death.

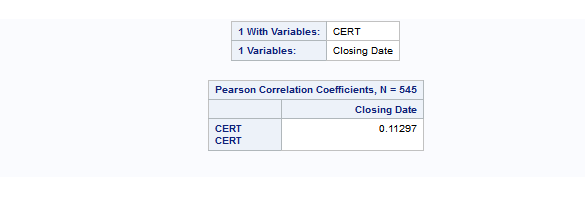
**Exploration analysis**

From the analysis of the scatter plot drawn below, we can observe the trend in the failure of these banks across the cities. The scatter shows the distributions of the banks across the years of 2016 to 2020, with most concentration of the trend between 2010 to 2012, the latest trend showing that the latest bank being after is King of Prussia, this information is important to the researcher to help them understand which banks are most likely to fail suppose investor was to pick them up based on their history and activity. The main dataset can be modelled further to predict which bank is most likely to fail next.

This could be done by using the variables on bank name and closing date, upon which the dataset is then split into the training set and the testing set. In order to most accuracy, 80% of the dataset can be put in for training, whereas the rest 20% can be put as test data. Some of the models that can be then used to predict the next failing bank include random forest, ARIMA and the Naive Bayes for classifications.



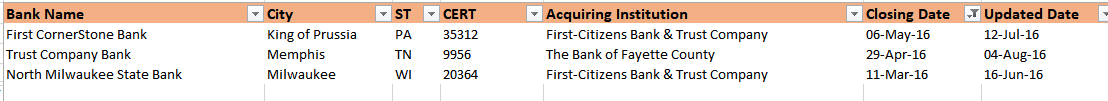
Also from the analysis below, it can be observed that the Pearson coefficient resulting from this analysis gives a correlation value of 0.11. Even though this value is not strong enough, it somewhat shows a small correlation between the closing date and the closing date, the result of which when plotted alongside each other will indicate an almost structured kind of population distribution among the variables in question. Inferentially interpreting this means that as the opening date of the bank increase, the closing date increases as well, but not to a large pedigree, meaning that even though the banks are closing at a given rate, this is not directly in line with their opening date, some may close after a quarter of an year, or semi-annually or after 2 years, so the consistency is not that big.

****

**Microsoft Excel data analysis**

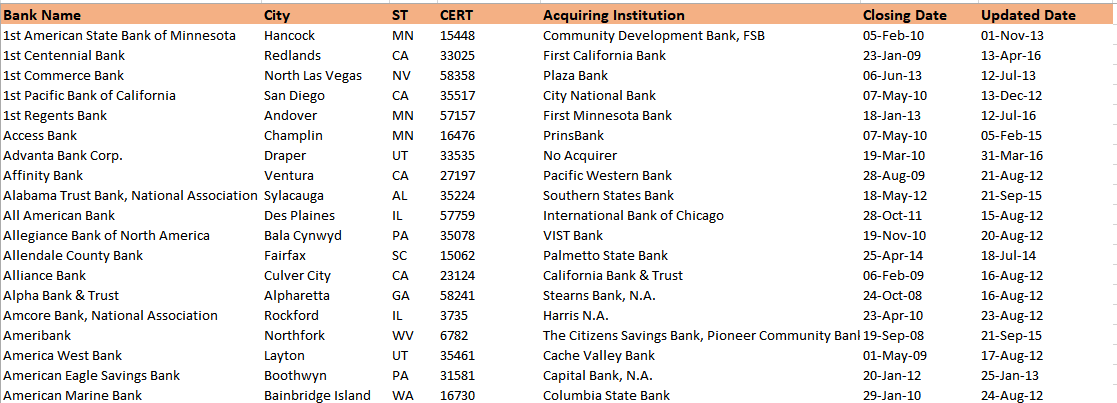
The next step in this dataset is to pass the dataset into an analysis through Microsoft excel dataset.

**Applying filters**

****

In the above exercise, filter was done on the dataset to try and establish a list of all the banks that were closed in the year 2016. This suiting information revealed a total of 3bnks that were closed in the same year. The filter command was applied on the slowing date column. This is a quick way of getting the count of all banks that closed in the same year.

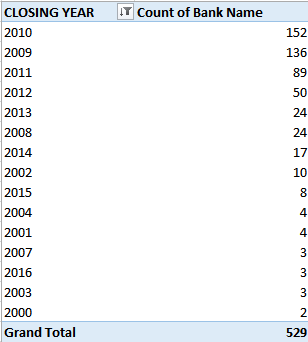
**Sorting data**

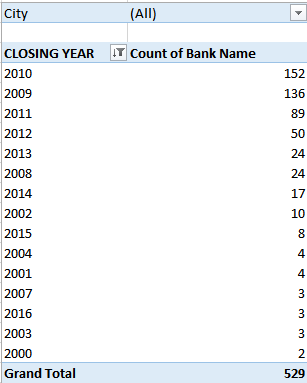
****

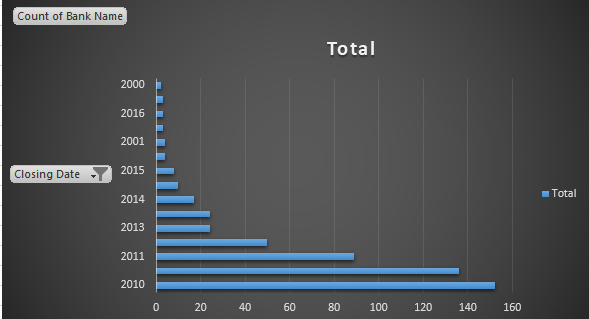
From the dataset of the failed banks lists, the researcher sought to establish the order of the banks by sorting them in Alphabetical order from A to Z. The list picked the first top 10 banks.

**Applying pivots tables and pivot charts**

The researcher also sought to find a quick summary count if banks and the years in which they were closing, the list was then ordered in ascending deciding order as shown below, then a graphical pivot chart derived from the same dataset. In the second part of the pivot, a filter by city is applied to the dataset where one can filter the count of these failed banks by the cities from which they originate,



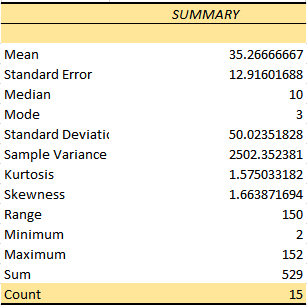




The analysis from these variables indicate that number of banks that failed in the United states hit a record high in the year 2010 and 2011.With the lowest year being recorded in the year 2000.

**Descriptive statistics and summary**

Further a descriptive statistics established on the above dataset revealed the following;



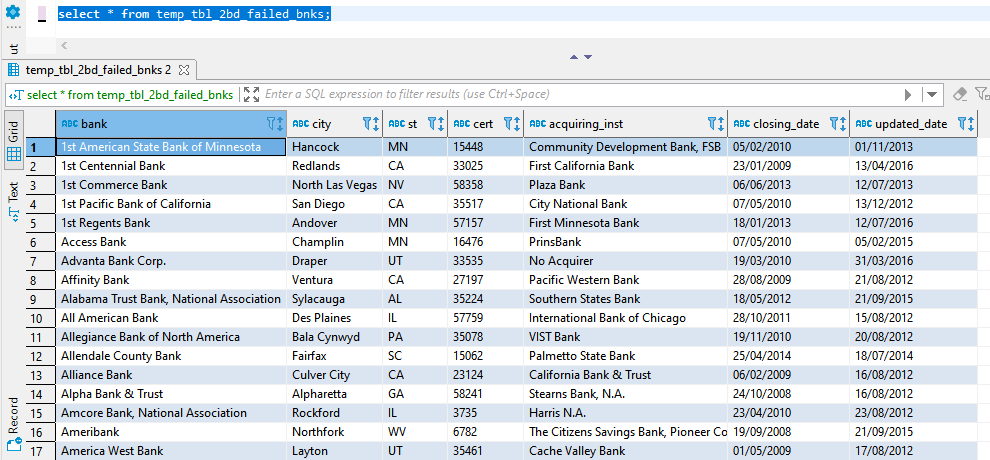
The result and summary of the count of banks from the summary statistics show that the skewed percentage is 1.66 which is a positive value of +1.66 indicating a positive correlation between the years and the number of numbers that were failing, this can also be interpreted to mean that in future, the numbers failing would potentially increase based on the provided summary, and as a result, this can also be used as a primary basis of for forecasting results in the near future.

**Analysis with structured Query Language**

The data is loaded on to and integrated database, which in this case is the Beaver, (open source) sql data management tool.

**Selecting all from the database;**

select \* from temp\_tbl\_2bd\_failed\_bnks;

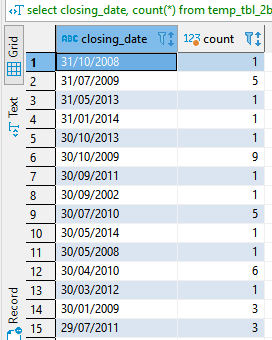
****

**Using Queries to group the failed banks into years**

select closing\_date, count(\*) from temp\_tbl\_2bd\_failed\_bnks

group by closing\_date

order by closing\_date desc

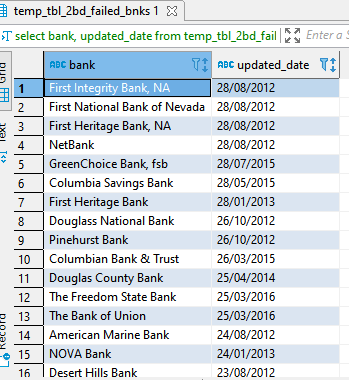
****

**Selecting banks and their updated dates after January first 2012**

select bank, updated\_date from temp\_tbl\_2bd\_failed\_bnks

where closing\_date >='2012-01-01'

order by updated\_date desc

****

**Summary, conclusion and recommendation**

Overall, it can be deduced that BI systems are very efficient in handling and managing large corporate data. Decisions support systems having been put in place by cannot effectively and wholesomely save managerial problems. On the other hand, intelligent support systems (ISS), have gone ahead to acquire and naturalize big data, crunch it and transform it into forms that can be effectively used in managing business problems.

Consequently, data is being produced daily and in very large amounts that corporate and organisation needs to keep track of, however, the available data points such social media messages, posts, comments and streams, blogs, feedback points and transitional logs need to be captured into a central unstructured database such as Mongo or Firebase, where they can be consumed by a BI application. The BI application will read the data, apply algorithms to it and then visualise the output to the user.

The result of the visualisations can consist of graphs, charts, counts, averages, means and medians. Further to this, BI have also been developed to help train models, by accepting a train set and a testing set in a ratio of 80:20. Some of the models that can be applied on the BI include Logistic regression, linear models, and KNN and Random forests. The results from this analysis can be used by the management to make effective decision on the required organisational processes and business activities, which in the short term and long term shall help the company to:

* Reduce overhead costs incurred during normal business activities
* Reduce risks by forecasting in the expected activities by modelling data and predicting it
* Minimise time required to make critical business decisions since the BI is real-time
* Patch revenue leakages that occur due to untapped data

The Bi system is hence worthwhile and should be accepted and adopted by most organisations since its significance overpower its limitations and the company can only get the best.

**REFERENCES**

Negash, S., & Gray, P. (2008). Business intelligence. In *Handbook on decision support systems 2* (pp. 175-193). Springer, Berlin, Heidelberg.

Lee, J. H., & Park, S. C. (2005). Intelligent profitable customers segmentation system based on business intelligence tools. *Expert* systems *with applications*, *29*(1), 145-152.

Moss, L. T., & Atre, S. (2003). *Business intelligence roadmap: the complete project lifecycle for decision-support applications*. Addison-Wesley Professional.

Loshin, D. (2012). *Business intelligence: the savvy manager's guide*. Newnes.