Capstone Project—Business Intelligence Solution for U.S. Organization

Alex N Huskey

MIS 480-1 – Capstone - Business Analytics and Information Systems

Colorado State University – Global Campus

Dr. Mamdouh Babi

August 6, 2023

Capstone Project—Business Intelligence Solution for U.S. Organization

As a Capstone for this program, I will compile a thorough proposal for a new Business Intelligence(BI) solution for a US based organization. In this scenario I have been hired as a consultant to present a BI solution for a U.S. Organization. The organization is interested in using business intelligence to help with strategic decision making and has asked me to demonstrate how BI tools can improve data analysis for issues or new theories. I will provide a written report as well as an oral presentation and recording to demonstrate how I would present my proposed BI solution to stakeholders of the U.S. Organization.

## Information on U.S. Organization

For this portfolio I am electing to look at the United States Department of Agriculture (USDA). The USDA acts as a governing body for the United States agriculture industry providing oversight and insight to United States based farmers. Per the USDA website:

We provide leadership on food, agriculture, natural resources, rural development, nutrition, and related issues based on public policy, the best available science, and effective management.We have a vision to provide economic opportunity through innovation, helping rural America to thrive; to promote agriculture production that better nourishes Americans while also helping feed others throughout the world; and to preserve our Nation's natural resources through conservation, restored forests, improved watersheds, and healthy private working lands (USDA, n.d.)

I chose this organization because I find the agriculture industry very interesting and has a large amount of very relevant applications and insights associated with it. The agriculture industry also has very thorough data, and has long been an essential section of the economy. I live and work in a small farming town which is heavily supported by the USDA and didn’t want to pass up on the opportunity to do some potentially relevant research and exercises. With the current state of the agriculture industry within the state there are plenty of future job opportunities available and relevant experience with the data would be beneficial long term.

## Problem Statement

Problem: Sales price of aquaculture fish can be volatile and it’s difficult to determine which product will continue to be the most profitable on a yearly basis.

Solution: Utilizing the BI tool R and RStudio for visual analysis and regression formulas to predict future profits from aquaculture to determine the most profitable aquaculture to focus on.

I believe with this dataset there will be very powerful statistical opportunities to build regression models. I plan to build various models to simulate future pricing to get an advanced look at which categories and regions would be the best to import/export the various aquaculture available. Linear regression models provide the most direct results for the solution. After the data has been split and broken down into usable sets it will be feasible to build the regression models for future predictions.

## Details on the Dataset

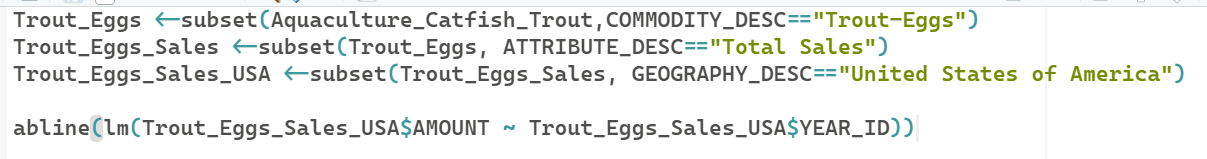
The dataset was found using the MIS480PublicDatasets.pdf, which led me to the website: <https://data.world/agriculture/aquaculture-data> . Which contains various other datasets but which I deemed to be too large to be able to compute information within a reasonable amount of time. The dataset currently chosen contains 20,000 records which is approximately 20 times larger than other datasets I’ve used in the past. The dataset I selected contains statistics on domestically grown catfish, trout and other fish including their import/export data. The data is currently in CSV format containing monthly prices from 1986 all the way up to early 2013. The dataset chosen contains over 20,000 records. The dataset contains a wide array of categories from types of fish, region the price is valid, how the fish is valued and multiple different categories of fish. Some categories are cents/pound, dollars per pound, and various bulk prices. I mainly chose this dataset because while still containing 20,000 records it’s significantly smaller than the other datasets available from the USDA and contains easily manipulated data properties for statistical analysis.

## BI Tool

For this portfolio I will be using the language R within the environment of RStudio. This dataset contains a significant quantity of historical raw data. In its current state it is excessively large and difficult to glean useful data from. With additional subsetting this data set becomes extremely tailored to work within R because it is a significant amount of numerical data, which R is extremely useful for powerful statistical problems. RStudio also has very high quality data visualization capabilities for a heavily statistically focused BI tool. The intended use of this data and its problem statement allow for limited visualization where charts and reports will suffice. If a BI tool more tailored to daily monitoring of data and progress might require a BI tool with dashboard capabilities. But a regression analysis and visualization is a perfect scenario to implement and utilize RStudio.

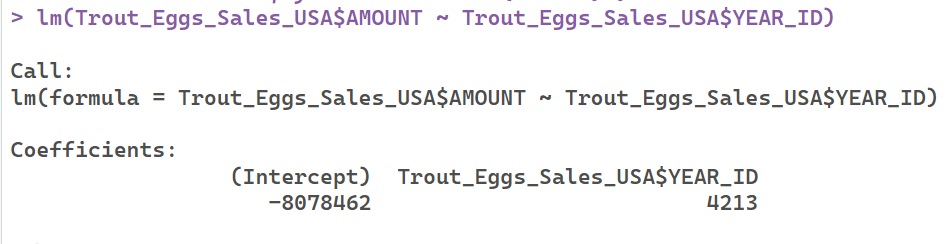
## Code of Functions Used

To start the usage of the BI tool, the data needs to be transferred into the BI tool. RStudio has an easier method to read in data and allows for the usage of CSV’s and spreadsheets. RStudio also stores the data and other items or functions which have been created to check and see what items users can work with. Reading the data in from CSV or spreadsheet is done with a simple “Import Dataset” button. Once data has been read, there are two main commands to be used to establish the regression of a specific data set.



*Figure 1.* RStudio commands to subset data and establish the regression

The top 3 commands filter down the larger CSV to only look for Trout Eggs, total sales, and in the United states. They are identified by the Attribute Desc filtering down to only the exact phrase within the column. Finally the lm() function builds the regression given the y-value initially followed by the x-value (y~x). In front of that the abline() command adds the function to plots. These two commands used in tandem create the regression for a specified set of data and plot the result. Without visualizing the plot the lm() allows for the linear regression to be built and provides the results in case a plot is not the proper analysis technique at the time.



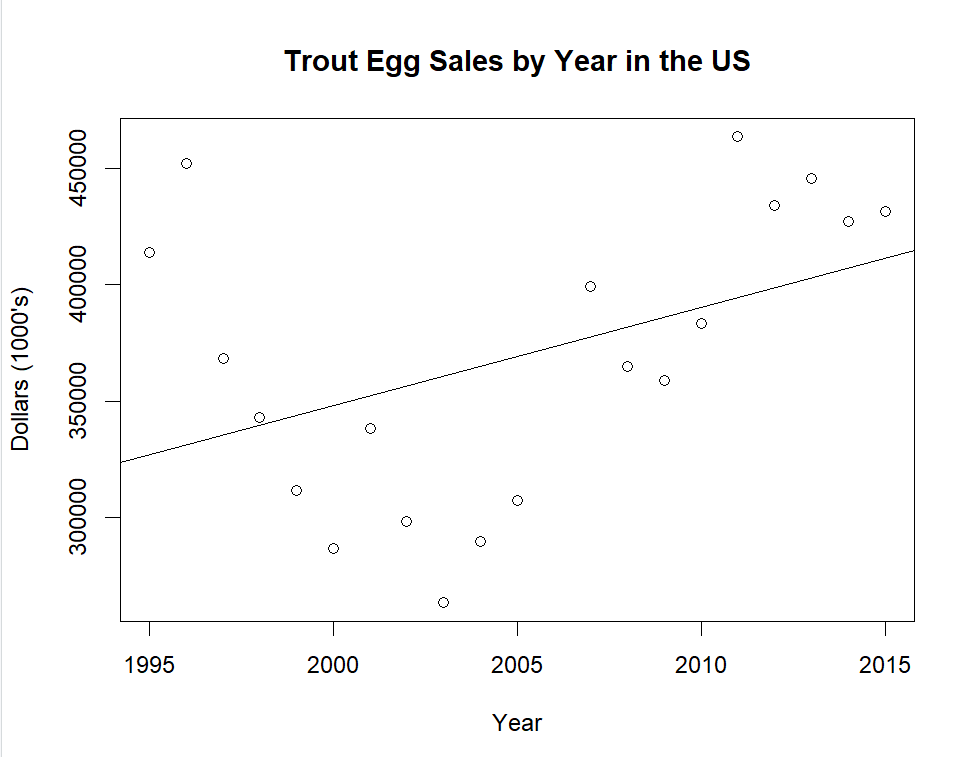
*Figure 2.* The lm() function results in linear regressions terms

In this figure, the lm() call returns the falling coefficients. These behave as traditional y=mx+b, in which the Intercept remains the same and the other coefficient represents the m-value. To use this result the intended predicted value, y, can be calculated by a year the prediction is requested for, x. If trying to calculate the sales for year 2020 the equation will be:

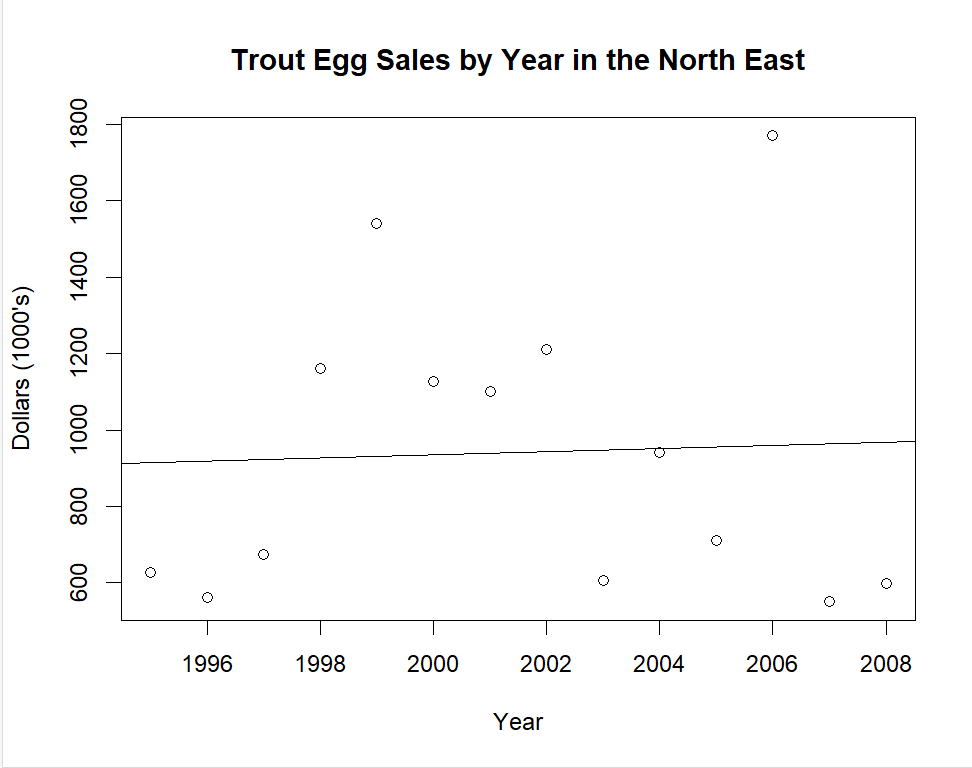
y=4,213(2020) - (8,078,462), resulting in y=431,798. For a total sales prediction in 2020 of $431,798.

## Data Visualizations Results and Analysis

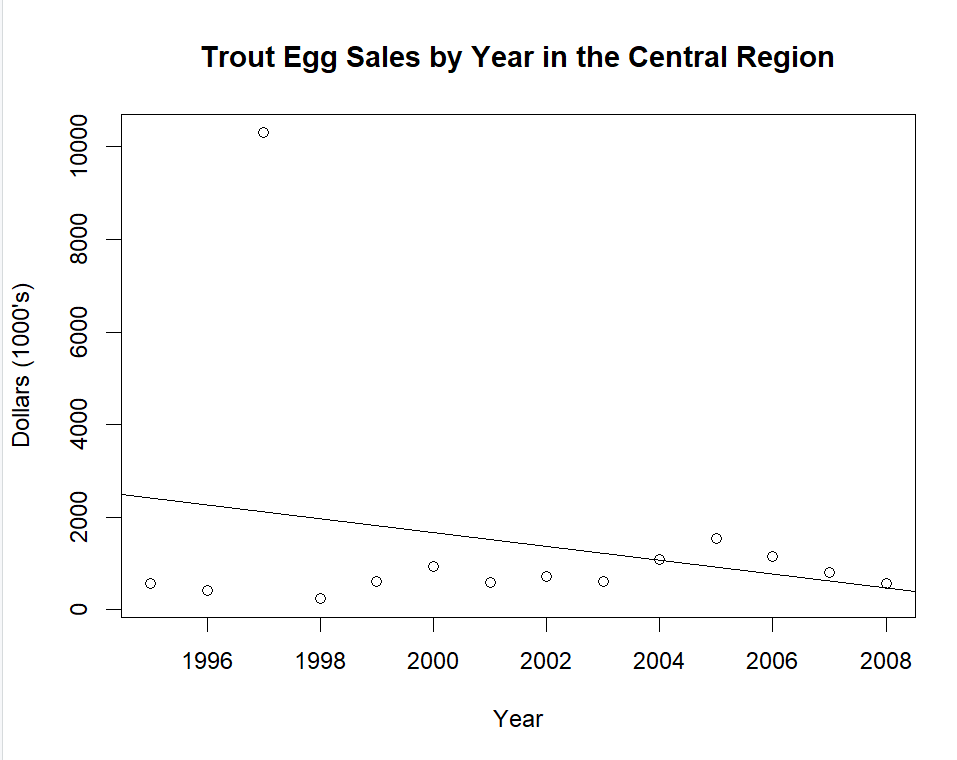
As of now, the main objective of this assignment is to build a regression model within R to simulate and predict future pricing. I think that will be doable for various categories of fish and regions. Likely utilizing histograms or other box plots for historical data as well as providing new plots with trends to more easily visualize how the data is generating a predicted result.



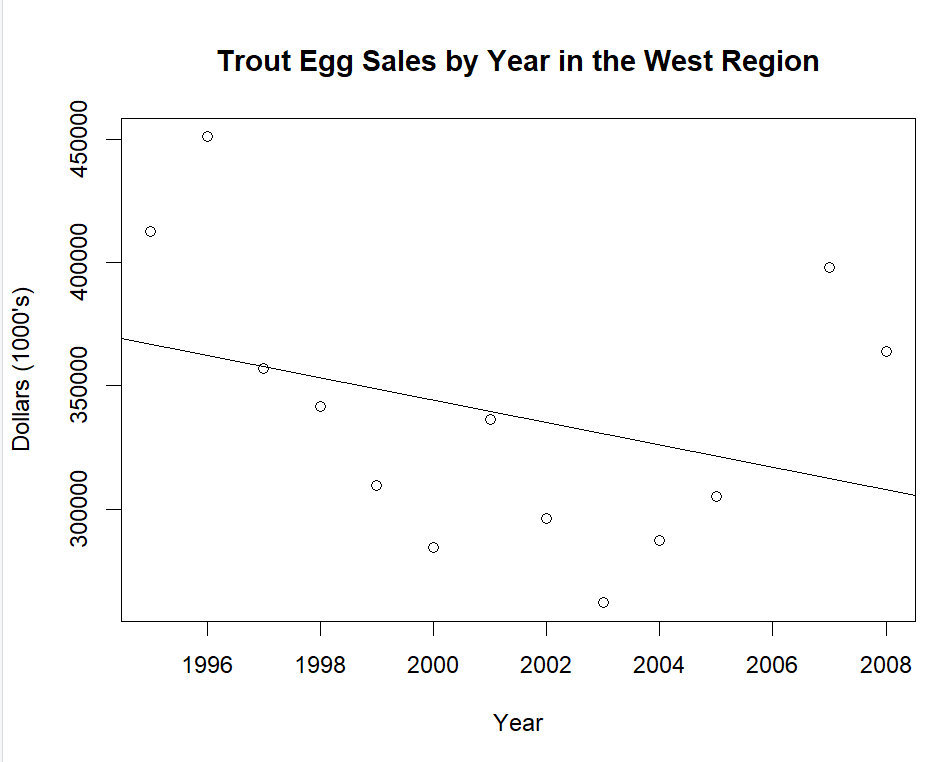
*Figure 3.* Total Value in thousands of Trout Egg sales in the US by year from 1995-2015, including regression analysis line



*Figure 4.* Total Value in thousands of Trout Egg sales in the NorthEast Region by year from 1995-2008, including regression analysis line



*Figure 5.* Total Value in thousands of Trout Egg sales in the Central Region by year from 1995-2008, including regression analysis line



*Figure 6.* Total Value in thousands of Trout Egg sales in the West Region by year from 1995-2008, including regression analysis line

## Outcomes and Benefits of Utilizing Business Intelligence

Utilizing business intelligence can be a difficult task to undergo for an organization. There are several challenges which must be considered such as technical limitations, financial limitations, and personnel/knowledge limitations. But the benefits and opportunities associated with utilizing business intelligence create ample opportunity to overcome and outweigh the challenges and risks of implementation. Business intelligence allows organizations to identify trends within the organization and outside of the organization to improve overall efficiency. Business intelligence also imparts data to ease decision making processes. Businesses face regular tough decisions on how to act and respond to certain market trends and situations. Poor decision making can be catastrophic to the long term health of an organization. However, making frequent correct decisions offers great benefit to an organization as it adapts and evolves. Reflecting and developing a process for business intelligence is a great practice for an organization for its long term success as well. Many different projects and processes need to be implemented and a difficult one like implementing business intelligence is a strong exercise for implementation. This implementation is a good exercise for developing and refining project management techniques for current and future projects. Utilizing and implementing business intelligence carries many considerable benefits for an organization

## Conclusion

While implementing R and Rstudio as a business intelligence tool can be costly and complicated, it provides many great benefits to an organization. RStudio is currently a free software with a strong user support base that allows for companies to implement it at an extremely affordable cost compared to other business intelligence solutions. R also has strong statistically relevant tools that allow companies to project revenues using regression analysis and creating easy to read visual tools. A lot of the challenges associated with R come with difficulty to learn the language. While it is a more difficult language to learn compared to SQL, SAS, or Tableau, it comes in far cheaper than the three of those options. R also has a wide variety of online resources available to help users learn and begin growing their skills. Personally, I do not find it much more difficult than writing excel formulas.

References

*About the U.S. Department of Agriculture*. USDA. (n.d.). https://www.usda.gov/our-agency/about-usda#:~:text=We%20provide%20leadership%20on%20food,available%20science%2C%20and%20effective%20management.

Bansal, A., & Srivastava, S. (2018). Tools Used in Data Analysis: A Comparative Study. *International Journal of Recent Research Aspects*, 5(1), 15–18.

Dennis, A., Wixom, B. H., & Roth, R. M. (n.d.). Systems Analysis and Design. Retrieved from [https://platform.virdocs.com/read/2045833/16/#/4/2/4[c07-sec-0001]/2](https://platform.virdocs.com/read/2045833/16/#/4/2/4%5Bc07-sec-0001%5D/2)

Sharma, A. (2020, June 12). *Why you need agile business intelligence and data analytics - dzone*. dzone.com. https://dzone.com/articles/what-is-agile-business-intelligence-and-data-analy

Venables, W. N., & Smith, D. M. (2023, April 21). An Introduction to R. Comprehensive R Archive Network (CRAN).