David Sacco

Fairfield University

Capstone Project

Business Analytics Capstone Project Proposal

22 May 2019

**Overview**

I would like to create a database that would house public mortgage data from Fannie Mae and Freddie Mac. If this is successful, I will then create an R Shiny application that will sit on top of the database, that will then be given to my professional advisor, Robert Sacco, who will then determine if there is a use for it in his line of work. Robert Sacco is an Executive Director at JP Morgan Chase for the Mortgage Banking division. He will be my main contact for any questions related to the dataset that I will be using.

The dataset that I intend to use has 56,000 rows and almost 80 columns. Professionals, such as Robert, must sift through excel sheets that have enormous amounts of data, and this project may help with speeding up this manual spreadsheet analysis with a few clicks on an R Shiny application.

**Goals**

1. Create a fully functional database using Oracle as the infrastructure, possibly PostgreSQL, also may use AWS for this project
2. Create an R Shiny application that will both visualize and export data that is useful to the end user

**Specifications**

This project will begin once I have the set criteria from Rob where he will give me the key indicators that he and his team look at while analyzing the data, and I will then begin to develop the R Shiny apps look and functionality. He will also tell me which data can be left out, as it may not all be useful to himself and his team.

**Milestones/Deadlines**

**ERD** – May 29th

**R Shiny App Format** – May 29th

**Database Completion** – June 7th

**R Shiny App Completion** – June 17th

**Additional Comments**

For the milestones/deadlines, I would like both the ERD and R Shiny App layout be completed by the same day, as those will serve as my backbone for creating my project. The ERD to me is the most challenging part to make sure that the database is normalized, and the R Shiny App’s layout will determine what needs to be added and how it must be created.

I would like the database to be completed before the R Shiny App, as I will have to test if the results that are being displayed from the app are accurate. Being completed with the project a week before the presentation will allow me to refine and tune my project if I find any issues with it or if I would like to add on to it with some backend scripting for data verification/quality.

I have left out the Python aspect of the project, as creating the database and the R Shiny app will be enough for the project. It will be enough work setting up both of those, and if I have time, I may do some Python at the end if I want to.