**Mortgage Delinquency Team Progress Report 1: October 19, 2019**

**Team:** 

**Domain:** Mortgage Delinquency/Default Rate Data

**Hypothesis:** To take key performance indicators and create a predicative model that will allow an accurate prediction of the default or delinquency rate of a mortgage within a 60-80% accuracy.

**Project Description:**

* Mortgage Delinquency/Default rate using key performance indicators (KPI) to build a data frame model which will be used for predictive analytics to determine delinquency/default of the mortgage along with current state of the economy.
* Utilizing statistical trend and regression analysis and methodologies to test the model. Utilizing test data to test the model to accurately forecast the probability that a mortgage with become delinquent/default.
* Present the results of the model indicating the accuracy of the model and the visualization of results.

**Meetings**: The team had meetings on Wednesdays during the week to discuss the current project status thru Go-To-Meeting. Other ad hoc meetings were conducted at the Reston Regional Library for in person meeting discussions. The team also would have lunch meetings during Saturday classes along with meeting after class when needed. Pending individual availability enough quorum attended the applicable meetings. Progress actions were communicated via email and/or phone calls for those members that could not attend meetings.

**Classes: With Actions and Tasks**

* Software Engineering
  + Proposal Written
    - * Data Exploration; locating- Freddie Mac Data source
* Data Sources
  + Architecture - Created
    - Project Pipeline design within the scope of the Data Science pipeline
  + Ingestion
    - PostgreSQL – setup and data ingested
      * Database setup and data ingested into database
      * Use SQL code to extract data
      * Python: Pandas Exploration with Data Extract from Database
  + Data Wrangling
    - Pandas
      * Use Python Pandas, NumPy, Matplotlib for Data Extract Exploration
    - Data format review
      * Changing the data format to relevant KPI
      * Review data elements within the dataset to determine KPIs and Non-KPIs
* Statistics
  + Progress Report I
  + Pandas and NumPy
* Visual Analytics
  + Progress Report II
* Applied Data Analytics
  + Demo and Final Report

**Task Actions:**

|  |  |  |
| --- | --- | --- |
| **Task** | **Task Done by** | **Notes** |
| Capstone Proposal | Joseph Welton |  |
| Architecture | Manish Pandey |  |
| Progress Report I | Joseph Welton |  |
| Database Research | Joseph Welton, Jitendra Patel, Nathaniel Rice, Manish Pandey, Terry Tsao |  |
| PostgreSQL Database Setup | Jitendra Patel, Nathaniel Rice, Terry Tsao | PostgreSQL/Python |
| PostgreSQL Database Data Dump | Jitendra Patel, Nathaniel Rice, Terry Tsao | PostgreSQL/Python |
| PostgreSQL - Data Extract: SQL Code actions | Joseph Welton, Jitendra Patel, Nathaniel Rice, Manish Pandey, Terry Tsao |  |
| Data Research-Locating Data Source(s) | Joseph Welton, Jitendra Patel, Nathaniel Rice, Manish Pandey, Terry Tsao | Freddie Mac Data |
| Data Exploration-Pandas | Joseph Welton, Jitendra Patel, Nathaniel Rice, Manish Pandey, Terry Tsao | Freddie Mac Data |
| Data Graphic Exploration -Python Pandas & Matplotlib | Joseph Welton, Jitendra Patel, Nathaniel Rice, Manish Pandey, Terry Tsao | Freddie Mac Data |
| Data Statistics Exploration -Python Pandas | Joseph Welton, Jitendra Patel, Nathaniel Rice, Manish Pandey, Terry Tsao | Freddie Mac Data |
| Algorithm Research | Manish Pandey | Scikit Learn Exploration |
| Modeling and Application | In progress |  |
| Pushing actions to GIT Hub | In progress |  |
| Scikit-Learn | In progress |  |
| Reporting and Visualization | In progress |  |
| Yellow Brick | In progress |  |
| Flask | In progress |  |
| Spark | In progress |  |
| Progress Report II | In progress |  |
| Demo and Final Report | In progress |  |

**Questions discovered during analyzing Data:**

1. Do we have sufficient/accurate data? We have over 15,000 mortgages that would be considered to meet the minimum requirements of being delinquent.
   1. Do we have the data that shows delinquent mortgages?
2. Do we have enough data for applicable actions?
   1. The Freddie Mac Data set has enough KPIs and volume of data to allow statistical actions to be performed using the Pandas package in Python.
3. What do the Statistics Exploration indicate?
   1. What are the KPI that show delinquent mortgages?
   2. Extracted the delinquent dataset from the Database created to perform a statistical analysis on the data to determine relevant KPI.
4. What do we expect the data to show?
   1. The initial data exploration of data extract from the databased showed unique correlation and statistical indicators for delinquent mortgages.
5. What visualization does the data show?
   1. Visual Exploration
   2. Data visualization of the initial data extract showed some interesting results allowing analytical actions of the supportive KPIs of delinquent mortgages.
6. When are we wanting to predict this delinquency?
   1. What are we trying to predict?
   2. What is the default threshold or the model if utilized to determine a good investment?
      1. What is a good investment?
   3. Delinquency at what point in time.
   4. Delinquency will happen when?
      1. Example: Something will happen at year 5
         1. Utilize delinquency data from year 1 and year 2 to train the model to predict what will happen at year 5
7. Who will this model serve?
8. Who is not delinquent in the dataset?
9. What are the non-KPIs within the dataset that need to be explored or taken inconsideration?
10. What will traditional KPIs within the dataset tell us?
    1. Expansion change KPIs for various model iterations
    2. What bias are in the KPIs
11. How are we going to use Pearson Correlation?
12. What is the instance?
    1. Is there more than one or what is the principal instance?
13. Uploading the SQL code for the Database to GitHub
    1. How many rows of data?
    2. What are the counts of data?
14. Using Data Exploration to determine which statistical probability model would best option to utilize for our dataset.

Project Name: **Mortgage Delinquent**

System Design

# Bibliography

Agency, F. H. (2017). *TREND: Federal Housing Finance Agency. House Price Index: House Price Index - All Transactions | State: New York | Seasonally Adjusted: Non-Seasonally Adj, 1975/1 - 2017/2. Data-Planet™ Statistical Datasets by Conquest Systems, Inc. Dataset-ID: 057-001-001*. Retrieved 10 18, 2019, from https://search.datacite.org/works/10.6068/dp15f08b7a62e41

Davis, M. A., Larson, W. D., Oliner, S. D., & Smith, B. (2019). *Mortgage Risk Since 1990*. Retrieved 10 18, 2019, from https://econpapers.repec.org/repec:hfa:wpaper:19-02

Liu, H. (2017). *Essays in Risk Management and Financial Econometrics*. Retrieved 10 18, 2019, from https://escholarship.org/uc/item/77r9f9z6

Ma, C. (2016). *House Price Expectations and Mortgage Default Decisions*. Retrieved 10 18, 2019, from https://papers.ssrn.com/sol3/delivery.cfm/ssrn\_id2831047\_code2494459.pdf?abstractid=2709563&mirid=1

Patrabansh, S. (2013). *A Study of First-Time Homebuyers*. Retrieved 10 18, 2019, from https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2533963

Initial Data Sources:

<https://www.fhfa.gov/DataTools/Downloads/Pages/Public-Use-Databases.aspx>

<https://www.federalreserve.gov/data/mdrm.htm>

<https://www.federalreserve.gov/data.htm>

<https://www.fanniemae.com/portal/funding-the-market/data/loan-performance-data.html>

<https://fred.stlouisfed.org/categories/32440>

<https://www.consumerfinance.gov/data-research/mortgage-performance-trends/download-the-data/>

<https://www.consumerfinance.gov/data-research/hmda/historic-data/>