SIADS 593: Milestone I

Team Project Proposal

version 2022.07.27.1.CT

**Instructions:** please make a ***copy*** of this template file (do not edit original).

## **Proposal Title:** TBA.

## 1. Team members

Please list your team members (2-3 max).

* Yueyai Ren (uniqname)
* Andrew Tunison (uniqname)
* Spencer Cheung(uniqname)

## 2. Project summary (Draft)

Summarize your proposed project in a few sentences.

#### What is your proposed project and why are you proposing it?

#### What are the question(s) you want to answer, or goal you want to achieve?

|  |
| --- |
| We aim to analyze the impact of various economic indicators, ranging from inflation, interest rates, or credit card delinquency rates on mortgage interest rates/ origination trends.  Understanding these relationships is crucial as it can help predict mortgage market trends and see their relationship with each other.  How do inflation or interest rate will impact the volume of mortgage?  Does credit card delinquency impacts mortgage origination trend?  Are there patterns that suggest shifts. In the economy that affect mortgage interest rate and others? |

## 3. Datasets (Draft)

#### Describe one primary dataset and at least one secondary dataset. If other secondary datasets will be used please describe them as well.

#### The proposed datasets should exhibit different features/columns and/or different access methods, e.g., \*.csv file, \*.json file, API retrieval, web scraping, etc. Different time periods, for example, with the same features/columns is not considered a different dataset. Remember, the focus of the project in this Milestone course is to give you the opportunity to practice your data manipulation skills, so feel free to challenge yourself.

#### If you're unsure if your data sets are "different enough" describe the datasets and request a review via the *#siads593\_[semester]\_001\_project* Slack channel.

#### **Please note:** all proposed datasets ***MUST*** be publicly available to all members of the class (students, instructors, course support personnel, etc.). Use of proprietary datasets for this project is ***not*** permitted.

## 3.1 Primary dataset description (Draft)

Describe your primary dataset. How is the data collected and how will you access it? Please share what features in the dataset are relevant to your topic. At a minimum, include the following information:

#### Short description (i.e., 1-3 sentences) of its key features

#### Estimated size (in records and/or bytes)

#### Location (give the URL or other access method)

#### Format (CSV, JSON, etc.)

#### Access method (download, web scraping, API, etc.)

|  |
| --- |
| Deciding on the suage of HMDA 2023 pipeline, CSV File,  consist of 11483889 rows × 99 columns,  size of 4.34GB,  obtain via HMDA ‘https://ffiec.cfpb.gov/data-publication/modified-lar/2023’  Description:  We will pull in multiple reports from Fred |

## 3.2 Secondary dataset(s) description (Draft)

Describe your secondary dataset(s). How is the data collected and how will you access it? Please share what features in the dataset(s) are relevant to your topic and describe the data types you’re expecting. At a minimum, for each secondary dataset include the following information:

#### Short description (i.e., 1-3 sentences) of its key features

#### Estimated size (in records and/or bytes)

#### Location (give the URL or other access method)

#### Format (CSV, JSON, etc.)

#### Access method (download, web scraping, API, etc.)

|  |
| --- |
| Multiple reports from FEDS could be the secondary datasets. |

## 3.3 [YES] Affirm: datasets are public. (Draft)

Please write YES in the above box to confirm that your primary and secondary datasets are accessible and available to your classmates and the instructional team.

## 4. Cleaning and manipulation (Draft)

Describe how you will need to manipulate your datasets: how will you handle missing or anomalous data? How will you join your primary and secondary datasets? What cleaning and manipulation challenges, if any, do you anticipate?

|  |
| --- |
| Seems some data are from FRED are Weekly and Quarterly, i will need to do a range set of sort to allocate correctly on time series analysis. |

## 5. Analysis (Draft)

Describe any analyses you plan to undertake. For each, please give the technique or approach and briefly explain what you expect to learn from it.

|  |
| --- |
| Review volume range from weekly, monthly and quarter basis to see if economic indicator will significantly impact origination and rates. |

## 6. Visualizations (Draft)

Describe in 1-3 sentences at least **two** data visualizations that you plan to create. Include the chart type (e.g. bar chart, scatterplot, SPLOM, etc.) as well as the variables (features) you intend to plot.

|  |
| --- |
| Considering a line graph with datetime, and heatmap. |

## 

## 7. Ethical considerations (Draft)

Does your choice of data raise any ethical issues? If so, briefly describe the concern and how you plan to mitigate it.

|  |
| --- |
| There shouldn’t be an ethical issue with HMDA as it showcase many details on a borrower basis, I will also review the # of ethnicity normalize to see if there’s a significant bias interms of origination dataset. |

## 8. Contributions (Draft)

Indicate the contribution that each team member will make to the project.

|  |
| --- |
| [Please use this space for your response. You may expand or contract this box as needed.] |

### 

## Changelog

(2022.07.27.1.CT) Update for 593

(2021.07.24.1.AW) Adjust title, number sections, simplify section headings, edit text