Sayaka Berkley

Dr. Valentine

Data 205 - 22017

Assignment #8

10/23/24

Project Plan

**DATA**

[Dynamic National Loan-Level Dataset](https://ffiec.cfpb.gov/data-publication/dynamic-national-loan-level-dataset/2023)

The dynamic files contain the national HMDA(Home Mortgage Disclosure Act) datasets, modified by the Bureau to protect applicant and borrower privacy, and updated weekly for all HMDA reporters. This dataset is good for the specific year of HMDA data I will work with from 2018 to 2022. The Dynamic dataset has been updated, and banks have submitted it since March 1, so I think this dataset is more accurate.

**Research goal(s) and questions**:

The goal of this analysis plan is “to contribute to the improvement of CFPB(Consumer Financial Protection Bureau)'s mortgage-related regulations and the soundness of the mortgage market”

The bursting of the housing bubble in the wake of the Lehman Shock in 2008 caused many consumers to be treated unfairly in relation to their mortgages, and this led to a serious problem in which many people lost their homes. In light of this situation, the CFPB aims to eradicate unfair practices and lack of transparencies practices in the mortgage market and to create an environment in which consumers can use mortgages with peace of mind.

1.    In-depth analysis of specific frauds.

* 1. Fraud patterns: In what cases are false income statements being made (e.g., self-employed, those with side jobs, etc.)? How fraudulent collateral valuations are carried out (e.g. fraudulent appraisals, overvaluing collateral properties, etc.)
  2. Scale of damage: How much financing was fraudulently carried out through false income declarations. The amount of loss incurred by financial institutions due to fraudulent collateral valuations
  3. Cause analysis: Why is this type of fraud so prevalent (e.g. lax screening standards, lenient penalties for fraud, etc.)

2. Analysis of specific regions or attributes: Focus on specific regions or attributes such as low-income groups or minorities

* 1. Unequal impact: Are people from specific regions or attributes more likely to be victims of fraud
  2. Are they being denied loans or forced to take out loans with high interest rates?

1. Systemic failures:
   1. Do existing regulations adequately protect people from certain regions or groups?
   2. Socio-economic factors: How do poverty, education level, place of residence, etc. affect the incidence of fraud?

**Tools, methods, and resources**: CFPB uses Python for data analysis; not allow to use public open analysis tools (Tableau), The Census demographic information.

**Risks and mitigations**: What could go wrong and how will you handle things that do go wrong?

1. Multivariate relationships: It is difficult to determine the causal relationship with a single factor because multiple factors are intricately intertwined in the occurrence of fraud.
2. External factors: It is impossible to completely rule out the impact of external factors such as economic fluctuations, policy changes, and financial regulations on the occurrence of fraud.
3. Privacy violations: Protection of personal information: As the data to be analyzed contains personal attribute information, it is necessary to give sufficient consideration to privacy protection. Appropriate measures such as anonymization and aggregation processing should be taken.

·**Milestones**: What will the progression of your project look like? Which parts depend on other parts?

1. 10/23 Submit the Project plan and creating a slide for presentation in the class Clarification of goals
2. 10/30 Start building modes/ Visualizations - Assumption of hypothesis
3. 11/6 Submit the models or visualizations
4. 11/13 Working with Fraud patterns and creating a slide for presentation in the class Model construction /Model evaluation
5. 11/20 Analysis of specific regions or attributes - Model construction /Model evaluation
6. 12/4 Creating a rough draft – creating a slide for Dry Runs
7. 12/11 Submit final cleaned up GitHub repo
8. 12/18 Submit Final report and PowerPoint