Data Appendix

I. Dataset 1: PRIMARY MORTGAGE MARKET SURVEY®

Derived from the Freddie Mac website, PRIMARY MORTGAGE MARKET SURVEY® is a dataset downloaded as an Excel spreadsheet. The dataset was then converted into a CSV format for the purposes of the project. Each row in the dataset represents weekly U.S. mortgage rates. The unit of observation is "interest rate." The project is focusing on analyzing trends between 30-year and 15-year fixed rate mortgages (FRM) and forecasting the next decade.

Dataset Uncleaned:

Column	Type	Description
Week	Datetime	The date for the recorded week's mortgage rates.
U.S. 30 yr FRM	Float	The average interest rate for a 30-year fixed-rate mortgage (FRM) in the U.S.
30 yr fees & points	Float	The average fees and points associated with the 30-year FRM.
U.S. 15 yr FRM	Float	The average interest rate for a 15-year fixed-rate mortgage (FRM) in the U.S.
15 yr fees & points	Float	The average fees and points associated with the 15-year FRM.
U.S. 5/1 ARM	Float	The average interest rate for a 5/1 adjustable-rate mortgage (ARM) in the U.S.
5/1 ARM fees & points	Float	The average fees and points associated with the 5/1 ARM.
U.S. 5/1 ARM margin	Float	The margin for the 5/1 ARM, which is added to the index rate to calculate the fully indexed rate.

30 yr FRM/ 5/1 ARM spread	Float	The difference (spread) between the 30-year FRM interest rate and the 5/1 ARM interest rate.

Dataset after cleaning:

Column	Type	Description
Week	Datetime	The date for the recorded week's mortgage rates.
U.S. 30 yr FRM	Float	The average interest rate for a 30-year fixed-rate mortgage (FRM) in the U.S.
U.S. 15 yr FRM	Float	The average interest rate for a 15-year fixed-rate mortgage (FRM) in the U.S.

II.   Variables in Dataset

The Freddie Mac dataset contains the following variables in the csv:
- Week - The date for the recorded week's mortgage rates.
- U.S. 30 yr FRM - The average interest rate for a 30-year fixed-rate mortgage (FRM) in the U.S.
- U.S. 15 yr FRM - The average interest rate for a 15-year fixed-rate mortgage (FRM) in the U.S.

III.   Summary Statistics on Numerical Variables

	Week	U.S. 30 yr FRM	U.S. 15 yr FRM
count	1729	1729.000000	1729.000000
mean	2008-03-20 08:28:52.330827008	5.800145	5.235662
min	1991-08-30 00:00:00	2.650000	2.100000
25%	1999-12-10 00:00:00	4.170000	3.370000
50%	2008-03-20 00:00:00	6.010000	5.420000
75%	2016-06-30 00:00:00	7.110000	6.650000
max	2024-10-10 00:00:00	9.250000	8.890000
std	NaN	1.693495	1.795265

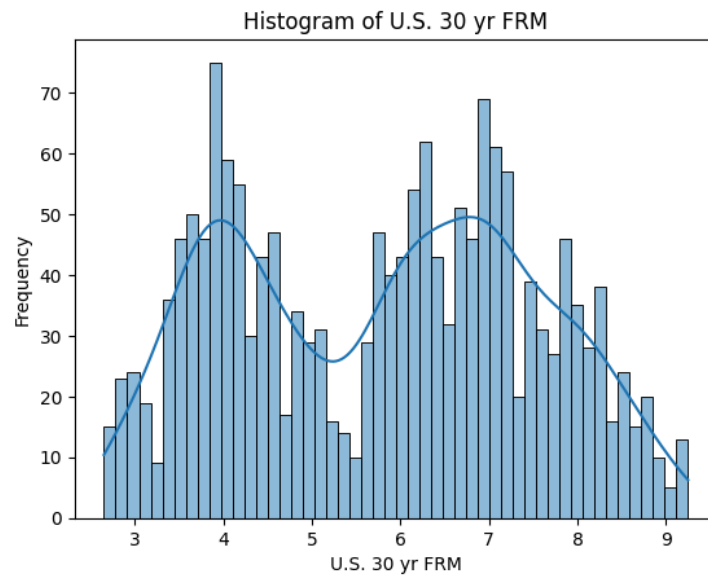## IV.   Frequency Tables on Numerical Variables

For the 30-year FRM:

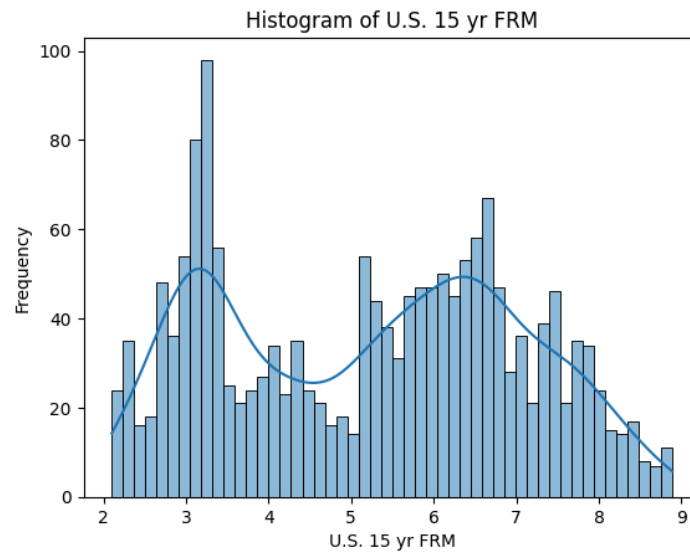Figure 1. Histogram of U.S. 30 year FRM

For the 15-year FRM:

Figure 2. Histogram of U.S. 15 year FRM

V.   Coding Scheme

Data preprocessing began by converting the Excel files to CSV format for easier manipulation. The raw CSV files were then imported into Google Colab for cleaning, where irrelevant columns were removed, retaining only the key variables: "Week,"

"30-year FRM," and "15-year FRM". The missing values were addressed by removing rows with "NaN" values in the primary columns of interest. Preliminary exploratory data analysis (EDA) was conducted to gain insights into trends, patterns, and any anomalies in the dataset. This analysis was supported by visualizations, which provided a clearer understanding of the data distribution and relationships between variables. For modeling and forecasting, SARIMA was chosen to forecast trends in the dataset with seasonal patterns. The results from the SARIMA model were then analyzed, with conclusions drawn based on the forecasting output. Moreover, the model performance will be assessed by comparing the forecasted rates to actual mortgage rates from 2015-2024. Further analysis was conducted to refine the model and better understand future trends.
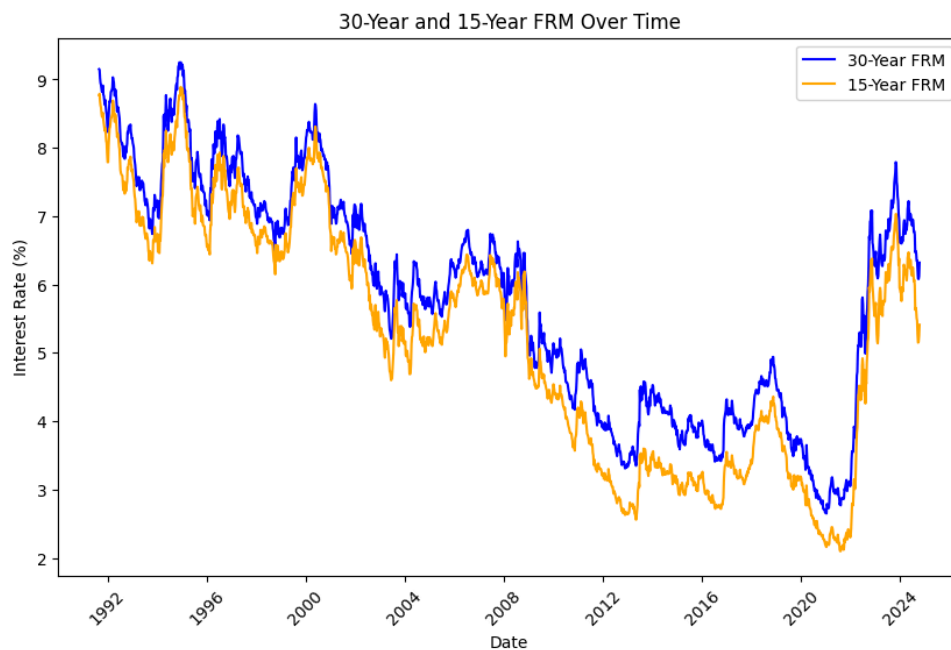
VI.   Exploratory Data Analysis

Figure 3. 30-Year and 15-Year FRM Overtime

This figure captures the interest rates of 30-year and 15-year FRM coded with their own respective colors, blue and orange. The plot depicts the ebbs and flows of each interest rate over time. For the most part, the 30-year and 15-year FRM remain quite close except the 30-year FRM remains higher than 15-year FRM throughout the years. Some notable drops in interest rates occurred in 2003, 2013, 2017, and most recently, 2021 and 2022. After the COVID-19 pandemic, there appears to be a rapid, sharp increase in interest rates and then presently there is a downward trend once again.
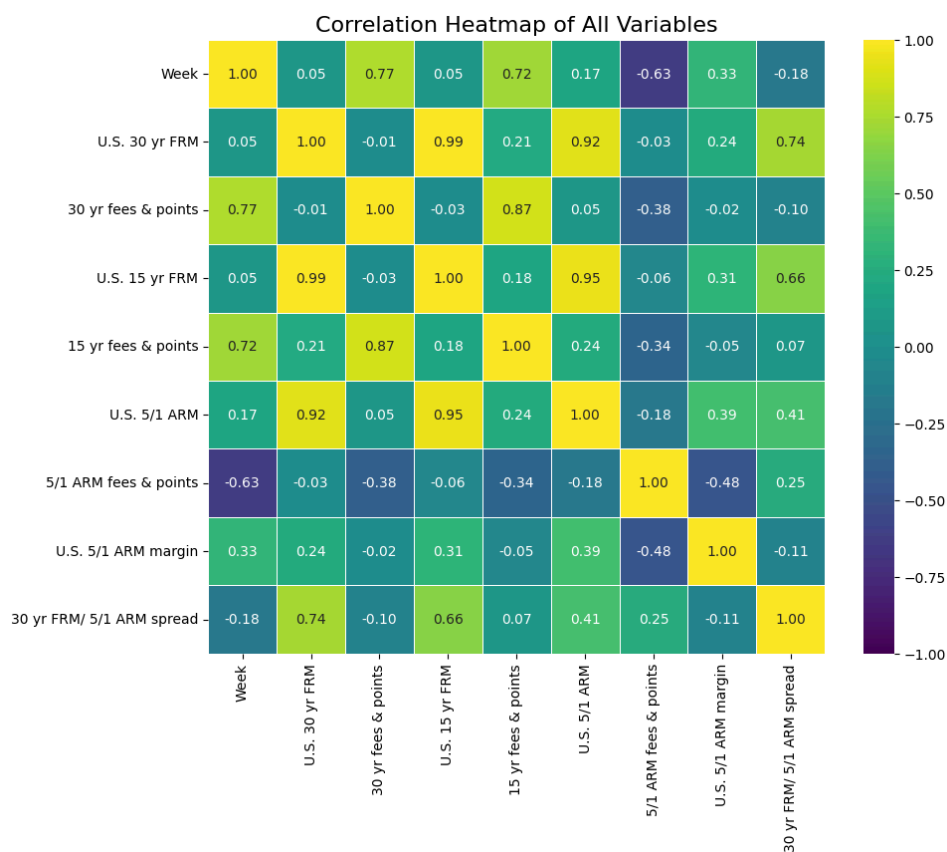
Figure 4. Correlation Heatmap of All Variables

The heatmap above features correlations for every variable included in the project such as Week, U.S. 30-yr FRM, and more. This figure reveals the level of strength, or weakness, between the variables involved in the project itself. The variables with the highest correlation of 0.99 are U.S. 30-year FRM and 15-year FRM–it makes sense that there is a very high positive correlation here and their relationship is depicted in Figure 3. Another variable with a high positive correlation is between U.S 15-year FRM and U.S. 5/1 ARM with a value of 0.95. The variable with the lowest correlation of -0.03 is U.S. 30 yr FRM and 5.1 ARM fees & points.