# Democratizing Residential Real Estate: Harnessing Big Data

## Problem and Opportunity Statement

There are large data sets related to real estate (GIS, Price History, Property Tax Information, etc.). Yet there’s a growing problem of the general populous being lost than ever before about the housing market with mix signals and misinformation from credible and uncredible sources. Hence, we have opportunity to solve the cyclical nature of real estate with data analytics intervention for information transparency and social stability.

### Description of the Statement:

Unaffordable Housing Crisis is affecting hundreds of millions of Americans every day. The cyclical nature of real estate market is largely caused by speculative and asymmetric information narrated by the Wall Street Elites in the media who are incentivized by profit from selling the mortgage-backed securities in the financial market to domestic and foreign investors to fulfill its own capitalistic greed in exchange of foundation of American families: Home. We must seek to democratize the housing market by providing free data analytics application accessible to the mass rather than having blind faith from so-called market experts from the media.

All large problems can be broken down to small parts and steps. To begin, the question becomes:

How do we harness large data sets to understand individual sub-market performance in residential real estate?

## Ideal Dataset

### Property Information Indices & Cluster information:

The basis of property information is the foundation of the data analysis for residential real estate. There are 208M single family residence and 37M multifamily units in the US. There is massive amount of data that are being generated daily.

This data can be obtained by major listing service platforms such as Redfin, Zillow, and Realtor.com

The property information indices will include three key variables: variety (5 > multiplex, SFR, etc.), volume (transaction volume per day, month, and year), velocity (how fast house is being resold). This information will harness all the market data.

### Geographic Information Systems (GIS):

The location is one of, if not, the most important factors in real estate analysis. The GIS data help us to understand by visualize the locality intelligence. The rise of open data sources of local and federal government is more accessible than ever before. Some of the public information available as variables are population migration by neighborhood to the location of public amenities. The GIS can also be used to walkability of neighborhood by foot, figure out commute times, population characteristics etc.

### Macroeconomic Factors:

There are countless macroeconomic factors. However, there are major players and variables that have higher weight for the coefficient of the real estate macro environment. All this data can be gathered from the federal government agency websites.

Some of the notable variables are:

Wage Growth/Housing Appreciation: This data indicates the housing affordability.

New Residential Building Permitting Data: This is a forward-looking indicator of housing supply.

Federal Bond Yield to Maturity (1 month to 30 years): This is a forward indicator of housing demand as it is related to mortgage payment interest rate.

Mortgage Application: This is a forward-looking indicator of housing demand

Stock Market (S&P 500, Dow Jones, Nasdaq): This is a forward indicator of labor market.

Current and Long-Term Inflation & GDP Growth: This is a current and forward indicator for the US economy.

## Analyses to Perform

### Descriptive (What Happened?):

We need to perform historical data analysis using property information indices and GIS data as far back as possible at the highest frequency (probably Monthly) to get the most accurate picture of descriptive analysis.

### Diagnostic (Why did it Happened?):

After the descriptive data analysis, we can investigate time periods of the past housing crash and boom to further scrutinize the macroeconomic conditions and correlations to the data set.

### Predictive (What will Happen?):

After the diagnostic analysis, we can investigate the correlations of past housing crash and boom compared to current market data to see where we are standing currently and where we’re heading based on the immediate indicators.

### Prescriptive (How can we make it Happen?):

After the diagnostic and predictive analyses, we learned the good indicators and conditions where we are of the real estate market cycle. If this analysis is freely available to the mass in a clear and simple communicative manner, the public can control the narrative of the market and making the economy condition favorable to housing stability.

# References:

1. <https://www.statista.com/statistics/1042111/single-family-vs-multifamily-homes-usa/#:~:text=As%20of%20October%202021%2C%20there,only%2037.8%20million%20multifamily%20units>