# Capstone Project

## MBS Realestate

Rajesh Nagarjunan

Date: 29 Jan 2021

# Realestate Project – Problem Statement

## Problem / Painpoint:

MBS (Mortgage Backed Securities) trading requires a very good understanding of user base which they seem to be missing !

## Approach & Learning:

1. Understanding the problem space and domain is key
2. Explore solution space to see how much we can get from this data
3. Explore Models, Accuracy, Error, Validate

## Solution :

1. How would this look like and what is our USP (Unique Selling Point)
2. Solution space and limitless possibilities for expansion
3. Learnings

"If you can film an idea in your mind, follow that film idea shot for shot, scene for scene, that idea is worth making.

- Craig Mapp

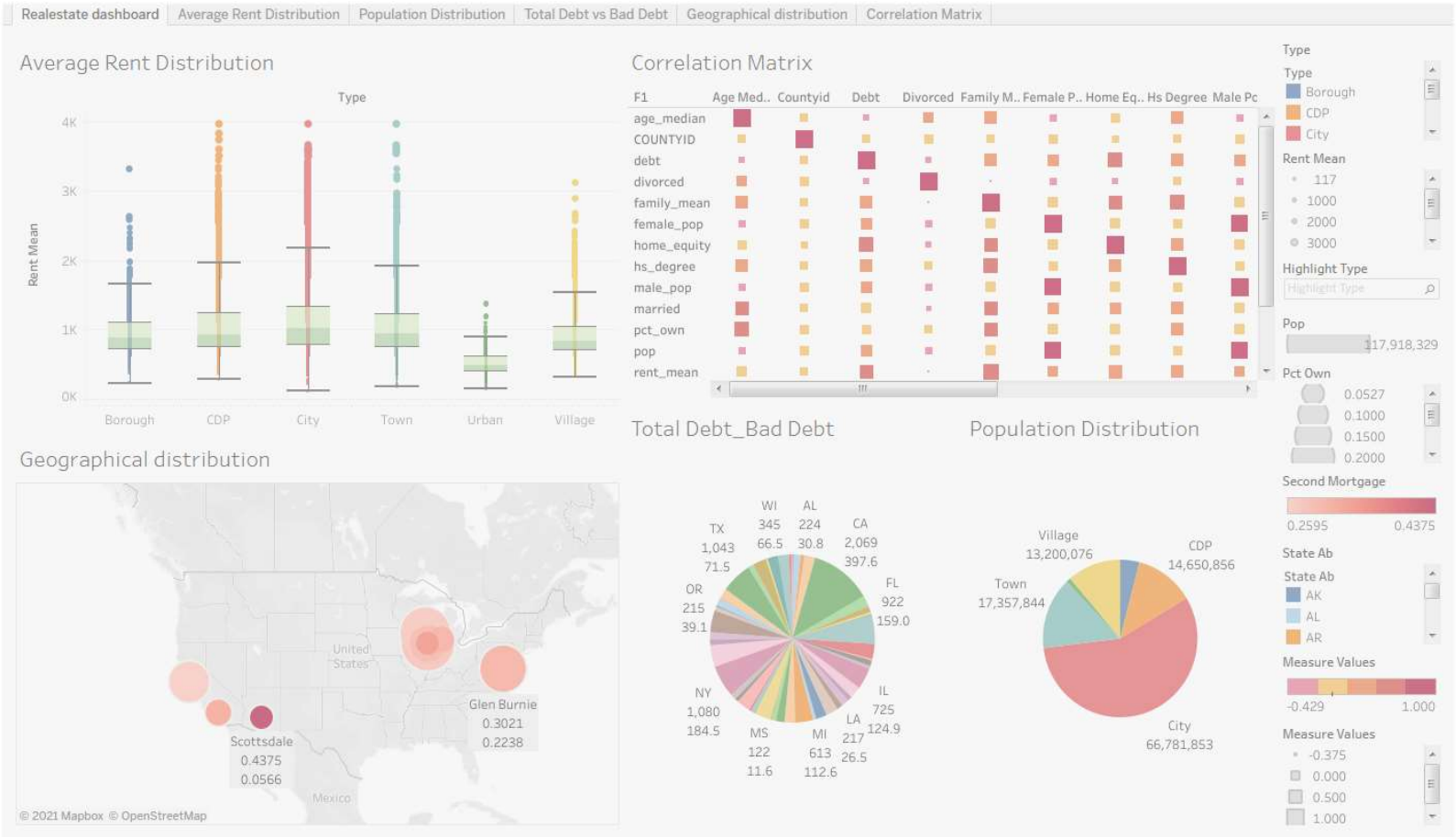
# Realestate Project – Problem Statement

- MBS Structuring : Complex business that cannot be performed without good understanding of what is happening in the market
- MBS Trading : Requires insights on how the trends are moving and without this the traders are blindsided
- Pain Point 1: Huge amount of data exists but a good summary of key features identified through analytics is missing
- Pain Point 2: Need a daily report of the Analytic features that require to be applied to the product strategy, first thing every morning on desk. And with this a dashboard for boardroom discussions

# What would the MBS traders like to see ?

- I would want to see if my product structure still fits the market in terms of its original purpose
- I would want to understand any change in trends day on day as we move forward
- I would want to get an understanding of demographics by age, sex, location etc
- I would want to know that I have full control over my client finances and can report back to regulators with confidence that there is adequate guard rails for financial risks / avoid melt down?

# Final Outcome



# Approach

## List of activites

- Data Import
- Data Profiling
- Missing Value Treatment
- Maximum Debt Mortgage
- Bad Debt vs Total Debt
- Second Mortgage Distribution
- Home Equity Distribution
- Good and Bad Debt Distribution
- Analyse
  - Population Density, Median Age, Group Bins, Marital Status
  - Overall Income Level, Corelation Analysis

# Week 1 – Understanding Key Data Elements

| Variables | Description |
| --- | --- |
| Second mortgage | Households with a second mortgage statistics |
| Home equity | Households with a home equity loan statistics |
| Debt | Households with any type of debt statistics |
| Mortgage Costs | Statistics regarding mortgage payments, home equity loans, utilities, and property taxes |
| Home Owner Costs | Sum of utilities, and property taxes statistics |
| Gross Rent | Contract rent plus the estimated average monthly cost of utility features |
| High school Graduation | High school graduation statistics |
| Population Demographics | Population demographics statistics |
| Age Demographics | Age demographic statistics |
| Household Income | Total income of people residing in the household |
| Family Income | Total income of people related to the householder |

# Outputs for Analysis

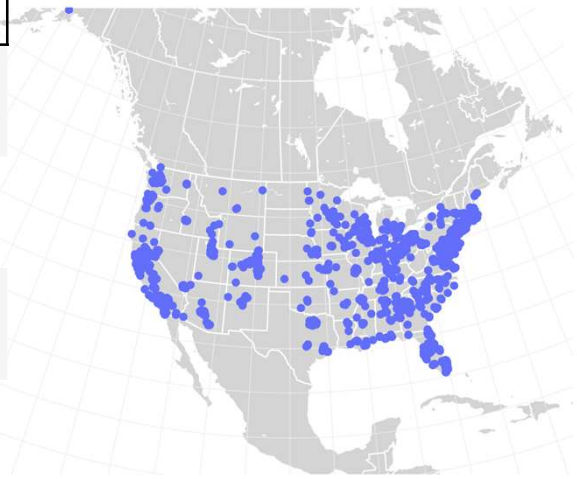
- Check the heatmap to see how correlated are the data


Looking at this it seemed like there was too much data variants which correlated like mean, Deviation, Avg, etc.. Which could be ideally removed if their correlation was similar to just the value

# Week 2

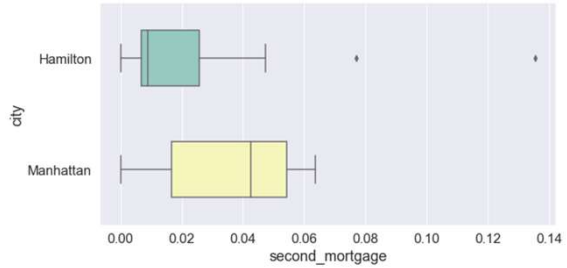
- **Top 2,500 locations where the percentage of households with a second mortgage is the highest and percent ownership is above 10 percent.**

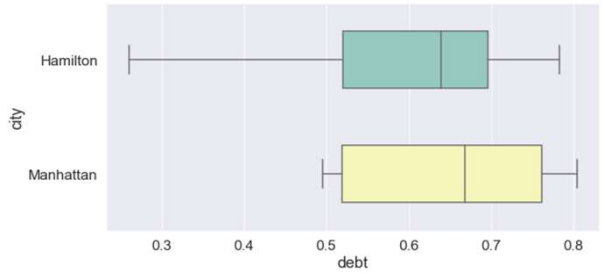
| | place | pct_own | second_mortgage | lat | lng | |
|---|---|---|---|---|---|---|
| **0** | Worcester City | 0.20247 | 0.43363 | 42.254262 | -71.800347 | |
| **1** | Harbor Hills | 0.15618 | 0.31818 | 40.751809 | -73.853582 | |
| **2** | Glen Burnie | 0.22380 | 0.30212 | 39.127273 | -76.635265 | |
| **3** | Egypt Lake-leto | 0.11618 | 0.28972 | 28.029063 | -82.495395 | |
| **4** | Lincolnwood | 0.14228 | 0.28899 | 41.967289 | -87.652434 | |

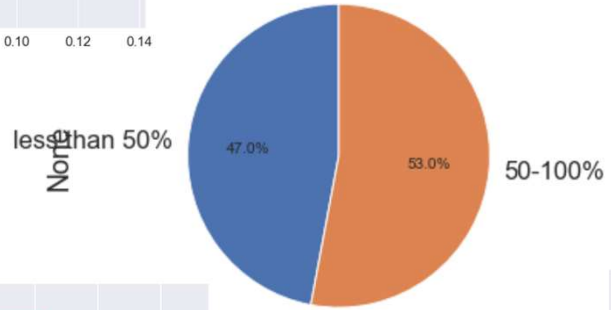


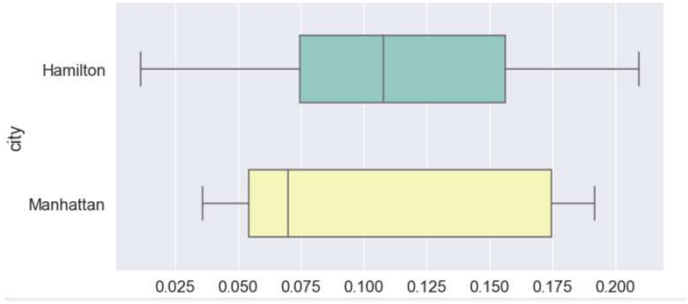
# Week 2

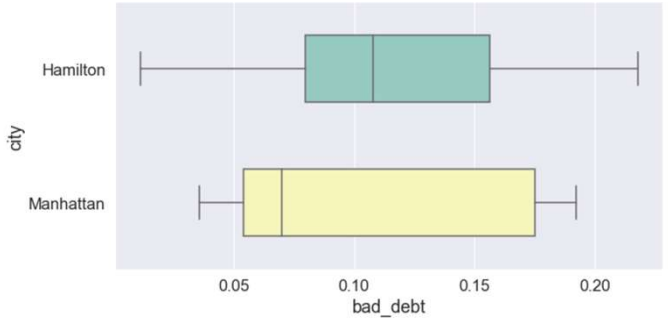
## Debt vs Bad debt comparison & distribution of 2nd mortgage, home equity, good debt, and bad debt for different cities



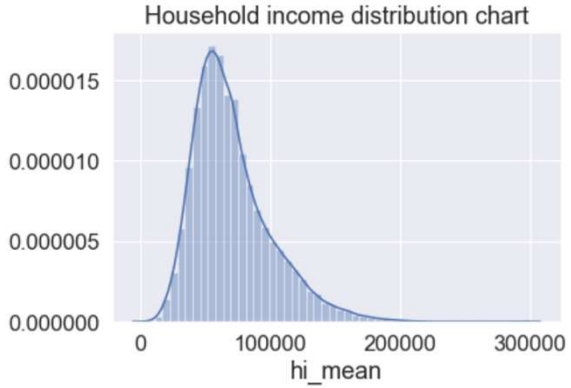



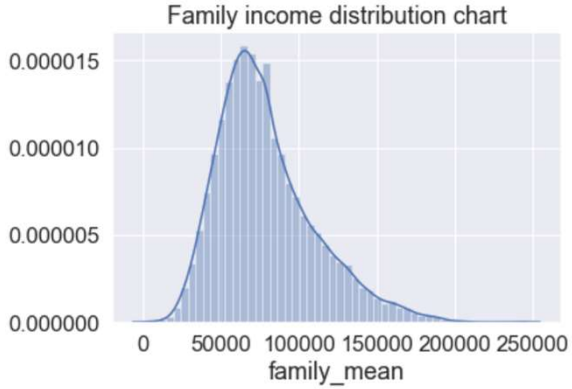



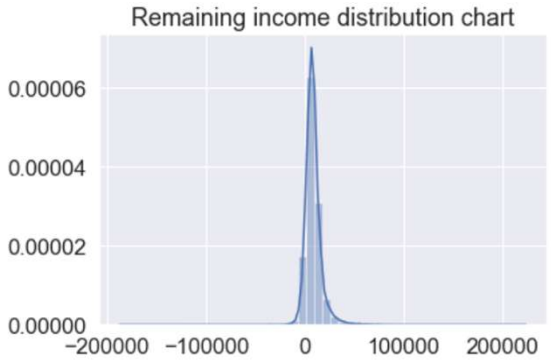



# Week 3

- **Collated income distribution chart for family income, house hold income, and remaining income¶**

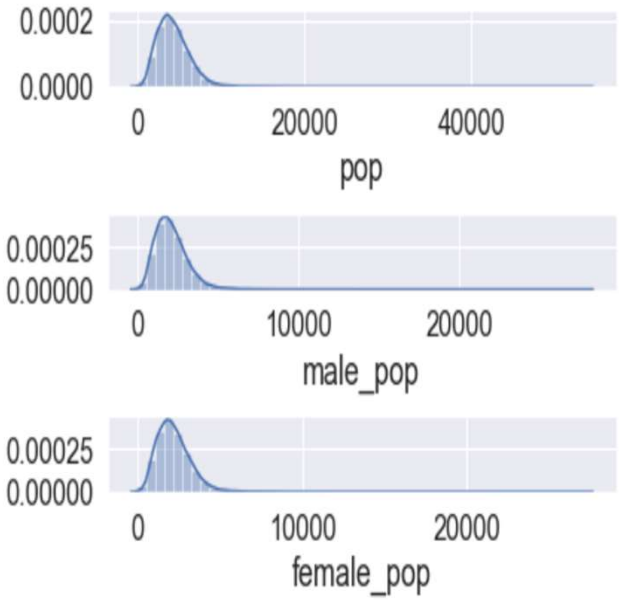


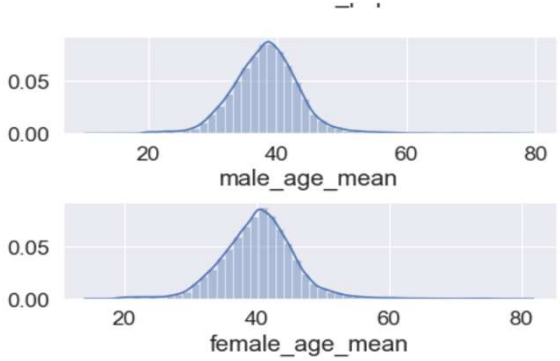



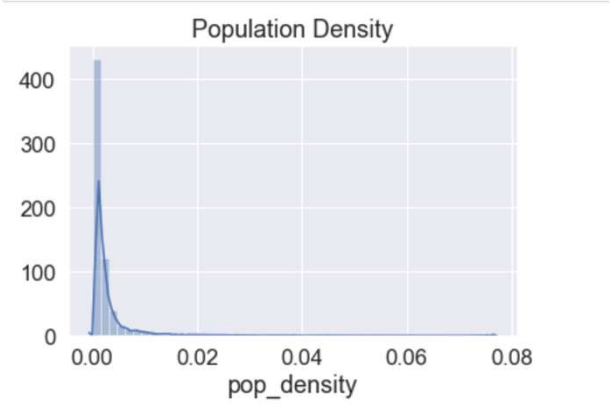



# Week 3

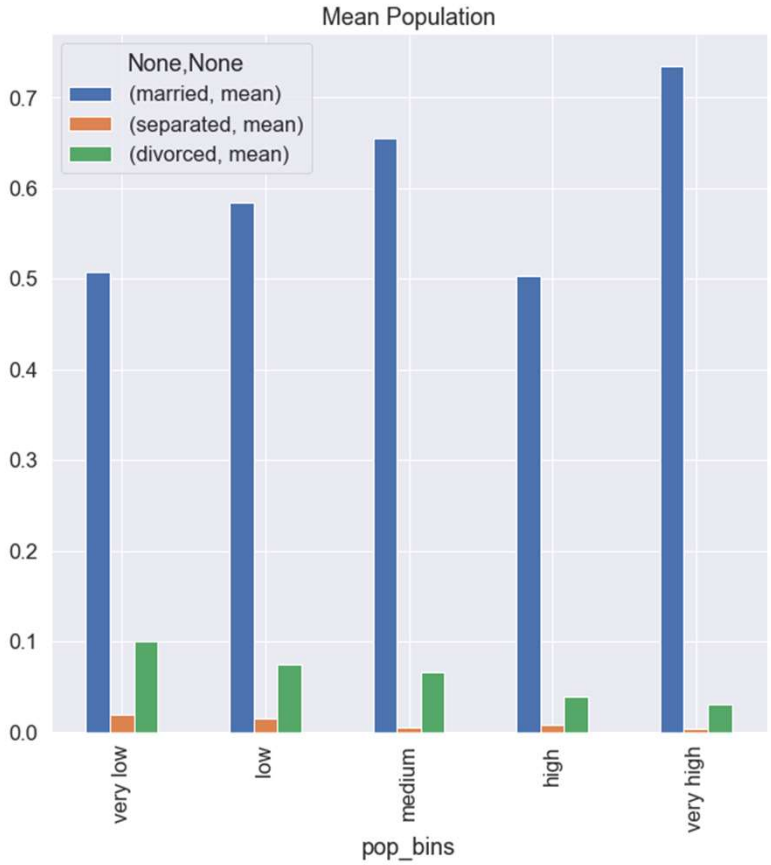
- **EDA to come out with insights into population density and age.:¶**

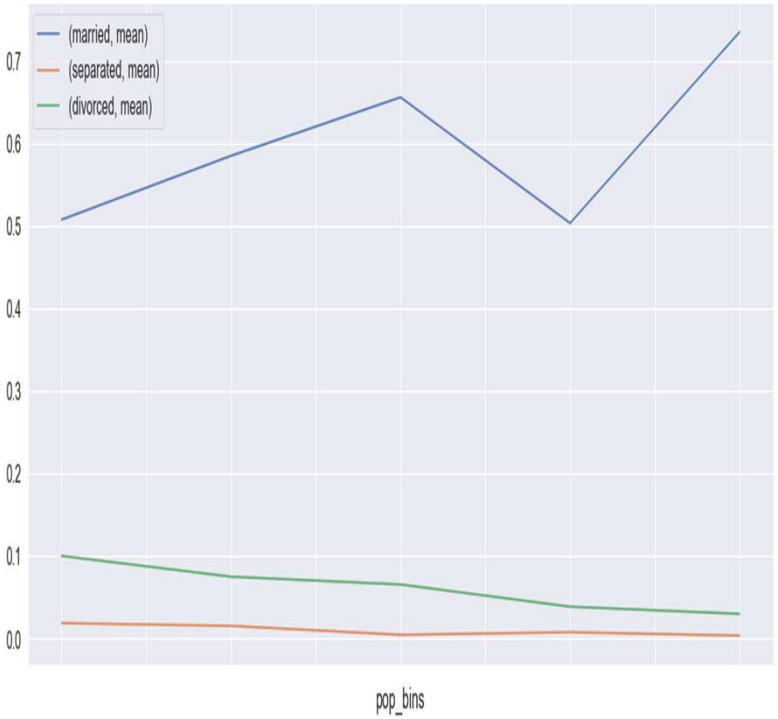


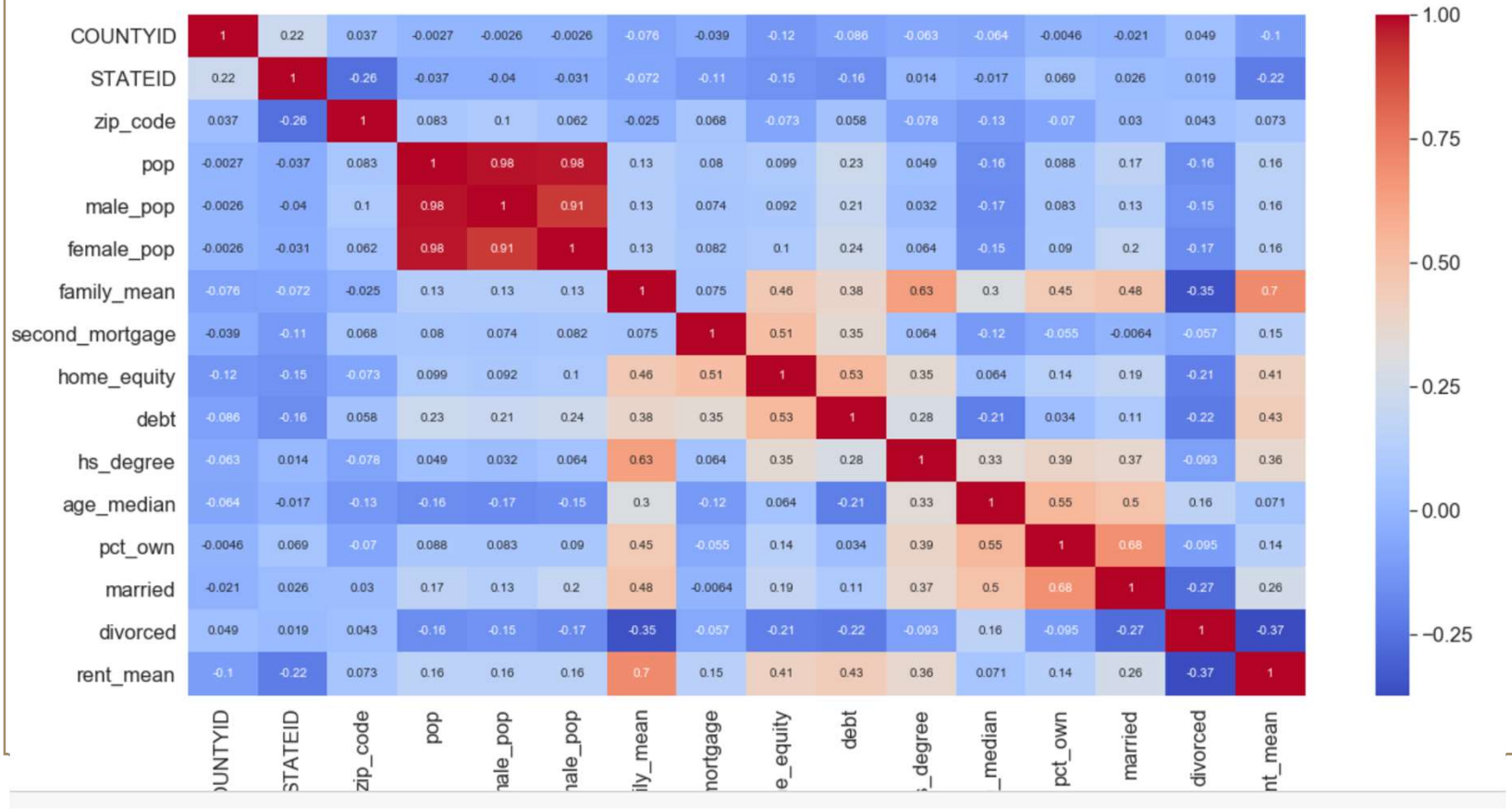
# Week 3


Mean Population
None,None
(married, mean)
(separated, mean)
(divorced, mean)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
very low
low
medium
high
very high
pop_bins


(married, mean)
(separated, mean)
(divorced, mean)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
pop_bins

# Week 3

- Correlation between fields



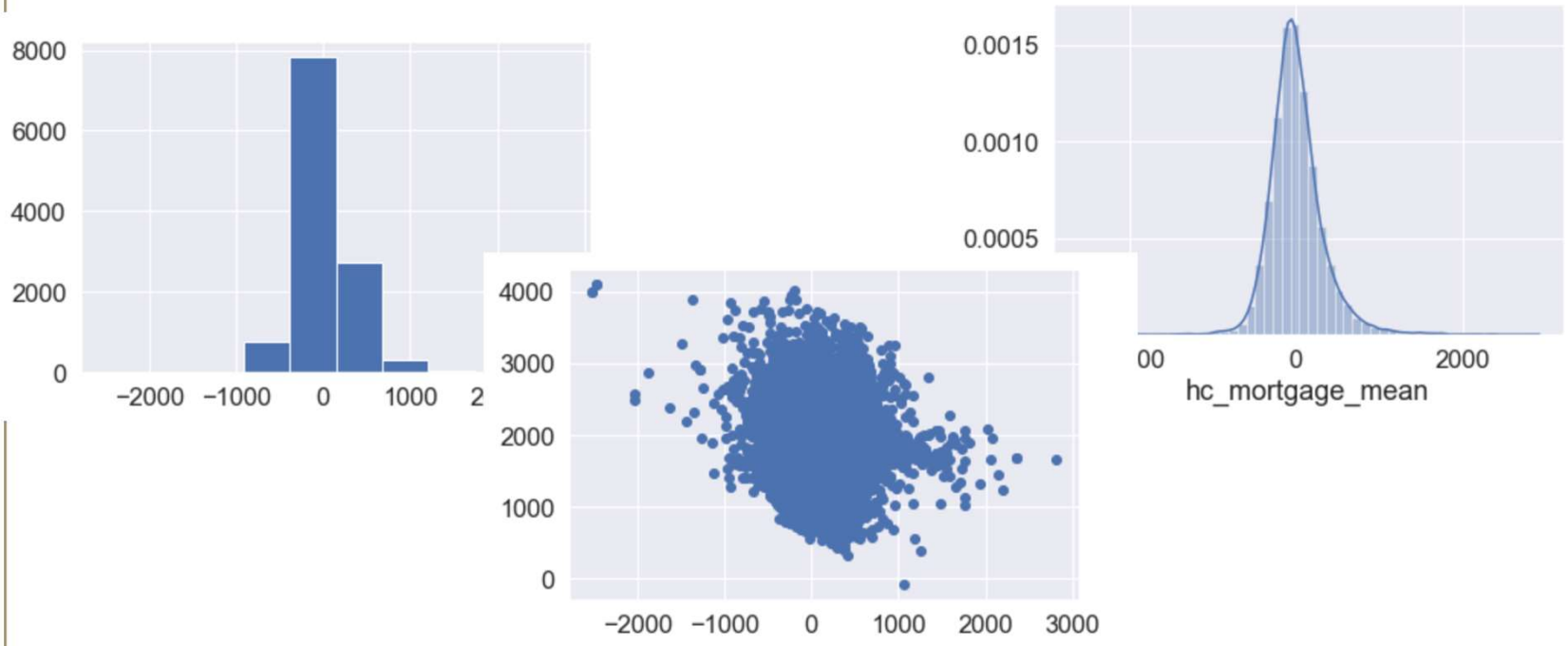
# Week 4

## Learning

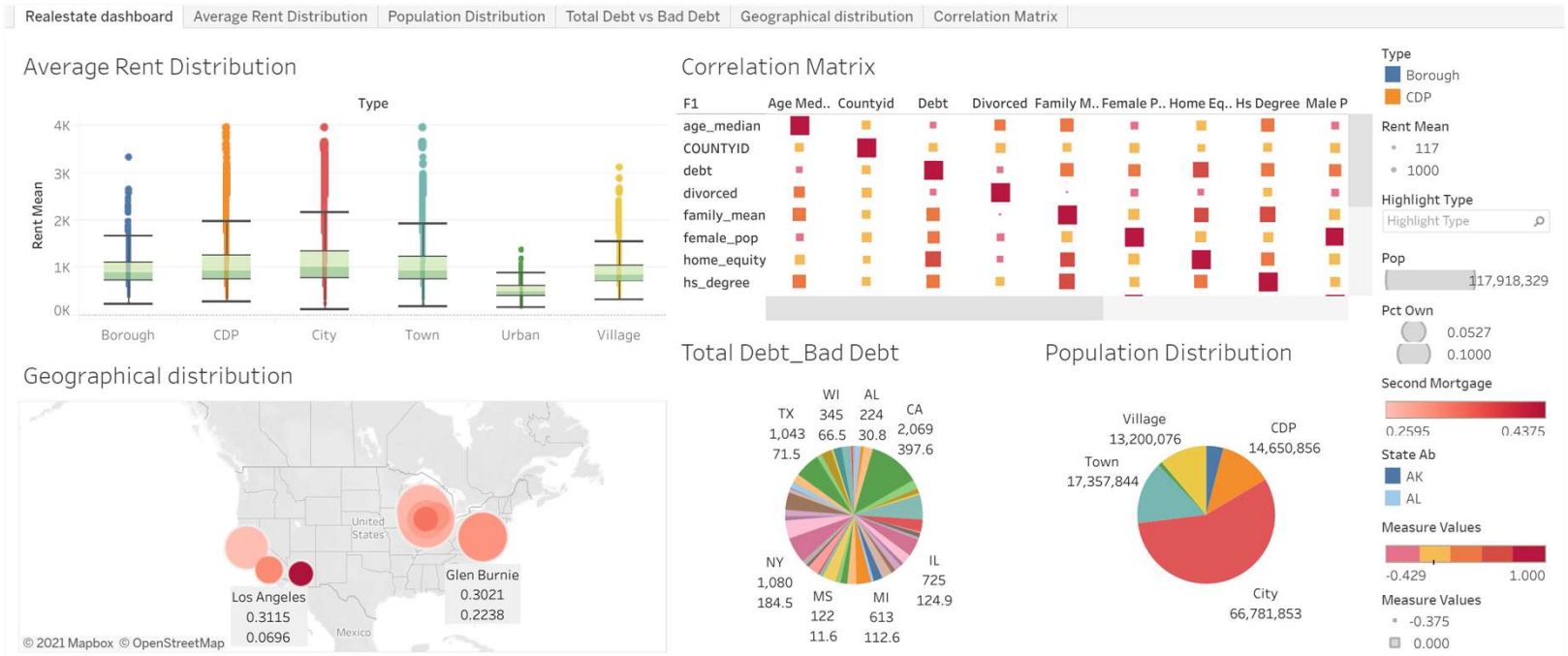
- Factor Analysis to find Latent variables
- Modelling
  - Nation Level
  - State Level
- Variable Verification
- Build a linear Regression model to predict the total monthly expenditure for home mortgages loan. • Variables should have significant impact on predicting Monthly mortgage and owner costs
- Utilize all predictor variable to start with initial hypothesis
- R square of 60 percent and above should be achieved
- Ensure Multi-collinearity does not exist in dependent variables
- Test if predicted variable is normally distributed

# Week 4 Variable Verification

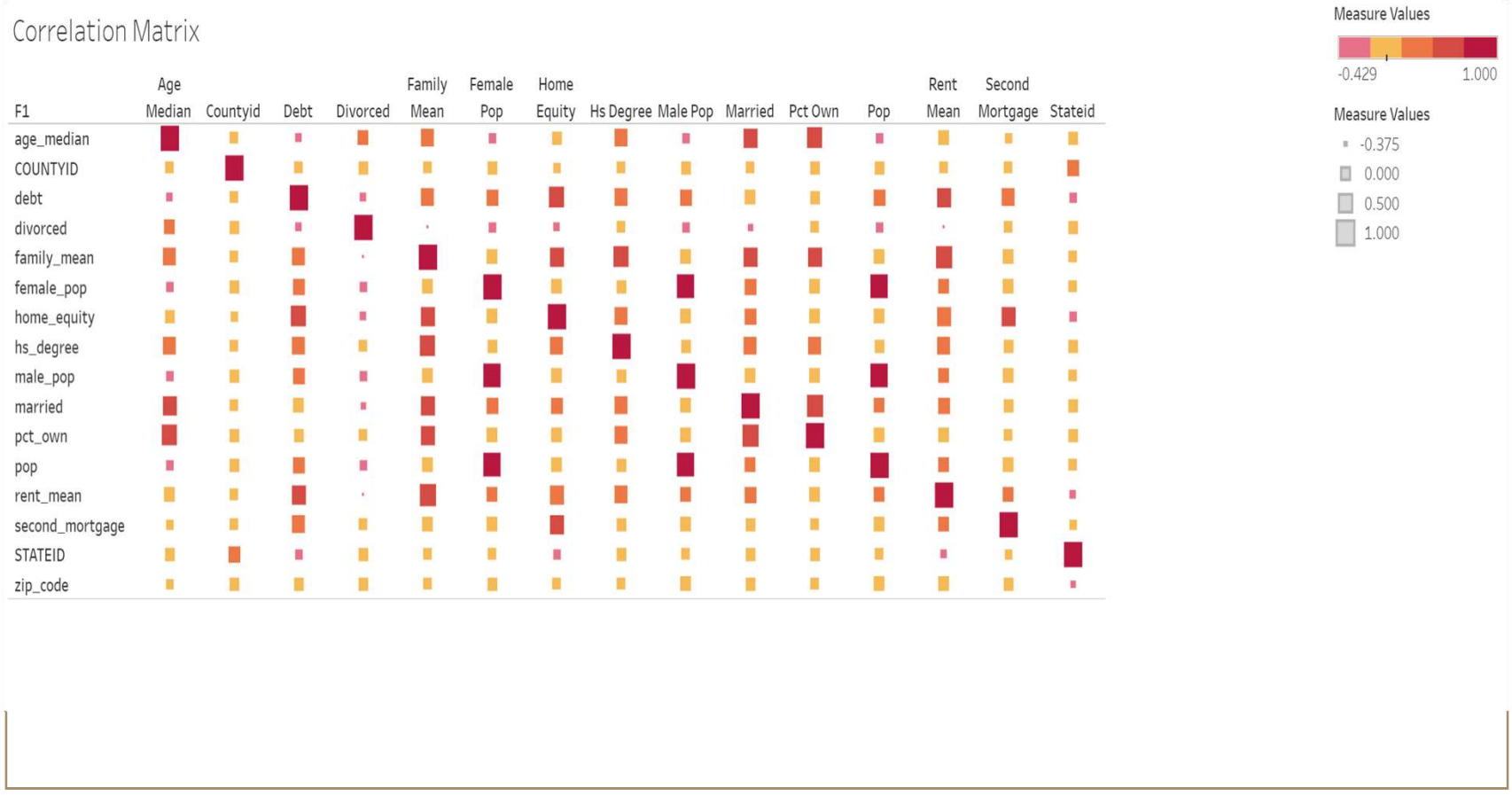
- **Checking Residuals of the Linear Regression model O/P**
- **Normal Distribution of Residuals**



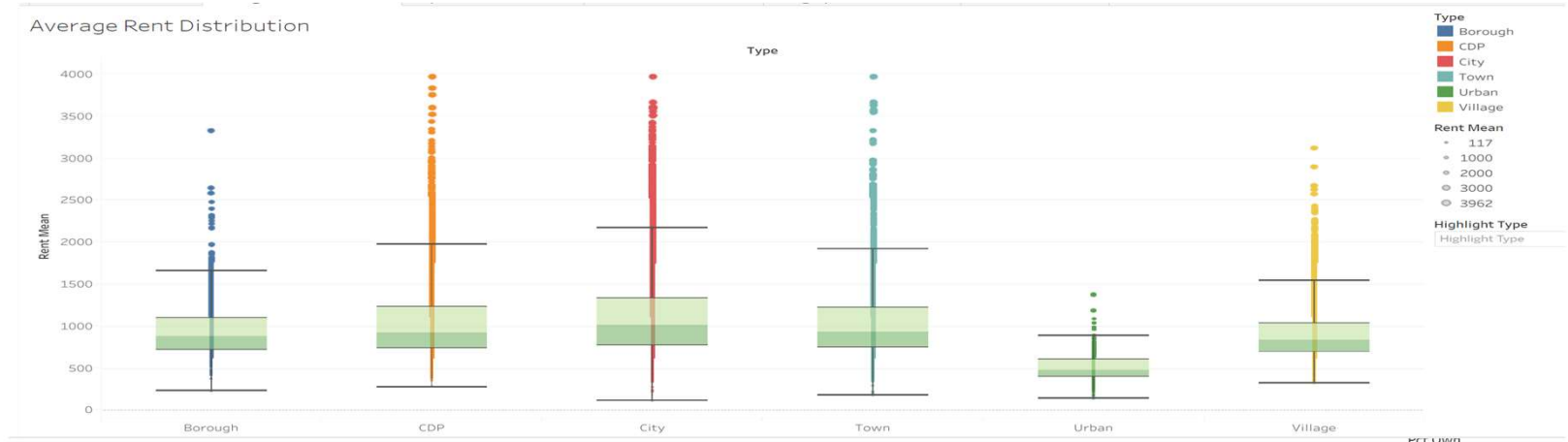
# Week 4

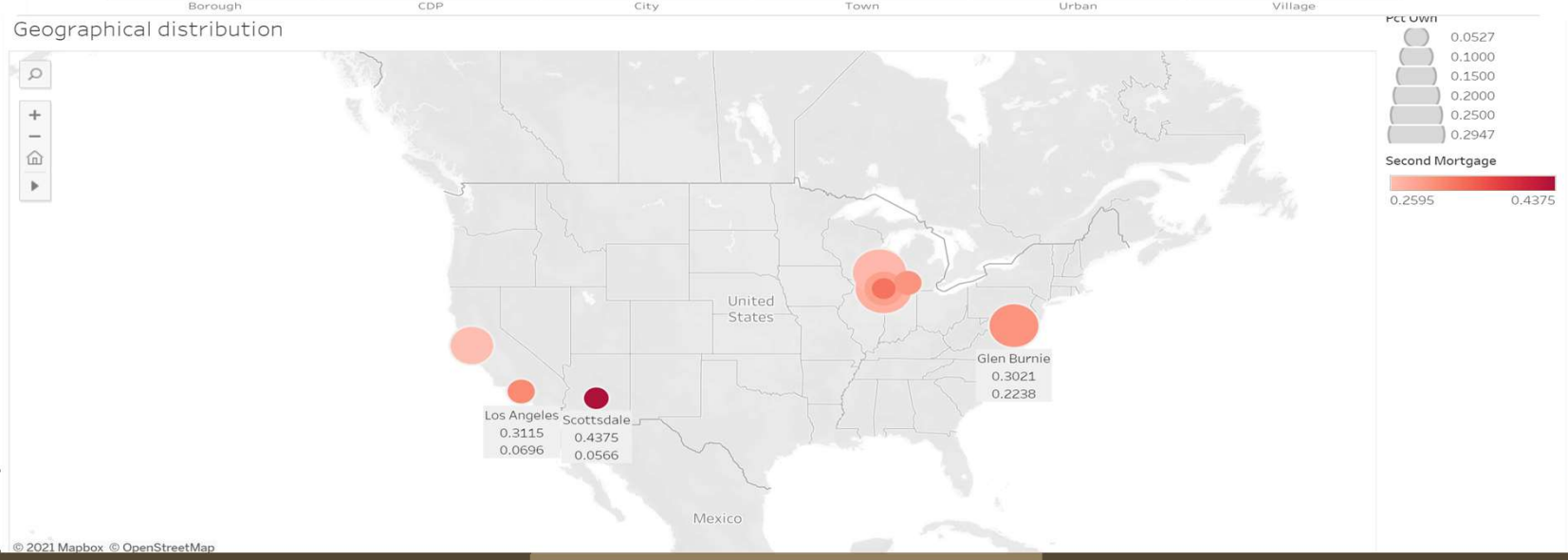

Realestate dashboard
Average Rent Distribution
Population Distribution
Total Debt vs Bad Debt
Geographical distribution
Correlation Matrix
Average Rent Distribution
Type
Rent Mean
4K
3K
2K
1K
0K
Borough
CDP
City
Town
Urban
Village
Correlation Matrix
F1
Age Med..
Countyid
Debt
Divorced
Family M..
Female P..
Home Eq..
Hs Degree
Male P
age_median
COUNTYID
debt
divorced
family_mean
female_pop
home_equity
hs_degree
Geographical distribution
United States
Mexico
Los Angeles
0.3115
0.0696
Glen Burnie
0.3021
0.2238
© 2021 Mapbox © OpenStreetMap
Total Debt_Bad Debt
WI
AL
TX
345
224
CA
1,043
66.5
30.8
2,069
71.5
397.6
NY
1,080
184.5
MS
122
11.6
MI
613
112.6
IL
725
124.9
Population Distribution
Village
13,200,076
CDP
14,650,856
Town
17,357,844
City
66,781,853
Type
Borough
CDP
Rent Mean
117
1000
Highlight Type
Highlight Type
Pop
117,918,329
Pct Own
0.0527
0.1000
Second Mortgage
0.2595
0.4375
State Ab
AK
AL
Measure Values
-0.429
1.000
Measure Values
-0.375
0.000

# Correlation Matrix

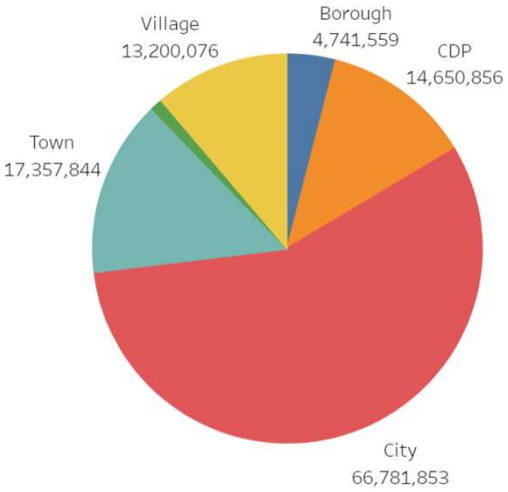


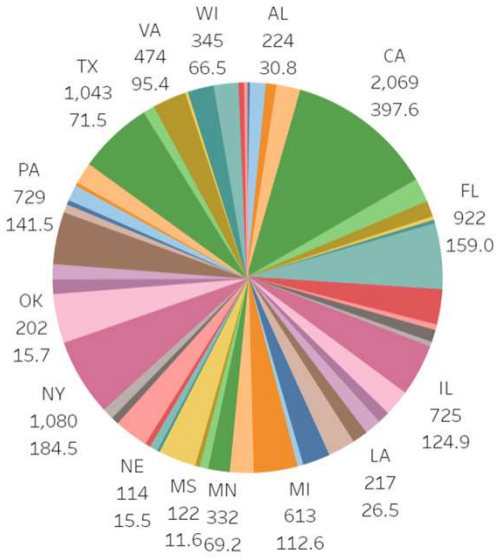
# Average Rent & Geographic Distribution



# Population Distribution & Debt vs Bad Debt



# Key Takeaways

**ONE:**
**Understanding Painpoints / Problem**

**TWO:**
**Domain Challenge**

**THREE:**
**User Journey / Goals / Objective**

**FOUR:**
**Analysis Pralysis**

**FIVE:**
**Repeatable cleansing, analysis and modelling**

**SIX:**
**Powerful Dashboard & Insight**

# Thank You

**Github Link**

CountMeIn (github.com)

**Rajesh Nagarjunan**