== REAL ESTATE PROJECT

Problem Statement

A banking institution requires actionable insights into mortgage-backed securities, geographic business investment, and real estate analysis.

The mortgage bank would like to identify potential monthly mortgage expenses for each region based on monthly family income and rental of the real estate.

A statistical model needs to be created to predict the potential demand in dollars amount of loan for each of the region in the USA. Also, there is a need to create a dashboard which would refresh periodically post data retrieval from the agencies.

The dashboard must demonstrate relationships and trends for the key metrics as follows: number of loans, average rental income, monthly mortgage and owner’s cost, family income vs mortgage cost comparison across different regions. The metrics described here do not limit the dashboard to these few.

Dataset Description

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

Project Task: Week 1

Data Import and Preparation:

Import data.

Figure out the primary key and look for the requirement of indexing.

Gauge the fill rate of the variables and devise plans for missing value treatment. Please explain explicitly the reason for the treatment chosen for each variable.

Exploratory Data Analysis (EDA):

4.Perform debt analysis. You may take the following steps:

a) Explore the top 2,500 locations where the percentage of households with a second mortgage is the highest and percent ownership is above 10 percent. Visualize using geo-map. You may keep the upper limit for the percent of households with a second mortgage to 50 percent

b) Use the following bad debt equation: Bad Debt = P (Second Mortgage ∩ Home Equity Loan) Bad Debt = second\_mortgage + home\_equity - home\_equity\_second\_mortgage c) Create pie charts to show overall debt and bad debt

d) Create Box and whisker plot and analyze the distribution for 2nd mortgage, home equity, good debt, and bad debt for different cities

e) Create a collated income distribution chart for family income, house hold income, and remaining income

Project Task: Week 2

Exploratory Data Analysis (EDA):

1. Perform EDA and come out with insights into population density and age. You may have to derive new fields (make sure to weight averages for accurate measurements):

a) Use pop and ALand variables to create a new field called population density

b) Use male\_age\_median, female\_age\_median, male\_pop, and female\_pop to create a new field called median age c) Visualize the findings using appropriate chart type

2. Create bins for population into a new variable by selecting appropriate class interval so that the number of categories don’t exceed 5 for the ease of analysis.

a) Analyze the married, separated, and divorced population for these population brackets

b) Visualize using appropriate chart type

3. Please detail your observations for rent as a percentage of income at an overall level, and for different states.

4. Perform correlation analysis for all the relevant variables by creating a heatmap. Describe your findings.

Project Task: Week 3

Data Pre-processing:

1. The economic multivariate data has a significant number of measured variables. The goal is to find where the measured variables depend on a number of smaller unobserved common factors or latent variables. 2. Each variable is assumed to be dependent upon a linear combination of the common factors, and the coefficients are known as loadings. Each measured variable also includes a component due to independent random variability, known as “specific variance” because it is specific to one variable. Obtain the common factors and then plot the loadings. Use factor analysis to find latent variables in our dataset and gain insight into the linear relationships in the data. Following are the list of latent variables:

• Highschool graduation rates

• Median population age

• Second mortgage statistics

• Percent own

• Bad debt expense

Project Task: Week 4

Data Modeling :

1. Build a linear Regression model to predict the total monthly expenditure for home mortgages loan. Please refer ‘deplotment\_RE.xlsx’. Column hc\_mortgage\_mean is predicted variable. This is the mean monthly mortgage and owner costs of specified geographical location. Note: Exclude loans from prediction model which have NaN (Not a Number) values for hc\_mortgage\_mean.

a) Run a model at a Nation level. If the accuracy levels and R square are not satisfactory proceed to below step.

b) Run another model at State level. There are 52 states in USA.

c) Keep below considerations while building a linear regression model. Data Modeling :

• 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

Data Reporting:

2. Create a dashboard in tableau by choosing appropriate chart types and metrics useful for the business. The dashboard must entail the following:

a) Box plot of distribution of average rent by type of place (village, urban, town, etc.).

b) Pie charts to show overall debt and bad debt.

c) Explore the top 2,500 locations where the percentage of households with a second mortgage is the highest and percent ownership is above 10 percent. Visualize using geo-map.

d) Heat map for correlation matrix.

e) Pie chart to show the population distribution across different types of places (village, urban, town etc.)

+\*In[1]:\*+

[source, ipython3]

----

#import libraries

import pandas as pd

import numpy as np

from itertools import cycle

import time

import random

from math import \*

#plotting libraries

import matplotlib.pyplot as plt

import seaborn as sns

import matplotlib

from pandas.plotting import scatter\_matrix

%matplotlib inline

sns.set(style="white", color\_codes=True)

sns.set(font\_scale=1.5)

----

+\*In[2]:\*+

[source, ipython3]

----

df\_train=pd.read\_csv("train.csv")

df\_test=pd.read\_csv("test.csv")

----

+\*In[3]:\*+

[source, ipython3]

----

df\_train.columns

----

+\*Out[3]:\*+

----Index(['UID', 'BLOCKID', 'SUMLEVEL', 'COUNTYID', 'STATEID', 'state',

'state\_ab', 'city', 'place', 'type', 'primary', 'zip\_code', 'area\_code',

'lat', 'lng', 'ALand', 'AWater', 'pop', 'male\_pop', 'female\_pop',

'rent\_mean', 'rent\_median', 'rent\_stdev', 'rent\_sample\_weight',

'rent\_samples', 'rent\_gt\_10', 'rent\_gt\_15', 'rent\_gt\_20', 'rent\_gt\_25',

'rent\_gt\_30', 'rent\_gt\_35', 'rent\_gt\_40', 'rent\_gt\_50',

'universe\_samples', 'used\_samples', 'hi\_mean', 'hi\_median', 'hi\_stdev',

'hi\_sample\_weight', 'hi\_samples', 'family\_mean', 'family\_median',

'family\_stdev', 'family\_sample\_weight', 'family\_samples',

'hc\_mortgage\_mean', 'hc\_mortgage\_median', 'hc\_mortgage\_stdev',

'hc\_mortgage\_sample\_weight', 'hc\_mortgage\_samples', 'hc\_mean',

'hc\_median', 'hc\_stdev', 'hc\_samples', 'hc\_sample\_weight',

'home\_equity\_second\_mortgage', 'second\_mortgage', 'home\_equity', 'debt',

'second\_mortgage\_cdf', 'home\_equity\_cdf', 'debt\_cdf', 'hs\_degree',

'hs\_degree\_male', 'hs\_degree\_female', 'male\_age\_mean',

'male\_age\_median', 'male\_age\_stdev', 'male\_age\_sample\_weight',

'male\_age\_samples', 'female\_age\_mean', 'female\_age\_median',

'female\_age\_stdev', 'female\_age\_sample\_weight', 'female\_age\_samples',

'pct\_own', 'married', 'married\_snp', 'separated', 'divorced'],

dtype='object')----

+\*In[4]:\*+

[source, ipython3]

----

df\_test.columns

----

+\*Out[4]:\*+

----Index(['UID', 'BLOCKID', 'SUMLEVEL', 'COUNTYID', 'STATEID', 'state',

'state\_ab', 'city', 'place', 'type', 'primary', 'zip\_code', 'area\_code',

'lat', 'lng', 'ALand', 'AWater', 'pop', 'male\_pop', 'female\_pop',

'rent\_mean', 'rent\_median', 'rent\_stdev', 'rent\_sample\_weight',

'rent\_samples', 'rent\_gt\_10', 'rent\_gt\_15', 'rent\_gt\_20', 'rent\_gt\_25',

'rent\_gt\_30', 'rent\_gt\_35', 'rent\_gt\_40', 'rent\_gt\_50',

'universe\_samples', 'used\_samples', 'hi\_mean', 'hi\_median', 'hi\_stdev',

'hi\_sample\_weight', 'hi\_samples', 'family\_mean', 'family\_median',

'family\_stdev', 'family\_sample\_weight', 'family\_samples',

'hc\_mortgage\_mean', 'hc\_mortgage\_median', 'hc\_mortgage\_stdev',

'hc\_mortgage\_sample\_weight', 'hc\_mortgage\_samples', 'hc\_mean',

'hc\_median', 'hc\_stdev', 'hc\_samples', 'hc\_sample\_weight',

'home\_equity\_second\_mortgage', 'second\_mortgage', 'home\_equity', 'debt',

'second\_mortgage\_cdf', 'home\_equity\_cdf', 'debt\_cdf', 'hs\_degree',

'hs\_degree\_male', 'hs\_degree\_female', 'male\_age\_mean',

'male\_age\_median', 'male\_age\_stdev', 'male\_age\_sample\_weight',

'male\_age\_samples', 'female\_age\_mean', 'female\_age\_median',

'female\_age\_stdev', 'female\_age\_sample\_weight', 'female\_age\_samples',

'pct\_own', 'married', 'married\_snp', 'separated', 'divorced'],

dtype='object')----

+\*In[5]:\*+

[source, ipython3]

----

df\_train.head()

----

+\*Out[5]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |BLOCKID |SUMLEVEL |COUNTYID |STATEID |state |state\_ab |city

|place |type |... |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|0 |267822 |NaN |140 |53 |36 |New York |NY |Hamilton |Hamilton |City

|... |44.48629 |45.33333 |22.51276 |685.33845 |2618.0 |0.79046 |0.57851

|0.01882 |0.01240 |0.08770

|1 |246444 |NaN |140 |141 |18 |Indiana |IN |South Bend |Roseland |City

|... |36.48391 |37.58333 |23.43353 |267.23367 |1284.0 |0.52483 |0.34886

|0.01426 |0.01426 |0.09030

|2 |245683 |NaN |140 |63 |18 |Indiana |IN |Danville |Danville |City |...

|42.15810 |42.83333 |23.94119 |707.01963 |3238.0 |0.85331 |0.64745

|0.02830 |0.01607 |0.10657

|3 |279653 |NaN |140 |127 |72 |Puerto Rico |PR |San Juan |Guaynabo

|Urban |... |47.77526 |50.58333 |24.32015 |362.20193 |1559.0 |0.65037

|0.47257 |0.02021 |0.02021 |0.10106

|4 |247218 |NaN |140 |161 |20 |Kansas |KS |Manhattan |Manhattan City

|City |... |24.17693 |21.58333 |11.10484 |1854.48652 |3051.0 |0.13046

|0.12356 |0.00000 |0.00000 |0.03109

|===

5 rows × 80 columns

----

+\*In[6]:\*+

[source, ipython3]

----

df\_test.head()

----

+\*Out[6]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |BLOCKID |SUMLEVEL |COUNTYID |STATEID |state |state\_ab |city

|place |type |... |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|0 |255504 |NaN |140 |163 |26 |Michigan |MI |Detroit |Dearborn Heights

City |CDP |... |34.78682 |33.75000 |21.58531 |416.48097 |1938.0 |0.70252

|0.28217 |0.05910 |0.03813 |0.14299

|1 |252676 |NaN |140 |1 |23 |Maine |ME |Auburn |Auburn City |City |...

|44.23451 |46.66667 |22.37036 |532.03505 |1950.0 |0.85128 |0.64221

|0.02338 |0.00000 |0.13377

|2 |276314 |NaN |140 |15 |42 |Pennsylvania |PA |Pine City |Millerton

|Borough |... |41.62426 |44.50000 |22.86213 |453.11959 |1879.0 |0.81897

|0.59961 |0.01746 |0.01358 |0.10026

|3 |248614 |NaN |140 |231 |21 |Kentucky |KY |Monticello |Monticello City

|City |... |44.81200 |48.00000 |21.03155 |263.94320 |1081.0 |0.84609

|0.56953 |0.05492 |0.04694 |0.12489

|4 |286865 |NaN |140 |355 |48 |Texas |TX |Corpus Christi |Edroy |Town

|... |40.66618 |42.66667 |21.30900 |709.90829 |2956.0 |0.79077 |0.57620

|0.01726 |0.00588 |0.16379

|===

5 rows × 80 columns

----

+\*In[7]:\*+

[source, ipython3]

----

len(df\_train)

----

+\*Out[7]:\*+

----27321----

+\*In[8]:\*+

[source, ipython3]

----

len(df\_test)

----

+\*Out[8]:\*+

----11709----

+\*In[9]:\*+

[source, ipython3]

----

pd.set\_option('max\_columns', 90)

pd.set\_option('max\_rows', 90)

plt.style.use('bmh')

color\_pal = plt.rcParams['axes.prop\_cycle'].by\_key()['color']

color\_cycle = cycle(plt.rcParams['axes.prop\_cycle'].by\_key()['color'])

----

+\*In[10]:\*+

[source, ipython3]

----

df\_train.dtypes

----

+\*Out[10]:\*+

----UID int64

BLOCKID float64

SUMLEVEL int64

COUNTYID int64

STATEID int64

state object

state\_ab object

city object

place object

type object

primary object

zip\_code int64

area\_code int64

lat float64

lng float64

ALand float64

AWater int64

pop int64

male\_pop int64

female\_pop int64

rent\_mean float64

rent\_median float64

rent\_stdev float64

rent\_sample\_weight float64

rent\_samples float64

rent\_gt\_10 float64

rent\_gt\_15 float64

rent\_gt\_20 float64

rent\_gt\_25 float64

rent\_gt\_30 float64

rent\_gt\_35 float64

rent\_gt\_40 float64

rent\_gt\_50 float64

universe\_samples int64

used\_samples int64

hi\_mean float64

hi\_median float64

hi\_stdev float64

hi\_sample\_weight float64

hi\_samples float64

family\_mean float64

family\_median float64

family\_stdev float64

family\_sample\_weight float64

family\_samples float64

hc\_mortgage\_mean float64

hc\_mortgage\_median float64

hc\_mortgage\_stdev float64

hc\_mortgage\_sample\_weight float64

hc\_mortgage\_samples float64

hc\_mean float64

hc\_median float64

hc\_stdev float64

hc\_samples float64

hc\_sample\_weight float64

home\_equity\_second\_mortgage float64

second\_mortgage float64

home\_equity float64

debt float64

second\_mortgage\_cdf float64

home\_equity\_cdf float64

debt\_cdf float64

hs\_degree float64

hs\_degree\_male float64

hs\_degree\_female float64

male\_age\_mean float64

male\_age\_median float64

male\_age\_stdev float64

male\_age\_sample\_weight float64

male\_age\_samples float64

female\_age\_mean float64

female\_age\_median float64

female\_age\_stdev float64

female\_age\_sample\_weight float64

female\_age\_samples float64

pct\_own float64

married float64

married\_snp float64

separated float64

divorced float64

dtype: object----

+\*In[11]:\*+

[source, ipython3]

----

df\_test.dtypes

----

+\*Out[11]:\*+

----UID int64

BLOCKID float64

SUMLEVEL int64

COUNTYID int64

STATEID int64

state object

state\_ab object

city object

place object

type object

primary object

zip\_code int64

area\_code int64

lat float64

lng float64

ALand int64

AWater int64

pop int64

male\_pop int64

female\_pop int64

rent\_mean float64

rent\_median float64

rent\_stdev float64

rent\_sample\_weight float64

rent\_samples float64

rent\_gt\_10 float64

rent\_gt\_15 float64

rent\_gt\_20 float64

rent\_gt\_25 float64

rent\_gt\_30 float64

rent\_gt\_35 float64

rent\_gt\_40 float64

rent\_gt\_50 float64

universe\_samples int64

used\_samples int64

hi\_mean float64

hi\_median float64

hi\_stdev float64

hi\_sample\_weight float64

hi\_samples float64

family\_mean float64

family\_median float64

family\_stdev float64

family\_sample\_weight float64

family\_samples float64

hc\_mortgage\_mean float64

hc\_mortgage\_median float64

hc\_mortgage\_stdev float64

hc\_mortgage\_sample\_weight float64

hc\_mortgage\_samples float64

hc\_mean float64

hc\_median float64

hc\_stdev float64

hc\_samples float64

hc\_sample\_weight float64

home\_equity\_second\_mortgage float64

second\_mortgage float64

home\_equity float64

debt float64

second\_mortgage\_cdf float64

home\_equity\_cdf float64

debt\_cdf float64

hs\_degree float64

hs\_degree\_male float64

hs\_degree\_female float64

male\_age\_mean float64

male\_age\_median float64

male\_age\_stdev float64

male\_age\_sample\_weight float64

male\_age\_samples float64

female\_age\_mean float64

female\_age\_median float64

female\_age\_stdev float64

female\_age\_sample\_weight float64

female\_age\_samples float64

pct\_own float64

married float64

married\_snp float64

separated float64

divorced float64

dtype: object----

+\*In[12]:\*+

[source, ipython3]

----

df\_train.describe()

----

+\*Out[12]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |BLOCKID |SUMLEVEL |COUNTYID |STATEID |zip\_code |area\_code |lat

|lng |ALand |AWater |pop |male\_pop |female\_pop |rent\_mean |rent\_median

|rent\_stdev |rent\_sample\_weight |rent\_samples |rent\_gt\_10 |rent\_gt\_15

|rent\_gt\_20 |rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35 |rent\_gt\_40 |rent\_gt\_50

|universe\_samples |used\_samples |hi\_mean |hi\_median |hi\_stdev

|hi\_sample\_weight |hi\_samples |family\_mean |family\_median |family\_stdev

|family\_sample\_weight |family\_samples |hc\_mortgage\_mean

|hc\_mortgage\_median |hc\_mortgage\_stdev |hc\_mortgage\_sample\_weight

|hc\_mortgage\_samples |hc\_mean |hc\_median |hc\_stdev |hc\_samples

|hc\_sample\_weight |home\_equity\_second\_mortgage |second\_mortgage

|home\_equity |debt |second\_mortgage\_cdf |home\_equity\_cdf |debt\_cdf

|hs\_degree |hs\_degree\_male |hs\_degree\_female |male\_age\_mean

|male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|count |27321.000000 |0.0 |27321.0 |27321.000000 |27321.000000

|27321.000000 |27321.000000 |27321.000000 |27321.000000 |2.732100e+04

|2.732100e+04 |27321.000000 |27321.000000 |27321.000000 |27007.000000

|27007.000000 |27007.000000 |27007.000000 |27007.000000 |27007.000000

|27007.000000 |27007.000000 |27007.000000 |27007.000000 |27007.000000

|27007.000000 |27007.000000 |27321.000000 |27321.000000 |27053.000000

|27053.000000 |27053.000000 |27053.000000 |27053.000000 |27023.000000

|27023.000000 |27023.000000 |27023.000000 |27023.000000 |26748.000000

|26748.000000 |26748.000000 |26748.000000 |26748.000000 |26721.000000

|26721.000000 |26721.000000 |26721.000000 |26721.000000 |26864.000000

|26864.000000 |26864.000000 |26864.000000 |26864.000000 |26864.000000

|26864.000000 |27131.000000 |27121.000000 |27098.000000 |27132.000000

|27132.000000 |27132.000000 |27132.000000 |27132.000000 |27115.000000

|27115.000000 |27115.000000 |27115.000000 |27115.000000 |27053.000000

|27130.000000 |27130.000000 |27130.000000 |27130.000000

|mean |257331.996303 |NaN |140.0 |85.646426 |28.271806 |50081.999524

|596.507668 |37.508813 |-91.288394 |1.295106e+08 |6.521754e+06

|4316.032685 |2123.924820 |2192.107866 |1055.129032 |1007.672789

|394.256202 |295.979447 |548.005702 |0.957824 |0.867134 |0.739429

|0.612959 |0.499994 |0.411007 |0.345424 |0.254469 |574.269390

|528.533546 |70441.191421 |57580.508964 |54429.005158 |923.580372

|1607.974384 |78987.539104 |69279.801465 |50728.337493 |533.686966

|1063.665988 |1629.856392 |1551.455735 |622.559191 |287.552519

|669.827389 |540.549473 |513.383968 |218.604647 |370.284570 |254.722233

|0.025695 |0.029947 |0.100847 |0.629190 |0.467957 |0.477485 |0.499458

|0.858459 |0.852136 |0.864931 |38.339988 |38.074193 |21.500301

|535.457318 |2138.719962 |40.319803 |40.355099 |22.178745 |544.238432

|2208.761903 |0.640434 |0.508300 |0.047537 |0.019089 |0.100248

|std |21343.859725 |NaN |0.0 |98.333097 |16.392846 |29558.115660

|232.497482 |5.588268 |16.343816 |1.275531e+09 |2.186781e+08

|2169.226173 |1114.948893 |1101.895160 |437.430562 |443.797814

|187.190303 |272.203470 |461.547524 |0.063186 |0.109655 |0.143799

|0.160305 |0.164006 |0.160201 |0.153217 |0.137742 |466.009996

|450.622720 |30166.895308 |29128.465950 |17619.932892 |453.057675

|751.096015 |31386.178602 |33472.030541 |14239.749880 |290.603105

|560.873112 |623.206122 |652.619435 |238.068593 |195.340264 |464.411215

|221.339933 |231.392365 |91.456509 |250.727935 |189.912748 |0.031331

|0.034134 |0.069304 |0.156267 |0.294956 |0.256125 |0.264138 |0.112420

|0.120746 |0.112273 |5.602570 |7.874651 |2.540576 |312.922652

|1104.593574 |5.886317 |8.039585 |2.540257 |283.546896 |1089.316999

|0.226640 |0.136860 |0.037640 |0.020796 |0.049055

|min |220342.000000 |NaN |140.0 |1.000000 |1.000000 |602.000000

|201.000000 |17.929085 |-165.453872 |4.113400e+04 |0.000000e+00

|0.000000 |0.000000 |0.000000 |117.150000 |104.000000 |18.257420

|0.343000 |4.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000

|0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |4999.846690

|4790.000000 |1825.741860 |0.114260 |3.000000 |5374.842520 |5278.000000

|1825.741860 |0.199960 |3.000000 |234.650000 |237.000000 |36.514840

|0.198400 |1.000000 |53.594610 |53.000000 |18.257420 |2.000000 |0.614040

|0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000

|0.186520 |0.000000 |0.000000 |12.145830 |9.750000 |0.962770 |0.745760

|3.000000 |16.008330 |13.250000 |0.556780 |0.664700 |2.000000 |0.000000

|0.000000 |0.000000 |0.000000 |0.000000

|25% |238816.000000 |NaN |140.0 |29.000000 |13.000000 |26554.000000

|405.000000 |33.899064 |-97.816067 |1.799408e+06 |0.000000e+00

|2885.000000 |1403.000000 |1454.000000 |743.153540 |702.000000

|263.662575 |101.922785 |221.000000 |0.940625 |0.819330 |0.662085

|0.517115 |0.396230 |0.307095 |0.243325 |0.160775 |250.000000

|209.000000 |49149.660560 |37424.000000 |42093.741360 |600.290760

|1096.000000 |56859.372910 |46166.000000 |40887.774050 |331.677595

|687.000000 |1158.312197 |1067.000000 |440.432127 |148.116155

|346.000000 |389.284170 |361.000000 |154.444740 |193.000000 |120.818180

|0.004990 |0.007680 |0.049247 |0.538460 |0.248910 |0.265270 |0.281195

|0.807890 |0.795270 |0.818025 |35.020857 |32.833330 |20.581183

|346.200507 |1416.000000 |36.892050 |34.916670 |21.312135 |355.995825

|1471.000000 |0.502780 |0.425102 |0.020810 |0.004530 |0.065800

|50% |257220.000000 |NaN |140.0 |63.000000 |28.000000 |47715.000000

|614.000000 |38.755183 |-86.554374 |4.866940e+06 |2.756300e+04

|4042.000000 |1978.000000 |2056.000000 |953.193930 |897.000000

|346.397060 |219.210100 |424.000000 |0.977070 |0.888160 |0.758170

|0.625000 |0.503790 |0.408600 |0.338620 |0.242950 |454.000000

|408.000000 |64020.023850 |51278.000000 |52213.886470 |863.714170

|1519.000000 |72876.445610 |62416.000000 |49679.731230 |490.868190

|986.000000 |1460.483290 |1371.000000 |589.364540 |253.549800

|590.000000 |478.798920 |449.000000 |198.699610 |327.000000 |213.030300

|0.018515 |0.022500 |0.094400 |0.648315 |0.419310 |0.466705 |0.491890

|0.889040 |0.883920 |0.895935 |38.336880 |37.833330 |21.906380

|490.967750 |1986.000000 |40.373320 |40.583330 |22.514410 |503.643890

|2066.000000 |0.690840 |0.526665 |0.038840 |0.013460 |0.095205

|75% |275818.000000 |NaN |140.0 |109.000000 |42.000000 |77093.000000

|801.000000 |41.380606 |-79.782503 |3.359820e+07 |5.239880e+05

|5430.000000 |2668.000000 |2764.000000 |1259.900165 |1198.000000

|475.601650 |408.709870 |742.000000 |1.000000 |0.940680 |0.837300

|0.722290 |0.608515 |0.515145 |0.440915 |0.335690 |771.000000

|718.000000 |85812.383150 |70734.000000 |65329.560620 |1179.293470

|2016.000000 |96010.265100 |84712.000000 |60415.096305 |685.226575

|1349.000000 |1982.588285 |1877.000000 |788.063712 |387.225985

|895.000000 |631.398210 |600.000000 |266.510900 |500.000000 |342.572420

|0.036943 |0.042733 |0.143492 |0.737525 |0.554115 |0.678620 |0.718510

|0.939580 |0.941070 |0.944650 |41.402437 |42.916670 |22.954955

|666.267473 |2672.250000 |43.567120 |45.416670 |23.575260 |680.275055

|2772.000000 |0.817460 |0.605760 |0.065100 |0.027487 |0.129000

|max |294334.000000 |NaN |140.0 |840.000000 |72.000000 |99925.000000

|989.000000 |67.074018 |-65.379332 |1.039510e+11 |2.453228e+10

|53812.000000 |27962.000000 |27250.000000 |3962.342290 |3972.000000

|1556.383030 |3060.247900 |6281.000000 |1.000000 |1.000000 |1.000000

|1.000000 |1.000000 |1.000000 |1.000000 |1.000000 |6648.000000

|6094.000000 |297142.857100 |296897.000000 |135902.619500 |10931.975610

|20395.000000 |242857.142900 |242720.000000 |111256.702500 |6904.496890

|14938.000000 |4462.342290 |4472.000000 |1596.206270 |4226.744200

|11670.000000 |1700.179110 |1702.000000 |820.968550 |11330.000000

|7107.064500 |1.000000 |1.000000 |1.000000 |1.000000 |1.000000 |1.000000

|1.000000 |1.000000 |1.000000 |1.000000 |77.759920 |80.166670 |31.060950

|12017.070440 |27962.000000 |79.837390 |82.250000 |30.241270

|6197.995200 |27250.000000 |1.000000 |1.000000 |0.714290 |0.714290

|1.000000

|===

----

+\*In[13]:\*+

[source, ipython3]

----

df\_test.describe()

----

+\*Out[13]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |BLOCKID |SUMLEVEL |COUNTYID |STATEID |zip\_code |area\_code |lat

|lng |ALand |AWater |pop |male\_pop |female\_pop |rent\_mean |rent\_median

|rent\_stdev |rent\_sample\_weight |rent\_samples |rent\_gt\_10 |rent\_gt\_15

|rent\_gt\_20 |rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35 |rent\_gt\_40 |rent\_gt\_50

|universe\_samples |used\_samples |hi\_mean |hi\_median |hi\_stdev

|hi\_sample\_weight |hi\_samples |family\_mean |family\_median |family\_stdev

|family\_sample\_weight |family\_samples |hc\_mortgage\_mean

|hc\_mortgage\_median |hc\_mortgage\_stdev |hc\_mortgage\_sample\_weight

|hc\_mortgage\_samples |hc\_mean |hc\_median |hc\_stdev |hc\_samples

|hc\_sample\_weight |home\_equity\_second\_mortgage |second\_mortgage

|home\_equity |debt |second\_mortgage\_cdf |home\_equity\_cdf |debt\_cdf

|hs\_degree |hs\_degree\_male |hs\_degree\_female |male\_age\_mean

|male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|count |11709.000000 |0.0 |11709.0 |11709.000000 |11709.000000

|11709.000000 |11709.000000 |11709.000000 |11709.000000 |1.170900e+04

|1.170900e+04 |11709.000000 |11709.000000 |11709.000000 |11561.000000

|11561.000000 |11561.000000 |11561.00000 |11561.000000 |11560.000000

|11560.000000 |11560.000000 |11560.000000 |11560.000000 |11560.000000

|11560.000000 |11560.000000 |11709.000000 |11709.000000 |11587.000000

|11587.000000 |11587.000000 |11587.000000 |11587.000000 |11573.000000

|11573.000000 |11573.000000 |11573.000000 |11573.000000 |11441.000000

|11441.000000 |11441.000000 |11441.000000 |11441.000000 |11419.000000

|11419.000000 |11419.000000 |11419.000000 |11419.000000 |11489.000000

|11489.000000 |11489.000000 |11489.000000 |11489.000000 |11489.000000

|11489.000000 |11624.000000 |11620.000000 |11604.000000 |11625.000000

|11625.000000 |11625.000000 |11625.000000 |11625.000000 |11613.000000

|11613.000000 |11613.000000 |11613.000000 |11613.000000 |11587.000000

|11625.000000 |11625.000000 |11625.000000 |11625.000000

|mean |257525.004783 |NaN |140.0 |85.710650 |28.489196 |50123.418396

|593.598514 |37.405491 |-91.340229 |1.095500e+08 |5.156069e+06

|4367.205995 |2152.510804 |2214.695192 |1054.143003 |1007.017646

|394.613338 |304.51603 |563.476256 |0.957482 |0.867770 |0.742615

|0.614405 |0.501188 |0.412992 |0.347003 |0.255507 |588.795969

|542.688189 |70169.909595 |57361.971779 |54164.666604 |935.084700

|1624.344093 |78684.992592 |69049.818630 |50408.173385 |540.262293

|1073.081483 |1636.445391 |1559.639018 |621.742098 |289.285332

|673.433004 |538.906730 |512.067869 |217.949778 |369.762326 |255.189048

|0.025789 |0.030187 |0.101570 |0.631615 |0.467226 |0.475517 |0.494432

|0.855912 |0.849148 |0.863003 |38.149424 |37.833111 |21.431971

|542.945584 |2168.064430 |40.111999 |40.131864 |22.148145 |550.411243

|2233.003186 |0.634194 |0.505632 |0.047960 |0.019346 |0.099191

|std |21466.372658 |NaN |0.0 |99.304334 |16.607262 |29775.134038

|232.074263 |5.625904 |16.407818 |7.624940e+08 |1.522649e+08

|2121.779736 |1086.382137 |1086.438040 |434.549555 |441.484366

|189.193868 |281.31471 |474.563369 |0.063603 |0.107789 |0.142514

|0.161556 |0.165759 |0.161312 |0.153982 |0.137658 |477.469706

|463.283992 |30619.277296 |29661.241996 |17794.261539 |457.759256

|747.394839 |31979.019465 |34130.762923 |14349.930513 |289.029814

|550.898356 |634.770720 |664.567754 |240.815700 |197.175161 |461.505232

|226.307832 |237.514474 |93.108675 |249.644673 |190.267726 |0.030513

|0.033644 |0.070412 |0.157634 |0.296905 |0.257148 |0.264962 |0.114424

|0.122605 |0.113205 |5.579728 |7.795907 |2.582541 |296.016752

|1074.723594 |5.851192 |7.972026 |2.554907 |280.992521 |1072.017063

|0.232232 |0.139774 |0.038693 |0.021428 |0.048525

|min |220336.000000 |NaN |140.0 |1.000000 |1.000000 |601.000000

|201.000000 |17.965835 |-166.770979 |8.299000e+03 |0.000000e+00

|0.000000 |0.000000 |0.000000 |147.548100 |104.000000 |18.257420

|0.39279 |3.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000

|0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |4999.846690

|4790.000000 |1825.741860 |0.399920 |3.000000 |5374.842520 |5278.000000

|1825.741860 |0.266610 |4.000000 |349.500000 |349.000000 |36.514840

|0.595190 |2.000000 |53.594610 |53.000000 |18.257420 |2.000000 |0.491230

|0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000

|0.000000 |0.000000 |0.199710 |17.009880 |9.750000 |0.737110 |0.745760

|4.000000 |15.360240 |12.833330 |0.737110 |0.251910 |3.000000 |0.000000

|0.000000 |0.000000 |0.000000 |0.000000

|25% |238819.000000 |NaN |140.0 |29.000000 |13.000000 |25570.000000

|404.000000 |33.919813 |-97.816561 |1.718660e+06 |0.000000e+00

|2937.000000 |1433.000000 |1484.000000 |741.389730 |704.000000

|262.377940 |103.86843 |226.000000 |0.940410 |0.820913 |0.665775

|0.517220 |0.397740 |0.307948 |0.241998 |0.160375 |255.000000

|216.000000 |48814.166430 |36953.500000 |41662.440610 |611.598530

|1110.000000 |56140.036620 |45709.000000 |40413.475230 |338.046690

|694.000000 |1152.337490 |1068.000000 |436.938690 |147.242890

|343.000000 |386.273775 |357.000000 |152.652175 |189.000000 |118.787880

|0.005060 |0.007790 |0.049700 |0.541060 |0.246060 |0.263960 |0.274550

|0.802980 |0.790218 |0.813850 |34.916000 |32.666670 |20.507130

|355.219790 |1445.000000 |36.729210 |34.750000 |21.270920 |363.225840

|1499.000000 |0.492500 |0.422020 |0.020890 |0.004500 |0.064590

|50% |257651.000000 |NaN |140.0 |61.000000 |28.000000 |47362.000000

|612.000000 |38.618092 |-86.643344 |4.835000e+06 |2.270900e+04

|4119.000000 |2010.000000 |2090.000000 |952.526270 |897.000000

|349.497450 |228.96877 |441.000000 |0.976970 |0.889180 |0.763485

|0.628110 |0.507090 |0.412875 |0.342330 |0.243710 |470.000000

|424.000000 |63788.482430 |51013.000000 |51925.227180 |877.368400

|1530.000000 |72809.895350 |61971.000000 |49401.698830 |496.572350

|996.000000 |1463.893720 |1374.000000 |586.516070 |255.414250

|593.000000 |474.995830 |445.000000 |198.361260 |327.000000 |212.090910

|0.018780 |0.022600 |0.095440 |0.650070 |0.418330 |0.461850 |0.487770

|0.886430 |0.881020 |0.893695 |38.200730 |37.833330 |21.884600

|499.653480 |2020.000000 |40.196960 |40.333330 |22.472990 |509.103610

|2099.000000 |0.687640 |0.525270 |0.038680 |0.013870 |0.094350

|75% |276300.000000 |NaN |140.0 |109.000000 |42.000000 |77406.000000

|787.000000 |41.232973 |-79.697311 |3.204540e+07 |4.864500e+05

|5474.000000 |2690.000000 |2792.000000 |1259.756750 |1194.000000

|475.718140 |420.81563 |763.000000 |1.000000 |0.939660 |0.839375

|0.726447 |0.612313 |0.517088 |0.444723 |0.340120 |790.000000

|741.000000 |85416.924520 |70484.500000 |64897.947475 |1194.786860

|2031.000000 |95623.665980 |84319.000000 |60297.436260 |689.158350

|1358.000000 |1990.646240 |1885.000000 |787.554270 |387.587270

|908.000000 |629.517360 |598.000000 |265.684575 |501.000000 |345.170125

|0.037270 |0.043150 |0.143860 |0.740560 |0.553320 |0.676590 |0.714090

|0.940100 |0.940183 |0.944935 |41.180250 |42.583330 |22.938350

|676.560290 |2696.000000 |43.496490 |45.333330 |23.549450 |685.883910

|2800.000000 |0.815235 |0.605660 |0.065340 |0.027910 |0.128400

|max |294333.000000 |NaN |140.0 |810.000000 |72.000000 |99929.000000

|989.000000 |64.804269 |-65.695344 |5.520166e+10 |1.212570e+10

|39454.000000 |27962.000000 |15466.000000 |3962.342290 |3972.000000

|1720.718990 |4112.12237 |7634.000000 |1.000000 |1.000000 |1.000000

|1.000000 |1.000000 |1.000000 |1.000000 |1.000000 |7634.000000

|7336.000000 |221622.723500 |242249.000000 |124534.013900 |8133.778720

|12316.000000 |242857.142900 |242720.000000 |105579.486100 |4888.944600

|6658.000000 |4462.342290 |4472.000000 |1814.113980 |1936.551660

|5033.000000 |1700.179110 |1702.000000 |782.862850 |3965.000000

|2878.131310 |1.000000 |1.000000 |1.000000 |1.000000 |1.000000 |1.000000

|1.000000 |1.000000 |1.000000 |1.000000 |83.358330 |83.333330 |27.920410

|12017.070440 |27962.000000 |90.107940 |90.166670 |29.626680

|4145.557870 |15466.000000 |1.000000 |1.000000 |0.714290 |0.714290

|0.362750

|===

----

+\*In[14]:\*+

[source, ipython3]

----

df\_train.info

----

+\*Out[14]:\*+

----<bound method DataFrame.info of UID BLOCKID SUMLEVEL COUNTYID STATEID state state\_ab \

0 267822 NaN 140 53 36 New York NY

1 246444 NaN 140 141 18 Indiana IN

2 245683 NaN 140 63 18 Indiana IN

3 279653 NaN 140 127 72 Puerto Rico PR

4 247218 NaN 140 161 20 Kansas KS

... ... ... ... ... ... ... ...

27316 279212 NaN 140 43 72 Puerto Rico PR

27317 277856 NaN 140 91 42 Pennsylvania PA

27318 233000 NaN 140 87 8 Colorado CO

27319 287425 NaN 140 439 48 Texas TX

27320 265371 NaN 140 3 32 Nevada NV

city place type primary zip\_code area\_code \

0 Hamilton Hamilton City tract 13346 315

1 South Bend Roseland City tract 46616 574

2 Danville Danville City tract 46122 317

3 San Juan Guaynabo Urban tract 927 787

4 Manhattan Manhattan City City tract 66502 785

... ... ... ... ... ... ...

27316 Coamo Coamo Urban tract 769 787

27317 Blue Bell Blue Bell Borough tract 19422 215

27318 Weldona Saddle Ridge City tract 80653 970

27319 Colleyville Colleyville City Town tract 76034 817

27320 Las Vegas Paradise City tract 89123 702

lat lng ALand AWater pop male\_pop \

0 42.840812 -75.501524 2.021834e+08 1699120 5230 2612

1 41.701441 -86.266614 1.560828e+06 100363 2633 1349

2 39.792202 -86.515246 6.956160e+07 284193 6881 3643

3 18.396103 -66.104169 1.105793e+06 0 2700 1141

4 39.195573 -96.569366 2.554403e+06 0 5637 2586

... ... ... ... ... ... ...

27316 18.076060 -66.358379 6.970300e+05 0 1847 909

27317 40.158138 -75.307271 5.077337e+06 11786 4155 2116

27318 40.410316 -103.814003 1.323262e+09 17577610 2829 1465

27319 32.904866 -97.162151 1.865230e+07 158882 11542 5727

27320 36.064754 -115.152237 7.796308e+06 0 3726 1815

female\_pop rent\_mean rent\_median rent\_stdev rent\_sample\_weight \

0 2618 769.38638 784.0 232.63967 272.34441

1 1284 804.87924 848.0 253.46747 312.58622

2 3238 742.77365 703.0 323.39011 291.85520

3 1559 803.42018 782.0 297.39258 259.30316

4 3051 938.56493 881.0 392.44096 1005.42886

... ... ... ... ... ...

27316 938 439.42839 419.0 140.29970 170.00000

27317 2039 1813.19253 1788.0 492.92300 64.84927

27318 1364 849.39107 834.0 336.47530 120.91448

27319 5815 1972.45746 1843.0 633.02173 19.16328

27320 1911 949.84199 924.0 198.82109 555.87526

rent\_samples rent\_gt\_10 rent\_gt\_15 rent\_gt\_20 rent\_gt\_25 \

0 362.0 0.86761 0.79155 0.59155 0.45634

1 513.0 0.97410 0.93227 0.69920 0.69920

2 378.0 0.95238 0.88624 0.79630 0.66667

3 368.0 0.94693 0.87151 0.69832 0.61732

4 1704.0 0.99286 0.98247 0.91688 0.84740

... ... ... ... ... ...

27316 170.0 1.00000 1.00000 1.00000 0.83333

27317 471.0 0.85435 0.63261 0.50000 0.37391

27318 195.0 0.93846 0.71282 0.54359 0.44615

27319 157.0 1.00000 1.00000 0.75796 0.61146

27320 1031.0 0.94956 0.87779 0.83705 0.63337

rent\_gt\_30 rent\_gt\_35 rent\_gt\_40 rent\_gt\_50 universe\_samples \

0 0.42817 0.18592 0.15493 0.12958 387

1 0.55179 0.41235 0.39044 0.27888 542

2 0.39153 0.39153 0.28307 0.15873 459

3 0.51397 0.46927 0.35754 0.32961 438

4 0.78247 0.60974 0.55455 0.44416 1725

... ... ... ... ... ...

27316 0.79012 0.79012 0.72222 0.62963 278

27317 0.30870 0.30870 0.26304 0.23478 484

27318 0.29744 0.23077 0.16923 0.09231 237

27319 0.50318 0.50318 0.27389 0.27389 234

27320 0.51115 0.41901 0.27934 0.10572 1055

used\_samples hi\_mean hi\_median hi\_stdev hi\_sample\_weight \

0 355 63125.28406 48120.0 49042.01206 1290.96240

1 502 41931.92593 35186.0 31639.50203 838.74664

2 378 84942.68317 74964.0 56811.62186 1155.20980

3 358 48733.67116 37845.0 45100.54010 928.32193

4 1540 31834.15466 22497.0 34046.50907 1548.67477

... ... ... ... ... ...

27316 162 18515.67021 13317.0 23914.42656 648.39020

27317 460 119889.08320 108284.0 77625.25547 518.53683

27318 195 79890.25113 73350.0 58132.65778 529.41812

27319 157 165510.27110 148548.0 102038.58810 960.70051

27320 1031 51648.18703 38072.0 46305.44046 1061.78932

hi\_samples family\_mean family\_median family\_stdev \

0 2024.0 67994.14790 53245.0 47667.30119

1 1127.0 50670.10337 43023.0 34715.57548

2 2488.0 95262.51431 85395.0 49292.67664

3 1267.0 56401.68133 44399.0 41082.90515

4 1983.0 54053.42396 50272.0 39609.12605

... ... ... ... ...

27316 774.0 20889.14617 16760.0 23488.17854

27317 1431.0 118896.06830 113313.0 66663.51722

27318 1077.0 88878.57034 81864.0 53510.48475

27319 4009.0 167148.83770 175952.0 77638.35136

27320 1409.0 54886.07827 42544.0 39352.40334

family\_sample\_weight family\_samples hc\_mortgage\_mean \

0 884.33516 1491.0 1414.80295

1 375.28798 554.0 864.41390

2 709.74925 1889.0 1506.06758

3 490.18479 729.0 1175.28642

4 244.08903 395.0 1192.58759

... ... ... ...

27316 346.58143 446.0 770.11560

27317 388.97237 1151.0 2210.84055

27318 375.21237 871.0 1671.07908

27319 719.65942 3452.0 3074.83088

27320 474.38547 698.0 1455.42340

hc\_mortgage\_median hc\_mortgage\_stdev hc\_mortgage\_sample\_weight \

0 1223.0 641.22898 377.83135

1 784.0 482.27020 316.88320

2 1361.0 731.89394 699.41354

3 1101.0 428.98751 261.28471

4 1125.0 327.49674 76.61052

... ... ... ...

27316 828.0 157.85227 58.00000

27317 2202.0 713.03361 154.95504

27318 1588.0 742.67822 182.53725

27319 3188.0 1121.07013 536.61873

27320 1364.0 629.41356 106.84849

hc\_mortgage\_samples hc\_mean hc\_median hc\_stdev hc\_samples \

0 867.0 570.01530 558.0 270.11299 770.0

1 356.0 351.98293 336.0 125.40457 229.0

2 1491.0 556.45986 532.0 184.42175 538.0

3 437.0 288.04047 247.0 185.55887 392.0

4 134.0 443.68855 444.0 76.12674 124.0

... ... ... ... ... ...

27316 58.0 160.86544 145.0 94.04517 438.0

27317 619.0 712.16631 663.0 281.77621 328.0

27318 488.0 536.04921 467.0 306.82251 352.0

27319 2481.0 1076.86881 1037.0 432.32205 1294.0

27320 232.0 540.26838 454.0 256.39951 122.0

hc\_sample\_weight home\_equity\_second\_mortgage second\_mortgage \

0 499.29293 0.01588 0.02077

1 189.60606 0.02222 0.02222

2 323.35354 0.00000 0.00000

3 314.90566 0.01086 0.01086

4 79.55556 0.05426 0.05426

... ... ... ...

27316 366.09000 0.00000 0.00000

27317 174.96970 0.00845 0.02112

27318 240.36899 0.02024 0.02024

27319 525.92451 0.05801 0.07550

27320 79.07071 0.01412 0.01412

home\_equity debt second\_mortgage\_cdf home\_equity\_cdf debt\_cdf \

0 0.08919 0.52963 0.43658 0.49087 0.73341

1 0.04274 0.60855 0.42174 0.70823 0.58120

2 0.09512 0.73484 1.00000 0.46332 0.28704

3 0.01086 0.52714 0.53057 0.82530 0.73727

4 0.05426 0.51938 0.18332 0.65545 0.74967

... ... ... ... ... ...

27316 0.00000 0.11694 1.00000 1.00000 0.98762

27317 0.19641 0.65364 0.43301 0.12376 0.47934

27318 0.07857 0.58095 0.44186 0.54095 0.63916

27319 0.12556 0.65722 0.10759 0.33196 0.47094

27320 0.18362 0.65537 0.50190 0.15035 0.47521

hs\_degree hs\_degree\_male hs\_degree\_female male\_age\_mean \

0 0.89288 0.85880 0.92434 42.48574

1 0.90487 0.86947 0.94187 34.84728

2 0.94288 0.94616 0.93952 39.38154

3 0.91500 0.90755 0.92043 48.64749

4 1.00000 1.00000 1.00000 26.07533

... ... ... ... ...

27316 0.60155 0.56962 0.63222 42.14011

27317 0.95737 0.95772 0.95701 37.75495

27318 0.93555 0.92200 0.94887 40.01134

27319 0.98540 0.98883 0.98201 40.75409

27320 0.87370 0.87166 0.87576 34.81046

male\_age\_median male\_age\_stdev male\_age\_sample\_weight \

0 44.00000 22.97306 696.42136

1 32.00000 20.37452 323.90204

2 40.83333 22.89769 888.29730

3 48.91667 23.05968 274.98956

4 22.41667 11.84399 1296.89877

... ... ... ...

27316 41.66667 23.76478 216.22207

27317 38.83333 21.45832 530.17185

27318 42.00000 23.08048 345.30911

27319 46.66667 22.48690 1305.28070

27320 32.50000 20.16376 461.52440

male\_age\_samples female\_age\_mean female\_age\_median female\_age\_stdev \

0 2612.0 44.48629 45.33333 22.51276

1 1349.0 36.48391 37.58333 23.43353

2 3643.0 42.15810 42.83333 23.94119

3 1141.0 47.77526 50.58333 24.32015

4 2586.0 24.17693 21.58333 11.10484

... ... ... ... ...

27316 909.0 42.73154 40.16667 24.79821

27317 2116.0 38.21269 39.50000 21.84826

27318 1465.0 43.40218 46.33333 23.40858

27319 5727.0 39.25921 43.41667 21.36235

27320 1815.0 34.45345 29.83333 19.77208

female\_age\_sample\_weight female\_age\_samples pct\_own married \

0 685.33845 2618.0 0.79046 0.57851

1 267.23367 1284.0 0.52483 0.34886

2 707.01963 3238.0 0.85331 0.64745

3 362.20193 1559.0 0.65037 0.47257

4 1854.48652 3051.0 0.13046 0.12356

... ... ... ... ...

27316 230.87898 938.0 0.60422 0.24603

27317 496.20427 2039.0 0.68072 0.61127

27318 316.52078 1364.0 0.78508 0.70451

27319 1373.94120 5815.0 0.93970 0.75503

27320 526.73261 1911.0 0.27912 0.34426

married\_snp separated divorced

0 0.01882 0.01240 0.08770

1 0.01426 0.01426 0.09030

2 0.02830 0.01607 0.10657

3 0.02021 0.02021 0.10106

4 0.00000 0.00000 0.03109

... ... ... ...

27316 0.03042 0.02249 0.14683

27317 0.05003 0.02473 0.04888

27318 0.01386 0.00520 0.07712

27319 0.02287 0.00915 0.05261

27320 0.03825 0.03005 0.13320

[27321 rows x 80 columns]>----

+\*In[15]:\*+

[source, ipython3]

----

df\_test.info

----

+\*Out[15]:\*+

----<bound method DataFrame.info of UID BLOCKID SUMLEVEL COUNTYID STATEID state state\_ab \

0 255504 NaN 140 163 26 Michigan MI

1 252676 NaN 140 1 23 Maine ME

2 276314 NaN 140 15 42 Pennsylvania PA

3 248614 NaN 140 231 21 Kentucky KY

4 286865 NaN 140 355 48 Texas TX

... ... ... ... ... ... ... ...

11704 238088 NaN 140 105 12 Florida FL

11705 242811 NaN 140 31 17 Illinois IL

11706 250127 NaN 140 9 25 Massachusetts MA

11707 241096 NaN 140 27 19 Iowa IA

11708 287763 NaN 140 453 48 Texas TX

city place type primary zip\_code \

0 Detroit Dearborn Heights City CDP tract 48239

1 Auburn Auburn City City tract 4210

2 Pine City Millerton Borough tract 14871

3 Monticello Monticello City City tract 42633

4 Corpus Christi Edroy Town tract 78410

... ... ... ... ... ...

11704 Lakeland Crystal Springs City tract 33810

11705 Chicago Chicago City Village tract 60609

11706 Lawrence Methuen Town City City tract 1841

11707 Carroll Carroll City City tract 51401

11708 Austin Sunset Valley City Town tract 78745

area\_code lat lng ALand AWater pop male\_pop \

0 313 42.346422 -83.252823 2711280 39555 3417 1479

1 207 44.100724 -70.257832 14778785 2705204 3796 1846

2 607 41.948556 -76.783808 258903666 863840 3944 2065

3 606 36.746009 -84.766870 501694825 2623067 2508 1427

4 361 27.882461 -97.678586 13796057 497689 6230 3274

... ... ... ... ... ... ... ...

11704 863 28.226068 -82.068886 92582775 1166617 5611 2697

11705 773 41.804936 -87.667304 327029 0 2695 1504

11706 978 42.737778 -71.131761 5225804 393810 7392 3669

11707 712 42.081366 -94.866175 11066759 0 5945 2732

11708 512 30.219013 -97.774728 1990126 0 4117 2070

female\_pop rent\_mean rent\_median rent\_stdev rent\_sample\_weight \

0 1938 858.57169 859.0 232.39082 276.07497

1 1950 832.68625 750.0 267.22342 183.32299

2 1879 816.00639 755.0 416.25699 141.39063

3 1081 418.68937 385.0 156.92024 88.95960

4 2956 1031.63763 997.0 326.76727 277.39844

... ... ... ... ... ...

11704 2914 1458.82449 1603.0 566.90682 29.43733

11705 1191 700.53513 661.0 254.66700 480.86455

11706 3723 1069.70567 1138.0 488.13975 207.29615

11707 3213 696.93368 576.0 595.16228 503.83775

11708 2047 950.09294 864.0 333.82364 417.07457

rent\_samples rent\_gt\_10 rent\_gt\_15 rent\_gt\_20 rent\_gt\_25 \

0 424.0 1.00000 0.95696 0.85316 0.85316

1 245.0 1.00000 1.00000 0.86611 0.67364

2 217.0 0.97573 0.93204 0.78641 0.71845

3 93.0 1.00000 0.93548 0.93548 0.64516

4 624.0 0.72276 0.66506 0.53526 0.38301

... ... ... ... ... ...

11704 99.0 1.00000 1.00000 1.00000 0.62626

11705 592.0 1.00000 0.90034 0.85911 0.63058

11706 506.0 0.85375 0.83004 0.77273 0.56324

11707 590.0 0.96886 0.92042 0.83045 0.69723

11708 675.0 1.00000 0.97481 0.86074 0.73926

rent\_gt\_30 rent\_gt\_35 rent\_gt\_40 rent\_gt\_50 universe\_samples \

0 0.85316 0.85316 0.76962 0.63544 435

1 0.30962 0.30962 0.30962 0.27197 275

2 0.63592 0.47573 0.43689 0.32524 245

3 0.55914 0.46237 0.46237 0.36559 153

4 0.18910 0.16667 0.14263 0.11058 660

... ... ... ... ... ...

11704 0.62626 0.35354 0.18182 0.09091 147

11705 0.53952 0.41237 0.35223 0.19931 618

11706 0.47431 0.33399 0.30237 0.02569 539

11707 0.62284 0.43772 0.33737 0.33737 663

11708 0.44593 0.38370 0.27852 0.25778 682

used\_samples hi\_mean hi\_median hi\_stdev hi\_sample\_weight \

0 395 48899.52121 38746.0 44392.20902 798.02401

1 239 72335.33234 61008.0 51895.81159 922.82969

2 206 58501.15901 51648.0 45245.27248 893.07759

3 93 38237.55059 31612.0 34527.61607 775.17947

4 624 114456.07790 94211.0 81950.95692 836.30759

... ... ... ... ... ...

11704 99 57723.48180 48192.0 41301.62188 1636.68434

11705 582 35249.76522 27396.0 28889.72217 683.94534

11706 506 89549.15374 75357.0 66560.76837 1339.55365

11707 578 57877.26387 41838.0 49745.93715 1605.79897

11708 675 58006.33817 44179.0 49189.98590 902.67611

hi\_samples family\_mean family\_median family\_stdev \

0 1180.0 53802.87122 45167.0 43756.56479

1 1722.0 85642.22095 74759.0 49156.72870

2 1461.0 65694.06582 57186.0 44239.31893

3 957.0 44156.38709 34687.0 34899.74300

4 2404.0 123527.02420 103898.0 72173.55823

... ... ... ... ...

11704 2496.0 70786.81912 59194.0 40582.36046

11705 838.0 38912.54156 32554.0 29796.19973

11706 2739.0 99484.96572 89050.0 62721.62266

11707 2596.0 75066.29009 72135.0 47200.66016

11708 1396.0 54913.24441 42469.0 41016.08651

family\_sample\_weight family\_samples hc\_mortgage\_mean \

0 464.30972 769.0 1139.24548

1 482.99945 1147.0 1533.25988

2 619.73962 1084.0 1254.54462

3 535.21987 689.0 862.65763

4 507.42257 1738.0 1996.41425

... ... ... ...

11704 945.85894 1685.0 1269.83033

11705 415.51917 555.0 1406.83478

11706 853.61856 1986.0 1791.63902

11707 782.93088 1568.0 1182.30365

11708 581.04758 877.0 1364.17379

hc\_mortgage\_median hc\_mortgage\_stdev hc\_mortgage\_sample\_weight \

0 1109.0 336.47710 262.67011

1 1438.0 536.61118 373.96188

2 1089.0 596.85204 340.45884

3 749.0 624.42157 299.56752

4 1907.0 740.21168 319.97570

... ... ... ...

11704 1119.0 689.35735 608.62709

11705 1224.0 621.89533 62.54709

11706 1794.0 656.68467 548.16568

11707 1059.0 587.01032 796.11244

11708 1318.0 463.57052 217.49287

hc\_mortgage\_samples hc\_mean hc\_median hc\_stdev hc\_samples \

0 474.0 488.51323 436.0 192.75147 271.0

1 937.0 661.31296 668.0 201.31365 510.0

2 552.0 397.44466 356.0 189.40372 664.0

3 337.0 200.88113 180.0 91.56490 467.0

4 1102.0 867.57713 804.0 376.20236 642.0

... ... ... ... ... ...

11704 1024.0 536.66053 500.0 267.25752 1325.0

11705 139.0 487.66419 465.0 220.16444 81.0

11706 1634.0 654.78088 612.0 256.84182 566.0

11707 1267.0 369.29903 334.0 133.20792 666.0

11708 456.0 550.78197 555.0 199.13527 258.0

hc\_sample\_weight home\_equity\_second\_mortgage second\_mortgage \

0 189.18182 0.06443 0.06443

1 279.69697 0.01175 0.01175

2 534.16737 0.01069 0.01316

3 454.85404 0.00995 0.00995

4 333.91919 0.00000 0.00000

... ... ... ...

11704 914.89899 0.02043 0.03619

11705 47.09727 0.05909 0.05909

11706 299.83838 0.02727 0.02727

11707 556.40404 0.03570 0.03570

11708 163.55556 0.00000 0.00000

home\_equity debt second\_mortgage\_cdf home\_equity\_cdf debt\_cdf \

0 0.07651 0.63624 0.14111 0.55087 0.51965

1 0.14375 0.64755 0.52310 0.26442 0.49359

2 0.06497 0.45395 0.51066 0.60484 0.83848

3 0.01741 0.41915 0.53770 0.80931 0.87403

4 0.03440 0.63188 1.00000 0.74519 0.52943

... ... ... ... ... ...

11704 0.04044 0.43593 0.29592 0.71860 0.85762

11705 0.08182 0.63182 0.16199 0.52552 0.52957

11706 0.13545 0.74273 0.37297 0.29411 0.26972

11707 0.07967 0.65546 0.30010 0.53579 0.47507

11708 0.05042 0.63866 1.00000 0.67315 0.51407

hs\_degree hs\_degree\_male hs\_degree\_female male\_age\_mean \

0 0.91047 0.92010 0.90391 33.37131

1 0.94290 0.92832 0.95736 43.88680

2 0.89238 0.86003 0.92463 39.81661

3 0.60908 0.56584 0.65947 41.81638

4 0.86297 0.87969 0.84466 42.13301

... ... ... ... ...

11704 0.92097 0.95007 0.89480 51.03535

11705 0.54890 0.49817 0.60965 32.94145

11706 0.94057 0.94000 0.94105 35.85743

11707 0.91407 0.92428 0.90634 39.18219

11708 0.78685 0.80615 0.76820 35.56404

male\_age\_median male\_age\_stdev male\_age\_sample\_weight \

0 27.83333 22.36768 334.30978

1 46.08333 22.90302 427.10824

2 41.91667 24.29111 499.10080

3 43.00000 24.65325 333.57733

4 43.75000 22.69502 833.57435

... ... ... ...

11704 55.50000 22.41099 704.65208

11705 29.83333 20.52061 408.44261

11706 34.91667 22.49430 880.48254

11707 40.25000 24.86317 636.20201

11708 35.00000 21.67509 522.45931

male\_age\_samples female\_age\_mean female\_age\_median female\_age\_stdev \

0 1479.0 34.78682 33.75000 21.58531

1 1846.0 44.23451 46.66667 22.37036

2 2065.0 41.62426 44.50000 22.86213

3 1427.0 44.81200 48.00000 21.03155

4 3274.0 40.66618 42.66667 21.30900

... ... ... ... ...

11704 2697.0 53.51255 59.58333 23.23426

11705 1504.0 33.14169 32.83333 20.24698

11706 3669.0 43.53905 43.66667 23.17995

11707 2732.0 45.63179 48.16667 24.84209

11708 2070.0 35.99955 35.41667 20.68049

female\_age\_sample\_weight female\_age\_samples pct\_own married \

0 416.48097 1938.0 0.70252 0.28217

1 532.03505 1950.0 0.85128 0.64221

2 453.11959 1879.0 0.81897 0.59961

3 263.94320 1081.0 0.84609 0.56953

4 709.90829 2956.0 0.79077 0.57620

... ... ... ... ...

11704 699.33353 2914.0 0.93121 0.65969

11705 306.63915 1191.0 0.33122 0.42882

11706 900.13903 3723.0 0.84372 0.50269

11707 693.82905 3213.0 0.83330 0.66699

11708 559.30291 2047.0 0.52587 0.51922

married\_snp separated divorced

0 0.05910 0.03813 0.14299

1 0.02338 0.00000 0.13377

2 0.01746 0.01358 0.10026

3 0.05492 0.04694 0.12489

4 0.01726 0.00588 0.16379

... ... ... ...

11704 0.02135 0.02135 0.08780

11705 0.07781 0.02829 0.05305

11706 0.00108 0.00108 0.07294

11707 0.02738 0.00000 0.04694

11708 0.08066 0.02520 0.10586

[11709 rows x 80 columns]>----

+\*In[16]:\*+

[source, ipython3]

----

df\_train.info()

----

+\*Out[16]:\*+

----

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 27321 entries, 0 to 27320

Data columns (total 80 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 UID 27321 non-null int64

1 BLOCKID 0 non-null float64

2 SUMLEVEL 27321 non-null int64

3 COUNTYID 27321 non-null int64

4 STATEID 27321 non-null int64

5 state 27321 non-null object

6 state\_ab 27321 non-null object

7 city 27321 non-null object

8 place 27321 non-null object

9 type 27321 non-null object

10 primary 27321 non-null object

11 zip\_code 27321 non-null int64

12 area\_code 27321 non-null int64

13 lat 27321 non-null float64

14 lng 27321 non-null float64

15 ALand 27321 non-null float64

16 AWater 27321 non-null int64

17 pop 27321 non-null int64

18 male\_pop 27321 non-null int64

19 female\_pop 27321 non-null int64

20 rent\_mean 27007 non-null float64

21 rent\_median 27007 non-null float64

22 rent\_stdev 27007 non-null float64

23 rent\_sample\_weight 27007 non-null float64

24 rent\_samples 27007 non-null float64

25 rent\_gt\_10 27007 non-null float64

26 rent\_gt\_15 27007 non-null float64

27 rent\_gt\_20 27007 non-null float64

28 rent\_gt\_25 27007 non-null float64

29 rent\_gt\_30 27007 non-null float64

30 rent\_gt\_35 27007 non-null float64

31 rent\_gt\_40 27007 non-null float64

32 rent\_gt\_50 27007 non-null float64

33 universe\_samples 27321 non-null int64

34 used\_samples 27321 non-null int64

35 hi\_mean 27053 non-null float64

36 hi\_median 27053 non-null float64

37 hi\_stdev 27053 non-null float64

38 hi\_sample\_weight 27053 non-null float64

39 hi\_samples 27053 non-null float64

40 family\_mean 27023 non-null float64

41 family\_median 27023 non-null float64

42 family\_stdev 27023 non-null float64

43 family\_sample\_weight 27023 non-null float64

44 family\_samples 27023 non-null float64

45 hc\_mortgage\_mean 26748 non-null float64

46 hc\_mortgage\_median 26748 non-null float64

47 hc\_mortgage\_stdev 26748 non-null float64

48 hc\_mortgage\_sample\_weight 26748 non-null float64

49 hc\_mortgage\_samples 26748 non-null float64

50 hc\_mean 26721 non-null float64

51 hc\_median 26721 non-null float64

52 hc\_stdev 26721 non-null float64

53 hc\_samples 26721 non-null float64

54 hc\_sample\_weight 26721 non-null float64

55 home\_equity\_second\_mortgage 26864 non-null float64

56 second\_mortgage 26864 non-null float64

57 home\_equity 26864 non-null float64

58 debt 26864 non-null float64

59 second\_mortgage\_cdf 26864 non-null float64

60 home\_equity\_cdf 26864 non-null float64

61 debt\_cdf 26864 non-null float64

62 hs\_degree 27131 non-null float64

63 hs\_degree\_male 27121 non-null float64

64 hs\_degree\_female 27098 non-null float64

65 male\_age\_mean 27132 non-null float64

66 male\_age\_median 27132 non-null float64

67 male\_age\_stdev 27132 non-null float64

68 male\_age\_sample\_weight 27132 non-null float64

69 male\_age\_samples 27132 non-null float64

70 female\_age\_mean 27115 non-null float64

71 female\_age\_median 27115 non-null float64

72 female\_age\_stdev 27115 non-null float64

73 female\_age\_sample\_weight 27115 non-null float64

74 female\_age\_samples 27115 non-null float64

75 pct\_own 27053 non-null float64

76 married 27130 non-null float64

77 married\_snp 27130 non-null float64

78 separated 27130 non-null float64

79 divorced 27130 non-null float64

dtypes: float64(62), int64(12), object(6)

memory usage: 16.7+ MB

----

+\*In[17]:\*+

[source, ipython3]

----

df\_test.info()

----

+\*Out[17]:\*+

----

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 11709 entries, 0 to 11708

Data columns (total 80 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 UID 11709 non-null int64

1 BLOCKID 0 non-null float64

2 SUMLEVEL 11709 non-null int64

3 COUNTYID 11709 non-null int64

4 STATEID 11709 non-null int64

5 state 11709 non-null object

6 state\_ab 11709 non-null object

7 city 11709 non-null object

8 place 11709 non-null object

9 type 11709 non-null object

10 primary 11709 non-null object

11 zip\_code 11709 non-null int64

12 area\_code 11709 non-null int64

13 lat 11709 non-null float64

14 lng 11709 non-null float64

15 ALand 11709 non-null int64

16 AWater 11709 non-null int64

17 pop 11709 non-null int64

18 male\_pop 11709 non-null int64

19 female\_pop 11709 non-null int64

20 rent\_mean 11561 non-null float64

21 rent\_median 11561 non-null float64

22 rent\_stdev 11561 non-null float64

23 rent\_sample\_weight 11561 non-null float64

24 rent\_samples 11561 non-null float64

25 rent\_gt\_10 11560 non-null float64

26 rent\_gt\_15 11560 non-null float64

27 rent\_gt\_20 11560 non-null float64

28 rent\_gt\_25 11560 non-null float64

29 rent\_gt\_30 11560 non-null float64

30 rent\_gt\_35 11560 non-null float64

31 rent\_gt\_40 11560 non-null float64

32 rent\_gt\_50 11560 non-null float64

33 universe\_samples 11709 non-null int64

34 used\_samples 11709 non-null int64

35 hi\_mean 11587 non-null float64

36 hi\_median 11587 non-null float64

37 hi\_stdev 11587 non-null float64

38 hi\_sample\_weight 11587 non-null float64

39 hi\_samples 11587 non-null float64

40 family\_mean 11573 non-null float64

41 family\_median 11573 non-null float64

42 family\_stdev 11573 non-null float64

43 family\_sample\_weight 11573 non-null float64

44 family\_samples 11573 non-null float64

45 hc\_mortgage\_mean 11441 non-null float64

46 hc\_mortgage\_median 11441 non-null float64

47 hc\_mortgage\_stdev 11441 non-null float64

48 hc\_mortgage\_sample\_weight 11441 non-null float64

49 hc\_mortgage\_samples 11441 non-null float64

50 hc\_mean 11419 non-null float64

51 hc\_median 11419 non-null float64

52 hc\_stdev 11419 non-null float64

53 hc\_samples 11419 non-null float64

54 hc\_sample\_weight 11419 non-null float64

55 home\_equity\_second\_mortgage 11489 non-null float64

56 second\_mortgage 11489 non-null float64

57 home\_equity 11489 non-null float64

58 debt 11489 non-null float64

59 second\_mortgage\_cdf 11489 non-null float64

60 home\_equity\_cdf 11489 non-null float64

61 debt\_cdf 11489 non-null float64

62 hs\_degree 11624 non-null float64

63 hs\_degree\_male 11620 non-null float64

64 hs\_degree\_female 11604 non-null float64

65 male\_age\_mean 11625 non-null float64

66 male\_age\_median 11625 non-null float64

67 male\_age\_stdev 11625 non-null float64

68 male\_age\_sample\_weight 11625 non-null float64

69 male\_age\_samples 11625 non-null float64

70 female\_age\_mean 11613 non-null float64

71 female\_age\_median 11613 non-null float64

72 female\_age\_stdev 11613 non-null float64

73 female\_age\_sample\_weight 11613 non-null float64

74 female\_age\_samples 11613 non-null float64

75 pct\_own 11587 non-null float64

76 married 11625 non-null float64

77 married\_snp 11625 non-null float64

78 separated 11625 non-null float64

79 divorced 11625 non-null float64

dtypes: float64(61), int64(13), object(6)

memory usage: 7.1+ MB

----

+\*In[18]:\*+

[source, ipython3]

----

len(df\_train["UID"].unique())

----

+\*Out[18]:\*+

----27161----

+\*In[19]:\*+

[source, ipython3]

----

df\_train[df\_train.duplicated()]

----

+\*Out[19]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |BLOCKID |SUMLEVEL |COUNTYID |STATEID |state |state\_ab |city

|place |type |primary |zip\_code |area\_code |lat |lng |ALand |AWater |pop

|male\_pop |female\_pop |rent\_mean |rent\_median |rent\_stdev

|rent\_sample\_weight |rent\_samples |rent\_gt\_10 |rent\_gt\_15 |rent\_gt\_20

|rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35 |rent\_gt\_40 |rent\_gt\_50

|universe\_samples |used\_samples |hi\_mean |hi\_median |hi\_stdev

|hi\_sample\_weight |hi\_samples |family\_mean |family\_median |family\_stdev

|family\_sample\_weight |family\_samples |hc\_mortgage\_mean

|hc\_mortgage\_median |hc\_mortgage\_stdev |hc\_mortgage\_sample\_weight

|hc\_mortgage\_samples |hc\_mean |hc\_median |hc\_stdev |hc\_samples

|hc\_sample\_weight |home\_equity\_second\_mortgage |second\_mortgage

|home\_equity |debt |second\_mortgage\_cdf |home\_equity\_cdf |debt\_cdf

|hs\_degree |hs\_degree\_male |hs\_degree\_female |male\_age\_mean

|male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|1623 |230058 |NaN |140 |73 |6 |California |CA |Oceanside |Camp

Pendleton North |City |tract |92058 |760 |33.380238 |-117.390100

|548969793.0 |10394327 |39454 |27962 |11492 |1783.71119 |1829.0

|591.18192 |1216.05870 |6281.0 |1.00000 |0.95061 |0.88989 |0.88021

|0.81638 |0.74056 |0.67345 |0.47407 |6648 |6094 |49895.41569 |41615.0

|33995.24006 |4991.10244 |6661.0 |49650.75721 |41981.0 |30501.50747

|4888.94460 |6603.0 |NaN |NaN |NaN |NaN |NaN |53.59461 |53.0 |24.62293

|13.0 |6.37000 |0.00000 |0.00000 |0.00000 |0.00000 |1.00000 |1.00000

|1.0000 |0.98810 |0.99079 |0.98418 |20.48470 |21.50000 |8.28680

|12017.07044 |27962.0 |19.99315 |22.41667 |11.62088 |3406.53918 |11492.0

|0.00107 |0.33566 |0.07245 |0.00250 |0.01032

|1907 |292484 |NaN |140 |25 |55 |Wisconsin |WI |Madison |Madison City

|City |tract |53703 |608 |43.073403 |-89.395430 |175748.0 |0 |3274 |1293

|1981 |1191.78679 |956.0 |737.36689 |768.38091 |1497.0 |1.00000 |0.98039

|0.96757 |0.96757 |0.96154 |0.90422 |0.84992 |0.78733 |1505 |1326

|14917.36079 |9758.0 |15534.75750 |1202.89330 |1505.0 |22527.10680

|7965.0 |21141.68686 |23.49730 |37.0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |0.87107 |0.89950

|0.82353 |24.13033 |22.08333 |7.84307 |755.14608 |1293.0 |22.03226

|21.08333 |5.13435 |1365.86300 |1981.0 |0.00000 |0.00773 |0.00000

|0.00000 |0.01160

|2447 |268401 |NaN |140 |61 |36 |New York |NY |Long Island City |New

York City |City |tract |11101 |718 |40.747948 |-73.962778 |106753.0

|317001 |0 |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN

|4161 |284060 |NaN |140 |113 |48 |Texas |TX |Dallas |University Park

City |Town |tract |75231 |972 |32.861585 |-96.767449 |510363.0 |0 |1909

|843 |1066 |1213.02668 |1203.0 |290.83832 |316.85951 |1038.0 |0.95922

|0.73689 |0.45534 |0.36990 |0.23204 |0.17864 |0.14757 |0.13010 |1047

|1030 |79270.10248 |68038.0 |55756.27549 |483.19883 |1055.0

|100978.20690 |90224.0 |55196.10143 |105.55705 |324.0 |1147.00000

|1143.0 |77.95565 |3.18072 |8.0 |NaN |NaN |NaN |NaN |NaN |0.00000

|0.00000 |0.00000 |1.00000 |1.00000 |1.00000 |0.0000 |0.99032 |0.97895

|1.00000 |35.42484 |31.83333 |16.43998 |201.07788 |843.0 |35.57082

|31.00000 |14.89626 |248.71488 |1066.0 |0.00419 |0.39327 |0.01423

|0.00000 |0.07245

|5066 |274254 |NaN |140 |109 |40 |Oklahoma |OK |Oklahoma City |Oklahoma

City City |CDP |tract |73102 |405 |35.467896 |-97.519872 |502382.0 |0

|231 |118 |113 |1271.16422 |1162.0 |606.43543 |63.90225 |167.0 |0.85802

|0.74691 |0.53086 |0.36420 |0.26543 |0.23457 |0.09259 |0.06790 |167 |162

|81877.33198 |70115.0 |70199.06235 |72.14662 |167.0 |118098.41760

|78761.0 |82478.99612 |6.19967 |15.0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |0.97685 |0.95690

|1.00000 |35.34160 |30.91667 |13.04167 |23.62712 |118.0 |36.16616

|29.83333 |13.14478 |23.22171 |113.0 |0.00000 |0.22881 |0.10169 |0.00000

|0.17797

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |...

|26769 |252187 |NaN |140 |33 |24 |Maryland |MD |Morningside |Andrews Afb

|CDP |tract |20746 |301 |38.809803 |-76.869158 |17906686.0 |91637 |3226

|1539 |1687 |2280.98449 |2269.0 |510.64895 |88.78295 |870.0 |1.00000

|0.97816 |0.84483 |0.76667 |0.68736 |0.65402 |0.53908 |0.41839 |897 |870

|82662.88655 |60463.0 |67873.52085 |494.61043 |897.0 |84291.31660

|64987.0 |62763.07723 |412.58421 |779.0 |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |0.97069 |0.98204

|0.96010 |22.72053 |23.25000 |14.95983 |370.62341 |1539.0 |21.92741

|22.50000 |15.50144 |375.14523 |1687.0 |0.00000 |0.78735 |0.06761

|0.01200 |0.00872

|26872 |293566 |NaN |140 |133 |55 |Wisconsin |WI |Brookfield |Pewaukee

City |City |tract |53045 |262 |43.078027 |-88.182293 |13727249.0 |294858

|4696 |2272 |2424 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |8 |0 |129992.83710 |109988.0 |81681.78977 |475.25565 |1644.0

|133057.97610 |118006.0 |69640.11261 |359.62834 |1437.0 |2045.53494

|1926.0 |807.56153 |328.93201 |1186.0 |768.66947 |703.0 |283.19072

|450.0 |216.02871 |0.02812 |0.04218 |0.14914 |0.72494 |0.25256 |0.24632

|0.3093 |0.98027 |0.96891 |0.99134 |42.21981 |46.83333 |22.88785

|568.09867 |2272.0 |39.92907 |44.33333 |22.25252 |593.35393 |2424.0

|0.99468 |0.77148 |0.00561 |0.00000 |0.02021

|26910 |222470 |NaN |140 |11 |4 |Arizona |AZ |Morenci |Clifton |CDP

|tract |85540 |928 |32.992986 |-109.332389 |319216540.0 |7585923 |3484

|1774 |1710 |485.41250 |402.0 |294.21131 |1060.68993 |1130.0 |0.30419

|0.18822 |0.16414 |0.10259 |0.09456 |0.07315 |0.05263 |0.04282 |1135

|1121 |76747.54126 |71812.0 |42368.41300 |521.58236 |1140.0 |81709.38475

|79242.0 |42554.58414 |332.87604 |782.0 |NaN |NaN |NaN |NaN |NaN

|549.50000 |549.0 |36.51484 |5.0 |2.47475 |0.00000 |0.00000 |0.00000

|0.00000 |1.00000 |1.00000 |1.0000 |0.89761 |0.92466 |0.86900 |28.81256

|28.16667 |18.06907 |402.01514 |1774.0 |28.24603 |27.83333 |17.42918

|392.61849 |1710.0 |0.00517 |0.46198 |0.01717 |0.00082 |0.21668

|27175 |235725 |NaN |140 |57 |12 |Florida |FL |Tampa |Pebble Creek |City

|tract |33647 |813 |28.149447 |-82.356517 |2464707.0 |50567 |1105 |457

|648 |1144.84723 |1135.0 |204.04871 |133.90349 |455.0 |1.00000 |0.90549

|0.64176 |0.57143 |0.30330 |0.15824 |0.12967 |0.11868 |455 |455

|59561.19893 |57254.0 |25207.50919 |246.98533 |455.0 |63367.26166

|67142.0 |24330.79882 |95.26054 |204.0 |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |0.99474 |0.98817

|1.00000 |32.09070 |29.75000 |16.72733 |94.84358 |457.0 |29.08800

|28.08333 |14.65116 |144.78344 |648.0 |0.00000 |0.25806 |0.10753

|0.10753 |0.09946

|27176 |247777 |NaN |140 |61 |21 |Kentucky |KY |Brownsville |Brownsville

City |City |tract |42210 |270 |37.197158 |-86.156329 |175916489.0

|2094374 |17 |11 |6 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |19.60934 |19.75000 |0.96277 |7.86957

|11.0 |19.39847 |19.00000 |1.49474 |3.39130 |6.0 |NaN |0.00000 |0.00000

|0.00000 |0.00000

|===

160 rows × 80 columns

----

+\*In[20]:\*+

[source, ipython3]

----

df\_train["UID"].isnull().sum

----

+\*Out[20]:\*+

----<bound method Series.sum of 0 False

1 False

2 False

3 False

4 False

...

27316 False

27317 False

27318 False

27319 False

27320 False

Name: UID, Length: 27321, dtype: bool>----

+\*In[21]:\*+

[source, ipython3]

----

df\_train.isnull().sum(axis=0)[0:90]

----

+\*Out[21]:\*+

----UID 0

BLOCKID 27321

SUMLEVEL 0

COUNTYID 0

STATEID 0

state 0

state\_ab 0

city 0

place 0

type 0

primary 0

zip\_code 0

area\_code 0

lat 0

lng 0

ALand 0

AWater 0

pop 0

male\_pop 0

female\_pop 0

rent\_mean 314

rent\_median 314

rent\_stdev 314

rent\_sample\_weight 314

rent\_samples 314

rent\_gt\_10 314

rent\_gt\_15 314

rent\_gt\_20 314

rent\_gt\_25 314

rent\_gt\_30 314

rent\_gt\_35 314

rent\_gt\_40 314

rent\_gt\_50 314

universe\_samples 0

used\_samples 0

hi\_mean 268

hi\_median 268

hi\_stdev 268

hi\_sample\_weight 268

hi\_samples 268

family\_mean 298

family\_median 298

family\_stdev 298

family\_sample\_weight 298

family\_samples 298

hc\_mortgage\_mean 573

hc\_mortgage\_median 573

hc\_mortgage\_stdev 573

hc\_mortgage\_sample\_weight 573

hc\_mortgage\_samples 573

hc\_mean 600

hc\_median 600

hc\_stdev 600

hc\_samples 600

hc\_sample\_weight 600

home\_equity\_second\_mortgage 457

second\_mortgage 457

home\_equity 457

debt 457

second\_mortgage\_cdf 457

home\_equity\_cdf 457

debt\_cdf 457

hs\_degree 190

hs\_degree\_male 200

hs\_degree\_female 223

male\_age\_mean 189

male\_age\_median 189

male\_age\_stdev 189

male\_age\_sample\_weight 189

male\_age\_samples 189

female\_age\_mean 206

female\_age\_median 206

female\_age\_stdev 206

female\_age\_sample\_weight 206

female\_age\_samples 206

pct\_own 268

married 191

married\_snp 191

separated 191

divorced 191

dtype: int64----

+\*In[22]:\*+

[source, ipython3]

----

#removing duplicates

train\_df = df\_train.drop\_duplicates()

----

+\*In[23]:\*+

[source, ipython3]

----

train\_df[train\_df.duplicated()]

----

+\*Out[23]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |BLOCKID |SUMLEVEL |COUNTYID |STATEID |state |state\_ab |city

|place |type |primary |zip\_code |area\_code |lat |lng |ALand |AWater |pop

|male\_pop |female\_pop |rent\_mean |rent\_median |rent\_stdev

|rent\_sample\_weight |rent\_samples |rent\_gt\_10 |rent\_gt\_15 |rent\_gt\_20

|rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35 |rent\_gt\_40 |rent\_gt\_50

|universe\_samples |used\_samples |hi\_mean |hi\_median |hi\_stdev

|hi\_sample\_weight |hi\_samples |family\_mean |family\_median |family\_stdev

|family\_sample\_weight |family\_samples |hc\_mortgage\_mean

|hc\_mortgage\_median |hc\_mortgage\_stdev |hc\_mortgage\_sample\_weight

|hc\_mortgage\_samples |hc\_mean |hc\_median |hc\_stdev |hc\_samples

|hc\_sample\_weight |home\_equity\_second\_mortgage |second\_mortgage

|home\_equity |debt |second\_mortgage\_cdf |home\_equity\_cdf |debt\_cdf

|hs\_degree |hs\_degree\_male |hs\_degree\_female |male\_age\_mean

|male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|===

----

+\*In[24]:\*+

[source, ipython3]

----

train\_df["UID"].isnull().sum(axis=0)

----

+\*Out[24]:\*+

----0----

+\*In[25]:\*+

[source, ipython3]

----

len(train\_df["UID"].unique())

----

+\*Out[25]:\*+

----27161----

# 2. After dropping duplicate rows. we identify UID has all unique values with no null values. hence we can consider UID as the primary key.

+\*In[26]:\*+

[source, ipython3]

----

train\_df.isnull().sum(axis=0)[0:90]

----

+\*Out[26]:\*+

----UID 0

BLOCKID 27161

SUMLEVEL 0

COUNTYID 0

STATEID 0

state 0

state\_ab 0

city 0

place 0

type 0

primary 0

zip\_code 0

area\_code 0

lat 0

lng 0

ALand 0

AWater 0

pop 0

male\_pop 0

female\_pop 0

rent\_mean 242

rent\_median 242

rent\_stdev 242

rent\_sample\_weight 242

rent\_samples 242

rent\_gt\_10 242

rent\_gt\_15 242

rent\_gt\_20 242

rent\_gt\_25 242

rent\_gt\_30 242

rent\_gt\_35 242

rent\_gt\_40 242

rent\_gt\_50 242

universe\_samples 0

used\_samples 0

hi\_mean 207

hi\_median 207

hi\_stdev 207

hi\_sample\_weight 207

hi\_samples 207

family\_mean 230

family\_median 230

family\_stdev 230

family\_sample\_weight 230

family\_samples 230

hc\_mortgage\_mean 442

hc\_mortgage\_median 442

hc\_mortgage\_stdev 442

hc\_mortgage\_sample\_weight 442

hc\_mortgage\_samples 442

hc\_mean 478

hc\_median 478

hc\_stdev 478

hc\_samples 478

hc\_sample\_weight 478

home\_equity\_second\_mortgage 360

second\_mortgage 360

home\_equity 360

debt 360

second\_mortgage\_cdf 360

home\_equity\_cdf 360

debt\_cdf 360

hs\_degree 145

hs\_degree\_male 154

hs\_degree\_female 171

male\_age\_mean 148

male\_age\_median 148

male\_age\_stdev 148

male\_age\_sample\_weight 148

male\_age\_samples 148

female\_age\_mean 161

female\_age\_median 161

female\_age\_stdev 161

female\_age\_sample\_weight 161

female\_age\_samples 161

pct\_own 207

married 150

married\_snp 150

separated 150

divorced 150

dtype: int64----

+\*In[27]:\*+

[source, ipython3]

----

#checking duplicate rows in test dataset

df\_test[df\_test.duplicated()]

----

+\*Out[27]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |BLOCKID |SUMLEVEL |COUNTYID |STATEID |state |state\_ab |city

|place |type |primary |zip\_code |area\_code |lat |lng |ALand |AWater |pop

|male\_pop |female\_pop |rent\_mean |rent\_median |rent\_stdev

|rent\_sample\_weight |rent\_samples |rent\_gt\_10 |rent\_gt\_15 |rent\_gt\_20

|rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35 |rent\_gt\_40 |rent\_gt\_50

|universe\_samples |used\_samples |hi\_mean |hi\_median |hi\_stdev

|hi\_sample\_weight |hi\_samples |family\_mean |family\_median |family\_stdev

|family\_sample\_weight |family\_samples |hc\_mortgage\_mean

|hc\_mortgage\_median |hc\_mortgage\_stdev |hc\_mortgage\_sample\_weight

|hc\_mortgage\_samples |hc\_mean |hc\_median |hc\_stdev |hc\_samples

|hc\_sample\_weight |home\_equity\_second\_mortgage |second\_mortgage

|home\_equity |debt |second\_mortgage\_cdf |home\_equity\_cdf |debt\_cdf

|hs\_degree |hs\_degree\_male |hs\_degree\_female |male\_age\_mean

|male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|460 |265339 |NaN |140 |3 |32 |Nevada |NV |Las Vegas |Winchester |City

|tract |89119 |702 |36.111448 |-115.159125 |612803 |0 |724 |490 |234

|751.50703 |736.0 |172.34699 |446.71620 |501.0 |0.95781 |0.91772

|0.87764 |0.77637 |0.57806 |0.45992 |0.43460 |0.29325 |501 |474

|32947.89919 |24961.0 |37213.92684 |434.21963 |501.0 |27064.98901

|27879.0 |16391.07515 |58.49710 |73.0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |0.89028 |0.88703

|0.90000 |47.99025 |46.58333 |14.22954 |108.69207 |490.0 |33.57247

|32.50000 |17.36519 |49.31407 |234.0 |0.00000 |0.22857 |0.11020 |0.06122

|0.26327

|897 |250903 |NaN |140 |25 |25 |Massachusetts |MA |Cambridge |Cambridge

City |City |tract |2139 |617 |42.359478 |-71.121263 |1158117 |0 |6234

|2518 |3716 |1964.19623 |2025.0 |769.92869 |137.01400 |646.0 |1.00000

|0.98262 |0.90521 |0.87204 |0.77725 |0.68562 |0.60506 |0.52607 |646 |633

|54583.21233 |33796.0 |53650.39665 |427.80531 |661.0 |72208.68011

|70655.0 |51468.81482 |75.59543 |158.0 |1943.06061 |1911.0 |265.85378

|2.97595 |15.0 |NaN |NaN |NaN |NaN |NaN |0.00000 |0.00000 |0.33333

|1.00000 |1.00000 |0.01506 |0.00000 |0.94958 |0.93675 |0.96093 |23.68423

|21.33333 |9.46337 |1279.90666 |2518.0 |22.53871 |20.75000 |7.40442

|2069.57453 |3716.0 |0.02169 |0.10879 |0.05440 |0.00204 |0.00409

|1340 |224994 |NaN |140 |29 |6 |California |CA |Edwards |Edwards Afb

|City |tract |93523 |661 |34.916328 |-117.874053 |871289334 |6131489

|2617 |1429 |1188 |1559.30305 |1497.0 |353.27412 |55.18693 |353.0

|0.93123 |0.83954 |0.71920 |0.57880 |0.46991 |0.36676 |0.25501 |0.13181

|692 |349 |72236.54800 |57439.0 |55874.77821 |391.76273 |697.0

|72935.01519 |58768.0 |50609.31079 |376.88353 |678.0 |NaN |NaN |NaN |NaN

|NaN |53.59461 |53.0 |24.62293 |5.0 |2.45000 |0.00000 |0.00000 |0.00000

|0.00000 |1.00000 |1.00000 |1.00000 |0.99251 |1.00000 |0.98393 |21.64262

|23.83333 |13.59898 |331.04457 |1429.0 |21.51140 |24.58333 |13.71310

|253.66584 |1188.0 |0.00615 |0.77842 |0.02411 |0.01148 |0.01148

|1400 |251389 |NaN |140 |3 |24 |Maryland |MD |Linthicum |Ferndale |CDP

|tract |21090 |410 |39.174278 |-76.671513 |13238827 |1056 |22 |14 |8

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |0 |0

|112499.50000 |112499.0 |9128.70929 |0.79987 |4.0 |112499.50000

|112499.0 |9128.70929 |0.79987 |4.0 |NaN |NaN |NaN |NaN |NaN |457.00000

|456.0 |32.04814 |4.0 |1.97980 |0.00000 |0.00000 |0.00000 |0.00000

|1.00000 |1.00000 |1.00000 |0.66667 |0.60000 |0.75000 |39.10110

|41.41667 |27.64420 |2.29033 |14.0 |57.44250 |39.58333 |24.03228

|1.18411 |8.0 |1.00000 |0.60000 |0.10000 |0.00000 |0.00000

|3065 |278467 |NaN |140 |101 |42 |Pennsylvania |PA |Philadelphia

|Philadelphia City |Borough |tract |19140 |215 |40.017529 |-75.143907

|416088 |0 |0 |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN

|3326 |257215 |NaN |140 |15 |29 |Missouri |MO |Warsaw |Lincoln City

|City |tract |65355 |660 |38.301823 |-93.172919 |216350358 |11553444

|1791 |969 |822 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |40 |0 |46327.26702 |38996.0 |37598.54588 |639.37037 |912.0

|59055.53460 |49136.0 |37206.65193 |376.65545 |576.0 |987.37333 |894.0

|505.55658 |341.06437 |439.0 |302.55139 |266.0 |157.28782 |433.0

|392.55141 |0.02179 |0.02179 |0.06422 |0.50344 |0.42637 |0.60834

|0.77412 |0.77672 |0.76712 |0.78674 |54.77733 |60.16667 |21.65187

|291.87013 |969.0 |59.60751 |60.50000 |16.70192 |197.45064 |822.0

|0.97767 |0.62311 |0.04320 |0.01944 |0.16523

|3331 |288512 |NaN |140 |35 |49 |Utah |UT |Salt Lake City |Salt Lake

City City |City |tract |84116 |801 |40.788810 |-111.992473 |39295590

|83499 |0 |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN

|4225 |266664 |NaN |140 |27 |36 |New York |NY |Poughquag |Hopewell

Junction |City |tract |12570 |845 |41.578838 |-73.722796 |3206552 |62329

|1843 |1843 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |0.70051 |0.70051 |NaN |43.06144 |43.50000 |10.84415

|379.82442 |1843.0 |NaN |NaN |NaN |NaN |NaN |NaN |0.30982 |0.30982

|0.04883 |0.19045

|5278 |278468 |NaN |140 |101 |42 |Pennsylvania |PA |Philadelphia |Colwyn

|Borough |tract |19145 |215 |39.905538 |-75.170715 |1286699 |0 |0 |0 |0

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |0 |0

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN

|5933 |231120 |NaN |140 |85 |6 |California |CA |Palo Alto |Stanford

|City |tract |94304 |650 |37.438545 |-122.174414 |1537202 |0 |3156 |1551

|1605 |2710.53509 |2731.0 |809.42736 |149.16704 |1449.0 |0.91803

|0.83321 |0.62651 |0.51461 |0.41625 |0.35068 |0.30150 |0.23093 |1485

|1403 |127847.38950 |100562.0 |93188.38962 |527.78028 |1682.0

|145366.06790 |124344.0 |74766.72145 |147.34926 |681.0 |NaN |NaN |NaN

|NaN |NaN |563.91660 |742.0 |450.07116 |197.0 |97.00515 |0.00000

|0.00000 |0.00000 |0.00000 |1.00000 |1.00000 |1.00000 |0.98607 |0.98118

|0.99057 |40.62226 |36.41667 |25.09848 |342.83838 |1551.0 |46.22367

|38.66667 |26.56767 |370.08556 |1605.0 |0.07817 |0.54481 |0.07626

|0.02280 |0.08648

|6288 |279063 |NaN |140 |23 |72 |Puerto Rico |PR |Cabo Rojo |Pole Ojea

|Urban |tract |623 |787 |17.968741 |-67.178217 |4912039 |12647781 |0 |0

|0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |0

|0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN

|6675 |267259 |NaN |140 |47 |36 |New York |NY |Brooklyn |New York City

|City |tract |11206 |718 |40.695256 |-73.940438 |121875 |0 |4150 |1978

|2172 |921.48679 |822.0 |566.74462 |797.31158 |1305.0 |0.95585 |0.91251

|0.79886 |0.64432 |0.49550 |0.45544 |0.38757 |0.22404 |1325 |1223

|42596.36726 |31519.0 |43411.50503 |1011.17505 |1416.0 |46019.48414

|36656.0 |40398.42249 |591.94683 |822.0 |2987.61392 |2618.0 |928.44641

|14.97281 |91.0 |NaN |NaN |NaN |NaN |NaN |0.00000 |0.00000 |0.00000

|1.00000 |1.00000 |1.00000 |0.00000 |0.62059 |0.59738 |0.63831 |30.69737

|27.25000 |20.08702 |454.80826 |1978.0 |35.31304 |31.83333 |20.52124

|539.64459 |2172.0 |0.12475 |0.31003 |0.13701 |0.04520 |0.07415

|7010 |281345 |NaN |140 |99 |46 |South Dakota |SD |Sioux Falls |Anderson

|CDP |tract |57110 |605 |43.523666 |-96.626164 |18540314 |176830 |1884

|913 |971 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |0 |0 |127236.15820 |112491.0 |72942.47436 |157.37355 |592.0

|127329.34000 |115191.0 |63183.65857 |132.84990 |557.0 |1889.64551

|1700.0 |800.47019 |156.79735 |480.0 |589.75510 |519.0 |215.86997 |112.0

|62.50505 |0.02365 |0.02365 |0.17736 |0.81081 |0.40792 |0.16507 |0.13928

|0.95756 |0.92512 |0.99133 |35.71392 |37.58333 |20.78729 |226.28395

|913.0 |32.92737 |35.16667 |21.17193 |221.34904 |971.0 |1.00000 |0.79370

|0.00860 |0.00000 |0.03438

|7186 |278597 |NaN |140 |119 |42 |Pennsylvania |PA |Lewisburg |Linntown

|Borough |tract |17837 |570 |40.951827 |-76.889360 |1622772 |0 |3476

|1800 |1676 |1029.05796 |1011.0 |96.92486 |2.76695 |8.0 |0.00000

|0.00000 |0.00000 |0.00000 |0.00000 |0.00000 |0.00000 |0.00000 |8 |8

|71889.08854 |47395.0 |86368.78152 |37.22853 |45.0 |242857.14290

|242720.0 |25316.86942 |0.26661 |4.0 |1498.36725 |1464.0 |192.26619

|12.12133 |37.0 |NaN |NaN |NaN |NaN |NaN |0.00000 |0.00000 |0.00000

|1.00000 |1.00000 |1.00000 |0.00000 |1.00000 |1.00000 |1.00000 |20.46232

|19.83333 |4.57539 |1216.20979 |1800.0 |20.50639 |19.50000 |6.69220

|1053.59149 |1676.0 |0.77778 |0.00278 |0.00056 |0.00000 |0.01333

|7500 |253706 |NaN |140 |73 |26 |Michigan |MI |Mount Pleasant |Mount

Pleasant City |CDP |tract |48858 |989 |43.573104 |-84.742742 |11336700

|9295 |8134 |3642 |4492 |912.02693 |686.0 |662.39378 |1454.71815 |1984.0

|1.00000 |0.97004 |0.91921 |0.90316 |0.81808 |0.71054 |0.67951 |0.56875

|1984 |1869 |20165.34995 |13227.0 |21172.10025 |1695.82951 |2016.0

|31262.24460 |24626.0 |28228.65602 |241.78851 |320.0 |NaN |NaN |NaN |NaN

|NaN |443.16097 |335.0 |192.25770 |32.0 |27.45455 |0.00000 |0.00000

|0.00000 |0.00000 |1.00000 |1.00000 |1.00000 |0.95939 |0.96000 |0.95870

|22.24897 |21.33333 |7.65018 |2117.14648 |3642.0 |23.00380 |20.66667

|11.57989 |2903.95751 |4492.0 |0.01672 |0.03808 |0.00000 |0.00000

|0.02491

|8858 |284740 |NaN |140 |141 |48 |Texas |TX |El Paso |Fort Bliss |Town

|tract |79901 |915 |31.752474 |-106.480779 |609273 |0 |2422 |1037 |1385

|335.92123 |352.0 |145.10767 |689.45000 |743.0 |0.89502 |0.78331

|0.74563 |0.58546 |0.45760 |0.34993 |0.21938 |0.13863 |790 |743

|19871.75923 |12537.0 |27319.94721 |725.05540 |871.0 |27361.09927

|16733.0 |31576.81773 |392.32244 |470.0 |NaN |NaN |NaN |NaN |NaN

|236.89895 |222.0 |85.11771 |81.0 |77.43000 |0.00000 |0.00000 |0.00000

|0.00000 |1.00000 |1.00000 |1.00000 |0.28211 |0.29901 |0.27148 |33.03870

|25.00000 |22.97857 |248.68171 |1037.0 |39.22728 |38.08333 |26.40984

|332.38824 |1385.0 |0.10858 |0.44740 |0.09055 |0.06924 |0.07324

|9006 |268339 |NaN |140 |61 |36 |New York |NY |New York |New York City

|City |tract |10009 |212 |40.722162 |-73.968250 |182345 |470253 |5557

|2592 |2965 |915.15641 |605.0 |951.48003 |1623.62685 |2126.0 |0.94267

|0.83612 |0.75968 |0.65600 |0.52891 |0.38700 |0.22981 |0.10798 |2145

|2093 |44210.57792 |20453.0 |63951.05025 |1594.06485 |2145.0

|47973.89834 |29432.0 |53946.23089 |995.53558 |1384.0 |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|0.60063 |0.56294 |0.62743 |38.61424 |36.16667 |23.23509 |695.96519

|2592.0 |47.11630 |47.83333 |20.35484 |751.01445 |2965.0 |0.00000

|0.32161 |0.02238 |0.02238 |0.06122

|9137 |268730 |NaN |140 |65 |36 |New York |NY |Rome |Rome City |City

|tract |13440 |315 |43.182307 |-75.482859 |2173458 |0 |1685 |1685 |0

|3324.65000 |3313.0 |229.67915 |0.39279 |4.0 |1.00000 |1.00000 |1.00000

|1.00000 |0.00000 |0.00000 |0.00000 |0.00000 |4 |4 |197499.85000

|194267.0 |28394.61657 |0.39992 |4.0 |182499.85000 |181311.0

|18427.87065 |0.39992 |4.0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |0.75288 |0.75288 |NaN |39.95128

|37.25000 |14.15392 |402.97579 |1685.0 |NaN |NaN |NaN |NaN |NaN |0.00000

|0.19525 |0.19050 |0.05519 |0.12819

|9231 |265330 |NaN |140 |3 |32 |Nevada |NV |Las Vegas |Winchester |City

|tract |89169 |702 |36.125959 |-115.141625 |683026 |0 |4729 |2737 |1992

|671.36697 |656.0 |187.52303 |1632.24267 |1814.0 |0.98268 |0.92682

|0.78045 |0.72849 |0.63128 |0.55642 |0.46536 |0.35922 |1856 |1790

|24957.28171 |20522.0 |22777.27105 |1617.20766 |1856.0 |26716.59876

|21956.0 |24127.47665 |625.70936 |781.0 |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |0.66274 |0.68173

|0.63826 |30.58418 |28.66667 |20.25123 |658.23330 |2737.0 |28.62225

|28.08333 |18.97010 |444.48906 |1992.0 |0.00000 |0.28295 |0.16334

|0.10022 |0.15061

|9287 |251576 |NaN |140 |5 |24 |Maryland |MD |Baltimore |Towson |CDP

|tract |21212 |410 |39.387974 |-76.618651 |2185712 |1315 |5007 |2162

|2845 |1336.02198 |1223.0 |605.33370 |34.67664 |96.0 |1.00000 |1.00000

|0.86747 |0.86747 |0.75904 |0.75904 |0.75904 |0.75904 |103 |83

|20568.45568 |14603.0 |17204.91573 |87.66033 |108.0 |40248.55193

|52427.0 |19468.07418 |8.49965 |12.0 |NaN |NaN |NaN |NaN |NaN |224.50000

|224.0 |18.25742 |5.0 |5.00000 |0.00000 |0.00000 |0.00000 |0.00000

|1.00000 |1.00000 |1.00000 |0.85477 |0.88235 |0.81905 |21.97592

|19.91667 |9.89882 |1318.24163 |2162.0 |20.93667 |19.58333 |8.74794

|1732.25545 |2845.0 |0.05397 |0.03846 |0.03290 |0.00000 |0.01529

|9396 |238950 |NaN |140 |53 |13 |Georgia |GA |Cusseta

|Cusseta-chattahoochee |City |tract |31805 |706 |32.418280 |-84.750912

|303578182 |1283826 |525 |427 |98 |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |0.97887 |0.97087 |1.00000 |23.36435

|22.16667 |4.47736 |218.35867 |427.0 |24.41960 |20.91667 |6.05302

|48.12957 |98.0 |NaN |0.11710 |0.11710 |0.01171 |0.03981

|9406 |279895 |NaN |140 |5 |44 |Rhode Island |RI |Middletown |Melville

|CDP |tract |2842 |401 |41.537245 |-71.310086 |2334457 |3083628 |1249

|724 |525 |1650.92692 |1549.0 |509.14613 |86.96682 |504.0 |1.00000

|0.95609 |0.87226 |0.63872 |0.48104 |0.33932 |0.22954 |0.16168 |504 |501

|67799.84133 |65720.0 |36023.31233 |257.41942 |504.0 |68726.57033

|67295.0 |35064.23114 |152.12434 |308.0 |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |0.96514 |0.99404

|0.92097 |31.18203 |29.00000 |16.09624 |160.97682 |724.0 |31.26418

|30.75000 |18.98342 |118.13191 |525.0 |0.00000 |0.50332 |0.09967

|0.00000 |0.09302

|9538 |225615 |NaN |140 |37 |6 |California |CA |Los Angeles |Vernon City

|City |tract |90057 |213 |34.063766 |-118.269130 |229680 |0 |4403 |2289

|2114 |821.42205 |848.0 |398.36147 |858.86409 |1409.0 |0.99145 |0.96154

|0.86111 |0.76496 |0.56481 |0.46439 |0.38746 |0.23932 |1409 |1404

|31373.46716 |22292.0 |27874.74886 |1272.90184 |1458.0 |36114.32264

|31193.0 |25828.34266 |771.19103 |895.0 |2150.61111 |1934.0 |496.31855

|9.72144 |49.0 |NaN |NaN |NaN |NaN |NaN |0.00000 |0.00000 |0.00000

|1.00000 |1.00000 |1.00000 |0.00000 |0.50320 |0.50706 |0.49928 |34.26073

|30.75000 |20.75802 |647.85451 |2289.0 |39.97799 |35.58333 |25.46371

|473.79652 |2114.0 |0.05292 |0.38684 |0.09105 |0.02895 |0.03474

|9770 |227404 |NaN |140 |37 |6 |California |CA |Los Angeles |El Segundo

City |City |tract |90045 |310 |33.942143 |-118.417330 |17299747 |2367417

|0 |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN

|9801 |283415 |NaN |140 |29 |48 |Texas |TX |San Antonio |Lackland Afb

|Town |tract |78227 |210 |29.372491 |-98.666017 |16145474 |4426 |1332

|894 |438 |1379.85766 |1381.0 |245.91291 |47.03635 |229.0 |0.99111

|0.96444 |0.91556 |0.80444 |0.72444 |0.50222 |0.34667 |0.20444 |238 |225

|50990.39025 |47412.0 |27651.97230 |172.58550 |238.0 |51451.42372

|47693.0 |25964.24378 |168.49520 |234.0 |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |1.00000 |1.00000

|1.00000 |22.26905 |22.00000 |7.73509 |343.91811 |894.0 |18.95394

|21.91667 |11.55560 |118.90004 |438.0 |0.00000 |0.37121 |0.09848

|0.00000 |0.00000

|10395 |240129 |NaN |140 |179 |13 |Georgia |GA |Fort Stewart |Fort

Stewart |City |tract |31315 |912 |31.940065 |-81.568158 |203407794

|972935 |6465 |3041 |3424 |1201.87699 |1187.0 |290.68785 |444.77827

|1703.0 |1.00000 |0.98533 |0.95982 |0.85842 |0.69962 |0.56186 |0.46875

|0.30421 |1754 |1568 |39039.93471 |34882.0 |27928.00400 |1435.08966

|1762.0 |39379.81618 |35512.0 |24833.02933 |1312.10071 |1636.0

|1787.00000 |1781.0 |160.24070 |1.58717 |8.0 |NaN |NaN |NaN |NaN |NaN

|0.00000 |0.00000 |1.00000 |1.00000 |1.00000 |0.00000 |0.00000 |0.96809

|0.99267 |0.94625 |19.64485 |19.41667 |13.25630 |657.38771 |3041.0

|19.95132 |20.50000 |13.65045 |814.71125 |3424.0 |0.00264 |0.84226

|0.03793 |0.00482 |0.02830

|10617 |245342 |NaN |140 |3 |18 |Indiana |IN |Fort Wayne |Fort Wayne

City |City |tract |46825 |260 |41.118061 |-85.110168 |3928112 |0 |630

|259 |371 |288.68848 |294.0 |178.14926 |51.97980 |54.0 |1.00000 |0.86000

|0.64000 |0.56000 |0.34000 |0.34000 |0.00000 |0.00000 |54 |50

|24655.77285 |11974.0 |28730.36014 |50.06375 |66.0 |88999.50000 |88753.0

|8231.03882 |2.39962 |12.0 |NaN |NaN |NaN |NaN |NaN |274.50000 |274.0

|18.25742 |12.0 |12.00000 |0.00000 |0.00000 |0.00000 |0.00000 |1.00000

|1.00000 |1.00000 |0.67073 |0.34783 |0.79661 |23.00788 |20.00000

|10.08713 |142.00790 |259.0 |26.71830 |19.41667 |18.74821 |198.87402

|371.0 |0.29870 |0.04247 |0.00000 |0.00000 |0.00000

|10682 |236210 |NaN |140 |73 |12 |Florida |FL |Tallahassee |Tallahassee

City |City |tract |32304 |850 |30.443219 |-84.300765 |1472129 |476 |5231

|2195 |3036 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |0.91176 |0.76923 |1.00000 |19.17985 |19.00000 |2.14627 |1207.87083

|2195.0 |19.12433 |18.91667 |1.38846 |1661.70780 |3036.0 |NaN |0.00091

|0.00091 |0.00046 |0.00046

|10961 |265934 |NaN |140 |1 |36 |New York |NY |Albany |Colonie |City

|tract |12203 |518 |42.690537 |-73.828376 |1830774 |84514 |5033 |2457

|2576 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|0 |0 |41249.50000 |45287.0 |8103.06869 |8.00000 |8.0 |NaN |NaN |NaN

|NaN |NaN |1471.31678 |1473.0 |344.07942 |2.38395 |8.0 |NaN |NaN |NaN

|NaN |NaN |0.00000 |0.00000 |0.00000 |1.00000 |1.00000 |1.00000 |0.00000

|0.80488 |0.67568 |0.91111 |20.24287 |19.75000 |3.49572 |1591.68084

|2457.0 |20.34341 |19.66667 |5.36003 |1646.97751 |2576.0 |1.00000

|0.00651 |0.00651 |0.00000 |0.00000

|11100 |265797 |NaN |140 |23 |32 |Nevada |NV |Beatty |Beatty |City

|tract |89003 |775 |37.312404 |-116.430732 |10826415150 |16598890 |0 |0

|0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |0

|0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN

|11543 |267675 |NaN |140 |47 |36 |New York |NY |Brooklyn |New York City

|City |tract |11212 |718 |40.670130 |-73.908859 |195831 |0 |4318 |1772

|2546 |610.68727 |434.0 |506.85855 |1313.88139 |1563.0 |0.97193 |0.91578

|0.77741 |0.63971 |0.46992 |0.36898 |0.33757 |0.28810 |1573 |1496

|26932.85425 |16399.0 |31321.38523 |1219.53900 |1591.0 |32650.49890

|25726.0 |31093.21593 |827.65173 |1181.0 |4038.22035 |3998.0 |367.72638

|2.67657 |18.0 |NaN |NaN |NaN |NaN |NaN |0.00000 |0.00000 |0.00000

|1.00000 |1.00000 |1.00000 |0.00000 |0.66134 |0.75184 |0.61609 |28.53376

|23.66667 |18.42455 |539.19631 |1772.0 |37.65214 |37.00000 |20.92169

|698.63402 |2546.0 |0.02453 |0.29500 |0.04282 |0.02776 |0.03013

|11554 |288512 |NaN |140 |35 |49 |Utah |UT |Salt Lake City |Salt Lake

City City |City |tract |84116 |801 |40.788810 |-111.992473 |39295590

|83499 |0 |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |0 |0 |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN |NaN

|NaN |NaN |NaN |NaN |NaN |NaN

|===

----

+\*In[28]:\*+

[source, ipython3]

----

len(df\_test)

----

+\*Out[28]:\*+

----11709----

+\*In[29]:\*+

[source, ipython3]

----

df\_test.head()

----

+\*Out[29]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |BLOCKID |SUMLEVEL |COUNTYID |STATEID |state |state\_ab |city

|place |type |primary |zip\_code |area\_code |lat |lng |ALand |AWater |pop

|male\_pop |female\_pop |rent\_mean |rent\_median |rent\_stdev

|rent\_sample\_weight |rent\_samples |rent\_gt\_10 |rent\_gt\_15 |rent\_gt\_20

|rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35 |rent\_gt\_40 |rent\_gt\_50

|universe\_samples |used\_samples |hi\_mean |hi\_median |hi\_stdev

|hi\_sample\_weight |hi\_samples |family\_mean |family\_median |family\_stdev

|family\_sample\_weight |family\_samples |hc\_mortgage\_mean

|hc\_mortgage\_median |hc\_mortgage\_stdev |hc\_mortgage\_sample\_weight

|hc\_mortgage\_samples |hc\_mean |hc\_median |hc\_stdev |hc\_samples

|hc\_sample\_weight |home\_equity\_second\_mortgage |second\_mortgage

|home\_equity |debt |second\_mortgage\_cdf |home\_equity\_cdf |debt\_cdf

|hs\_degree |hs\_degree\_male |hs\_degree\_female |male\_age\_mean

|male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|0 |255504 |NaN |140 |163 |26 |Michigan |MI |Detroit |Dearborn Heights

City |CDP |tract |48239 |313 |42.346422 |-83.252823 |2711280 |39555

|3417 |1479 |1938 |858.57169 |859.0 |232.39082 |276.07497 |424.0

|1.00000 |0.95696 |0.85316 |0.85316 |0.85316 |0.85316 |0.76962 |0.63544

|435 |395 |48899.52121 |38746.0 |44392.20902 |798.02401 |1180.0

|53802.87122 |45167.0 |43756.56479 |464.30972 |769.0 |1139.24548 |1109.0

|336.47710 |262.67011 |474.0 |488.51323 |436.0 |192.75147 |271.0

|189.18182 |0.06443 |0.06443 |0.07651 |0.63624 |0.14111 |0.55087

|0.51965 |0.91047 |0.92010 |0.90391 |33.37131 |27.83333 |22.36768

|334.30978 |1479.0 |34.78682 |33.75000 |21.58531 |416.48097 |1938.0

|0.70252 |0.28217 |0.05910 |0.03813 |0.14299

|1 |252676 |NaN |140 |1 |23 |Maine |ME |Auburn |Auburn City |City |tract

|4210 |207 |44.100724 |-70.257832 |14778785 |2705204 |3796 |1846 |1950

|832.68625 |750.0 |267.22342 |183.32299 |245.0 |1.00000 |1.00000

|0.86611 |0.67364 |0.30962 |0.30962 |0.30962 |0.27197 |275 |239

|72335.33234 |61008.0 |51895.81159 |922.82969 |1722.0 |85642.22095

|74759.0 |49156.72870 |482.99945 |1147.0 |1533.25988 |1438.0 |536.61118

|373.96188 |937.0 |661.31296 |668.0 |201.31365 |510.0 |279.69697

|0.01175 |0.01175 |0.14375 |0.64755 |0.52310 |0.26442 |0.49359 |0.94290

|0.92832 |0.95736 |43.88680 |46.08333 |22.90302 |427.10824 |1846.0

|44.23451 |46.66667 |22.37036 |532.03505 |1950.0 |0.85128 |0.64221

|0.02338 |0.00000 |0.13377

|2 |276314 |NaN |140 |15 |42 |Pennsylvania |PA |Pine City |Millerton

|Borough |tract |14871 |607 |41.948556 |-76.783808 |258903666 |863840

|3944 |2065 |1879 |816.00639 |755.0 |416.25699 |141.39063 |217.0

|0.97573 |0.93204 |0.78641 |0.71845 |0.63592 |0.47573 |0.43689 |0.32524

|245 |206 |58501.15901 |51648.0 |45245.27248 |893.07759 |1461.0

|65694.06582 |57186.0 |44239.31893 |619.73962 |1084.0 |1254.54462

|1089.0 |596.85204 |340.45884 |552.0 |397.44466 |356.0 |189.40372 |664.0

|534.16737 |0.01069 |0.01316 |0.06497 |0.45395 |0.51066 |0.60484

|0.83848 |0.89238 |0.86003 |0.92463 |39.81661 |41.91667 |24.29111

|499.10080 |2065.0 |41.62426 |44.50000 |22.86213 |453.11959 |1879.0

|0.81897 |0.59961 |0.01746 |0.01358 |0.10026

|3 |248614 |NaN |140 |231 |21 |Kentucky |KY |Monticello |Monticello City

|City |tract |42633 |606 |36.746009 |-84.766870 |501694825 |2623067

|2508 |1427 |1081 |418.68937 |385.0 |156.92024 |88.95960 |93.0 |1.00000

|0.93548 |0.93548 |0.64516 |0.55914 |0.46237 |0.46237 |0.36559 |153 |93

|38237.55059 |31612.0 |34527.61607 |775.17947 |957.0 |44156.38709

|34687.0 |34899.74300 |535.21987 |689.0 |862.65763 |749.0 |624.42157

|299.56752 |337.0 |200.88113 |180.0 |91.56490 |467.0 |454.85404 |0.00995

|0.00995 |0.01741 |0.41915 |0.53770 |0.80931 |0.87403 |0.60908 |0.56584

|0.65947 |41.81638 |43.00000 |24.65325 |333.57733 |1427.0 |44.81200

|48.00000 |21.03155 |263.94320 |1081.0 |0.84609 |0.56953 |0.05492

|0.04694 |0.12489

|4 |286865 |NaN |140 |355 |48 |Texas |TX |Corpus Christi |Edroy |Town

|tract |78410 |361 |27.882461 |-97.678586 |13796057 |497689 |6230 |3274

|2956 |1031.63763 |997.0 |326.76727 |277.39844 |624.0 |0.72276 |0.66506

|0.53526 |0.38301 |0.18910 |0.16667 |0.14263 |0.11058 |660 |624

|114456.07790 |94211.0 |81950.95692 |836.30759 |2404.0 |123527.02420

|103898.0 |72173.55823 |507.42257 |1738.0 |1996.41425 |1907.0 |740.21168

|319.97570 |1102.0 |867.57713 |804.0 |376.20236 |642.0 |333.91919

|0.00000 |0.00000 |0.03440 |0.63188 |1.00000 |0.74519 |0.52943 |0.86297

|0.87969 |0.84466 |42.13301 |43.75000 |22.69502 |833.57435 |3274.0

|40.66618 |42.66667 |21.30900 |709.90829 |2956.0 |0.79077 |0.57620

|0.01726 |0.00588 |0.16379

|===

----

+\*In[30]:\*+

[source, ipython3]

----

len(df\_test["UID"].unique())

----

+\*Out[30]:\*+

----11677----

+\*In[31]:\*+

[source, ipython3]

----

test\_df = df\_test.drop\_duplicates()

----

+\*In[32]:\*+

[source, ipython3]

----

len(test\_df)

----

+\*Out[32]:\*+

----11677----

+\*In[33]:\*+

[source, ipython3]

----

test\_df.isnull().sum(axis=0)[0:90]

----

+\*Out[33]:\*+

----UID 0

BLOCKID 11677

SUMLEVEL 0

COUNTYID 0

STATEID 0

state 0

state\_ab 0

city 0

place 0

type 0

primary 0

zip\_code 0

area\_code 0

lat 0

lng 0

ALand 0

AWater 0

pop 0

male\_pop 0

female\_pop 0

rent\_mean 134

rent\_median 134

rent\_stdev 134

rent\_sample\_weight 134

rent\_samples 134

rent\_gt\_10 135

rent\_gt\_15 135

rent\_gt\_20 135

rent\_gt\_25 135

rent\_gt\_30 135

rent\_gt\_35 135

rent\_gt\_40 135

rent\_gt\_50 135

universe\_samples 0

used\_samples 0

hi\_mean 112

hi\_median 112

hi\_stdev 112

hi\_sample\_weight 112

hi\_samples 112

family\_mean 125

family\_median 125

family\_stdev 125

family\_sample\_weight 125

family\_samples 125

hc\_mortgage\_mean 245

hc\_mortgage\_median 245

hc\_mortgage\_stdev 245

hc\_mortgage\_sample\_weight 245

hc\_mortgage\_samples 245

hc\_mean 267

hc\_median 267

hc\_stdev 267

hc\_samples 267

hc\_sample\_weight 267

home\_equity\_second\_mortgage 204

second\_mortgage 204

home\_equity 204

debt 204

second\_mortgage\_cdf 204

home\_equity\_cdf 204

debt\_cdf 204

hs\_degree 78

hs\_degree\_male 82

hs\_degree\_female 96

male\_age\_mean 77

male\_age\_median 77

male\_age\_stdev 77

male\_age\_sample\_weight 77

male\_age\_samples 77

female\_age\_mean 87

female\_age\_median 87

female\_age\_stdev 87

female\_age\_sample\_weight 87

female\_age\_samples 87

pct\_own 112

married 77

married\_snp 77

separated 77

divorced 77

dtype: int64----

#3.missing value treatment

+\*In[34]:\*+

[source, ipython3]

----

train\_df.nunique()

----

+\*Out[34]:\*+

----UID 27161

BLOCKID 0

SUMLEVEL 1

COUNTYID 296

STATEID 52

state 52

state\_ab 52

city 6916

place 9912

type 6

primary 1

zip\_code 12744

area\_code 274

lat 27158

lng 27160

ALand 27131

AWater 16488

pop 7648

male\_pop 4507

female\_pop 4606

rent\_mean 26806

rent\_median 2335

rent\_stdev 26792

rent\_sample\_weight 26693

rent\_samples 2178

rent\_gt\_10 8373

rent\_gt\_15 13983

rent\_gt\_20 16650

rent\_gt\_25 17592

rent\_gt\_30 17828

rent\_gt\_35 17619

rent\_gt\_40 17320

rent\_gt\_50 16540

universe\_samples 2200

used\_samples 2124

hi\_mean 26938

hi\_median 23111

hi\_stdev 26934

hi\_sample\_weight 26941

hi\_samples 3491

family\_mean 26914

family\_median 23606

family\_stdev 26912

family\_sample\_weight 26915

family\_samples 2700

hc\_mortgage\_mean 26598

hc\_mortgage\_median 3051

hc\_mortgage\_stdev 26585

hc\_mortgage\_sample\_weight 26556

hc\_mortgage\_samples 2193

hc\_mean 26502

hc\_median 1407

hc\_stdev 26491

hc\_samples 1298

hc\_sample\_weight 24056

home\_equity\_second\_mortgage 7223

second\_mortgage 7821

home\_equity 14578

debt 20088

second\_mortgage\_cdf 13751

home\_equity\_cdf 18225

debt\_cdf 21118

hs\_degree 17917

hs\_degree\_male 18007

hs\_degree\_female 17386

male\_age\_mean 26811

male\_age\_median 682

male\_age\_stdev 26403

male\_age\_sample\_weight 27008

male\_age\_samples 4506

female\_age\_mean 26771

female\_age\_median 665

female\_age\_stdev 26399

female\_age\_sample\_weight 26990

female\_age\_samples 4605

pct\_own 22302

married 20282

married\_snp 10350

separated 6190

divorced 13688

dtype: int64----

+\*In[35]:\*+

[source, ipython3]

----

test\_df.nunique()

----

+\*Out[35]:\*+

----UID 11677

BLOCKID 0

SUMLEVEL 1

COUNTYID 246

STATEID 52

state 52

state\_ab 52

city 4369

place 5902

type 6

primary 1

zip\_code 7788

area\_code 268

lat 11676

lng 11677

ALand 11677

AWater 7026

pop 5692

male\_pop 3663

female\_pop 3770

rent\_mean 11499

rent\_median 1930

rent\_stdev 11494

rent\_sample\_weight 11495

rent\_samples 1838

rent\_gt\_10 5038

rent\_gt\_15 7715

rent\_gt\_20 8756

rent\_gt\_25 9156

rent\_gt\_30 9200

rent\_gt\_35 9212

rent\_gt\_40 9105

rent\_gt\_50 8749

universe\_samples 1857

used\_samples 1805

hi\_mean 11559

hi\_median 10768

hi\_stdev 11556

hi\_sample\_weight 11562

hi\_samples 2935

family\_mean 11550

family\_median 10883

family\_stdev 11548

family\_sample\_weight 11550

family\_samples 2314

hc\_mortgage\_mean 11376

hc\_mortgage\_median 2564

hc\_mortgage\_stdev 11370

hc\_mortgage\_sample\_weight 11387

hc\_mortgage\_samples 1881

hc\_mean 11323

hc\_median 1214

hc\_stdev 11318

hc\_samples 1141

hc\_sample\_weight 10851

home\_equity\_second\_mortgage 4914

second\_mortgage 5348

home\_equity 8236

debt 9944

second\_mortgage\_cdf 7397

home\_equity\_cdf 9166

debt\_cdf 10158

hs\_degree 9666

hs\_degree\_male 9536

hs\_degree\_female 9356

male\_age\_mean 11570

male\_age\_median 596

male\_age\_stdev 11505

male\_age\_sample\_weight 11600

male\_age\_samples 3662

female\_age\_mean 11552

female\_age\_median 595

female\_age\_stdev 11472

female\_age\_sample\_weight 11587

female\_age\_samples 3769

pct\_own 10578

married 10215

married\_snp 6829

separated 4512

divorced 8273

dtype: int64----

Since "BLOCKID" has all null values, "SUMLEVEL" and "primary" has same value in all the rows, they dont impact variance hence they these columns can be dropped

+\*In[36]:\*+

[source, ipython3]

----

train\_df.drop(['BLOCKID','SUMLEVEL','primary'], axis=1, inplace=True)

----

+\*Out[36]:\*+

----

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py:4163: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy

return super().drop(

----

+\*In[37]:\*+

[source, ipython3]

----

train\_df.nunique()

----

+\*Out[37]:\*+

----UID 27161

COUNTYID 296

STATEID 52

state 52

state\_ab 52

city 6916

place 9912

type 6

zip\_code 12744

area\_code 274

lat 27158

lng 27160

ALand 27131

AWater 16488

pop 7648

male\_pop 4507

female\_pop 4606

rent\_mean 26806

rent\_median 2335

rent\_stdev 26792

rent\_sample\_weight 26693

rent\_samples 2178

rent\_gt\_10 8373

rent\_gt\_15 13983

rent\_gt\_20 16650

rent\_gt\_25 17592

rent\_gt\_30 17828

rent\_gt\_35 17619

rent\_gt\_40 17320

rent\_gt\_50 16540

universe\_samples 2200

used\_samples 2124

hi\_mean 26938

hi\_median 23111

hi\_stdev 26934

hi\_sample\_weight 26941

hi\_samples 3491

family\_mean 26914

family\_median 23606

family\_stdev 26912

family\_sample\_weight 26915

family\_samples 2700

hc\_mortgage\_mean 26598

hc\_mortgage\_median 3051

hc\_mortgage\_stdev 26585

hc\_mortgage\_sample\_weight 26556

hc\_mortgage\_samples 2193

hc\_mean 26502

hc\_median 1407

hc\_stdev 26491

hc\_samples 1298

hc\_sample\_weight 24056

home\_equity\_second\_mortgage 7223

second\_mortgage 7821

home\_equity 14578

debt 20088

second\_mortgage\_cdf 13751

home\_equity\_cdf 18225

debt\_cdf 21118

hs\_degree 17917

hs\_degree\_male 18007

hs\_degree\_female 17386

male\_age\_mean 26811

male\_age\_median 682

male\_age\_stdev 26403

male\_age\_sample\_weight 27008

male\_age\_samples 4506

female\_age\_mean 26771

female\_age\_median 665

female\_age\_stdev 26399

female\_age\_sample\_weight 26990

female\_age\_samples 4605

pct\_own 22302

married 20282

married\_snp 10350

separated 6190

divorced 13688

dtype: int64----

+\*In[38]:\*+

[source, ipython3]

----

test\_df.drop(['BLOCKID','SUMLEVEL','primary'], axis=1, inplace=True)

----

+\*In[39]:\*+

[source, ipython3]

----

test\_df.nunique()

----

+\*Out[39]:\*+

----UID 11677

COUNTYID 246

STATEID 52

state 52

state\_ab 52

city 4369

place 5902

type 6

zip\_code 7788

area\_code 268

lat 11676

lng 11677

ALand 11677

AWater 7026

pop 5692

male\_pop 3663

female\_pop 3770

rent\_mean 11499

rent\_median 1930

rent\_stdev 11494

rent\_sample\_weight 11495

rent\_samples 1838

rent\_gt\_10 5038

rent\_gt\_15 7715

rent\_gt\_20 8756

rent\_gt\_25 9156

rent\_gt\_30 9200

rent\_gt\_35 9212

rent\_gt\_40 9105

rent\_gt\_50 8749

universe\_samples 1857

used\_samples 1805

hi\_mean 11559

hi\_median 10768

hi\_stdev 11556

hi\_sample\_weight 11562

hi\_samples 2935

family\_mean 11550

family\_median 10883

family\_stdev 11548

family\_sample\_weight 11550

family\_samples 2314

hc\_mortgage\_mean 11376

hc\_mortgage\_median 2564

hc\_mortgage\_stdev 11370

hc\_mortgage\_sample\_weight 11387

hc\_mortgage\_samples 1881

hc\_mean 11323

hc\_median 1214

hc\_stdev 11318

hc\_samples 1141

hc\_sample\_weight 10851

home\_equity\_second\_mortgage 4914

second\_mortgage 5348

home\_equity 8236

debt 9944

second\_mortgage\_cdf 7397

home\_equity\_cdf 9166

debt\_cdf 10158

hs\_degree 9666

hs\_degree\_male 9536

hs\_degree\_female 9356

male\_age\_mean 11570

male\_age\_median 596

male\_age\_stdev 11505

male\_age\_sample\_weight 11600

male\_age\_samples 3662

female\_age\_mean 11552

female\_age\_median 595

female\_age\_stdev 11472

female\_age\_sample\_weight 11587

female\_age\_samples 3769

pct\_own 10578

married 10215

married\_snp 6829

separated 4512

divorced 8273

dtype: int64----

+\*In[40]:\*+

[source, ipython3]

----

train\_df.isnull().sum(axis=0)[0:90]

----

+\*Out[40]:\*+

----UID 0

COUNTYID 0

STATEID 0

state 0

state\_ab 0

city 0

place 0

type 0

zip\_code 0

area\_code 0

lat 0

lng 0

ALand 0

AWater 0

pop 0

male\_pop 0

female\_pop 0

rent\_mean 242

rent\_median 242

rent\_stdev 242

rent\_sample\_weight 242

rent\_samples 242

rent\_gt\_10 242

rent\_gt\_15 242

rent\_gt\_20 242

rent\_gt\_25 242

rent\_gt\_30 242

rent\_gt\_35 242

rent\_gt\_40 242

rent\_gt\_50 242

universe\_samples 0

used\_samples 0

hi\_mean 207

hi\_median 207

hi\_stdev 207

hi\_sample\_weight 207

hi\_samples 207

family\_mean 230

family\_median 230

family\_stdev 230

family\_sample\_weight 230

family\_samples 230

hc\_mortgage\_mean 442

hc\_mortgage\_median 442

hc\_mortgage\_stdev 442

hc\_mortgage\_sample\_weight 442

hc\_mortgage\_samples 442

hc\_mean 478

hc\_median 478

hc\_stdev 478

hc\_samples 478

hc\_sample\_weight 478

home\_equity\_second\_mortgage 360

second\_mortgage 360

home\_equity 360

debt 360

second\_mortgage\_cdf 360

home\_equity\_cdf 360

debt\_cdf 360

hs\_degree 145

hs\_degree\_male 154

hs\_degree\_female 171

male\_age\_mean 148

male\_age\_median 148

male\_age\_stdev 148

male\_age\_sample\_weight 148

male\_age\_samples 148

female\_age\_mean 161

female\_age\_median 161

female\_age\_stdev 161

female\_age\_sample\_weight 161

female\_age\_samples 161

pct\_own 207

married 150

married\_snp 150

separated 150

divorced 150

dtype: int64----

# POP indicates Male and Female population of the geographic location as per data dictionary. checking if population is 0 in few localities

+\*In[41]:\*+

[source, ipython3]

----

train\_df["pop"][train\_df["pop"]==0].count()

----

+\*Out[41]:\*+

----142----

#when population is 0, survey cannot happen hence removing these rows from the dataframe

+\*In[42]:\*+

[source, ipython3]

----

train\_df = train\_df.loc[train\_df["pop"] != 0]

----

+\*In[43]:\*+

[source, ipython3]

----

train\_df["pop"][train\_df["pop"]==0].count()

----

+\*Out[43]:\*+

----0----

+\*In[44]:\*+

[source, ipython3]

----

train\_df.isnull().sum(axis=0)[0:90]

----

+\*Out[44]:\*+

----UID 0

COUNTYID 0

STATEID 0

state 0

state\_ab 0

city 0

place 0

type 0

zip\_code 0

area\_code 0

lat 0

lng 0

ALand 0

AWater 0

pop 0

male\_pop 0

female\_pop 0

rent\_mean 100

rent\_median 100

rent\_stdev 100

rent\_sample\_weight 100

rent\_samples 100

rent\_gt\_10 100

rent\_gt\_15 100

rent\_gt\_20 100

rent\_gt\_25 100

rent\_gt\_30 100

rent\_gt\_35 100

rent\_gt\_40 100

rent\_gt\_50 100

universe\_samples 0

used\_samples 0

hi\_mean 65

hi\_median 65

hi\_stdev 65

hi\_sample\_weight 65

hi\_samples 65

family\_mean 88

family\_median 88

family\_stdev 88

family\_sample\_weight 88

family\_samples 88

hc\_mortgage\_mean 300

hc\_mortgage\_median 300

hc\_mortgage\_stdev 300

hc\_mortgage\_sample\_weight 300

hc\_mortgage\_samples 300

hc\_mean 336

hc\_median 336

hc\_stdev 336

hc\_samples 336

hc\_sample\_weight 336

home\_equity\_second\_mortgage 218

second\_mortgage 218

home\_equity 218

debt 218

second\_mortgage\_cdf 218

home\_equity\_cdf 218

debt\_cdf 218

hs\_degree 3

hs\_degree\_male 12

hs\_degree\_female 29

male\_age\_mean 6

male\_age\_median 6

male\_age\_stdev 6

male\_age\_sample\_weight 6

male\_age\_samples 6

female\_age\_mean 19

female\_age\_median 19

female\_age\_stdev 19

female\_age\_sample\_weight 19

female\_age\_samples 19

pct\_own 65

married 8

married\_snp 8

separated 8

divorced 8

dtype: int64----

# following same steps for test dataset

+\*In[45]:\*+

[source, ipython3]

----

test\_df["pop"][test\_df["pop"]==0].count()

----

+\*Out[45]:\*+

----74----

+\*In[46]:\*+

[source, ipython3]

----

test\_df = test\_df.loc[test\_df["pop"] != 0]

----

+\*In[47]:\*+

[source, ipython3]

----

test\_df["pop"][test\_df["pop"]==0].count()

----

+\*Out[47]:\*+

----0----

+\*In[48]:\*+

[source, ipython3]

----

test\_df.isnull().sum(axis=0)[0:90]

----

+\*Out[48]:\*+

----UID 0

COUNTYID 0

STATEID 0

state 0

state\_ab 0

city 0

place 0

type 0

zip\_code 0

area\_code 0

lat 0

lng 0

ALand 0

AWater 0

pop 0

male\_pop 0

female\_pop 0

rent\_mean 60

rent\_median 60

rent\_stdev 60

rent\_sample\_weight 60

rent\_samples 60

rent\_gt\_10 61

rent\_gt\_15 61

rent\_gt\_20 61

rent\_gt\_25 61

rent\_gt\_30 61

rent\_gt\_35 61

rent\_gt\_40 61

rent\_gt\_50 61

universe\_samples 0

used\_samples 0

hi\_mean 38

hi\_median 38

hi\_stdev 38

hi\_sample\_weight 38

hi\_samples 38

family\_mean 51

family\_median 51

family\_stdev 51

family\_sample\_weight 51

family\_samples 51

hc\_mortgage\_mean 171

hc\_mortgage\_median 171

hc\_mortgage\_stdev 171

hc\_mortgage\_sample\_weight 171

hc\_mortgage\_samples 171

hc\_mean 193

hc\_median 193

hc\_stdev 193

hc\_samples 193

hc\_sample\_weight 193

home\_equity\_second\_mortgage 130

second\_mortgage 130

home\_equity 130

debt 130

second\_mortgage\_cdf 130

home\_equity\_cdf 130

debt\_cdf 130

hs\_degree 4

hs\_degree\_male 8

hs\_degree\_female 22

male\_age\_mean 3

male\_age\_median 3

male\_age\_stdev 3

male\_age\_sample\_weight 3

male\_age\_samples 3

female\_age\_mean 13

female\_age\_median 13

female\_age\_stdev 13

female\_age\_sample\_weight 13

female\_age\_samples 13

pct\_own 38

married 3

married\_snp 3

separated 3

divorced 3

dtype: int64----

+\*In[49]:\*+

[source, ipython3]

----

train\_df.describe()

----

+\*Out[49]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |COUNTYID |STATEID |zip\_code |area\_code |lat |lng |ALand |AWater

|pop |male\_pop |female\_pop |rent\_mean |rent\_median |rent\_stdev

|rent\_sample\_weight |rent\_samples |rent\_gt\_10 |rent\_gt\_15 |rent\_gt\_20

|rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35 |rent\_gt\_40 |rent\_gt\_50

|universe\_samples |used\_samples |hi\_mean |hi\_median |hi\_stdev

|hi\_sample\_weight |hi\_samples |family\_mean |family\_median |family\_stdev

|family\_sample\_weight |family\_samples |hc\_mortgage\_mean

|hc\_mortgage\_median |hc\_mortgage\_stdev |hc\_mortgage\_sample\_weight

|hc\_mortgage\_samples |hc\_mean |hc\_median |hc\_stdev |hc\_samples

|hc\_sample\_weight |home\_equity\_second\_mortgage |second\_mortgage

|home\_equity |debt |second\_mortgage\_cdf |home\_equity\_cdf |debt\_cdf

|hs\_degree |hs\_degree\_male |hs\_degree\_female |male\_age\_mean

|male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|count |27019.000000 |27019.000000 |27019.000000 |27019.000000

|27019.000000 |27019.000000 |27019.000000 |2.701900e+04 |2.701900e+04

|27019.000000 |27019.000000 |27019.000000 |26919.000000 |26919.000000

|26919.000000 |26919.000000 |26919.000000 |26919.000000 |26919.000000

|26919.000000 |26919.000000 |26919.000000 |26919.000000 |26919.000000

|26919.000000 |27019.000000 |27019.000000 |26954.000000 |26954.000000

|26954.000000 |26954.000000 |26954.000000 |26931.000000 |26931.000000

|26931.000000 |26931.000000 |26931.000000 |26719.000000 |26719.000000

|26719.000000 |26719.000000 |26719.000000 |26683.000000 |26683.000000

|26683.000000 |26683.000000 |26683.000000 |26801.000000 |26801.000000

|26801.000000 |26801.000000 |26801.000000 |26801.000000 |26801.000000

|27016.000000 |27007.00000 |26990.000000 |27013.000000 |27013.000000

|27013.000000 |27013.000000 |27013.000000 |27000.000000 |27000.000000

|27000.000000 |27000.000000 |27000.000000 |26954.000000 |27011.000000

|27011.000000 |27011.000000 |27011.000000

|mean |257310.991673 |85.592139 |28.246567 |50132.614568 |596.381509

|37.530584 |-91.306019 |1.302186e+08 |6.495841e+06 |4347.275140

|2137.612569 |2209.662571 |1055.344777 |1007.852781 |394.419068

|294.987443 |546.209926 |0.957880 |0.867144 |0.739364 |0.612774

|0.499869 |0.410941 |0.345364 |0.254420 |577.016063 |530.991710

|70500.318611 |57624.340766 |54501.465013 |924.151692 |1610.055502

|79045.667606 |69333.490290 |50783.707702 |533.755939 |1064.829899

|1629.260500 |1550.778996 |622.686711 |287.836433 |670.452562

|540.646148 |513.453810 |218.821586 |370.768542 |255.061125 |0.025689

|0.029951 |0.100899 |0.629621 |0.466927 |0.476681 |0.499327 |0.858819

|0.85247 |0.865284 |38.373671 |38.114800 |21.534969 |534.194158

|2138.087365 |40.354614 |40.395275 |22.213313 |544.289544 |2211.217519

|0.642269 |0.509312 |0.047344 |0.019073 |0.100385

|std |21340.758358 |98.176550 |16.371166 |29539.429679 |232.438332

|5.576366 |16.329382 |1.282239e+09 |2.198107e+08 |2113.892839

|1069.804237 |1084.636509 |437.447399 |443.768889 |187.186282

|269.963679 |455.957474 |0.062576 |0.109240 |0.143514 |0.160059

|0.163692 |0.159890 |0.152947 |0.137500 |459.049696 |444.617372

|30093.686767 |29048.553867 |17567.473532 |450.125360 |747.872712

|31316.868952 |33412.036731 |14188.695534 |286.566037 |557.243789

|622.506368 |651.804936 |237.789770 |195.244877 |464.212715 |221.286456

|231.348701 |91.291683 |250.560450 |189.830532 |0.030694 |0.033558

|0.068667 |0.154410 |0.294238 |0.255305 |0.263379 |0.111772 |0.11998

|0.111384 |5.561133 |7.844144 |2.454032 |287.630780 |1069.448516

|5.822023 |7.985358 |2.447802 |280.681202 |1083.432414 |0.224184

|0.135701 |0.037156 |0.020744 |0.048808

|min |220342.000000 |1.000000 |1.000000 |602.000000 |201.000000

|17.929085 |-165.453872 |4.113400e+04 |0.000000e+00 |3.000000 |0.000000

|0.000000 |117.150000 |104.000000 |18.257420 |0.343000 |4.000000

|0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000

|0.000000 |0.000000 |0.000000 |4999.846690 |4790.000000 |1825.741860

|0.114260 |3.000000 |5374.842520 |5278.000000 |1825.741860 |0.199960

|3.000000 |234.650000 |237.000000 |36.514840 |0.198400 |1.000000

|53.594610 |53.000000 |18.257420 |2.000000 |0.614040 |0.000000 |0.000000

|0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.186520 |0.00000

|0.000000 |12.145830 |9.750000 |0.962770 |0.745760 |3.000000 |16.008330

|13.250000 |0.556780 |0.664700 |2.000000 |0.000000 |0.000000 |0.000000

|0.000000 |0.000000

|25% |238819.500000 |29.000000 |13.000000 |27023.000000 |405.000000

|33.910638 |-97.820012 |1.806570e+06 |0.000000e+00 |2915.500000

|1418.000000 |1473.000000 |743.435305 |702.000000 |263.748280

|101.988565 |221.000000 |0.940590 |0.819205 |0.661825 |0.516985

|0.396230 |0.307265 |0.243400 |0.160940 |255.000000 |213.500000

|49231.516002 |37476.000000 |42155.831497 |601.275388 |1098.000000

|56935.821225 |46256.000000 |40961.471880 |332.312015 |688.000000

|1158.213850 |1067.000000 |440.539475 |148.403580 |347.000000

|389.327340 |361.000000 |154.546580 |194.000000 |121.313130 |0.005120

|0.007770 |0.049560 |0.538850 |0.248560 |0.265020 |0.281600 |0.808420

|0.79569 |0.818510 |35.049430 |32.916670 |20.601000 |346.537640

|1419.000000 |36.932603 |35.000000 |21.324497 |356.481745 |1474.000000

|0.505040 |0.426550 |0.020825 |0.004555 |0.066015

|50% |257187.000000 |63.000000 |28.000000 |47904.000000 |614.000000

|38.767808 |-86.591218 |4.879552e+06 |2.767400e+04 |4063.000000

|1986.000000 |2067.000000 |953.199390 |897.000000 |346.519970

|218.913110 |423.000000 |0.977060 |0.888020 |0.758110 |0.624570

|0.503700 |0.408570 |0.338580 |0.242970 |458.000000 |412.000000

|64072.176470 |51325.000000 |52266.974870 |863.899590 |1520.000000

|72903.375000 |62487.000000 |49724.439880 |491.240280 |988.000000

|1460.251170 |1371.000000 |589.416430 |253.813570 |590.000000

|478.860720 |449.000000 |198.798520 |327.000000 |213.353540 |0.018560

|0.022560 |0.094530 |0.648330 |0.418630 |0.466030 |0.491860 |0.889115

|0.88402 |0.895965 |38.354740 |37.916670 |21.913890 |490.980820

|1986.000000 |40.389660 |40.583330 |22.520945 |503.969830 |2068.000000

|0.691585 |0.527230 |0.038770 |0.013460 |0.095330

|75% |275779.000000 |109.000000 |42.000000 |77095.000000 |801.000000

|41.397686 |-79.811109 |3.376139e+07 |5.239845e+05 |5443.000000

|2673.000000 |2773.000000 |1260.078330 |1198.000000 |475.816770

|408.070310 |739.500000 |1.000000 |0.940555 |0.837145 |0.722145

|0.608065 |0.514865 |0.440830 |0.335655 |773.000000 |719.500000

|85838.910645 |70794.000000 |65349.873875 |1179.247160 |2018.000000

|96042.317750 |84769.000000 |60431.458305 |685.329205 |1350.000000

|1981.337375 |1876.000000 |788.049765 |387.330945 |896.000000

|631.376975 |600.000000 |266.645185 |501.000000 |342.763425 |0.036960

|0.042780 |0.143560 |0.737320 |0.553460 |0.677170 |0.718080 |0.939450

|0.94102 |0.944547 |41.409970 |43.000000 |22.960520 |665.967200

|2673.000000 |43.574030 |45.416670 |23.579895 |680.324678 |2774.000000

|0.817673 |0.606055 |0.064895 |0.027460 |0.129030

|max |294334.000000 |840.000000 |72.000000 |99925.000000 |989.000000

|67.074018 |-65.379332 |1.039510e+11 |2.453228e+10 |53812.000000

|27962.000000 |27250.000000 |3962.342290 |3972.000000 |1556.383030

|3060.247900 |6281.000000 |1.000000 |1.000000 |1.000000 |1.000000

|1.000000 |1.000000 |1.000000 |1.000000 |6648.000000 |6094.000000

|297142.857100 |296897.000000 |135902.619500 |10931.975610 |20395.000000

|242857.142900 |242720.000000 |111256.702500 |6904.496890 |14938.000000

|4462.342290 |4472.000000 |1596.206270 |4226.744200 |11670.000000

|1700.179110 |1702.000000 |820.968550 |11330.000000 |7107.064500

|1.000000 |1.000000 |1.000000 |1.000000 |1.000000 |1.000000 |1.000000

|1.000000 |1.00000 |1.000000 |77.759920 |80.166670 |31.060950

|12017.070440 |27962.000000 |79.837390 |82.250000 |30.241270

|6197.995200 |27250.000000 |1.000000 |1.000000 |0.714290 |0.714290

|1.000000

|===

----

+\*In[50]:\*+

[source, ipython3]

----

train\_df.dtypes

----

+\*Out[50]:\*+

----UID int64

COUNTYID int64

STATEID int64

state object

state\_ab object

city object

place object

type object

zip\_code int64

area\_code int64

lat float64

lng float64

ALand float64

AWater int64

pop int64

male\_pop int64

female\_pop int64

rent\_mean float64

rent\_median float64

rent\_stdev float64

rent\_sample\_weight float64

rent\_samples float64

rent\_gt\_10 float64

rent\_gt\_15 float64

rent\_gt\_20 float64

rent\_gt\_25 float64

rent\_gt\_30 float64

rent\_gt\_35 float64

rent\_gt\_40 float64

rent\_gt\_50 float64

universe\_samples int64

used\_samples int64

hi\_mean float64

hi\_median float64

hi\_stdev float64

hi\_sample\_weight float64

hi\_samples float64

family\_mean float64

family\_median float64

family\_stdev float64

family\_sample\_weight float64

family\_samples float64

hc\_mortgage\_mean float64

hc\_mortgage\_median float64

hc\_mortgage\_stdev float64

hc\_mortgage\_sample\_weight float64

hc\_mortgage\_samples float64

hc\_mean float64

hc\_median float64

hc\_stdev float64

hc\_samples float64

hc\_sample\_weight float64

home\_equity\_second\_mortgage float64

second\_mortgage float64

home\_equity float64

debt float64

second\_mortgage\_cdf float64

home\_equity\_cdf float64

debt\_cdf float64

hs\_degree float64

hs\_degree\_male float64

hs\_degree\_female float64

male\_age\_mean float64

male\_age\_median float64

male\_age\_stdev float64

male\_age\_sample\_weight float64

male\_age\_samples float64

female\_age\_mean float64

female\_age\_median float64

female\_age\_stdev float64

female\_age\_sample\_weight float64

female\_age\_samples float64

pct\_own float64

married float64

married\_snp float64

separated float64

divorced float64

dtype: object----

#all the columns which have null values have float64 data type

There are total 8 samples present in the data

1. rent\_samples - rent\_mean, rent\_median, rent\_stdev, rent\_sample\_weight

2. universe\_samples

3. used\_samples - rent\_gt\_10, rent\_gt\_15, rent\_gt\_20, rent\_gt\_25, rent\_gt\_30, rent\_gt\_35, rent\_gt\_40, rent\_gt\_50

4. hi\_samples - hi\_mean, hi\_median,hi\_stdev, hi\_sample\_weight

5. family\_samples - family\_mean, family\_median, family\_stdev, family\_sample\_weight

6. hc\_mortgage\_samples - hc\_mortgage\_mean, hc\_mortgage\_median, hc\_mortgage\_stdev, hc\_mortgage\_sample\_weight

7. hc\_samples - hc\_mean, hc\_median, hc\_stdev, hc\_samples, hc\_sample\_weight,

8. male\_age\_samples - male\_age\_mean, male\_age\_median, male\_age\_stdev, male\_age\_sample\_weight

9. female\_age\_samples - female\_age\_mean, female\_age\_median, female\_age\_stdev, female\_age\_sample\_weight

a. When we observe the NULL values, the number of null value in samples is same as that of their respective fields as listed above.

b. Since the minimum value in most of the columns is 0, 0 or any other values cannot be used to replace the NULL values.

c. hence the best possible option is to fill the mean value.

+\*In[51]:\*+

[source, ipython3]

----

train\_df.shape

----

+\*Out[51]:\*+

----(27019, 77)----

+\*In[52]:\*+

[source, ipython3]

----

col\_ref = ['rent\_mean', 'rent\_median', 'rent\_stdev', 'rent\_sample\_weight',

'rent\_samples', 'rent\_gt\_10', 'rent\_gt\_15', 'rent\_gt\_20', 'rent\_gt\_25',

'rent\_gt\_30', 'rent\_gt\_35', 'rent\_gt\_40', 'rent\_gt\_50',

'universe\_samples', 'used\_samples', 'hi\_mean', 'hi\_median', 'hi\_stdev',

'hi\_sample\_weight', 'hi\_samples', 'family\_mean', 'family\_median',

'family\_stdev', 'family\_sample\_weight', 'family\_samples',

'hc\_mortgage\_mean', 'hc\_mortgage\_median', 'hc\_mortgage\_stdev',

'hc\_mortgage\_sample\_weight', 'hc\_mortgage\_samples', 'hc\_mean',

'hc\_median', 'hc\_stdev', 'hc\_samples', 'hc\_sample\_weight',

'home\_equity\_second\_mortgage', 'second\_mortgage', 'home\_equity', 'debt',

'second\_mortgage\_cdf', 'home\_equity\_cdf', 'debt\_cdf', 'hs\_degree',

'hs\_degree\_male', 'hs\_degree\_female', 'male\_age\_mean',

'male\_age\_median', 'male\_age\_stdev', 'male\_age\_sample\_weight',

'male\_age\_samples', 'female\_age\_mean', 'female\_age\_median',

'female\_age\_stdev', 'female\_age\_sample\_weight', 'female\_age\_samples',

'pct\_own', 'married', 'married\_snp', 'separated', 'divorced']

----

+\*In[53]:\*+

[source, ipython3]

----

for col in col\_ref:

train\_df[col].fillna(value = train\_df[col].mean(), inplace = True)

----

+\*In[54]:\*+

[source, ipython3]

----

train\_df.isnull().sum(axis=0)[0:90]

----

+\*Out[54]:\*+

----UID 0

COUNTYID 0

STATEID 0

state 0

state\_ab 0

city 0

place 0

type 0

zip\_code 0

area\_code 0

lat 0

lng 0

ALand 0

AWater 0

pop 0

male\_pop 0

female\_pop 0

rent\_mean 0

rent\_median 0

rent\_stdev 0

rent\_sample\_weight 0

rent\_samples 0

rent\_gt\_10 0

rent\_gt\_15 0

rent\_gt\_20 0

rent\_gt\_25 0

rent\_gt\_30 0

rent\_gt\_35 0

rent\_gt\_40 0

rent\_gt\_50 0

universe\_samples 0

used\_samples 0

hi\_mean 0

hi\_median 0

hi\_stdev 0

hi\_sample\_weight 0

hi\_samples 0

family\_mean 0

family\_median 0

family\_stdev 0

family\_sample\_weight 0

family\_samples 0

hc\_mortgage\_mean 0

hc\_mortgage\_median 0

hc\_mortgage\_stdev 0

hc\_mortgage\_sample\_weight 0

hc\_mortgage\_samples 0

hc\_mean 0

hc\_median 0

hc\_stdev 0

hc\_samples 0

hc\_sample\_weight 0

home\_equity\_second\_mortgage 0

second\_mortgage 0

home\_equity 0

debt 0

second\_mortgage\_cdf 0

home\_equity\_cdf 0

debt\_cdf 0

hs\_degree 0

hs\_degree\_male 0

hs\_degree\_female 0

male\_age\_mean 0

male\_age\_median 0

male\_age\_stdev 0

male\_age\_sample\_weight 0

male\_age\_samples 0

female\_age\_mean 0

female\_age\_median 0

female\_age\_stdev 0

female\_age\_sample\_weight 0

female\_age\_samples 0

pct\_own 0

married 0

married\_snp 0

separated 0

divorced 0

dtype: int64----

+\*In[55]:\*+

[source, ipython3]

----

for col in col\_ref:

test\_df[col].fillna(value = test\_df[col].mean(), inplace = True)

test\_df.isnull().sum(axis=0)[0:90]

----

+\*Out[55]:\*+

----UID 0

COUNTYID 0

STATEID 0

state 0

state\_ab 0

city 0

place 0

type 0

zip\_code 0

area\_code 0

lat 0

lng 0

ALand 0

AWater 0

pop 0

male\_pop 0

female\_pop 0

rent\_mean 0

rent\_median 0

rent\_stdev 0

rent\_sample\_weight 0

rent\_samples 0

rent\_gt\_10 0

rent\_gt\_15 0

rent\_gt\_20 0

rent\_gt\_25 0

rent\_gt\_30 0

rent\_gt\_35 0

rent\_gt\_40 0

rent\_gt\_50 0

universe\_samples 0

used\_samples 0

hi\_mean 0

hi\_median 0

hi\_stdev 0

hi\_sample\_weight 0

hi\_samples 0

family\_mean 0

family\_median 0

family\_stdev 0

family\_sample\_weight 0

family\_samples 0

hc\_mortgage\_mean 0

hc\_mortgage\_median 0

hc\_mortgage\_stdev 0

hc\_mortgage\_sample\_weight 0

hc\_mortgage\_samples 0

hc\_mean 0

hc\_median 0

hc\_stdev 0

hc\_samples 0

hc\_sample\_weight 0

home\_equity\_second\_mortgage 0

second\_mortgage 0

home\_equity 0

debt 0

second\_mortgage\_cdf 0

home\_equity\_cdf 0

debt\_cdf 0

hs\_degree 0

hs\_degree\_male 0

hs\_degree\_female 0

male\_age\_mean 0

male\_age\_median 0

male\_age\_stdev 0

male\_age\_sample\_weight 0

male\_age\_samples 0

female\_age\_mean 0

female\_age\_median 0

female\_age\_stdev 0

female\_age\_sample\_weight 0

female\_age\_samples 0

pct\_own 0

married 0

married\_snp 0

separated 0

divorced 0

dtype: int64----

+\*In[56]:\*+

[source, ipython3]

----

train\_df.head()

----

+\*Out[56]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |COUNTYID |STATEID |state |state\_ab |city |place |type |zip\_code

|area\_code |lat |lng |ALand |AWater |pop |male\_pop |female\_pop

|rent\_mean |rent\_median |rent\_stdev |rent\_sample\_weight |rent\_samples

|rent\_gt\_10 |rent\_gt\_15 |rent\_gt\_20 |rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35

|rent\_gt\_40 |rent\_gt\_50 |universe\_samples |used\_samples |hi\_mean

|hi\_median |hi\_stdev |hi\_sample\_weight |hi\_samples |family\_mean

|family\_median |family\_stdev |family\_sample\_weight |family\_samples

|hc\_mortgage\_mean |hc\_mortgage\_median |hc\_mortgage\_stdev

|hc\_mortgage\_sample\_weight |hc\_mortgage\_samples |hc\_mean |hc\_median

|hc\_stdev |hc\_samples |hc\_sample\_weight |home\_equity\_second\_mortgage

|second\_mortgage |home\_equity |debt |second\_mortgage\_cdf

|home\_equity\_cdf |debt\_cdf |hs\_degree |hs\_degree\_male |hs\_degree\_female

|male\_age\_mean |male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|0 |267822 |53 |36 |New York |NY |Hamilton |Hamilton |City |13346 |315

|42.840812 |-75.501524 |202183361.0 |1699120 |5230 |2612 |2618

|769.38638 |784.0 |232.63967 |272.34441 |362.0 |0.86761 |0.79155

|0.59155 |0.45634 |0.42817 |0.18592 |0.15493 |0.12958 |387 |355

|63125.28406 |48120.0 |49042.01206 |1290.96240 |2024.0 |67994.14790

|53245.0 |47667.30119 |884.33516 |1491.0 |1414.80295 |1223.0 |641.22898

|377.83135 |867.0 |570.01530 |558.0 |270.11299 |770.0 |499.29293

|0.01588 |0.02077 |0.08919 |0.52963 |0.43658 |0.49087 |0.73341 |0.89288

|0.85880 |0.92434 |42.48574 |44.00000 |22.97306 |696.42136 |2612.0

|44.48629 |45.33333 |22.51276 |685.33845 |2618.0 |0.79046 |0.57851

|0.01882 |0.01240 |0.08770

|1 |246444 |141 |18 |Indiana |IN |South Bend |Roseland |City |46616 |574

|41.701441 |-86.266614 |1560828.0 |100363 |2633 |1349 |1284 |804.87924

|848.0 |253.46747 |312.58622 |513.0 |0.97410 |0.93227 |0.69920 |0.69920

|0.55179 |0.41235 |0.39044 |0.27888 |542 |502 |41931.92593 |35186.0

|31639.50203 |838.74664 |1127.0 |50670.10337 |43023.0 |34715.57548

|375.28798 |554.0 |864.41390 |784.0 |482.27020 |316.88320 |356.0

|351.98293 |336.0 |125.40457 |229.0 |189.60606 |0.02222 |0.02222

|0.04274 |0.60855 |0.42174 |0.70823 |0.58120 |0.90487 |0.86947 |0.94187

|34.84728 |32.00000 |20.37452 |323.90204 |1349.0 |36.48391 |37.58333

|23.43353 |267.23367 |1284.0 |0.52483 |0.34886 |0.01426 |0.01426

|0.09030

|2 |245683 |63 |18 |Indiana |IN |Danville |Danville |City |46122 |317

|39.792202 |-86.515246 |69561595.0 |284193 |6881 |3643 |3238 |742.77365

|703.0 |323.39011 |291.85520 |378.0 |0.95238 |0.88624 |0.79630 |0.66667

|0.39153 |0.39153 |0.28307 |0.15873 |459 |378 |84942.68317 |74964.0

|56811.62186 |1155.20980 |2488.0 |95262.51431 |85395.0 |49292.67664

|709.74925 |1889.0 |1506.06758 |1361.0 |731.89394 |699.41354 |1491.0

|556.45986 |532.0 |184.42175 |538.0 |323.35354 |0.00000 |0.00000

|0.09512 |0.73484 |1.00000 |0.46332 |0.28704 |0.94288 |0.94616 |0.93952

|39.38154 |40.83333 |22.89769 |888.29730 |3643.0 |42.15810 |42.83333

|23.94119 |707.01963 |3238.0 |0.85331 |0.64745 |0.02830 |0.01607

|0.10657

|3 |279653 |127 |72 |Puerto Rico |PR |San Juan |Guaynabo |Urban |927

|787 |18.396103 |-66.104169 |1105793.0 |0 |2700 |1141 |1559 |803.42018

|782.0 |297.39258 |259.30316 |368.0 |0.94693 |0.87151 |0.69832 |0.61732

|0.51397 |0.46927 |0.35754 |0.32961 |438 |358 |48733.67116 |37845.0

|45100.54010 |928.32193 |1267.0 |56401.68133 |44399.0 |41082.90515

|490.18479 |729.0 |1175.28642 |1101.0 |428.98751 |261.28471 |437.0

|288.04047 |247.0 |185.55887 |392.0 |314.90566 |0.01086 |0.01086

|0.01086 |0.52714 |0.53057 |0.82530 |0.73727 |0.91500 |0.90755 |0.92043

|48.64749 |48.91667 |23.05968 |274.98956 |1141.0 |47.77526 |50.58333

|24.32015 |362.20193 |1559.0 |0.65037 |0.47257 |0.02021 |0.02021

|0.10106

|4 |247218 |161 |20 |Kansas |KS |Manhattan |Manhattan City |City |66502

|785 |39.195573 |-96.569366 |2554403.0 |0 |5637 |2586 |3051 |938.56493

|881.0 |392.44096 |1005.42886 |1704.0 |0.99286 |0.98247 |0.91688

|0.84740 |0.78247 |0.60974 |0.55455 |0.44416 |1725 |1540 |31834.15466

|22497.0 |34046.50907 |1548.67477 |1983.0 |54053.42396 |50272.0

|39609.12605 |244.08903 |395.0 |1192.58759 |1125.0 |327.49674 |76.61052

|134.0 |443.68855 |444.0 |76.12674 |124.0 |79.55556 |0.05426 |0.05426

|0.05426 |0.51938 |0.18332 |0.65545 |0.74967 |1.00000 |1.00000 |1.00000

|26.07533 |22.41667 |11.84399 |1296.89877 |2586.0 |24.17693 |21.58333

|11.10484 |1854.48652 |3051.0 |0.13046 |0.12356 |0.00000 |0.00000

|0.03109

|===

----

+\*In[57]:\*+

[source, ipython3]

----

test\_df.head()

----

+\*Out[57]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |COUNTYID |STATEID |state |state\_ab |city |place |type |zip\_code

|area\_code |lat |lng |ALand |AWater |pop |male\_pop |female\_pop

|rent\_mean |rent\_median |rent\_stdev |rent\_sample\_weight |rent\_samples

|rent\_gt\_10 |rent\_gt\_15 |rent\_gt\_20 |rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35

|rent\_gt\_40 |rent\_gt\_50 |universe\_samples |used\_samples |hi\_mean

|hi\_median |hi\_stdev |hi\_sample\_weight |hi\_samples |family\_mean

|family\_median |family\_stdev |family\_sample\_weight |family\_samples

|hc\_mortgage\_mean |hc\_mortgage\_median |hc\_mortgage\_stdev

|hc\_mortgage\_sample\_weight |hc\_mortgage\_samples |hc\_mean |hc\_median

|hc\_stdev |hc\_samples |hc\_sample\_weight |home\_equity\_second\_mortgage

|second\_mortgage |home\_equity |debt |second\_mortgage\_cdf

|home\_equity\_cdf |debt\_cdf |hs\_degree |hs\_degree\_male |hs\_degree\_female

|male\_age\_mean |male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|0 |255504 |163 |26 |Michigan |MI |Detroit |Dearborn Heights City |CDP

|48239 |313 |42.346422 |-83.252823 |2711280 |39555 |3417 |1479 |1938

|858.57169 |859.0 |232.39082 |276.07497 |424.0 |1.00000 |0.95696

|0.85316 |0.85316 |0.85316 |0.85316 |0.76962 |0.63544 |435 |395

|48899.52121 |38746.0 |44392.20902 |798.02401 |1180.0 |53802.87122

|45167.0 |43756.56479 |464.30972 |769.0 |1139.24548 |1109.0 |336.47710

|262.67011 |474.0 |488.51323 |436.0 |192.75147 |271.0 |189.18182

|0.06443 |0.06443 |0.07651 |0.63624 |0.14111 |0.55087 |0.51965 |0.91047

|0.92010 |0.90391 |33.37131 |27.83333 |22.36768 |334.30978 |1479.0

|34.78682 |33.75000 |21.58531 |416.48097 |1938.0 |0.70252 |0.28217

|0.05910 |0.03813 |0.14299

|1 |252676 |1 |23 |Maine |ME |Auburn |Auburn City |City |4210 |207

|44.100724 |-70.257832 |14778785 |2705204 |3796 |1846 |1950 |832.68625

|750.0 |267.22342 |183.32299 |245.0 |1.00000 |1.00000 |0.86611 |0.67364

|0.30962 |0.30962 |0.30962 |0.27197 |275 |239 |72335.33234 |61008.0

|51895.81159 |922.82969 |1722.0 |85642.22095 |74759.0 |49156.72870

|482.99945 |1147.0 |1533.25988 |1438.0 |536.61118 |373.96188 |937.0

|661.31296 |668.0 |201.31365 |510.0 |279.69697 |0.01175 |0.01175

|0.14375 |0.64755 |0.52310 |0.26442 |0.49359 |0.94290 |0.92832 |0.95736

|43.88680 |46.08333 |22.90302 |427.10824 |1846.0 |44.23451 |46.66667

|22.37036 |532.03505 |1950.0 |0.85128 |0.64221 |0.02338 |0.00000

|0.13377

|2 |276314 |15 |42 |Pennsylvania |PA |Pine City |Millerton |Borough

|14871 |607 |41.948556 |-76.783808 |258903666 |863840 |3944 |2065 |1879

|816.00639 |755.0 |416.25699 |141.39063 |217.0 |0.97573 |0.93204

|0.78641 |0.71845 |0.63592 |0.47573 |0.43689 |0.32524 |245 |206

|58501.15901 |51648.0 |45245.27248 |893.07759 |1461.0 |65694.06582

|57186.0 |44239.31893 |619.73962 |1084.0 |1254.54462 |1089.0 |596.85204

|340.45884 |552.0 |397.44466 |356.0 |189.40372 |664.0 |534.16737

|0.01069 |0.01316 |0.06497 |0.45395 |0.51066 |0.60484 |0.83848 |0.89238

|0.86003 |0.92463 |39.81661 |41.91667 |24.29111 |499.10080 |2065.0

|41.62426 |44.50000 |22.86213 |453.11959 |1879.0 |0.81897 |0.59961

|0.01746 |0.01358 |0.10026

|3 |248614 |231 |21 |Kentucky |KY |Monticello |Monticello City |City

|42633 |606 |36.746009 |-84.766870 |501694825 |2623067 |2508 |1427 |1081

|418.68937 |385.0 |156.92024 |88.95960 |93.0 |1.00000 |0.93548 |0.93548

|0.64516 |0.55914 |0.46237 |0.46237 |0.36559 |153 |93 |38237.55059

|31612.0 |34527.61607 |775.17947 |957.0 |44156.38709 |34687.0

|34899.74300 |535.21987 |689.0 |862.65763 |749.0 |624.42157 |299.56752

|337.0 |200.88113 |180.0 |91.56490 |467.0 |454.85404 |0.00995 |0.00995

|0.01741 |0.41915 |0.53770 |0.80931 |0.87403 |0.60908 |0.56584 |0.65947

|41.81638 |43.00000 |24.65325 |333.57733 |1427.0 |44.81200 |48.00000

|21.03155 |263.94320 |1081.0 |0.84609 |0.56953 |0.05492 |0.04694

|0.12489

|4 |286865 |355 |48 |Texas |TX |Corpus Christi |Edroy |Town |78410 |361

|27.882461 |-97.678586 |13796057 |497689 |6230 |3274 |2956 |1031.63763

|997.0 |326.76727 |277.39844 |624.0 |0.72276 |0.66506 |0.53526 |0.38301

|0.18910 |0.16667 |0.14263 |0.11058 |660 |624 |114456.07790 |94211.0

|81950.95692 |836.30759 |2404.0 |123527.02420 |103898.0 |72173.55823

|507.42257 |1738.0 |1996.41425 |1907.0 |740.21168 |319.97570 |1102.0

|867.57713 |804.0 |376.20236 |642.0 |333.91919 |0.00000 |0.00000

|0.03440 |0.63188 |1.00000 |0.74519 |0.52943 |0.86297 |0.87969 |0.84466

|42.13301 |43.75000 |22.69502 |833.57435 |3274.0 |40.66618 |42.66667

|21.30900 |709.90829 |2956.0 |0.79077 |0.57620 |0.01726 |0.00588

|0.16379

|===

----

+\*In[58]:\*+

[source, ipython3]

----

train\_df.describe()

----

+\*Out[58]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |COUNTYID |STATEID |zip\_code |area\_code |lat |lng |ALand |AWater

|pop |male\_pop |female\_pop |rent\_mean |rent\_median |rent\_stdev

|rent\_sample\_weight |rent\_samples |rent\_gt\_10 |rent\_gt\_15 |rent\_gt\_20

|rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35 |rent\_gt\_40 |rent\_gt\_50

|universe\_samples |used\_samples |hi\_mean |hi\_median |hi\_stdev

|hi\_sample\_weight |hi\_samples |family\_mean |family\_median |family\_stdev

|family\_sample\_weight |family\_samples |hc\_mortgage\_mean

|hc\_mortgage\_median |hc\_mortgage\_stdev |hc\_mortgage\_sample\_weight

|hc\_mortgage\_samples |hc\_mean |hc\_median |hc\_stdev |hc\_samples

|hc\_sample\_weight |home\_equity\_second\_mortgage |second\_mortgage

|home\_equity |debt |second\_mortgage\_cdf |home\_equity\_cdf |debt\_cdf

|hs\_degree |hs\_degree\_male |hs\_degree\_female |male\_age\_mean

|male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|count |27019.000000 |27019.000000 |27019.000000 |27019.000000

|27019.000000 |27019.000000 |27019.000000 |2.701900e+04 |2.701900e+04

|27019.000000 |27019.000000 |27019.000000 |27019.000000 |27019.000000

|27019.000000 |27019.000000 |27019.000000 |27019.00000 |27019.000000

|27019.000000 |27019.000000 |27019.000000 |27019.000000 |27019.000000

|27019.000000 |27019.000000 |27019.000000 |27019.000000 |27019.000000

|27019.000000 |27019.000000 |27019.000000 |27019.000000 |27019.000000

|27019.000000 |27019.000000 |27019.000000 |27019.000000 |27019.000000

|27019.000000 |27019.000000 |27019.000000 |27019.000000 |27019.000000

|27019.000000 |27019.000000 |27019.000000 |27019.000000 |27019.000000

|27019.000000 |27019.000000 |27019.000000 |27019.000000 |27019.000000

|27019.000000 |27019.000000 |27019.000000 |27019.000000 |27019.000000

|27019.000000 |27019.000000 |27019.000000 |27019.000000 |27019.000000

|27019.000000 |27019.000000 |27019.000000 |27019.000000 |27019.000000

|27019.000000 |27019.000000 |27019.000000

|mean |257310.991673 |85.592139 |28.246567 |50132.614568 |596.381509

|37.530584 |-91.306019 |1.302186e+08 |6.495841e+06 |4347.275140

|2137.612569 |2209.662571 |1055.344777 |1007.852781 |394.419068

|294.987443 |546.209926 |0.95788 |0.867144 |0.739364 |0.612774 |0.499869

|0.410941 |0.345364 |0.254420 |577.016063 |530.991710 |70500.318611

|57624.340766 |54501.465013 |924.151692 |1610.055502 |79045.667606

|69333.490290 |50783.707702 |533.755939 |1064.829899 |1629.260500

|1550.778996 |622.686711 |287.836433 |670.452562 |540.646148 |513.453810

|218.821586 |370.768542 |255.061125 |0.025689 |0.029951 |0.100899

|0.629621 |0.466927 |0.476681 |0.499327 |0.858819 |0.852470 |0.865284

|38.373671 |38.114800 |21.534969 |534.194158 |2138.087365 |40.354614

|40.395275 |22.213313 |544.289544 |2211.217519 |0.642269 |0.509312

|0.047344 |0.019073 |0.100385

|std |21340.758358 |98.176550 |16.371166 |29539.429679 |232.438332

|5.576366 |16.329382 |1.282239e+09 |2.198107e+08 |2113.892839

|1069.804237 |1084.636509 |436.637101 |442.946881 |186.839550

|269.463616 |455.112889 |0.06246 |0.109037 |0.143248 |0.159762 |0.163389

|0.159593 |0.152663 |0.137245 |459.049696 |444.617372 |30057.465220

|29013.590268 |17546.328862 |449.583578 |746.972553 |31265.826459

|33357.579382 |14165.569774 |286.098971 |556.335553 |619.040656

|648.176109 |236.465911 |194.157880 |461.628280 |219.906175 |229.905656

|90.722248 |248.997572 |188.646458 |0.030570 |0.033422 |0.068390

|0.153785 |0.293048 |0.254273 |0.262314 |0.111765 |0.119953 |0.111325

|5.560516 |7.843273 |2.453759 |287.598841 |1069.329761 |5.819975

|7.982550 |2.446941 |280.582493 |1083.051393 |0.223914 |0.135681

|0.037151 |0.020741 |0.048800

|min |220342.000000 |1.000000 |1.000000 |602.000000 |201.000000

|17.929085 |-165.453872 |4.113400e+04 |0.000000e+00 |3.000000 |0.000000

|0.000000 |117.150000 |104.000000 |18.257420 |0.343000 |4.000000

|0.00000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000

|0.000000 |0.000000 |0.000000 |4999.846690 |4790.000000 |1825.741860

|0.114260 |3.000000 |5374.842520 |5278.000000 |1825.741860 |0.199960

|3.000000 |234.650000 |237.000000 |36.514840 |0.198400 |1.000000

|53.594610 |53.000000 |18.257420 |2.000000 |0.614040 |0.000000 |0.000000

|0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.186520 |0.000000

|0.000000 |12.145830 |9.750000 |0.962770 |0.745760 |3.000000 |16.008330

|13.250000 |0.556780 |0.664700 |2.000000 |0.000000 |0.000000 |0.000000

|0.000000 |0.000000

|25% |238819.500000 |29.000000 |13.000000 |27023.000000 |405.000000

|33.910638 |-97.820012 |1.806570e+06 |0.000000e+00 |2915.500000

|1418.000000 |1473.000000 |744.534925 |703.000000 |264.120350

|102.400075 |222.000000 |0.94079 |0.819540 |0.662380 |0.517455 |0.396690

|0.307690 |0.243820 |0.161290 |255.000000 |213.500000 |49273.848990

|37516.000000 |42194.979190 |602.038105 |1099.000000 |57006.613295

|46330.000000 |40986.939865 |332.802170 |689.000000 |1160.927890

|1069.000000 |442.127815 |149.570035 |350.000000 |390.466480 |362.500000

|155.020205 |195.000000 |122.472090 |0.005275 |0.007880 |0.050000

|0.539875 |0.249890 |0.266520 |0.283635 |0.808425 |0.795725 |0.818615

|35.050285 |32.916670 |20.601185 |346.553985 |1419.000000 |36.935390

|35.000000 |21.324830 |356.593825 |1474.000000 |0.505685 |0.426585

|0.020830 |0.004560 |0.066045

|50% |257187.000000 |63.000000 |28.000000 |47904.000000 |614.000000

|38.767808 |-86.591218 |4.879552e+06 |2.767400e+04 |4063.000000

|1986.000000 |2067.000000 |955.026680 |899.000000 |347.346940

|220.273470 |426.000000 |0.97683 |0.887600 |0.757580 |0.624030 |0.503030

|0.409300 |0.339220 |0.243510 |458.000000 |412.000000 |64166.450590

|51396.000000 |52313.707630 |865.043200 |1522.000000 |73032.951660

|62587.000000 |49789.669880 |492.272740 |989.000000 |1467.589120

|1379.000000 |592.895460 |256.665360 |596.000000 |481.516610 |452.000000

|199.790600 |332.000000 |215.767680 |0.018800 |0.022800 |0.095180

|0.646560 |0.421370 |0.468720 |0.496080 |0.889110 |0.883930 |0.895830

|38.355400 |37.916670 |21.913680 |491.108090 |1987.000000 |40.387160

|40.583330 |22.519290 |504.099040 |2070.000000 |0.690900 |0.527210

|0.038780 |0.013470 |0.095350

|75% |275779.000000 |109.000000 |42.000000 |77095.000000 |801.000000

|41.397686 |-79.811109 |3.376139e+07 |5.239845e+05 |5443.000000

|2673.000000 |2773.000000 |1259.016375 |1197.000000 |475.060710

|406.830555 |738.000000 |1.00000 |0.940320 |0.836810 |0.721760 |0.607640

|0.514490 |0.440425 |0.335240 |773.000000 |719.500000 |85791.313770

|70698.000000 |65320.677295 |1178.064865 |2016.000000 |95959.471190

|84625.500000 |60393.714810 |684.521770 |1349.000000 |1972.362855

|1870.000000 |785.740955 |385.330270 |891.000000 |628.154350 |597.000000

|265.485970 |498.000000 |340.855660 |0.036780 |0.042550 |0.143115

|0.736355 |0.552715 |0.674830 |0.716285 |0.939440 |0.940995 |0.944525

|41.407320 |43.000000 |22.960415 |665.943575 |2673.000000 |43.570560

|45.416670 |23.579335 |680.258950 |2773.000000 |0.817390 |0.606035

|0.064885 |0.027450 |0.129030

|max |294334.000000 |840.000000 |72.000000 |99925.000000 |989.000000

|67.074018 |-65.379332 |1.039510e+11 |2.453228e+10 |53812.000000

|27962.000000 |27250.000000 |3962.342290 |3972.000000 |1556.383030

|3060.247900 |6281.000000 |1.00000 |1.000000 |1.000000 |1.000000

|1.000000 |1.000000 |1.000000 |1.000000 |6648.000000 |6094.000000

|297142.857100 |296897.000000 |135902.619500 |10931.975610 |20395.000000

|242857.142900 |242720.000000 |111256.702500 |6904.496890 |14938.000000

|4462.342290 |4472.000000 |1596.206270 |4226.744200 |11670.000000

|1700.179110 |1702.000000 |820.968550 |11330.000000 |7107.064500

|1.000000 |1.000000 |1.000000 |1.000000 |1.000000 |1.000000 |1.000000

|1.000000 |1.000000 |1.000000 |77.759920 |80.166670 |31.060950

|12017.070440 |27962.000000 |79.837390 |82.250000 |30.241270

|6197.995200 |27250.000000 |1.000000 |1.000000 |0.714290 |0.714290

|1.000000

|===

----

+\*In[59]:\*+

[source, ipython3]

----

test\_df.describe()

----

+\*Out[59]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |COUNTYID |STATEID |zip\_code |area\_code |lat |lng |ALand |AWater

|pop |male\_pop |female\_pop |rent\_mean |rent\_median |rent\_stdev

|rent\_sample\_weight |rent\_samples |rent\_gt\_10 |rent\_gt\_15 |rent\_gt\_20

|rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35 |rent\_gt\_40 |rent\_gt\_50

|universe\_samples |used\_samples |hi\_mean |hi\_median |hi\_stdev

|hi\_sample\_weight |hi\_samples |family\_mean |family\_median |family\_stdev

|family\_sample\_weight |family\_samples |hc\_mortgage\_mean

|hc\_mortgage\_median |hc\_mortgage\_stdev |hc\_mortgage\_sample\_weight

|hc\_mortgage\_samples |hc\_mean |hc\_median |hc\_stdev |hc\_samples

|hc\_sample\_weight |home\_equity\_second\_mortgage |second\_mortgage

|home\_equity |debt |second\_mortgage\_cdf |home\_equity\_cdf |debt\_cdf

|hs\_degree |hs\_degree\_male |hs\_degree\_female |male\_age\_mean

|male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|count |11603.000000 |11603.000000 |11603.000000 |11603.000000

|11603.000000 |11603.000000 |11603.000000 |1.160300e+04 |1.160300e+04

|11603.000000 |11603.000000 |11603.000000 |11603.000000 |11603.000000

|11603.000000 |11603.000000 |11603.000000 |11603.000000 |11603.000000

|11603.000000 |11603.000000 |11603.000000 |11603.000000 |11603.000000

|11603.000000 |11603.000000 |11603.000000 |11603.000000 |11603.000000

|11603.000000 |11603.000000 |11603.000000 |11603.000000 |11603.000000

|11603.000000 |11603.000000 |11603.000000 |11603.000000 |11603.000000

|11603.000000 |11603.000000 |11603.000000 |11603.000000 |11603.000000

|11603.000000 |11603.000000 |11603.000000 |11603.000000 |11603.000000

|11603.000000 |11603.000000 |11603.000000 |11603.000000 |11603.000000

|11603.000000 |11603.000000 |11603.000000 |11603.000000 |11603.000000

|11603.000000 |11603.000000 |11603.000000 |11603.000000 |11603.000000

|11603.000000 |11603.000000 |11603.00000 |11603.000000 |11603.000000

|11603.000000 |11603.000000 |11603.000000

|mean |257510.397570 |85.809015 |28.467379 |50171.778247 |593.220633

|37.420387 |-91.359068 |1.079633e+08 |5.187847e+06 |4399.982591

|2168.605619 |2231.376971 |1053.847576 |1006.738370 |394.563417

|304.139569 |562.926276 |0.957542 |0.867768 |0.742558 |0.614289

|0.501172 |0.412973 |0.347015 |0.255512 |592.700250 |546.274239

|70191.203114 |57377.772071 |54193.528896 |935.673418 |1625.779334

|78691.369385 |69048.913262 |50439.623539 |540.616037 |1074.098165

|1636.093291 |1559.312194 |621.876346 |289.465464 |673.863016

|539.056097 |512.200876 |218.017066 |369.976775 |255.330882 |0.025821

|0.030225 |0.101575 |0.631771 |0.466584 |0.475065 |0.494432 |0.856022

|0.849294 |0.863076 |38.165621 |37.851545 |21.446900 |542.772254

|2169.166466 |40.128878 |40.152660 |22.160832 |550.062635 |2233.87981

|0.634921 |0.506016 |0.047903 |0.019342 |0.099274

|std |21477.738651 |99.370939 |16.593153 |29752.939038 |232.055148

|5.611077 |16.385709 |7.518341e+08 |1.529580e+08 |2096.839269

|1075.209593 |1073.483604 |432.648578 |439.522568 |188.589326

|279.708246 |472.640225 |0.062845 |0.107240 |0.141992 |0.161011

|0.165217 |0.160824 |0.153471 |0.137183 |475.825441 |461.964752

|30534.158349 |29577.185584 |17742.552979 |456.538850 |745.490305

|31848.049461 |34001.610628 |14294.308424 |288.115199 |549.309314

|629.741457 |659.398291 |238.980486 |195.666021 |457.991800 |224.396536

|235.516933 |92.261146 |247.513357 |188.658288 |0.030349 |0.033462

|0.069453 |0.155812 |0.294861 |0.255231 |0.263050 |0.114183 |0.122286

|0.112874 |5.556051 |7.778184 |2.543056 |295.277373 |1074.643806

|5.817413 |7.947693 |2.519982 |278.991858 |1070.87898 |0.230859

|0.139157 |0.038568 |0.021412 |0.048434

|min |220336.000000 |1.000000 |1.000000 |601.000000 |201.000000

|17.965835 |-166.770979 |8.299000e+03 |0.000000e+00 |3.000000 |0.000000

|0.000000 |147.548100 |104.000000 |18.257420 |0.392790 |3.000000

|0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000

|0.000000 |0.000000 |0.000000 |4999.846690 |4790.000000 |1825.741860

|0.399920 |3.000000 |5374.842520 |5278.000000 |1825.741860 |0.266610

|4.000000 |349.500000 |349.000000 |36.514840 |0.595190 |2.000000

|53.594610 |53.000000 |18.257420 |2.000000 |0.491230 |0.000000 |0.000000

|0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000 |0.000000

|0.199710 |17.009880 |9.750000 |0.737110 |0.745760 |4.000000 |15.360240

|12.833330 |0.737110 |0.251910 |3.00000 |0.000000 |0.000000 |0.000000

|0.000000 |0.000000

|25% |238812.000000 |29.000000 |13.000000 |25941.500000 |404.000000

|33.921714 |-97.828018 |1.721184e+06 |0.000000e+00 |2966.000000

|1445.000000 |1498.000000 |742.264780 |705.000000 |263.017880

|104.532310 |227.000000 |0.940800 |0.821225 |0.666670 |0.517670

|0.398345 |0.308640 |0.242355 |0.160850 |259.000000 |219.000000

|48910.114490 |37053.500000 |41763.873790 |613.257530 |1113.000000

|56256.077425 |45850.000000 |40498.623400 |339.093040 |697.000000

|1156.064170 |1072.000000 |439.424845 |148.682150 |347.000000

|387.919530 |359.000000 |153.313620 |192.000000 |120.383305 |0.005395

|0.007995 |0.050335 |0.542715 |0.247835 |0.266470 |0.277095 |0.803205

|0.790590 |0.814410 |34.945030 |32.666670 |20.514145 |355.323875

|1445.500000 |36.754580 |34.750000 |21.279365 |363.528895 |1501.00000

|0.494280 |0.422590 |0.020890 |0.004505 |0.064690

|50% |257639.000000 |61.000000 |28.000000 |47394.000000 |612.000000

|38.618817 |-86.669044 |4.823082e+06 |2.276400e+04 |4135.000000

|2020.000000 |2098.000000 |955.519670 |900.000000 |350.601730

|230.134190 |443.000000 |0.976680 |0.888680 |0.762540 |0.626820

|0.505680 |0.412973 |0.343270 |0.244540 |473.000000 |427.000000

|63949.915200 |51143.000000 |52039.454510 |879.026000 |1533.000000

|72976.408950 |62112.000000 |49500.776830 |497.880930 |999.000000

|1476.259080 |1385.000000 |591.748660 |259.673510 |601.000000

|478.901980 |449.000000 |199.906370 |332.000000 |215.535350 |0.019120

|0.022900 |0.096390 |0.647980 |0.421390 |0.466650 |0.492630 |0.886340

|0.880990 |0.893450 |38.201950 |37.833330 |21.886490 |499.893890

|2020.000000 |40.197990 |40.333330 |22.473000 |509.482470 |2100.00000

|0.686950 |0.525420 |0.038670 |0.013870 |0.094450

|75% |276258.000000 |109.000000 |42.000000 |77409.000000 |787.000000

|41.238961 |-79.761644 |3.234604e+07 |4.862615e+05 |5485.500000

|2697.000000 |2798.500000 |1256.801020 |1191.000000 |474.663570

|418.681650 |760.000000 |1.000000 |0.939330 |0.839010 |0.725580

|0.611620 |0.516340 |0.444005 |0.339680 |794.000000 |743.000000

|85347.211905 |70409.500000 |64861.531265 |1193.625590 |2030.000000

|95517.047185 |84156.000000 |60278.246910 |688.456160 |1357.000000

|1980.944995 |1874.500000 |784.023775 |385.106825 |901.000000

|625.409615 |593.500000 |264.149530 |498.000000 |342.616165 |0.037040

|0.042880 |0.143140 |0.739380 |0.551965 |0.673695 |0.711170 |0.940070

|0.940130 |0.944780 |41.183985 |42.666670 |22.939240 |676.322005

|2697.000000 |43.496235 |45.333330 |23.548580 |685.461915 |2798.50000

|0.814970 |0.605670 |0.065275 |0.027905 |0.128400

|max |294333.000000 |810.000000 |72.000000 |99929.000000 |989.000000

|64.804269 |-65.695344 |5.520166e+10 |1.212570e+10 |39454.000000

|27962.000000 |15466.000000 |3962.342290 |3972.000000 |1720.718990

|4112.122370 |7634.000000 |1.000000 |1.000000 |1.000000 |1.000000

|1.000000 |1.000000 |1.000000 |1.000000 |7634.000000 |7336.000000

|221622.723500 |242249.000000 |124534.013900 |8133.778720 |12316.000000

|242857.142900 |242720.000000 |105579.486100 |4888.944600 |6658.000000

|4462.342290 |4472.000000 |1814.113980 |1936.551660 |5033.000000

|1700.179110 |1702.000000 |782.862850 |3965.000000 |2878.131310

|1.000000 |1.000000 |1.000000 |1.000000 |1.000000 |1.000000 |1.000000

|1.000000 |1.000000 |1.000000 |83.358330 |83.333330 |27.920410

|12017.070440 |27962.000000 |90.107940 |90.166670 |29.626680

|4145.557870 |15466.00000 |1.000000 |1.000000 |0.714290 |0.714290

|0.362750

|===

----

Exploratory Data Analysis (EDA):

4. debt analysis.

a) Explore the top 2,500 locations where the percentage of households with a second mortgage is the highest and percent ownership is above 10 percent. Visualize using geo-map. You may keep the upper limit for the percent of households with a second mortgage to 50 percent

+\*In[60]:\*+

[source, ipython3]

----

train\_df.nlargest(2500, ['second\_mortgage', 'pct\_own'])

----

+\*Out[60]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |COUNTYID |STATEID |state |state\_ab |city |place |type |zip\_code

|area\_code |lat |lng |ALand |AWater |pop |male\_pop |female\_pop

|rent\_mean |rent\_median |rent\_stdev |rent\_sample\_weight |rent\_samples

|rent\_gt\_10 |rent\_gt\_15 |rent\_gt\_20 |rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35

|rent\_gt\_40 |rent\_gt\_50 |universe\_samples |used\_samples |hi\_mean

|hi\_median |hi\_stdev |hi\_sample\_weight |hi\_samples |family\_mean

|family\_median |family\_stdev |family\_sample\_weight |family\_samples

|hc\_mortgage\_mean |hc\_mortgage\_median |hc\_mortgage\_stdev

|hc\_mortgage\_sample\_weight |hc\_mortgage\_samples |hc\_mean |hc\_median

|hc\_stdev |hc\_samples |hc\_sample\_weight |home\_equity\_second\_mortgage

|second\_mortgage |home\_equity |debt |second\_mortgage\_cdf

|home\_equity\_cdf |debt\_cdf |hs\_degree |hs\_degree\_male |hs\_degree\_female

|male\_age\_mean |male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|17289 |249744 |115 |22 |Louisiana |LA |Fort Polk |Fort Polk North |City

|71459 |337 |31.099609 |-93.202088 |17983022.0 |17262 |2391 |1250 |1141

|1015.01390 |966.0 |204.94664 |371.10883 |886.0 |1.00000 |0.85402

|0.76322 |0.41839 |0.37011 |0.30805 |0.26322 |0.15287 |934 |870

|51307.09657 |46500.0 |32444.96129 |699.35004 |941.0 |51669.55301

|46593.0 |32222.92387 |638.85288 |860.0 |3824.65000 |3813.0 |229.67915

|1.38878 |7.0 |540.646148 |513.45381 |218.821586 |370.768542 |255.061125

|1.00000 |1.00000 |1.00000 |1.00000 |0.00000 |0.00000 |0.00000 |0.97861

|0.97593 |0.98228 |22.20676 |23.08333 |10.11288 |440.77883 |1250.0

|20.50383 |23.00000 |10.77346 |338.31385 |1141.0 |0.01213 |0.95068

|0.03253 |0.00000 |0.02099

|1211 |247937 |93 |21 |Kentucky |KY |Fort Knox |Fort Knox |City |40121

|502 |37.879246 |-85.961748 |10129479.0 |7700 |3937 |1808 |2129

|1196.96888 |1173.0 |329.32165 |342.04691 |1006.0 |0.97001 |0.88418

|0.69907 |0.49121 |0.36711 |0.29886 |0.19752 |0.13650 |1110 |967

|61478.21044 |53248.0 |37451.20023 |670.39860 |1125.0 |63273.30137

|55270.0 |36398.66402 |590.13663 |1035.0 |1412.00000 |1406.0 |114.79148

|5.96386 |15.0 |540.646148 |513.45381 |218.821586 |370.768542

|255.061125 |1.00000 |1.00000 |1.00000 |1.00000 |0.00000 |0.00000

|0.00000 |0.93870 |0.94002 |0.93736 |24.24531 |25.91667 |15.13769

|372.04805 |1808.0 |22.33822 |20.91667 |14.99343 |509.74299 |2129.0

|0.01067 |0.85223 |0.04275 |0.00000 |0.05576

|7413 |290029 |510 |51 |Virginia |VA |Alexandria |Bailey's Crossroads

|Town |22311 |703 |38.828707 |-77.121251 |889826.0 |13370 |3904 |2020

|1884 |1694.14150 |1678.0 |449.98984 |232.87054 |1413.0 |0.98633

|0.95108 |0.76043 |0.65612 |0.55036 |0.42518 |0.33309 |0.27122 |1413

|1390 |74471.20336 |60591.0 |50854.52041 |767.39763 |1420.0 |62928.66788

|46820.0 |44669.09955 |502.38117 |811.0 |1412.00000 |1406.0 |114.79148

|2.78313 |7.0 |540.646148 |513.45381 |218.821586 |370.768542 |255.061125

|1.00000 |1.00000 |1.00000 |1.00000 |0.00000 |0.00000 |0.00000 |0.87954

|0.86780 |0.89487 |34.85774 |34.66667 |16.18076 |460.88590 |2020.0

|30.27329 |28.91667 |16.61773 |432.84916 |1884.0 |0.00666 |0.45629

|0.19114 |0.06993 |0.04254

|16834 |248394 |163 |21 |Kentucky |KY |Fort Knox |Fort Knox |City |40121

|502 |37.901474 |-85.986474 |4114191.0 |8502 |2661 |1280 |1381

|1118.09185 |1039.0 |272.97314 |262.35860 |718.0 |1.00000 |0.93454

|0.78552 |0.57103 |0.49025 |0.36351 |0.24234 |0.11838 |796 |718

|48184.50734 |44822.0 |24841.69744 |578.68043 |798.0 |50819.49924

|46983.0 |23710.51317 |526.18373 |744.0 |449.50000 |449.0 |36.51484

|2.00000 |2.0 |540.646148 |513.45381 |218.821586 |370.768542 |255.061125

|1.00000 |1.00000 |1.00000 |1.00000 |0.00000 |0.00000 |0.00000 |0.96933

|0.96314 |0.97500 |22.84809 |24.50000 |13.06332 |289.82619 |1280.0

|23.05701 |25.25000 |13.75385 |312.77660 |1381.0 |0.00225 |0.82381

|0.03214 |0.00000 |0.01429

|14014 |264403 |31 |34 |New Jersey |NJ |Passaic |Garfield City |City

|7055 |973 |40.867944 |-74.114633 |480161.0 |77188 |5103 |2577 |2526

|999.11901 |986.0 |334.07761 |641.55220 |1375.0 |0.97047 |0.87879

|0.82595 |0.75214 |0.66511 |0.58275 |0.48096 |0.35120 |1430 |1287

|37443.40200 |28053.0 |36885.04214 |1124.57237 |1453.0 |40967.82646

|29340.0 |38042.45182 |722.67532 |968.0 |1787.00000 |1781.0 |160.24070

|2.77756 |14.0 |649.500000 |649.00000 |36.514840 |9.000000 |4.454550

|0.00000 |0.60870 |0.00000 |0.60870 |0.00041 |1.00000 |0.58076 |0.48728

|0.44205 |0.54190 |30.29470 |30.50000 |19.42202 |614.88027 |2577.0

|26.57222 |25.66667 |19.34716 |553.14428 |2526.0 |0.01157 |0.32288

|0.04792 |0.00913 |0.04678

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |...

|18169 |264444 |31 |34 |New Jersey |NJ |Paterson |Paterson City |City

|7503 |973 |40.903835 |-74.168169 |428448.0 |0 |2255 |1045 |1210

|1135.38669 |1143.0 |359.95409 |162.97036 |495.0 |1.00000 |0.92128

|0.82340 |0.80638 |0.66383 |0.62979 |0.61702 |0.41277 |495 |470

|36925.00941 |29793.0 |29426.63484 |481.28659 |582.0 |34416.76722

|29243.0 |25984.10193 |410.39039 |493.0 |2045.36963 |2020.0 |628.63272

|19.46281 |73.0 |899.466870 |984.00000 |223.395850 |14.000000 |6.929290

|0.00000 |0.06897 |0.00000 |0.83908 |0.12605 |1.00000 |0.09940 |0.60116

|0.55442 |0.64526 |30.18005 |29.91667 |18.40715 |252.32969 |1045.0

|31.03831 |26.50000 |20.91646 |323.09904 |1210.0 |0.15801 |0.37701

|0.11765 |0.03877 |0.03075

|25969 |235440 |33 |12 |Florida |FL |Pensacola |Ferry Pass |City |32514

|850 |30.541237 |-87.174514 |10066937.0 |9148345 |7091 |3605 |3486

|897.73467 |876.0 |316.76068 |615.97895 |1047.0 |1.00000 |0.84432

|0.81948 |0.78606 |0.59503 |0.40497 |0.34862 |0.25979 |1086 |1047

|62708.13895 |44336.0 |60701.65368 |1779.49562 |2754.0 |79912.21836

|63325.0 |60740.59162 |811.12993 |1552.0 |1524.35045 |1249.0 |919.50572

|615.61314 |1107.0 |513.623140 |447.00000 |269.847410 |561.000000

|388.777780 |0.02878 |0.06894 |0.07674 |0.66367 |0.12630 |0.54991

|0.45534 |0.95044 |0.96614 |0.93574 |36.96246 |33.33333 |22.50707

|918.09385 |3605.0 |41.07527 |39.33333 |24.33266 |934.84769 |3486.0

|0.59088 |0.47429 |0.06143 |0.05250 |0.09464

|26946 |244139 |97 |17 |Illinois |IL |Lindenhurst |Lindenhurst |Village

|60046 |847 |42.436767 |-88.034492 |8949666.0 |572563 |4183 |2080 |2103

|2033.66667 |2031.0 |288.09600 |1.17836 |12.0 |1.00000 |0.50000 |0.50000

|0.50000 |0.50000 |0.50000 |0.00000 |0.00000 |31 |12 |120907.40110

|108434.0 |71406.87338 |407.87341 |1279.0 |117889.54060 |107283.0

|64962.32882 |368.20382 |1159.0 |2410.47524 |2264.0 |828.78074

|256.43525 |1146.0 |789.851500 |747.00000 |305.183940 |102.000000

|55.030300 |0.06891 |0.06891 |0.25000 |0.91827 |0.12634 |0.05223

|0.03059 |0.98308 |0.97122 |0.99513 |35.97593 |38.75000 |21.77225

|523.03991 |2080.0 |34.32062 |36.75000 |20.00299 |523.51854 |2103.0

|0.97466 |0.68847 |0.00000 |0.00000 |0.03562

|2269 |234575 |9 |12 |Florida |FL |Melbourne |Palm Shores |City |32935

|321 |28.196504 |-80.666296 |11458985.0 |5893971 |6776 |3266 |3510

|1157.53838 |1080.0 |560.44721 |168.51905 |383.0 |0.96345 |0.87728

|0.65796 |0.41514 |0.32376 |0.27937 |0.25849 |0.25849 |405 |383

|84699.00808 |67790.0 |63663.66833 |1321.58196 |2771.0 |94471.25910

|83692.0 |58893.16048 |798.97839 |1966.0 |1757.84090 |1500.0 |854.13059

|517.73702 |1457.0 |493.056110 |470.00000 |305.422630 |909.000000

|557.598160 |0.06889 |0.06889 |0.18681 |0.61581 |0.12640 |0.14305

|0.56561 |0.93470 |0.94793 |0.92299 |46.84182 |50.41667 |23.70760

|793.18994 |3266.0 |47.10980 |49.75000 |21.80041 |875.53605 |3510.0

|0.85854 |0.63879 |0.00463 |0.00107 |0.10214

|11737 |293317 |101 |55 |Wisconsin |WI |Racine |Elmwood Park |City

|53403 |262 |42.703447 |-87.805437 |2351306.0 |0 |5818 |2616 |3202

|799.89978 |803.0 |242.99490 |447.99099 |676.0 |0.95414 |0.88757

|0.60355 |0.51036 |0.32988 |0.32988 |0.27811 |0.19970 |680 |676

|57723.67239 |55360.0 |39418.66328 |1316.08029 |2248.0 |69199.74963

|63663.0 |38218.44076 |733.82674 |1426.0 |1139.31165 |1072.0 |412.50334

|729.28491 |1179.0 |508.047130 |473.00000 |154.668640 |389.000000

|227.383840 |0.05038 |0.06888 |0.11735 |0.75191 |0.12644 |0.36525

|0.24972 |0.81051 |0.76861 |0.84363 |35.68616 |35.41667 |20.19046

|642.75698 |2616.0 |37.44296 |35.83333 |23.30099 |802.73591 |3202.0

|0.69429 |0.49164 |0.01290 |0.00000 |0.15050

|===

2500 rows × 77 columns

----

+\*In[61]:\*+

[source, ipython3]

----

top\_2500 = train\_df[['state', 'lat', 'lng', 'second\_mortgage', 'pct\_own', 'place', 'state', 'city', 'COUNTYID', 'STATEID', 'home\_equity', 'home\_equity\_second\_mortgage', 'debt', 'hi\_median', 'family\_median']].nlargest(2563, ['second\_mortgage', 'pct\_own'])

top\_2500

----

+\*Out[61]:\*+

----

[cols=",,,,,,,,,,,,,,,",options="header",]

|===

| |state |lat |lng |second\_mortgage |pct\_own |place |state |city

|COUNTYID |STATEID |home\_equity |home\_equity\_second\_mortgage |debt

|hi\_median |family\_median

|17289 |Louisiana |31.099609 |-93.202088 |1.00000 |0.01213 |Fort Polk

North |Louisiana |Fort Polk |115 |22 |1.00000 |1.00000 |1.00000 |46500.0

|46593.0

|1211 |Kentucky |37.879246 |-85.961748 |1.00000 |0.01067 |Fort Knox

|Kentucky |Fort Knox |93 |21 |1.00000 |1.00000 |1.00000 |53248.0

|55270.0

|7413 |Virginia |38.828707 |-77.121251 |1.00000 |0.00666 |Bailey's

Crossroads |Virginia |Alexandria |510 |51 |1.00000 |1.00000 |1.00000

|60591.0 |46820.0

|16834 |Kentucky |37.901474 |-85.986474 |1.00000 |0.00225 |Fort Knox

|Kentucky |Fort Knox |163 |21 |1.00000 |1.00000 |1.00000 |44822.0

|46983.0

|14014 |New Jersey |40.867944 |-74.114633 |0.60870 |0.01157 |Garfield

City |New Jersey |Passaic |31 |34 |0.00000 |0.00000 |0.60870 |28053.0

|29340.0

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |...

|21116 |Connecticut |41.582888 |-73.063650 |0.06832 |0.39632 |Oakville

|Connecticut |Waterbury |9 |9 |0.08696 |0.06832 |0.71118 |28730.0

|28843.0

|3071 |New Jersey |40.876773 |-74.124589 |0.06832 |0.27669 |Garfield

City |New Jersey |Clifton |31 |34 |0.08696 |0.04658 |0.53727 |42736.0

|46066.0

|15191 |Florida |28.506292 |-81.470967 |0.06830 |0.16311 |Orlovista

|Florida |Orlando |95 |12 |0.08751 |0.06830 |0.49733 |47956.0 |49757.0

|19842 |Illinois |39.467576 |-88.374644 |0.06829 |0.60836 |Mattoon City

|Illinois |Mattoon |29 |17 |0.02894 |0.02894 |0.68403 |40142.0 |50786.0

|11899 |California |34.403252 |-116.935898 |0.06827 |0.77511 |Lucerne

Valley |California |Lucerne Valley |71 |6 |0.18266 |0.06827 |0.59133

|24633.0 |42856.0

|===

2563 rows × 15 columns

----

+\*In[62]:\*+

[source, ipython3]

----

top\_2500.pct\_own.unique

----

+\*Out[62]:\*+

----<bound method Series.unique of 17289 0.01213

1211 0.01067

7413 0.00666

16834 0.00225

14014 0.01157

...

21116 0.39632

3071 0.27669

15191 0.16311

19842 0.60836

11899 0.77511

Name: pct\_own, Length: 2563, dtype: float64>----

+\*In[63]:\*+

[source, ipython3]

----

train\_df[train\_df.pct\_own > 0.1]

----

+\*Out[63]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |COUNTYID |STATEID |state |state\_ab |city |place |type |zip\_code

|area\_code |lat |lng |ALand |AWater |pop |male\_pop |female\_pop

|rent\_mean |rent\_median |rent\_stdev |rent\_sample\_weight |rent\_samples

|rent\_gt\_10 |rent\_gt\_15 |rent\_gt\_20 |rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35

|rent\_gt\_40 |rent\_gt\_50 |universe\_samples |used\_samples |hi\_mean

|hi\_median |hi\_stdev |hi\_sample\_weight |hi\_samples |family\_mean

|family\_median |family\_stdev |family\_sample\_weight |family\_samples

|hc\_mortgage\_mean |hc\_mortgage\_median |hc\_mortgage\_stdev

|hc\_mortgage\_sample\_weight |hc\_mortgage\_samples |hc\_mean |hc\_median

|hc\_stdev |hc\_samples |hc\_sample\_weight |home\_equity\_second\_mortgage

|second\_mortgage |home\_equity |debt |second\_mortgage\_cdf

|home\_equity\_cdf |debt\_cdf |hs\_degree |hs\_degree\_male |hs\_degree\_female

|male\_age\_mean |male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|0 |267822 |53 |36 |New York |NY |Hamilton |Hamilton |City |13346 |315

|42.840812 |-75.501524 |2.021834e+08 |1699120 |5230 |2612 |2618

|769.38638 |784.0 |232.63967 |272.34441 |362.0 |0.86761 |0.79155

|0.59155 |0.45634 |0.42817 |0.18592 |0.15493 |0.12958 |387 |355

|63125.28406 |48120.0 |49042.01206 |1290.96240 |2024.0 |67994.14790

|53245.0 |47667.30119 |884.33516 |1491.0 |1414.80295 |1223.0 |641.22898

|377.83135 |867.0 |570.01530 |558.0 |270.11299 |770.0 |499.29293

|0.01588 |0.02077 |0.08919 |0.52963 |0.43658 |0.49087 |0.73341 |0.89288

|0.85880 |0.92434 |42.48574 |44.00000 |22.97306 |696.42136 |2612.0

|44.48629 |45.33333 |22.51276 |685.33845 |2618.0 |0.79046 |0.57851

|0.01882 |0.01240 |0.08770

|1 |246444 |141 |18 |Indiana |IN |South Bend |Roseland |City |46616 |574

|41.701441 |-86.266614 |1.560828e+06 |100363 |2633 |1349 |1284

|804.87924 |848.0 |253.46747 |312.58622 |513.0 |0.97410 |0.93227

|0.69920 |0.69920 |0.55179 |0.41235 |0.39044 |0.27888 |542 |502

|41931.92593 |35186.0 |31639.50203 |838.74664 |1127.0 |50670.10337

|43023.0 |34715.57548 |375.28798 |554.0 |864.41390 |784.0 |482.27020

|316.88320 |356.0 |351.98293 |336.0 |125.40457 |229.0 |189.60606

|0.02222 |0.02222 |0.04274 |0.60855 |0.42174 |0.70823 |0.58120 |0.90487

|0.86947 |0.94187 |34.84728 |32.00000 |20.37452 |323.90204 |1349.0

|36.48391 |37.58333 |23.43353 |267.23367 |1284.0 |0.52483 |0.34886

|0.01426 |0.01426 |0.09030

|2 |245683 |63 |18 |Indiana |IN |Danville |Danville |City |46122 |317

|39.792202 |-86.515246 |6.956160e+07 |284193 |6881 |3643 |3238

|742.77365 |703.0 |323.39011 |291.85520 |378.0 |0.95238 |0.88624

|0.79630 |0.66667 |0.39153 |0.39153 |0.28307 |0.15873 |459 |378

|84942.68317 |74964.0 |56811.62186 |1155.20980 |2488.0 |95262.51431

|85395.0 |49292.67664 |709.74925 |1889.0 |1506.06758 |1361.0 |731.89394

|699.41354 |1491.0 |556.45986 |532.0 |184.42175 |538.0 |323.35354

|0.00000 |0.00000 |0.09512 |0.73484 |1.00000 |0.46332 |0.28704 |0.94288

|0.94616 |0.93952 |39.38154 |40.83333 |22.89769 |888.29730 |3643.0

|42.15810 |42.83333 |23.94119 |707.01963 |3238.0 |0.85331 |0.64745

|0.02830 |0.01607 |0.10657

|3 |279653 |127 |72 |Puerto Rico |PR |San Juan |Guaynabo |Urban |927

|787 |18.396103 |-66.104169 |1.105793e+06 |0 |2700 |1141 |1559

|803.42018 |782.0 |297.39258 |259.30316 |368.0 |0.94693 |0.87151

|0.69832 |0.61732 |0.51397 |0.46927 |0.35754 |0.32961 |438 |358

|48733.67116 |37845.0 |45100.54010 |928.32193 |1267.0 |56401.68133

|44399.0 |41082.90515 |490.18479 |729.0 |1175.28642 |1101.0 |428.98751

|261.28471 |437.0 |288.04047 |247.0 |185.55887 |392.0 |314.90566

|0.01086 |0.01086 |0.01086 |0.52714 |0.53057 |0.82530 |0.73727 |0.91500

|0.90755 |0.92043 |48.64749 |48.91667 |23.05968 |274.98956 |1141.0

|47.77526 |50.58333 |24.32015 |362.20193 |1559.0 |0.65037 |0.47257

|0.02021 |0.02021 |0.10106

|4 |247218 |161 |20 |Kansas |KS |Manhattan |Manhattan City |City |66502

|785 |39.195573 |-96.569366 |2.554403e+06 |0 |5637 |2586 |3051

|938.56493 |881.0 |392.44096 |1005.42886 |1704.0 |0.99286 |0.98247

|0.91688 |0.84740 |0.78247 |0.60974 |0.55455 |0.44416 |1725 |1540

|31834.15466 |22497.0 |34046.50907 |1548.67477 |1983.0 |54053.42396

|50272.0 |39609.12605 |244.08903 |395.0 |1192.58759 |1125.0 |327.49674

|76.61052 |134.0 |443.68855 |444.0 |76.12674 |124.0 |79.55556 |0.05426

|0.05426 |0.05426 |0.51938 |0.18332 |0.65545 |0.74967 |1.00000 |1.00000

|1.00000 |26.07533 |22.41667 |11.84399 |1296.89877 |2586.0 |24.17693

|21.58333 |11.10484 |1854.48652 |3051.0 |0.13046 |0.12356 |0.00000

|0.00000 |0.03109

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |...

|27316 |279212 |43 |72 |Puerto Rico |PR |Coamo |Coamo |Urban |769 |787

|18.076060 |-66.358379 |6.970300e+05 |0 |1847 |909 |938 |439.42839

|419.0 |140.29970 |170.00000 |170.0 |1.00000 |1.00000 |1.00000 |0.83333

|0.79012 |0.79012 |0.72222 |0.62963 |278 |162 |18515.67021 |13317.0

|23914.42656 |648.39020 |774.0 |20889.14617 |16760.0 |23488.17854

|346.58143 |446.0 |770.11560 |828.0 |157.85227 |58.00000 |58.0

|160.86544 |145.0 |94.04517 |438.0 |366.09000 |0.00000 |0.00000 |0.00000

|0.11694 |1.00000 |1.00000 |0.98762 |0.60155 |0.56962 |0.63222 |42.14011

|41.66667 |23.76478 |216.22207 |909.0 |42.73154 |40.16667 |24.79821

|230.87898 |938.0 |0.60422 |0.24603 |0.03042 |0.02249 |0.14683

|27317 |277856 |91 |42 |Pennsylvania |PA |Blue Bell |Blue Bell |Borough

|19422 |215 |40.158138 |-75.307271 |5.077337e+06 |11786 |4155 |2116

|2039 |1813.19253 |1788.0 |492.92300 |64.84927 |471.0 |0.85435 |0.63261

|0.50000 |0.37391 |0.30870 |0.30870 |0.26304 |0.23478 |484 |460

|119889.08320 |108284.0 |77625.25547 |518.53683 |1431.0 |118896.06830

|113313.0 |66663.51722 |388.97237 |1151.0 |2210.84055 |2202.0 |713.03361

|154.95504 |619.0 |712.16631 |663.0 |281.77621 |328.0 |174.96970

|0.00845 |0.02112 |0.19641 |0.65364 |0.43301 |0.12376 |0.47934 |0.95737

|0.95772 |0.95701 |37.75495 |38.83333 |21.45832 |530.17185 |2116.0

|38.21269 |39.50000 |21.84826 |496.20427 |2039.0 |0.68072 |0.61127

|0.05003 |0.02473 |0.04888

|27318 |233000 |87 |8 |Colorado |CO |Weldona |Saddle Ridge |City |80653

|970 |40.410316 |-103.814003 |1.323262e+09 |17577610 |2829 |1465 |1364

|849.39107 |834.0 |336.47530 |120.91448 |195.0 |0.93846 |0.71282

|0.54359 |0.44615 |0.29744 |0.23077 |0.16923 |0.09231 |237 |195

|79890.25113 |73350.0 |58132.65778 |529.41812 |1077.0 |88878.57034

|81864.0 |53510.48475 |375.21237 |871.0 |1671.07908 |1588.0 |742.67822

|182.53725 |488.0 |536.04921 |467.0 |306.82251 |352.0 |240.36899

|0.02024 |0.02024 |0.07857 |0.58095 |0.44186 |0.54095 |0.63916 |0.93555

|0.92200 |0.94887 |40.01134 |42.00000 |23.08048 |345.30911 |1465.0

|43.40218 |46.33333 |23.40858 |316.52078 |1364.0 |0.78508 |0.70451

|0.01386 |0.00520 |0.07712

|27319 |287425 |439 |48 |Texas |TX |Colleyville |Colleyville City |Town

|76034 |817 |32.904866 |-97.162151 |1.865230e+07 |158882 |11542 |5727

|5815 |1972.45746 |1843.0 |633.02173 |19.16328 |157.0 |1.00000 |1.00000

|0.75796 |0.61146 |0.50318 |0.50318 |0.27389 |0.27389 |234 |157

|165510.27110 |148548.0 |102038.58810 |960.70051 |4009.0 |167148.83770

|175952.0 |77638.35136 |719.65942 |3452.0 |3074.83088 |3188.0

|1121.07013 |536.61873 |2481.0 |1076.86881 |1037.0 |432.32205 |1294.0

|525.92451 |0.05801 |0.07550 |0.12556 |0.65722 |0.10759 |0.33196

|0.47094 |0.98540 |0.98883 |0.98201 |40.75409 |46.66667 |22.48690

|1305.28070 |5727.0 |39.25921 |43.41667 |21.36235 |1373.94120 |5815.0

|0.93970 |0.75503 |0.02287 |0.00915 |0.05261

|27320 |265371 |3 |32 |Nevada |NV |Las Vegas |Paradise |City |89123 |702

|36.064754 |-115.152237 |7.796308e+06 |0 |3726 |1815 |1911 |949.84199

|924.0 |198.82109 |555.87526 |1031.0 |0.94956 |0.87779 |0.83705 |0.63337

|0.51115 |0.41901 |0.27934 |0.10572 |1055 |1031 |51648.18703 |38072.0

|46305.44046 |1061.78932 |1409.0 |54886.07827 |42544.0 |39352.40334

|474.38547 |698.0 |1455.42340 |1364.0 |629.41356 |106.84849 |232.0

|540.26838 |454.0 |256.39951 |122.0 |79.07071 |0.01412 |0.01412 |0.18362

|0.65537 |0.50190 |0.15035 |0.47521 |0.87370 |0.87166 |0.87576 |34.81046

|32.50000 |20.16376 |461.52440 |1815.0 |34.45345 |29.83333 |19.77208

|526.73261 |1911.0 |0.27912 |0.34426 |0.03825 |0.03005 |0.13320

|===

26344 rows × 77 columns

----

+\*In[64]:\*+

[source, ipython3]

----

top\_2500[top\_2500.pct\_own > 0.1].head()

----

+\*Out[64]:\*+

----

[cols=",,,,,,,,,,,,,,,",options="header",]

|===

| |state |lat |lng |second\_mortgage |pct\_own |place |state |city

|COUNTYID |STATEID |home\_equity |home\_equity\_second\_mortgage |debt

|hi\_median |family\_median

|3285 |Virginia |37.297357 |-78.396452 |0.50000 |0.62069 |Farmville

|Virginia |Farmville |147 |51 |0.00000 |0.00000 |0.50000 |23236.0

|59954.0

|11980 |Massachusetts |42.254262 |-71.800347 |0.43363 |0.20247

|Worcester City |Massachusetts |Worcester |27 |25 |0.43363 |0.43363

|0.84956 |29037.0 |40476.0

|26018 |New York |40.751809 |-73.853582 |0.31818 |0.15618 |Harbor Hills

|New York |Corona |81 |36 |0.40341 |0.31818 |0.78409 |46106.0 |40462.0

|7829 |Maryland |39.127273 |-76.635265 |0.30212 |0.22380 |Glen Burnie

|Maryland |Glen Burnie |3 |24 |0.35689 |0.27739 |0.87633 |50164.0

|50705.0

|2077 |Florida |28.029063 |-82.495395 |0.28972 |0.11618 |Egypt Lake-leto

|Florida |Tampa |57 |12 |0.38785 |0.28972 |0.78972 |38340.0 |39980.0

|===

----

#b) Use the following bad debt equation: Bad Debt = P (Second Mortgage ∩ Home Equity Loan) Bad Debt = second\_mortgage + home\_equity - home\_equity\_second\_mortgage

c) Create pie charts to show overall debt and bad debt

+\*In[65]:\*+

[source, ipython3]

----

train\_df['bad\_debt']=train\_df['second\_mortgage']+train\_df['home\_equity']-train\_df['home\_equity\_second\_mortgage']

----

+\*In[66]:\*+

[source, ipython3]

----

train\_df['bins'] = pd.cut(train\_df['bad\_debt'],bins=[0,0.10,1], labels=["less than 50%","50-100%"],)

train\_df.groupby(['bins']).size().plot(kind='pie',subplots=True,figsize=(8, 3), colors=['green','red'], startangle=90, autopct='%1.1f%%')

plt.axis('equal')

plt.title('Bad Debt in Overall Debt')

plt.show()

----

+\*Out[66]:\*+

----

![png](output\_76\_0.png)

----

#d) Create Box and whisker plot and analyze the distribution for 2nd mortgage, home equity, good debt, and bad debt for different cities

+\*In[67]:\*+

[source, ipython3]

----

top\_2500['Bad\_Debt'] = top\_2500['second\_mortgage'] + top\_2500['home\_equity'] - top\_2500['home\_equity\_second\_mortgage']

top\_2500['Good\_Debt'] = top\_2500['debt'] - top\_2500['Bad\_Debt']

top\_2500.head(15)

----

+\*Out[67]:\*+

----

[cols=",,,,,,,,,,,,,,,,,",options="header",]

|===

| |state |lat |lng |second\_mortgage |pct\_own |place |state |city

|COUNTYID |STATEID |home\_equity |home\_equity\_second\_mortgage |debt

|hi\_median |family\_median |Bad\_Debt |Good\_Debt

|17289 |Louisiana |31.099609 |-93.202088 |1.00000 |0.01213 |Fort Polk

North |Louisiana |Fort Polk |115 |22 |1.00000 |1.00000 |1.00000 |46500.0

|46593.0 |1.00000 |0.00000

|1211 |Kentucky |37.879246 |-85.961748 |1.00000 |0.01067 |Fort Knox

|Kentucky |Fort Knox |93 |21 |1.00000 |1.00000 |1.00000 |53248.0

|55270.0 |1.00000 |0.00000

|7413 |Virginia |38.828707 |-77.121251 |1.00000 |0.00666 |Bailey's

Crossroads |Virginia |Alexandria |510 |51 |1.00000 |1.00000 |1.00000

|60591.0 |46820.0 |1.00000 |0.00000

|16834 |Kentucky |37.901474 |-85.986474 |1.00000 |0.00225 |Fort Knox

|Kentucky |Fort Knox |163 |21 |1.00000 |1.00000 |1.00000 |44822.0

|46983.0 |1.00000 |0.00000

|14014 |New Jersey |40.867944 |-74.114633 |0.60870 |0.01157 |Garfield

City |New Jersey |Passaic |31 |34 |0.00000 |0.00000 |0.60870 |28053.0

|29340.0 |0.60870 |0.00000

|20807 |California |34.067232 |-118.292902 |0.56000 |0.03542 |West

Hollywood City |California |Los Angeles |37 |6 |0.44000 |0.00000

|1.00000 |35334.0 |37921.0 |1.00000 |0.00000

|3285 |Virginia |37.297357 |-78.396452 |0.50000 |0.62069 |Farmville

|Virginia |Farmville |147 |51 |0.00000 |0.00000 |0.50000 |23236.0

|59954.0 |0.50000 |0.00000

|10812 |Maryland |38.944502 |-76.969572 |0.50000 |0.03476 |Mount Rainier

City |Maryland |Mt Rainier |33 |24 |0.50000 |0.50000 |1.00000 |38457.0

|38673.0 |0.50000 |0.50000

|2350 |District of Columbia |38.813245 |-77.023847 |0.50000 |0.01374

|Washington City |District of Columbia |Oxon Hill |1 |11 |0.50000

|0.50000 |1.00000 |33418.0 |35937.0 |0.50000 |0.50000

|3671 |California |34.183584 |-118.444345 |0.47059 |0.03302 |Burbank

City |California |Van Nuys |37 |6 |0.25490 |0.25490 |1.00000 |29324.0

|27698.0 |0.47059 |0.52941

|21706 |Arizona |33.458658 |-111.955104 |0.43750 |0.05660 |Tempe City

|Arizona |Scottsdale |13 |4 |0.43750 |0.43750 |0.54688 |40883.0 |59657.0

|0.43750 |0.10938

|11980 |Massachusetts |42.254262 |-71.800347 |0.43363 |0.20247

|Worcester City |Massachusetts |Worcester |27 |25 |0.43363 |0.43363

|0.84956 |29037.0 |40476.0 |0.43363 |0.41593

|12896 |Pennsylvania |39.952954 |-75.202767 |0.39024 |0.05041

|Millbourne |Pennsylvania |Philadelphia |101 |42 |0.21951 |0.00000

|0.93902 |12881.0 |50622.0 |0.60975 |0.32927

|7453 |Texas |30.285534 |-97.747727 |0.36364 |0.01737 |Austin City

|Texas |Austin |453 |48 |0.36364 |0.36364 |0.75758 |8309.0 |9587.0

|0.36364 |0.39394

|15589 |Georgia |33.740759 |-84.401777 |0.34783 |0.04026 |Atlanta City

|Georgia |Atlanta |121 |13 |0.34783 |0.34783 |0.69565 |17818.0 |20712.0

|0.34783 |0.34782

|===

----

+\*In[68]:\*+

[source, ipython3]

----

train\_df['bins'] = pd.cut(top\_2500['Bad\_Debt'],bins=[0,0.10,1], labels=["less than 50%","50-100%"],)

train\_df.groupby(['bins']).size().plot(kind='pie',subplots=True,figsize=(8, 3), colors=['green','red'], startangle=90, autopct='%1.1f%%')

plt.axis('equal')

plt.title('Bad Debt in Overall Debt of top 2500')

plt.show()

----

+\*Out[68]:\*+

----

![png](output\_79\_0.png)

----

+\*In[69]:\*+

[source, ipython3]

----

l1 = list(top\_2500['Bad\_Debt'] )

l2 = list(top\_2500['Good\_Debt'] )

l3 = sum(zip(l1, l2+[0]), ())

size = 10

explode = [0.4] \* size

explode = tuple(explode)

explode\_bd = [0.5] \* size\*2

explode\_bd = tuple(explode\_bd)

labels\_D = ['GD', 'BD'] \* size

labels\_D = tuple(labels\_D)

----

#limiting the donut chart to 10 cities

+\*In[70]:\*+

[source, ipython3]

----

labels = list(top\_2500.place[:10])

debt = list(top\_2500.debt[:10])

sns.set\_style("whitegrid")

gd\_bd = l3[:20]

plt.figure(figsize = (15, 15))

plt.pie(debt, labels = labels, startangle = 90, frame = True, radius =25, autopct='%1.1f%%', pctdistance=0.85, labeldistance = 0.9)

plt.pie(gd\_bd, labels = labels\_D, startangle = 90, frame = True, radius = 20, autopct='%1.1f%%', pctdistance=0.80, labeldistance = 0.85)

centre\_circle = plt.Circle((0,0),15,color='black', fc='white',linewidth=0)

fig = plt.gcf()

fig.gca().add\_artist(centre\_circle)

plt.axis('equal')

plt.tight\_layout()

plt.show()

----

+\*Out[70]:\*+

----

![png](output\_82\_0.png)

----

+\*In[71]:\*+

[source, ipython3]

----

second\_mortgage = list(top\_2500.second\_mortgage)

home\_equity = list(top\_2500.home\_equity)

Good\_Debt = list(top\_2500.Good\_Debt)

Bad\_Debt = list(top\_2500.Bad\_Debt)

----

limiting the cities to 20 to fit the box plots in the notebook

+\*In[72]:\*+

[source, ipython3]

----

top\_2500['city'].value\_counts()[:20].index

----

+\*Out[72]:\*+

----Index(['Chicago', 'Los Angeles', 'Washington', 'Brooklyn', 'Milwaukee',

'Aurora', 'Jacksonville', 'Charlotte', 'Bronx', 'Denver', 'Las Vegas',

'Minneapolis', 'Cincinnati', 'Sacramento', 'Colorado Springs',

'Baltimore', 'Long Beach', 'San Diego', 'New Orleans', 'Columbus'],

dtype='object')----

+\*In[73]:\*+

[source, ipython3]

----

cities = ['Chicago', 'Los Angeles', 'Washington', 'Brooklyn', 'Aurora',

'Milwaukee', 'Jacksonville', 'Bronx', 'Denver', 'Charlotte',

'Las Vegas', 'Cincinnati', 'Minneapolis', 'Baltimore', 'Sacramento',

'Long Beach', 'Colorado Springs', 'Columbus', 'San Diego',

'New Orleans']

----

+\*In[74]:\*+

[source, ipython3]

----

boxplot\_df = top\_2500[top\_2500['city'].isin (cities)]

sns.set\_style("whitegrid")

plt.figure(figsize = (30, 10))

sns.boxplot(x='city',y='second\_mortgage',data=boxplot\_df).set\_title('Second Mortgage distribution by cities', fontsize = 30)

plt.show()

----

+\*Out[74]:\*+

----

![png](output\_87\_0.png)

----

+\*In[75]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

plt.figure(figsize = (25, 10))

sns.boxplot(x='city',y='home\_equity',data=boxplot\_df).set\_title('Home Equity distribution by cities', fontsize = 30)

plt.show()

----

+\*Out[75]:\*+

----

![png](output\_88\_0.png)

----

+\*In[76]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

plt.figure(figsize = (25, 10))

sns.boxplot(x='city',y='Good\_Debt',data=boxplot\_df).set\_title('Good Debt distribution by cities', fontsize = 30)

plt.show()

----

+\*Out[76]:\*+

----

![png](output\_89\_0.png)

----

e) Create a collated income distribution chart for family income, house hold income, and remaining income

+\*In[77]:\*+

[source, ipython3]

----

top\_2500['remaining\_income'] = top\_2500['family\_median'] - top\_2500['hi\_median']

----

+\*In[78]:\*+

[source, ipython3]

----

income\_chart = round(top\_2500[['city', 'hi\_median', 'family\_median', 'remaining\_income']], 2)

income\_chart

----

+\*Out[78]:\*+

----

[cols=",,,,",options="header",]

|===

| |city |hi\_median |family\_median |remaining\_income

|17289 |Fort Polk |46500.0 |46593.0 |93.0

|1211 |Fort Knox |53248.0 |55270.0 |2022.0

|7413 |Alexandria |60591.0 |46820.0 |-13771.0

|16834 |Fort Knox |44822.0 |46983.0 |2161.0

|14014 |Passaic |28053.0 |29340.0 |1287.0

|... |... |... |... |...

|21116 |Waterbury |28730.0 |28843.0 |113.0

|3071 |Clifton |42736.0 |46066.0 |3330.0

|15191 |Orlando |47956.0 |49757.0 |1801.0

|19842 |Mattoon |40142.0 |50786.0 |10644.0

|11899 |Lucerne Valley |24633.0 |42856.0 |18223.0

|===

2563 rows × 4 columns

----

+\*In[79]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

plt.figure(figsize = (10,10))

sns.boxplot(data=top\_2500[['family\_median', 'hi\_median', 'remaining\_income']], palette=color\_pal).set\_title('Collated Income distribution', fontsize = 20)

plt.show()

----

+\*Out[79]:\*+

----

![png](output\_93\_0.png)

----

Exploratory Data Analysis (EDA):

1. Perform EDA and come out with insights into population density and age. You may have to derive new fields (make sure to weight averages for accurate measurements):

a) Use pop and ALand variables to create a new field called population density

b) Use male\_age\_median, female\_age\_median, male\_pop, and female\_pop to create a new field called median age c) Visualize the findings using appropriate chart type

+\*In[80]:\*+

[source, ipython3]

----

density\_eda\_df = train\_df[['state', 'city', 'place', 'ALand', 'pop', 'male\_age\_median', 'female\_age\_median', 'male\_pop', 'female\_pop']]

density\_eda\_df.head()

----

+\*Out[80]:\*+

----

[cols=",,,,,,,,,",options="header",]

|===

| |state |city |place |ALand |pop |male\_age\_median |female\_age\_median

|male\_pop |female\_pop

|0 |New York |Hamilton |Hamilton |202183361.0 |5230 |44.00000 |45.33333

|2612 |2618

|1 |Indiana |South Bend |Roseland |1560828.0 |2633 |32.00000 |37.58333

|1349 |1284

|2 |Indiana |Danville |Danville |69561595.0 |6881 |40.83333 |42.83333

|3643 |3238

|3 |Puerto Rico |San Juan |Guaynabo |1105793.0 |2700 |48.91667 |50.58333

|1141 |1559

|4 |Kansas |Manhattan |Manhattan City |2554403.0 |5637 |22.41667

|21.58333 |2586 |3051

|===

----

+\*In[81]:\*+

[source, ipython3]

----

density\_eda\_df['pop\_density'] = density\_eda\_df['pop'] / density\_eda\_df['ALand']

density\_eda\_df['median\_age'] = (density\_eda\_df['male\_age\_median'] \* density\_eda\_df['male\_pop'] + density\_eda\_df['female\_age\_median'] \* density\_eda\_df['female\_pop']) / density\_eda\_df['pop']

density\_eda\_df.nlargest(300, 'pop\_density')

----

+\*Out[81]:\*+

----

<ipython-input-81-3a90d48a6089>:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy

density\_eda\_df['pop\_density'] = density\_eda\_df['pop'] / density\_eda\_df['ALand']

<ipython-input-81-3a90d48a6089>:2: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy

density\_eda\_df['median\_age'] = (density\_eda\_df['male\_age\_median'] \* density\_eda\_df['male\_pop'] + density\_eda\_df['female\_age\_median'] \* density\_eda\_df['female\_pop']) / density\_eda\_df['pop']

[cols=",,,,,,,,,,,",options="header",]

|===

| |state |city |place |ALand |pop |male\_age\_median |female\_age\_median

|male\_pop |female\_pop |pop\_density |median\_age

|14417 |New York |New York |Mount Vernon City |212070.0 |16231 |33.66667

|36.08333 |8315 |7916 |0.076536 |34.845296

|21050 |New York |New York |New York City |182091.0 |13162 |38.83333

|34.66667 |5597 |7565 |0.072283 |36.438498

|10251 |New York |New York |Mount Vernon City |169349.0 |12189 |33.25000

|35.33333 |6110 |6079 |0.071976 |34.289016

|1546 |New York |New York |New York City |183653.0 |12427 |37.00000

|41.83333 |5425 |7002 |0.067666 |39.723342

|21795 |California |San Francisco |Daly City City |61277.0 |4101

|49.08333 |55.00000 |2465 |1636 |0.066926 |51.443650

|... |... |... |... |... |... |... |... |... |... |... |...

|20136 |California |San Rafael |San Rafael City |323250.0 |7330

|28.33333 |27.91667 |3860 |3470 |0.022676 |28.136084

|19093 |New York |New York |New York City |59282.0 |1343 |40.25000

|50.91667 |547 |796 |0.022654 |46.572166

|24412 |New York |Brooklyn |New York City |193832.0 |4386 |25.83333

|29.66667 |1868 |2518 |0.022628 |28.034048

|3347 |Virginia |Alexandria |Bailey's Crossroads |188743.0 |4257

|38.91667 |36.00000 |2230 |2027 |0.022554 |37.527877

|19464 |California |San Francisco |Sausalito City |226078.0 |5065

|42.16667 |41.00000 |2233 |2832 |0.022404 |41.514348

|===

300 rows × 11 columns

----

+\*In[82]:\*+

[source, ipython3]

----

pop\_density\_300 = density\_eda\_df.nlargest(300, 'pop\_density')

----

+\*In[83]:\*+

[source, ipython3]

----

cities = pop\_density\_300['city'].unique()

cities

----

+\*Out[83]:\*+

----array(['New York', 'San Francisco', 'Bronx', 'Brooklyn', 'Elmhurst',

'Flushing', 'Jackson Heights', 'Long Island City', 'Forest Hills',

'Trenton', 'Chicago', 'Corona', 'Jamaica', 'Woodside', 'Glendale',

'Los Angeles', 'Astoria', 'Honolulu', 'Richmond Hill', 'Baltimore',

'East Elmhurst', 'Rego Park', 'Guttenberg', 'Union City',

'West New York', 'Boston', 'Washington', 'Ridgewood', 'Arlington',

'Ozone Park', 'Corte Madera', 'Philadelphia', 'Hoboken',

'Staten Island', 'San Rafael', 'Alexandria'], dtype=object)----

+\*In[84]:\*+

[source, ipython3]

----

len(cities)

----

+\*Out[84]:\*+

----36----

+\*In[85]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

plt.figure(figsize = (30, 10))

sns.boxplot(x = 'place', y = 'pop\_density', data=density\_eda\_df.nlargest(300, 'pop\_density'), palette=color\_pal)

plt.show()

----

+\*Out[85]:\*+

----

![png](output\_100\_0.png)

----

#state wise distribution

+\*In[86]:\*+

[source, ipython3]

----

list(density\_eda\_df.nsmallest(450, 'pop\_density').state.unique())

----

+\*Out[86]:\*+

----['Florida',

'Alaska',

'Montana',

'Utah',

'Puerto Rico',

'Kentucky',

'Oregon',

'Nevada',

'Colorado',

'Idaho',

'California',

'New Mexico',

'Maine',

'South Dakota',

'Wyoming',

'Texas',

'Nebraska',

'Kansas',

'North Dakota',

'Arizona',

'Washington',

'South Carolina',

'New York',

'Tennessee',

'Oklahoma',

'Minnesota',

'Louisiana',

'Michigan',

'Wisconsin',

'North Carolina',

'Maryland',

'Georgia',

'Mississippi',

'New Hampshire',

'Missouri',

'Virginia']----

+\*In[87]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

plt.figure(figsize = (40, 10))

sns.boxplot(x = 'state', y = 'pop\_density', data=density\_eda\_df.nlargest(26585, 'pop\_density'), palette=color\_pal, order = ['New York', 'California', 'Illinois', 'Hawaii', 'New Jersey', 'Massachusetts', 'District of Columbia', 'Virginia',

'Pennsylvania', 'Florida', 'Puerto Rico', 'Maryland', 'Connecticut', 'Washington', 'Colorado', 'Wisconsin',

'Delaware', 'Oregon', 'Texas']).set\_title('Population Density Distribution of THICKLY populated States', fontsize = 30)

plt.show()

----

+\*Out[87]:\*+

----

![png](output\_103\_0.png)

----

+\*In[88]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

plt.figure(figsize = (40, 10))

sns.boxplot(x = 'state', y = 'pop\_density', data=density\_eda\_df.nsmallest(27019, 'pop\_density'), palette=color\_pal, order = ['Alaska', 'Montana', 'Utah', 'Oregon', 'Nevada', 'Colorado', 'Idaho', 'California', 'New Mexico',

'Maine', 'South Dakota', 'Wyoming', 'Nebraska', 'Texas', 'Kansas', 'North Dakota', 'Arizona',

'Washington', 'New York', 'Oklahoma', 'Minnesota', 'Louisiana', 'Michigan', 'Florida', 'Wisconsin', 'Mississippi',

'New Hampshire', 'Georgia', 'Missouri', 'Virginia', 'Alabama', 'Arkansas']).set\_title('Population Density Distribution of THINLY populated States', fontsize = 30)

plt.show()

----

+\*Out[88]:\*+

----

![png](output\_104\_0.png)

----

2. Create bins for population into a new variable by selecting appropriate class interval so that the number of categories don’t exceed 5 for the ease of analysis.

a) Analyze the married, separated, and divorced population for these population brackets

b) Visualize using appropriate chart type

+\*In[89]:\*+

[source, ipython3]

----

age\_df = train\_df[['state', 'city', 'place', 'pop', 'male\_pop', 'female\_pop', 'male\_age\_median', 'female\_age\_median', 'married', 'separated', 'divorced']]

train\_df.male\_age\_median.unique()

----

+\*Out[89]:\*+

----array([44. , 32. , 40.83333 , 48.91667 , 22.41667 ,

41.41667 , 40. , 53.08333 , 30.66667 , 47.33333 ,

34.33333 , 46.91667 , 49.75 , 34.66667 , 42.58333 ,

45.83333 , 44.16667 , 32.5 , 30.41667 , 27.41667 ,

30.08333 , 41.16667 , 38.75 , 30. , 31.16667 ,

46.75 , 36.66667 , 38.16667 , 34.91667 , 40.16667 ,

27.66667 , 39.33333 , 42.83333 , 36.41667 , 41.91667 ,

44.5 , 51.75 , 43.41667 , 51.66667 , 34. ,

64.08333 , 51.41667 , 20.25 , 29. , 28. ,

41.25 , 49.83333 , 34.5 , 24.91667 , 45.41667 ,

28.16667 , 34.08333 , 36.91667 , 46.66667 , 36.16667 ,

36.75 , 38.5 , 36.08333 , 47.5 , 51.16667 ,

48.16667 , 33. , 25.25 , 37.08333 , 42.66667 ,

40.25 , 29.75 , 38.41667 , 37.41667 , 42. ,

44.08333 , 36.5 , 32.16667 , 35.91667 , 39.5 ,

37.75 , 38.58333 , 21.25 , 35.33333 , 40.41667 ,

46.08333 , 54.41667 , 41.5 , 37.83333 , 19.66667 ,

31.41667 , 41.75 , 32.41667 , 26.66667 , 39.83333 ,

31.91667 , 34.58333 , 35.58333 , 52.58333 , 40.75 ,

37.33333 , 33.08333 , 40.58333 , 36.25 , 42.16667 ,

32.91667 , 45.58333 , 46.16667 , 25. , 29.66667 ,

42.33333 , 45.08333 , 31. , 52.66667 , 37.66667 ,

39.25 , 34.83333 , 32.58333 , 46.5 , 22.66667 ,

49.16667 , 35.5 , 32.66667 , 37.25 , 22. ,

51.33333 , 41.33333 , 43.25 , 35.83333 , 37. ,

41.08333 , 29.33333 , 34.75 , 34.41667 , 45.25 ,

35.75 , 74.16667 , 30.5 , 52.83333 , 23.08333 ,

30.75 , 36.33333 , 33.41667 , 46. , 37.5 ,

30.58333 , 36. , 35.25 , 34.25 , 36.83333 ,

24.83333 , 48. , 33.83333 , 44.83333 , 42.5 ,

43.66667 , 43.33333 , 33.75 , 46.25 , 57.83333 ,

31.83333 , 49.08333 , 43.5 , 30.33333 , 47.08333 ,

52.5 , 27. , 29.16667 , 44.41667 , 38. ,

36.58333 , 21.83333 , 42.75 , 29.5 , 29.58333 ,

34.16667 , 35.08333 , 33.16667 , 39.58333 , 23.5 ,

33.66667 , 43.16667 , 31.58333 , 40.91667 , 27.75 ,

48.41667 , 26.75 , 45.16667 , 44.33333 , 41.66667 ,

38.33333 , 33.5 , 25.83333 , 39.66667 , 55.58333 ,

25.08333 , 53.33333 , 45.75 , 48.66667 , 26.83333 ,

45.91667 , 28.66667 , 35.41667 , 26.08333 , 26.16667 ,

31.66667 , 43. , 31.33333 , 35. , 29.41667 ,

44.66667 , 42.41667 , 23.83333 , 38.91667 , 22.08333 ,

40.5 , 44.58333 , 28.75 , 26.58333 , 30.16667 ,

33.91667 , 32.33333 , 19.25 , 43.83333 , 39.08333 ,

32.08333 , 39.16667 , 24. , 32.25 , 28.08333 ,

29.08333 , 51.58333 , 48.75 , 20.91667 , 24.5 ,

42.91667 , 43.75 , 50.33333 , 67.16667 , 48.83333 ,

59.08333 , 22.5 , 29.91667 , 46.83333 , 50.66667 ,

39. , 44.91667 , 56. , 38.25 , 46.41667 ,

48.58333 , 52.25 , 41.58333 , 45.66667 , 25.91667 ,

33.33333 , 30.83333 , 19.75 , 39.91667 , 33.58333 ,

49.58333 , 47.16667 , 32.75 , 70.5 , 40.33333 ,

54.08333 , 35.66667 , 52. , 31.5 , 25.16667 ,

22.16667 , 45.5 , 48.5 , 23.58333 , 43.91667 ,

47.58333 , 24.08333 , 21.08333 , 43.58333 , 37.91667 ,

37.16667 , 33.25 , 47.25 , 20.66667 , 37.58333 ,

47.75 , 56.91667 , 50.5 , 24.33333 , 38.08333 ,

28.91667 , 43.08333 , 44.25 , 27.83333 , 40.08333 ,

42.08333 , 29.25 , 27.08333 , 56.25 , 80.16667 ,

38.83333 , 47.66667 , 41.83333 , 57. , 23.16667 ,

40.66667 , 47.91667 , 30.25 , 27.25 , 39.75 ,

27.91667 , 49.25 , 46.58333 , 45. , 50.41667 ,

54.16667 , 71.41667 , 31.25 , 50.75 , 18.91667 ,

50.58333 , 35.16667 , 38.66667 , 44.75 , 28.25 ,

21.5 , 25.41667 , 20.33333 , 49.66667 , 16.16667 ,

41. , 65. , 23.41667 , 28.58333 , 32.83333 ,

22.33333 , 51. , 25.33333 , 52.08333 , 26.33333 ,

24.25 , 39.41667 , 53.41667 , 51.91667 , 26.25 ,

28.41667 , 50.91667 , 21.91667 , 29.83333 , 62.33333 ,

48.25 , 23. , 62.08333 , 28.33333 , 31.75 ,

54.83333 , 53. , 60.66667 , 49. , 15.58333 ,

56.66667 , 24.75 , 24.58333 , 46.33333 , 30.91667 ,

23.91667 , 23.25 , 49.91667 , 51.5 , 27.5 ,

55.75 , 17.83333 , 67.75 , 27.33333 , 51.25 ,

42.25 , 22.83333 , 21.16667 , 47. , 24.66667 ,

23.66667 , 19.41667 , 63.83333 , 52.41667 , 19.5 ,

24.41667 , 26.5 , 21.75 , 45.33333 , 53.66667 ,

52.91667 , 67.08333 , 47.83333 , 58.83333 , 50.16667 ,

15.25 , 49.41667 , 28.5 , 53.83333 , 21.33333 ,

63.58333 , 50.83333 , 51.83333 , 64.58333 , 27.16667 ,

22.25 , 24.16667 , 71.33333 , 62.5 , 58.5 ,

18.41667 , 52.75 , 50.25 , 58.16667 , 22.91667 ,

31.08333 , 28.83333 , 51.08333 , 22.58333 , 20. ,

57.91667 , 27.58333 , 26.41667 , 59.25 , 16.5 ,

47.41667 , 60.75 , 19.16667 , 25.58333 , 20.58333 ,

19.91667 , 26. , 19.33333 , 56.75 , 54.25 ,

21.66667 , 65.16667 , 19.58333 , 63.33333 , 25.5 ,

53.5 , 49.33333 , 55.25 , 22.75 , 21.41667 ,

52.33333 , 17.66667 , 62.91667 , 49.5 , 50.08333 ,

71.91667 , 57.66667 , 60.08333 , 75.58333 , 59.66667 ,

21.58333 , 76.33333 , 53.75 , 61.41667 , 60.33333 ,

23.75 , 53.25 , 50. , 56.83333 , 69.83333 ,

68.83333 , 48.08333 , 54.58333 , 58. , 57.25 ,

25.66667 , 60.5 , 57.08333 , 54. , 55.16667 ,

20.83333 , 48.33333 , 73. , 26.91667 , 57.75 ,

54.33333 , 62.75 , 64.66667 , 64.75 , 60.25 ,

56.16667 , 55. , 61.33333 , 20.5 , 52.16667 ,

64.41667 , 59.91667 , 15.16667 , 55.41667 , 57.58333 ,

61.25 , 54.75 , 23.33333 , 59.58333 , 70.91667 ,

58.66667 , 70.08333 , 17.91667 , 55.83333 , 60.41667 ,

60.58333 , 73.41667 , 20.08333 , 56.41667 , 18.08333 ,

67.83333 , 53.58333 , 55.08333 , 70.33333 , 20.75 ,

55.66667 , 64.83333 , 15.08333 , 62.58333 , 71.58333 ,

25.75 , 71.83333 , 17. , 64.33333 , 56.33333 ,

54.66667 , 69.08333 , 65.58333 , 38.11479967, 66.33333 ,

66. , 61.75 , 17.58333 , 54.5 , 67.41667 ,

71.25 , 12.16667 , 63.25 , 66.75 , 21. ,

76.83333 , 59.16667 , 65.41667 , 57.5 , 18.33333 ,

66.25 , 53.16667 , 74.66667 , 66.08333 , 73.91667 ,

77.08333 , 20.41667 , 58.08333 , 59.33333 , 73.16667 ,

77.83333 , 65.33333 , 13.58333 , 65.5 , 15.91667 ,

56.58333 , 58.41667 , 63.5 , 72.75 , 59.83333 ,

63.91667 , 56.08333 , 74.41667 , 65.66667 , 55.33333 ,

55.91667 , 61.83333 , 59. , 57.33333 , 60.16667 ,

68.25 , 16.83333 , 19.83333 , 10. , 20.16667 ,

64. , 70.25 , 68.66667 , 59.41667 , 66.5 ,

62. , 63.08333 , 59.5 , 67.58333 , 62.83333 ,

19. , 53.91667 , 62.25 , 78. , 58.75 ,

62.16667 , 73.66667 , 18.75 , 15.33333 , 59.75 ,

66.41667 , 72.33333 , 75. , 67. , 62.66667 ,

54.91667 , 61.16667 , 72.66667 , 62.41667 , 18.16667 ,

77.75 , 60.91667 , 68.91667 , 19.08333 , 65.75 ,

77.25 , 75.16667 , 74.83333 , 18.66667 , 56.5 ,

9.83333 , 58.91667 , 69.25 , 17.41667 , 13.5 ,

67.66667 , 68.41667 , 57.16667 , 58.33333 , 72.58333 ,

78.25 , 67.5 , 14.75 , 61.08333 , 55.5 ,

71.08333 , 64.91667 , 58.58333 , 70. , 68.33333 ,

16.75 , 61. , 14.91667 , 72.91667 , 17.33333 ,

18.83333 , 69.91667 , 69.58333 , 69.16667 , 66.16667 ,

63. , 13.16667 , 67.33333 , 71. , 65.08333 ,

63.75 , 69.33333 , 68. , 75.66667 , 77.66667 ,

65.91667 , 74.5 , 60.83333 , 73.33333 , 60. ,

63.41667 , 73.25 , 64.25 , 75.5 , 76.91667 ,

71.75 , 70.75 , 73.58333 , 16. , 9.75 ,

74. , 63.66667 , 16.33333 , 73.83333 , 68.16667 ,

12.5 , 65.25 , 18.25 , 71.66667 , 61.58333 ,

16.41667 , 17.5 , 17.25 ])----

+\*In[90]:\*+

[source, ipython3]

----

bins = [0, 12,18, 35, 55, 100]

labels = ['kids', 'Young Adult', 'Youth ' , 'Adult', 'Senior']

----

+\*In[91]:\*+

[source, ipython3]

----

age\_df['male\_population\_bracket'] = pd.cut(age\_df['male\_age\_median'], bins, labels = labels)

age\_df['female\_population\_bracket'] = pd.cut(age\_df['female\_age\_median'], bins, labels = labels)

age\_df.head()

----

+\*Out[91]:\*+

----

<ipython-input-91-e772c5b3b9dc>:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy

age\_df['male\_population\_bracket'] = pd.cut(age\_df['male\_age\_median'], bins, labels = labels)

<ipython-input-91-e772c5b3b9dc>:2: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy

age\_df['female\_population\_bracket'] = pd.cut(age\_df['female\_age\_median'], bins, labels = labels)

[cols=",,,,,,,,,,,,,",options="header",]

|===

| |state |city |place |pop |male\_pop |female\_pop |male\_age\_median

|female\_age\_median |married |separated |divorced

|male\_population\_bracket |female\_population\_bracket

|0 |New York |Hamilton |Hamilton |5230 |2612 |2618 |44.00000 |45.33333

|0.57851 |0.01240 |0.08770 |Adult |Adult

|1 |Indiana |South Bend |Roseland |2633 |1349 |1284 |32.00000 |37.58333

|0.34886 |0.01426 |0.09030 |Youth |Adult

|2 |Indiana |Danville |Danville |6881 |3643 |3238 |40.83333 |42.83333

|0.64745 |0.01607 |0.10657 |Adult |Adult

|3 |Puerto Rico |San Juan |Guaynabo |2700 |1141 |1559 |48.91667

|50.58333 |0.47257 |0.02021 |0.10106 |Adult |Adult

|4 |Kansas |Manhattan |Manhattan City |5637 |2586 |3051 |22.41667

|21.58333 |0.12356 |0.00000 |0.03109 |Youth |Youth

|===

----

+\*In[92]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

plt.figure(figsize = (40, 10))

ax = sns.barplot(x = 'state', y = 'married', hue = 'male\_population\_bracket', data = age\_df, palette=color\_pal,

order = ['New York', 'Ohio', 'Georgia', 'New Jersey', 'Texas', 'Colorado', 'Virginia', 'California', 'Mississippi', 'Oregon', 'Arizona', 'Utah', 'Pennsylvania', 'Michigan', 'Arkansas', 'Maine', 'Florida', 'North Carolina',

'Minnesota', 'Iowa', 'Connecticut', 'Maryland', 'Wisconsin', 'Louisiana', 'Alabama', 'Massachusetts', 'Washington', 'Illinois', 'Tennessee', 'Indiana', 'District of Columbia', 'South Carolina', 'Oklahoma', 'Delaware'])

ax.set\_title('Married Male population by state', fontsize = 30)

plt.show()

----

+\*Out[92]:\*+

----

![png](output\_109\_0.png)

----

Ohio and Georgia has under-age marriage among male kid population. Nearly half the states have married young adults.

+\*In[93]:\*+

[source, ipython3]

----

age\_df.city.unique()

----

+\*Out[93]:\*+

----array(['Hamilton', 'South Bend', 'Danville', ..., 'Blue Bell', 'Weldona',

'Colleyville'], dtype=object)----

+\*In[94]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

plt.figure(figsize = (40, 10))

ax = sns.barplot(x = 'state', y = 'married', hue = 'female\_population\_bracket', data = age\_df, palette=color\_pal,

order = ['New York', 'Ohio', 'Georgia', 'New Jersey', 'Texas', 'Colorado', 'Virginia', 'California', 'Mississippi', 'Oregon', 'Arizona', 'Utah', 'Pennsylvania', 'Michigan', 'Arkansas', 'Maine', 'Florida', 'North Carolina',

'Minnesota', 'Iowa', 'Connecticut', 'Maryland', 'Wisconsin', 'Louisiana', 'Alabama', 'Massachusetts', 'Washington', 'Illinois', 'Tennessee', 'Indiana', 'District of Columbia', 'South Carolina', 'Oklahoma', 'Delaware'])

ax.set\_title('Married Female population by state', fontsize = 30)

plt.show()

----

+\*Out[94]:\*+

----

![png](output\_112\_0.png)

----

Newyork, Texas and California are having married female young adult population. No kid population is married in any of the states

+\*In[95]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

plt.figure(figsize = (40, 10))

ax = sns.barplot(x = 'state', y = 'separated', hue = 'male\_population\_bracket', data = age\_df, palette=color\_pal,

order = ['New York', 'Ohio', 'Georgia', 'New Jersey', 'Texas', 'Colorado', 'Virginia', 'California', 'Mississippi', 'Oregon', 'Arizona', 'Utah', 'Pennsylvania', 'Michigan', 'Arkansas', 'Maine', 'Florida', 'North Carolina',

'Minnesota', 'Iowa', 'Connecticut', 'Maryland', 'Wisconsin', 'Louisiana', 'Alabama', 'Massachusetts', 'Washington', 'Illinois', 'Tennessee', 'Indiana', 'District of Columbia', 'South Carolina', 'Oklahoma', 'Delaware'])

ax.set\_title('Separated Male population by state', fontsize = 30)

plt.show()

----

+\*Out[95]:\*+

----

![png](output\_114\_0.png)

----

peaks - Youth - 1. Virginia, 2. Connecticut, 3. Illinois

Young Adult - 1. Virginia, 2. Connecticut, 3. Tennessee

+\*In[96]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

plt.figure(figsize = (40, 10))

ax = sns.barplot(x = 'state', y = 'separated', hue = 'female\_population\_bracket', data = age\_df, palette=color\_pal,

order = ['New York', 'Ohio', 'Georgia', 'New Jersey', 'Texas', 'Colorado', 'Virginia', 'California', 'Mississippi', 'Oregon', 'Arizona', 'Utah', 'Pennsylvania', 'Michigan', 'Arkansas', 'Maine', 'Florida', 'North Carolina',

'Minnesota', 'Iowa', 'Connecticut', 'Maryland', 'Wisconsin', 'Louisiana', 'Alabama', 'Massachusetts', 'Washington', 'Illinois', 'Tennessee', 'Indiana', 'District of Columbia', 'South Carolina', 'Oklahoma', 'Delaware'])

ax.set\_title('Separated Female population by state', fontsize = 30)

plt.show()

----

+\*Out[96]:\*+

----

![png](output\_116\_0.png)

----

peaks - Youth - 1.Tennessee , 2. Texas, 3. Mississippi

Young Adult - 1. Texas, 2. California, 3. Newyork

+\*In[97]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

plt.figure(figsize = (40, 10))

ax = sns.barplot(x = 'state', y = 'divorced', hue = 'male\_population\_bracket', data = age\_df, palette=color\_pal,

order = ['New York', 'Ohio', 'Georgia', 'New Jersey', 'Texas', 'Colorado', 'Virginia', 'California', 'Mississippi', 'Oregon', 'Arizona', 'Utah', 'Pennsylvania', 'Michigan', 'Arkansas', 'Maine', 'Florida', 'North Carolina',

'Minnesota', 'Iowa', 'Connecticut', 'Maryland', 'Wisconsin', 'Louisiana', 'Alabama', 'Massachusetts', 'Washington', 'Illinois', 'Tennessee', 'Indiana', 'District of Columbia', 'South Carolina', 'Oklahoma', 'Delaware'])

ax.set\_title('Divorced Male population by state', fontsize = 30)

plt.show()

----

+\*Out[97]:\*+

----

![png](output\_118\_0.png)

----

peaks - Kids - 1. Ohio, 2. Georgia

Young Adult - 1. Louisiana 2. Minnesota, 3. District of Columbia

Youth - 1. Oklahoma, 2. Louisiana 3. Mississippi

Senior - 1. Oklahoma 2. Mississippi 3. Louisiana

+\*In[98]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

plt.figure(figsize = (40, 10))

ax = sns.barplot(x = 'state', y = 'divorced', hue = 'female\_population\_bracket', data = age\_df, palette=color\_pal,

order = ['New York', 'Ohio', 'Georgia', 'New Jersey', 'Texas', 'Colorado', 'Virginia', 'California', 'Mississippi', 'Oregon', 'Arizona', 'Utah', 'Pennsylvania', 'Michigan', 'Arkansas', 'Maine', 'Florida', 'North Carolina',

'Minnesota', 'Iowa', 'Connecticut', 'Maryland', 'Wisconsin', 'Louisiana', 'Alabama', 'Massachusetts', 'Washington', 'Illinois', 'Tennessee', 'Indiana', 'District of Columbia', 'South Carolina', 'Oklahoma', 'Delaware'])

ax.set\_title('Divorced Female population by state', fontsize = 30)

plt.show()

----

+\*Out[98]:\*+

----

![png](output\_120\_0.png)

----

peaks - Young Adult - 1. Texas 2. Newyork

Youth - 1. Oklahoma, 2. Louisiana 3. Mississippi, Texas, Arkansas

Senior - 1. Oklahoma 2. Louisiana 3. Mississippi#3. Please detail your observations for rent as a percentage of income at an overall level, and for different states.

+\*In[99]:\*+

[source, ipython3]

----

rent\_df = train\_df[['state', 'city', 'rent\_median', 'hi\_median', 'family\_median']]

Overall\_rent\_percentage = (rent\_df['rent\_median'].sum() / rent\_df['hi\_median'].sum()) \* 100

round(Overall\_rent\_percentage, 2)

----

+\*Out[99]:\*+

----1.75----

+\*In[100]:\*+

[source, ipython3]

----

rent\_df['ov\_rent\_pcnt'] = round((rent\_df['rent\_median'] / rent\_df['hi\_median']) \* 100, 2)

----

+\*Out[100]:\*+

----

<ipython-input-100-bc97e1f802ff>:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy

rent\_df['ov\_rent\_pcnt'] = round((rent\_df['rent\_median'] / rent\_df['hi\_median']) \* 100, 2)

----

+\*In[101]:\*+

[source, ipython3]

----

rent\_df.head()

----

+\*Out[101]:\*+

----

[cols=",,,,,,",options="header",]

|===

| |state |city |rent\_median |hi\_median |family\_median |ov\_rent\_pcnt

|0 |New York |Hamilton |784.0 |48120.0 |53245.0 |1.63

|1 |Indiana |South Bend |848.0 |35186.0 |43023.0 |2.41

|2 |Indiana |Danville |703.0 |74964.0 |85395.0 |0.94

|3 |Puerto Rico |San Juan |782.0 |37845.0 |44399.0 |2.07

|4 |Kansas |Manhattan |881.0 |22497.0 |50272.0 |3.92

|===

----

+\*In[102]:\*+

[source, ipython3]

----

print(list(rent\_df.nlargest(500, 'ov\_rent\_pcnt').state.unique()))

print(len(list(rent\_df.nlargest(500, 'ov\_rent\_pcnt').state.unique())))

----

+\*Out[102]:\*+

----

['Florida', 'Arizona', 'Virginia', 'Illinois', 'Georgia', 'Texas', 'North Carolina', 'Louisiana', 'Ohio', 'California', 'Indiana', 'Wisconsin', 'New York', 'Washington', 'Oregon', 'Pennsylvania', 'Michigan', 'Tennessee', 'Maryland', 'Mississippi', 'Alabama', 'Iowa', 'Puerto Rico', 'New Jersey', 'South Carolina', 'Wyoming', 'Hawaii', 'Utah', 'Missouri', 'Connecticut', 'Minnesota', 'Massachusetts', 'Colorado', 'Kansas', 'Oklahoma', 'District of Columbia', 'New Mexico', 'Kentucky', 'Maine', 'Arkansas', 'Vermont', 'Rhode Island']

42

----

+\*In[103]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

plt.figure(figsize = (40,10))

ax = sns.boxplot(x = 'state', y = 'ov\_rent\_pcnt', data=rent\_df.nlargest(27019, 'ov\_rent\_pcnt'), palette=color\_pal,

order = ['Georgia', 'Texas', 'California', 'New York', 'Florida', 'Washington', 'Oregon', 'Pennsylvania', 'Maryland', 'Virginia', 'Mississippi', 'Alabama', 'Michigan', 'Louisiana',

'Iowa', 'Puerto Rico', 'New Jersey', 'Illinois', 'Arizona', 'North Carolina', 'South Carolina', 'Tennessee', 'Ohio', 'Wisconsin', 'Missouri', 'Connecticut', 'Minnesota',

'Massachusetts', 'Indiana', 'Colorado', 'Kansas', 'Oklahoma', 'District of Columbia', 'New Mexico', 'Hawaii', 'Maine', 'Arkansas', 'Vermont', 'Rhode Island', 'Kentucky']

).set\_title('Rent as percentage of House Hold Income by State', fontsize = 30)

plt.show()

----

+\*Out[103]:\*+

----

![png](output\_127\_0.png)

----

#4. Perform correlation analysis for all the relevant variables by creating a heatmap. Describe your findings.

+\*In[104]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

corr = train\_df.corr()

mask = np.zeros\_like(corr, dtype=np.bool)

mask[np.triu\_indices\_from(mask)] = True

kot = corr[corr>=.6]

plt.figure(figsize=(45,20))

sns.heatmap(kot, cmap="Greens", annot = True, mask = mask, linewidths=1, linecolor='black').set\_title('Positive correlation Heat Map', fontsize = 20)

plt.grid('on', )

plt.show()

----

+\*Out[104]:\*+

----

![png](output\_129\_0.png)

----

+\*In[105]:\*+

[source, ipython3]

----

sns.set\_style("whitegrid")

kot = corr[corr <=-.3]

plt.figure(figsize=(45,20))

sns.heatmap(kot, cmap="Blues", annot = True, mask = mask, linewidths=1, linecolor='green').set\_title('Negative correlation Heat Map', fontsize = 20)

plt.grid('on', )

plt.show()

----

+\*Out[105]:\*+

----

![png](output\_130\_0.png)

----

+\*In[106]:\*+

[source, ipython3]

----

cor=train\_df[['COUNTYID','STATEID','zip\_code','type','pop', 'family\_mean',

'second\_mortgage', 'home\_equity', 'debt','hs\_degree',

'male\_age\_median','female\_age\_median','pct\_own', 'married','separated', 'divorced']].corr()

----

+\*In[107]:\*+

[source, ipython3]

----

plt.figure(figsize=(20,10))

sns.heatmap(cor,annot=True,cmap='coolwarm')

plt.show()

----

+\*Out[107]:\*+

----

![png](output\_132\_0.png)

----

High positive correaltion is noticed between pop, male\_pop and female\_pop

High positive correaltion is noticed between rent\_mean,hi\_mean, family\_mean,hc\_meanData Pre-processing:

1. The economic multivariate data has a significant number of measured variables. The goal is to find where the measured variables depend on a number of smaller unobserved common factors or latent variables. 2. Each variable is assumed to be dependent upon a linear combination of the common factors, and the coefficients are known as loadings. Each measured variable also includes a component due to independent random variability, known as “specific variance” because it is specific to one variable. Obtain the common factors and then plot the loadings. Use factor analysis to find latent variables in our dataset and gain insight into the linear relationships in the data. Following are the list of latent variables:

• Highschool graduation rates

• Median population age

• Second mortgage statistics

• Percent own

• Bad debt expense

+\*In[108]:\*+

[source, ipython3]

----

train\_df.info()

----

+\*Out[108]:\*+

----

<class 'pandas.core.frame.DataFrame'>

Int64Index: 27019 entries, 0 to 27320

Data columns (total 79 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 UID 27019 non-null int64

1 COUNTYID 27019 non-null int64

2 STATEID 27019 non-null int64

3 state 27019 non-null object

4 state\_ab 27019 non-null object

5 city 27019 non-null object

6 place 27019 non-null object

7 type 27019 non-null object

8 zip\_code 27019 non-null int64

9 area\_code 27019 non-null int64

10 lat 27019 non-null float64

11 lng 27019 non-null float64

12 ALand 27019 non-null float64

13 AWater 27019 non-null int64

14 pop 27019 non-null int64

15 male\_pop 27019 non-null int64

16 female\_pop 27019 non-null int64

17 rent\_mean 27019 non-null float64

18 rent\_median 27019 non-null float64

19 rent\_stdev 27019 non-null float64

20 rent\_sample\_weight 27019 non-null float64

21 rent\_samples 27019 non-null float64

22 rent\_gt\_10 27019 non-null float64

23 rent\_gt\_15 27019 non-null float64

24 rent\_gt\_20 27019 non-null float64

25 rent\_gt\_25 27019 non-null float64

26 rent\_gt\_30 27019 non-null float64

27 rent\_gt\_35 27019 non-null float64

28 rent\_gt\_40 27019 non-null float64

29 rent\_gt\_50 27019 non-null float64

30 universe\_samples 27019 non-null int64

31 used\_samples 27019 non-null int64

32 hi\_mean 27019 non-null float64

33 hi\_median 27019 non-null float64

34 hi\_stdev 27019 non-null float64

35 hi\_sample\_weight 27019 non-null float64

36 hi\_samples 27019 non-null float64

37 family\_mean 27019 non-null float64

38 family\_median 27019 non-null float64

39 family\_stdev 27019 non-null float64

40 family\_sample\_weight 27019 non-null float64

41 family\_samples 27019 non-null float64

42 hc\_mortgage\_mean 27019 non-null float64

43 hc\_mortgage\_median 27019 non-null float64

44 hc\_mortgage\_stdev 27019 non-null float64

45 hc\_mortgage\_sample\_weight 27019 non-null float64

46 hc\_mortgage\_samples 27019 non-null float64

47 hc\_mean 27019 non-null float64

48 hc\_median 27019 non-null float64

49 hc\_stdev 27019 non-null float64

50 hc\_samples 27019 non-null float64

51 hc\_sample\_weight 27019 non-null float64

52 home\_equity\_second\_mortgage 27019 non-null float64

53 second\_mortgage 27019 non-null float64

54 home\_equity 27019 non-null float64

55 debt 27019 non-null float64

56 second\_mortgage\_cdf 27019 non-null float64

57 home\_equity\_cdf 27019 non-null float64

58 debt\_cdf 27019 non-null float64

59 hs\_degree 27019 non-null float64

60 hs\_degree\_male 27019 non-null float64

61 hs\_degree\_female 27019 non-null float64

62 male\_age\_mean 27019 non-null float64

63 male\_age\_median 27019 non-null float64

64 male\_age\_stdev 27019 non-null float64

65 male\_age\_sample\_weight 27019 non-null float64

66 male\_age\_samples 27019 non-null float64

67 female\_age\_mean 27019 non-null float64

68 female\_age\_median 27019 non-null float64

69 female\_age\_stdev 27019 non-null float64

70 female\_age\_sample\_weight 27019 non-null float64

71 female\_age\_samples 27019 non-null float64

72 pct\_own 27019 non-null float64

73 married 27019 non-null float64

74 married\_snp 27019 non-null float64

75 separated 27019 non-null float64

76 divorced 27019 non-null float64

77 bad\_debt 27019 non-null float64

78 bins 2563 non-null category

dtypes: category(1), float64(62), int64(11), object(5)

memory usage: 17.6+ MB

----

+\*In[109]:\*+

[source, ipython3]

----

train\_df['Bad\_Debt'] = train\_df['second\_mortgage'] + train\_df['home\_equity'] - train\_df['home\_equity\_second\_mortgage']

----

+\*In[110]:\*+

[source, ipython3]

----

for col in train\_df.columns:

print(col,' = ' ,train\_df[col].dtype)

----

+\*Out[110]:\*+

----

UID = int64

COUNTYID = int64

STATEID = int64

state = object

state\_ab = object

city = object

place = object

type = object

zip\_code = int64

area\_code = int64

lat = float64

lng = float64

ALand = float64

AWater = int64

pop = int64

male\_pop = int64

female\_pop = int64

rent\_mean = float64

rent\_median = float64

rent\_stdev = float64

rent\_sample\_weight = float64

rent\_samples = float64

rent\_gt\_10 = float64

rent\_gt\_15 = float64

rent\_gt\_20 = float64

rent\_gt\_25 = float64

rent\_gt\_30 = float64

rent\_gt\_35 = float64

rent\_gt\_40 = float64

rent\_gt\_50 = float64

universe\_samples = int64

used\_samples = int64

hi\_mean = float64

hi\_median = float64

hi\_stdev = float64

hi\_sample\_weight = float64

hi\_samples = float64

family\_mean = float64

family\_median = float64

family\_stdev = float64

family\_sample\_weight = float64

family\_samples = float64

hc\_mortgage\_mean = float64

hc\_mortgage\_median = float64

hc\_mortgage\_stdev = float64

hc\_mortgage\_sample\_weight = float64

hc\_mortgage\_samples = float64

hc\_mean = float64

hc\_median = float64

hc\_stdev = float64

hc\_samples = float64

hc\_sample\_weight = float64

home\_equity\_second\_mortgage = float64

second\_mortgage = float64

home\_equity = float64

debt = float64

second\_mortgage\_cdf = float64

home\_equity\_cdf = float64

debt\_cdf = float64

hs\_degree = float64

hs\_degree\_male = float64

hs\_degree\_female = float64

male\_age\_mean = float64

male\_age\_median = float64

male\_age\_stdev = float64

male\_age\_sample\_weight = float64

male\_age\_samples = float64

female\_age\_mean = float64

female\_age\_median = float64

female\_age\_stdev = float64

female\_age\_sample\_weight = float64

female\_age\_samples = float64

pct\_own = float64

married = float64

married\_snp = float64

separated = float64

divorced = float64

bad\_debt = float64

bins = category

Bad\_Debt = float64

----

+\*In[111]:\*+

[source, ipython3]

----

cat\_variables = ['state',

'state\_ab',

'city',

'place',

'type',

]

----

+\*In[112]:\*+

[source, ipython3]

----

num\_variables = ['UID','COUNTYID','STATEID', 'zip\_code', 'area\_code','lat','lng','ALand','AWater','pop','male\_pop', 'female\_pop',

'rent\_mean','rent\_median','rent\_stdev','rent\_sample\_weight','rent\_samples','rent\_gt\_10','rent\_gt\_15','rent\_gt\_20','rent\_gt\_25',

'rent\_gt\_30','rent\_gt\_35','rent\_gt\_40','rent\_gt\_50','universe\_samples','used\_samples','hi\_mean','hi\_median','hi\_stdev','hi\_sample\_weight',

'hi\_samples','family\_mean','family\_median','family\_stdev','family\_sample\_weight','family\_samples','hc\_mortgage\_mean','hc\_mortgage\_median',

'hc\_mortgage\_stdev','hc\_mortgage\_sample\_weight','hc\_mortgage\_samples', 'hc\_mean', 'hc\_median','hc\_stdev','hc\_samples','hc\_sample\_weight',

'home\_equity\_second\_mortgage','second\_mortgage','home\_equity','debt','second\_mortgage\_cdf','home\_equity\_cdf','debt\_cdf','hs\_degree',

'hs\_degree\_male','hs\_degree\_female','male\_age\_mean','male\_age\_median','male\_age\_stdev','male\_age\_sample\_weight','male\_age\_samples',

'female\_age\_mean','female\_age\_median','female\_age\_stdev','female\_age\_sample\_weight','female\_age\_samples','pct\_own','married',

'married\_snp','separated','divorced','Bad\_Debt']

----

+\*In[113]:\*+

[source, ipython3]

----

fa\_train\_df = train\_df[num\_variables]

fa\_train\_df

----

+\*Out[113]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |COUNTYID |STATEID |zip\_code |area\_code |lat |lng |ALand |AWater

|pop |male\_pop |female\_pop |rent\_mean |rent\_median |rent\_stdev

|rent\_sample\_weight |rent\_samples |rent\_gt\_10 |rent\_gt\_15 |rent\_gt\_20

|rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35 |rent\_gt\_40 |rent\_gt\_50

|universe\_samples |used\_samples |hi\_mean |hi\_median |hi\_stdev

|hi\_sample\_weight |hi\_samples |family\_mean |family\_median |family\_stdev

|family\_sample\_weight |family\_samples |hc\_mortgage\_mean

|hc\_mortgage\_median |hc\_mortgage\_stdev |hc\_mortgage\_sample\_weight

|hc\_mortgage\_samples |hc\_mean |hc\_median |hc\_stdev |hc\_samples

|hc\_sample\_weight |home\_equity\_second\_mortgage |second\_mortgage

|home\_equity |debt |second\_mortgage\_cdf |home\_equity\_cdf |debt\_cdf

|hs\_degree |hs\_degree\_male |hs\_degree\_female |male\_age\_mean

|male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced |Bad\_Debt

|0 |267822 |53 |36 |13346 |315 |42.840812 |-75.501524 |2.021834e+08

|1699120 |5230 |2612 |2618 |769.38638 |784.0 |232.63967 |272.34441

|362.0 |0.86761 |0.79155 |0.59155 |0.45634 |0.42817 |0.18592 |0.15493

|0.12958 |387 |355 |63125.28406 |48120.0 |49042.01206 |1290.96240

|2024.0 |67994.14790 |53245.0 |47667.30119 |884.33516 |1491.0

|1414.80295 |1223.0 |641.22898 |377.83135 |867.0 |570.01530 |558.0

|270.11299 |770.0 |499.29293 |0.01588 |0.02077 |0.08919 |0.52963

|0.43658 |0.49087 |0.73341 |0.89288 |0.85880 |0.92434 |42.48574

|44.00000 |22.97306 |696.42136 |2612.0 |44.48629 |45.33333 |22.51276

|685.33845 |2618.0 |0.79046 |0.57851 |0.01882 |0.01240 |0.08770 |0.09408

|1 |246444 |141 |18 |46616 |574 |41.701441 |-86.266614 |1.560828e+06

|100363 |2633 |1349 |1284 |804.87924 |848.0 |253.46747 |312.58622 |513.0

|0.97410 |0.93227 |0.69920 |0.69920 |0.55179 |0.41235 |0.39044 |0.27888

|542 |502 |41931.92593 |35186.0 |31639.50203 |838.74664 |1127.0

|50670.10337 |43023.0 |34715.57548 |375.28798 |554.0 |864.41390 |784.0

|482.27020 |316.88320 |356.0 |351.98293 |336.0 |125.40457 |229.0

|189.60606 |0.02222 |0.02222 |0.04274 |0.60855 |0.42174 |0.70823

|0.58120 |0.90487 |0.86947 |0.94187 |34.84728 |32.00000 |20.37452

|323.90204 |1349.0 |36.48391 |37.58333 |23.43353 |267.23367 |1284.0

|0.52483 |0.34886 |0.01426 |0.01426 |0.09030 |0.04274

|2 |245683 |63 |18 |46122 |317 |39.792202 |-86.515246 |6.956160e+07

|284193 |6881 |3643 |3238 |742.77365 |703.0 |323.39011 |291.85520 |378.0

|0.95238 |0.88624 |0.79630 |0.66667 |0.39153 |0.39153 |0.28307 |0.15873

|459 |378 |84942.68317 |74964.0 |56811.62186 |1155.20980 |2488.0

|95262.51431 |85395.0 |49292.67664 |709.74925 |1889.0 |1506.06758

|1361.0 |731.89394 |699.41354 |1491.0 |556.45986 |532.0 |184.42175

|538.0 |323.35354 |0.00000 |0.00000 |0.09512 |0.73484 |1.00000 |0.46332

|0.28704 |0.94288 |0.94616 |0.93952 |39.38154 |40.83333 |22.89769

|888.29730 |3643.0 |42.15810 |42.83333 |23.94119 |707.01963 |3238.0

|0.85331 |0.64745 |0.02830 |0.01607 |0.10657 |0.09512

|3 |279653 |127 |72 |927 |787 |18.396103 |-66.104169 |1.105793e+06 |0

|2700 |1141 |1559 |803.42018 |782.0 |297.39258 |259.30316 |368.0

|0.94693 |0.87151 |0.69832 |0.61732 |0.51397 |0.46927 |0.35754 |0.32961

|438 |358 |48733.67116 |37845.0 |45100.54010 |928.32193 |1267.0

|56401.68133 |44399.0 |41082.90515 |490.18479 |729.0 |1175.28642 |1101.0

|428.98751 |261.28471 |437.0 |288.04047 |247.0 |185.55887 |392.0

|314.90566 |0.01086 |0.01086 |0.01086 |0.52714 |0.53057 |0.82530

|0.73727 |0.91500 |0.90755 |0.92043 |48.64749 |48.91667 |23.05968

|274.98956 |1141.0 |47.77526 |50.58333 |24.32015 |362.20193 |1559.0

|0.65037 |0.47257 |0.02021 |0.02021 |0.10106 |0.01086

|4 |247218 |161 |20 |66502 |785 |39.195573 |-96.569366 |2.554403e+06 |0

|5637 |2586 |3051 |938.56493 |881.0 |392.44096 |1005.42886 |1704.0

|0.99286 |0.98247 |0.91688 |0.84740 |0.78247 |0.60974 |0.55455 |0.44416

|1725 |1540 |31834.15466 |22497.0 |34046.50907 |1548.67477 |1983.0

|54053.42396 |50272.0 |39609.12605 |244.08903 |395.0 |1192.58759 |1125.0

|327.49674 |76.61052 |134.0 |443.68855 |444.0 |76.12674 |124.0 |79.55556

|0.05426 |0.05426 |0.05426 |0.51938 |0.18332 |0.65545 |0.74967 |1.00000

|1.00000 |1.00000 |26.07533 |22.41667 |11.84399 |1296.89877 |2586.0

|24.17693 |21.58333 |11.10484 |1854.48652 |3051.0 |0.13046 |0.12356

|0.00000 |0.00000 |0.03109 |0.05426

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |... |... |... |... |... |... |... |... |... |... |...

|... |... |... |...

|27316 |279212 |43 |72 |769 |787 |18.076060 |-66.358379 |6.970300e+05 |0

|1847 |909 |938 |439.42839 |419.0 |140.29970 |170.00000 |170.0 |1.00000

|1.00000 |1.00000 |0.83333 |0.79012 |0.79012 |0.72222 |0.62963 |278 |162

|18515.67021 |13317.0 |23914.42656 |648.39020 |774.0 |20889.14617

|16760.0 |23488.17854 |346.58143 |446.0 |770.11560 |828.0 |157.85227

|58.00000 |58.0 |160.86544 |145.0 |94.04517 |438.0 |366.09000 |0.00000

|0.00000 |0.00000 |0.11694 |1.00000 |1.00000 |0.98762 |0.60155 |0.56962

|0.63222 |42.14011 |41.66667 |23.76478 |216.22207 |909.0 |42.73154

|40.16667 |24.79821 |230.87898 |938.0 |0.60422 |0.24603 |0.03042

|0.02249 |0.14683 |0.00000

|27317 |277856 |91 |42 |19422 |215 |40.158138 |-75.307271 |5.077337e+06

|11786 |4155 |2116 |2039 |1813.19253 |1788.0 |492.92300 |64.84927 |471.0

|0.85435 |0.63261 |0.50000 |0.37391 |0.30870 |0.30870 |0.26304 |0.23478

|484 |460 |119889.08320 |108284.0 |77625.25547 |518.53683 |1431.0

|118896.06830 |113313.0 |66663.51722 |388.97237 |1151.0 |2210.84055

|2202.0 |713.03361 |154.95504 |619.0 |712.16631 |663.0 |281.77621 |328.0

|174.96970 |0.00845 |0.02112 |0.19641 |0.65364 |0.43301 |0.12376

|0.47934 |0.95737 |0.95772 |0.95701 |37.75495 |38.83333 |21.45832

|530.17185 |2116.0 |38.21269 |39.50000 |21.84826 |496.20427 |2039.0

|0.68072 |0.61127 |0.05003 |0.02473 |0.04888 |0.20908

|27318 |233000 |87 |8 |80653 |970 |40.410316 |-103.814003 |1.323262e+09

|17577610 |2829 |1465 |1364 |849.39107 |834.0 |336.47530 |120.91448

|195.0 |0.93846 |0.71282 |0.54359 |0.44615 |0.29744 |0.23077 |0.16923

|0.09231 |237 |195 |79890.25113 |73350.0 |58132.65778 |529.41812 |1077.0

|88878.57034 |81864.0 |53510.48475 |375.21237 |871.0 |1671.07908 |1588.0

|742.67822 |182.53725 |488.0 |536.04921 |467.0 |306.82251 |352.0

|240.36899 |0.02024 |0.02024 |0.07857 |0.58095 |0.44186 |0.54095

|0.63916 |0.93555 |0.92200 |0.94887 |40.01134 |42.00000 |23.08048

|345.30911 |1465.0 |43.40218 |46.33333 |23.40858 |316.52078 |1364.0

|0.78508 |0.70451 |0.01386 |0.00520 |0.07712 |0.07857

|27319 |287425 |439 |48 |76034 |817 |32.904866 |-97.162151 |1.865230e+07

|158882 |11542 |5727 |5815 |1972.45746 |1843.0 |633.02173 |19.16328

|157.0 |1.00000 |1.00000 |0.75796 |0.61146 |0.50318 |0.50318 |0.27389

|0.27389 |234 |157 |165510.27110 |148548.0 |102038.58810 |960.70051

|4009.0 |167148.83770 |175952.0 |77638.35136 |719.65942 |3452.0

|3074.83088 |3188.0 |1121.07013 |536.61873 |2481.0 |1076.86881 |1037.0

|432.32205 |1294.0 |525.92451 |0.05801 |0.07550 |0.12556 |0.65722

|0.10759 |0.33196 |0.47094 |0.98540 |0.98883 |0.98201 |40.75409

|46.66667 |22.48690 |1305.28070 |5727.0 |39.25921 |43.41667 |21.36235

|1373.94120 |5815.0 |0.93970 |0.75503 |0.02287 |0.00915 |0.05261

|0.14305

|27320 |265371 |3 |32 |89123 |702 |36.064754 |-115.152237 |7.796308e+06

|0 |3726 |1815 |1911 |949.84199 |924.0 |198.82109 |555.87526 |1031.0

|0.94956 |0.87779 |0.83705 |0.63337 |0.51115 |0.41901 |0.27934 |0.10572

|1055 |1031 |51648.18703 |38072.0 |46305.44046 |1061.78932 |1409.0

|54886.07827 |42544.0 |39352.40334 |474.38547 |698.0 |1455.42340 |1364.0

|629.41356 |106.84849 |232.0 |540.26838 |454.0 |256.39951 |122.0

|79.07071 |0.01412 |0.01412 |0.18362 |0.65537 |0.50190 |0.15035 |0.47521

|0.87370 |0.87166 |0.87576 |34.81046 |32.50000 |20.16376 |461.52440

|1815.0 |34.45345 |29.83333 |19.77208 |526.73261 |1911.0 |0.27912

|0.34426 |0.03825 |0.03005 |0.13320 |0.18362

|===

27019 rows × 73 columns

----

+\*In[114]:\*+

[source, ipython3]

----

fa\_train\_df = fa\_train\_df[fa\_train\_df.columns[~fa\_train\_df.columns.isin(['COUNTYID','STATEID','lat','lng',])]]

----

+\*In[115]:\*+

[source, ipython3]

----

from sklearn.decomposition import FactorAnalysis

from factor\_analyzer import FactorAnalyzer

import warnings

warnings.filterwarnings('ignore')

----

+\*In[116]:\*+

[source, ipython3]

----

# Create factor analysis object and perform factor analysis

fa = FactorAnalyzer( rotation=None, n\_factors = 25)

fa.fit(fa\_train\_df)

# Check Eigenvalues

ev, v = fa.get\_eigenvalues()

ev

----

+\*Out[116]:\*+

----array([1.54796148e+01, 1.20114545e+01, 8.16503440e+00, 4.58170513e+00,

3.93327207e+00, 3.06205512e+00, 2.14869445e+00, 1.54832600e+00,

1.48291050e+00, 1.34492326e+00, 1.30003998e+00, 1.14591772e+00,

1.03538491e+00, 9.75175971e-01, 8.52888983e-01, 7.72717495e-01,

7.36442979e-01, 7.10461310e-01, 5.85800941e-01, 5.74126431e-01,

5.49480399e-01, 5.01685229e-01, 4.88383281e-01, 4.34177308e-01,

3.90237917e-01, 3.60209056e-01, 3.20246923e-01, 3.14284882e-01,

3.04847630e-01, 2.56738815e-01, 2.46368352e-01, 2.30358325e-01,

2.02215142e-01, 1.94047669e-01, 1.89621136e-01, 1.67526187e-01,

1.63672935e-01, 1.34963988e-01, 1.31819080e-01, 1.21717923e-01,

1.13224839e-01, 9.93809906e-02, 9.28682291e-02, 8.97772284e-02,

5.77534793e-02, 5.57259631e-02, 4.73337725e-02, 4.59558058e-02,

3.77244885e-02, 3.26460169e-02, 2.80009706e-02, 2.45163429e-02,

2.01138452e-02, 1.65389184e-02, 1.55250503e-02, 1.50860496e-02,

1.43903548e-02, 1.00258009e-02, 8.64889229e-03, 6.91923862e-03,

4.71901343e-03, 3.58180967e-03, 3.15718830e-03, 2.65807540e-03,

2.51478427e-03, 1.21836400e-03, 4.45426788e-04, 1.49884856e-15,

2.08074481e-16])----

+\*In[117]:\*+

[source, ipython3]

----

print(sorted(ev,))

----

+\*Out[117]:\*+

----

[2.0807448066841072e-16, 1.498848559097174e-15, 0.00044542678809433194, 0.0012183640027078502, 0.0025147842731133223, 0.002658075395064511, 0.003157188299615511, 0.003581809666936499, 0.004719013434512463, 0.006919238620539721, 0.008648892286418538, 0.010025800927756309, 0.014390354751365436, 0.015086049566099874, 0.015525050251326718, 0.016538918363320745, 0.020113845245985654, 0.024516342901951826, 0.028000970635954768, 0.032646016896399614, 0.037724488509944684, 0.04595580576165212, 0.04733377248836864, 0.05572596309911661, 0.05775347928851887, 0.08977722837898074, 0.09286822905836241, 0.09938099064486902, 0.11322483851494092, 0.12171792260337373, 0.13181908004906936, 0.13496398756405287, 0.16367293526870524, 0.1675261865849374, 0.1896211363338694, 0.1940476689533682, 0.20221514210358602, 0.2303583245959046, 0.2463683521213322, 0.256738815396396, 0.30484762990460396, 0.3142848823775822, 0.3202469233976112, 0.3602090563305132, 0.3902379173993906, 0.4341773078827861, 0.4883832814129391, 0.5016852291481207, 0.5494803985315786, 0.5741264306618326, 0.5858009410479698, 0.7104613095608943, 0.7364429791803128, 0.7727174950281911, 0.8528889826103567, 0.9751759705094343, 1.0353849074772141, 1.1459177202891617, 1.3000399785982577, 1.3449232560900142, 1.4829105032885388, 1.5483259956617834, 2.148694445967115, 3.0620551163213, 3.933272072279956, 4.581705131133648, 8.165034401557909, 12.011454487988365, 15.479614758736105]

----

+\*In[118]:\*+

[source, ipython3]

----

loadings = fa.loadings\_

----

+\*In[119]:\*+

[source, ipython3]

----

xvals = range(1, fa\_train\_df.shape[1]+1)

----

+\*In[120]:\*+

[source, ipython3]

----

sns.set()

plt.figure(figsize = (15,10))

plt.scatter(xvals, ev)

plt.plot(xvals, ev)

plt.title('Scree plot')

plt.xlabel('Factors')

plt.ylabel('Eigen Value')

plt.grid(color = 'green', )

plt.grid(b=True, which='minor', color='r', linestyle='--')

plt.minorticks\_on()

plt.show()

----

+\*Out[120]:\*+

----

![png](output\_147\_0.png)

----

+\*In[121]:\*+

[source, ipython3]

----

Factors = pd.DataFrame.from\_records(loadings)

Factors = Factors.add\_prefix('Factor ')

Factors.index = fa\_train\_df.columns

Factors

----

+\*Out[121]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |Factor 0 |Factor 1 |Factor 2 |Factor 3 |Factor 4 |Factor 5 |Factor 6

|Factor 7 |Factor 8 |Factor 9 |Factor 10 |Factor 11 |Factor 12 |Factor

13 |Factor 14 |Factor 15 |Factor 16 |Factor 17 |Factor 18 |Factor 19

|Factor 20 |Factor 21 |Factor 22 |Factor 23 |Factor 24

|UID |-0.077596 |-0.005912 |-0.160150 |-0.089152 |0.064134 |-0.031105

|-0.093358 |-0.053239 |-0.268283 |-0.118066 |-0.032843 |0.157924

|-0.032106 |0.000194 |0.031465 |0.057326 |-0.001170 |-0.105931 |0.033559

|-0.004247 |0.012583 |0.011706 |0.040383 |0.011193 |0.065961

|zip\_code |-0.033732 |0.090385 |0.058919 |-0.104747 |-0.046177

|-0.106389 |0.056367 |0.089799 |0.503184 |0.439620 |0.008504 |-0.062270

|-0.026260 |0.084042 |-0.196527 |0.019611 |-0.147151 |0.209139 |0.281170

|0.025317 |0.094464 |-0.036035 |-0.076482 |0.003758 |0.013530

|area\_code |0.020546 |0.016520 |0.017181 |0.026291 |0.001856 |-0.067501

|0.049580 |0.014817 |-0.002822 |-0.026493 |-0.024004 |0.036723 |0.037551

|-0.016187 |0.018446 |0.043434 |-0.052663 |0.028599 |-0.040905

|-0.005239 |0.014698 |0.000794 |0.051050 |0.006409 |0.021286

|ALand |-0.041496 |-0.046072 |-0.118814 |-0.045479 |0.045941 |-0.048366

|0.042653 |-0.141977 |0.153832 |0.179443 |0.175461 |0.154310 |0.613838

|-0.193436 |0.077201 |0.014949 |0.041178 |-0.042211 |-0.009742

|-0.004645 |0.010756 |0.005932 |0.003186 |0.005975 |0.008023

|AWater |-0.008581 |-0.022861 |-0.031445 |-0.025663 |0.026467 |-0.025605

|0.016392 |-0.086698 |0.106793 |0.110095 |0.114547 |0.111514 |0.458340

|-0.141245 |0.073024 |0.007404 |0.057686 |-0.038448 |-0.011868 |0.010131

|0.005423 |0.005800 |-0.011218 |-0.006881 |-0.010858

|pop |0.312113 |0.929652 |-0.061383 |0.033231 |0.023261 |-0.112052

|0.063915 |0.001917 |0.003662 |-0.003424 |0.012954 |-0.001188 |0.008015

|-0.008466 |-0.007038 |-0.072809 |-0.076266 |-0.029874 |-0.008743

|0.008268 |-0.016011 |0.002133 |0.012780 |0.015892 |0.024582

|male\_pop |0.301582 |0.903810 |-0.055534 |0.009515 |0.025334 |-0.141054

|0.079914 |-0.024738 |0.047926 |-0.041836 |0.029123 |-0.016465 |0.018345

|0.002746 |-0.038779 |-0.127444 |-0.094943 |-0.049652 |0.019366

|0.001767 |0.052539 |-0.022504 |0.140845 |-0.008538 |0.026544

|female\_pop |0.310734 |0.920013 |-0.064818 |0.055320 |0.020320

|-0.079129 |0.045636 |0.028034 |-0.039979 |0.034443 |-0.003461 |0.013849

|-0.002458 |-0.019086 |0.024346 |-0.016045 |-0.054380 |-0.009140

|-0.035672 |0.014157 |-0.081628 |0.025819 |-0.111322 |0.038394 |0.020916

|rent\_mean |0.743732 |-0.075226 |0.357637 |0.143488 |0.098671 |-0.178126

|0.048161 |0.097916 |0.282941 |0.030286 |-0.070475 |-0.020834 |-0.074173

|0.051596 |0.359647 |-0.041092 |0.080788 |0.017370 |-0.037338 |-0.005924

|-0.033079 |-0.105966 |0.005268 |0.018521 |-0.001666

|rent\_median |0.702145 |-0.069626 |0.348256 |0.129354 |0.079666

|-0.190621 |0.051746 |0.103009 |0.274145 |0.015888 |-0.077879 |-0.015184

|-0.072675 |0.041918 |0.368117 |-0.027578 |0.075804 |0.014979 |-0.026380

|-0.004806 |-0.035936 |-0.103081 |0.037910 |-0.002415 |-0.002763

|rent\_stdev |0.562597 |-0.060136 |0.263551 |0.155247 |0.205355 |0.006132

|0.078388 |0.002392 |0.100702 |0.096935 |0.024553 |-0.036866 |-0.027338

|0.028754 |0.062381 |-0.099783 |0.060140 |0.024965 |-0.061780 |-0.043081

|-0.007428 |0.007144 |-0.080823 |0.047440 |0.000186

|rent\_sample\_weight |-0.409924 |0.490141 |0.229983 |-0.183515 |0.267992

|0.451188 |-0.068046 |0.031652 |-0.105220 |0.069236 |0.021202 |-0.004693

|-0.000026 |0.015017 |-0.046443 |0.048355 |0.038905 |-0.062870 |0.096083

|-0.016796 |0.033485 |0.038484 |-0.023769 |0.015997 |-0.051440

|rent\_samples |-0.127223 |0.573900 |0.456014 |-0.183095 |0.393583

|0.470636 |-0.064904 |0.049391 |0.041971 |0.079113 |-0.001222 |-0.010950

|-0.022773 |-0.002314 |0.045733 |0.083400 |0.010944 |-0.071977 |0.011977

|0.002077 |0.014357 |-0.007272 |0.018805 |-0.011321 |-0.007354

|rent\_gt\_10 |-0.060862 |0.116914 |0.269571 |0.295956 |-0.163787

|0.023336 |-0.141162 |0.222715 |0.049153 |-0.102712 |-0.277010 |0.077470

|0.103022 |0.074926 |-0.024326 |0.016017 |0.010501 |-0.008361 |-0.020849

|0.175284 |0.034652 |0.000436 |0.013600 |-0.006971 |0.000002

|rent\_gt\_15 |-0.123098 |0.137628 |0.422956 |0.500764 |-0.182356

|0.023056 |-0.175298 |0.288777 |0.087061 |-0.127117 |-0.409682 |0.136080

|0.199593 |0.125668 |-0.098676 |-0.009342 |0.026742 |-0.051564 |0.005591

|0.202211 |0.044914 |0.041403 |-0.012099 |0.002966 |0.003586

|rent\_gt\_20 |-0.220565 |0.120693 |0.462668 |0.596757 |-0.139327

|0.016579 |-0.125816 |0.142957 |0.038781 |-0.053926 |-0.202577 |0.078359

|0.112817 |0.059201 |-0.088891 |-0.011684 |0.002500 |-0.023582 |0.025378

|-0.148306 |-0.034412 |-0.008099 |-0.003426 |0.011984 |-0.002359

|rent\_gt\_25 |-0.284389 |0.110596 |0.482168 |0.679329 |-0.116806

|-0.008329 |-0.107101 |0.030134 |-0.000744 |0.003411 |-0.044531

|0.042597 |0.047766 |0.017771 |-0.079539 |0.005800 |-0.013041 |-0.022690

|0.047091 |-0.324936 |-0.067746 |-0.022755 |0.000838 |0.003320

|-0.003978

|rent\_gt\_30 |-0.307219 |0.099020 |0.465401 |0.696905 |-0.102486

|-0.044812 |-0.095878 |-0.079463 |-0.020972 |0.057118 |0.104505

|0.010968 |-0.026561 |-0.014568 |-0.023071 |0.036029 |-0.008515

|-0.013102 |0.029588 |-0.148207 |-0.028743 |-0.008306 |0.023772

|-0.014173 |0.003329

|rent\_gt\_35 |-0.302743 |0.083318 |0.453447 |0.702086 |-0.081946

|-0.071889 |-0.099313 |-0.151310 |-0.040916 |0.088948 |0.203128

|-0.016464 |-0.079442 |-0.033408 |0.023819 |0.044051 |-0.003900

|-0.015953 |0.024470 |0.026239 |0.010609 |0.007159 |0.026911 |-0.019642

|0.005325

|rent\_gt\_40 |-0.300053 |0.072453 |0.442581 |0.689908 |-0.071711

|-0.080620 |-0.104171 |-0.203306 |-0.058224 |0.106647 |0.265623

|-0.034652 |-0.109347 |-0.050150 |0.056113 |0.046545 |-0.002465

|-0.013042 |0.007878 |0.188847 |0.051323 |0.023486 |0.009473 |-0.011889

|0.001328

|rent\_gt\_50 |-0.276521 |0.047079 |0.403543 |0.594196 |-0.041872

|-0.071204 |-0.089914 |-0.196118 |-0.062742 |0.089239 |0.228474

|-0.038836 |-0.100954 |-0.041455 |0.056096 |0.009041 |0.006175

|-0.003344 |-0.010745 |0.174374 |0.040398 |0.019194 |-0.023189 |0.006349

|-0.001262

|universe\_samples |-0.143267 |0.593654 |0.423208 |-0.171029 |0.399718

|0.469989 |-0.054246 |0.044716 |0.031337 |0.097714 |-0.004516 |-0.005891

|-0.019126 |0.006495 |0.051604 |0.079448 |0.012167 |-0.065907 |-0.004857

|0.001775 |0.005564 |-0.002962 |0.028068 |-0.000107 |0.005198

|used\_samples |-0.117594 |0.576890 |0.447635 |-0.181448 |0.389943

|0.475756 |-0.059811 |0.063943 |0.033514 |0.097683 |-0.002279 |-0.003744

|-0.023034 |-0.000804 |0.051868 |0.079673 |0.005407 |-0.066749 |0.008095

|0.004140 |0.008146 |-0.009797 |0.023884 |-0.006553 |0.001534

|hi\_mean |0.943077 |-0.147234 |0.012830 |-0.016259 |0.075587 |-0.160192

|0.019857 |0.031430 |0.026844 |0.049514 |0.024191 |0.030956 |-0.036495

|0.046583 |-0.008005 |0.028764 |0.053238 |-0.129164 |0.087000 |-0.011258

|0.010741 |0.036897 |0.031681 |-0.012132 |-0.001006

|hi\_median |0.914089 |-0.134482 |0.000849 |-0.044412 |0.023795

|-0.202795 |0.017210 |0.057389 |0.032112 |0.027604 |0.012355 |0.029403

|-0.044373 |0.027352 |0.083487 |0.030222 |0.031583 |-0.136950 |0.137335

|-0.015794 |0.023742 |0.098948 |0.035826 |-0.013347 |-0.024666

|hi\_stdev |0.876351 |-0.157372 |0.041420 |0.076418 |0.212801 |-0.016012

|0.023156 |-0.051502 |-0.010527 |0.120590 |0.067162 |0.036426 |-0.015836

|0.088972 |-0.234188 |0.009632 |0.111065 |-0.080473 |-0.059437 |0.017253

|-0.020713 |-0.111890 |0.001455 |0.012576 |0.049315

|hi\_sample\_weight |-0.181810 |0.874210 |-0.155880 |0.071248 |0.112159

|0.320707 |-0.043901 |0.002109 |-0.013437 |-0.004016 |-0.010081

|-0.012818 |0.000725 |-0.011081 |0.018485 |0.073685 |0.126280 |0.067786

|0.016810 |-0.012155 |0.031177 |-0.018010 |0.012209 |0.001977 |-0.083943

|hi\_samples |0.324648 |0.878582 |-0.164777 |0.048530 |0.113702 |0.194103

|-0.054376 |0.018041 |0.004287 |-0.007809 |0.013741 |0.016707 |-0.014277

|-0.013613 |0.011188 |0.024084 |0.095716 |-0.003128 |0.037055 |0.007587

|-0.010470 |0.009071 |-0.043990 |-0.076235 |0.033106

|family\_mean |0.935178 |-0.172472 |0.003265 |-0.000343 |0.136463

|-0.059044 |-0.052536 |-0.018556 |0.014840 |0.043187 |0.035745 |0.035569

|-0.029683 |0.066970 |-0.087927 |0.002420 |0.092553 |-0.155202 |0.079878

|-0.002794 |0.018630 |0.038105 |-0.021730 |0.012884 |-0.021882

|family\_median |0.910606 |-0.171049 |0.001261 |-0.010209 |0.120264

|-0.089111 |-0.046609 |-0.009182 |0.007303 |0.034042 |0.028573 |0.034865

|-0.040297 |0.053334 |-0.012804 |-0.000446 |0.085958 |-0.165071

|0.142942 |-0.014071 |0.030631 |0.109892 |-0.024045 |0.013491 |-0.052544

|family\_stdev |0.805823 |-0.140787 |0.050887 |0.070713 |0.224514

|0.068981 |-0.035138 |-0.071102 |-0.000779 |0.108754 |0.071569 |0.043840

|-0.001024 |0.101275 |-0.302114 |-0.003852 |0.121734 |-0.056082

|-0.117201 |0.037406 |-0.027333 |-0.167118 |-0.002328 |0.009062

|0.062778

|family\_sample\_weight |-0.140428 |0.878434 |-0.224347 |0.099347

|-0.049966 |-0.005757 |0.128523 |0.051524 |-0.032503 |0.009753

|-0.035769 |-0.016530 |0.009008 |-0.047274 |0.040612 |0.117883 |0.010855

|0.100770 |-0.036946 |-0.016266 |0.015120 |-0.046237 |0.082949

|-0.036866 |-0.029045

|family\_samples |0.400397 |0.834953 |-0.254217 |0.089026 |-0.052727

|-0.091326 |0.070138 |0.056535 |-0.036151 |0.002544 |0.001163 |0.020848

|-0.001391 |-0.034793 |0.019167 |0.055504 |-0.023188 |0.021816 |0.002882

|-0.001299 |-0.012898 |0.000585 |0.007088 |-0.106631 |0.070284

|hc\_mortgage\_mean |0.786079 |-0.135507 |0.393283 |0.106683 |0.278957

|-0.012644 |0.183037 |-0.035127 |0.004454 |0.009594 |-0.060698

|-0.050543 |0.026053 |-0.018442 |-0.051109 |0.017993 |-0.029092

|0.137653 |-0.093131 |-0.020739 |-0.002245 |0.205076 |0.039683

|-0.070429 |-0.007470

|hc\_mortgage\_median |0.765527 |-0.133562 |0.400878 |0.098939 |0.271786

|-0.027394 |0.177086 |-0.023716 |-0.000839 |0.008921 |-0.057653

|-0.045212 |0.016999 |-0.022767 |-0.012193 |0.010633 |-0.021350

|0.116296 |-0.043621 |-0.030491 |0.010488 |0.201186 |0.026872 |-0.046879

|0.010114

|hc\_mortgage\_stdev |0.683793 |-0.137525 |0.172544 |0.141481 |0.249126

|0.079814 |0.168542 |-0.083499 |0.035745 |0.049698 |-0.025426 |-0.056344

|0.050006 |0.037108 |-0.173016 |0.012487 |-0.005696 |0.110470 |-0.187341

|0.026976 |-0.053127 |0.006498 |0.041420 |-0.074803 |-0.070813

|hc\_mortgage\_sample\_weight |0.122760 |0.644690 |-0.503408 |0.033589

|-0.313583 |-0.082331 |-0.146519 |0.078965 |-0.018945 |-0.089690

|0.081457 |0.015102 |-0.030165 |-0.067130 |-0.007526 |-0.000601

|0.144421 |0.037422 |0.030352 |-0.002051 |-0.015614 |-0.049222

|-0.049294 |-0.103096 |-0.069136

|hc\_mortgage\_samples |0.591376 |0.605718 |-0.276426 |0.070738 |-0.243665

|-0.152876 |-0.079666 |0.097313 |-0.013606 |-0.089725 |0.074260

|0.022361 |-0.011336 |-0.090238 |-0.026688 |-0.070310 |0.031036

|-0.002489 |0.037925 |0.006933 |-0.058564 |0.012652 |-0.112697

|-0.190893 |0.033225

|hc\_mean |0.703671 |-0.168746 |0.296476 |0.108174 |0.345752 |0.019805

|0.079445 |-0.080169 |-0.303686 |-0.163484 |-0.104620 |0.122433

|0.029319 |-0.211804 |0.023887 |-0.018890 |-0.045676 |0.133824 |0.171231

|0.020887 |0.044351 |-0.104619 |-0.028002 |0.029833 |-0.005489

|hc\_median |0.672333 |-0.159122 |0.296020 |0.097642 |0.330089 |0.012652

|0.069733 |-0.072126 |-0.298154 |-0.162534 |-0.103671 |0.118434

|0.022532 |-0.204182 |0.053091 |-0.012645 |-0.042503 |0.107562 |0.174651

|0.007203 |0.057038 |-0.057565 |-0.026772 |0.036164 |0.009779

|hc\_stdev |0.539171 |-0.127850 |0.138949 |0.140513 |0.343838 |0.047749

|0.120487 |-0.104627 |-0.100727 |-0.021393 |-0.051407 |0.065706

|0.029122 |-0.059850 |-0.089958 |-0.033217 |-0.004170 |0.102016

|-0.001400 |0.034864 |-0.044335 |-0.103353 |0.006337 |-0.024414

|-0.037448

|hc\_samples |0.115084 |0.415911 |-0.713382 |0.289415 |0.074781

|-0.024762 |0.055232 |-0.259235 |0.019330 |-0.086046 |-0.101258

|-0.004137 |0.015120 |0.119364 |0.016998 |0.088080 |0.179604 |0.124573

|0.046272 |-0.019835 |0.070635 |0.030151 |-0.002631 |0.078659 |0.036619

|hc\_sample\_weight |-0.094535 |0.414708 |-0.736990 |0.240305 |0.007335

|-0.038162 |0.053710 |-0.242682 |0.077969 |-0.048947 |-0.080672

|-0.030211 |0.007781 |0.160294 |0.009581 |0.115498 |0.233028 |0.126968

|0.004538 |-0.029584 |0.069104 |0.054917 |0.041855 |0.095707 |0.015725

|home\_equity\_second\_mortgage |0.182939 |0.089819 |0.298799 |-0.152366

|-0.597072 |0.305237 |0.274132 |-0.282874 |0.111337 |0.058131 |-0.132461

|0.228965 |-0.116887 |-0.087164 |0.005511 |0.012344 |0.005451 |-0.037323

|-0.008596 |-0.001575 |0.003065 |0.006392 |-0.008834 |0.003143 |0.050544

|second\_mortgage |0.205049 |0.084046 |0.325027 |-0.144140 |-0.617343

|0.317388 |0.302815 |-0.308750 |0.127357 |0.068142 |-0.152298 |0.261751

|-0.137387 |-0.116484 |-0.003998 |0.007903 |0.011124 |-0.044950

|-0.034058 |-0.008256 |0.002245 |0.029099 |-0.009777 |0.014162 |0.059590

|home\_equity |0.597270 |-0.024601 |0.302039 |-0.058138 |-0.472422

|0.308463 |0.119319 |-0.061510 |-0.154390 |-0.092062 |0.072989

|-0.277572 |0.176937 |0.217383 |0.059987 |0.028992 |-0.033103 |0.013161

|0.069564 |0.001790 |0.002070 |-0.021903 |-0.015863 |0.003219 |0.078622

|debt |0.498359 |0.160430 |0.445734 |-0.190601 |-0.373983 |-0.015836

|-0.127786 |0.365619 |-0.007259 |-0.043242 |0.169447 |-0.049081

|-0.044325 |-0.240323 |-0.093375 |0.051835 |0.121381 |0.093210

|-0.049738 |-0.010196 |0.055184 |0.013403 |0.044601 |0.110984 |0.006547

|second\_mortgage\_cdf |-0.330222 |-0.097416 |-0.133053 |0.072059

|0.576977 |-0.232965 |-0.172573 |0.131613 |-0.054993 |-0.086875

|0.059412 |-0.170615 |0.093132 |0.088379 |0.041424 |0.051398 |-0.023824

|0.003737 |0.012183 |-0.009681 |0.001102 |0.023343 |-0.026359 |0.015422

|0.164106

|home\_equity\_cdf |-0.640792 |0.013454 |-0.250292 |0.031886 |0.465966

|-0.299437 |-0.074851 |0.007179 |0.159812 |0.061361 |-0.082318 |0.225375

|-0.142015 |-0.161366 |-0.025371 |0.017162 |0.013997 |-0.020843

|-0.032709 |-0.017116 |0.012603 |0.049292 |-0.004854 |0.018666 |0.115928

|debt\_cdf |-0.495487 |-0.163209 |-0.484044 |0.206375 |0.369605 |0.055073

|0.119242 |-0.371927 |-0.019438 |0.056671 |-0.177365 |0.062148 |0.035617

|0.259304 |0.087008 |-0.050374 |-0.108693 |-0.066543 |0.054619 |0.021049

|-0.051223 |-0.037680 |-0.020370 |-0.100071 |-0.048353

|hs\_degree |0.670908 |-0.120471 |-0.198753 |-0.064568 |-0.064310

|0.248797 |-0.582513 |-0.071816 |0.047801 |-0.024305 |0.099416 |0.211410

|-0.007389 |0.111062 |0.059463 |-0.015124 |-0.106962 |0.110286

|-0.044256 |-0.020172 |0.003896 |0.024129 |0.014703 |0.006628 |0.004712

|hs\_degree\_male |0.658640 |-0.108295 |-0.163464 |-0.063705 |-0.047075

|0.236792 |-0.532922 |-0.055792 |0.046802 |-0.005658 |0.085419 |0.177490

|-0.008020 |0.093823 |0.050016 |-0.011370 |-0.091607 |0.088051

|-0.038528 |-0.016402 |0.016606 |0.019135 |-0.011139 |0.006890 |0.020281

|hs\_degree\_female |0.630183 |-0.123833 |-0.218944 |-0.060291 |-0.072029

|0.226036 |-0.531327 |-0.075906 |0.047525 |-0.049315 |0.093716 |0.186558

|-0.003167 |0.096407 |0.039390 |-0.023367 |-0.075668 |0.076303

|-0.024287 |-0.012313 |-0.008496 |0.009779 |0.040768 |0.001573

|-0.014737

|male\_age\_mean |0.248133 |-0.308191 |-0.562167 |0.389360 |0.117669

|0.414329 |0.083315 |0.038049 |0.200102 |-0.123423 |0.020715 |-0.098300

|-0.017035 |-0.127143 |-0.028698 |0.025692 |-0.042128 |-0.026657

|0.035427 |0.049501 |-0.142909 |0.008792 |0.002250 |0.070520 |-0.007089

|male\_age\_median |0.322751 |-0.303836 |-0.575763 |0.377029 |0.063067

|0.341058 |0.096193 |0.047614 |0.204064 |-0.149950 |0.034004 |-0.098023

|-0.010714 |-0.159193 |-0.031614 |0.094065 |-0.086013 |-0.044338

|0.071911 |0.067181 |-0.233854 |0.030168 |0.066284 |0.065961 |0.013644

|male\_age\_stdev |0.073658 |-0.090839 |-0.489653 |0.300919 |-0.068554

|0.188529 |0.219820 |0.284719 |-0.315475 |0.401054 |0.010483 |0.140075

|-0.008632 |0.070703 |0.093452 |-0.087160 |-0.010772 |0.048340

|-0.025575 |0.001735 |-0.025749 |0.025745 |0.022509 |0.021341 |0.019151

|male\_age\_sample\_weight |0.237000 |0.854174 |0.015497 |-0.009960

|0.038307 |-0.187754 |-0.003804 |-0.137815 |0.065946 |-0.080594

|0.013516 |-0.050478 |0.033437 |0.034724 |-0.057250 |-0.160516

|-0.155909 |-0.065235 |-0.026104 |-0.007607 |0.076084 |-0.013261

|0.100948 |0.097478 |-0.039410

|male\_age\_samples |0.301566 |0.903692 |-0.055642 |0.008980 |0.025266

|-0.141338 |0.079566 |-0.025378 |0.048611 |-0.043285 |0.029158

|-0.017003 |0.018231 |0.001932 |-0.038970 |-0.126436 |-0.095141

|-0.049264 |0.019403 |0.000158 |0.053376 |-0.022775 |0.141337 |-0.010053

|0.026760

|female\_age\_mean |0.187575 |-0.289390 |-0.542939 |0.426231 |0.109478

|0.475294 |0.086704 |0.089186 |0.139126 |-0.102958 |0.009912 |-0.097478

|-0.034340 |-0.112340 |0.020185 |-0.117785 |-0.040184 |-0.071121

|-0.086601 |-0.054057 |0.232044 |-0.018214 |-0.095888 |-0.017496

|0.006628

|female\_age\_median |0.259249 |-0.290627 |-0.575975 |0.415916 |0.042546

|0.373794 |0.097912 |0.081746 |0.132389 |-0.117720 |0.015176 |-0.091400

|-0.021626 |-0.126839 |0.014612 |-0.065312 |-0.072259 |-0.071008

|-0.059793 |-0.047301 |0.180856 |-0.006073 |-0.030583 |-0.043091

|0.016122

|female\_age\_stdev |-0.015700 |-0.059259 |-0.416693 |0.245375 |-0.039117

|0.221330 |0.229888 |0.293886 |-0.322785 |0.381499 |0.044898 |0.140723

|-0.002036 |0.110680 |0.069392 |-0.190618 |0.022481 |0.031862 |-0.013356

|0.008647 |0.001220 |0.026354 |0.027304 |0.056621 |0.006561

|female\_age\_sample\_weight |0.245384 |0.884640 |0.014483 |0.027539

|0.037269 |-0.129181 |-0.037002 |-0.094032 |-0.021313 |-0.009695

|-0.015596 |-0.018243 |0.012265 |0.011760 |-0.006195 |-0.044451

|-0.107100 |-0.018706 |-0.064782 |0.019587 |-0.092752 |0.040910

|-0.160367 |0.157797 |-0.052247

|female\_age\_samples |0.310562 |0.920557 |-0.064260 |0.054877 |0.020327

|-0.079881 |0.045765 |0.026761 |-0.037626 |0.029405 |-0.002391 |0.013846

|-0.002320 |-0.018670 |0.023121 |-0.018146 |-0.051398 |-0.008185

|-0.034131 |0.014959 |-0.081787 |0.026668 |-0.113102 |0.036456 |0.018882

|pct\_own |0.471871 |-0.128450 |-0.633571 |0.221310 |-0.302482 |-0.232346

|0.000248 |0.037118 |-0.070408 |0.008216 |0.022697 |0.033970 |0.002943

|-0.024705 |-0.047584 |0.002298 |-0.001005 |-0.069253 |0.043301

|0.009656 |0.023069 |-0.005692 |0.038363 |0.036181 |-0.071521

|married |0.542058 |-0.035213 |-0.494500 |0.134888 |-0.134294 |-0.107210

|0.142693 |0.202272 |-0.047197 |0.130806 |0.006321 |0.107970 |-0.018050

|0.022228 |0.029829 |0.449888 |-0.196928 |-0.037457 |-0.018585

|-0.004241 |0.018864 |-0.059395 |0.026615 |-0.017890 |-0.009988

|married\_snp |-0.351425 |0.050869 |0.279614 |0.008633 |0.165876

|0.003290 |0.495271 |0.178275 |0.173296 |-0.380023 |0.374532 |0.326201

|-0.000531 |0.241963 |-0.039769 |0.005818 |-0.026444 |0.039738 |0.021422

|0.005618 |0.032252 |-0.007097 |-0.018418 |0.019493 |-0.028472

|separated |-0.351776 |0.023794 |0.144953 |0.025631 |0.075014 |0.063910

|0.290905 |0.145116 |0.053165 |-0.268306 |0.251502 |0.253850 |-0.007436

|0.162599 |0.005309 |0.001681 |0.032917 |0.025847 |0.016182 |0.006211

|-0.029677 |0.029229 |-0.032887 |-0.013359 |0.012634

|divorced |-0.378481 |-0.033288 |-0.204218 |0.019137 |-0.041945

|0.315502 |-0.078960 |0.052693 |0.109751 |-0.008094 |0.052182 |0.003946

|-0.025122 |-0.056326 |-0.034802 |-0.231377 |0.117379 |0.076072

|0.173436 |0.047973 |-0.135809 |0.021200 |0.112202 |-0.041263 |0.040202

|Bad\_Debt |0.594729 |-0.023288 |0.314028 |-0.056402 |-0.479608 |0.308619

|0.131244 |-0.070376 |-0.139319 |-0.082933 |0.061720 |-0.250629

|0.159471 |0.192679 |0.051243 |0.023928 |-0.028123 |0.011881 |0.052500

|-0.001134 |0.001060 |-0.010502 |-0.012201 |0.006618 |0.050163

|===

----

+\*In[122]:\*+

[source, ipython3]

----

fa = FactorAnalyzer( rotation="varimax", n\_factors = 12)

fa.fit(fa\_train\_df)

loadings = fa.loadings\_

----

+\*In[123]:\*+

[source, ipython3]

----

Factors = pd.DataFrame.from\_records(loadings)

Factors = Factors.add\_prefix('Factor ')

Factors.index = fa\_train\_df.columns

Factors

----

+\*Out[123]:\*+

----

[cols=",,,,,,,,,,,,",options="header",]

|===

| |Factor 0 |Factor 1 |Factor 2 |Factor 3 |Factor 4 |Factor 5 |Factor 6

|Factor 7 |Factor 8 |Factor 9 |Factor 10 |Factor 11

|UID |-0.112586 |0.004241 |-0.036947 |-0.101178 |-0.061646 |0.078145

|-0.064224 |-0.071115 |-0.106840 |-0.057157 |0.047487 |0.332240

|zip\_code |-0.020936 |0.051244 |0.033846 |-0.033193 |-0.101020

|-0.019504 |0.052504 |0.018441 |-0.079273 |-0.004742 |0.006307

|-0.476403

|area\_code |0.039809 |0.030462 |-0.049947 |0.001307 |-0.022502

|-0.061328 |0.002226 |-0.001699 |-0.023979 |0.039786 |-0.006290

|0.017645

|ALand |-0.036622 |-0.025379 |-0.041890 |-0.041833 |0.018012 |0.010368

|-0.020354 |-0.106733 |-0.048692 |-0.109154 |0.008241 |-0.100266

|AWater |0.001326 |-0.016861 |-0.009558 |-0.013094 |-0.005574 |0.012457

|-0.003680 |-0.061850 |-0.019305 |-0.052723 |-0.009578 |-0.079733

|pop |0.121939 |0.971792 |0.111528 |-0.008067 |-0.103230 |-0.018836

|0.033275 |0.040760 |0.014982 |0.021534 |-0.003083 |-0.031701

|male\_pop |0.119778 |0.945261 |0.089806 |-0.021777 |-0.103149 |-0.043817

|0.029500 |0.031564 |0.020785 |0.003297 |-0.065809 |-0.050886

|female\_pop |0.118522 |0.955360 |0.129842 |0.005611 |-0.098871 |0.009357

|0.036188 |0.048156 |0.008775 |0.039712 |0.060216 |-0.010441

|rent\_mean |0.808782 |0.062231 |-0.100383 |0.038873 |-0.018580 |0.071305

|0.063942 |0.172360 |0.033088 |0.161125 |-0.119805 |-0.262949

|rent\_median |0.762394 |0.060685 |-0.110384 |0.031670 |-0.032132

|0.055003 |0.066331 |0.174879 |0.027543 |0.163562 |-0.122946 |-0.246594

|rent\_stdev |0.668352 |0.029767 |0.063406 |0.089784 |0.052039 |0.045229

|0.018203 |0.052872 |0.053749 |0.015774 |0.004122 |-0.144227

|rent\_sample\_weight |-0.324103 |0.206623 |0.780014 |0.041462 |-0.116369

|-0.089410 |-0.028917 |-0.018226 |0.000565 |0.002496 |0.031967 |0.092314

|rent\_samples |0.013260 |0.294747 |0.922634 |0.051875 |-0.138046

|-0.072719 |0.011538 |0.060214 |-0.009602 |0.060847 |-0.078222

|-0.033673

|rent\_gt\_10 |-0.028781 |0.046464 |0.032087 |0.207633 |-0.037483

|0.003976 |0.040677 |0.083309 |0.029072 |0.578115 |-0.020765 |0.005552

|rent\_gt\_15 |-0.015051 |0.030140 |0.060974 |0.377503 |-0.046665

|-0.039391 |0.039916 |0.059244 |0.020013 |0.794749 |-0.025421 |-0.009420

|rent\_gt\_20 |-0.043850 |-0.001858 |0.084753 |0.586947 |-0.037424

|-0.100848 |0.027831 |0.029511 |0.011257 |0.630222 |-0.013985 |-0.001994

|rent\_gt\_25 |-0.065369 |-0.015178 |0.083076 |0.747218 |-0.037764

|-0.131653 |0.013806 |0.016694 |0.003233 |0.439481 |-0.007397 |0.002033

|rent\_gt\_30 |-0.074697 |-0.016458 |0.059100 |0.866916 |-0.043346

|-0.126150 |0.007877 |0.008093 |-0.009922 |0.263261 |-0.008569

|-0.009015

|rent\_gt\_35 |-0.062914 |-0.020261 |0.042529 |0.938167 |-0.049149

|-0.109510 |-0.007174 |0.007933 |-0.014811 |0.123465 |-0.013615

|-0.010573

|rent\_gt\_40 |-0.061378 |-0.026337 |0.039248 |0.941419 |-0.055728

|-0.098016 |-0.010974 |0.002384 |-0.012505 |0.048443 |-0.021338

|-0.003827

|rent\_gt\_50 |-0.050104 |-0.045853 |0.048031 |0.833201 |-0.063082

|-0.093540 |-0.017717 |-0.013352 |-0.010863 |0.027105 |-0.028608

|0.007593

|universe\_samples |-0.006127 |0.316879 |0.919778 |0.052528 |-0.131339

|-0.075801 |0.003853 |0.033029 |-0.017007 |0.054165 |-0.048898

|-0.036417

|used\_samples |0.018020 |0.300070 |0.922395 |0.046027 |-0.136913

|-0.067961 |0.014756 |0.065323 |-0.010605 |0.062609 |-0.050214

|-0.037838

|hi\_mean |0.847878 |0.106410 |-0.264597 |-0.202405 |0.043355 |0.267842

|0.045623 |0.138570 |0.076554 |-0.071853 |0.035324 |-0.063679

|hi\_median |0.795632 |0.118389 |-0.306857 |-0.224085 |0.016670 |0.253463

|0.053948 |0.172051 |0.072272 |-0.057371 |0.020122 |-0.063943

|hi\_stdev |0.851093 |0.058902 |-0.101032 |-0.093312 |0.112266 |0.268055

|0.015212 |0.022788 |0.075740 |-0.111689 |0.085987 |-0.035039

|hi\_sample\_weight |-0.328824 |0.757185 |0.474313 |0.057306 |0.124525

|-0.004819 |-0.007304 |-0.078986 |-0.015715 |0.052414 |0.060433

|0.055198

|hi\_samples |0.090486 |0.902417 |0.307074 |-0.054175 |0.123619 |0.146601

|0.027785 |0.027230 |0.011895 |0.016373 |0.047768 |0.034954

|family\_mean |0.843559 |0.067521 |-0.178556 |-0.181603 |0.098713

|0.345158 |0.026594 |0.091627 |0.080106 |-0.086037 |0.020459 |-0.021867

|family\_median |0.817412 |0.067301 |-0.200844 |-0.185023 |0.078389

|0.324822 |0.024878 |0.102869 |0.071085 |-0.083900 |0.017448 |-0.014851

|family\_stdev |0.773865 |0.041602 |-0.019026 |-0.077581 |0.134260

|0.299889 |0.017363 |0.010967 |0.077892 |-0.093516 |0.058882 |-0.016105

|family\_sample\_weight |-0.310430 |0.843632 |0.131328 |0.035218 |0.009630

|-0.164623 |0.011643 |-0.046685 |-0.029039 |0.057454 |0.111539 |0.002197

|family\_samples |0.125283 |0.952476 |-0.040137 |-0.074422 |0.041767

|0.054593 |0.044502 |0.058256 |0.016679 |0.016110 |0.110571 |0.012480

|hc\_mortgage\_mean |0.938839 |-0.009591 |0.055698 |0.008556 |0.003559

|-0.020011 |0.062270 |0.048934 |0.123749 |0.038825 |-0.049568 |-0.000502

|hc\_mortgage\_median |0.922268 |-0.013643 |0.052785 |0.012035 |-0.015125

|-0.027745 |0.057974 |0.062418 |0.114663 |0.040705 |-0.049313 |0.001890

|hc\_mortgage\_stdev |0.767894 |0.004274 |0.031681 |-0.014528 |0.149501

|0.045629 |0.048531 |-0.057484 |0.127810 |-0.021525 |0.015411 |-0.035724

|hc\_mortgage\_sample\_weight |-0.302301 |0.761043 |-0.235995 |-0.105659

|0.138672 |0.224861 |0.028704 |0.143002 |0.005886 |-0.023568 |0.063531

|0.050010

|hc\_mortgage\_samples |0.209518 |0.799823 |-0.267430 |-0.105242 |0.086782

|0.244059 |0.089724 |0.245646 |0.071486 |0.015003 |0.029999 |0.040386

|hc\_mean |0.853213 |-0.040512 |0.049257 |-0.006198 |0.012949 |0.019390

|0.032121 |0.048521 |-0.012959 |0.022563 |-0.024999 |0.466890

|hc\_median |0.815786 |-0.039625 |0.049551 |-0.005243 |-0.001234

|0.019782 |0.028322 |0.054867 |-0.011607 |0.026410 |-0.027975 |0.445896

|hc\_stdev |0.675426 |-0.012413 |0.061338 |0.003860 |0.109676 |0.010447

|-0.008474 |-0.091568 |-0.022804 |-0.032449 |0.021379 |0.177777

|hc\_samples |-0.111502 |0.609176 |-0.244100 |-0.066836 |0.407314

|0.122120 |-0.099459 |-0.452241 |-0.063341 |-0.080225 |0.082302

|0.081150

|hc\_sample\_weight |-0.318933 |0.556084 |-0.235377 |-0.061118 |0.365835

|0.076620 |-0.107165 |-0.449671 |-0.072684 |-0.089843 |0.081407

|-0.011604

|home\_equity\_second\_mortgage |0.036599 |0.030742 |0.040411 |0.004534

|-0.082194 |-0.004268 |0.908958 |0.058745 |0.153899 |0.035261 |-0.049769

|-0.052803

|second\_mortgage |0.065682 |0.025479 |0.039618 |0.016296 |-0.079209

|-0.014882 |0.976251 |0.056392 |0.129721 |0.038877 |-0.051754 |-0.057523

|home\_equity |0.375872 |0.033172 |-0.021182 |-0.028242 |0.001330

|0.158436 |0.386714 |0.183863 |0.786357 |0.056795 |-0.013655 |0.054132

|debt |0.315265 |0.158853 |0.005795 |-0.029053 |-0.221256 |0.129182

|0.239939 |0.736828 |0.212319 |0.114754 |-0.074490 |-0.034893

|second\_mortgage\_cdf |-0.100654 |-0.121830 |0.086851 |0.021597

|-0.002031 |-0.117485 |-0.710529 |-0.162084 |-0.179477 |-0.031884

|-0.064833 |0.050131

|home\_equity\_cdf |-0.394038 |-0.065048 |0.045667 |0.031864 |-0.033492

|-0.212633 |-0.376034 |-0.241832 |-0.703281 |-0.059286 |-0.040679

|-0.060494

|debt\_cdf |-0.327573 |-0.157327 |0.000048 |0.026008 |0.249880 |-0.098175

|-0.231827 |-0.764342 |-0.206187 |-0.115445 |0.112790 |0.056673

|hs\_degree |0.350821 |0.031892 |-0.032624 |-0.161822 |0.210988 |0.852633

|0.075164 |0.128001 |0.024523 |-0.025890 |-0.056890 |0.086873

|hs\_degree\_male |0.366297 |0.035117 |-0.015986 |-0.156136 |0.192157

|0.788455 |0.070100 |0.130132 |0.034586 |-0.022132 |-0.046232 |0.067114

|hs\_degree\_female |0.319065 |0.031724 |-0.062496 |-0.167708 |0.222855

|0.785307 |0.070453 |0.114143 |0.031383 |-0.034470 |-0.065432 |0.091673

|male\_age\_mean |0.130007 |-0.100396 |-0.125561 |-0.071276 |0.901081

|0.114619 |-0.050413 |-0.103890 |-0.002384 |-0.047621 |0.113277

|0.011406

|male\_age\_median |0.171431 |-0.062783 |-0.219382 |-0.094629 |0.856643

|0.122802 |-0.032505 |-0.064210 |0.008233 |-0.055755 |0.101876 |0.017033

|male\_age\_stdev |-0.031280 |0.029293 |-0.163320 |-0.028767 |0.307175

|0.047329 |-0.013750 |-0.067438 |-0.010779 |-0.020143 |0.851393

|0.031709

|male\_age\_sample\_weight |0.091173 |0.867808 |0.094634 |0.029740

|-0.180253 |-0.004665 |0.005865 |-0.036570 |0.031076 |0.002143

|-0.184896 |-0.043683

|male\_age\_samples |0.119554 |0.945260 |0.089463 |-0.022140 |-0.102808

|-0.043884 |0.029499 |0.031810 |0.020736 |0.002975 |-0.067573 |-0.050494

|female\_age\_mean |0.080113 |-0.108953 |-0.074722 |-0.035048 |0.876815

|0.103326 |-0.052421 |-0.096406 |0.012274 |-0.008558 |0.196012 |0.041083

|female\_age\_median |0.113651 |-0.069617 |-0.195973 |-0.058611 |0.860788

|0.110726 |-0.036211 |-0.068926 |0.021816 |-0.020197 |0.180851 |0.042759

|female\_age\_stdev |-0.089413 |0.013231 |-0.070376 |-0.025801 |0.264838

|-0.005727 |-0.026240 |-0.057190 |0.007983 |-0.028968 |0.749226

|0.037402

|female\_age\_sample\_weight |0.090135 |0.887633 |0.141556 |0.057187

|-0.184485 |0.043602 |0.015043 |-0.021810 |0.019510 |0.032224 |-0.075683

|-0.000642

|female\_age\_samples |0.118244 |0.956105 |0.129218 |0.005359 |-0.097668

|0.007946 |0.036233 |0.048683 |0.008951 |0.039798 |0.054904 |-0.009528

|pct\_own |0.145171 |0.202319 |-0.713170 |-0.147304 |0.324685 |0.283814

|0.008941 |0.004265 |0.028978 |-0.074282 |0.245691 |0.048210

|married |0.292649 |0.231916 |-0.433378 |-0.231699 |0.288202 |0.171294

|0.023045 |0.048801 |0.008950 |-0.049518 |0.325218 |-0.029076

|married\_snp |-0.102363 |-0.074157 |0.198606 |0.098548 |-0.010016

|-0.549773 |0.006961 |0.085132 |-0.082308 |0.001901 |-0.113763

|-0.009934

|separated |-0.191783 |-0.083068 |0.156562 |0.084529 |0.043272

|-0.407015 |0.007689 |0.089521 |-0.076785 |0.010449 |-0.047630 |0.056189

|divorced |-0.425760 |-0.115228 |0.163963 |0.002845 |0.262838 |0.024669

|0.006741 |-0.009231 |-0.060479 |-0.011539 |0.045327 |-0.020197

|Bad\_Debt |0.379525 |0.031390 |-0.019505 |-0.021507 |-0.000966 |0.149182

|0.420141 |0.185981 |0.759613 |0.058887 |-0.015178 |0.047499

|===

----

+\*In[124]:\*+

[source, ipython3]

----

Factors\_df = round(Factors.loc[['hs\_degree', 'hs\_degree\_male', 'hs\_degree\_female',"male\_age\_median", "female\_age\_median", "home\_equity\_second\_mortgage", 'second\_mortgage', 'second\_mortgage\_cdf', 'pct\_own', 'Bad\_Debt'], :], 2)

----

+\*In[125]:\*+

[source, ipython3]

----

len(fa\_train\_df.columns)

----

+\*Out[125]:\*+

----69----

+\*In[126]:\*+

[source, ipython3]

----

# Get variance of each factors

fact\_variance = fa.get\_factor\_variance()

fact\_variance

----

+\*Out[126]:\*+

----(array([12.05075025, 11.61628359, 5.09717084, 4.75177647, 4.40186725,

3.56509927, 2.96710152, 2.09472564, 1.99794727, 1.86158467,

1.77481955, 1.05669516]),

array([0.17464855, 0.16835194, 0.07387204, 0.06886633, 0.06379518,

0.05166811, 0.04300147, 0.03035834, 0.02895576, 0.02697949,

0.02572202, 0.01531442]),

array([0.17464855, 0.34300049, 0.41687253, 0.48573886, 0.54953403,

0.60120214, 0.64420361, 0.67456195, 0.70351771, 0.7304972 ,

0.75621922, 0.77153364]))----

+\*In[127]:\*+

[source, ipython3]

----

Factor\_variance = pd.DataFrame.from\_records(fact\_variance)

Factor\_variance = Factor\_variance.add\_prefix('Factor ')

Factor\_variance.index = ['SS Loadings', 'Proportion Var', 'Cumulative Var']

round(Factor\_variance, 2)

----

+\*Out[127]:\*+

----

[cols=",,,,,,,,,,,,",options="header",]

|===

| |Factor 0 |Factor 1 |Factor 2 |Factor 3 |Factor 4 |Factor 5 |Factor 6

|Factor 7 |Factor 8 |Factor 9 |Factor 10 |Factor 11

|SS Loadings |12.05 |11.62 |5.10 |4.75 |4.40 |3.57 |2.97 |2.09 |2.00

|1.86 |1.77 |1.06

|Proportion Var |0.17 |0.17 |0.07 |0.07 |0.06 |0.05 |0.04 |0.03 |0.03

|0.03 |0.03 |0.02

|Cumulative Var |0.17 |0.34 |0.42 |0.49 |0.55 |0.60 |0.64 |0.67 |0.70

|0.73 |0.76 |0.77

|===

----

Data Modeling :

1. Build a linear Regression model to predict the total monthly expenditure for home mortgages loan. Please refer ‘deplotment\_RE.xlsx’. Column hc\_mortgage\_mean is predicted variable. This is the mean monthly mortgage and owner costs of specified geographical location. Note: Exclude loans from prediction model which have NaN (Not a Number) values for hc\_mortgage\_mean.a) Run a model at a Nation level. If the accuracy levels and R square are not satisfactory proceed to below step.

b) Run another model at State level. There are 52 states in USA.

c) Keep below considerations while building a linear regression model. Data Modeling :

• 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

+\*In[128]:\*+

[source, ipython3]

----

train\_df.columns

----

+\*Out[128]:\*+

----Index(['UID', 'COUNTYID', 'STATEID', 'state', 'state\_ab', 'city', 'place',

'type', 'zip\_code', 'area\_code', 'lat', 'lng', 'ALand', 'AWater', 'pop',

'male\_pop', 'female\_pop', 'rent\_mean', 'rent\_median', 'rent\_stdev',

'rent\_sample\_weight', 'rent\_samples', 'rent\_gt\_10', 'rent\_gt\_15',

'rent\_gt\_20', 'rent\_gt\_25', 'rent\_gt\_30', 'rent\_gt\_35', 'rent\_gt\_40',

'rent\_gt\_50', 'universe\_samples', 'used\_samples', 'hi\_mean',

'hi\_median', 'hi\_stdev', 'hi\_sample\_weight', 'hi\_samples',

'family\_mean', 'family\_median', 'family\_stdev', 'family\_sample\_weight',

'family\_samples', 'hc\_mortgage\_mean', 'hc\_mortgage\_median',

'hc\_mortgage\_stdev', 'hc\_mortgage\_sample\_weight', 'hc\_mortgage\_samples',

'hc\_mean', 'hc\_median', 'hc\_stdev', 'hc\_samples', 'hc\_sample\_weight',

'home\_equity\_second\_mortgage', 'second\_mortgage', 'home\_equity', 'debt',

'second\_mortgage\_cdf', 'home\_equity\_cdf', 'debt\_cdf', 'hs\_degree',

'hs\_degree\_male', 'hs\_degree\_female', 'male\_age\_mean',

'male\_age\_median', 'male\_age\_stdev', 'male\_age\_sample\_weight',

'male\_age\_samples', 'female\_age\_mean', 'female\_age\_median',

'female\_age\_stdev', 'female\_age\_sample\_weight', 'female\_age\_samples',

'pct\_own', 'married', 'married\_snp', 'separated', 'divorced',

'bad\_debt', 'bins', 'Bad\_Debt'],

dtype='object')----

+\*In[129]:\*+

[source, ipython3]

----

train\_df.drop(['state\_ab','bins','bad\_debt','Bad\_Debt'],axis=1,inplace=True)

----

+\*In[130]:\*+

[source, ipython3]

----

train\_df.head()

----

+\*Out[130]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |UID |COUNTYID |STATEID |state |city |place |type |zip\_code |area\_code

|lat |lng |ALand |AWater |pop |male\_pop |female\_pop |rent\_mean

|rent\_median |rent\_stdev |rent\_sample\_weight |rent\_samples |rent\_gt\_10

|rent\_gt\_15 |rent\_gt\_20 |rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35 |rent\_gt\_40

|rent\_gt\_50 |universe\_samples |used\_samples |hi\_mean |hi\_median

|hi\_stdev |hi\_sample\_weight |hi\_samples |family\_mean |family\_median

|family\_stdev |family\_sample\_weight |family\_samples |hc\_mortgage\_mean

|hc\_mortgage\_median |hc\_mortgage\_stdev |hc\_mortgage\_sample\_weight

|hc\_mortgage\_samples |hc\_mean |hc\_median |hc\_stdev |hc\_samples

|hc\_sample\_weight |home\_equity\_second\_mortgage |second\_mortgage

|home\_equity |debt |second\_mortgage\_cdf |home\_equity\_cdf |debt\_cdf

|hs\_degree |hs\_degree\_male |hs\_degree\_female |male\_age\_mean

|male\_age\_median |male\_age\_stdev |male\_age\_sample\_weight

|male\_age\_samples |female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|0 |267822 |53 |36 |New York |Hamilton |Hamilton |City |13346 |315

|42.840812 |-75.501524 |202183361.0 |1699120 |5230 |2612 |2618

|769.38638 |784.0 |232.63967 |272.34441 |362.0 |0.86761 |0.79155

|0.59155 |0.45634 |0.42817 |0.18592 |0.15493 |0.12958 |387 |355

|63125.28406 |48120.0 |49042.01206 |1290.96240 |2024.0 |67994.14790

|53245.0 |47667.30119 |884.33516 |1491.0 |1414.80295 |1223.0 |641.22898

|377.83135 |867.0 |570.01530 |558.0 |270.11299 |770.0 |499.29293

|0.01588 |0.02077 |0.08919 |0.52963 |0.43658 |0.49087 |0.73341 |0.89288

|0.85880 |0.92434 |42.48574 |44.00000 |22.97306 |696.42136 |2612.0

|44.48629 |45.33333 |22.51276 |685.33845 |2618.0 |0.79046 |0.57851

|0.01882 |0.01240 |0.08770

|1 |246444 |141 |18 |Indiana |South Bend |Roseland |City |46616 |574

|41.701441 |-86.266614 |1560828.0 |100363 |2633 |1349 |1284 |804.87924

|848.0 |253.46747 |312.58622 |513.0 |0.97410 |0.93227 |0.69920 |0.69920

|0.55179 |0.41235 |0.39044 |0.27888 |542 |502 |41931.92593 |35186.0

|31639.50203 |838.74664 |1127.0 |50670.10337 |43023.0 |34715.57548

|375.28798 |554.0 |864.41390 |784.0 |482.27020 |316.88320 |356.0

|351.98293 |336.0 |125.40457 |229.0 |189.60606 |0.02222 |0.02222

|0.04274 |0.60855 |0.42174 |0.70823 |0.58120 |0.90487 |0.86947 |0.94187

|34.84728 |32.00000 |20.37452 |323.90204 |1349.0 |36.48391 |37.58333

|23.43353 |267.23367 |1284.0 |0.52483 |0.34886 |0.01426 |0.01426

|0.09030

|2 |245683 |63 |18 |Indiana |Danville |Danville |City |46122 |317

|39.792202 |-86.515246 |69561595.0 |284193 |6881 |3643 |3238 |742.77365

|703.0 |323.39011 |291.85520 |378.0 |0.95238 |0.88624 |0.79630 |0.66667

|0.39153 |0.39153 |0.28307 |0.15873 |459 |378 |84942.68317 |74964.0

|56811.62186 |1155.20980 |2488.0 |95262.51431 |85395.0 |49292.67664

|709.74925 |1889.0 |1506.06758 |1361.0 |731.89394 |699.41354 |1491.0

|556.45986 |532.0 |184.42175 |538.0 |323.35354 |0.00000 |0.00000

|0.09512 |0.73484 |1.00000 |0.46332 |0.28704 |0.94288 |0.94616 |0.93952

|39.38154 |40.83333 |22.89769 |888.29730 |3643.0 |42.15810 |42.83333

|23.94119 |707.01963 |3238.0 |0.85331 |0.64745 |0.02830 |0.01607

|0.10657

|3 |279653 |127 |72 |Puerto Rico |San Juan |Guaynabo |Urban |927 |787

|18.396103 |-66.104169 |1105793.0 |0 |2700 |1141 |1559 |803.42018 |782.0

|297.39258 |259.30316 |368.0 |0.94693 |0.87151 |0.69832 |0.61732

|0.51397 |0.46927 |0.35754 |0.32961 |438 |358 |48733.67116 |37845.0

|45100.54010 |928.32193 |1267.0 |56401.68133 |44399.0 |41082.90515

|490.18479 |729.0 |1175.28642 |1101.0 |428.98751 |261.28471 |437.0

|288.04047 |247.0 |185.55887 |392.0 |314.90566 |0.01086 |0.01086

|0.01086 |0.52714 |0.53057 |0.82530 |0.73727 |0.91500 |0.90755 |0.92043

|48.64749 |48.91667 |23.05968 |274.98956 |1141.0 |47.77526 |50.58333

|24.32015 |362.20193 |1559.0 |0.65037 |0.47257 |0.02021 |0.02021

|0.10106

|4 |247218 |161 |20 |Kansas |Manhattan |Manhattan City |City |66502 |785

|39.195573 |-96.569366 |2554403.0 |0 |5637 |2586 |3051 |938.56493 |881.0

|392.44096 |1005.42886 |1704.0 |0.99286 |0.98247 |0.91688 |0.84740

|0.78247 |0.60974 |0.55455 |0.44416 |1725 |1540 |31834.15466 |22497.0

|34046.50907 |1548.67477 |1983.0 |54053.42396 |50272.0 |39609.12605

|244.08903 |395.0 |1192.58759 |1125.0 |327.49674 |76.61052 |134.0

|443.68855 |444.0 |76.12674 |124.0 |79.55556 |0.05426 |0.05426 |0.05426

|0.51938 |0.18332 |0.65545 |0.74967 |1.00000 |1.00000 |1.00000 |26.07533

|22.41667 |11.84399 |1296.89877 |2586.0 |24.17693 |21.58333 |11.10484

|1854.48652 |3051.0 |0.13046 |0.12356 |0.00000 |0.00000 |0.03109

|===

----

+\*In[131]:\*+

[source, ipython3]

----

train\_df.shape

----

+\*Out[131]:\*+

----(27019, 76)----

+\*In[132]:\*+

[source, ipython3]

----

test\_df.shape

----

+\*Out[132]:\*+

----(11603, 77)----

+\*In[133]:\*+

[source, ipython3]

----

test\_df.columns

----

+\*Out[133]:\*+

----Index(['UID', 'COUNTYID', 'STATEID', 'state', 'state\_ab', 'city', 'place',

'type', 'zip\_code', 'area\_code', 'lat', 'lng', 'ALand', 'AWater', 'pop',

'male\_pop', 'female\_pop', 'rent\_mean', 'rent\_median', 'rent\_stdev',

'rent\_sample\_weight', 'rent\_samples', 'rent\_gt\_10', 'rent\_gt\_15',

'rent\_gt\_20', 'rent\_gt\_25', 'rent\_gt\_30', 'rent\_gt\_35', 'rent\_gt\_40',

'rent\_gt\_50', 'universe\_samples', 'used\_samples', 'hi\_mean',

'hi\_median', 'hi\_stdev', 'hi\_sample\_weight', 'hi\_samples',

'family\_mean', 'family\_median', 'family\_stdev', 'family\_sample\_weight',

'family\_samples', 'hc\_mortgage\_mean', 'hc\_mortgage\_median',

'hc\_mortgage\_stdev', 'hc\_mortgage\_sample\_weight', 'hc\_mortgage\_samples',

'hc\_mean', 'hc\_median', 'hc\_stdev', 'hc\_samples', 'hc\_sample\_weight',

'home\_equity\_second\_mortgage', 'second\_mortgage', 'home\_equity', 'debt',

'second\_mortgage\_cdf', 'home\_equity\_cdf', 'debt\_cdf', 'hs\_degree',

'hs\_degree\_male', 'hs\_degree\_female', 'male\_age\_mean',

'male\_age\_median', 'male\_age\_stdev', 'male\_age\_sample\_weight',

'male\_age\_samples', 'female\_age\_mean', 'female\_age\_median',

'female\_age\_stdev', 'female\_age\_sample\_weight', 'female\_age\_samples',

'pct\_own', 'married', 'married\_snp', 'separated', 'divorced'],

dtype='object')----

+\*In[134]:\*+

[source, ipython3]

----

test\_df.drop(['state\_ab'],axis = 1, inplace = True)

----

+\*In[135]:\*+

[source, ipython3]

----

test\_df.shape

----

+\*Out[135]:\*+

----(11603, 76)----

+\*In[136]:\*+

[source, ipython3]

----

num\_2\_cat = ['UID','COUNTYID', 'STATEID', 'zip\_code', 'area\_code', 'lat', 'lng']

----

+\*In[137]:\*+

[source, ipython3]

----

train\_df[num\_2\_cat].dtypes

----

+\*Out[137]:\*+

----UID int64

COUNTYID int64

STATEID int64

zip\_code int64

area\_code int64

lat float64

lng float64

dtype: object----

+\*In[138]:\*+

[source, ipython3]

----

for col in num\_2\_cat:

train\_df[col] = train\_df[col].astype('category')

test\_df[col] = test\_df[col].astype('category')

----

+\*In[139]:\*+

[source, ipython3]

----

train\_df[num\_2\_cat].dtypes

----

+\*Out[139]:\*+

----UID category

COUNTYID category

STATEID category

zip\_code category

area\_code category

lat category

lng category

dtype: object----

+\*In[140]:\*+

[source, ipython3]

----

test\_df[num\_2\_cat].dtypes

----

+\*Out[140]:\*+

----UID category

COUNTYID category

STATEID category

zip\_code category

area\_code category

lat category

lng category

dtype: object----

+\*In[141]:\*+

[source, ipython3]

----

train\_df.dtypes

----

+\*Out[141]:\*+

----UID category

COUNTYID category

STATEID category

state object

city object

place object

type object

zip\_code category

area\_code category

lat category

lng category

ALand float64

AWater int64

pop int64

male\_pop int64

female\_pop int64

rent\_mean float64

rent\_median float64

rent\_stdev float64

rent\_sample\_weight float64

rent\_samples float64

rent\_gt\_10 float64

rent\_gt\_15 float64

rent\_gt\_20 float64

rent\_gt\_25 float64

rent\_gt\_30 float64

rent\_gt\_35 float64

rent\_gt\_40 float64

rent\_gt\_50 float64

universe\_samples int64

used\_samples int64

hi\_mean float64

hi\_median float64

hi\_stdev float64

hi\_sample\_weight float64

hi\_samples float64

family\_mean float64

family\_median float64

family\_stdev float64

family\_sample\_weight float64

family\_samples float64

hc\_mortgage\_mean float64

hc\_mortgage\_median float64

hc\_mortgage\_stdev float64

hc\_mortgage\_sample\_weight float64

hc\_mortgage\_samples float64

hc\_mean float64

hc\_median float64

hc\_stdev float64

hc\_samples float64

hc\_sample\_weight float64

home\_equity\_second\_mortgage float64

second\_mortgage float64

home\_equity float64

debt float64

second\_mortgage\_cdf float64

home\_equity\_cdf float64

debt\_cdf float64

hs\_degree float64

hs\_degree\_male float64

hs\_degree\_female float64

male\_age\_mean float64

male\_age\_median float64

male\_age\_stdev float64

male\_age\_sample\_weight float64

male\_age\_samples float64

female\_age\_mean float64

female\_age\_median float64

female\_age\_stdev float64

female\_age\_sample\_weight float64

female\_age\_samples float64

pct\_own float64

married float64

married\_snp float64

separated float64

divorced float64

dtype: object----

+\*In[142]:\*+

[source, ipython3]

----

obj\_2\_cat = ['state', 'city', 'place', 'type']

----

+\*In[143]:\*+

[source, ipython3]

----

for col in obj\_2\_cat:

train\_df[col] = train\_df[col].astype('category')

test\_df[col] = test\_df[col].astype('category')

----

+\*In[144]:\*+

[source, ipython3]

----

train\_df.info()

----

+\*Out[144]:\*+

----

<class 'pandas.core.frame.DataFrame'>

Int64Index: 27019 entries, 0 to 27320

Data columns (total 76 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 UID 27019 non-null category

1 COUNTYID 27019 non-null category

2 STATEID 27019 non-null category

3 state 27019 non-null category

4 city 27019 non-null category

5 place 27019 non-null category

6 type 27019 non-null category

7 zip\_code 27019 non-null category

8 area\_code 27019 non-null category

9 lat 27019 non-null category

10 lng 27019 non-null category

11 ALand 27019 non-null float64

12 AWater 27019 non-null int64

13 pop 27019 non-null int64

14 male\_pop 27019 non-null int64

15 female\_pop 27019 non-null int64

16 rent\_mean 27019 non-null float64

17 rent\_median 27019 non-null float64

18 rent\_stdev 27019 non-null float64

19 rent\_sample\_weight 27019 non-null float64

20 rent\_samples 27019 non-null float64

21 rent\_gt\_10 27019 non-null float64

22 rent\_gt\_15 27019 non-null float64

23 rent\_gt\_20 27019 non-null float64

24 rent\_gt\_25 27019 non-null float64

25 rent\_gt\_30 27019 non-null float64

26 rent\_gt\_35 27019 non-null float64

27 rent\_gt\_40 27019 non-null float64

28 rent\_gt\_50 27019 non-null float64

29 universe\_samples 27019 non-null int64

30 used\_samples 27019 non-null int64

31 hi\_mean 27019 non-null float64

32 hi\_median 27019 non-null float64

33 hi\_stdev 27019 non-null float64

34 hi\_sample\_weight 27019 non-null float64

35 hi\_samples 27019 non-null float64

36 family\_mean 27019 non-null float64

37 family\_median 27019 non-null float64

38 family\_stdev 27019 non-null float64

39 family\_sample\_weight 27019 non-null float64

40 family\_samples 27019 non-null float64

41 hc\_mortgage\_mean 27019 non-null float64

42 hc\_mortgage\_median 27019 non-null float64

43 hc\_mortgage\_stdev 27019 non-null float64

44 hc\_mortgage\_sample\_weight 27019 non-null float64

45 hc\_mortgage\_samples 27019 non-null float64

46 hc\_mean 27019 non-null float64

47 hc\_median 27019 non-null float64

48 hc\_stdev 27019 non-null float64

49 hc\_samples 27019 non-null float64

50 hc\_sample\_weight 27019 non-null float64

51 home\_equity\_second\_mortgage 27019 non-null float64

52 second\_mortgage 27019 non-null float64

53 home\_equity 27019 non-null float64

54 debt 27019 non-null float64

55 second\_mortgage\_cdf 27019 non-null float64

56 home\_equity\_cdf 27019 non-null float64

57 debt\_cdf 27019 non-null float64

58 hs\_degree 27019 non-null float64

59 hs\_degree\_male 27019 non-null float64

60 hs\_degree\_female 27019 non-null float64

61 male\_age\_mean 27019 non-null float64

62 male\_age\_median 27019 non-null float64

63 male\_age\_stdev 27019 non-null float64

64 male\_age\_sample\_weight 27019 non-null float64

65 male\_age\_samples 27019 non-null float64

66 female\_age\_mean 27019 non-null float64

67 female\_age\_median 27019 non-null float64

68 female\_age\_stdev 27019 non-null float64

69 female\_age\_sample\_weight 27019 non-null float64

70 female\_age\_samples 27019 non-null float64

71 pct\_own 27019 non-null float64

72 married 27019 non-null float64

73 married\_snp 27019 non-null float64

74 separated 27019 non-null float64

75 divorced 27019 non-null float64

dtypes: category(11), float64(59), int64(6)

memory usage: 21.2 MB

----

+\*In[145]:\*+

[source, ipython3]

----

kwargs = dict(hist\_kws={'alpha':1}, kde\_kws={'linewidth':3})

plt.figure(figsize=(10,10), dpi= 80)

sns.distplot(train\_df.hc\_mortgage\_mean, color="green", label="hc\_mortgage\_mean", \*\*kwargs)

plt.legend();

----

+\*Out[145]:\*+

----

![png](output\_174\_0.png)

----

Target Variable "hc\_mortgage\_mean" has a Positive Skew.

+\*In[146]:\*+

[source, ipython3]

----

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, mean\_absolute\_error, r2\_score, SCORERS

----

+\*In[147]:\*+

[source, ipython3]

----

lin\_reg = LinearRegression()

----

+\*In[148]:\*+

[source, ipython3]

----

cols\_2\_drop = ['UID', 'state', 'COUNTYID', 'STATEID', 'city', 'place', 'type', 'zip\_code', 'area\_code', 'lat', 'lng']

----

+\*In[149]:\*+

[source, ipython3]

----

train\_df.drop(cols\_2\_drop, axis=1, inplace=True)

test\_df.drop(cols\_2\_drop, axis=1, inplace=True)

----

+\*In[150]:\*+

[source, ipython3]

----

train\_y = train\_df['hc\_mortgage\_mean']

test\_y = test\_df['hc\_mortgage\_mean']

----

+\*In[151]:\*+

[source, ipython3]

----

train\_X = train\_df.drop(columns=['hc\_mortgage\_mean'])

test\_X = test\_df.drop(columns=['hc\_mortgage\_mean'])

----

+\*In[152]:\*+

[source, ipython3]

----

print(train\_X.shape, train\_y.shape, test\_X.shape,test\_y.shape)

----

+\*Out[152]:\*+

----

(27019, 64) (27019,) (11603, 64) (11603,)

----

+\*In[153]:\*+

[source, ipython3]

----

lin\_reg.fit(train\_X,train\_y)

----

+\*Out[153]:\*+

----LinearRegression()----

+\*In[154]:\*+

[source, ipython3]

----

pred\_y = lin\_reg.predict(test\_X)

----

+\*In[155]:\*+

[source, ipython3]

----

pred\_y

----

+\*Out[155]:\*+

----array([1140.29712564, 1512.54387678, 1221.32683601, ..., 1853.70247831,

1158.76921634, 1387.54097806])----

+\*In[156]:\*+

[source, ipython3]

----

# model evaluation for testing set

mae = mean\_absolute\_error(test\_y, pred\_y)

mse = mean\_squared\_error(test\_y, pred\_y)

r2 = r2\_score(test\_y, pred\_y)

adj\_rsqrd = 1 - (1-r2)\*(len(test\_y) - 1) / (len(test\_y) - (test\_X.shape[1] - 1) - 1)

print('Mean Absolute Error is {}'.format(round(mae, 3)))

print('Mean Squared Error is {}'.format(round(mse, 3)))

print('Root Mean Squared Error is {}'.format(round(mse\*\*(0.5), 3)))

print('R-Square score is {}'.format(round(r2, 3)))

print('Adjusted R-Square score is {}'.format(round(adj\_rsqrd, 3)))

----

+\*Out[156]:\*+

----

Mean Absolute Error is 44.429

Mean Squared Error is 5015.094

Root Mean Squared Error is 70.817

R-Square score is 0.987

Adjusted R-Square score is 0.987

----

== The accuracy levels and R square at a Nation level.

Since R-square is 98.7%. model is accurate.State level regression

analysis is not required

+\*In[157]:\*+

[source, ipython3]

----

correlated\_features = set()

correlation\_matrix = train\_df.drop('hc\_mortgage\_mean', axis=1).corr()

----

+\*In[158]:\*+

[source, ipython3]

----

correlation\_matrix

----

+\*Out[158]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |ALand |AWater |pop |male\_pop |female\_pop |rent\_mean |rent\_median

|rent\_stdev |rent\_sample\_weight |rent\_samples |rent\_gt\_10 |rent\_gt\_15

|rent\_gt\_20 |rent\_gt\_25 |rent\_gt\_30 |rent\_gt\_35 |rent\_gt\_40 |rent\_gt\_50

|universe\_samples |used\_samples |hi\_mean |hi\_median |hi\_stdev

|hi\_sample\_weight |hi\_samples |family\_mean |family\_median |family\_stdev

|family\_sample\_weight |family\_samples |hc\_mortgage\_median

|hc\_mortgage\_stdev |hc\_mortgage\_sample\_weight |hc\_mortgage\_samples

|hc\_mean |hc\_median |hc\_stdev |hc\_samples |hc\_sample\_weight

|home\_equity\_second\_mortgage |second\_mortgage |home\_equity |debt

|second\_mortgage\_cdf |home\_equity\_cdf |debt\_cdf |hs\_degree

|hs\_degree\_male |hs\_degree\_female |male\_age\_mean |male\_age\_median

|male\_age\_stdev |male\_age\_sample\_weight |male\_age\_samples

|female\_age\_mean |female\_age\_median |female\_age\_stdev

|female\_age\_sample\_weight |female\_age\_samples |pct\_own |married

|married\_snp |separated |divorced

|ALand |1.000000 |0.413671 |-0.036081 |-0.025621 |-0.045048 |-0.067422

|-0.065765 |-0.033652 |-0.046308 |-0.067790 |-0.098697 |-0.097052

|-0.084723 |-0.076661 |-0.062345 |-0.056424 |-0.049869 |-0.049702

|-0.059325 |-0.067041 |-0.028725 |-0.030005 |-0.018642 |-0.028512

|-0.042085 |-0.028141 |-0.029566 |-0.018214 |-0.013156 |-0.028408

|-0.057968 |-0.015473 |-0.010268 |-0.038050 |-0.056751 |-0.058165

|-0.006400 |0.051854 |0.067830 |-0.042531 |-0.044410 |-0.080306

|-0.117135 |0.047465 |0.087734 |0.106694 |-0.002706 |-0.006499 |0.003046

|0.043723 |0.050695 |0.034123 |-0.024953 |-0.025676 |0.017361 |0.030907

|0.026841 |-0.044301 |-0.045249 |0.050627 |0.028043 |0.003994 |-0.005773

|0.023420

|AWater |0.413671 |1.000000 |-0.016652 |-0.013612 |-0.019027 |-0.009582

|-0.009393 |0.002468 |-0.017056 |-0.020252 |-0.029499 |-0.028691

|-0.032941 |-0.028790 |-0.023977 |-0.023381 |-0.019539 |-0.017993

|-0.018840 |-0.020311 |-0.002232 |-0.002207 |0.000784 |-0.018335

|-0.019306 |-0.002117 |-0.002489 |0.001692 |-0.014358 |-0.015563

|-0.010897 |0.005096 |-0.014672 |-0.016985 |-0.010590 |-0.010922

|0.004735 |0.008908 |0.010834 |-0.014830 |-0.014817 |-0.024547

|-0.041554 |0.016077 |0.025872 |0.035456 |0.006169 |0.005950 |0.005889

|0.007767 |0.010537 |-0.001746 |-0.012675 |-0.013631 |-0.005620

|-0.001258 |-0.006398 |-0.019013 |-0.019089 |0.010259 |-0.000761

|0.005268 |-0.001177 |0.007678

|pop |-0.036081 |-0.016652 |1.000000 |0.980916 |0.981439 |0.162924

|0.156708 |0.118136 |0.247401 |0.406672 |0.064624 |0.060280 |0.021006

|-0.001991 |-0.013999 |-0.022439 |-0.029887 |-0.046119 |0.424884

|0.414677 |0.173868 |0.180338 |0.131390 |0.710296 |0.897350 |0.134395

|0.130222 |0.111936 |0.784932 |0.930581 |0.110547 |0.084803 |0.645214

|0.774241 |0.057414 |0.056097 |0.055962 |0.458678 |0.386986 |0.084978

|0.084867 |0.105696 |0.247888 |-0.152876 |-0.132071 |-0.254184 |0.047779

|0.056287 |0.037697 |-0.192439 |-0.155431 |-0.027974 |0.912969 |0.980439

|-0.196112 |-0.163807 |-0.026764 |0.923600 |0.980851 |0.094828 |0.171907

|-0.039066 |-0.084397 |-0.164737

|male\_pop |-0.025621 |-0.013612 |0.980916 |1.000000 |0.925422 |0.160013

|0.154592 |0.108759 |0.224763 |0.383447 |0.051021 |0.045027 |0.006744

|-0.015356 |-0.026806 |-0.033588 |-0.040599 |-0.056091 |0.399228

|0.388983 |0.175732 |0.182920 |0.128455 |0.673333 |0.857665 |0.133820

|0.129612 |0.108743 |0.757128 |0.898409 |0.107893 |0.083352 |0.620826

|0.747064 |0.048706 |0.046975 |0.052662 |0.441285 |0.375081 |0.081578

|0.081213 |0.100935 |0.236359 |-0.146775 |-0.123673 |-0.243430 |0.030239

|0.037469 |0.028657 |-0.205537 |-0.166188 |-0.075688 |0.942592 |0.999556

|-0.199602 |-0.161795 |-0.063224 |0.871078 |0.925742 |0.093538 |0.140432

|-0.004272 |-0.076845 |-0.150692

|female\_pop |-0.045048 |-0.019027 |0.981439 |0.925422 |1.000000

|0.159705 |0.152937 |0.122969 |0.260481 |0.414376 |0.075626 |0.073071

|0.034287 |0.011266 |-0.000844 |-0.010603 |-0.018204 |-0.034559

|0.434305 |0.424517 |0.165529 |0.171050 |0.129374 |0.720200 |0.902945

|0.129938 |0.125956 |0.110901 |0.783012 |0.927525 |0.109032 |0.083065

|0.645148 |0.772103 |0.063856 |0.062996 |0.057125 |0.458686 |0.384260

|0.085155 |0.085298 |0.106441 |0.249992 |-0.153178 |-0.135416 |-0.255289

|0.063292 |0.072744 |0.045205 |-0.172326 |-0.139011 |0.020133 |0.849621

|0.924931 |-0.185339 |-0.159668 |0.010198 |0.940876 |0.998539 |0.092555

|0.196526 |-0.071924 |-0.088690 |-0.172432

|rent\_mean |-0.067422 |-0.009582 |0.162924 |0.160013 |0.159705 |1.000000

|0.976525 |0.655634 |-0.390870 |-0.016698 |0.101404 |0.113636 |0.054619

|0.011170 |-0.003730 |0.002104 |-0.002797 |-0.004467 |-0.038948

|-0.011577 |0.755901 |0.752720 |0.646668 |-0.277481 |0.105899 |0.702878

|0.699791 |0.566237 |-0.218601 |0.165236 |0.751698 |0.566718 |-0.150295

|0.290788 |0.596115 |0.578463 |0.448193 |-0.143196 |-0.298275 |0.129245

|0.153188 |0.413819 |0.437968 |-0.185721 |-0.428932 |-0.459536 |0.363260

|0.372288 |0.328594 |0.043793 |0.093017 |-0.112621 |0.131686 |0.159888

|0.003875 |0.045239 |-0.172810 |0.128762 |0.159953 |0.140698 |0.256218

|-0.107838 |-0.188070 |-0.377479

|rent\_median |-0.065765 |-0.009393 |0.156708 |0.154592 |0.152937

|0.976525 |1.000000 |0.569635 |-0.385527 |-0.021336 |0.099573 |0.114488

|0.057516 |0.013348 |-0.002426 |0.002495 |-0.003136 |-0.006820

|-0.042554 |-0.016339 |0.719634 |0.722718 |0.600498 |-0.270814 |0.094178

|0.660093 |0.660342 |0.520912 |-0.201366 |0.159185 |0.715387 |0.527718

|-0.141133 |0.277301 |0.563220 |0.548361 |0.416664 |-0.145708 |-0.289261

|0.129134 |0.151884 |0.391680 |0.425381 |-0.180258 |-0.405522 |-0.446454

|0.333124 |0.341960 |0.300575 |0.024865 |0.074968 |-0.115311 |0.125258

|0.154501 |-0.013918 |0.029037 |-0.175715 |0.120780 |0.153199 |0.132248

|0.242848 |-0.095903 |-0.174123 |-0.361252

|rent\_stdev |-0.033652 |0.002468 |0.118136 |0.108759 |0.122969 |0.655634

|0.569635 |1.000000 |-0.180206 |0.066356 |-0.014277 |0.040207 |0.040415

|0.034501 |0.020990 |0.024603 |0.027684 |0.040993 |0.052482 |0.068045

|0.549108 |0.514547 |0.577372 |-0.162431 |0.118084 |0.560007 |0.547461

|0.529284 |-0.197155 |0.099145 |0.632512 |0.526635 |-0.171261 |0.177666

|0.508916 |0.488165 |0.441140 |-0.080714 |-0.216093 |0.060740 |0.081253

|0.308594 |0.272806 |-0.105882 |-0.318893 |-0.283658 |0.270215 |0.283391

|0.238460 |0.110195 |0.116358 |-0.012700 |0.095979 |0.108416 |0.101978

|0.099196 |-0.025764 |0.108126 |0.122759 |0.049597 |0.130537 |-0.070423

|-0.136935 |-0.270542

|rent\_sample\_weight |-0.046308 |-0.017056 |0.247401 |0.224763 |0.260481

|-0.390870 |-0.385527 |-0.180206 |1.000000 |0.802171 |0.054400 |0.090267

|0.126070 |0.141612 |0.122964 |0.104341 |0.098875 |0.090872 |0.804015

|0.794939 |-0.477683 |-0.487644 |-0.370814 |0.664129 |0.371471

|-0.424231 |-0.423041 |-0.293984 |0.398751 |0.109174 |-0.260194

|-0.265882 |0.041720 |-0.153277 |-0.211770 |-0.197356 |-0.183569

|-0.067776 |-0.003988 |0.002982 |-0.007619 |-0.157704 |-0.074597

|0.109338 |0.183686 |0.082890 |-0.256984 |-0.236590 |-0.263558

|-0.270018 |-0.342550 |-0.133379 |0.237580 |0.224624 |-0.207395

|-0.303488 |-0.034040 |0.287024 |0.260621 |-0.613327 |-0.440904

|0.216546 |0.204450 |0.213381

|rent\_samples |-0.067790 |-0.020252 |0.406672 |0.383447 |0.414376

|-0.016698 |-0.021336 |0.066356 |0.802171 |1.000000 |0.104060 |0.146575

|0.157698 |0.150303 |0.128980 |0.111501 |0.102544 |0.094939 |0.993165

|0.996060 |-0.222960 |-0.246676 |-0.121752 |0.637302 |0.527963

|-0.175231 |-0.186920 |-0.057653 |0.375106 |0.233305 |0.061970

|-0.001861 |-0.040622 |-0.037641 |0.030317 |0.036377 |0.023233

|-0.156547 |-0.154598 |0.081712 |0.081188 |-0.005212 |0.131981 |0.039186

|0.029133 |-0.138391 |-0.118516 |-0.092197 |-0.140690 |-0.296783

|-0.362014 |-0.269415 |0.373061 |0.383353 |-0.253734 |-0.350294

|-0.177937 |0.415831 |0.414716 |-0.686417 |-0.393193 |0.213205 |0.141390

|0.052663

|rent\_gt\_10 |-0.098697 |-0.029499 |0.064624 |0.051021 |0.075626

|0.101404 |0.099573 |-0.014277 |0.054400 |0.104060 |1.000000 |0.616881

|0.453908 |0.368267 |0.313162 |0.280104 |0.256106 |0.217351 |0.097047

|0.103616 |-0.099055 |-0.094160 |-0.103651 |0.089561 |0.051091

|-0.100939 |-0.106370 |-0.076538 |0.079352 |0.035376 |0.008888

|-0.029236 |0.006068 |0.031835 |-0.010413 |-0.008710 |-0.042615

|-0.094810 |-0.095496 |0.072386 |0.078542 |0.070639 |0.149861 |-0.062505

|-0.074909 |-0.149331 |-0.051368 |-0.044790 |-0.056946 |-0.097346

|-0.105726 |-0.060407 |0.056100 |0.050008 |-0.070305 |-0.082919

|-0.048293 |0.080626 |0.075889 |-0.098433 |-0.093731 |0.037412 |0.041863

|-0.010049

|rent\_gt\_15 |-0.097052 |-0.028691 |0.060280 |0.045027 |0.073071

|0.113636 |0.114488 |0.040207 |0.090267 |0.146575 |0.616881 |1.000000

|0.744549 |0.610684 |0.524014 |0.465119 |0.424119 |0.364631 |0.139181

|0.144656 |-0.151717 |-0.151815 |-0.132125 |0.103202 |0.028208

|-0.151779 |-0.154037 |-0.107757 |0.086499 |0.004376 |0.030748

|-0.031588 |-0.052304 |-0.014924 |-0.002964 |0.000613 |-0.041870

|-0.144934 |-0.143755 |0.082194 |0.090659 |0.062408 |0.145918 |-0.051230

|-0.059755 |-0.152264 |-0.125355 |-0.115279 |-0.130872 |-0.132694

|-0.147678 |-0.083468 |0.061170 |0.044639 |-0.091147 |-0.113285

|-0.070668 |0.088634 |0.073455 |-0.180554 |-0.179243 |0.095322 |0.089480

|-0.005079

|rent\_gt\_20 |-0.084723 |-0.032941 |0.021006 |0.006744 |0.034287

|0.054619 |0.057516 |0.040415 |0.126070 |0.157698 |0.453908 |0.744549

|1.000000 |0.826560 |0.710902 |0.630039 |0.576479 |0.499112 |0.152704

|0.153777 |-0.239929 |-0.249370 |-0.181409 |0.109454 |-0.021501

|-0.234311 |-0.237036 |-0.154807 |0.087680 |-0.052591 |0.002191

|-0.058445 |-0.107276 |-0.099913 |-0.029155 |-0.024360 |-0.046491

|-0.156173 |-0.141499 |0.059369 |0.069729 |0.008607 |0.071959 |-0.011099

|0.002145 |-0.080238 |-0.217589 |-0.204976 |-0.219565 |-0.140507

|-0.163784 |-0.079638 |0.037116 |0.006635 |-0.089798 |-0.119396

|-0.058901 |0.063714 |0.034750 |-0.237966 |-0.251667 |0.151538 |0.129711

|0.020269

|rent\_gt\_25 |-0.076661 |-0.028790 |-0.001991 |-0.015356 |0.011266

|0.011170 |0.013348 |0.034501 |0.141612 |0.150303 |0.368267 |0.610684

|0.826560 |1.000000 |0.862830 |0.770179 |0.707255 |0.616151 |0.147621

|0.145012 |-0.289617 |-0.302585 |-0.210589 |0.107589 |-0.052434

|-0.282232 |-0.283321 |-0.185136 |0.087492 |-0.082005 |-0.021375

|-0.080582 |-0.132449 |-0.145507 |-0.050387 |-0.044528 |-0.057454

|-0.151306 |-0.128810 |0.036333 |0.047963 |-0.033336 |0.015669 |0.019903

|0.048483 |-0.025196 |-0.274682 |-0.259821 |-0.274354 |-0.145335

|-0.173021 |-0.072503 |0.023858 |-0.015494 |-0.092474 |-0.124816

|-0.053383 |0.049357 |0.010886 |-0.260684 |-0.289002 |0.179466 |0.153450

|0.023998

|rent\_gt\_30 |-0.062345 |-0.023977 |-0.013999 |-0.026806 |-0.000844

|-0.003730 |-0.002426 |0.020990 |0.122964 |0.128980 |0.313162 |0.524014

|0.710902 |0.862830 |1.000000 |0.895242 |0.820029 |0.714114 |0.127862

|0.122984 |-0.302042 |-0.314324 |-0.215522 |0.097869 |-0.069646

|-0.296304 |-0.295054 |-0.192933 |0.085450 |-0.094067 |-0.039909

|-0.088359 |-0.136743 |-0.164833 |-0.068875 |-0.063033 |-0.062268

|-0.132232 |-0.103667 |0.023422 |0.034594 |-0.061938 |-0.021081

|0.036160 |0.077623 |0.011923 |-0.290700 |-0.277469 |-0.288790

|-0.147319 |-0.173988 |-0.069339 |0.017909 |-0.026792 |-0.100590

|-0.131239 |-0.053060 |0.042263 |-0.001076 |-0.252927 |-0.290348

|0.182522 |0.156383 |0.025301

|rent\_gt\_35 |-0.056424 |-0.023381 |-0.022439 |-0.033588 |-0.010603

|0.002104 |0.002495 |0.024603 |0.104341 |0.111501 |0.280104 |0.465119

|0.630039 |0.770179 |0.895242 |1.000000 |0.915587 |0.798695 |0.110784

|0.105207 |-0.288393 |-0.299737 |-0.200399 |0.080274 |-0.081008

|-0.282767 |-0.279171 |-0.181607 |0.069781 |-0.102416 |-0.037621

|-0.083934 |-0.142182 |-0.172005 |-0.061477 |-0.055139 |-0.055362

|-0.123431 |-0.096212 |0.005665 |0.016197 |-0.072611 |-0.040281

|0.051551 |0.088144 |0.033015 |-0.284830 |-0.271853 |-0.282907

|-0.142846 |-0.168354 |-0.069966 |0.014255 |-0.034174 |-0.102545

|-0.131202 |-0.053321 |0.037121 |-0.010729 |-0.240134 |-0.288658

|0.174162 |0.146173 |0.019348

|rent\_gt\_40 |-0.049869 |-0.019539 |-0.029887 |-0.040599 |-0.018204

|-0.002797 |-0.003136 |0.027684 |0.098875 |0.102544 |0.256106 |0.424119

|0.576479 |0.707255 |0.820029 |0.915587 |1.000000 |0.871955 |0.102296

|0.095925 |-0.284943 |-0.295376 |-0.194744 |0.069298 |-0.088560

|-0.276746 |-0.272742 |-0.175851 |0.056497 |-0.110252 |-0.039732

|-0.082991 |-0.144599 |-0.177969 |-0.057354 |-0.051641 |-0.050551

|-0.119105 |-0.092358 |0.000213 |0.009993 |-0.075330 |-0.049929

|0.055545 |0.091189 |0.044054 |-0.275408 |-0.262218 |-0.274618

|-0.141201 |-0.166731 |-0.073007 |0.010454 |-0.041119 |-0.103614

|-0.132240 |-0.056169 |0.034196 |-0.018249 |-0.234997 |-0.289330

|0.164097 |0.137820 |0.014230

|rent\_gt\_50 |-0.049702 |-0.017993 |-0.046119 |-0.056091 |-0.034559

|-0.004467 |-0.006820 |0.040993 |0.090872 |0.094939 |0.217351 |0.364631

|0.499112 |0.616151 |0.714114 |0.798695 |0.871955 |1.000000 |0.094986

|0.087978 |-0.267504 |-0.276742 |-0.175684 |0.046404 |-0.096304

|-0.253700 |-0.247714 |-0.157452 |0.023014 |-0.124488 |-0.031879

|-0.072646 |-0.153330 |-0.182581 |-0.041567 |-0.037503 |-0.036779

|-0.119509 |-0.096747 |-0.005162 |0.000192 |-0.074420 |-0.059033

|0.063165 |0.089769 |0.052375 |-0.253331 |-0.239286 |-0.254049

|-0.135374 |-0.161571 |-0.074132 |-0.001272 |-0.056527 |-0.098938

|-0.128332 |-0.055382 |0.023776 |-0.034479 |-0.228085 |-0.293197

|0.149117 |0.125471 |0.009755

|universe\_samples |-0.059325 |-0.018840 |0.424884 |0.399228 |0.434305

|-0.038948 |-0.042554 |0.052482 |0.804015 |0.993165 |0.097047 |0.139181

|0.152704 |0.147621 |0.127862 |0.110784 |0.102296 |0.094986 |1.000000

|0.994518 |-0.242521 |-0.266912 |-0.134864 |0.664145 |0.544680

|-0.193977 |-0.205422 |-0.070175 |0.404443 |0.250155 |0.040369

|-0.012188 |-0.020701 |-0.033879 |0.007688 |0.014138 |0.013353

|-0.121878 |-0.113263 |0.070341 |0.069003 |-0.026763 |0.102497 |0.050125

|0.051103 |-0.108717 |-0.129694 |-0.104421 |-0.150843 |-0.288532

|-0.354316 |-0.237388 |0.385697 |0.398849 |-0.245186 |-0.341325

|-0.146247 |0.431320 |0.433142 |-0.680462 |-0.381922 |0.207102 |0.144319

|0.062983

|used\_samples |-0.067041 |-0.020311 |0.414677 |0.388983 |0.424517

|-0.011577 |-0.016339 |0.068045 |0.794939 |0.996060 |0.103616 |0.144656

|0.153777 |0.145012 |0.122984 |0.105207 |0.095925 |0.087978 |0.994518

|1.000000 |-0.215397 |-0.239544 |-0.114231 |0.638670 |0.533821

|-0.168981 |-0.181174 |-0.051454 |0.379400 |0.241263 |0.064905 |0.002804

|-0.034627 |-0.029196 |0.033161 |0.038914 |0.026808 |-0.153115

|-0.152480 |0.083914 |0.083164 |-0.000971 |0.137321 |0.033092 |0.022051

|-0.143072 |-0.111098 |-0.084933 |-0.134381 |-0.291465 |-0.355815

|-0.245749 |0.372792 |0.388629 |-0.248507 |-0.344168 |-0.153574

|0.419200 |0.423429 |-0.680498 |-0.375701 |0.204754 |0.137527 |0.053767

|hi\_mean |-0.028725 |-0.002232 |0.173868 |0.175732 |0.165529 |0.755901

|0.719634 |0.549108 |-0.477683 |-0.222960 |-0.099055 |-0.151717

|-0.239929 |-0.289617 |-0.302042 |-0.288393 |-0.284943 |-0.267504

|-0.242521 |-0.215397 |1.000000 |0.979702 |0.895607 |-0.349215 |0.156278

|0.961709 |0.949692 |0.794422 |-0.272464 |0.265611 |0.756335 |0.648276

|-0.001732 |0.463268 |0.677319 |0.651384 |0.518828 |0.035331 |-0.156633

|0.082510 |0.101047 |0.476613 |0.422640 |-0.227276 |-0.515186 |-0.428350

|0.583437 |0.577860 |0.551221 |0.203388 |0.286687 |0.059374 |0.117184

|0.175599 |0.131128 |0.214398 |-0.031619 |0.104465 |0.165483 |0.481080

|0.533542 |-0.296233 |-0.317903 |-0.396216

|hi\_median |-0.030005 |-0.002207 |0.180338 |0.182920 |0.171050 |0.752720

|0.722718 |0.514547 |-0.487644 |-0.246676 |-0.094160 |-0.151815

|-0.249370 |-0.302585 |-0.314324 |-0.299737 |-0.295376 |-0.276742

|-0.266912 |-0.239544 |0.979702 |1.000000 |0.808238 |-0.353781 |0.146834

|0.923569 |0.935694 |0.692930 |-0.254407 |0.275657 |0.718150 |0.578333

|0.027529 |0.477199 |0.639566 |0.619180 |0.462524 |0.030784 |-0.152378

|0.091840 |0.108533 |0.468386 |0.432203 |-0.233457 |-0.504066 |-0.441210

|0.562464 |0.555444 |0.532398 |0.169760 |0.261416 |0.051456 |0.122703

|0.182912 |0.098061 |0.189832 |-0.040775 |0.107838 |0.171136 |0.494538

|0.536703 |-0.296818 |-0.312044 |-0.391262

|hi\_stdev |-0.018642 |0.000784 |0.131390 |0.128455 |0.129374 |0.646668

|0.600498 |0.577372 |-0.370814 |-0.121752 |-0.103651 |-0.132125

|-0.181409 |-0.210589 |-0.215522 |-0.200399 |-0.194744 |-0.175684

|-0.134864 |-0.114231 |0.895607 |0.808238 |1.000000 |-0.277588 |0.164149

|0.917030 |0.861275 |0.945962 |-0.279600 |0.198516 |0.751027 |0.731735

|-0.079634 |0.350089 |0.682836 |0.647638 |0.593203 |0.049629 |-0.137736

|0.039888 |0.060132 |0.420615 |0.325396 |-0.173092 |-0.463083 |-0.319317

|0.553318 |0.551734 |0.521099 |0.264896 |0.312344 |0.090284 |0.082886

|0.127649 |0.203050 |0.248528 |0.022221 |0.083581 |0.128690 |0.377004

|0.442274 |-0.253932 |-0.284381 |-0.348822

|hi\_sample\_weight |-0.028512 |-0.018335 |0.710296 |0.673333 |0.720200

|-0.277481 |-0.270814 |-0.162431 |0.664129 |0.637302 |0.089561 |0.103202

|0.109454 |0.107589 |0.097869 |0.080274 |0.069298 |0.046404 |0.664145

|0.638670 |-0.349215 |-0.353781 |-0.277588 |1.000000 |0.826679

|-0.321883 |-0.326976 |-0.221204 |0.860847 |0.654357 |-0.287855

|-0.214002 |0.602444 |0.388001 |-0.259025 |-0.248109 |-0.171369

|0.505851 |0.535909 |0.000271 |-0.011950 |-0.139094 |-0.051924

|-0.004663 |0.123886 |0.072597 |-0.114904 |-0.110718 |-0.114302

|-0.065206 |-0.108775 |0.041628 |0.616392 |0.673026 |-0.023754

|-0.075653 |0.080794 |0.674606 |0.720637 |-0.177136 |-0.092904 |0.054834

|0.077240 |0.169202

|hi\_samples |-0.042085 |-0.019306 |0.897350 |0.857665 |0.902945

|0.105899 |0.094178 |0.118084 |0.371471 |0.527963 |0.051091 |0.028208

|-0.021501 |-0.052434 |-0.069646 |-0.081008 |-0.088560 |-0.096304

|0.544680 |0.533821 |0.156278 |0.146834 |0.164149 |0.826679 |1.000000

|0.168046 |0.155475 |0.174621 |0.739903 |0.895762 |0.082918 |0.104652

|0.669830 |0.756384 |0.074946 |0.071382 |0.086831 |0.552013 |0.464226

|0.059491 |0.055499 |0.110933 |0.193775 |-0.136666 |-0.147467 |-0.183770

|0.202019 |0.203347 |0.185177 |0.010053 |0.014505 |0.053006 |0.754327

|0.857410 |0.012513 |0.008989 |0.046180 |0.813252 |0.903620 |0.079529

|0.184851 |-0.113021 |-0.091849 |-0.044028

|family\_mean |-0.028141 |-0.002117 |0.134395 |0.133820 |0.129938

|0.702878 |0.660093 |0.560007 |-0.424231 |-0.175231 |-0.100939

|-0.151779 |-0.234311 |-0.282232 |-0.296304 |-0.282767 |-0.276746

|-0.253700 |-0.193977 |-0.168981 |0.961709 |0.923569 |0.917030

|-0.321883 |0.168046 |1.000000 |0.979933 |0.855575 |-0.331517 |0.211392

|0.747320 |0.671703 |-0.021858 |0.425687 |0.689790 |0.661363 |0.549511

|0.049525 |-0.146074 |0.058327 |0.077101 |0.465874 |0.381264 |-0.203528

|-0.508388 |-0.380650 |0.638384 |0.629432 |0.607017 |0.261157 |0.320837

|0.050293 |0.090582 |0.133879 |0.196665 |0.254205 |-0.025327 |0.085941

|0.130209 |0.450855 |0.483535 |-0.318561 |-0.324546 |-0.359034

|family\_median |-0.029566 |-0.002489 |0.130222 |0.129612 |0.125956

|0.699791 |0.660342 |0.547461 |-0.423041 |-0.186920 |-0.106370

|-0.154037 |-0.237036 |-0.283321 |-0.295054 |-0.279171 |-0.272742

|-0.247714 |-0.205422 |-0.181174 |0.949692 |0.935694 |0.861275

|-0.326976 |0.155475 |0.979933 |1.000000 |0.770106 |-0.327557 |0.206297

|0.733640 |0.620760 |-0.012113 |0.420533 |0.675614 |0.652717 |0.520649

|0.046621 |-0.143901 |0.055799 |0.073732 |0.452519 |0.372158 |-0.197199

|-0.490767 |-0.374106 |0.613345 |0.603621 |0.584153 |0.238433 |0.302184

|0.050041 |0.086878 |0.129669 |0.174890 |0.235868 |-0.026282 |0.082225

|0.126238 |0.452139 |0.476421 |-0.314633 |-0.315453 |-0.352417

|family\_stdev |-0.018214 |0.001692 |0.111936 |0.108743 |0.110901

|0.566237 |0.520912 |0.529284 |-0.293984 |-0.057653 |-0.076538

|-0.107757 |-0.154807 |-0.185136 |-0.192933 |-0.181607 |-0.175851

|-0.157452 |-0.070175 |-0.051454 |0.794422 |0.692930 |0.945962

|-0.221204 |0.174621 |0.855575 |0.770106 |1.000000 |-0.272769 |0.165911

|0.682403 |0.712575 |-0.085189 |0.307165 |0.628698 |0.591165 |0.570258

|0.040727 |-0.131411 |0.037690 |0.057847 |0.392093 |0.301182 |-0.160879

|-0.438774 |-0.290308 |0.558611 |0.555873 |0.527433 |0.268170 |0.294836

|0.064529 |0.073495 |0.108791 |0.216262 |0.240852 |0.009337 |0.076970

|0.110842 |0.309182 |0.366904 |-0.240988 |-0.263647 |-0.294044

|family\_sample\_weight |-0.013156 |-0.014358 |0.784932 |0.757128

|0.783012 |-0.218601 |-0.201366 |-0.197155 |0.398751 |0.375106 |0.079352

|0.086499 |0.087680 |0.087492 |0.085450 |0.069781 |0.056497 |0.023014

|0.404443 |0.379400 |-0.272464 |-0.254407 |-0.279600 |0.860847 |0.739903

|-0.331517 |-0.327557 |-0.272769 |1.000000 |0.791573 |-0.277379

|-0.224707 |0.683764 |0.492317 |-0.287267 |-0.276274 |-0.200686

|0.536271 |0.577789 |0.016873 |0.006078 |-0.144055 |-0.009193 |-0.046072

|0.128440 |0.021560 |-0.236470 |-0.231381 |-0.228401 |-0.146880

|-0.138018 |0.104320 |0.678043 |0.757464 |-0.126613 |-0.113699 |0.094391

|0.705780 |0.784018 |-0.001088 |0.103197 |0.058545 |0.054943 |0.044689

|family\_samples |-0.028408 |-0.015563 |0.930581 |0.898409 |0.927525

|0.165236 |0.159185 |0.099145 |0.109174 |0.233305 |0.035376 |0.004376

|-0.052591 |-0.082005 |-0.094067 |-0.102416 |-0.110252 |-0.124488

|0.250155 |0.241263 |0.265611 |0.275657 |0.198516 |0.654357 |0.895762

|0.211392 |0.206297 |0.165911 |0.791573 |1.000000 |0.108146 |0.112783

|0.732987 |0.866272 |0.075870 |0.071082 |0.070981 |0.583697 |0.492629

|0.070757 |0.068728 |0.136073 |0.233427 |-0.182879 |-0.175591 |-0.229376

|0.154026 |0.153759 |0.142422 |-0.026546 |0.030067 |0.134560 |0.781136

|0.898807 |-0.039914 |0.020127 |0.081382 |0.815770 |0.928739 |0.286603

|0.392347 |-0.146165 |-0.134334 |-0.161526

|hc\_mortgage\_median |-0.057968 |-0.010897 |0.110547 |0.107893 |0.109032

|0.751698 |0.715387 |0.632512 |-0.260194 |0.061970 |0.008888 |0.030748

|0.002191 |-0.021375 |-0.039909 |-0.037621 |-0.039732 |-0.031879

|0.040369 |0.064905 |0.756335 |0.718150 |0.751027 |-0.287855 |0.082918

|0.747320 |0.733640 |0.682403 |-0.277379 |0.108146 |1.000000 |0.737910

|-0.313073 |0.190980 |0.789440 |0.768558 |0.596558 |-0.176544 |-0.360998

|0.112608 |0.145431 |0.461155 |0.393487 |-0.149006 |-0.463435 |-0.410406

|0.321544 |0.336098 |0.285128 |0.076708 |0.119335 |-0.086772 |0.083962

|0.107861 |0.034644 |0.066310 |-0.133261 |0.086811 |0.109191 |0.059133

|0.211894 |-0.077229 |-0.171184 |-0.399903

|hc\_mortgage\_stdev |-0.015473 |0.005096 |0.084803 |0.083352 |0.083065

|0.566718 |0.527718 |0.526635 |-0.265882 |-0.001861 |-0.029236

|-0.031588 |-0.058445 |-0.080582 |-0.088359 |-0.083934 |-0.082991

|-0.072646 |-0.012188 |0.002804 |0.648276 |0.578333 |0.731735 |-0.214002

|0.104652 |0.671703 |0.620760 |0.712575 |-0.224707 |0.112783 |0.737910

|1.000000 |-0.207533 |0.186540 |0.615663 |0.575139 |0.605976 |-0.016048

|-0.171131 |0.068890 |0.096458 |0.382562 |0.234569 |-0.149581 |-0.415584

|-0.233094 |0.342810 |0.348296 |0.316141 |0.242516 |0.267509 |0.031111

|0.056977 |0.083205 |0.193392 |0.216632 |-0.028162 |0.057690 |0.083186

|0.150820 |0.275825 |-0.114287 |-0.181585 |-0.299325

|hc\_mortgage\_sample\_weight |-0.010268 |-0.014672 |0.645214 |0.620826

|0.645148 |-0.150295 |-0.141133 |-0.171261 |0.041720 |-0.040622

|0.006068 |-0.052304 |-0.107276 |-0.132449 |-0.136743 |-0.142182

|-0.144599 |-0.153330 |-0.020701 |-0.034627 |-0.001732 |0.027529

|-0.079634 |0.602444 |0.669830 |-0.021858 |-0.012113 |-0.085189

|0.683764 |0.732987 |-0.313073 |-0.207533 |1.000000 |0.775357 |-0.261982

|-0.256402 |-0.200387 |0.600970 |0.618890 |0.016712 |0.001961 |-0.005236

|0.140116 |-0.164725 |-0.052005 |-0.116143 |0.170579 |0.151087 |0.178979

|0.054245 |0.090028 |0.164076 |0.537322 |0.620993 |0.050709 |0.097415

|0.130341 |0.567275 |0.646093 |0.417911 |0.313649 |-0.210699 |-0.119283

|0.068633

|hc\_mortgage\_samples |-0.038050 |-0.016985 |0.774241 |0.747064 |0.772103

|0.290788 |0.277301 |0.177666 |-0.153277 |-0.037641 |0.031835 |-0.014924

|-0.099913 |-0.145507 |-0.164833 |-0.172005 |-0.177969 |-0.182581

|-0.033879 |-0.029196 |0.463268 |0.477199 |0.350089 |0.388001 |0.756384

|0.425687 |0.420533 |0.307165 |0.492317 |0.866272 |0.190980 |0.186540

|0.775357 |1.000000 |0.170012 |0.159694 |0.116000 |0.468750 |0.347090

|0.116680 |0.118307 |0.291805 |0.401935 |-0.291311 |-0.354113 |-0.408305

|0.383772 |0.372426 |0.369662 |0.063356 |0.139922 |0.108386 |0.638037

|0.747289 |0.034903 |0.117314 |0.041243 |0.670180 |0.773233 |0.495645

|0.455726 |-0.276123 |-0.226232 |-0.177670

|hc\_mean |-0.056751 |-0.010590 |0.057414 |0.048706 |0.063856 |0.596115

|0.563220 |0.508916 |-0.211770 |0.030317 |-0.010413 |-0.002964

|-0.029155 |-0.050387 |-0.068875 |-0.061477 |-0.057354 |-0.041567

|0.007688 |0.033161 |0.677319 |0.639566 |0.682836 |-0.259025 |0.074946

|0.689790 |0.675614 |0.628698 |-0.287267 |0.075870 |0.789440 |0.615663

|-0.261982 |0.170012 |1.000000 |0.977424 |0.698084 |-0.109651 |-0.341865

|0.035365 |0.057139 |0.363104 |0.299659 |-0.086879 |-0.382132 |-0.299048

|0.361028 |0.366544 |0.330576 |0.124945 |0.156751 |-0.036624 |0.026409

|0.048728 |0.091112 |0.114765 |-0.082837 |0.044602 |0.063858 |0.101780

|0.200965 |-0.117806 |-0.168849 |-0.339450

|hc\_median |-0.058165 |-0.010922 |0.056097 |0.046975 |0.062996 |0.578463

|0.548361 |0.488165 |-0.197356 |0.036377 |-0.008710 |0.000613 |-0.024360

|-0.044528 |-0.063033 |-0.055139 |-0.051641 |-0.037503 |0.014138

|0.038914 |0.651384 |0.619180 |0.647638 |-0.248109 |0.071382 |0.661363

|0.652717 |0.591165 |-0.276274 |0.071082 |0.768558 |0.575139 |-0.256402

|0.159694 |0.977424 |1.000000 |0.615674 |-0.112179 |-0.338423 |0.035906

|0.055926 |0.348039 |0.291054 |-0.081349 |-0.366463 |-0.292040 |0.343761

|0.349991 |0.313458 |0.105605 |0.136843 |-0.042331 |0.026227 |0.046996

|0.076609 |0.098670 |-0.085710 |0.044942 |0.063010 |0.089084 |0.186182

|-0.111775 |-0.161710 |-0.330823

|hc\_stdev |-0.006400 |0.004735 |0.055962 |0.052662 |0.057125 |0.448193

|0.416664 |0.441140 |-0.183569 |0.023233 |-0.042615 |-0.041870

|-0.046491 |-0.057454 |-0.062268 |-0.055362 |-0.050551 |-0.036779

|0.013353 |0.026808 |0.518828 |0.462524 |0.593203 |-0.171369 |0.086831

|0.549511 |0.520649 |0.570258 |-0.200686 |0.070981 |0.596558 |0.605976

|-0.200387 |0.116000 |0.698084 |0.615674 |1.000000 |0.009142 |-0.160170

|-0.014840 |0.001137 |0.218933 |0.117595 |-0.038630 |-0.236812

|-0.103972 |0.268591 |0.267995 |0.251899 |0.200446 |0.214164 |0.033829

|0.028686 |0.052685 |0.155349 |0.165716 |-0.008316 |0.035750 |0.057041

|0.106419 |0.186482 |-0.074546 |-0.122186 |-0.222098

|hc\_samples |0.051854 |0.008908 |0.458678 |0.441285 |0.458686 |-0.143196

|-0.145708 |-0.080714 |-0.067776 |-0.156547 |-0.094810 |-0.144934

|-0.156173 |-0.151306 |-0.132232 |-0.123431 |-0.119105 |-0.119509

|-0.121878 |-0.153115 |0.035331 |0.030784 |0.049629 |0.505851 |0.552013

|0.049525 |0.046621 |0.040727 |0.536271 |0.583697 |-0.176544 |-0.016048

|0.600970 |0.468750 |-0.109651 |-0.112179 |0.009142 |1.000000 |0.947436

|-0.166741 |-0.179927 |-0.164893 |-0.370614 |0.060779 |0.142091

|0.409329 |0.124201 |0.103495 |0.138267 |0.392873 |0.393688 |0.299959

|0.380567 |0.441480 |0.372159 |0.394625 |0.236860 |0.399983 |0.459247

|0.473480 |0.373015 |-0.192828 |-0.120478 |0.056935

|hc\_sample\_weight |0.067830 |0.010834 |0.386986 |0.375081 |0.384260

|-0.298275 |-0.289261 |-0.216093 |-0.003988 |-0.154598 |-0.095496

|-0.143755 |-0.141499 |-0.128810 |-0.103667 |-0.096212 |-0.092358

|-0.096747 |-0.113263 |-0.152480 |-0.156633 |-0.152378 |-0.137736

|0.535909 |0.464226 |-0.146074 |-0.143901 |-0.131411 |0.577789 |0.492629

|-0.360998 |-0.171131 |0.618890 |0.347090 |-0.341865 |-0.338423

|-0.160170 |0.947436 |1.000000 |-0.175441 |-0.192572 |-0.268877

|-0.439638 |0.097917 |0.258132 |0.475531 |-0.007127 |-0.028404 |0.016343

|0.323074 |0.314752 |0.282744 |0.327314 |0.375248 |0.309130 |0.323402

|0.238005 |0.338267 |0.384723 |0.400465 |0.283773 |-0.133442 |-0.054778

|0.147810

|home\_equity\_second\_mortgage |-0.042531 |-0.014830 |0.084978 |0.081578

|0.085155 |0.129245 |0.129134 |0.060740 |0.002982 |0.081712 |0.072386

|0.082194 |0.059369 |0.036333 |0.023422 |0.005665 |0.000213 |-0.005162

|0.070341 |0.083914 |0.082510 |0.091840 |0.039888 |0.000271 |0.059491

|0.058327 |0.055799 |0.037690 |0.016873 |0.070757 |0.112608 |0.068890

|0.016712 |0.116680 |0.035365 |0.035906 |-0.014840 |-0.166741 |-0.175441

|1.000000 |0.931887 |0.516285 |0.334901 |-0.686036 |-0.467785 |-0.337642

|0.063771 |0.066669 |0.055905 |-0.138166 |-0.115535 |-0.095698 |0.064325

|0.081552 |-0.136623 |-0.119886 |-0.099382 |0.071663 |0.085216

|-0.053974 |-0.009606 |0.008342 |-0.008629 |-0.042251

|second\_mortgage |-0.044410 |-0.014817 |0.084867 |0.081213 |0.085298

|0.153188 |0.151884 |0.081253 |-0.007619 |0.081188 |0.078542 |0.090659

|0.069729 |0.047963 |0.034594 |0.016197 |0.009993 |0.000192 |0.069003

|0.083164 |0.101047 |0.108533 |0.060132 |-0.011950 |0.055499 |0.077101

|0.073732 |0.057847 |0.006078 |0.068728 |0.145431 |0.096458 |0.001961

|0.118307 |0.057139 |0.055926 |0.001137 |-0.179927 |-0.192572 |0.931887

|1.000000 |0.507355 |0.355311 |-0.732504 |-0.463977 |-0.358418 |0.065923

|0.069282 |0.057209 |-0.134994 |-0.111047 |-0.097751 |0.065701 |0.081182

|-0.135032 |-0.117000 |-0.103851 |0.073838 |0.085356 |-0.055017

|-0.011215 |0.007998 |-0.010866 |-0.058995

|home\_equity |-0.080306 |-0.024547 |0.105696 |0.100935 |0.106441

|0.413819 |0.391680 |0.308594 |-0.157704 |-0.005212 |0.070639 |0.062408

|0.008607 |-0.033336 |-0.061938 |-0.072611 |-0.075330 |-0.074420

|-0.026763 |-0.000971 |0.476613 |0.468386 |0.420615 |-0.139094 |0.110933

|0.465874 |0.452519 |0.392093 |-0.144055 |0.136073 |0.461155 |0.382562

|-0.005236 |0.291805 |0.363104 |0.348039 |0.218933 |-0.164893 |-0.268877

|0.516285 |0.507355 |1.000000 |0.534620 |-0.477237 |-0.938310 |-0.532900

|0.360495 |0.359307 |0.337041 |0.023198 |0.070223 |-0.029166 |0.075778

|0.100871 |0.010571 |0.049697 |-0.051914 |0.080716 |0.106486 |0.143811

|0.190072 |-0.159945 |-0.158590 |-0.211249

|debt |-0.117135 |-0.041554 |0.247888 |0.236359 |0.249992 |0.437968

|0.425381 |0.272806 |-0.074597 |0.131981 |0.149861 |0.145918 |0.071959

|0.015669 |-0.021081 |-0.040281 |-0.049929 |-0.059033 |0.102497

|0.137321 |0.422640 |0.432203 |0.325396 |-0.051924 |0.193775 |0.381264

|0.372158 |0.301182 |-0.009193 |0.233427 |0.393487 |0.234569 |0.140116

|0.401935 |0.299659 |0.291054 |0.117595 |-0.370614 |-0.439638 |0.334901

|0.355311 |0.534620 |1.000000 |-0.401072 |-0.566776 |-0.968423 |0.282577

|0.290578 |0.253457 |-0.261394 |-0.199182 |-0.198644 |0.192619 |0.236642

|-0.272968 |-0.227554 |-0.201750 |0.211217 |0.250049 |0.028166 |0.106946

|-0.093309 |-0.121939 |-0.227943

|second\_mortgage\_cdf |0.047465 |0.016077 |-0.152876 |-0.146775

|-0.153178 |-0.185721 |-0.180258 |-0.105882 |0.109338 |0.039186

|-0.062505 |-0.051230 |-0.011099 |0.019903 |0.036160 |0.051551 |0.055545

|0.063165 |0.050125 |0.033092 |-0.227276 |-0.233457 |-0.173092

|-0.004663 |-0.136666 |-0.203528 |-0.197199 |-0.160879 |-0.046072

|-0.182879 |-0.149006 |-0.149581 |-0.164725 |-0.291311 |-0.086879

|-0.081349 |-0.038630 |0.060779 |0.097917 |-0.686036 |-0.732504

|-0.477237 |-0.401072 |1.000000 |0.551102 |0.384299 |-0.207536

|-0.199927 |-0.200735 |0.020379 |-0.014004 |-0.041459 |-0.108244

|-0.146706 |0.022031 |-0.012852 |-0.024776 |-0.115445 |-0.153265

|-0.180715 |-0.150196 |0.114806 |0.101325 |0.057750

|home\_equity\_cdf |0.087734 |0.025872 |-0.132071 |-0.123673 |-0.135416

|-0.428932 |-0.405522 |-0.318893 |0.183686 |0.029133 |-0.074909

|-0.059755 |0.002145 |0.048483 |0.077623 |0.088144 |0.091189 |0.089769

|0.051103 |0.022051 |-0.515186 |-0.504066 |-0.463083 |0.123886

|-0.147467 |-0.508388 |-0.490767 |-0.438774 |0.128440 |-0.175591

|-0.463435 |-0.415584 |-0.052005 |-0.354113 |-0.382132 |-0.366463

|-0.236812 |0.142091 |0.258132 |-0.467785 |-0.463977 |-0.938310

|-0.566776 |0.551102 |1.000000 |0.554807 |-0.417490 |-0.413137

|-0.393838 |-0.063845 |-0.109111 |-0.023195 |-0.088298 |-0.123576

|-0.051567 |-0.090353 |0.002085 |-0.100319 |-0.135458 |-0.215688

|-0.233642 |0.201287 |0.189726 |0.205626

|debt\_cdf |0.106694 |0.035456 |-0.254184 |-0.243430 |-0.255289

|-0.459536 |-0.446454 |-0.283658 |0.082890 |-0.138391 |-0.149331

|-0.152264 |-0.080238 |-0.025196 |0.011923 |0.033015 |0.044054 |0.052375

|-0.108717 |-0.143072 |-0.428350 |-0.441210 |-0.319317 |0.072597

|-0.183770 |-0.380650 |-0.374106 |-0.290308 |0.021560 |-0.229376

|-0.410406 |-0.233094 |-0.116143 |-0.408305 |-0.299048 |-0.292040

|-0.103972 |0.409329 |0.475531 |-0.337642 |-0.358418 |-0.532900

|-0.968423 |0.384299 |0.554807 |1.000000 |-0.255933 |-0.266428

|-0.226004 |0.293288 |0.228730 |0.240333 |-0.203234 |-0.243679 |0.305600

|0.258229 |0.247648 |-0.219901 |-0.255519 |-0.006041 |-0.083662

|0.075354 |0.110526 |0.249586

|hs\_degree |-0.002706 |0.006169 |0.047779 |0.030239 |0.063292 |0.363260

|0.333124 |0.270215 |-0.256984 |-0.118516 |-0.051368 |-0.125355

|-0.217589 |-0.274682 |-0.290700 |-0.284830 |-0.275408 |-0.253331

|-0.129694 |-0.111098 |0.583437 |0.562464 |0.553318 |-0.114904 |0.202019

|0.638384 |0.613345 |0.558611 |-0.236470 |0.154026 |0.321544 |0.342810

|0.170579 |0.383772 |0.361028 |0.343761 |0.268591 |0.124201 |-0.007127

|0.063771 |0.065923 |0.360495 |0.282577 |-0.207536 |-0.417490 |-0.255933

|1.000000 |0.966645 |0.959991 |0.310830 |0.341150 |0.070959 |0.026258

|0.030482 |0.275871 |0.306959 |-0.002686 |0.052253 |0.062774 |0.392744

|0.372597 |-0.454777 |-0.333383 |-0.092639

|hs\_degree\_male |-0.006499 |0.005950 |0.056287 |0.037469 |0.072744

|0.372288 |0.341960 |0.283391 |-0.236590 |-0.092197 |-0.044790

|-0.115279 |-0.204976 |-0.259821 |-0.277469 |-0.271853 |-0.262218

|-0.239286 |-0.104421 |-0.084933 |0.577860 |0.555444 |0.551734

|-0.110718 |0.203347 |0.629432 |0.603621 |0.555873 |-0.231381 |0.153759

|0.336098 |0.348296 |0.151087 |0.372426 |0.366544 |0.349991 |0.267995

|0.103495 |-0.028404 |0.066669 |0.069282 |0.359307 |0.290578 |-0.199927

|-0.413137 |-0.266428 |0.966645 |1.000000 |0.864498 |0.287734 |0.315717

|0.057181 |0.033157 |0.037486 |0.261924 |0.288082 |-0.003362 |0.061404

|0.072353 |0.359141 |0.357293 |-0.437306 |-0.327920 |-0.102871

|hs\_degree\_female |0.003046 |0.005889 |0.037697 |0.028657 |0.045205

|0.328594 |0.300575 |0.238460 |-0.263558 |-0.140690 |-0.056946

|-0.130872 |-0.219565 |-0.274354 |-0.288790 |-0.282907 |-0.274618

|-0.254049 |-0.150843 |-0.134381 |0.551221 |0.532398 |0.521099

|-0.114302 |0.185177 |0.607017 |0.584153 |0.527433 |-0.228401 |0.142422

|0.285128 |0.316141 |0.178979 |0.369662 |0.330576 |0.313458 |0.251899

|0.138267 |0.016343 |0.055905 |0.057209 |0.337041 |0.253457 |-0.200735

|-0.393838 |-0.226004 |0.959991 |0.864498 |1.000000 |0.319591 |0.349134

|0.070654 |0.023507 |0.028875 |0.275343 |0.309509 |-0.004624 |0.035812

|0.045271 |0.402457 |0.358228 |-0.429327 |-0.313358 |-0.071506

|male\_age\_mean |0.043723 |0.007767 |-0.192439 |-0.205537 |-0.172326

|0.043793 |0.024865 |0.110195 |-0.270018 |-0.296783 |-0.097346

|-0.132694 |-0.140507 |-0.145335 |-0.147319 |-0.142846 |-0.141201

|-0.135374 |-0.288532 |-0.291465 |0.203388 |0.169760 |0.264896

|-0.065206 |0.010053 |0.261157 |0.238433 |0.268170 |-0.146880 |-0.026546

|0.076708 |0.242516 |0.054245 |0.063356 |0.124945 |0.105605 |0.200446

|0.392873 |0.323074 |-0.138166 |-0.134994 |0.023198 |-0.261394 |0.020379

|-0.063845 |0.293288 |0.310830 |0.287734 |0.319591 |1.000000 |0.947964

|0.398579 |-0.271108 |-0.205628 |0.856579 |0.837644 |0.322018 |-0.242556

|-0.172608 |0.462326 |0.412265 |-0.159333 |-0.099688 |0.201965

|male\_age\_median |0.050695 |0.010537 |-0.155431 |-0.166188 |-0.139011

|0.093017 |0.074968 |0.116358 |-0.342550 |-0.362014 |-0.105726

|-0.147678 |-0.163784 |-0.173021 |-0.173988 |-0.168354 |-0.166731

|-0.161571 |-0.354316 |-0.355815 |0.286687 |0.261416 |0.312344

|-0.108775 |0.014505 |0.320837 |0.302184 |0.294836 |-0.138018 |0.030067

|0.119335 |0.267509 |0.090028 |0.139922 |0.156751 |0.136843 |0.214164

|0.393688 |0.314752 |-0.115535 |-0.111047 |0.070223 |-0.199182

|-0.014004 |-0.109111 |0.228730 |0.341150 |0.315717 |0.349134 |0.947964

|1.000000 |0.387385 |-0.239679 |-0.166262 |0.791338 |0.831498 |0.285148

|-0.216830 |-0.139405 |0.531994 |0.512755 |-0.188195 |-0.122179

|0.165970

|male\_age\_stdev |0.034123 |-0.001746 |-0.027974 |-0.075688 |0.020133

|-0.112621 |-0.115311 |-0.012700 |-0.133379 |-0.269415 |-0.060407

|-0.083468 |-0.079638 |-0.072503 |-0.069339 |-0.069966 |-0.073007

|-0.074132 |-0.237388 |-0.245749 |0.059374 |0.051456 |0.090284 |0.041628

|0.053006 |0.050293 |0.050041 |0.064529 |0.104320 |0.134560 |-0.086772

|0.031111 |0.164076 |0.108386 |-0.036624 |-0.042331 |0.033829 |0.299959

|0.282744 |-0.095698 |-0.097751 |-0.029166 |-0.198644 |-0.041459

|-0.023195 |0.240333 |0.070959 |0.057181 |0.070654 |0.398579 |0.387385

|1.000000 |-0.194169 |-0.075722 |0.455531 |0.458741 |0.772007 |-0.106518

|0.012751 |0.425840 |0.454104 |-0.166787 |-0.042954 |0.125313

|male\_age\_sample\_weight |-0.024953 |-0.012675 |0.912969 |0.942592

|0.849621 |0.131686 |0.125258 |0.095979 |0.237580 |0.373061 |0.056100

|0.061170 |0.037116 |0.023858 |0.017909 |0.014255 |0.010454 |-0.001272

|0.385697 |0.372792 |0.117184 |0.122703 |0.082886 |0.616392 |0.754327

|0.090582 |0.086878 |0.073495 |0.678043 |0.781136 |0.083962 |0.056977

|0.537322 |0.638037 |0.026409 |0.026227 |0.028686 |0.380567 |0.327314

|0.064325 |0.065701 |0.075778 |0.192619 |-0.108244 |-0.088298 |-0.203234

|0.026258 |0.033157 |0.023507 |-0.271108 |-0.239679 |-0.194169 |1.000000

|0.943010 |-0.262578 |-0.227533 |-0.172248 |0.881196 |0.849774 |0.025696

|0.020088 |-0.013030 |-0.083714 |-0.184634

|male\_age\_samples |-0.025676 |-0.013631 |0.980439 |0.999556 |0.924931

|0.159888 |0.154501 |0.108416 |0.224624 |0.383353 |0.050008 |0.044639

|0.006635 |-0.015494 |-0.026792 |-0.034174 |-0.041119 |-0.056527

|0.398849 |0.388629 |0.175599 |0.182912 |0.127649 |0.673026 |0.857410

|0.133879 |0.129669 |0.108791 |0.757464 |0.898807 |0.107861 |0.083205

|0.620993 |0.747289 |0.048728 |0.046996 |0.052685 |0.441480 |0.375248

|0.081552 |0.081182 |0.100871 |0.236642 |-0.146706 |-0.123576 |-0.243679

|0.030482 |0.037486 |0.028875 |-0.205628 |-0.166262 |-0.075722 |0.943010

|1.000000 |-0.199351 |-0.161553 |-0.066185 |0.870606 |0.925248 |0.093061

|0.140494 |-0.004274 |-0.076879 |-0.150758

|female\_age\_mean |0.017361 |-0.005620 |-0.196112 |-0.199602 |-0.185339

|0.003875 |-0.013918 |0.101978 |-0.207395 |-0.253734 |-0.070305

|-0.091147 |-0.089798 |-0.092474 |-0.100590 |-0.102545 |-0.103614

|-0.098938 |-0.245186 |-0.248507 |0.131128 |0.098061 |0.203050

|-0.023754 |0.012513 |0.196665 |0.174890 |0.216262 |-0.126613 |-0.039914

|0.034644 |0.193392 |0.050709 |0.034903 |0.091112 |0.076609 |0.155349

|0.372159 |0.309130 |-0.136623 |-0.135032 |0.010571 |-0.272968 |0.022031

|-0.051567 |0.305600 |0.275871 |0.261924 |0.275343 |0.856579 |0.791338

|0.455531 |-0.262578 |-0.199351 |1.000000 |0.949709 |0.399228 |-0.269523

|-0.185610 |0.414969 |0.331997 |-0.139054 |-0.077411 |0.184776

|female\_age\_median |0.030907 |-0.001258 |-0.163807 |-0.161795 |-0.159668

|0.045239 |0.029037 |0.099196 |-0.303488 |-0.350294 |-0.082919

|-0.113285 |-0.119396 |-0.124816 |-0.131239 |-0.131202 |-0.132240

|-0.128332 |-0.341325 |-0.344168 |0.214398 |0.189832 |0.248528

|-0.075653 |0.008989 |0.254205 |0.235868 |0.240852 |-0.113699 |0.020127

|0.066310 |0.216632 |0.097415 |0.117314 |0.114765 |0.098670 |0.165716

|0.394625 |0.323402 |-0.119886 |-0.117000 |0.049697 |-0.227554

|-0.012852 |-0.090353 |0.258229 |0.306959 |0.288082 |0.309509 |0.837644

|0.831498 |0.458741 |-0.227533 |-0.161553 |0.949709 |1.000000 |0.370185

|-0.254180 |-0.159902 |0.512314 |0.432651 |-0.179475 |-0.109790

|0.138756

|female\_age\_stdev |0.026841 |-0.006398 |-0.026764 |-0.063224 |0.010198

|-0.172810 |-0.175715 |-0.025764 |-0.034040 |-0.177937 |-0.048293

|-0.070668 |-0.058901 |-0.053383 |-0.053060 |-0.053321 |-0.056169

|-0.055382 |-0.146247 |-0.153574 |-0.031619 |-0.040775 |0.022221

|0.080794 |0.046180 |-0.025327 |-0.026282 |0.009337 |0.094391 |0.081382

|-0.133261 |-0.028162 |0.130341 |0.041243 |-0.082837 |-0.085710

|-0.008316 |0.236860 |0.238005 |-0.099382 |-0.103851 |-0.051914

|-0.201750 |-0.024776 |0.002085 |0.247648 |-0.002686 |-0.003362

|-0.004624 |0.322018 |0.285148 |0.772007 |-0.172248 |-0.066185 |0.399228

|0.370185 |1.000000 |-0.117473 |0.010213 |0.318024 |0.298611 |-0.059466

|0.007100 |0.166414

|female\_age\_sample\_weight |-0.044301 |-0.019013 |0.923600 |0.871078

|0.940876 |0.128762 |0.120780 |0.108126 |0.287024 |0.415831 |0.080626

|0.088634 |0.063714 |0.049357 |0.042263 |0.037121 |0.034196 |0.023776

|0.431320 |0.419200 |0.104465 |0.107838 |0.083581 |0.674606 |0.813252

|0.085941 |0.082225 |0.076970 |0.705780 |0.815770 |0.086811 |0.057690

|0.567275 |0.670180 |0.044602 |0.044942 |0.035750 |0.399983 |0.338267

|0.071663 |0.073838 |0.080716 |0.211217 |-0.115445 |-0.100319 |-0.219901

|0.052253 |0.061404 |0.035812 |-0.242556 |-0.216830 |-0.106518 |0.881196

|0.870606 |-0.269523 |-0.254180 |-0.117473 |1.000000 |0.942253 |0.017581

|0.069773 |-0.073133 |-0.090515 |-0.197474

|female\_age\_samples |-0.045249 |-0.019089 |0.980851 |0.925742 |0.998539

|0.159953 |0.153199 |0.122759 |0.260621 |0.414716 |0.075889 |0.073455

|0.034750 |0.010886 |-0.001076 |-0.010729 |-0.018249 |-0.034479

|0.433142 |0.423429 |0.165483 |0.171136 |0.128690 |0.720637 |0.903620

|0.130209 |0.126238 |0.110842 |0.784018 |0.928739 |0.109191 |0.083186

|0.646093 |0.773233 |0.063858 |0.063010 |0.057041 |0.459247 |0.384723

|0.085216 |0.085356 |0.106486 |0.250049 |-0.153265 |-0.135458 |-0.255519

|0.062774 |0.072353 |0.045271 |-0.172608 |-0.139405 |0.012751 |0.849774

|0.925248 |-0.185610 |-0.159902 |0.010213 |0.942253 |1.000000 |0.092161

|0.193570 |-0.068055 |-0.088227 |-0.170718

|pct\_own |0.050627 |0.010259 |0.094828 |0.093538 |0.092555 |0.140698

|0.132248 |0.049597 |-0.613327 |-0.686417 |-0.098433 |-0.180554

|-0.237966 |-0.260684 |-0.252927 |-0.240134 |-0.234997 |-0.228085

|-0.680462 |-0.680498 |0.481080 |0.494538 |0.377004 |-0.177136 |0.079529

|0.450855 |0.452139 |0.309182 |-0.001088 |0.286603 |0.059133 |0.150820

|0.417911 |0.495645 |0.101780 |0.089084 |0.106419 |0.473480 |0.400465

|-0.053974 |-0.055017 |0.143811 |0.028166 |-0.180715 |-0.215688

|-0.006041 |0.392744 |0.359141 |0.402457 |0.462326 |0.531994 |0.425840

|0.025696 |0.093061 |0.414969 |0.512314 |0.318024 |0.017581 |0.092161

|1.000000 |0.686640 |-0.390745 |-0.287860 |-0.103050

|married |0.028043 |-0.000761 |0.171907 |0.140432 |0.196526 |0.256218

|0.242848 |0.130537 |-0.440904 |-0.393193 |-0.093731 |-0.179243

|-0.251667 |-0.289002 |-0.290348 |-0.288658 |-0.289330 |-0.293197

|-0.381922 |-0.375701 |0.533542 |0.536703 |0.442274 |-0.092904 |0.184851

|0.483535 |0.476421 |0.366904 |0.103197 |0.392347 |0.211894 |0.275825

|0.313649 |0.455726 |0.200965 |0.186182 |0.186482 |0.373015 |0.283773

|-0.009606 |-0.011215 |0.190072 |0.106946 |-0.150196 |-0.233642

|-0.083662 |0.372597 |0.357293 |0.358228 |0.412265 |0.512755 |0.454104

|0.020088 |0.140494 |0.331997 |0.432651 |0.298611 |0.069773 |0.193570

|0.686640 |1.000000 |-0.254288 |-0.222314 |-0.277228

|married\_snp |0.003994 |0.005268 |-0.039066 |-0.004272 |-0.071924

|-0.107838 |-0.095903 |-0.070423 |0.216546 |0.213205 |0.037412 |0.095322

|0.151538 |0.179466 |0.182522 |0.174162 |0.164097 |0.149117 |0.207102

|0.204754 |-0.296233 |-0.296818 |-0.253932 |0.054834 |-0.113021

|-0.318561 |-0.314633 |-0.240988 |0.058545 |-0.146165 |-0.077229

|-0.114287 |-0.210699 |-0.276123 |-0.117806 |-0.111775 |-0.074546

|-0.192828 |-0.133442 |0.008342 |0.007998 |-0.159945 |-0.093309

|0.114806 |0.201287 |0.075354 |-0.454777 |-0.437306 |-0.429327

|-0.159333 |-0.188195 |-0.166787 |-0.013030 |-0.004274 |-0.139054

|-0.179475 |-0.059466 |-0.073133 |-0.068055 |-0.390745 |-0.254288

|1.000000 |0.675407 |0.060225

|separated |-0.005773 |-0.001177 |-0.084397 |-0.076845 |-0.088690

|-0.188070 |-0.174123 |-0.136935 |0.204450 |0.141390 |0.041863 |0.089480

|0.129711 |0.153450 |0.156383 |0.146173 |0.137820 |0.125471 |0.144319

|0.137527 |-0.317903 |-0.312044 |-0.284381 |0.077240 |-0.091849

|-0.324546 |-0.315453 |-0.263647 |0.054943 |-0.134334 |-0.171184

|-0.181585 |-0.119283 |-0.226232 |-0.168849 |-0.161710 |-0.122186

|-0.120478 |-0.054778 |-0.008629 |-0.010866 |-0.158590 |-0.121939

|0.101325 |0.189726 |0.110526 |-0.333383 |-0.327920 |-0.313358

|-0.099688 |-0.122179 |-0.042954 |-0.083714 |-0.076879 |-0.077411

|-0.109790 |0.007100 |-0.090515 |-0.088227 |-0.287860 |-0.222314

|0.675407 |1.000000 |0.133960

|divorced |0.023420 |0.007678 |-0.164737 |-0.150692 |-0.172432

|-0.377479 |-0.361252 |-0.270542 |0.213381 |0.052663 |-0.010049

|-0.005079 |0.020269 |0.023998 |0.025301 |0.019348 |0.014230 |0.009755

|0.062983 |0.053767 |-0.396216 |-0.391262 |-0.348822 |0.169202

|-0.044028 |-0.359034 |-0.352417 |-0.294044 |0.044689 |-0.161526

|-0.399903 |-0.299325 |0.068633 |-0.177670 |-0.339450 |-0.330823

|-0.222098 |0.056935 |0.147810 |-0.042251 |-0.058995 |-0.211249

|-0.227943 |0.057750 |0.205626 |0.249586 |-0.092639 |-0.102871

|-0.071506 |0.201965 |0.165970 |0.125313 |-0.184634 |-0.150758 |0.184776

|0.138756 |0.166414 |-0.197474 |-0.170718 |-0.103050 |-0.277228

|0.060225 |0.133960 |1.000000

|===

----

+\*In[159]:\*+

[source, ipython3]

----

for i in range(len(correlation\_matrix.columns)):

for j in range(i):

if abs(correlation\_matrix.iloc[i, j]) > 0.8:

colname = correlation\_matrix.columns[i]

correlated\_features.add(colname)

----

+\*In[160]:\*+

[source, ipython3]

----

correlated\_features

----

+\*Out[160]:\*+

----{'debt\_cdf',

'family\_mean',

'family\_median',

'family\_sample\_weight',

'family\_samples',

'family\_stdev',

'female\_age\_mean',

'female\_age\_median',

'female\_age\_sample\_weight',

'female\_age\_samples',

'female\_pop',

'hc\_median',

'hc\_mortgage\_samples',

'hc\_sample\_weight',

'hi\_median',

'hi\_samples',

'hi\_stdev',

'home\_equity\_cdf',

'hs\_degree\_female',

'hs\_degree\_male',

'male\_age\_median',

'male\_age\_sample\_weight',

'male\_age\_samples',

'male\_pop',

'rent\_gt\_25',

'rent\_gt\_30',

'rent\_gt\_35',

'rent\_gt\_40',

'rent\_gt\_50',

'rent\_median',

'rent\_samples',

'second\_mortgage',

'universe\_samples',

'used\_samples'}----

+\*In[161]:\*+

[source, ipython3]

----

corr\_list = ['debt\_cdf', 'family\_mean', 'family\_median', 'family\_sample\_weight', 'family\_samples', 'family\_stdev', 'female\_age\_mean',

'female\_age\_median', 'female\_age\_sample\_weight', 'female\_age\_samples', 'female\_pop', 'hc\_median', 'hc\_mortgage\_samples', 'hc\_sample\_weight',

'hi\_median', 'hi\_samples', 'hi\_stdev', 'home\_equity\_cdf', 'hs\_degree\_female', 'hs\_degree\_male', 'male\_age\_median', 'male\_age\_sample\_weight',

'male\_age\_samples', 'male\_pop', 'rent\_gt\_25', 'rent\_gt\_30', 'rent\_gt\_35', 'rent\_gt\_40', 'rent\_gt\_50', 'rent\_median', 'rent\_samples',

'second\_mortgage', 'universe\_samples', 'used\_samples']

----

+\*In[162]:\*+

[source, ipython3]

----

train\_df.drop(corr\_list, axis=1, inplace=True)

test\_df.drop(corr\_list, axis=1, inplace=True)

----

+\*In[163]:\*+

[source, ipython3]

----

print(train\_df.shape, test\_df.shape)

----

+\*Out[163]:\*+

----

(27019, 31) (11603, 31)

----

performing regression after removing multi-collinear variable

+\*In[164]:\*+

[source, ipython3]

----

train\_df.head()

----

+\*Out[164]:\*+

----

[cols=",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,",options="header",]

|===

| |ALand |AWater |pop |rent\_mean |rent\_stdev |rent\_sample\_weight

|rent\_gt\_10 |rent\_gt\_15 |rent\_gt\_20 |hi\_mean |hi\_sample\_weight

|hc\_mortgage\_mean |hc\_mortgage\_median |hc\_mortgage\_stdev

|hc\_mortgage\_sample\_weight |hc\_mean |hc\_stdev |hc\_samples

|home\_equity\_second\_mortgage |home\_equity |debt |second\_mortgage\_cdf

|hs\_degree |male\_age\_mean |male\_age\_stdev |female\_age\_stdev |pct\_own

|married |married\_snp |separated |divorced

|0 |202183361.0 |1699120 |5230 |769.38638 |232.63967 |272.34441 |0.86761

|0.79155 |0.59155 |63125.28406 |1290.96240 |1414.80295 |1223.0

|641.22898 |377.83135 |570.01530 |270.11299 |770.0 |0.01588 |0.08919

|0.52963 |0.43658 |0.89288 |42.48574 |22.97306 |22.51276 |0.79046

|0.57851 |0.01882 |0.01240 |0.08770

|1 |1560828.0 |100363 |2633 |804.87924 |253.46747 |312.58622 |0.97410

|0.93227 |0.69920 |41931.92593 |838.74664 |864.41390 |784.0 |482.27020

|316.88320 |351.98293 |125.40457 |229.0 |0.02222 |0.04274 |0.60855

|0.42174 |0.90487 |34.84728 |20.37452 |23.43353 |0.52483 |0.34886

|0.01426 |0.01426 |0.09030

|2 |69561595.0 |284193 |6881 |742.77365 |323.39011 |291.85520 |0.95238

|0.88624 |0.79630 |84942.68317 |1155.20980 |1506.06758 |1361.0

|731.89394 |699.41354 |556.45986 |184.42175 |538.0 |0.00000 |0.09512

|0.73484 |1.00000 |0.94288 |39.38154 |22.89769 |23.94119 |0.85331

|0.64745 |0.02830 |0.01607 |0.10657

|3 |1105793.0 |0 |2700 |803.42018 |297.39258 |259.30316 |0.94693

|0.87151 |0.69832 |48733.67116 |928.32193 |1175.28642 |1101.0 |428.98751

|261.28471 |288.04047 |185.55887 |392.0 |0.01086 |0.01086 |0.52714

|0.53057 |0.91500 |48.64749 |23.05968 |24.32015 |0.65037 |0.47257

|0.02021 |0.02021 |0.10106

|4 |2554403.0 |0 |5637 |938.56493 |392.44096 |1005.42886 |0.99286

|0.98247 |0.91688 |31834.15466 |1548.67477 |1192.58759 |1125.0

|327.49674 |76.61052 |443.68855 |76.12674 |124.0 |0.05426 |0.05426

|0.51938 |0.18332 |1.00000 |26.07533 |11.84399 |11.10484 |0.13046

|0.12356 |0.00000 |0.00000 |0.03109

|===

----

+\*In[165]:\*+

[source, ipython3]

----

train\_X = train\_df.drop(columns=['hc\_mortgage\_mean'])

train\_y = train\_df['hc\_mortgage\_mean']

----

+\*In[166]:\*+

[source, ipython3]

----

lin\_reg.fit(train\_X, train\_y)

----

+\*Out[166]:\*+

----LinearRegression()----

+\*In[167]:\*+

[source, ipython3]

----

test\_X = test\_df.drop(columns=['hc\_mortgage\_mean'])

test\_y = test\_df['hc\_mortgage\_mean']

----

+\*In[168]:\*+

[source, ipython3]

----

pred\_y = lin\_reg.predict(test\_X)

----

+\*In[169]:\*+

[source, ipython3]

----

# model evaluation for testing set

mae = mean\_absolute\_error(test\_y, pred\_y)

mse = mean\_squared\_error(test\_y, pred\_y)

r2 = r2\_score(test\_y, pred\_y)

adj\_rsqrd = 1 - (1-r2)\*(len(test\_y) - 1) / (len(test\_y) - (test\_X.shape[1] - 1) - 1)

print('Mean Absolute Error is {}'.format(round(mae, 3)))

print('Mean Squared Error is {}'.format(round(mse, 3)))

print('Root Mean Squared Error is {}'.format(round(mse\*\*(0.5), 3)))

print('R-Square score is {}'.format(round(r2, 3)))

print('Adjusted R-Square score is {}'.format(round(adj\_rsqrd, 3)))

----

+\*Out[169]:\*+

----

Mean Absolute Error is 44.819

Mean Squared Error is 5141.608

Root Mean Squared Error is 71.705

R-Square score is 0.987

Adjusted R-Square score is 0.987

----

+\*In[170]:\*+

[source, ipython3]

----

sorted(SCORERS.keys())

----

+\*Out[170]:\*+

----['accuracy',

'adjusted\_mutual\_info\_score',

'adjusted\_rand\_score',

'average\_precision',

'balanced\_accuracy',

'completeness\_score',

'explained\_variance',

'f1',

'f1\_macro',

'f1\_micro',

'f1\_samples',

'f1\_weighted',

'fowlkes\_mallows\_score',

'homogeneity\_score',

'jaccard',

'jaccard\_macro',

'jaccard\_micro',

'jaccard\_samples',

'jaccard\_weighted',

'max\_error',

'mutual\_info\_score',

'neg\_brier\_score',

'neg\_log\_loss',

'neg\_mean\_absolute\_error',

'neg\_mean\_gamma\_deviance',

'neg\_mean\_poisson\_deviance',

'neg\_mean\_squared\_error',

'neg\_mean\_squared\_log\_error',

'neg\_median\_absolute\_error',

'neg\_root\_mean\_squared\_error',

'normalized\_mutual\_info\_score',

'precision',

'precision\_macro',

'precision\_micro',

'precision\_samples',

'precision\_weighted',

'r2',

'recall',

'recall\_macro',

'recall\_micro',

'recall\_samples',

'recall\_weighted',

'roc\_auc',

'roc\_auc\_ovo',

'roc\_auc\_ovo\_weighted',

'roc\_auc\_ovr',

'roc\_auc\_ovr\_weighted',

'v\_measure\_score']----

+\*In[171]:\*+

[source, ipython3]

----

import random

randomlist = []

for i in range(0,100):

n = random.randint(1,len(test\_df))

randomlist.append(n)

print(randomlist)

----

+\*Out[171]:\*+

----

[4447, 4293, 6909, 5268, 653, 5747, 6175, 11502, 11070, 4851, 6336, 9986, 7467, 6265, 8348, 8630, 8271, 9496, 11207, 10017, 6324, 8437, 509, 9311, 8738, 658, 9620, 953, 8416, 5024, 9939, 4291, 5794, 3516, 3178, 3865, 102, 8585, 9307, 731, 6918, 9581, 5982, 10467, 9070, 251, 10913, 8247, 7105, 9938, 7197, 1272, 769, 1177, 4847, 9159, 430, 7118, 8720, 9596, 10788, 10817, 2643, 8695, 222, 2224, 5538, 11538, 10621, 8451, 6963, 6435, 602, 6523, 11102, 10607, 3091, 10069, 3389, 1759, 7052, 10900, 4912, 866, 11038, 8201, 3345, 3721, 498, 1312, 10993, 8345, 3333, 10970, 8319, 8295, 7067, 1522, 4067, 8502]

----

+\*In[172]:\*+

[source, ipython3]

----

pre\_out = []

out = []

for i in randomlist:

data\_in = [list(test\_X.iloc[i])]

pre\_data\_out = lin\_reg.predict(data\_in)

data\_out = test\_y.iloc[i]

print(i, pre\_data\_out, data\_out)

pre\_out.append(pre\_data\_out)

out.append(data\_out)

----

+\*Out[172]:\*+

----

4447 [1078.69377576] 1086.4188900000001

4293 [786.09495242] 739.61163

6909 [1565.47212696] 1559.68985

5268 [2036.29265667] 1926.39242

653 [1826.33920845] 1778.8648899999998

5747 [2631.1294642] 2663.08793

6175 [1265.62127071] 1203.76602

11502 [2257.08784535] 2272.6168399999997

11070 [1398.95780998] 1355.35852

4851 [1565.51625888] 1566.79724

6336 [847.47157519] 858.7098699999999

9986 [2194.25799951] 2163.43284

7467 [1018.04127466] 1061.90173

6265 [1259.1128559] 1224.85055

8348 [1404.18697123] 1517.15876

8630 [1989.69405879] 1921.88825

8271 [1075.47725982] 1137.12158

9496 [1228.29937518] 1256.53099

11207 [1762.32893384] 1748.4288800000002

10017 [1828.49436631] 1750.13017

6324 [2979.76297875] 2888.2056399999997

8437 [1157.49912753] 1173.0045699999998

509 [1112.38818691] 1106.7562300000002

9311 [1184.86544839] 1196.6210800000001

8738 [2750.08236544] 2794.7814

658 [3106.66117086] 2882.1154899999997

9620 [2117.51893364] 2207.9828399999997

953 [1986.41166065] 2037.75016

8416 [955.30273534] 907.89755

5024 [1220.46362472] 1230.33555

9939 [1654.09986705] 1574.16648

4291 [1975.37719403] 1937.10588

5794 [923.08474755] 903.30174

3516 [1070.40831171] 1168.0677699999999

3178 [1377.61862563] 1349.36614

3865 [1194.53052854] 1162.80653

102 [1170.30204322] 1242.32844

8585 [1233.7722339] 1208.8901

9307 [1601.58967648] 1636.0932913514725

731 [952.681987] 951.7347

6918 [3004.24455235] 3347.62202

9581 [1412.7733369] 1484.99036

5982 [1254.45297669] 1259.25535

10467 [1372.54744852] 1323.34017

9070 [1169.10988504] 1183.4733800000001

251 [1964.68978145] 1997.5215600000001

10913 [1562.49163624] 1579.5503199999998

8247 [1178.41762374] 1216.2907300000002

7105 [1136.213933] 1160.24343

9938 [1175.00831484] 1187.2947900000001

7197 [1629.08206496] 1694.2871

1272 [2505.55628662] 2471.23854

769 [756.13382631] 772.32453

1177 [1510.7208324] 1474.7979699999999

4847 [2270.80046922] 2211.42531

9159 [850.56791519] 846.8262699999999

430 [1420.35292099] 1405.88276

7118 [1121.95721194] 1105.0771300000001

8720 [2173.39134113] 2119.42645

9596 [898.01341461] 870.29218

10788 [1279.80145202] 1289.5588

10817 [1157.31481896] 1191.98445

2643 [856.70629848] 809.35722

8695 [2019.03615568] 1970.42272

222 [1605.28564861] 1593.94955

2224 [1422.01869696] 1372.9438

5538 [1605.75499028] 1636.0932913514725

11538 [1207.22897498] 1222.12725

10621 [1187.77946481] 1195.8859699999998

8451 [1200.87603369] 1196.3459599999999

6963 [2225.9811698] 2385.0237399999996

6435 [1566.92168782] 1554.93104

602 [1626.35525526] 1636.0932913514725

6523 [1980.79180286] 1857.9066300000002

11102 [1405.72970307] 1408.69084

10607 [2644.04740355] 2627.6603600000003

3091 [1922.42527392] 1849.16795

10069 [1490.63554942] 1508.76795

3389 [1640.80200366] 1612.85813

1759 [1009.0193493] 1050.37573

7052 [1658.75683925] 1662.2681300000002

10900 [1602.51900655] 1636.0932913514725

4912 [1122.40159455] 1088.6359

866 [1044.52827719] 1052.53169

11038 [3116.09010767] 3063.5122300000003

8201 [1371.08659666] 1364.97237

3345 [1362.02412833] 1360.42649

3721 [1895.19122809] 1960.5756399999998

498 [1288.96765202] 1399.51799

1312 [1805.02409165] 1756.2159600000002

10993 [1520.58833948] 1520.3708900000001

8345 [1430.76414938] 1456.65572

3333 [999.38492233] 981.63145

10970 [1887.49106503] 1903.7696899999999

8319 [1637.77669102] 1636.0932913514725

8295 [3014.39436184] 3013.8814

7067 [2087.80582529] 2139.35322

1522 [1107.03296476] 1122.26788

4067 [2705.87077344] 2693.72426

8502 [855.14916403] 720.48133

----

+\*In[173]:\*+

[source, ipython3]

----

pre\_out

----

+\*Out[173]:\*+

----[array([1078.69377576]),

array([786.09495242]),

array([1565.47212696]),

array([2036.29265667]),

array([1826.33920845]),

array([2631.1294642]),

array([1265.62127071]),

array([2257.08784535]),

array([1398.95780998]),

array([1565.51625888]),

array([847.47157519]),

array([2194.25799951]),

array([1018.04127466]),

array([1259.1128559]),

array([1404.18697123]),

array([1989.69405879]),

array([1075.47725982]),

array([1228.29937518]),

array([1762.32893384]),

array([1828.49436631]),

array([2979.76297875]),

array([1157.49912753]),

array([1112.38818691]),

array([1184.86544839]),

array([2750.08236544]),

array([3106.66117086]),

array([2117.51893364]),

array([1986.41166065]),

array([955.30273534]),

array([1220.46362472]),

array([1654.09986705]),

array([1975.37719403]),

array([923.08474755]),

array([1070.40831171]),

array([1377.61862563]),

array([1194.53052854]),

array([1170.30204322]),

array([1233.7722339]),

array([1601.58967648]),

array([952.681987]),

array([3004.24455235]),

array([1412.7733369]),

array([1254.45297669]),

array([1372.54744852]),

array([1169.10988504]),

array([1964.68978145]),

array([1562.49163624]),

array([1178.41762374]),

array([1136.213933]),

array([1175.00831484]),

array([1629.08206496]),

array([2505.55628662]),

array([756.13382631]),

array([1510.7208324]),

array([2270.80046922]),

array([850.56791519]),

array([1420.35292099]),

array([1121.95721194]),

array([2173.39134113]),

array([898.01341461]),

array([1279.80145202]),

array([1157.31481896]),

array([856.70629848]),

array([2019.03615568]),

array([1605.28564861]),

array([1422.01869696]),

array([1605.75499028]),

array([1207.22897498]),

array([1187.77946481]),

array([1200.87603369]),

array([2225.9811698]),

array([1566.92168782]),

array([1626.35525526]),

array([1980.79180286]),

array([1405.72970307]),

array([2644.04740355]),

array([1922.42527392]),

array([1490.63554942]),

array([1640.80200366]),

array([1009.0193493]),

array([1658.75683925]),

array([1602.51900655]),

array([1122.40159455]),

array([1044.52827719]),

array([3116.09010767]),

array([1371.08659666]),

array([1362.02412833]),

array([1895.19122809]),

array([1288.96765202]),

array([1805.02409165]),

array([1520.58833948]),

array([1430.76414938]),

array([999.38492233]),

array([1887.49106503]),

array([1637.77669102]),

array([3014.39436184]),

array([2087.80582529]),

array([1107.03296476]),

array([2705.87077344]),

array([855.14916403])]----

+\*In[174]:\*+

[source, ipython3]

----

fig, ax = plt.subplots(figsize=(20,10))

ax.scatter(pre\_out, out, edgecolors=(0, 1, 0))

ax.plot([min(out), max(out)], [min(out), max(out)], 'r--', lw=3)

ax.set\_xlabel('Predicted')

ax.set\_ylabel('Actual')

plt.show()

----

+\*Out[174]:\*+

----

![png](output\_206\_0.png)

----

+\*In[175]:\*+

[source, ipython3]

----

# model evaluation for testing set

mae = mean\_absolute\_error(test\_y, pred\_y)

mse = mean\_squared\_error(test\_y, pred\_y)

r2 = r2\_score(test\_y, pred\_y)

adj\_rsqrd = 1 - (1-r2)\*(len(test\_y) - 1) / (len(test\_y) - (test\_X.shape[1] - 1) - 1)

print('Mean Absolute Error is {}'.format(round(mae, 3)))

print('Mean Squared Error is {}'.format(round(mse, 3)))

print('Root Mean Squared Error is {}'.format(round(mse\*\*(0.5), 3)))

print('R-Square score is {}'.format(round(r2, 3)))

print('Adjusted R-Square score is {}'.format(round(adj\_rsqrd, 3)))

----

+\*Out[175]:\*+

----

Mean Absolute Error is 44.819

Mean Squared Error is 5141.608

Root Mean Squared Error is 71.705

R-Square score is 0.987

Adjusted R-Square score is 0.987

----

== Values are still intact. Target variable(Dependent Variable) is highly depending on Independent Variables. hence model accuracy is successfully established even after removing multicollinear variables.

