tool-time-part-2

April 18, 2024

[137]:
```
import pandas as pd
import numpy as np
import importlib
```

[138]:
```
import plotly
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
plotly.__version__
```

[138]: '5.20.0'

[139]:
```
init_notebook_mode(connected=True)
pd.set_option('display.max_rows', 4000)
pd.set_option('display.max_columns', 500)
```

[140]:
```
HDLo = pd.read_csv('C:/Users/shett/Downloads/data/Home_Depot_Lowes_Data.csv')
region = pd.read_csv('C:/Users/shett/Downloads/data/state_region.csv')
Property_tax = pd.read_csv('C:/Users/shett/Downloads/data/Property_Tax.csv')
highways = pd.read_csv('C:/Users/shett/Downloads/data/highways.csv')
```

[141]:
```
merge = pd.merge(HDLo,region, left_on='state', right_on='State Code',␣
↪how='inner')
HDLo_NE = merge[merge.Division == 'New England']
len(HDLo_NE)
```

[141]: 67

[142]:
```
#HDLo_NE with Highways
HDLo_NE = pd.merge(HDLo_NE, highways, how='left', left_on=['areaname','state'],␣
↪right_on = ['County','state'])
HDLo_NE.head()
```

[142]:

	areaname	county	state	r1	r2	Lcount	HDcount	pop_2000	pop_2010	\
0	Fairfield	9001	CT	1	1	1	6	882567.0	916829.0	
1	Hartford	9003	CT	1	1	5	9	857183.0	894014.0	
2	Litchfield	9005	CT	1	1	1	2	182193.0	189927.0	
3	Middlesex	9007	CT	1	1	1	1	155071.0	165676.0	
4	New Haven	9009	CT	1	1	5	7	824008.0	862477.0	

```
   income_2000  income_2010  pct_U18_2000  pct_U18_2010  pctcollege_2000  \
0        77690       100179          25.6          24.8             39.9
1        62144        78826          24.6          22.8             29.6
2        66445        84422          24.6          21.6             27.5
3        71319        90666          23.2          21.2             33.8
4        60549        77451          24.5          22.4             27.6

   pctcollege_2010  ownhome_2000  ownhome_2010  density_2000  density_2010  \
0             43.4          69.2          68.6        1409.9        1467.2
1             32.8          64.2          65.5        1166.2        1216.2
2             32.4          75.2          76.3         198.0         206.3
3             36.8          72.1          74.4         420.2         448.6
4             32.0          63.1          63.4        1359.7        1426.7

   pctwhite_2000  pctwhite_2010  pctblack_2000  pctblack_2010        State  \
0           79.3           74.8           11.3           10.8  Connecticut
1           76.9           72.4           13.9           13.3  Connecticut
2           95.8           93.9            1.4            1.3  Connecticut
3           91.3           89.2            5.0            4.7  Connecticut
4           79.4           74.8           13.3           12.7  Connecticut

  State Code     Region     Division      County  Interstate Highways  \
0         CT  Northeast  New England   Fairfield                    1
1         CT  Northeast  New England    Hartford                    1
2         CT  Northeast  New England  Litchfield                    0
3         CT  Northeast  New England   Middlesex                    1
4         CT  Northeast  New England   New Haven                    2

   U.S Highways  Toll Roads
0             2           0
1             5           0
2             4           0
3             2           0
4             2           0
```

[143]:

```python
#merge of property tax
HDLo_NE = pd.merge(HDLo_NE, Property_tax, how='left',␣
 ↪left_on=['areaname','state'], right_on = ['County','state'])
HDLo_NE = pd.DataFrame(HDLo_NE)
```

[144]:

```python
#remove collinearity
HDLo_NE['population'] = abs(HDLo_NE['pop_2010'] - HDLo_NE['pop_2000'])
HDLo_NE['income'] = abs(HDLo_NE['income_2010'] - HDLo_NE['income_2000'])
HDLo_NE['pct_U18'] = abs(HDLo_NE['pct_U18_2010'] - HDLo_NE['pct_U18_2000'])
HDLo_NE['pctcollege'] = abs(HDLo_NE['pctcollege_2010'] -␣
 ↪HDLo_NE['pctcollege_2000'])
```

```
HDLo_NE['ownhome'] = abs(HDLo_NE['ownhome_2010'] - HDLo_NE['ownhome_2000'])
HDLo_NE['density'] = abs(HDLo_NE['density_2010'] - HDLo_NE['density_2000'])
HDLo_NE['pctwhite'] = abs(HDLo_NE['pctwhite_2010'] - HDLo_NE['pctwhite_2000'])
HDLo_NE['pctblack'] = abs(HDLo_NE['pctblack_2010'] - HDLo_NE['pctblack_2000'])
```

[145]:
```
#remove *_2010 and *_2000 fields
HDLo_NE = HDLo_NE[HDLo_NE.columns.drop(list(HDLo_NE.filter(regex=('(.*_2010|.
↪*_2000)'))))]
```

[146]:
```
#stats on the columns
HDLo_NE.describe()
```

[146]:
```
               county    r1    r2     Lcount    HDcount  Interstate Highways  \
count       67.000000  67.0  67.0  67.000000  67.000000            67.000000
mean     30460.253731   1.0   1.0   1.104478   1.716418             1.014925
std      13095.644716   0.0   0.0   1.558305   2.165892             1.007435
min       9001.000000   1.0   1.0   0.000000   0.000000             0.000000
25%      23018.000000   1.0   1.0   0.000000   0.000000             0.000000
50%      25019.000000   1.0   1.0   0.000000   1.000000             1.000000
75%      44004.000000   1.0   1.0   2.000000   2.500000             2.000000
max      50027.000000   1.0   1.0   6.000000   9.000000             3.000000

       U.S Highways  Toll Roads  Median Home Value  \
count     67.000000   67.000000          67.000000
mean       1.447761    0.238806      245101.492537
std        1.034021    0.429572      129118.171728
min        0.000000    0.000000       95800.000000
25%        1.000000    0.000000      189500.000000
50%        1.000000    0.000000      215800.000000
75%        2.000000    0.000000      264850.000000
max        5.000000    1.000000      966600.000000

       Median Annual Property Tax Payment  \
count                           67.000000
mean                          3630.761194
std                           1269.875884
min                           1339.000000
25%                           2837.000000
50%                           3669.000000
75%                           4448.000000
max                           7057.000000

       Average Effective Property Tax Rate    population        income  \
count                            67.000000     67.000000     67.000000
mean                              0.015909   7841.388060  16056.134328
std                               0.004555  10516.675451   4443.460152
min                               0.003200      0.000000   7467.000000
```

```
25%                               0.012900    943.000000  12961.000000
50%                               0.016100   3292.000000  15380.000000
75%                               0.018500  10304.500000  19044.000000
max                               0.026400  47589.000000  30214.000000

         pct_U18  pctcollege    ownhome     density   pctwhite   pctblack
count  67.000000   67.000000  67.000000   67.000000  67.000000  67.000000
mean    3.038806    3.623881   1.073134   25.385075   1.995522   0.116418
std     0.970154    1.335992   0.944615   88.542438   1.440058   0.191957
min     0.700000    0.400000   0.100000    0.000000   0.200000   0.000000
25%     2.600000    2.800000   0.400000    1.700000   1.000000   0.000000
50%     3.000000    3.800000   0.800000    6.100000   1.400000   0.000000
75%     3.550000    4.400000   1.500000   24.200000   2.550000   0.100000
max     6.600000    7.100000   4.800000  724.100000   6.700000   1.100000
```

[147]:
```python
#ensure there are no empty values
HDLo_NE.isnull().values.any()
```

[147]: False

[148]:
```python
#Question 1
#1. Create dummy variables to identify if HomeDepot or Lowes is present in that␣
↪county
## reference https://towardsdatascience.com/
↪the-dummys-guide-to-creating-dummy-variables-f21faddb1d40
HDLo_NE['HD_present'] = np.where(HDLo_NE['HDcount'] > 0, 1, 0)
HDLo_NE['Lo_present'] = np.where(HDLo_NE['Lcount'] > 0, 1,0)
HDLo_NE['Store_present'] = np.where(HDLo_NE['Lcount'] > 0,1, np.
↪where(HDLo_NE['HDcount'] > 0,1,0))
```

[149]:
```python
from ydata_profiling import ProfileReport
```

[150]:
```python
ProfileReport(HDLo_NE)
```

```
Summarize dataset:   0%|          | 0/5 [00:00<?, ?it/s]

Generate report structure:   0%|          | 0/1 [00:00<?, ?it/s]

Render HTML:   0%|          | 0/1 [00:00<?, ?it/s]

<IPython.core.display.HTML object>
```

[150]:

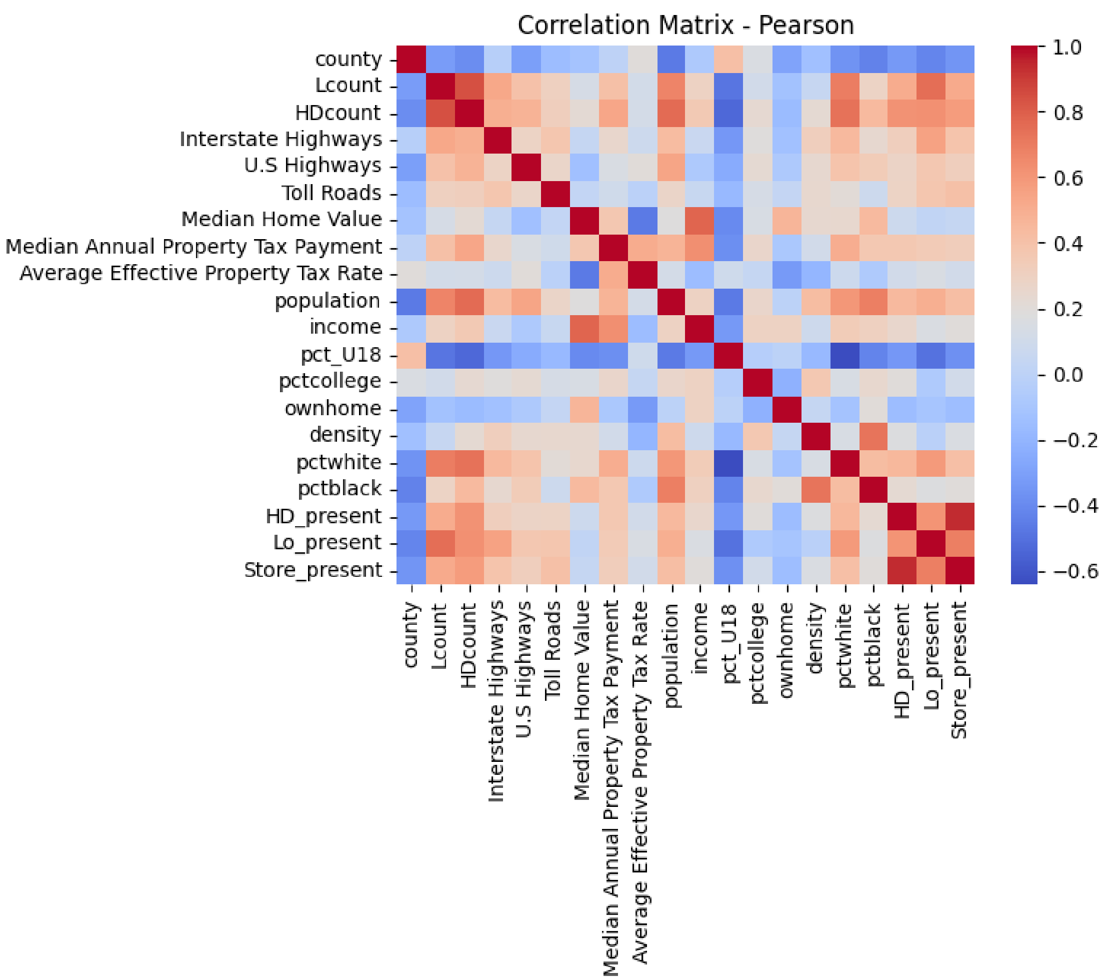
[151]:
```python
HDLo_NE_Corr = HDLo_NE[['county', 'Lcount', 'HDcount','Interstate Highways',
       'U.S Highways', 'Toll Roads', 'Median Home Value',
       'Median Annual Property Tax Payment',
       'Average Effective Property Tax Rate', 'population', 'income',
       'pct_U18', 'pctcollege', 'ownhome', 'density', 'pctwhite', 'pctblack',
```

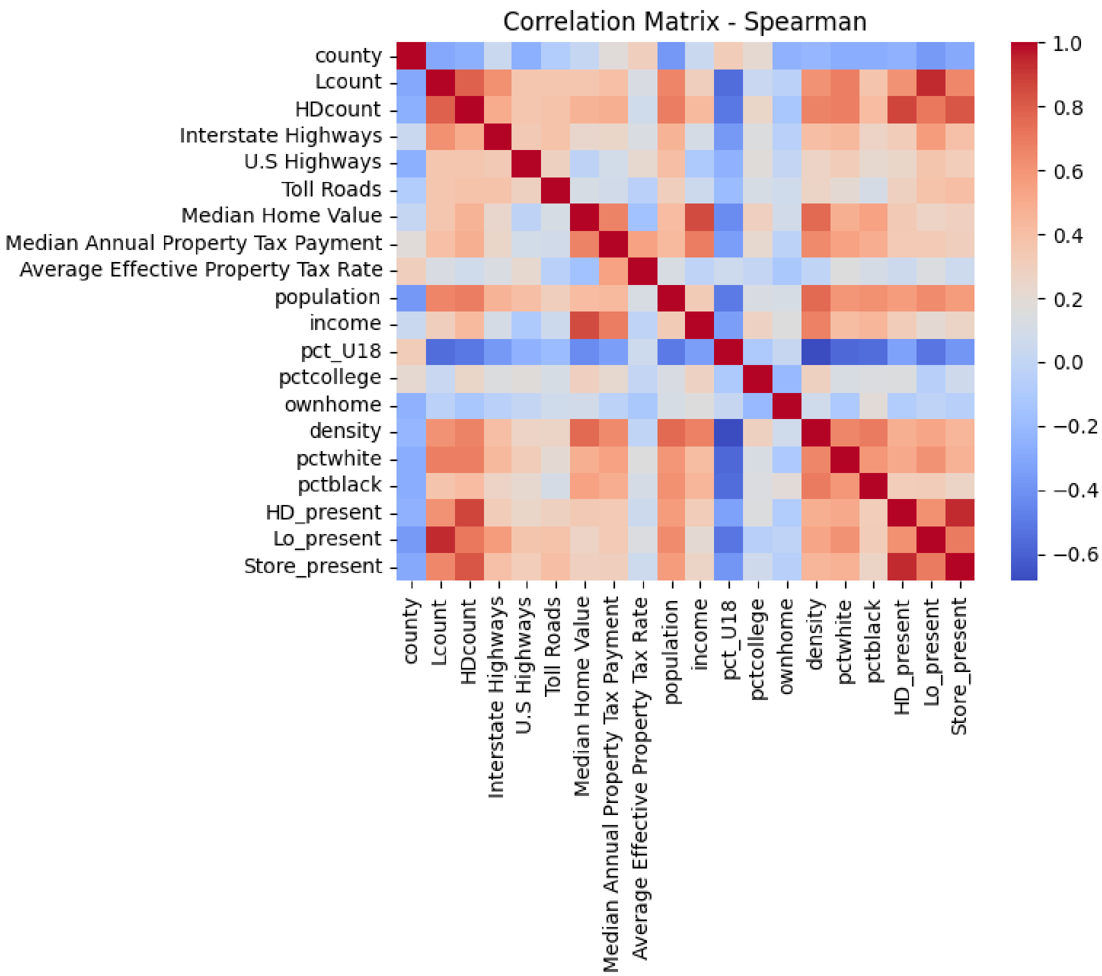
```
       'HD_present', 'Lo_present', 'Store_present']]
correlation_matrix_pearson = HDLo_NE_Corr.corr(method='pearson')
correlation_matrix_spearman = HDLo_NE_Corr.corr(method='spearman')

import seaborn as sns
import matplotlib.pyplot as plt

sns.heatmap(correlation_matrix_pearson, annot=False, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Matrix - Pearson")
plt.show()

sns.heatmap(correlation_matrix_spearman, annot=False, cmap='coolwarm', fmt=".
 ↪2f")
plt.title("Correlation Matrix - Spearman")
plt.show()
```





[152]:
```
HDLo_NE[:5]
```

[152]:

	areaname	county	state	r1	r2	Lcount	HDcount	State	State Code	\
0	Fairfield	9001	CT	1	1	1	6	Connecticut	CT	
1	Hartford	9003	CT	1	1	5	9	Connecticut	CT	
2	Litchfield	9005	CT	1	1	1	2	Connecticut	CT	
3	Middlesex	9007	CT	1	1	1	1	Connecticut	CT	
4	New Haven	9009	CT	1	1	5	7	Connecticut	CT	

	Region	Division	County_x	Interstate Highways	U.S Highways	\
0	Northeast	New England	Fairfield	1	2	
1	Northeast	New England	Hartford	1	5	
2	Northeast	New England	Litchfield	0	4	
3	Northeast	New England	Middlesex	1	2	
4	Northeast	New England	New Haven	2	2	

```
   Toll Roads    County_y  Median Home Value  \
0           0   Fairfield             413400
1           0    Hartford             234900
2           0  Litchfield             249500
3           0   Middlesex             283800
4           0   New Haven             244000

   Median Annual Property Tax Payment  Average Effective Property Tax Rate  \
0                                7057                               0.0171
1                                5035                               0.0214
2                                4639                               0.0186
3                                5298                               0.0187
4                                5486                               0.0225

   population  income  pct_U18  pctcollege  ownhome  density  pctwhite  \
0     34262.0   22489      0.8         3.5      0.6     57.3       4.5
1     36831.0   16682      1.8         3.2      1.3     50.0       4.5
2      7734.0   17977      3.0         4.9      1.1      8.3       1.9
3     10605.0   19347      2.0         3.0      2.3     28.4       2.1
4     38469.0   16902      2.1         4.4      0.3     67.0       4.6

   pctblack  HD_present  Lo_present  Store_present
0       0.5           1           1              1
1       0.6           1           1              1
2       0.1           1           1              1
3       0.3           1           1              1
4       0.6           1           1              1
```

[153]:
```
correlation_matrix_spearman
```

[153]:
```
                                       county    Lcount   HDcount  \
county                               1.000000 -0.301158 -0.261994
Lcount                              -0.301158  1.000000  0.784832
HDcount                             -0.261994  0.784832  1.000000
Interstate Highways                  0.041474  0.613492  0.498809
U.S Highways                        -0.256219  0.358264  0.361506
Toll Roads                          -0.079647  0.360025  0.379462
Median Home Value                    0.018756  0.357390  0.467380
Median Annual Property Tax Payment   0.180204  0.401103  0.481583
Average Effective Property Tax Rate  0.297255  0.126710  0.079112
population                          -0.375848  0.664338  0.685781
income                               0.045574  0.303131  0.423973
pct_U18                              0.320518 -0.556485 -0.506080
pctcollege                           0.217483  0.037873  0.246832
ownhome                             -0.251501 -0.037538 -0.127321
density                             -0.216622  0.607229  0.671242
pctwhite                            -0.270296  0.679246  0.679314
```

pctblack	-0.271004	0.371954	0.418589
HD_present	-0.255323	0.608749	0.871572
Lo_present	-0.361552	0.940384	0.708069
Store_present	-0.290956	0.650105	0.816762

	Interstate Highways	U.S Highways \
county	0.041474	-0.256219
Lcount	0.613492	0.358264
HDcount	0.498809	0.361506
Interstate Highways	1.000000	0.336786
U.S Highways	0.336786	1.000000
Toll Roads	0.377585	0.288754
Median Home Value	0.234511	-0.020572
Median Annual Property Tax Payment	0.249112	0.092393
Average Effective Property Tax Rate	0.134944	0.219455
population	0.464271	0.403533
income	0.107880	-0.103073
pct_U18	-0.368248	-0.243665
pctcollege	0.149913	0.175764
ownhome	-0.040953	0.010950
density	0.402246	0.266531
pctwhite	0.431107	0.321978
pctblack	0.264199	0.218438
HD_present	0.319414	0.250282
Lo_present	0.571338	0.368051
Store_present	0.392139	0.312365

	Toll Roads	Median Home Value \
county	-0.079647	0.018756
Lcount	0.360025	0.357390
HDcount	0.379462	0.467380
Interstate Highways	0.377585	0.234511
U.S Highways	0.288754	-0.020572
Toll Roads	1.000000	0.115851
Median Home Value	0.115851	1.000000
Median Annual Property Tax Payment	0.082364	0.670351
Average Effective Property Tax Rate	-0.038924	-0.161809
population	0.296868	0.414957
income	0.059736	0.854857
pct_U18	-0.194727	-0.428741
pctcollege	0.114971	0.290570
ownhome	0.074301	0.081084
density	0.260670	0.757378
pctwhite	0.208325	0.486396
pctblack	0.110813	0.540721
HD_present	0.287370	0.338303
Lo_present	0.375526	0.267301

```
Store_present                              0.404960          0.292581

                                    Median Annual Property Tax Payment  \
county                                                        0.180204
Lcount                                                        0.401103
HDcount                                                       0.481583
Interstate Highways                                           0.249112
U.S Highways                                                  0.092393
Toll Roads                                                    0.082364
Median Home Value                                             0.670351
Median Annual Property Tax Payment                            1.000000
Average Effective Property Tax Rate                           0.546300
population                                                    0.422504
income                                                        0.687232
pct_U18                                                      -0.346758
pctcollege                                                    0.221956
ownhome                                                      -0.029485
density                                                       0.642189
pctwhite                                                      0.542614
pctblack                                                      0.493366
HD_present                                                    0.335114
Lo_present                                                    0.331426
Store_present                                                 0.299086

                                    Average Effective Property Tax Rate  \
county                                                         0.297255
Lcount                                                         0.126710
HDcount                                                        0.079112
Interstate Highways                                            0.134944
U.S Highways                                                   0.219455
Toll Roads                                                    -0.038924
Median Home Value                                             -0.161809
Median Annual Property Tax Payment                             0.546300
Average Effective Property Tax Rate                            1.000000
population                                                     0.121377
income                                                        -0.017642
pct_U18                                                        0.062587
pctcollege                                                     0.009412
ownhome                                                       -0.115670
density                                                       -0.010288
pctwhite                                                       0.159504
pctblack                                                       0.111281
HD_present                                                     0.053466
Lo_present                                                     0.141397
Store_present                                                  0.060964

                                    population    income   pct_U18  \
```

county	-0.375848	0.045574	0.320518
Lcount	0.664338	0.303131	-0.556485
HDcount	0.685781	0.423973	-0.506080
Interstate Highways	0.464271	0.107880	-0.368248
U.S Highways	0.403533	-0.103073	-0.243665
Toll Roads	0.296868	0.059736	-0.194727
Median Home Value	0.414957	0.854857	-0.428741
Median Annual Property Tax Payment	0.422504	0.687232	-0.346758
Average Effective Property Tax Rate	0.121377	-0.017642	0.062587
population	1.000000	0.329037	-0.499785
income	0.329037	1.000000	-0.341983
pct_U18	-0.499785	-0.341983	1.000000
pctcollege	0.127812	0.273166	-0.098751
ownhome	0.112526	0.163745	0.022209
density	0.757717	0.672274	-0.682991
pctwhite	0.598640	0.412732	-0.571887
pctblack	0.615512	0.448508	-0.556095
HD_present	0.568093	0.320749	-0.334544
Lo_present	0.635034	0.208588	-0.520283
Store_present	0.572159	0.258447	-0.380618

	pctcollege	ownhome	density	pctwhite \
county	0.217483	-0.251501	-0.216622	-0.270296
Lcount	0.037873	-0.037538	0.607229	0.679246
HDcount	0.246832	-0.127321	0.671242	0.679314
Interstate Highways	0.149913	-0.040953	0.402246	0.431107
U.S Highways	0.175764	0.010950	0.266531	0.321978
Toll Roads	0.114971	0.074301	0.260670	0.208325
Median Home Value	0.290570	0.081084	0.757378	0.486396
Median Annual Property Tax Payment	0.221956	-0.029485	0.642189	0.542614
Average Effective Property Tax Rate	0.009412	-0.115670	-0.010288	0.159504
population	0.127812	0.112526	0.757717	0.598640
income	0.273166	0.163745	0.672274	0.412732
pct_U18	-0.098751	0.022209	-0.682991	-0.571887
pctcollege	1.000000	-0.205401	0.289528	0.121117
ownhome	-0.205401	1.000000	0.073264	-0.106393
density	0.289528	0.073264	1.000000	0.652968
pctwhite	0.121117	-0.106393	0.652968	1.000000
pctblack	0.143612	0.188303	0.694399	0.589209
HD_present	0.151631	-0.079080	0.482729	0.508633
Lo_present	-0.053317	-0.017789	0.531523	0.610770
Store_present	0.068284	-0.052073	0.451072	0.470921

	pctblack	HD_present	Lo_present \
county	-0.271004	-0.255323	-0.361552
Lcount	0.371954	0.608749	0.940384
HDcount	0.418589	0.871572	0.708069

```
Interstate Highways                 0.264199    0.319414    0.571338
U.S Highways                        0.218438    0.250282    0.368051
Toll Roads                          0.110813    0.287370    0.375526
Median Home Value                   0.540721    0.338303    0.267301
Median Annual Property Tax Payment  0.493366    0.335114    0.331426
Average Effective Property Tax Rate 0.111281    0.053466    0.141397
population                          0.615512    0.568093    0.635034
income                              0.448508    0.320749    0.208588
pct_U18                            -0.556095   -0.334544   -0.520283
pctcollege                          0.143612    0.151631   -0.053317
ownhome                             0.188303   -0.079080   -0.017789
density                             0.694399    0.482729    0.531523
pctwhite                            0.589209    0.508633    0.610770
pctblack                            1.000000    0.311060    0.321096
HD_present                          0.311060    1.000000    0.614145
Lo_present                          0.321096    0.614145    1.000000
Store_present                       0.259238    0.937114    0.691319

                                    Store_present
county                                  -0.290956
Lcount                                   0.650105
HDcount                                  0.816762
Interstate Highways                      0.392139
U.S Highways                             0.312365
Toll Roads                               0.404960
Median Home Value                        0.292581
Median Annual Property Tax Payment       0.299086
Average Effective Property Tax Rate      0.060964
population                               0.572159
income                                   0.258447
pct_U18                                 -0.380618
pctcollege                               0.068284
ownhome                                 -0.052073
density                                  0.451072
pctwhite                                 0.470921
pctblack                                 0.259238
HD_present                               0.937114
Lo_present                               0.691319
Store_present                            1.000000
```

[154]:

```
#create feature for all highway count by county
HDLo_NE['highway_count'] = HDLo_NE['U.S Highways'] + HDLo_NE['Toll Roads'] +
    HDLo_NE['Interstate Highways']
```

[155]:

```
#prediction for HD stores
import pandas as pd
import statsmodels.api as sm
```

```
HDLo = HDLo.dropna()
X = HDLo_NE.filter(['highway_count', 'Median Home Value' , 'Median Annual␣
↪Property Tax Payment' , 'Average Effective Property Tax Rate' ,
                    'population' , 'income' , 'pct_U18' , 'ownhome' ,␣
↪'density'])
y = HDLo_NE.filter(['Store_present'])

from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.feature_selection import RFE
from sklearn.linear_model import LogisticRegression
from sklearn import metrics
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
#from mlxtend.plotting import plot_confusion_matrix

import matplotlib.pyplot as plt
from sklearn.metrics import classification_report
import joblib
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4,␣
↪random_state=0)

logreg = LogisticRegression()
logreg.fit(X_train, y_train)

predictions = logreg.predict(X_test)
accuracy = accuracy_score(y_test, predictions)

#scoring the model
print('Accuracy score: ')
print(accuracy)
print(classification_report(y_test, predictions))

#Get ALL results from the model
import statsmodels.api as sm
X2 = sm.add_constant(X)
est = sm.OLS(y, X2)
est2 = est.fit()
print(est2.summary())

#cross validation matrix of accuracy
confusion_matrix = confusion_matrix(y_test, predictions)
print(confusion_matrix)
print('#TRUE NEGATIVE | FALSE POSITIVE')
print('FALSE NEGATIVE | #TRUE POSITIVE')
# fig, ax = plot_confusion_matrix(conf_mat=confusion_matrix)
# plt.show()
```

```python
#save model to use on HD's without store
filename = 'Store_log_predic.sav'
joblib.dump(logreg, filename)
```

```
Accuracy score:
0.6666666666666666
              precision    recall  f1-score   support

           0       0.62      0.45      0.53        11
           1       0.68      0.81      0.74        16

    accuracy                           0.67        27
   macro avg       0.65      0.63      0.63        27
weighted avg       0.66      0.67      0.65        27

                            OLS Regression Results
==============================================================================
Dep. Variable:          Store_present   R-squared:                       0.344
Model:                            OLS   Adj. R-squared:                  0.240
Method:                 Least Squares   F-statistic:                     3.321
Date:                Thu, 18 Apr 2024   Prob (F-statistic):            0.00253
Time:                        10:27:11   Log-Likelihood:                -31.041
No. Observations:                  67   AIC:                             82.08
Df Residuals:                      57   BIC:                             104.1
Df Model:                           9
Covariance Type:            nonrobust
================================================================================
=======================
                                         coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-----------------------
const                                  0.6336      0.472      1.343      0.185
-0.311       1.578
highway_count                          0.0877      0.037      2.347      0.022
0.013       0.163
Median Home Value                  -1.314e-06    9.6e-07     -1.369      0.176
-3.24e-06    6.08e-07
Median Annual Property Tax Payment  7.573e-05      0.000      0.659      0.512
-0.000       0.000
Average Effective Property Tax Rate  -19.5028     26.151     -0.746      0.459
-71.869      32.864
population                          2.497e-06   7.43e-06      0.336      0.738
-1.24e-05    1.74e-05
income                              2.746e-05    2.6e-05      1.056      0.295
-2.46e-05    7.95e-05
```

```
pct_U18                                 -0.0918      0.069     -1.322      0.192
-0.231       0.047
ownhome                                 -0.0335      0.068     -0.489      0.627
-0.171       0.104
density                              -7.929e-05      0.001     -0.110      0.913
-0.002       0.001
==============================================================================
Omnibus:                       12.546   Durbin-Watson:                   1.959
Prob(Omnibus):                  0.002   Jarque-Bera (JB):                3.623
Skew:                          -0.144   Prob(JB):                        0.163
Kurtosis:                       1.898   Cond. No.                     1.42e+08
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.42e+08. This might indicate that there are
strong multicollinearity or other numerical problems.
[[ 5  6]
 [ 3 13]]
#TRUE NEGATIVE | FALSE POSITIVE
FALSE NEGATIVE | #TRUE POSITIVE
```

```
C:\Users\shett\AppData\Local\Programs\Python\Python311\Lib\site-
packages\sklearn\utils\validation.py:1300: DataConversionWarning:

A column-vector y was passed when a 1d array was expected. Please change the
shape of y to (n_samples, ), for example using ravel().

C:\Users\shett\AppData\Local\Programs\Python\Python311\Lib\site-
packages\sklearn\linear_model\_logistic.py:469: ConvergenceWarning:

lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
```

[155]: ['Store_log_predic.sav']

[156]:
```
loaded_model = joblib.load(filename)

#filter and then predict on stores where HD is not present
Store_no = HDLo_NE[(HDLo_NE.Store_present == 0)]
```

```
Store_X = Store_no.filter(['highway_count', 'Median Home Value' , 'Median␣
↪Annual Property Tax Payment' , 'Average Effective Property Tax Rate' ,
'population' , 'income' , 'pct_U18' , 'ownhome' , 'density'])
Store_predict = loaded_model.predict_proba(Store_X)
Store_predict = pd.DataFrame(Store_predict)
Store_predict = Store_predict.round(4)
New_stores = pd.concat([Store_predict.reset_index(drop=True), Store_no.
↪reset_index(drop=True)], axis=1)
New_stores = New_stores.sort_values(by=[1], ascending=False)
New_stores
```

[156]:

	0	1	areaname	county	state	r1	r2	Lcount	HDcount	\
9	0.0000	1.0000	Nantucket	25019	MA	1	1	0	0	
8	0.0021	0.9979	Dukes	25007	MA	1	1	0	0	
13	0.1436	0.8564	Caledonia	50005	VT	1	1	0	0	
6	0.2597	0.7403	Waldo	23027	ME	1	1	0	0	
3	0.2786	0.7214	Oxford	23017	ME	1	1	0	0	
0	0.2900	0.7100	Tolland	9013	CT	1	1	0	0	
2	0.3003	0.6997	Lincoln	23015	ME	1	1	0	0	
18	0.3340	0.6660	Orange	50017	VT	1	1	0	0	
22	0.3713	0.6287	Windsor	50027	VT	1	1	0	0	
15	0.3901	0.6099	Franklin	50011	VT	1	1	0	0	
20	0.4275	0.5725	Washington	50023	VT	1	1	0	0	
21	0.4428	0.5572	Windham	50025	VT	1	1	0	0	
11	0.4795	0.5205	Bristol	44001	RI	1	1	0	0	
5	0.5296	0.4704	Somerset	23025	ME	1	1	0	0	
19	0.5594	0.4406	Orleans	50019	VT	1	1	0	0	
16	0.6455	0.3545	Grand Isle	50013	VT	1	1	0	0	
7	0.6865	0.3135	Washington	23029	ME	1	1	0	0	
1	0.7226	0.2774	Franklin	23007	ME	1	1	0	0	
17	0.7304	0.2696	Lamoille	50015	VT	1	1	0	0	
10	0.7340	0.2660	Coos	33007	NH	1	1	0	0	
12	0.7555	0.2445	Addison	50001	VT	1	1	0	0	
4	0.8067	0.1933	Piscataquis	23021	ME	1	1	0	0	
14	0.8830	0.1170	Essex	50009	VT	1	1	0	0	

	State	State Code	Region	Division	County_x	\
9	Massachusetts	MA	Northeast	New England	Nantucket	
8	Massachusetts	MA	Northeast	New England	Dukes	
13	Vermont	VT	Northeast	New England	Caledonia	
6	Maine	ME	Northeast	New England	Waldo	
3	Maine	ME	Northeast	New England	Oxford	
0	Connecticut	CT	Northeast	New England	Tolland	
2	Maine	ME	Northeast	New England	Lincoln	
18	Vermont	VT	Northeast	New England	Orange	
22	Vermont	VT	Northeast	New England	Windsor	

15	Vermont	VT	Northeast	New England	Franklin
20	Vermont	VT	Northeast	New England	Washington
21	Vermont	VT	Northeast	New England	Windham
11	Rhode Island	RI	Northeast	New England	Bristol
5	Maine	ME	Northeast	New England	Somerset
19	Vermont	VT	Northeast	New England	Orleans
16	Vermont	VT	Northeast	New England	Grand Isle
7	Maine	ME	Northeast	New England	Washington
1	Maine	ME	Northeast	New England	Franklin
17	Vermont	VT	Northeast	New England	Lamoille
10	New Hampshire	NH	Northeast	New England	Coos
12	Vermont	VT	Northeast	New England	Addison
4	Maine	ME	Northeast	New England	Piscataquis
14	Vermont	VT	Northeast	New England	Essex

	Interstate Highways	U.S Highways	Toll Roads	County_y \
9	0	0	0	Nantucket
8	0	0	0	Dukes
13	2	2	0	Caledonia
6	0	2	0	Waldo
3	0	2	0	Oxford
0	0	1	0	Tolland
2	1	1	0	Lincoln
18	2	2	0	Orange
22	2	1	0	Windsor
15	1	1	0	Franklin
20	1	2	0	Washington
21	1	1	0	Windham
11	0	0	0	Bristol
5	0	1	0	Somerset
19	1	1	0	Orleans
16	0	1	0	Grand Isle
7	0	1	0	Washington
1	0	0	0	Franklin
17	0	0	0	Lamoille
10	0	2	0	Coos
12	0	1	0	Addison
4	0	0	0	Piscataquis
14	0	1	0	Essex

	Median Home Value	Median Annual Property Tax Payment \
9	966600	3112
8	656000	3669
13	164200	3005
6	158400	1911
3	132900	1715
0	247800	5133

2	209700	2109
18	191700	3534
22	215800	4231
15	206500	3326
20	213200	4065
21	209500	4110
11	330000	5099
5	107100	1408
19	156800	2658
16	259600	4129
7	107400	1396
1	132400	1591
17	220300	3860
10	122000	2856
12	236400	4238
4	106800	1339
14	123800	2098

	Average Effective Property Tax Rate	population	income	pct_U18	\
9	0.0032	652.0	30214	1.5	
8	0.0056	1548.0	22797	3.5	
13	0.0183	1525.0	8309	3.5	
6	0.0121	2506.0	9610	3.2	
3	0.0129	3078.0	8815	2.9	
0	0.0207	16327.0	20192	2.9	
2	0.0101	841.0	13923	3.9	
18	0.0184	710.0	14658	4.7	
22	0.0196	748.0	13728	3.4	
15	0.0161	2329.0	14455	3.4	
20	0.0191	1495.0	16810	2.8	
21	0.0196	297.0	11272	3.6	
11	0.0155	773.0	24268	2.5	
5	0.0131	1340.0	7467	3.3	
19	0.0170	954.0	12802	3.8	
16	0.0159	69.0	17789	4.5	
7	0.0130	1085.0	11860	2.9	
1	0.0120	1301.0	11235	3.8	
17	0.0175	1242.0	17081	2.0	
10	0.0234	56.0	12649	3.9	
12	0.0179	847.0	17827	4.5	
4	0.0125	300.0	9933	4.2	
14	0.0169	153.0	14109	6.6	

	pctcollege	ownhome	density	pctwhite	pctblack	HD_present	Lo_present	\
9	3.0	4.6	27.9	0.2	0.4	0	0	
8	1.7	4.8	16.1	3.1	0.3	0	0	
13	5.4	0.7	2.5	1.0	0.1	0	0	

6	1.0	1.3	3.4	0.8	0.0	0	0
3	1.5	0.3	1.5	1.5	0.0	0	0
0	3.2	1.8	39.6	2.5	0.4	0	0
2	4.1	2.0	1.9	0.9	0.0	0	0
18	4.4	0.3	1.1	1.0	0.0	0	0
22	3.8	0.8	0.6	1.4	0.0	0	0
15	4.3	0.3	4.0	0.5	0.1	0	0
20	4.6	1.6	2.4	0.9	0.1	0	0
21	2.8	0.1	0.7	1.4	0.1	0	0
11	6.2	0.7	38.1	1.1	0.0	0	0
5	2.8	1.3	0.3	0.9	0.0	0	0
19	3.9	1.5	1.7	0.6	0.0	0	0
16	4.4	0.1	2.1	2.1	0.0	0	0
7	3.8	2.3	0.4	1.4	0.0	0	0
1	3.9	0.6	0.7	0.7	0.0	0	0
17	1.1	0.3	2.9	0.6	0.0	0	0
10	4.8	0.1	0.0	1.2	0.0	0	0
12	2.0	0.8	1.3	1.6	0.0	0	0
4	0.8	2.1	0.1	0.9	0.0	0	0
14	5.3	0.7	0.2	0.7	0.0	0	0

	Store_present	highway_count
9	0	0
8	0	0
13	0	4
6	0	2
3	0	2
0	0	1
2	0	2
18	0	4
22	0	3
15	0	2
20	0	3
21	0	2
11	0	0
5	0	1
19	0	2
16	0	1
7	0	1
1	0	0
17	0	0
10	0	2
12	0	1
4	0	0
14	0	1

[157]:
```
from sklearn.ensemble import RandomForestClassifier

clf=RandomForestClassifier(n_estimators=1000, max_depth=2, warm_start=True)
clf.fit(X_train,y_train)

y_pred2=clf.predict(X_test)
```

```
C:\Users\shett\AppData\Local\Programs\Python\Python311\Lib\site-
packages\sklearn\base.py:1474: DataConversionWarning:

A column-vector y was passed when a 1d array was expected. Please change the
shape of y to (n_samples,), for example using ravel().
```

[158]:
```
#Eval for RandomForestClassifier
from sklearn import metrics
from sklearn.metrics import confusion_matrix
print("Accuracy:", metrics.accuracy_score(y_test,y_pred2))
print(classification_report(y_test, y_pred2))
confusion = confusion_matrix(y_test, y_pred2)
print(confusion)
print('#TRUE NEGATIVE | FALSE POSITIVE')
print('FALSE NEGATIVE | #TRUE POSITIVE')
```

```
Accuracy: 0.8148148148148148
              precision    recall  f1-score   support

           0       0.80      0.73      0.76        11
           1       0.82      0.88      0.85        16

    accuracy                           0.81        27
   macro avg       0.81      0.80      0.81        27
weighted avg       0.81      0.81      0.81        27

[[ 8  3]
 [ 2 14]]
#TRUE NEGATIVE | FALSE POSITIVE
FALSE NEGATIVE | #TRUE POSITIVE
```

[159]:
```
#save model to use on HD's without store
filename = 'RandomForest_stores.sav'
joblib.dump(clf, filename)
```

[159]: ['RandomForest_stores.sav']

[160]:
```
loaded_model = joblib.load(filename)
#filter and then predict on stores where HD is not present
```

```
Store_no = HDLo_NE[(HDLo_NE.Store_present == 0)]
Store_X = Store_no.filter(['highway_count', 'Median Home Value' , 'Median␣
 ↪Annual Property Tax Payment' , 'Average Effective Property Tax Rate' ,
 'population' , 'income' , 'pct_U18' , 'ownhome' , 'density'])
Store_predict = loaded_model.predict_proba(Store_X)
Store_predict = pd.DataFrame(Store_predict)
Store_predict = Store_predict.round(4)
New_stores = pd.concat([Store_predict.reset_index(drop=True), Store_no.
 ↪reset_index(drop=True)], axis=
1)
New_stores = New_stores.sort_values(by=[1], ascending=False)
New_stores
```

[160]:

```
         0       1     areaname  county state  r1  r2  Lcount  HDcount  \
0   0.2272  0.7728      Tolland    9013    CT   1   1       0        0
9   0.3862  0.6138    Nantucket   25019    MA   1   1       0        0
8   0.4094  0.5906        Dukes   25007    MA   1   1       0        0
11  0.4132  0.5868      Bristol   44001    RI   1   1       0        0
15  0.5167  0.4833     Franklin   50011    VT   1   1       0        0
18  0.5261  0.4739       Orange   50017    VT   1   1       0        0
10  0.5491  0.4509         Coos   33007    NH   1   1       0        0
3   0.5642  0.4358       Oxford   23017    ME   1   1       0        0
14  0.5817  0.4183        Essex   50009    VT   1   1       0        0
16  0.5830  0.4170   Grand Isle   50013    VT   1   1       0        0
20  0.5973  0.4027   Washington   50023    VT   1   1       0        0
6   0.5974  0.4026        Waldo   23027    ME   1   1       0        0
17  0.6160  0.3840     Lamoille   50015    VT   1   1       0        0
21  0.6195  0.3805      Windham   50025    VT   1   1       0        0
22  0.6196  0.3804      Windsor   50027    VT   1   1       0        0
13  0.6222  0.3778    Caledonia   50005    VT   1   1       0        0
12  0.6225  0.3775      Addison   50001    VT   1   1       0        0
7   0.6331  0.3669   Washington   23029    ME   1   1       0        0
2   0.6676  0.3324      Lincoln   23015    ME   1   1       0        0
19  0.6735  0.3265      Orleans   50019    VT   1   1       0        0
4   0.6950  0.3050  Piscataquis   23021    ME   1   1       0        0
1   0.7120  0.2880     Franklin   23007    ME   1   1       0        0
5   0.7377  0.2623     Somerset   23025    ME   1   1       0        0

            State State Code     Region     Division     County_x  \
0     Connecticut         CT  Northeast  New England      Tolland
9   Massachusetts         MA  Northeast  New England    Nantucket
8   Massachusetts         MA  Northeast  New England        Dukes
11   Rhode Island         RI  Northeast  New England      Bristol
15        Vermont         VT  Northeast  New England     Franklin
18        Vermont         VT  Northeast  New England       Orange
10  New Hampshire         NH  Northeast  New England         Coos
3           Maine         ME  Northeast  New England       Oxford
```

```
14        Vermont         VT  Northeast  New England        Essex
16        Vermont         VT  Northeast  New England   Grand Isle
20        Vermont         VT  Northeast  New England   Washington
6           Maine         ME  Northeast  New England        Waldo
17        Vermont         VT  Northeast  New England     Lamoille
21        Vermont         VT  Northeast  New England      Windham
22        Vermont         VT  Northeast  New England      Windsor
13        Vermont         VT  Northeast  New England    Caledonia
12        Vermont         VT  Northeast  New England      Addison
7           Maine         ME  Northeast  New England   Washington
2           Maine         ME  Northeast  New England      Lincoln
19        Vermont         VT  Northeast  New England      Orleans
4           Maine         ME  Northeast  New England  Piscataquis
1           Maine         ME  Northeast  New England     Franklin
5           Maine         ME  Northeast  New England     Somerset

    Interstate Highways  U.S Highways  Toll Roads     County_y  \
0                     0             1           0      Tolland
9                     0             0           0    Nantucket
8                     0             0           0        Dukes
11                    0             0           0      Bristol
15                    1             1           0     Franklin
18                    2             2           0       Orange
10                    0             2           0         Coos
3                     0             2           0       Oxford
14                    0             1           0        Essex
16                    0             1           0   Grand Isle
20                    1             2           0   Washington
6                     0             2           0        Waldo
17                    0             0           0     Lamoille
21                    1             1           0      Windham
22                    2             1           0      Windsor
13                    2             2           0    Caledonia
12                    0             1           0      Addison
7                     0             1           0   Washington
2                     1             1           0      Lincoln
19                    1             1           0      Orleans
4                     0             0           0  Piscataquis
1                     0             0           0     Franklin
5                     0             1           0     Somerset

    Median Home Value  Median Annual Property Tax Payment  \
0              247800                                5133
9              966600                                3112
8              656000                                3669
11             330000                                5099
15             206500                                3326
```

18	191700	3534
10	122000	2856
3	132900	1715
14	123800	2098
16	259600	4129
20	213200	4065
6	158400	1911
17	220300	3860
21	209500	4110
22	215800	4231
13	164200	3005
12	236400	4238
7	107400	1396
2	209700	2109
19	156800	2658
4	106800	1339
1	132400	1591
5	107100	1408

	Average Effective Property Tax Rate	population	income	pct_U18	\
0	0.0207	16327.0	20192	2.9	
9	0.0032	652.0	30214	1.5	
8	0.0056	1548.0	22797	3.5	
11	0.0155	773.0	24268	2.5	
15	0.0161	2329.0	14455	3.4	
18	0.0184	710.0	14658	4.7	
10	0.0234	56.0	12649	3.9	
3	0.0129	3078.0	8815	2.9	
14	0.0169	153.0	14109	6.6	
16	0.0159	69.0	17789	4.5	
20	0.0191	1495.0	16810	2.8	
6	0.0121	2506.0	9610	3.2	
17	0.0175	1242.0	17081	2.0	
21	0.0196	297.0	11272	3.6	
22	0.0196	748.0	13728	3.4	
13	0.0183	1525.0	8309	3.5	
12	0.0179	847.0	17827	4.5	
7	0.0130	1085.0	11860	2.9	
2	0.0101	841.0	13923	3.9	
19	0.0170	954.0	12802	3.8	
4	0.0125	300.0	9933	4.2	
1	0.0120	1301.0	11235	3.8	
5	0.0131	1340.0	7467	3.3	

	pctcollege	ownhome	density	pctwhite	pctblack	HD_present	Lo_present	\
0	3.2	1.8	39.6	2.5	0.4	0	0	
9	3.0	4.6	27.9	0.2	0.4	0	0	

8	1.7	4.8	16.1	3.1	0.3	0	0
11	6.2	0.7	38.1	1.1	0.0	0	0
15	4.3	0.3	4.0	0.5	0.1	0	0
18	4.4	0.3	1.1	1.0	0.0	0	0
10	4.8	0.1	0.0	1.2	0.0	0	0
3	1.5	0.3	1.5	1.5	0.0	0	0
14	5.3	0.7	0.2	0.7	0.0	0	0
16	4.4	0.1	2.1	2.1	0.0	0	0
20	4.6	1.6	2.4	0.9	0.1	0	0
6	1.0	1.3	3.4	0.8	0.0	0	0
17	1.1	0.3	2.9	0.6	0.0	0	0
21	2.8	0.1	0.7	1.4	0.1	0	0
22	3.8	0.8	0.6	1.4	0.0	0	0
13	5.4	0.7	2.5	1.0	0.1	0	0
12	2.0	0.8	1.3	1.6	0.0	0	0
7	3.8	2.3	0.4	1.4	0.0	0	0
2	4.1	2.0	1.9	0.9	0.0	0	0
19	3.9	1.5	1.7	0.6	0.0	0	0
4	0.8	2.1	0.1	0.9	0.0	0	0
1	3.9	0.6	0.7	0.7	0.0	0	0
5	2.8	1.3	0.3	0.9	0.0	0	0

	Store_present	highway_count
0	0	1
9	0	0
8	0	0
11	0	0
15	0	2
18	0	4
10	0	2
3	0	2
14	0	1
16	0	1
20	0	3
6	0	2
17	0	0
21	0	2
22	0	3
13	0	4
12	0	1
7	0	1
2	0	2
19	0	2
4	0	0
1	0	0
5	0	1

[161]:
```
#feature importance according to reandom forest classifier
feature_importance = pd.DataFrame(data=clf.feature_importances_, index=X_train.
 ↪columns.values, columns=['values'])
feature_importance.sort_values(['values'], ascending=False, inplace=True)
feature_importance
```

[161]:

	values
population	0.320684
density	0.195709
highway_count	0.094862
income	0.088037
Median Home Value	0.072051
pct_U18	0.069298
Median Annual Property Tax Payment	0.068037
ownhome	0.046559
Average Effective Property Tax Rate	0.044761

[162]:
```
#Decision Tree
from sklearn import tree
model = tree.DecisionTreeClassifier(criterion='entropy')
model.fit(X,y)
model.score(X,y)
#predict output
y_pred = model.predict(X_test)
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
from sklearn import tree
print("Accuracy is ", accuracy_score(y_test, y_pred))
confusion_mtrx = confusion_matrix(y_test, y_pred)
print(confusion_mtrx)
```

```
Accuracy is  1.0
[[11  0]
 [ 0 16]]
```

[163]:
```
#feature importance according to reandom forest classifier
feature_importance = pd.DataFrame(data=model.feature_importances_,␣
 ↪index=X_train.columns.values, columns=['values'])
feature_importance.sort_values(['values'], ascending=False, inplace=True)
feature_importance
```

[163]:

	values
population	0.511189
Median Home Value	0.173855
highway_count	0.104800
Average Effective Property Tax Rate	0.064339

```
Median Annual Property Tax Payment   0.058060
density                              0.047704
ownhome                              0.040054
income                               0.000000
pct_U18                              0.000000
```

[164]:
```
#save model to use on HD's without store
filename = 'Decsion_Tree_stores.sav'
joblib.dump(model, filename)
```

[164]:
```
['Decsion_Tree_stores.sav']
```

[165]:
```
loaded_model = joblib.load(filename)
#filter and then predict on stores where HD is not present
Store_no = HDLo_NE[(HDLo_NE.Store_present == 0)]
Store_X = Store_no.filter(['highway_count' , 'Median Home Value' , 'Median␣
 ↪Annual Property Tax Payment' , 'Average Effective Property Tax Rate' ,
 'population' , 'income' , 'pct_U18' , 'ownhome' , 'density'])
Store_predict = loaded_model.predict_proba(Store_X)
Store_predict = pd.DataFrame(Store_predict)
Store_predict = Store_predict.round(4)
New_stores = pd.concat([Store_predict.reset_index(drop=True), Store_no.
 ↪reset_index(drop=True)], axis=1)
New_stores = New_stores.sort_values(by=[1], ascending=False)
```

[166]:
```
New_stores
```

[166]:
```
      0    1     areaname  county state  r1  r2  Lcount  HDcount  \
0   1.0  0.0      Tolland    9013    CT   1   1       0        0
12  1.0  0.0      Addison   50001    VT   1   1       0        0
21  1.0  0.0      Windham   50025    VT   1   1       0        0
20  1.0  0.0   Washington   50023    VT   1   1       0        0
19  1.0  0.0      Orleans   50019    VT   1   1       0        0
18  1.0  0.0       Orange   50017    VT   1   1       0        0
17  1.0  0.0     Lamoille   50015    VT   1   1       0        0
16  1.0  0.0   Grand Isle   50013    VT   1   1       0        0
15  1.0  0.0     Franklin   50011    VT   1   1       0        0
14  1.0  0.0        Essex   50009    VT   1   1       0        0
13  1.0  0.0    Caledonia   50005    VT   1   1       0        0
11  1.0  0.0      Bristol   44001    RI   1   1       0        0
1   1.0  0.0     Franklin   23007    ME   1   1       0        0
10  1.0  0.0         Coos   33007    NH   1   1       0        0
9   1.0  0.0    Nantucket   25019    MA   1   1       0        0
8   1.0  0.0        Dukes   25007    MA   1   1       0        0
7   1.0  0.0   Washington   23029    ME   1   1       0        0
6   1.0  0.0        Waldo   23027    ME   1   1       0        0
5   1.0  0.0     Somerset   23025    ME   1   1       0        0
```

```
4   1.0  0.0  Piscataquis   23021    ME   1   1       0         0
3   1.0  0.0       Oxford   23017    ME   1   1       0         0
2   1.0  0.0      Lincoln   23015    ME   1   1       0         0
22  1.0  0.0      Windsor   50027    VT   1   1       0         0
```

```
             State State Code     Region     Division     County_x  \
0      Connecticut         CT  Northeast  New England      Tolland
12         Vermont         VT  Northeast  New England      Addison
21         Vermont         VT  Northeast  New England      Windham
20         Vermont         VT  Northeast  New England   Washington
19         Vermont         VT  Northeast  New England      Orleans
18         Vermont         VT  Northeast  New England       Orange
17         Vermont         VT  Northeast  New England     Lamoille
16         Vermont         VT  Northeast  New England   Grand Isle
15         Vermont         VT  Northeast  New England     Franklin
14         Vermont         VT  Northeast  New England        Essex
13         Vermont         VT  Northeast  New England    Caledonia
11    Rhode Island         RI  Northeast  New England      Bristol
1            Maine         ME  Northeast  New England     Franklin
10   New Hampshire         NH  Northeast  New England         Coos
9    Massachusetts         MA  Northeast  New England    Nantucket
8    Massachusetts         MA  Northeast  New England        Dukes
7            Maine         ME  Northeast  New England   Washington
6            Maine         ME  Northeast  New England        Waldo
5            Maine         ME  Northeast  New England     Somerset
4            Maine         ME  Northeast  New England  Piscataquis
3            Maine         ME  Northeast  New England       Oxford
2            Maine         ME  Northeast  New England      Lincoln
22         Vermont         VT  Northeast  New England      Windsor
```

```
    Interstate Highways  U.S Highways  Toll Roads     County_y  \
0                     0             1           0      Tolland
12                    0             1           0      Addison
21                    1             1           0      Windham
20                    1             2           0   Washington
19                    1             1           0      Orleans
18                    2             2           0       Orange
17                    0             0           0     Lamoille
16                    0             1           0   Grand Isle
15                    1             1           0     Franklin
14                    0             1           0        Essex
13                    2             2           0    Caledonia
11                    0             0           0      Bristol
1                     0             0           0     Franklin
10                    0             2           0         Coos
9                     0             0           0    Nantucket
8                     0             0           0        Dukes
```

```
7                    0            1          0   Washington
6                    0            2          0        Waldo
5                    0            1          0     Somerset
4                    0            0          0  Piscataquis
3                    0            2          0       Oxford
2                    1            1          0      Lincoln
22                   2            1          0      Windsor

    Median Home Value  Median Annual Property Tax Payment  \
0              247800                                5133
12             236400                                4238
21             209500                                4110
20             213200                                4065
19             156800                                2658
18             191700                                3534
17             220300                                3860
16             259600                                4129
15             206500                                3326
14             123800                                2098
13             164200                                3005
11             330000                                5099
1              132400                                1591
10             122000                                2856
9              966600                                3112
8              656000                                3669
7              107400                                1396
6              158400                                1911
5              107100                                1408
4              106800                                1339
3              132900                                1715
2              209700                                2109
22             215800                                4231

    Average Effective Property Tax Rate  population  income  pct_U18  \
0                                0.0207     16327.0   20192      2.9
12                               0.0179       847.0   17827      4.5
21                               0.0196       297.0   11272      3.6
20                               0.0191      1495.0   16810      2.8
19                               0.0170       954.0   12802      3.8
18                               0.0184       710.0   14658      4.7
17                               0.0175      1242.0   17081      2.0
16                               0.0159        69.0   17789      4.5
15                               0.0161      2329.0   14455      3.4
14                               0.0169       153.0   14109      6.6
13                               0.0183      1525.0    8309      3.5
11                               0.0155       773.0   24268      2.5
1                                0.0120      1301.0   11235      3.8
```

10	0.0234	56.0	12649	3.9
9	0.0032	652.0	30214	1.5
8	0.0056	1548.0	22797	3.5
7	0.0130	1085.0	11860	2.9
6	0.0121	2506.0	9610	3.2
5	0.0131	1340.0	7467	3.3
4	0.0125	300.0	9933	4.2
3	0.0129	3078.0	8815	2.9
2	0.0101	841.0	13923	3.9
22	0.0196	748.0	13728	3.4

	pctcollege	ownhome	density	pctwhite	pctblack	HD_present	Lo_present \
0	3.2	1.8	39.6	2.5	0.4	0	0
12	2.0	0.8	1.3	1.6	0.0	0	0
21	2.8	0.1	0.7	1.4	0.1	0	0
20	4.6	1.6	2.4	0.9	0.1	0	0
19	3.9	1.5	1.7	0.6	0.0	0	0
18	4.4	0.3	1.1	1.0	0.0	0	0
17	1.1	0.3	2.9	0.6	0.0	0	0
16	4.4	0.1	2.1	2.1	0.0	0	0
15	4.3	0.3	4.0	0.5	0.1	0	0
14	5.3	0.7	0.2	0.7	0.0	0	0
13	5.4	0.7	2.5	1.0	0.1	0	0
11	6.2	0.7	38.1	1.1	0.0	0	0
1	3.9	0.6	0.7	0.7	0.0	0	0
10	4.8	0.1	0.0	1.2	0.0	0	0
9	3.0	4.6	27.9	0.2	0.4	0	0
8	1.7	4.8	16.1	3.1	0.3	0	0
7	3.8	2.3	0.4	1.4	0.0	0	0
6	1.0	1.3	3.4	0.8	0.0	0	0
5	2.8	1.3	0.3	0.9	0.0	0	0
4	0.8	2.1	0.1	0.9	0.0	0	0
3	1.5	0.3	1.5	1.5	0.0	0	0
2	4.1	2.0	1.9	0.9	0.0	0	0
22	3.8	0.8	0.6	1.4	0.0	0	0

	Store_present	highway_count
0	0	1
12	0	1
21	0	2
20	0	3
19	0	2
18	0	4
17	0	0
16	0	1
15	0	2
14	0	1

13	0	4
11	0	0
1	0	0
10	0	2
9	0	0
8	0	0
7	0	1
6	0	2
5	0	1
4	0	0
3	0	2
2	0	2
22	0	3

[ ]:

```

```