Assignment 4

April 25, 2021

---

*You are currently looking at **version 1.1** of this notebook. To download notebooks and datafiles, as well as get help on Jupyter notebooks in the Coursera platform, visit the Jupyter Notebook FAQ course resource.*

0.1 Assignment 4 - Understanding and Predicting Property Maintenance Fines

This assignment is based on a data challenge from the Michigan Data Science Team (MDST).

The Michigan Data Science Team (MDST) and the Michigan Student Symposium for Interdisciplinary Statistical Sciences (MSSISS) have partnered with the City of Detroit to help solve one of the most pressing problems facing Detroit - blight. Blight violations are issued by the city to individuals who allow their properties to remain in a deteriorated condition. Every year, the city of Detroit issues millions of dollars in fines to residents and every year, many of these fines remain unpaid. Enforcing unpaid blight fines is a costly and tedious process, so the city wants to know: how can we increase blight ticket compliance?

The first step in answering this question is understanding when and why a resident might fail to comply with a blight ticket. This is where predictive modeling comes in. For this assignment, your task is to predict whether a given blight ticket will be paid on time.

All data for this assignment has been provided to us through the Detroit Open Data Portal. **Only the data already included in your Coursera directory can be used for training the model for this assignment.** Nonetheless, we encourage you to look into data from other Detroit datasets to help inform feature creation and model selection. We recommend taking a look at the following related datasets:

- Building Permits
- Trades Permits
- Improve Detroit: Submitted Issues
- DPD: Citizen Complaints
- Parcel Map

We provide you with two data files for use in training and validating your models: train.csv and test.csv. Each row in these two files corresponds to a single blight ticket, and includes information about when, why, and to whom each ticket was issued. The target variable is compliance, which is True if the ticket was paid early, on time, or within one month of the hearing data, False if the ticket was paid after the hearing date or not at all, and Null if the violator was found not responsible. Compliance, as well as a handful of other variables that will not be available at test-time, are only included in train.csv.

Note: All tickets where the violators were found not responsible are not considered during evaluation. They are included in the training set as an additional source of data for visualization, and to enable unsupervised and semi-supervised approaches. However, they are not included in the test set.

File descriptions (Use only this data for training your model!)

```
readonly/train.csv - the training set (all tickets issued 2004-2011)
readonly/test.csv - the test set (all tickets issued 2012-2016)
readonly/addresses.csv & readonly/latlons.csv - mapping from ticket id to addresses
 Note: misspelled addresses may be incorrectly geolocated.
```

Data fields

train.csv & test.csv

```
ticket_id - unique identifier for tickets
agency_name - Agency that issued the ticket
inspector_name - Name of inspector that issued the ticket
violator_name - Name of the person/organization that the ticket was issued to
violation_street_number, violation_street_name, violation_zip_code - Address where
mailing_address_str_number, mailing_address_str_name, city, state, zip_code, non_us
ticket_issued_date - Date and time the ticket was issued
hearing_date - Date and time the violator's hearing was scheduled
violation_code, violation_description - Type of violation
disposition - Judgment and judgement type
fine_amount - Violation fine amount, excluding fees
admin_fee - $20 fee assigned to responsible judgments
```

state_fee - $10 fee assigned to responsible judgments late_fee - 10% fee assigned to responsible judgments discount_amount - discount applied, if any clean_up_cost - DPW clean-up or graffiti removal cost judgment_amount - Sum of all fines and fees grafitti_status - Flag for graffiti violations

train.csv only

```
payment_amount - Amount paid, if any
payment_date - Date payment was made, if it was received
payment_status - Current payment status as of Feb 1 2017
balance_due - Fines and fees still owed
collection_status - Flag for payments in collections
compliance [target variable for prediction]
 Null = Not responsible
 0 = Responsible, non-compliant
 1 = Responsible, compliant
compliance_detail - More information on why each ticket was marked compliant or non
```

0.2 Evaluation

Your predictions will be given as the probability that the corresponding blight ticket will be paid on time.

The evaluation metric for this assignment is the Area Under the ROC Curve (AUC).

Your grade will be based on the AUC score computed for your classifier. A model which with an AUROC of 0.7 passes this assignment, over 0.75 will recieve full points. ___

For this assignment, create a function that trains a model to predict blight ticket compliance in Detroit using `readonly/train.csv`. Using this model, return a series of length 61001 with the data being the probability that each corresponding ticket from `readonly/test.csv` will be paid, and the index being the ticket_id.

Example:

```
ticket_id
   284932    0.531842
   285362    0.401958
   285361    0.105928
   285338    0.018572
             ...
   376499    0.208567
   376500    0.818759
   369851    0.018528
   Name: compliance, dtype: float32
```

0.2.1 Hints

- Make sure your code is working before submitting it to the autograder.
- Print out your result to see whether there is anything weird (e.g., all probabilities are the same).
- Generally the total runtime should be less than 10 mins. You should NOT use Neural Network related classifiers (e.g., MLPClassifier) in this question.
- Try to avoid global variables. If you have other functions besides blight_model, you should move those functions inside the scope of blight_model.
- Refer to the pinned threads in Week 4's discussion forum when there is something you could not figure it out.

In [38]:
```
import pandas as pd
import numpy as np



df_train = pd.read_csv('readonly/train.csv', encoding = "ISO-8859-1")
df_test = pd.read_csv('readonly/test.csv', encoding = "ISO-8859-1")
```

```
columns_to_remove_train = ['balance_due',
 'collection_status',
 'compliance_detail',
 'payment_amount',
 'payment_date',
 'payment_status']

columns_to_remove_all = ['violator_name', 'zip_code', 'country', 'city',
                         'inspector_name', 'violation_street_number', 'violat
                         'violation_zip_code', 'violation_description',
                         'mailing_address_str_number', 'mailing_address_str_n
                         'non_us_str_code',
                         'ticket_issued_date', 'hearing_date','grafitti_statu


df_train.drop(columns_to_remove_train, axis=1, inplace=True)
df_train.drop(columns_to_remove_all, axis=1, inplace=True)
df_test.drop(columns_to_remove_all, axis=1, inplace=True)


df_latlons = pd.read_csv('readonly/latlons.csv')

df_address =  pd.read_csv('readonly/addresses.csv')

#df_id_latlons = df_address.set_index('address').join(df_latlons.set_index
df_id_latlons = pd.merge(df_latlons,df_address,on='address')


#df_train = df_train.set_index('ticket_id').join(df_id_latlons.set_index(
df_train = pd.merge(df_train,df_id_latlons,on='ticket_id')
df_test = pd.merge(df_test,df_id_latlons,on='ticket_id')
#df_test = df_test.set_index('ticket_id').join(df_id_latlons.set_index('ti
vio_code_freq_top15 = df_train['violation_code'].value_counts().index[0:15
#print(vio_code_freq_top15)

df_train['violation_code_freq_top15'] = [list(vio_code_freq_top15).index(c
print(df_train)
# drop violation code

df_train.drop('violation_code', axis=1, inplace=True)

df_test['violation_code_freq_top15'] = [list(vio_code_freq_top15).index(c)
df_test.drop('violation_code', axis=1, inplace=True)



#df_train.grafitti_status.fillna('None', inplace=True)
#df_test.grafitti_status.fillna('None', inplace=True)
```

```
df_train = df_train[df_train.compliance.isnull() == False]

df_train.lat.fillna(method='pad', inplace=True)
df_train.lon.fillna(method='pad', inplace=True)
df_train.state.fillna(method='pad', inplace=True)
#print(df_train)

df_test.lat.fillna(method='pad', inplace=True)
df_test.lon.fillna(method='pad', inplace=True)
df_test.state.fillna(method='pad', inplace=True)

df_train.drop('address', axis=1, inplace=True)
df_test.drop('address', axis=1, inplace=True)




# So remove city and states...

one_hot_encode_columns = ['agency_name', 'state', 'disposition']

df_train = pd.get_dummies(df_train, columns=one_hot_encode_columns)
df_test = pd.get_dummies(df_test, columns=one_hot_encode_columns)
#print(df_test)



from sklearn.model_selection import train_test_split
#train_features = df_train.columns.drop('compliance')

#X_data, X_keep, y_data, y_keep = train_test_split(df_train[train_features
                                                    #df_train.compliance,
                                                    #random_state=0,
                                                    #test_size=0.05)

#print(X_data.shape, X_keep.shape)


#X_train, X_test, y_train, y_test = train_test_split(X_data[train_features
                                                    #y_data,
                                                    #random_state=0,
                                                    #test_size=0.2)


X_train, X_test_2, y_train, y_test_2 = train_test_split(df_train.drop('com
                                                    df_train['compliance']
                                                    random_state=0,
```

```
                                                              test_size=0.8)
```

```python
#print(X_train.shape, X_test.shape)

def blight_model():

    # Your code here
    from sklearn.ensemble import RandomForestClassifier
    from sklearn.ensemble import GradientBoostingClassifier
    from sklearn.metrics import roc_auc_score
    clf_RF = RandomForestClassifier(max_features = 10, random_state = 0)
    clf_RF.fit(X_train, y_train)
    clf_GDBT= GradientBoostingClassifier(learning_rate = 0.01, max_depth =
    clf_GDBT.fit(X_train, y_train)
    #print('Here training score RF: {}'.format(clf_RF.score(X_train, y_tra
    #print('Here test score RF: {}'.format(clf_RF.score(X_test_2, y_test_2
    #print('Here training score GDBT: {}'.format(clf_GDBT.score(X_train, y
    #print('Here test score GDBT: {}'.format(clf_GDBT.score(X_test_2, y_te
    roc_score = roc_auc_score(y_test_2, clf_GDBT.predict_proba(X_test_2)[:
    preds = pd.Series(data=clf_GDBT.predict_proba(X_test)[:,1], index=X_te
    #print(len(preds))


    return preds#preds# Your answer here
#print(blight_model())
```

```
/opt/conda/lib/python3.6/site-packages/IPython/core/interactiveshell.py:2717: Dtype
  interactivity=interactivity, compiler=compiler, result=result)
```

```
    ticket_id                                     agency_name state  \
0       22056  Buildings, Safety Engineering & Env Department    IL
1       27586  Buildings, Safety Engineering & Env Department    MI
2       22062  Buildings, Safety Engineering & Env Department    MI
3       22084  Buildings, Safety Engineering & Env Department    MI
4       22093  Buildings, Safety Engineering & Env Department    MI
5       22046  Buildings, Safety Engineering & Env Department    CA
6       18738  Buildings, Safety Engineering & Env Department    MI
7       18735  Buildings, Safety Engineering & Env Department    MI
8       18733  Buildings, Safety Engineering & Env Department    MI
9       28204  Buildings, Safety Engineering & Env Department    MI
10      18739  Buildings, Safety Engineering & Env Department    MI
11      18740  Buildings, Safety Engineering & Env Department    MI
12      18743  Buildings, Safety Engineering & Env Department    MI
13      18741  Buildings, Safety Engineering & Env Department    MI
```

14	18978	Buildings, Safety Engineering & Env Department	MI
15	18974	Buildings, Safety Engineering & Env Department	MI
16	18742	Buildings, Safety Engineering & Env Department	MI
17	28211	Buildings, Safety Engineering & Env Department	MI
18	18746	Buildings, Safety Engineering & Env Department	MI
19	18744	Buildings, Safety Engineering & Env Department	MI
20	18747	Buildings, Safety Engineering & Env Department	MI
21	26846	Buildings, Safety Engineering & Env Department	MI
22	26848	Buildings, Safety Engineering & Env Department	MI
23	28209	Buildings, Safety Engineering & Env Department	MI
24	18736	Buildings, Safety Engineering & Env Department	MI
25	18737	Buildings, Safety Engineering & Env Department	MI
26	19950	Health Department	MI
27	19311	Department of Public Works	MI
28	18645	Buildings, Safety Engineering & Env Department	MI
29	19310	Department of Public Works	MI
...	...	...	...
250276	285037	Department of Public Works	MI
250277	285034	Department of Public Works	MI
250278	285106	Department of Public Works	MI
250279	285040	Department of Public Works	MI
250280	285042	Department of Public Works	MI
250281	285097	Department of Public Works	MI
250282	285098	Department of Public Works	MI
250283	284740	Buildings, Safety Engineering & Env Department	MI
250284	284877	Detroit Police Department	MI
250285	284879	Department of Public Works	MI
250286	284880	Department of Public Works	MI
250287	284650	Department of Public Works	MI
250288	285125	Department of Public Works	MI
250289	284647	Department of Public Works	MI
250290	284881	Department of Public Works	MI
250291	284878	Department of Public Works	MI
250292	284338	Department of Public Works	OH
250293	284333	Department of Public Works	MI
250294	284882	Detroit Police Department	MI
250295	284883	Detroit Police Department	MI
250296	366178	Buildings, Safety Engineering & Env Department	MI
250297	366176	Buildings, Safety Engineering & Env Department	MI
250298	325560	Buildings, Safety Engineering & Env Department	MI
250299	325556	Buildings, Safety Engineering & Env Department	MI
250300	325558	Buildings, Safety Engineering & Env Department	MI
250301	325555	Buildings, Safety Engineering & Env Department	MI
250302	325557	Buildings, Safety Engineering & Env Department	MI
250303	325562	Buildings, Safety Engineering & Env Department	MI
250304	325559	Buildings, Safety Engineering & Env Department	MI
250305	325561	Buildings, Safety Engineering & Env Department	MI

	violation_code	disposition	fine_amount	\
0	9-1-36(a)	Responsible by Default	250.0	
1	61-63.0600	Responsible by Determination	750.0	
2	9-1-36(a)	Not responsible by Dismissal	250.0	
3	9-1-36(a)	Not responsible by City Dismissal	250.0	
4	9-1-36(a)	Not responsible by Dismissal	250.0	
5	9-1-36(a)	Responsible by Default	250.0	
6	61-63.0500	Responsible by Default	750.0	
7	61-63.0100	Responsible by Default	100.0	
8	61-63.0100	Responsible by Default	100.0	
9	61-63.0600	Responsible by Default	750.0	
10	61-63.0600	Not responsible by Dismissal	750.0	
11	61-63.0100	Not responsible by Dismissal	100.0	
12	61-63.0600	Responsible by Default	750.0	
13	61-63.0600	Responsible by Default	750.0	
14	61-63.0600	Responsible by Default	750.0	
15	61-63.0600	PENDING JUDGMENT	750.0	
16	61-63.0600	Not responsible by Dismissal	750.0	
17	61-63.0600	Not responsible by Dismissal	750.0	
18	61-63.0100	Responsible by Determination	100.0	
19	61-63.0100	Responsible by Determination	100.0	
20	61-63.0100	Not responsible by Dismissal	100.0	
21	61-63.0600	Responsible by Default	750.0	
22	61-63.0600	Responsible by Default	750.0	
23	61-63.0600	Responsible by Default	750.0	
24	61-63.0100	Not responsible by Dismissal	100.0	
25	61-63.0100	PENDING JUDGMENT	100.0	
26	9-1-103(C)	Responsible by Admission	100.0	
27	22-2-83(a)(b)(c)	Not responsible by Dismissal	3500.0	
28	9-1-36(a)	Responsible by Default	250.0	
29	22-2-83(a)(b)(c)	Not responsible by Dismissal	3500.0	
...	...	...	...	
250276	9-1-105	Responsible by Default	100.0	
250277	22-2-88(b)	Responsible by Default	500.0	
250278	22-2-88(b)	Responsible by Default	200.0	
250279	22-2-88(b)	Not responsible by City Dismissal	1000.0	
250280	9-1-110(a)	Not responsible by Dismissal	50.0	
250281	22-2-88(b)	Not responsible by Determination	500.0	
250282	9-1-110(a)	Not responsible by Determination	50.0	
250283	9-1-36(a)	Not responsible by Dismissal	250.0	
250284	9-1-36(a)	Not responsible by Dismissal	250.0	
250285	22-2-61	Not responsible by Dismissal	200.0	
250286	22-2-61	Not responsible by Dismissal	200.0	
250287	22-2-88(b)	Responsible by Default	1000.0	
250288	22-2-88(b)	Responsible by Default	500.0	
250289	22-2-88(b)	Not responsible by Dismissal	1000.0	
250290	22-2-61	Responsible by Determination	200.0	
250291	22-2-61	Not responsible by Dismissal	200.0	

250292	22-2-88(b)	Not responsible by Dismissal	200.0
250293	22-2-88(b)	Responsible by Default	200.0
250294	22-2-88(a)	Not responsible by Dismissal	100.0
250295	22-2-88(a)	Not responsible by Dismissal	100.0
250296	9-1-111	Not responsible by City Dismissal	100.0
250297	9-1-36(a)	Not responsible by City Dismissal	250.0
250298	9-1-43(a) - (Structu	Not responsible by City Dismissal	1000.0
250299	9-1-43(a) - (Structu	Not responsible by City Dismissal	1000.0
250300	9-1-43(a) - (Structu	Not responsible by City Dismissal	1000.0
250301	9-1-43(a) - (Structu	Not responsible by City Dismissal	1000.0
250302	9-1-43(a) - (Structu	Not responsible by City Dismissal	1000.0
250303	9-1-43(a) - (Structu	Not responsible by City Dismissal	1000.0
250304	9-1-43(a) - (Structu	Not responsible by City Dismissal	1000.0
250305	9-1-43(a) - (Structu	Not responsible by City Dismissal	1000.0

	admin_fee	state_fee	late_fee	discount_amount	clean_up_cost	\
0	20.0	10.0	25.0	0.0	0.0	
1	20.0	10.0	75.0	0.0	0.0	
2	0.0	0.0	0.0	0.0	0.0	
3	0.0	0.0	0.0	0.0	0.0	
4	0.0	0.0	0.0	0.0	0.0	
5	20.0	10.0	25.0	0.0	0.0	
6	20.0	10.0	75.0	0.0	0.0	
7	20.0	10.0	10.0	0.0	0.0	
8	20.0	10.0	10.0	0.0	0.0	
9	20.0	10.0	75.0	0.0	0.0	
10	0.0	0.0	0.0	0.0	0.0	
11	0.0	0.0	0.0	0.0	0.0	
12	20.0	10.0	75.0	0.0	0.0	
13	20.0	10.0	75.0	0.0	0.0	
14	20.0	10.0	75.0	0.0	0.0	
15	0.0	0.0	0.0	0.0	0.0	
16	0.0	0.0	0.0	0.0	0.0	
17	0.0	0.0	0.0	0.0	0.0	
18	20.0	10.0	10.0	0.0	0.0	
19	20.0	10.0	10.0	0.0	0.0	
20	0.0	0.0	0.0	0.0	0.0	
21	20.0	10.0	75.0	0.0	0.0	
22	20.0	10.0	75.0	0.0	0.0	
23	20.0	10.0	75.0	0.0	0.0	
24	0.0	0.0	0.0	0.0	0.0	
25	0.0	0.0	0.0	0.0	0.0	
26	20.0	10.0	0.0	0.0	0.0	
27	0.0	0.0	0.0	0.0	0.0	
28	20.0	10.0	25.0	0.0	0.0	
29	0.0	0.0	0.0	0.0	0.0	
...	...	...	...	...	...	
250276	20.0	10.0	10.0	0.0	0.0	

250277	20.0	10.0	50.0	0.0	0.0
250278	20.0	10.0	20.0	0.0	0.0
250279	0.0	0.0	0.0	0.0	0.0
250280	0.0	0.0	0.0	0.0	0.0
250281	0.0	0.0	0.0	0.0	0.0
250282	0.0	0.0	0.0	0.0	0.0
250283	0.0	0.0	0.0	0.0	0.0
250284	0.0	0.0	0.0	0.0	0.0
250285	0.0	0.0	0.0	0.0	0.0
250286	0.0	0.0	0.0	0.0	0.0
250287	20.0	10.0	100.0	0.0	0.0
250288	20.0	10.0	50.0	0.0	0.0
250289	0.0	0.0	0.0	0.0	0.0
250290	20.0	10.0	0.0	0.0	0.0
250291	0.0	0.0	0.0	0.0	0.0
250292	0.0	0.0	0.0	0.0	0.0
250293	20.0	10.0	20.0	0.0	0.0
250294	0.0	0.0	0.0	0.0	0.0
250295	0.0	0.0	0.0	0.0	0.0
250296	0.0	0.0	0.0	0.0	0.0
250297	0.0	0.0	0.0	0.0	0.0
250298	0.0	0.0	0.0	0.0	0.0
250299	0.0	0.0	0.0	0.0	0.0
250300	0.0	0.0	0.0	0.0	0.0
250301	0.0	0.0	0.0	0.0	0.0
250302	0.0	0.0	0.0	0.0	0.0
250303	0.0	0.0	0.0	0.0	0.0
250304	0.0	0.0	0.0	0.0	0.0
250305	0.0	0.0	0.0	0.0	0.0

	judgment_amount	compliance	address	lat \
0	305.0	0.0	2900 tyler, Detroit MI	42.390729
1	855.0	1.0	4311 central, Detroit MI	42.326937
2	0.0	NaN	1449 longfellow, Detroit MI	42.380516
3	0.0	NaN	1441 longfellow, Detroit MI	42.380570
4	0.0	NaN	2449 churchill, Detroit MI	42.145257
5	305.0	0.0	6478 northfield, Detroit MI	42.145257
6	855.0	0.0	8027 brentwood, Detroit MI	42.433466
7	140.0	0.0	8228 mt elliott, Detroit MI	42.388641
8	140.0	0.0	8228 mt elliott, Detroit MI	42.388641
9	855.0	0.0	15307 seven mile, Detroit MI	42.435773
10	0.0	NaN	18353 van dyke, Detroit MI	42.428590
11	0.0	NaN	9500 van dyke, Detroit MI	42.428590
12	855.0	0.0	9100 van dyke, Detroit MI	42.395765
13	855.0	0.0	20024 schaefer, Detroit MI	42.440190
14	855.0	0.0	9425 van dyke, Detroit MI	42.399222
15	0.0	NaN	9313 van dyke, Detroit MI	42.397677
16	0.0	NaN	7301 harper, Detroit MI	42.386298

```
17                   0.0          NaN       14935 warren, Detroit MI  42.386298
18                 140.0          1.0        14715 klenk, Detroit MI  42.360836
19                 140.0          1.0        14715 klenk, Detroit MI  42.360836
20                   0.0          NaN        14715 klenk, Detroit MI  42.360836
21                 855.0          0.0       16331 warren, Detroit MI  42.341729
22                 855.0          0.0     5650 livernois, Detroit MI  42.341620
23                 855.0          0.0   15733 seven mile, Detroit MI  42.435592
24                   0.0          NaN         18701 mack, Detroit MI  42.414639
25                   0.0          NaN         18701 mack, Detroit MI  42.414639
26                 130.0          0.0     13400 schaefer, Detroit MI  42.385741
27                   0.0          NaN    20525 pinehurst, Detroit MI  42.444927
28                 305.0          0.0            601 king, Detroit MI  42.383385
29                   0.0          NaN    20525 pinehurst, Detroit MI  42.444927
...                  ...          ...                            ...        ...
250276             140.0          0.0     13924 woodmont, Detroit MI  42.387270
250277             580.0          0.0     13924 woodmont, Detroit MI  42.387270
250278             250.0          0.0    20009 northlawn, Detroit MI  42.440228
250279               0.0          NaN    14413 grandmont, Detroit MI  42.393239
250280               0.0          NaN    14413 grandmont, Detroit MI  42.393239
250281               0.0          NaN    14413 grandmont, Detroit MI  42.393239
250282               0.0          NaN    14413 grandmont, Detroit MI  42.393239
250283               0.0          NaN  17191 edinborough, Detroit MI  42.416602
250284               0.0          NaN    1701 grand blvd, Detroit MI  42.376407
250285               0.0          NaN         19750 marx, Detroit MI  42.439841
250286               0.0          NaN         19751 marx, Detroit MI  42.439819
250287            1130.0          0.0        15725 steel, Detroit MI  42.406293
250288             580.0          0.0       7152 chicago, Detroit MI  42.366529
250289               0.0          NaN   19400 greenfield, Detroit MI  42.434137
250290             230.0          1.0   17403 mt elliott, Detroit MI  42.422081
250291               0.0          NaN    1610 state fair, Detroit MI  42.439749
250292               0.0          NaN        16211 coram, Detroit MI  42.438587
250293             250.0          0.0       15634 novara, Detroit MI  42.438867
250294               0.0          NaN     12210 st marys, Detroit MI  42.377159
250295               0.0          NaN      19044 chicago, Detroit MI  42.365113
250296               0.0          NaN       8020 puritan, Detroit MI  42.410031
250297               0.0          NaN       8020 puritan, Detroit MI  42.410031
250298               0.0          NaN    10701 meyers rd, Detroit MI  42.369982
250299               0.0          NaN    10701 meyers rd, Detroit MI  42.369982
250300               0.0          NaN    10701 meyers rd, Detroit MI  42.369982
250301               0.0          NaN  10701 santa maria, Detroit MI  42.418549
250302               0.0          NaN    10701 meyers rd, Detroit MI  42.369982
250303               0.0          NaN    10701 meyers rd, Detroit MI  42.369982
250304               0.0          NaN    10701 meyers rd, Detroit MI  42.369982
250305               0.0          NaN    10701 meyers rd, Detroit MI  42.369982

              lon  violation_code_freq_top15
0      -83.124268                          0
1      -83.135118                         -1
```

2	-83.096069	0
3	-83.095919	0
4	-83.208233	0
5	-83.208233	0
6	-83.023493	-1
7	-83.037858	-1
8	-83.037858	-1
9	-82.963348	-1
10	-83.024307	-1
11	-83.024307	-1
12	-83.022333	-1
13	-83.180488	-1
14	-83.023161	-1
15	-83.023098	-1
16	-83.026209	-1
17	-83.026209	-1
18	-82.930958	-1
19	-82.930958	-1
20	-82.930958	-1
21	-83.262245	-1
22	-83.127416	-1
23	-82.958282	-1
24	-82.911820	-1
25	-82.911820	-1
26	-83.178017	10
27	-83.168174	-1
28	-83.072582	0
29	-83.168174	-1
...	...	...
250276	-83.208856	7
250277	-83.208856	4
250278	-83.154829	4
250279	-83.211170	4
250280	-83.211170	8
250281	-83.211170	4
250282	-83.211170	8
250283	-83.234201	0
250284	-83.031859	0
250285	-83.083820	14
250286	-83.084577	14
250287	-83.171850	4
250288	-83.141897	4
250289	-83.199508	4
250290	-83.038656	14
250291	-83.087521	14
250292	-82.952976	4
250293	-82.959888	4
250294	-83.204497	-1

```
250295 -83.227270                                -1
250296 -83.150309                                -1
250297 -83.150309                                 0
250298 -83.168038                                -1
250299 -83.168038                                -1
250300 -83.168038                                -1
250301 -83.168063                                -1
250302 -83.168038                                -1
250303 -83.168038                                -1
250304 -83.168038                                -1
250305 -83.168038                                -1

[250306 rows x 17 columns]
Here training score RF: 0.9893044783587691
Here test score RF: 0.938508568926695
Here training score GDBT: 0.9450838128596447
Here test score GDBT: 0.9434732299224419
7994
0.792729070456
```

```
In [ ]: blight_model()
```

```
In [ ]:
```