Fines_On_Detrot

December 12, 2023

1 Data Science and Machine Learning

1.1 Property fines of maintenance in Detroid

1.2 Object

The object of our work is to implementation machine learning alghoritm to predict whether person is gonna pay the fine or no.

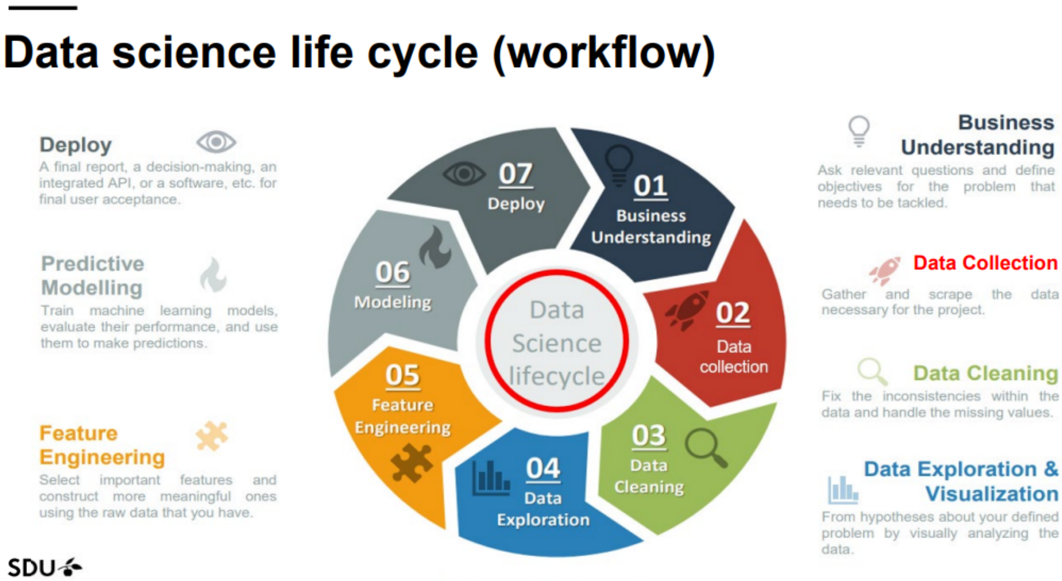
Students (with exam numbers): Iñigo Susila Uranga(4123840), Bohdan Kopčák(4123815), Mateusz Sopyła(4123873) Teachers : Julia Pahl, Jie Cai, Marina Meireles Pereira

About the data: https://midas.umich.edu/wp-content/uploads/sites/3/2017/09/understanding-blight-ticket.pdf

2 Introduction

In the city of Detroit, over 20% of properties are affected by property blight, leading the City to issue tickets to the owners of these parcels as a means of encouraging residents to maintain their properties. Despite this effort, the compliance rate for these blight tickets is alarmingly low. Our approach involves constructing a predictive model to forecast ticket compliance, conducting a thorough analysis of property owners who receive blight tickets, and examining the disparities in compliance among these diverse groups of residents.

2.0.1 The data was analyzed according to the scheme below:



3 Data Presentation

The Datasets about fines in Detroid includes information about blights tickests (when, why and to whom each ticket was issued), adresses issued each tickets and data regarding location of violation. Whole 3 data frames have been used for analysis. The following are presented name of columns for all datasets.

fines.csv: all tickets issued 2004-2011 addresses.csv: datasets mapping from the fine ticket id to addresses latlons.csv: datasets mapping from addresses to latitude/longitude coordinates.

[291]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import matplotlib as mpl
import seaborn as sns
from sklearn.cluster import KMeans
import folium
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from datetime import date
from six import StringIO
from IPython.display import Image
import pydotplus
from termcolor import colored
import random
from sklearn.metrics import ConfusionMatrixDisplay
from sklearn import tree
```

```
from sklearn.tree import export_graphviz
```

[292]:
```
df_fines = pd.read_csv("fines.csv", sep=',',skiprows=0,skipfooter=1,␣
 ↪na_values=None,engine='python')
df_latlons = pd.read_csv("latlons.csv", sep=',', skiprows=0, skipfooter=1,␣
 ↪na_values=None,engine='python')
df_addresses = pd.read_csv("addresses.csv", sep=',', skiprows=0,␣
 ↪na_values=None,engine='python')
```

[293]:
```
print('df_fines')
print(colored(df_fines.columns,'red'))
print("\ndf_latlons")
print(colored(df_latlons.columns,'green'))
print("\ndf_addresses")
print(colored(df_addresses.columns,'blue'))
```

```
df_fines
Index(['ticket_id', 'agency_name', 'inspector_name', 'violator_name',

       'violation_street_number', 'violation_street_name',

       'violation_zip_code', 'mailing_address_str_number',

       'mailing_address_str_name', 'city', 'state', 'zip_code',

       'non_us_str_code', 'country', 'ticket_issued_date', 'hearing_date',

       'violation_code', 'violation_description', 'disposition', 'fine_amount',

       'admin_fee', 'state_fee', 'late_fee', 'discount_amount',

       'clean_up_cost', 'judgment_amount', 'payment_amount', 'balance_due',

       'payment_date', 'payment_status', 'collection_status',

       'grafitti_status', 'compliance_detail', 'compliance'],

      dtype='object')

df_latlons
Index(['address', 'lat', 'lon'], dtype='object')

df_addresses
Index(['ticket_id', 'address'], dtype='object')
```

4 Data Cleaning

Data cleaning is the most crucial part of any project involving a large amount of data. Properly preparing data frames allows for more accurate predictions of the rate under investigation. Cleaning data includes identifying and correcting errors and inconsistencies in the data, as well as formatting the data in a way that would be suitable for the analysis. The following steps were taken to clean:

4.0.1 1. Changed values in column payment status:

We assumed these enumerations:

payment status	value assigned
NO PAYMENT APPLIED	0
PARTIAL PAYMENT APPLIED	1
PAID IN FULL	2

4.0.2 2. Changed values in column disposition as the following

disposition	value assigned
Not responsible by Dismissal	Not responsible
Not responsible by City Dismissal	Not responsible
Not responsible by Determination	Not responsible
Responsible by Default	Responsible
Responsible by Determination	Responsible
Responsible by Admission	Responsible
PENDING JUDGMENT	*omitted*
SET-ASIDE (PENDING JUDGMENT)	*omitted*

[294]:
```
not_responsible_tags = r'Not responsible by Dismissal\
                        |Not responsible by City Dismissal\
                        |Not responsible by Determination'
responsible_tags = r'Responsible by Default\
                    |Responsible by Determination\
                    |Responsible by Admission'
omitted_tags = (r'PENDING JUDGMENT|SET-ASIDE \(PENDING JUDGMENT\)')

# Filter out fines that were not resolved
df_fines = df_fines[df_fines.disposition.str.contains(omitted_tags) == False]

# Change labels to boolean vals
df_fines['disposition'] = df_fines['disposition'].map(
    lambda x:
        "Not responsible" if x in not_responsible_tags
        else "Responsible")
```

4.0.3 3. Removed all fines where the subject corrected the violation before his hearing, eg. when the case was dismissed.

[295]:
```
df_fines = df_fines[df_fines['disposition'] == "Responsible"]
```

4.0.4 4. We have changed compliancy 0/1 values to boolean

[296]:
```
df_fines['compliance'] = df_fines['compliance'].map(
    lambda x:
        False if x == 0.0
        else True)
```

4.0.5 5. We casted data to String and Datatime

Frst of all we are going to change the type of the rows to ones that we can later operate with, changing from object to String, Date, etc…

[297]:
```
df_fines['mailing_address_str_name'] = df_fines['mailing_address_str_name'].
 ↪astype(str)
df_fines['violation_street_name'] = df_fines['violation_street_name'].
 ↪astype(str)
df_fines['violator_name'] = df_fines['violator_name'].astype(str)
df_fines['ticket_issued_date'] = pd.to_datetime(df_fines['ticket_issued_date'])
```

[298]:
```
df_fines['ticket_year'] = df_fines['ticket_issued_date'].dt.year
df_fines['ticket_month'] = df_fines['ticket_issued_date'].dt.month
df_fines['ticket_day'] = df_fines['ticket_issued_date'].dt.day
df_fines['ticket_day_week'] = df_fines['ticket_issued_date'].dt.day_of_week
```

[299]:
```
df_fines = pd.get_dummies(df_fines,columns=['ticket_day_week'])
df_fines = df_fines.rename(columns={'ticket_day_week_0':'isMonday',
                                    'ticket_day_week_1': 'isTuesday',
                                    'ticket_day_week_2':'isWednesday',
                                    'ticket_day_week_3':'isThursday',
                                    'ticket_day_week_4':'isFriday',
                                    'ticket_day_week_5':'isSaturday',
                                    'ticket_day_week_6':'isSunday'})
```

4.0.6 `Violator_name` clean up

Data frame have a lot of misspelling. In this section we corrected misspelling in dataframe.

The data has to be analysed manualy and the discrepancies corrected.

[300]:
```
misspellings = {
    'INVESTMENT, ACORN' : ['INVESTMENT CO., ACORN',
                           'Investment, Acorn',
                           'ACORN INVESTMENT, *',
                           'INVESTMENTS, ACORN',
                           'Acorn Investments, *',
                           'COMPANY, ACORN INVESTMENT',
                           'ACORN INVESTMENTS, *',
                           'CO., ACORN INVESTMENT',
```

```
                             'Investment Co., Acorn',
                             'investment, acorn',
                             'INVESTMENT CO, ACORN',
                             'ACORN INVESTMENT CO., .',
                             'Company, Acorn Investment',
                             'ACORN INVESTMENT COMPANY, .',
                             'Co., Acorn Investment',
                             'INVESTMENT COMP., ACORN',
                             'COMPANY, ACORN  INVESTMENT',
                             'ACORN INVESTMENT CO., *',
                             'Co, Acorn Investment',
                             'INVESTMENT CO. , ACORN',
                             'Investment Company, Acorn',
                             'INVESTMENT COMPANY, ACORN',
                             'COMP., ACORN INVESTMENT',
                             'ACORN INVSTM. CO., .',
                             'CO, ACORN INVESTMENTS',
                             'CO, ACORN INVESTMENT',
                             'INVETMENT CO., ACORN',
                             'CO., ACORN  INVESTMENT',
                             'Co., Acorn Investment-',
                             'COMP, ACORN INVERTMENT',
                             'Investment, Acorns',
                             'Investment Co, Acorn',
                             'acorn investment, *',
                             'ACORN INVESMENTS, *',
                             'INNESTMENT, ACORN',
                             'ACORN INVESTMENT, CO',
                             'IVESTMENT, ACORN',
                             'INVESMENT, ACORN',
                             'Acorn Investment, *',
                             'ACORN  INVESTMENTS  CO., .',
                             'c/o ACORN INV., OAK MANAGEMENT',
                             'Investment, Company Acorn',
                             'COMPANY, ACORN   INVESTMENT',
                             'INVESTMENT CO, A CORN',
                             'INVESTMENT CO., A CORN',
                             'INVESTMENT COMPANY, A CORN',
                             'INVESTMENT, A CORN'],

    'BANK, WELLS FARGO' : ['BANK MINNESOTA, WELLS FARGO',
                             'BANK MINNESOTA, WELLS  FARGO',
                             'BANK MINNSOTA, WELL FARGO',
                             'BANK, W ELLS FARGO',
                             'BANK, WELL FARGO',
                             'BANK, WELLS  FARGO',
                             'BANK, WELLS FARGO',
```

```
                          'FARGO BANK, WELLS',
                          'FARGO HOME MORTGAGE, WELLS',
                          'FARGO, WELLS',
                          'Fargo Bank, Wells',
                          'MINNESOTA, WELL FARGO BANK',
                          'MINNESOTA, WELLS   FARGO BANK',
                          'WELL FARGO BANK MINNESOTA, .',
                          'WELLS  FARGO  BANKS  MINNESOTA, .'
                          'WELLS FARGO BANK, .',
                          'MINNESOTA, WELLS FARGO BANK,',
                          'BANK, N.A., WELLS FARGO',
                          'BANK  MINNESOTA, WELLS FARGO',
                          'MINNESOTA, WELLS FARGO BANK',
                          'Bank, Wells Fargo',
                          'BANK, NA, WELLS FARGO',
                          'C/O OCWEN FEDERAL BANK, FSD., WELLS FARGO BANK-MINN.
↪',
                          'BANK,NA TRUSTEE, WELLS FARGO',
                          'HOME MORTGAGE INC, WELLS FARGO',
                          'BANK N. A., WELLS FARGO',
                          'WELLS FARGO HOME MTG., INC., .',
                          'for option one mtg., wells fargo bank natl.assoc␣
↪trustee',
                          'TRUSTEE, WELLS FARGO BANK',
                          'Bank,NA, Wells Fargo',
                          'for option one mtg. loan, wells fargo bank m natl.
↪assoc. trustt',
                          'HOME MORTGAGE, WELLS FARGO',
                          'BANK  MINNESOTA  NA, WELLS FARGO',
                          'MINNESOTA, N.A., WELLS FARGO BANK',
                          'MINNNESOTA, N.A., WELLS FARGO BANK',
                          'BANK,,N.A., WELLS FARGO',
                          'BANK,N.A., WELLS FARGO',
                          'BANK , WELLS FARGO',
                          'WELLS FARGO HOME MORTGAGE, INC, .',
                          'BANK- MINNESOTA, WELLS FARGO',
                          'MORTGAGE, INC., WELLS FARGO  HOME',
                          'MTG INC, WELLS FARGO  HOME',
                          'MINNESOTA, WELLS FARGO',
                          'MORTGAGE, INC, WELLS FARGO HOMES',
                          'BANK-MINNESOTA, WELLS FARGO',
                          'NATIONAL ASSOCIATION, WELLS FARGO  MINNESOTA',
                          'HOME  MORTGAGA, WELLS FARGO',
                          'WELLS FARGO HOME MTG.. INC, .',
                          'C/O OCWEN FEDERAL BANK FSB, WELLS FARGO BANK MINN.',
                          'NATIONAL  ASSC, WELLS FARGO BANK MINNESOTA',
                          'MORTGAGE, INC, WELLS FARGO HOME',
```

```
                          'AMERICA INC, WELLS FARGO FINANCIAL',
                          'BANK  N.A, WELLS FARGO',
                          'BANK, N.A, WELLS FARGO',
                          'MORTGAGE INC., WELLS FARGO HOME',
                          'BANK MINN., WELLS FARGO',
                          'WELLS FARGO BANK MINNESOTA, .',
                          'MORTGAGE, INC., WELLS FARGO HOME',
                          'BANK MINNESOTA NATIONAL ASSOC., WELLS FARGO',
                          'BANK OF MINESOTA, WELLS FARGO',
                          'BANK N.A., WELLS FARGO',
                          'MORTGAGE, WELLS FARGO HOME',
                          'GRP/AG REAL ESTATE ASSET TRUST, WELLS FARGO␣
↪MINNESOTA NATL ASSOC. (TRUSTEE)',
                          'GRP/AG REAL ESTATE ASSET TRUST, WELLS FARGO␣
↪MINNESOTA NATL.  ASSOC (TRUSTEE)',
                          'MORTGAGE INC, WELLS FARGO',
                          'BANK MINNISOTA, WELLS FARGO',
                          'MINNESOTA, WELLS FARGO  BANK,',
                          'WELLS  FARGO  BANKS  MINNESOTA, .',
                          'WELLS FATGO BANK MINN., .',
                          'option one mtg.loan, wellsfarbo bank , Minnesota␣
↪natl.assoc.,trustee for'],
  'MANAGEMENT CO., WINGATE' : ['MANAGEMENT CO., WINGATE',
                                'MANAGEMENT, WINGATE',
                                'APTS/WINGATE, INDUSTRIAL BUILDING',
                                'WINGATE, EDDIE & JAMES',
                                'MANAGEMENT CO,, WINGATE',
                                'MANAGEMENT CO,., WINGATE',
                                'WINGATE MGT CO. INC, *',
                                'WINGATE, WILLIAM',
                                'Wingate Mgt. Co. Inc., *',
                                'WINGATE, NEIL',
                                'MGT. CO, WINGATE',
                                'Industrial Bldg. Apts/ Wingate, *'],
  'NEW YORK, BANK OF' : ['NEW YORK, BANK OF',
                          'NEW YORK, BANK  OF',
                          'NEW YORK, THE BANK OF',
                          'OF NEW YORK, THE BANK',
                          'BANK OF NEW YORK, .',
                          'NEW YORK , BANK OF',
                          'TRU, BANK OF NEW YORK',
                          'NEW YORK, BNK  OF',
                          'CERTFICATE HOLDERS CWABS2004-1, BANK OF NEW YORK␣
↪TRUSTEE',
                          'OF NEW YORK, BANK',
                          'NEW YORK, THE BANK  OF',
                          'NEW YORK-TRUSTEE , BANK OF',
```

```
                              'NEW YORK TRUST, BANK OF',
                              'New York, The Bank of',
                              'THE BANK OF NEW YORK, *',
                              'NEW YORK, Bank of'],
    'BANK, DEUTSCHE' : ['TRUST CO, DEUTCHE  BANK NATIONAL',
                        'NATIONAL TRUST CO., DEUTSCHE BANK',
                        'NATIONAL TRUST, DEUTSCHE BANK',
                        'NATIONAL TRUST CO, DEUTSCHE BANK',
                        'TRUST CO., DEUTSCHE BANK NATIONAL',
                        'TRUST COMPANY, DEUTSCHE BANK NATIONAL',
                        'COMPANY AMERICAS, DEUTSCHE BANK TRUST',
                        'COMPANY, DEUTSCHE BANK NATIONAL TRUST',
                        "CO, DEUTSCHE BANK NAT'L TRUST",
                        'NATIONAL TRUST COMPANY, DEUTSCHE BANK',
                        'TRUST CO, DEUTSCHE BANK NATIONAL',
                        'NATIONAL  TRUST  CO., DEUTSCHE BANK',
                        'TRUST CO., DEUTSCHE BANK',
                        '& TRUST CO., DEUTSCHE BANK',
                        "TRUST CO, DEUTSCHE BANK NAT'L",
                        'TRUST CO., DEUTSCHE NATIONAL',
                        'DEUTSCHE BANK NATIONAL TRUST C, .',
                        "CO., DEUTSHE BANK NAT'L TRUST",
                        'BANK TRUST CO., DEUTSCHE NATIONAL',
                        'NAT. TRUST, DEUTSCHE BANK',
                        'National Trust Co., Deutsche Bank',
                        'BANK NATIONAL TRUST CO., DEUTSCHE',
                        'TRUST CO, DEUTSHE BANK NATION AL',
                        'TRUST CO AMERICAS, DEUTSCHE BANK',
                        'DEUTSCHE BANK NAT. TRUST CO., .',
                        'NATIONAL TRUST  CO., DEUTSCHE BANK',
                        'F/K/A BANKERS TRUST CO., DEUTSCHE BANK TRUST CO. ␣
↪AMERICAS,',
                        'NATIONAL  TRUST, DEUTSCHE BANK',
                        'COMPANY, DEUTSCHE BANK TRUST',
                        'TRUST, DEUTSCHE  BANK NATIONAL',
                        'ASS, DEUTSCHE  BANK NATIONAL',
                        'BANK, DEUTSCHE NATIONAL',
                        'Company of the Americas, Deutsche Bank Trust',
                        'TRUST CO, DEUTSCHE  BANK NATIONAL',
                        "NAT'L TRUST CO., DEUTSCHE BANK",
                        'BANK NATIONAL TRUST CO, DEUTSCHE',
                        'BANK  NATIONAL TRUST CO, DEUTSCHE',
                        'C/O AMERE QUEST, DEUTSCHE BANK NATL  TRUST CO',
                        'NATIONAL  TRUST  CO., DEUTSHE BANK',
                        'NAT. TRUST CO., DEUTSCHE BANK',
                        'COMPANY, DEUTSCHE BANK  NATIONAL TRUST',
                        'NATIONAL. CO     , DEUTSCHE BANK',
```

```
                    'AMERICAS, DEUTSCHE BANK TRUST COMP.',
                    'DEUTSCHE BANK NATIONAL TRUST , 000000',
                    'DEUTSCHE BANK CO. OF CA, .',
                    'TRUST COMPANY, DEUTSCHE  BANK  NATIONAL',
                    'CO., DEUTSCHE BANK NATIONAL TRUST',
                    'CO, DEUTSCHE BANK NATIONAL TRUST',
                    'bank, deutsche',
                    'NATIONAL  TURST  CO., DEUTSCHE BANK',
                    'TRUST CO., DEUTSHE BANK NATIONAL',
                    'TRUST COMPANY AMERICAS, DEUTSCHE BANK',
                    'TRUST CO, DEUTSCHE BANK NAT'],
    'US Bank National Association' : ['ASSOCIATION, U.S. BANK NATIONAL',
                    'ASSOC, US BANK NATIONAL',
                    'ASSOC., US BANK NATIONAL',
                    'ASSOCIATION, US BANK NATIONAL',
                    'ASSOCIATION, U. S.BANK NATIONAL',
                    'ASSOC., U.S. BANK NATIONAL',
                    'C/O HOMEQ SERVICING CORP., US BANK NATIONAL ASSOC',
                    'C/O HOMEQ SERVICING CORP, US BANK NATIONAL ASSOC',
                    'ASS., US BANK NATIONAL',
                    'ASSOCATION, U.S. BANK NATIONAL',
                    'US BANK NATIONAL ASSOC, *',
                    'ASSOCIATION, US BANK NATIONSAL',
                    'N. A., US BANK',
                    'NA, US BANK',
                    'NATIONAL ASSOC., US BANK',
                    'NATIONAL ASSOCIATION, US BANK',
                    'TRUSTEE, US BANK NATIONAL ASSOCIATION',
                    'US BANH NTNL. ASSN., .',
                    'US BANK NTNL. ASSN., .',
                    'US Bank National Association'],
    'Lasalle Bank National Association' : [
                    'NATIONAL ASSOCIATION, LASALLE BANK',
                    'ASSOCIATION, LASALLE BANK NATIONAL',
                    'LASALLE BANK, .',
                    'Bank, LaSalle',
                    'Associates, LaSalle Bank National',
                    'NATIONAL ASSOC., LASALLE BANK',
                    'LASALLE BANK MIDWEST, .',
                    'NATIONAL TRUST ASSOCIATION, LASALLE BANK',
                    'National Association, LaSalle Bank',
                    "TRUSTEE, LASALLE BANK NAT'L ASSOC.",
                    'NATIONAL, LASALLE BANK',
                    'BANK NATIONAL ASSOCIATION, LASALLE',
                    'ASSOC., LASALLE BANK NATIONAL',
                    'MIDWEST, LASALLE BANK',
                    'LASALLE, GLENN',
```

```
                        'LASALLE BANK NATL. ASSN., .',
                        'BANK, LASALLE',
                        'ASS., LASALLE BANK NATIONAL',
                        'LASALLE BANK NATL. ASSC., .',
                        'LASALLE BANK NATL. ASSCN., .'],
    'BANK, WASHINGTON MUTUAL': [
                        'BANK, WASHINGTON MUTUAL',
                        'MUTUAL BANK, WASHINGTON',
                        'WASHINGTON MUTUAL BANK, .',
                        'TROTT&TROTT,PC, WASHINGTON  MUTUAL BANK',
                        'BANK F.A., WASHINGTON  MUTUAL',
                        'MUTUAL BANK F.A., WASHINGTON',

                        'BANK, WASHINGTON  MUTUAL',
                        'F A, WASHINGTON  MUTUAL BANK',
                        'BANK, F.A., WASHINGTON MUTUAL',
                        'BANK, FA, WASHINGTON  MUTUAL',],
    'FELLOWSHIP ESTATES LLC, .' : ['FELLOWSHIP ESTATES, L.L.C. , .',
                                   'FELLOWSHIP ESTATES, L.L.C., .',
                                   'FELLOWSHIP ESTATES, LLC, .'],
    'BANK, WACHOVIA' : ['BANK, WACHOVIA',
                        'Bank NA, Wachovia',
                        'OF DELAWARE, WACHOVIA BANK',
                        'WACHOVIA BANK, .'],
    'BANK, JP MORGAN CHASE' : ['BANK, J  P  MORGAN CHASE',
                               'BANK, J P MORGAN CHASE',
                               'BANK, J.P. MORGAN CHASE',
                               'BANK, JP MORGAN  CHASE',
                               'BANK, JP MORGAN CASE',
                               'BANK, JPMORGAN  CHASE',
                               'BANK, JPMORGAN CHASE',
                               'C-BASS MGT., J. P. MORGAN CHASE BANK',
                               'C-BASS MGT., J.P. MORGAN CHASE BANK',
                               'C/O OCWEN LOAN SERVICE LLC, J.P. MORGAN CHASE␣
↪BANK',
                               'C/OHOMECOMINGS FINANCIAL NETWO, JP MORGAN CHASE␣
↪BANK',
                               'C/Ohomecoming financial networ, JP.MORGAN CHASE␣
↪BANK',
                               'CAPITAL  INC., MORGAN STANLEY ABS',
                               'CAPITAL, MORGAN STANLEY DEAN WITTER',
                               'CHASE  BANK, J P MORGAN',
                               'CHASE BANK TRUSTEE, JP MORGAN',
                               'CHASE BANK, J P  MORGAN',
                               'CHASE BANK, J P MORGAN',
                               'CHASE BANK, J. P. MORGAN',
                               'CHASE BANK, J.P. MORGAN',
```

```
                                'CHASE BANK, JP MORGAN',
                                'CHASE BANK, N.A. AS TRUSTEE, J. P. MORGAN',
                                'CHASE BANK-TRUSTEE, JPMORGAN',
                                'CHASE, J P MORGAN',
                                'CHASE, JP MORGAN',
                                'Chase Bank, JP Morgan',
                                'J P MORGAN CHASE BANK, .',
                                'MORGAN CAHSE BANK, JP',
                                'MORGAN CHASE, JP',
                                'TROTT&TROTT P.C., JP MORGAN CHASE BANK IN C/O'],
       'MORTGAGE ELECTRONIC' : ["INC, MORTGAGE ELECTRONIC REG. SYS'S",
                                  "INC, MORTGAGE ELECTRONIC SYS'S",
                                  'MORTGAGE ELECTRONIC  REG. SYS., .',
                                  'MORTGAGE ELECTRONIC REG.SYS.,I, .',
                                  'MORTGAGE ELECTRONIC REGIST., *',
                                  'REG SYS. INC, MORTGAGE ELECTRONIC',
                                  'REG. SYS., INC., MORTGAGE ELECTRONIC',
                                  'REG. SYSTEM, MORTGAGE ELECTRONIC',
                                  'REGISITRATION SYS. INC, MORTGAGE ELECTRONIC',
                                  'REGIST  SYSTEM, MORTGAGE ELECTRONIC',
                                  'REGISTARTION SYSTEM INC., MORTGAGE␣
↪ELECTRONIC',
                                  'REGISTATION  SYSTEM  INC., MORTGAGE␣
↪ELECTRONIC',
                                  'REGISTATION SYSTEM INC, MORTGAGE ELECTRONIC',
                                  'REGISTRARION SYSTEM INC., MORTGAGE␣
↪ELECTRONIC',
                                  'REGISTRATION   INC, MORTGAGE ELECTRONIC',
                                  'REGISTRATION  INC, MORTGAGE ELECTRONIC',
                                  'REGISTRATION  SYSTEM   INC., MORTGAGE␣
↪ELECTRONIC',
                                  'REGISTRATION  SYSTEM  INC., MORTGAGE␣
↪ELECTRONIC',
                                  'REGISTRATION  SYSTEM INC, MORTGAGE␣
↪ELECTRONIC',
                                  'REGISTRATION  SYSTEM, MORTGAGE ELECTRONIC',
                                  'REGISTRATION  SYSTEM,INC, MORTGAGE␣
↪ELECTRONIC',
                                  'REGISTRATION INC, MORTGAGE ELECTRONIC',
                                  'REGISTRATION SY, MORTGAGE ELECTRONIC',
                                  'REGISTRATION SY., MORTGAGE ELECTRONIC',
                                  'REGISTRATION SYM. INC., MORTGAGE ELECTRONIC',
                                  "REGISTRATION SYS'S, INC., MORTGAGE␣
↪ELECTRONIC",
                                  'REGISTRATION SYS, MORTGAGE ELECTRONIC',
                                  'REGISTRATION SYS. INC, MORTGAGE ELECTRONIC',
```

```
                                 'REGISTRATION SYS., MORTGAGE ELECTRONIC',
                                 'REGISTRATION SYSTEM  INC, MORTGAGE␣
↪ELECTRONIC',
                                 'REGISTRATION SYSTEM  INC., MORTGAGE␣
↪ELECTRONIC',
                                 'REGISTRATION SYSTEM INC, MORTGAGE ELECTRONIC',
                                 'REGISTRATION SYSTEM INC., MORTGAGE␣
↪ELECTRONIC',
                                 'REGISTRATION SYSTEM, MORTGAGE ELECTRONIC',
                                 'REGISTRATION SYSTEM,INC, MORTGAGE ELECTRONIC',
                                 'REGISTRATION SYSTEM,INC., MORTGAGE␣
↪ELECTRONIC',
                                 'REGISTRATION SYSTEMS  INC., MORTGAGE␣
↪ELECTRONIC',
                                 'REGISTRATION SYSTEMS INC, MORTGAGE␣
↪ELECTRONIC',
                                 'REGISTRATION SYSTEMS INC., MORTGAGE␣
↪ELECTRONIC',
                                 'REGISTRATION SYSTEMS, INC, MORTGAGE␣
↪ELECTRONIC',
                                 'REGISTRATION SYSTEMS, INC., MORTGAGE␣
↪ELECTRONIC',
                                 'REGISTRATION SYSTEMS, MORTGAGE ELECTRONIC',
                                 'REGISTRATION SYSTEMS,INC  , MORTGAGE␣
↪ELECTRONIC',
                                 'REGISTRATION SYSTEMS,INC, MORTGAGE␣
↪ELECTRONIC',
                                 'REGISTRATION SYSTEMS,INC, MORTGAGE␣
↪ELECTRONICS',
                                 'REGISTRATION SYSTEMS,INC., MORTGAGE␣
↪ELECTRONIC',
                                 'REGISTRATION, MORTGAGE ELECTRONIC',
                                 'REGRISTRATION SYSTEM INC, MORTGAGE␣
↪ELECTRONIC',
                                 'Registration Systems, Mortgage Electronic',
                                 'SY, MORTGAGE ELECTRONIC REGISTRATION',
                                 'SYSTEM INC, MORTGAGE ELECTRONIC␣
↪REAGISTRATION',
                                 'SYSTEM INC., MORTGAGE ELECTRONIC␣
↪REGISTRATION',
                                 "SYSTEM'S INC, MORTGAGE ELECTRONIC REGIS.",
                                 'SYSTEM, INC, MORTGAGE ELECTRONIC␣
↪REGISTRATION',
                                 'SYSTEM, MORTGAGE ELECTRONIC  REGISTRAT',
                                 'SYSTEMS INC., MORTGAGE ELECTRONIC',
```

```
                                'SYSTEMS, INC., MORTGAGE ELECTRONIC␣
↪REGISTRATION',
                                'SYSTEMS, MORTGAGE ELECTRONIC  REGISTRATION',
                                'SYSTEMS, MORTGAGE ELECTRONIC REGISTRATI',
                                'SYSTEMS, MORTGAGE ELECTRONIC REGISTRATION',
                                'SYSTEMS,INC, MORTGAGE ELECTRONIC␣
↪REGISTRATION',
                                'SYSTEMS,INC., MORTGAGE ELECTRONIC␣
↪REGISTRATION',

                                'ELECTRONIC   REGISTATION  SYST, MORTGAGE',
                                'ELECTRONIC  INC, MORTGAGE',
                                'ELECTRONIC  REG. SYS.,  MORTGAGE',
                                'ELECTRONIC MORT. REG. SYSTEM, .',
                                'ELECTRONIC REG  SYS INC, MORTGAGE',
                                'ELECTRONIC REGISTRATION , SHD MORTGAGE',
                                'ELECTRONIC REGISTRATION SYSTEM, MORTGAGE',
                                'ELECTRONIC REGISTRATION, MORTGAGE',
                                'ELECTRONIC SYSTEMS, MORTGAGE',
                                'ELECTRONIC, MORTGAGE',
                                'JONATHAN EGHMAN, MTG. ELECTRONIC REGISTRATION␣
↪S IN C/O',
                                'MORTGAGE ELECTRONIC',
                                'REGISRATION SYSTEM INC., MTG. ELECTRONIC',
                                'REGISTRATION  SY, MORTOR ELECTRONIC',
                                'REGISTRATION SYS. INC, MORTAGE  ELECTRONIC',
                                'REGISTRATION SYSTEM INC., MORTAGE ELECTRONIC',
                                'REGISTRATION SYSTEM, INC., MORTGAGE ␣
↪ELECTRONIC',
                                'REGISTRATION SYSTEMS, MORTGAGE  ELECTRONIC',
                                'REGISTRATION SYSTEMS,INC, MORTGAGE ␣
↪ELECTRONIC',
                                'REGISTRATION SYSTEMS. INC., MORTGAGE ␣
↪ELECTRONIC',
                                'REGISTRATION, MORTGAGE  ELECTRONIC',
                                'SYSTEM, ELECTRONIC REGISTRATION',
                                'SYSTEMS INC., MTG. ELECTRONIC REGISTRATION',
                                'SYSTEMS, INC, ELECTRONIC REGISTRATION',
                                'SYSTEMS, MTG.ELECTRONIC REGISTRATION',
                                'TROTT&TROTT ATTY., MTG.ELECTRONIC␣
↪REGISTRATION SY IN C/O',
                                'system inc., MTG.Electronic registration',
                                'system llc, mtg.electronic registration',
                                'system llc., mtg.electronic registration'],
       'FEDERAL NATIONAL MTG. A.' : ['ASSN, FEDERAL NATIONAL MTG',
                                     'ASSOC, FEDERAL NATIONAL MORTGAGE',
```

```
                                       'ASSOC.,  FEDERAL NATIONAL MTG.',
                                       'ASSOC., FEDERAL NATIONAL MORTGAGE',
                                       'ASSOC., FEDERAL NATIONAL MTG.',
                                       'ASSOCIATION, FEDERAL NATIONAL MORTGAGE',
                                       'FEDERAL NATIONAL MORTGAGE ASSO, *',
                                       'FEDERAL NATIONAL MORTGAGE ASSO, .',
                                       'FEDERAL NATIONAL MTG. ASSN., .',
                                       'MORT ASSOCIATES, FEDERAL NATIONAL',
                                       'MORTGAGE  ASS., FEDERAL NATIONAL',
                                       'MORTGAGE  ASSOC, FEDERAL NATIONAL',
                                       'MORTGAGE ASS, FEDERAL NATIONAL',
                                       'MORTGAGE ASSN., FEDERAL NATIONAL',
                                       'MORTGAGE ASSO., FEDERAL NATIONAL',
                                       'MORTGAGE ASSOC, FEDERAL NATIONAL',
                                       'MORTGAGE ASSOC. , FEDERAL NATIONAL',
                                       'MORTGAGE ASSOC., FEDERAL NATIONAL',
                                       'MORTGAGE ASSOCATION , FEDERAL NATIONAL',
                                       'MORTGAGE ASSOCIATION, FEDERAL NATIONAL',
                                       'MORTGAGE, FEDERAL NATIONAL',
                                       'MTG ASSOC., FEDERAL NATIONAL',
                                       'MTG, FEDERAL NATIONAL',
                                       'MTG. ASSOC., FEDERAL NATIONAL',
                                       'Mortgage Assoc., Federal National',
                                       'Mortgage Association, Federal National',
                                       'NORTGAGE ASSOCIATION, FEDERAL NATIONAL'],
            'BANK, HSBC' : ['BANK  USA, HSBC',
                            'BANK (USA), HSBC',
                            'BANK USA, HSBC',
                            'CORP III.HSBC  MTG SERVICES, HOUSEHOLD FINANCE',
                            'CORP III.HSBC MTG SERVICES, HOUSEHOLD FINANCE',
                            'MORTGAGE, HSBC',
                            'USA, HSBC BANK'],
            'SEPTEMBER PROPERTIES' : ['PROPERTIES, SEPTEMBER', 'PROPERTYS,␣
↪SEPTEMBER', 'SEPTEMBER PROPERTIES, .'],
            'DETROIT NEIGHBORHOOD DEVELOPMENT' : ['CO, DET  NEIGHBORHOOD␣
↪DEVELOPMENT',
                                     'CO, DETROIT NEIGHBOR HOOD DEVELOPMENT',
                                     'CO, DETROIT NEIGHBORHOOD DEVELOPMENT',
                                     'CORP., DETROIT NEIGHBORHOOD DEVELOPMENT',
                                     'DETROIT NEIGHBORHOOD DEV, NORTHWEST',
                                     'DETROIT NEIGHBORHOOD DEV., *',
                                     'DETROIT NEIGHBORHOOD DEVEL COR, 000000',
                                     'DETROIT NEIGHBORHOOD DEVELOPE, .',
                                     'DETROIT NEIGHBORHOOD DEVELOPME, *',
                                     'DETROIT NEIGHBORHOOD DVL. CORP, *',
                                     'DETROIT NEIGHBORHOOD DVLP., *',
                                     'DETROIT NEIGHBORHOOD SERV., *',
```

```
                                   'DETROIT NEIGHBORHOOD, NORTHWEST',
                                   'DEV. CORP., DETROIT NEIGHBORHOOD',
                                   'DEV., DETROIT NEIGHBORHOOD',
                                   'DEVELOPMENT, DETROIT NEIGHBORHOOD',
                                   'DEVELOPEMENT CO., DETROIT NEIGHBORHOOD',
                                   'DEVELOPEMENT, DETROIT  NEIGHBORHOOD',
                                   'DEVELOPEMENT, DETROIT NEIGHBORHOOD',
                                   'DEVELOPMENR CORPORATION, DETROIT␣
↪NEIGHBORHOOD',
                                   'DEVELOPMENT  CO., DETROIT NEIGHBORHOOD',
                                   'DEVELOPMENT  CORP., DETROIT NEIGHBORHOOD',
                                   'DEVELOPMENT CO, DETROIT NEIGHBORHOOD',
                                   'DEVELOPMENT CO., DETROIT  NEIGHBORHOOD',
                                   'DEVELOPMENT CO., DETROIT NEIGHBORHOOD',
                                   'DEVELOPMENT CORP, DETROIT NEIGHBORHOOD',
                                   'DEVELOPMENT CORP., NORTHWEST DETROIT␣
↪NEIGHBORHOOD',
                                   'DEVELOPMENT, DETROIT NEIGHBOR',
                                   'DEVELOPMENT, DETROIT NEIGHBORHOOD',
                                   'DEVELOPMENT, NORTHWESTDETROIT␣
↪NEIGHBORHOOD',
                                   'DEVELOPMENT.CO, DETROIT  NEIGHBORHOOD',
                                   'Detroit Neighborhood Dev. Co. , *',
                                   'Dev. Corp., Cass Corridor Neighborhood',
                                   'Development Co., Detroit Neighborhood',
                                   'Development Corp, Detroit Neighborhood',
                                   'Development Corp., Detroit Neighborhood',
                                   'Development, Detroit Neighborhood',
                                   'NEIGHBOR DEV., DETROIT',
                                   'NEIGHBORHOOD  DEV, NORTHWEST DETROIT',
                                   'NEIGHBORHOOD  DEV., NORTHWEST  DETROIT',
                                   'NEIGHBORHOOD  DEV., NORTHWEST DETROIT',
                                   "NEIGHBORHOOD DEV'T, DETROIT",
                                   'NEIGHBORHOOD DEV, DETROIT',
                                   'NEIGHBORHOOD DEV. CO., DETROIT',
                                   'NEIGHBORHOOD DEV., DETROIT',
                                   'NEIGHBORHOOD DEV., NORTHWEST  DETROIT',
                                   'NEIGHBORHOOD DEV., NORTHWEST DETROIT',
                                   'NEIGHBORHOOD DEVELOPMENT CO, DETROIT',
                                   'NEIGHBORHOOD DEVELOPMENT CORP, *',
                                   'NEIGHBORHOOD DEVELOPMENT, DETROIT',
                                   'NEIGHBORHOOD INC., CORE CITY',
                                   'NEIGHBORHOOD, *',
                                   'NEIGHBORHOOD, CORE CITY',
                                   'NEIGHBORHOOD, INC., CORE CITY',
                                   'NEIGHBORHOOD, NORTHWEST DETROIT',
                                   'NEIGHBORHOOD, SOUTHWEST ALLIANCE FOR',
```

```
                                    'NEIGHBORHOODS INC., CORE CITY',
                                    'NEIGHBORHOODS, CORE CITY',
                                    'NORTHWEST DETROIT NEIGHBORHOOD, .',
                                    'development co, detroit neighborhood'],
        'KAY BEE KAY PROPERTIES, .' : ['PROPERTIES L.L.C, KAY BEE KAY',
                                       'PROPERTIES, KAY BEE KAY',
                                       'Properties, LLC, Kay Bee Kay']





}

misspellings_inv = {}
for key in misspellings.keys():
    for value in misspellings[key]:
        misspellings_inv[value] = key
```

[301]:
```
df_fines['violator_name'] = df_fines['violator_name'].map(
    lambda name:
        misspellings_inv[name] if name in misspellings_inv.keys()
        else name
)
```

[302]:
```
selected = []
query = "KAY BEE KAY"
keys = df_fines['violator_name'].unique()
for key in keys:
    if query.lower() in key.lower():
        selected.append(key)
sorted(selected)
```

[302]:
```
['KAY BEE KAY PROPERTIES, .']
```

4.1 Combine the data

In this section we combine every datasets in one 'df_join'. This step is essential to data explaration and in the last step prediction.

[303]:
```
df_join = pd.merge(df_addresses, df_latlons,
                    how='inner', on='address')
df_join = pd.merge(df_fines, df_join,
                    how='inner', on='ticket_id')

df_join
```

[303]:

```
       ticket_id                                     agency_name  \
0          22056  Buildings, Safety Engineering & Env Department
1          22046  Buildings, Safety Engineering & Env Department
2          18735  Buildings, Safety Engineering & Env Department
3          18733  Buildings, Safety Engineering & Env Department
4          18743  Buildings, Safety Engineering & Env Department
…              …                                               …
11404      70848  Buildings, Safety Engineering & Env Department
11405      71141  Buildings, Safety Engineering & Env Department
11406      70928                      Department of Public Works
11407      70826  Buildings, Safety Engineering & Env Department
11408      71143  Buildings, Safety Engineering & Env Department

           inspector_name                      violator_name  \
0         Sims, Martinzie  INVESTMENT INC., MIDWEST MORTGAGE
1         Sims, Martinzie                     KASIMU, UKWELI
2        Williams, Darrin    Rafee Auto Services L.L.C., RAF
3        Williams, Darrin    Rafee Auto Services L.L.C., RAF
4        Williams, Darrin                Gardner Resale, GAR
…                     …                                  …
11404  Karwowski, Stephen                   BEVELLE, EARLENE
11405     Matthews, Delos                     NAZARETH, JUDY
11406       Cato, Valesta                     SMITH, ARMELLA
11407  Karwowski, Stephen                          TROUP, R.
11408     Matthews, Delos                       RUTTAN, JAMES

       violation_street_number violation_street_name  violation_zip_code  \
0                         2900                 TYLER                 NaN
1                         6478            NORTHFIELD                 NaN
2                         8228            MT ELLIOTT                 NaN
3                         8228            MT ELLIOTT                 NaN
4                         9100              VAN DYKE                 NaN
…                            …                     …           …
11404                     4421                FISHER                 NaN
11405                    18104               CARDONI                 NaN
11406                    12956             WESTBROOK                 NaN
11407                    18808               GODDARD                 NaN
11408                    19930               ANDOVER                 NaN

       mailing_address_str_number mailing_address_str_name           city  \
0                             3.0                S. WICKER        CHICAGO
1                          2755.0                  E. 17TH       LOG BEACH
2                          8228.0              Mt. Elliott        Detroit
3                          8228.0              Mt. Elliott        Detroit
4                            91.0                 Van Dyke        Detroit
…                               …                        …        …
11404                      4421.0                   FISHER        DETROIT
```

```
11405                          2915.0            WESTHAMPTON         SOUTHFIELD
11406                         12956.0               WESTWOOD            DETROIT
11407                         19125.0         CANTERBURY RD.            DETROIT
11408                          3872.0            VENETIAN DR  HARRISON TOWNSHIP

       … isMonday isTuesday  isWednesday isThursday isFriday isSaturday  \
0      …    False      True        False      False    False      False
1      …    False     False        False      False    False       True
2      …    False     False         True      False    False      False
3      …    False     False         True      False    False      False
4      …    False     False         True      False    False      False
…    …        …       …          …        …      …      …
11404  …    False      True        False      False    False      False
11405  …    False      True        False      False    False      False
11406  …    False      True        False      False    False      False
11407  …    False      True        False      False    False      False
11408  …    False      True        False      False    False      False

      isSunday                      address        lat        lon
0        False         2900 tyler, Detroit MI  42.390729 -83.124268
1        False    6478 northfield, Detroit MI  42.145257 -83.208233
2        False    8228 mt elliott, Detroit MI  42.388641 -83.037858
3        False    8228 mt elliott, Detroit MI  42.388641 -83.037858
4        False      9100 van dyke, Detroit MI  42.395765 -83.022333
…          …                            …        …        …
11404    False        4421 fisher, Detroit MI  42.374758 -83.003211
11405    False      18104 cardoni, Detroit MI  42.426510 -83.090686
11406    False    12956 westbrook, Detroit MI  42.383586 -83.249566
11407    False      18808 goddard, Detroit MI  42.430707 -83.077756
11408    False      19930 andover, Detroit MI  42.440180 -83.098919

[11409 rows x 47 columns]
```

4.2 Data exploration

4.2.1 How many fines there are after data cleaning?

[304]:
```
len(df_fines.index)
```

[304]: 42023

4.2.2 Count numbers of fines by violator_name top 10

[305]:
```
df_fine_counter=df_fines['violator_name'].value_counts()
df_fine_counter.head(10)
```

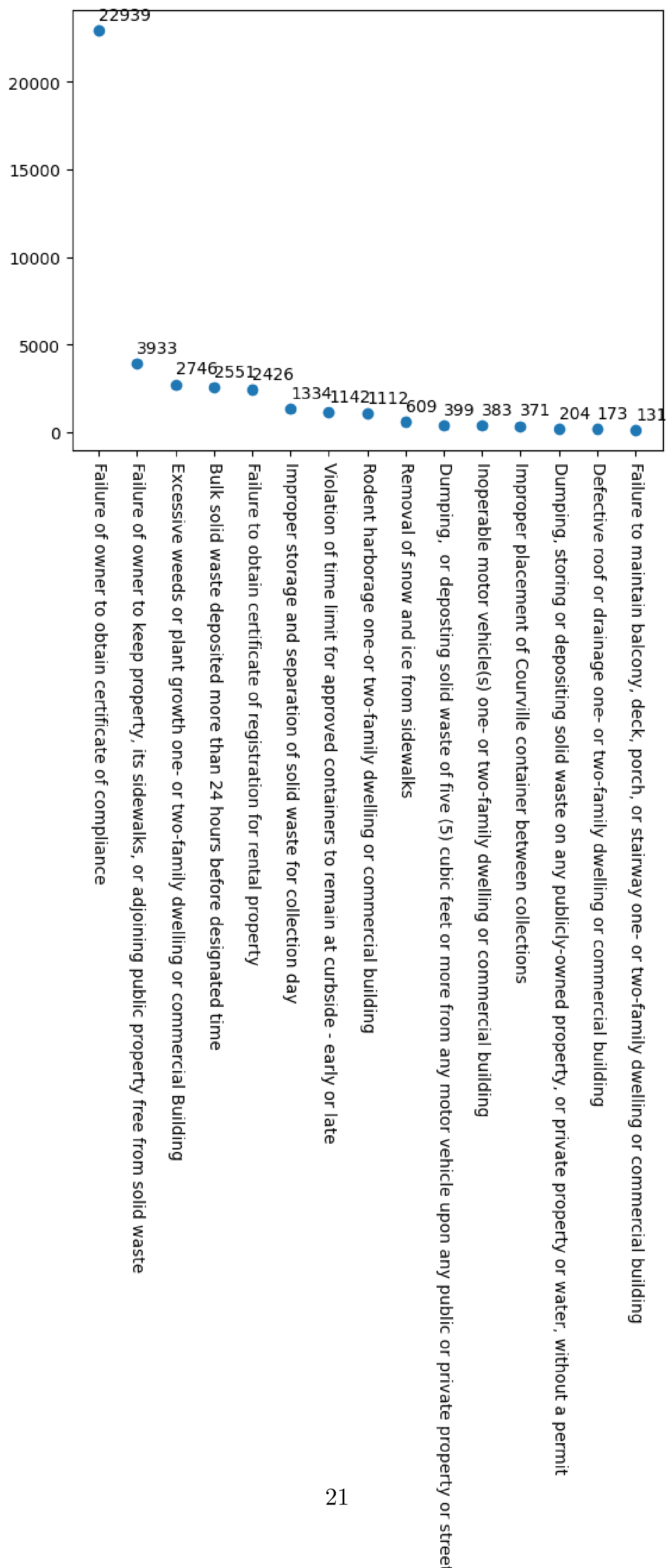
[305]:
```
violator_name
INVESTMENT, ACORN                        692
BANK, WELLS FARGO                        156
MORTGAGE ELECTRONIC                      138
DETROIT NEIGHBORHOOD DEVELOPMENT         131
BANK, DEUTSCHE                           121
MANAGEMENT CO., WINGATE                   62
NEW YORK, BANK OF                         61
BANK, JP MORGAN CHASE                     60
FEDERAL NATIONAL MTG. A.                  59
MILLER, JOHN                              50
Name: count, dtype: int64
```

4.2.3 Header of the DataFrame

4.2.4 Counts of most common fine types

We presented the quantities of each type of fines issued. The most of quantities was issued for 'Failure of owner to obtain certificate of compliance'.

[306]:
```
numberOfTimes = df_fines['violation_description'].value_counts()
number = pd.DataFrame({'description':numberOfTimes.index,'times':numberOfTimes.
 ↪values}).head(n=15) #e just want the first 15 because are the most relevant
plt.scatter(number['description'],number['times'])
plt.xticks(number['description'], rotation=-90)
for i, row in number.iterrows():
    plt.text(row['description'], row['times']+600, s= str(row['times']))
plt.show()
```


22939
3933
2746
2551
2426
1334
1142
1112
609
399
383
371
204
173
131
20000
15000
10000
5000
0
Failure of owner to obtain certificate of compliance
Failure of owner to keep property, its sidewalks, or adjoining public property free from solid waste
Excessive weeds or plant growth one- or two-family dwelling or commercial Building
Bulk solid waste deposited more than 24 hours before designated time
Failure to obtain certificate of registration for rental property
Improper storage and separation of solid waste for collection day
Violation of time limit for approved containers to remain at curbside - early or late
Rodent harborage one-or two-family dwelling or commercial building
Removal of snow and ice from sidewalks
Dumping, or deposting solid waste of five (5) cubic feet or more from any motor vehicle upon any public or private property or street
Inoperable motor vehicle(s) one- or two-family dwelling or commercial building
Improper placement of Courville container between collections
Dumping, storing or depositing solid waste on any publicly-owned property, or private property or water, without a permit
Defective roof or drainage one- or two-family dwelling or commercial building
Failure to maintain balcony, deck, porch, or stairway one- or two-family dwelling or commercial building

4.2.5 Which addresses recieved the most tickets? (Common offenders)

4.2.6 The top 10 streets where were the most fines issued.

[307]:
```
df_fine_counter_by_street=df_fines['violation_street_name'].value_counts()
df_fine_counter_by_street.head(10)
```

[307]:
```
violation_street_name
SEVEN MILE     845
MCNICHOLS      682
LIVERNOIS      447
EVERGREEN      380
FENKELL        303
JOY RD         299
WARREN         290
EIGHT MILE     264
PURITAN        254
GRAND RIVER    251
Name: count, dtype: int64
```

4.2.7 Where were the tickets issued?

In this section we are created a map with points which mean lokalization fines issued. This kind of grafs are very helpfull to imagin where is the biggest problem with not maintanenence propertys.

[308]:
```
import plotly.express as px
import pandas as pd
df_latlons.dropna(
    axis=0,
    how='any',
    subset=None,
    inplace=True
)

color_scale = [(0, 'orange'), (1,'red')]

fig = px.scatter_mapbox(df_latlons,
                        lat="lat",
                        lon="lon",

                        zoom=8,
                        height=800,
                        width=800)

fig.update_layout(mapbox_style="open-street-map")
```

```
fig.update_layout(margin={"r":0,"t":0,"l":0,"b":0})
fig.show()
```

5 Feature Engineering

In this part of the process we wanted to create new features from the raw data to enhance our dataset. The raw data ussualy isn't enough for a machine learning task, so this step is quite commonly needed. We tried to come up with an idea of what can be typical for blight violatiors, and then we tried to extract this new characteristic out of our raw data.

It was also at this step when we were really thankful for discovering the original paper that was posted after the competition held by the city of detroit on this dataset. We had a lot ouf our own ideas about new features, but this paper helped us to better understand our dataset, a thing for which we were much grateful down the line.

The paper can be accessed at https://midas.umich.edu/wp-content/uploads/sites/3/2017/09/understanding-blight-ticket.pdf

5.0.1 Is owner flag

In this part we added new column ('isOwner'). This column crudely indicates if the violator is also the owner of the blighted property. Our heuristic is that if 'mailing_address_str_number' equals 'violation_street_number' it indicates that the violator is the owner.

[309]:
```
df_fines['isOwner'] = np.where(df_fines['mailing_address_str_name'].astype(str).
 ↪str.upper() == df_fines['violation_street_name'].astype(str).str.upper(),␣
 ↪1,0)
df_fines['isOwner'] = np.where(df_fines['mailing_address_str_number'].
 ↪astype(float) == df_fines['violation_street_number'].astype(float),1,0)
df_join['isOwner'] = df_fines['isOwner']
```

5.0.2 Past compliance rate

We made an assumption that a person with many unpaid blight violations in the past may have more problems with paying the tickets in the future. To represent this assumption we added average compliance rate computed from the historical data.

[310]:
```
df_compliance_rate = df_join.
 ↪groupby('violator_name',as_index=False)['compliance'].mean()
df_compliance_rate.rename(columns={'compliance':'violator_compliance_rate'},␣
 ↪inplace=True)
df_join = df_join.merge(df_compliance_rate, how='left', on='violator_name')
```

5.0.3 Total money owed

It is quite clear if a subject already owes the city a lot in unpaid fines, then he is also quite unlikely to comply with a future ticket

[311]:
```
df_fine_total = df_join.groupby('violator_name',as_index=False)['balance_due'].
    ↪sum()
df_fine_total.rename(columns={'balance_due':'total_balance_due'}, inplace=True)
df_join = df_join.merge(df_fine_total, how='left', on='violator_name')
```

Spatial features The raw data contains a very good resource of spatial information that we have not used so far. We wanted to create a simple feature that would carry information about the region around the property. However, comparing streets (some short, some running through the whole city) or different quarters (often with senslessly drawn borders) didn't really really appeal to us.

Therefore we decided to create our own division of the area: with a use of K-means clustering we have divided the city into 75 clusters. We wanted to create homogenous relatively small clusters and their count couldn't be estimated and wasn't important to us. After that we calculated the average compliance rate within the cluster.

[312]:
```
data_for_clustering = df_join[['lat', 'lon']]



kmeans = KMeans(n_clusters=75,n_init='auto')
kmeans.fit(data_for_clustering[['lat', 'lon']])

df_join['cluster'] = kmeans.labels_

payment_rates = df_join.groupby('cluster')['compliance'].mean().
    ↪sort_values(ascending=False)
```

[313]:
```
map_points= folium.Map(data_for_clustering=[df_join['lat'].mean(),␣
    ↪df_join['lon'].mean()], zoom_start=12)
map_poligon = folium.Map(data_for_clustering=[df_join['lat'].mean(),␣
    ↪df_join['lon'].mean()], zoom_start=12)




colors_poligon= ['#1f77b4', '#ff7f0e', '#2ca02c', '#d62728', '#9467bd',␣
    ↪'#8c564b', '#e377c2', '#7f7f7f', '#bcbd22', '#17becf']
colors_points= ['#' + ''.join([random.choice('0123456789ABCDEF') for j in␣
    ↪range(6)]) for i in range(10)]
#poligon_map
for cluster_label in range(75):
    color = colors_points[cluster_label % 10]
    points = df_join[kmeans.labels_ == cluster_label][['lat', 'lon']].values.
    ↪tolist()
    folium.Polygon(locations=points, color=color, fill=True, fill_opacity=0.3).
    ↪add_to(map_poligon)

    center = np.mean(points, axis=0)
```

```
    folium.Marker(location=[center[0], center[1]],
                  icon=folium.DivIcon(html=f'<div style="font-size: 10pt; color:
↪ black;">{cluster_label}</div>')).add_to(map_poligon)

#points_map

for lat, lon, cluster in zip(df_join['lat'], df_join['lon'], kmeans.labels_):
    folium.CircleMarker(
        location=[lat, lon],
        radius=5,
        color=None,
        fill=True,
        fill_opacity=0.7,
        fill_color= colors_points[cluster%10]
    ).add_to(map_points)

map_poligon.fit_bounds(map_poligon.get_bounds())
map_points.fit_bounds(map_points.get_bounds())
```

Now with all the info we print first a map that each dot represents a fine and the colour is the cluster that is part of.

[314]:
```
map_points
```

[314]: <folium.folium.Map at 0x7ffad225e410>

Also we make another map but these one is with all the dot makeing the cluster with a poligon that represents the cluster and the numbered cluster for identifying the cluster.

[315]:
```
map_poligon
```

[315]: <folium.folium.Map at 0x7ffad2270580>

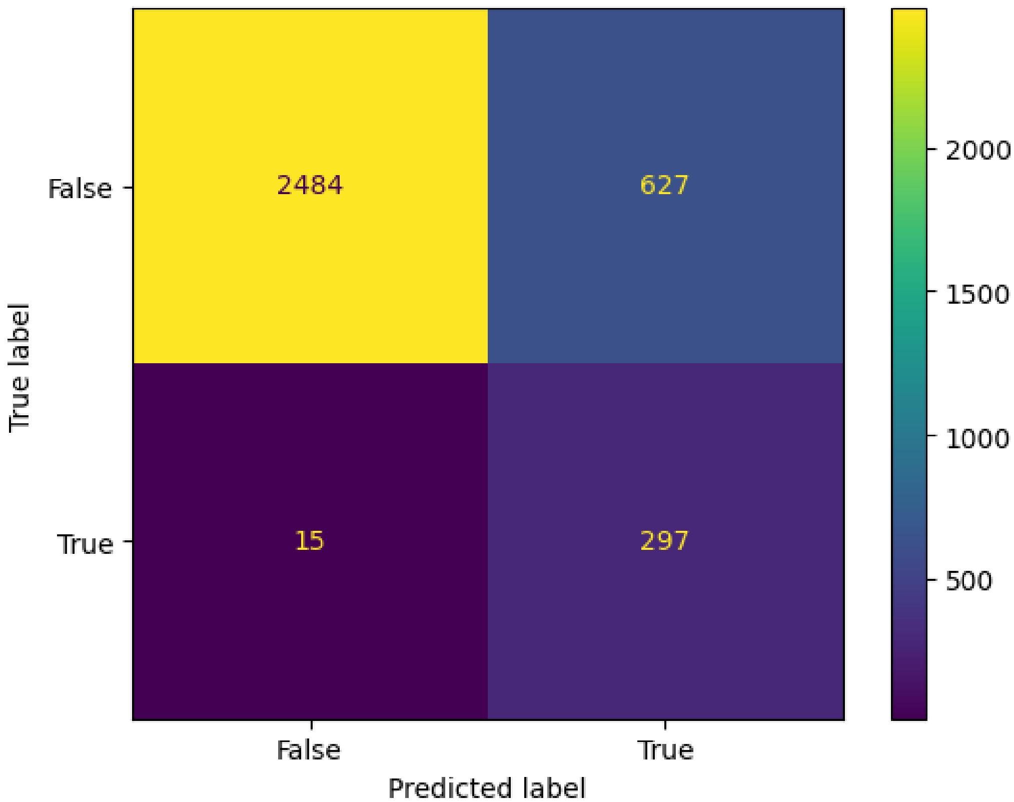
5.1 Predictive modeling

We use the Decision Tree to predict the fines

[316]:
```
df_join = df_join.merge(payment_rates, on='cluster', how='inner', suffixes=('',␣
↪'_cluster'))
df_join = df_join.rename(columns={'compliance_cluster':'cluster_compliance'})

dtv = df_join[['fine_amount','late_fee','isMonday', 'isTuesday', 'isWednesday',
       'isThursday', 'isFriday', 'isSaturday', 'isSunday', 'isOwner',␣
↪'cluster_compliance', 'violator_compliance_rate','total_balance_due']]
payStatus = df_join[['compliance']]
```

[318]:

```
clf = DecisionTreeClassifier(criterion='log_loss', class_weight="balanced",␣
↪max_depth=8)

train_x, test_x, train_y, test_y  = train_test_split(dtv,payStatus,test_size=0.
↪3,random_state=1)



clf = clf.fit(train_x,train_y)

prediction = clf.predict(test_x)

ConfusionMatrixDisplay.from_estimator(clf, test_x, test_y)

plt.show()

print("Accuracy:\t\t",metrics.accuracy_score(test_y, prediction))
print("Validation dataset mean:\t", 1 - test_y['compliance'].mean())
print("Difference:\t\t", metrics.accuracy_score(test_y, prediction) - (1 -␣
↪test_y['compliance'].mean()))

dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,
                filled=True, rounded=True,
                special_characters=True,feature_names = dtv.
↪columns,class_names=['0','1'])
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('tree.png')
```




```
Accuracy:                 0.8124452234881683
Validation dataset mean:          0.908851884312007
Difference:               -0.09640666082383875
```

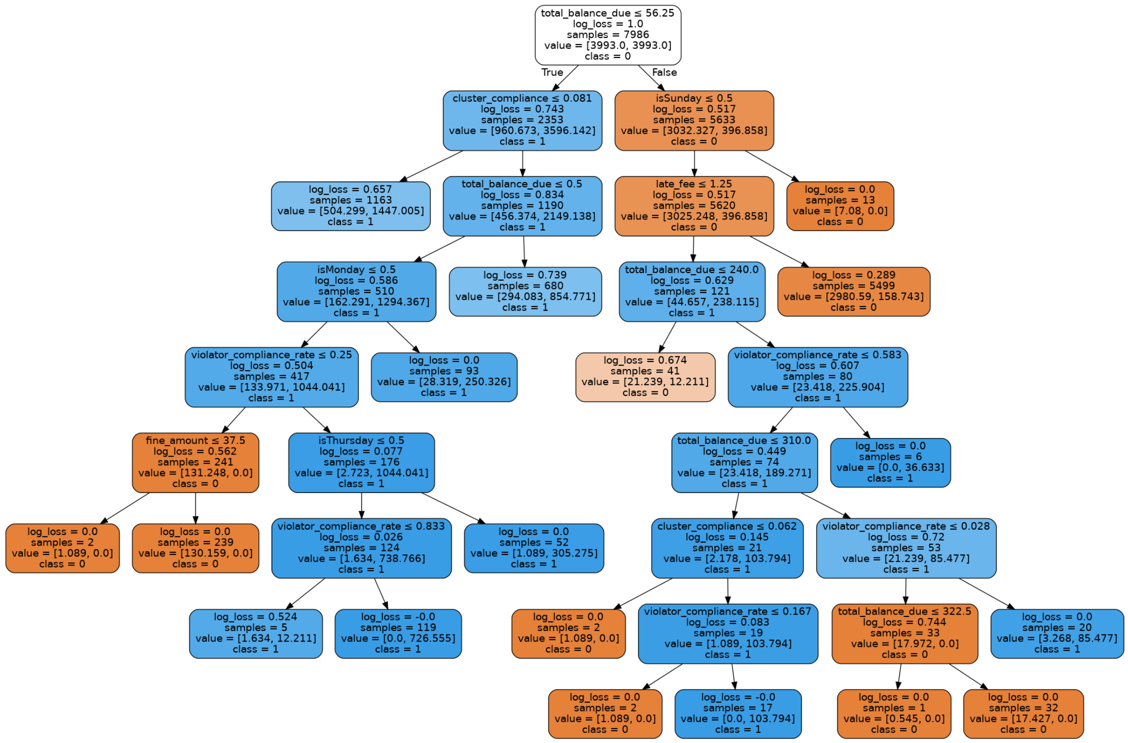
```
[318]: True
```

6 Conclusion

Our's reachers allows us to drowing conclusion about regarding situation property in Detroid. We detected the following issues: - 90% fines issued are unpaid - the biggest violators are companys this companys don't pay every fines which they should. - The of vast majority fines are issued with reason "Failure of owner to obtain certificate of compliance"

The next steps of reserch should be answers on the following question: - Why a lot of companys dont't pay the fines? - What regulation we can estabilish to improve the enforce the payment of the fines?

In our's project we used a few librarys allows clean data, visualization data on maps, created clasters, implement alghoritm to predict, generete decision trees.



Based on decistion tree diagram, we drawing conclusion that more important parameter to prediction whether fine will be pay or no is 'violetor_compliance_rate'. This value for each violator was calculeted based on how much fines paid or no by a particular violator. The second important column is 'late_fee' too this values was given to us by dataset. Last important column is 'Day_of_week' becouse, depending by day violetors are more likely to pay the fine.

To analysis of indyvidual parts of the Detroid as regard pay fines we created districts used a kmeans metod to classification. This metod is characterized by efficiency, simplicity and flexibility. Our's data don't have a outliers thats why kmean method is good choice. The results of the metod we can see on the maps. This metod generated clasters very well.

In this case we have used the Decision Tree because we think that is the one in wich we can visualize the process of predicting in the easiest why possible. Also we choosed this algorithm because it'a supervised learning algorithm, it's a non linear algorithm and it's a greedy algorithm so it is always goingo to choose the the closest solution.