In [1]:
```
import pandas as pd

df = pd.read_csv('Speeding/Violations_2017_Speeding_geoloc.csv')
```

In [2]:
```
from difflib import SequenceMatcher

#text = """there are
#some toyot in my text
#but I cannot find them"""

df = df[:100]
df['Make2'] = 'Make'
```

In [3]:
```
df.head()
```

Out[3]:

	Date Of Stop	Time Of Stop	Agency	SubAgency	Description	Location	Latitude	Longitude
0	8/28/2017	23:13:00	MCP	6th district, Gaithersburg / Montgomery Village	EXCEEDING POSTED MAXIMUM SPEED LIMIT: 39 MPH I...	LOST KNIFE RD. @ LOST KNIFE CIR	39.157853	-77.204095
1	8/28/2017	22:53:00	MCP	6th district, Gaithersburg / Montgomery Village	EXCEEDING POSTED MAXIMUM SPEED LIMIT: 45 MPH I...	WOODFIELD RD / EMORY GROVE RD	39.152928	-77.171317
2	8/29/2017	1:44:00	MCP	1st district, Rockville	EXCEEDING POSTED MAXIMUM SPEED LIMIT: 50 MPH I...	SEVEN LOCKS RD/CADBURY AVE	39.083948	-77.153498
3	8/29/2017	7:55:00	MCP	2nd district, Bethesda	EXCEEDING POSTED MAXIMUM SPEED LIMIT: 49 MPH I...	W/B UNIVERSITY BLVD @ NEWPORT MILL	39.033387	-77.072528
4	8/29/2017	8:07:00	MCP	2nd district, Bethesda	EXCEEDING POSTED MAXIMUM SPEED LIMIT: 49 MPH I...	W/B UNIVERSITY BLVD @ NEWPORT MILL	39.032983	-77.073839

5 rows × 36 columns

In [12]:
```
def fuzzy_search(search_key, text, strictness):
    lines = text.split("\n")
    for i, line in enumerate(lines):
        words = line.split()
        for word in words:
            similarity = SequenceMatcher(None, word, search_key)
            if similarity.ratio() > strictness:
                #return " '{}' matches: '{}' in line {}".format(search_key, word, i+
1)
                return search_key

make_list = ['ACURA', 'ASTON MARTIN', 'AUDI', 'BMW', 'BENTLEY', 'BUICK',
             'CADILLAC', 'CHEVROLET', 'CHRYSLER', 'DODGE', 'DUCATI', 'FERARRI',
             'FIAT', 'FORD', 'FREIGHTLINER', 'GEO', 'GMC', 'GILLIG', 'GRUMMAN',
             'HARLEY', 'HINO', 'HONDA', 'INFINITI', 'ISUZU', 'JAGUAR', 'HUMMER',
             'HYUNDAI', 'INTERNATIONAL', 'JEEP', 'KAWASAKI', 'KENWORTH',
             'KIA', 'LANDROVER', 'LEXUS', 'LINCOLN', 'MACK', 'MASERATI',
             'MAZDA', 'MERCEDES', 'MERCURY', 'MINI', 'MITSUBISHI', 'NISSAN',
             'OLDSMOBILE', 'PETERBILT', 'PLYMOUTH', 'POLARIS',
             'PONTIAC', 'PORSCHE', 'RANGE ROVER', 'SATURN', 'SCION',
             'SAAB', 'SMART', 'SUBARU', 'SUZUKI', 'TESLA', 'TOYOTA',
             'VOLKSWAGEN', 'VOLVO', 'YAMAHA']

not_found = 0

for index, text in df.iterrows():
    found = 0

    #Special fix for Chevrolet...
    if text['Make'] == 'CHEVY' or text['Make']=='CHEV':
        text['Make'] = 'CHEVROLET'

    #Fix for Volkswagen...
    if text['Make'] == 'VW':
        text['Make'] = 'VOLKSWAGEN'

    #fix for Oldsmobile...
    if text['Make'] == 'OLDS':
        text['Make'] = 'OLDSMOBILE'

    #RAM is a type of Dodge...
    if text['Make'] == 'RAM':
        text['Make'] = 'DODGE'

    #MERZ is abbreviation for Mercedes...
    if text['Make'] == 'MERZ':
        text['Make'] = 'MERCEDES'

    #MITZ and MITS is Mitsubishi...
    if text['Make'] == 'MITS' or text['Make'] == 'MITZ':
        text['Make'] = 'MITSUBISHI'

    #HD is Harley
    if text['Make'] == 'HD' or text['Make'] == 'H D':
        text['Make'] = 'HARLEY'

    #INTL is international:
    if text['Make'] == 'INTL':
        text['Make'] = 'INTERNATIONAL'
```

```
    #CADI is cadillac:
    if text['Make'] == 'CADI':
        text['Make'] = 'CADILLAC'

    #print(text)
    for item in make_list:
        if ((fuzzy_search(item, text['Make'], 0.65) != None) and found == 0):
            #print (fuzzy_search(item, text['Make'], 0.65))
            df.at[index, 'Make2'] = (fuzzy_search(item, text['Make'], 0.65))
            found = 1


    if found == 0:
        #print(text)
        not_found += 1
        df.at[index, 'Make2'] = 'Unknown'
```

In [13]:
```
not_found
```

Out[13]: 2

In [15]:
```
pd.set_option('display.max_colwidth', -1)
print(df[['Make','Make2']])
```

	Make	Make2
0	FORD	FORD
1	SUBARU	SUBARU
2	TOYT	TOYOTA
3	HONDA	HONDA
4	SUBARU	SUBARU
5	BMW	BMW
6	ACURA	ACURA
7	TOYOTA	TOYOTA
8	FORD	FORD
9	JEEP	JEEP
10	SUBA	SUBARU
11	TOYT	TOYOTA
12	TOYOTA	TOYOTA
13	ACURA	ACURA
14	MERCEDES	MERCEDES
15	FORD	FORD
16	TOYT	TOYOTA
17	BMW	BMW
18	LNDR	Unknown
19	HOND	HONDA
20	NISS	NISSAN
21	MAZDA	MAZDA
22	GMC	GMC
23	DODGE	DODGE
24	HYUN	HYUNDAI
25	TOYOTA	TOYOTA
26	JAG	JAGUAR
27	CADI	CADILLAC
28	TOYT	TOYOTA
29	HONDA	HONDA
..	...	...
70	HONDA	HONDA
71	TOYT	TOYOTA
72	HOND	HONDA
73	HONDA	HONDA
74	TOYT	TOYOTA
75	NISS	NISSAN
76	CHEV	CHEVROLET
77	FORD	FORD
78	DODGE	DODGE
79	JEEP	JEEP
80	NISS	NISSAN
81	TOYT	TOYOTA
82	TOYT	TOYOTA
83	TOYT	TOYOTA
84	NISSAN	NISSAN
85	FORD	FORD
86	LEXS	LEXUS
87	KIA	KIA
88	TOYT	TOYOTA
89	VOLV	VOLVO
90	CHEV	CHEVROLET
91	NISS	NISSAN
92	TOYT	TOYOTA
93	BMW	BMW
94	VOLV	VOLVO
95	HOND	HONDA
96	BMW	BMW
97	TOYOTA	TOYOTA

```
98  NISSAN     NISSAN
99  CHEVROLET  CHEVROLET

[100 rows x 2 columns]
```