In [1]:

```
########## Import All common Library and Setting up folder
import os
import time
import sys
from datetime import datetime

# Data science common library
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
import random
import csv
import sqlite3


# Common folder for all raw data
Path_Raw_Data=r'C:\Users\HP\Desktop\Data Scientist\DB_Sample_Data\\'

#update the folder name
Path=r'C:\Users\HP\Desktop\Data Scientist\fuzzywuzzy\Python'
Path_Data=Path+'\\Data\\'

# Changing path to currrent folder
os.chdir(Path)

# Setting sqlite3 DB
db_sqlite3=sqlite3.connect(Path_Data+'Python.db')

print(datetime.now())
```

```
2022-09-20 00:13:46.017352
```

In [2]:

```
########## Data Loading
print(datetime.now())
'''
df_sales = pd.read_csv(Path_Data+'sales.csv',delimiter=',',quotechar='"'
                       #,dtype={'Store':str}
                       ,encoding='ISO-8859-1',low_memory=False
                      )
# write DFto DB as table
df_sales.to_sql('df_sales',db_sqlite3,if_exists='replace')
'''

# reading from xls file
xls=pd.ExcelFile(Path_Data+'DB_TXN_List.xlsx')

DB_TXN_List=pd.read_excel(xls,'Data',dtype={'Marchant_name':str})
# write DFto DB as table
DB_TXN_List.to_sql('DB_TXN_List',db_sqlite3,if_exists='replace')

# reading from xls file
xls=pd.ExcelFile(Path_Data+'Merchant_List.xlsx')

Merchant_List=pd.read_excel(xls,'Data',dtype={'Marchant_name':str})
# write DFto DB as table
Merchant_List.to_sql('Merchant_List',db_sqlite3,if_exists='replace')

print(datetime.now())
```

```
2022-09-20 00:13:47.655537
2022-09-20 00:13:49.177383
```

In [3]:

```
print(DB_TXN_List)
print("\n")
print(Merchant_List)
```

```
    Marchant_name    AMT  CT
0            Noon  200.0   5
1         Noon db   34.0   1
2        Noon Abu   34.0   1
3       Noon Food  345.0  34
4        Nood gro   34.0   1
5          Amazon  123.1  16
6       Amazon db  121.0  18
7      Amazon Abu  118.9  21
8     Amazon Food  116.8  23
9      Amazon gro  114.7  26
10           Lulu  112.6  28
11        Lulu db  110.5  31
12       Lulu Abu  108.4  33
13      Lulu Food  106.3  36
14       Lulu gro  104.2  38
15            KML  102.1  41
16         KML db  100.0  43
17        KML Abu   97.9  46
18       KML Food   95.8  48
19        KML gro   93.7  51
20          DFGKJ   91.6  53
21       DFGKJ db   89.5  56
22      DFGKJ Abu   87.4  58
23     DFGKJ Food   85.3  61
24      DFGKJ gro   83.2  63
25       DFdghq23   81.1  66
26    DFdghq23 db   79.0  68
27   DFdghq23 Abu   76.9  71
28  DFdghq23 Food   74.8  73
29   DFdghq23 gro   72.7  76


  Marchant_name
0          Noon
1        Amazon
2          Lulu
3           KML
4         DFGKJ
5      DFdghq23
```

In [4]:

```
#pip install fuzzywuzzy
#pip install python-Levenshtein
from fuzzywuzzy import fuzz
from fuzzywuzzy import process
```

In [5]:

```
s1 = "mariners vs angels"
s2 = "los angeles angels of anaheim at seattle mariners"
print(fuzz.ratio(s1,s2))
print(fuzz.partial_ratio(s1,s2))
print(fuzz.token_sort_ratio(s1,s2))
print(fuzz.WRatio(s1,s2))
print(fuzz.token_set_ratio(s1,s2))
```

```
24
62
51
86
91
```

In [7]:

```
#ML_MName=Merchant_List
ML_MName=Merchant_List.iloc[:,0]
print(ML_MName)

DB_MName =DB_TXN_List.iloc[:,0]
print(DB_MName)
```

```
0        Noon
1      Amazon
2        Lulu
3         KML
4       DFGKJ
5    DFdghq23
Name: Marchant_name, dtype: object
0              Noon
1           Noon db
2          Noon Abu
3         Noon Food
4          Nood gro
5            Amazon
6         Amazon db
7        Amazon Abu
8       Amazon Food
9        Amazon gro
10             Lulu
11          Lulu db
12         Lulu Abu
13        Lulu Food
14         Lulu gro
15              KML
16           KML db
17          KML Abu
18         KML Food
19          KML gro
20            DFGKJ
21         DFGKJ db
22        DFGKJ Abu
23       DFGKJ Food
24        DFGKJ gro
25         DFdghq23
26      DFdghq23 db
27     DFdghq23 Abu
28    DFdghq23 Food
29     DFdghq23 gro
Name: Marchant_name, dtype: object
```

In [8]:

```
print(len(ML_MName),len(DB_MName))
```

```
6 30
```

In [9]:

```
Temp_ML_MName=[]
Temp_DB_MName=[]
Temp_ratio=[]
for a in range(len(ML_MName)):
    for b in range(len(DB_MName)):
        if(fuzz.token_set_ratio(ML_MName[a],DB_MName[b])>=90):
            #print(fuzz.token_set_ratio(ML_MName[a],DB_MName[b]))
            #print(DB_MName[b])
            Temp_ML_MName.append(ML_MName[a])
            Temp_DB_MName.append(DB_MName[b])
            Temp_ratio.append(fuzz.token_set_ratio(ML_MName[a],DB_MName[b]))

print(Temp_ML_MName)
print(Temp_DB_MName)
print(Temp_ratio)
```

```
['Noon', 'Noon', 'Noon', 'Noon', 'Amazon', 'Amazon', 'Amazon', 'Amazon',
'Amazon', 'Lulu', 'Lulu', 'Lulu', 'Lulu', 'Lulu', 'KML', 'KML', 'KML', 'KM
L', 'KML', 'DFGKJ', 'DFGKJ', 'DFGKJ', 'DFGKJ', 'DFGKJ', 'DFdghq23', 'DFdgh
q23', 'DFdghq23', 'DFdghq23', 'DFdghq23']
['Noon', 'Noon db', 'Noon Abu', 'Noon Food', 'Amazon', 'Amazon db', 'Amazo
n Abu', 'Amazon Food', 'Amazon gro', 'Lulu', 'Lulu db', 'Lulu Abu', 'Lulu
Food', 'Lulu gro', 'KML', 'KML db', 'KML Abu', 'KML Food', 'KML gro', 'DFG
KJ', 'DFGKJ db', 'DFGKJ Abu', 'DFGKJ Food', 'DFGKJ gro', 'DFdghq23', 'DFdg
hq23 db', 'DFdghq23 Abu', 'DFdghq23 Food', 'DFdghq23 gro']
[100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 10
0, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100]
```

In [10]:

```
df = pd.DataFrame(
    {'Marchant_name': Temp_ML_MName,
     'Marchant_name_DB': Temp_DB_MName,
     'Match_Ratio': Temp_ratio
    })
df
```

Out[10]:

	Marchant_name	Marchant_name_DB	Match_Ratio
0	Noon	Noon	100
1	Noon	Noon db	100
2	Noon	Noon Abu	100
3	Noon	Noon Food	100
4	Amazon	Amazon	100
5	Amazon	Amazon db	100
6	Amazon	Amazon Abu	100
7	Amazon	Amazon Food	100
8	Amazon	Amazon gro	100
9	Lulu	Lulu	100
10	Lulu	Lulu db	100
11	Lulu	Lulu Abu	100
12	Lulu	Lulu Food	100
13	Lulu	Lulu gro	100
14	KML	KML	100
15	KML	KML db	100
16	KML	KML Abu	100
17	KML	KML Food	100
18	KML	KML gro	100
19	DFGKJ	DFGKJ	100
20	DFGKJ	DFGKJ db	100
21	DFGKJ	DFGKJ Abu	100
22	DFGKJ	DFGKJ Food	100
23	DFGKJ	DFGKJ gro	100
24	DFdghq23	DFdghq23	100
25	DFdghq23	DFdghq23 db	100
26	DFdghq23	DFdghq23 Abu	100
27	DFdghq23	DFdghq23 Food	100
28	DFdghq23	DFdghq23 gro	100

In [12]:

```
df.to_csv("Final.csv",index = True,quotechar='"'
                 ,quoting=csv.QUOTE_ALL)
```