In [3]:
```
# Noah Manz
# 09-08-2022
# Code written for Rena Scott to drop duplicates of mailing addresses from 3 different .csv files

# Import relevant libraries
import pandas as pd
import os

# Load .csv files into memory
E1 = pd.read_csv(os.path.join(r'C:\Users\Noah\Desktop\Elections', 'E1.csv'))
E2= pd.read_csv(os.path.join(r'C:\Users\Noah\Desktop\Elections', 'E2.csv'))
E3 = pd.read_csv(os.path.join(r'C:\Users\Noah\Desktop\Elections', 'E3.csv'))
```

In [4]:
```
# Look at a piece of the dataframe
E1.head()
```

Out[4]:

	Voter ID	Last Name	First Name	Name Suffix	Mailing Address	Mailing City	Mailing Zip	Party	Status	Status Reason	Gender	Magis
0	[redacted]	[redacted]		NaN	[redacted]	KIRTLAND	87417	NO PARTY/INDEPENDENT/DECLINED TO SELECT	Active	Active Registrant	M	DIVISI
1	[redacted]	[redacted]		NaN	[redacted]	KIRTLAND	87417	CONSTITUTION	Inactive	Confirmation Mailing	M	DIVISI
2	[redacted]	[redacted]		NaN	[redacted]	FARMINGTON	87401-3761	NO PARTY/INDEPENDENT/DECLINED TO SELECT	Active	Active Registrant	F	DIVISI
3	[redacted]	[redacted]		NaN	[redacted]	FRUITLAND	87416-0613	NO PARTY/INDEPENDENT/DECLINED TO SELECT	Active	Active Registrant	F	DIVISI
4	[redacted]	[redacted]		NaN	[redacted]	FRUITLAND	87416-0613	DEMOCRATIC	Active	Active Registrant	M	DIVISI

In [2]:
```
# Combine all 3 .csv files into one dataframe
newfile = pd.concat([E1, E2, E3])

# Drop entires (keeping one) with duplicates in both the Mailing Address and Mailing City columns
NEWFILE = newfile.drop_duplicates(subset = ['Mailing Address', 'Mailing City'])
```

```
# Compare the shapes of the dataframes to see how many rows were dropped
print(f'The original file had {NEWFILE.shape[0]} entries.')
print(f'The new file has {newfile.shape[0]} entries.')

# Save the processed file as a new .csv file
NEWFILE.to_excel("Voter_List_Without_Duplicates.xlsx")
```