

```
In [ ]: #initial importing of data from excel files to DataFrame, can uncomment to switch
```

```
In [6]: import pandas as pd
```

```
df_13_14 = pd.read_excel("PDE_DATA_CLEAN/13_14.xlsx")
df_14_15 = pd.read_excel("PDE_DATA_CLEAN/14_15.xlsx")
#df_15_16 = pd.read_excel("PDE_DATA_CLEAN/15_16.xlsx")
#df_16_17 = pd.read_excel("PDE_DATA_CLEAN/16_17.xlsx")
#df_17_18 = pd.read_excel("PDE_DATA_CLEAN/17_18.xlsx")
#df_18_19 = pd.read_excel("PDE_DATA_CLEAN/18_19.xlsx")
#df_19_20 = pd.read_excel("PDE_DATA_CLEAN/19_20.xlsx")
#df_20_21 = pd.read_excel("PDE_DATA_CLEAN/20_21.xlsx")
#df_21_22 = pd.read_excel("PDE_DATA_CLEAN/21_22.xlsx")

print("Finished creating DataFrames.")
```

Finished creating DataFrames.

```
In [ ]: #Examining the contents of the DataFrames, can uncomment to switch DataFrame
```

```
In [7]: df_13_14.head()
df_14_15.head()
#df_15_16.head()
#df_16_17.head()
#df_17_18.head()
#df_18_19.head()
#df_19_20.head()
#df_20_21.head()
#df_21_22.head()
```

Out[7]:

	SY	Staff Snapshot Date	PublicID	Last Name	First Name	MiddleName	Suffix	Gender	AnnualSa
0	2014-15	2014-10-01 00:00:00	-2144618496	Eckert	Jennifer	NaN	NaN	F	545
1	2014-15	2014-10-01 00:00:00	-2144570305	McCann	Natalie	NaN	NaN	F	594
2	2014-15	2014-10-01 00:00:00	-2144567542	Olisky	Margaret	NaN	NaN	F	900
3	2014-15	2014-10-01 00:00:00	-2144509443	WEISS	LISA	M	NaN	F	596
4	2014-15	2014-10-01 00:00:00	-2144475810	VIECZOREK	MARIE	A	NaN	F	526

5 rows x 34 columns

```
In [ ]: #Examining the columns of the DataFrame
```

```
In [14]: len(df_13_14.columns)
         #df_14_15.columns
         #df_15_16.columns
         #df_16_17.columns
         #df_17_18.columns
         #df_18_19.columns
         #df_19_20.columns
         #df_20_21.columns
         #df_21_22.columns
```

```
Out[14]: 35
```

```
In [15]: len(df_14_15.columns)
```

```
Out[15]: 34
```

```
In [ ]: #How many records are there in df_13_14?
```

```
In [19]: len(df_13_14)
```

```
Out[19]: 190900
```

```
In [ ]: #How many records are there in df_14_15?
```

```
In [20]: len(df_14_15)
```

```
Out[20]: 188465
```

```
In [26]: #unique column names in both df_13_14 and df_14_15
         set(df_13_14.columns).intersection(set(df_14_15.columns))
```

```
Out[26]: {'AUN',
          'AnnualSalary',
          'AssignCd',
          'Assignment Description',
          'CategoryDescription',
          'DegreeCode',
          'DegreeDescription',
          'EDFCd',
          'EDFactsDescription',
          'FT/PT',
          'FTE%',
          'First Name',
          'Gender',
          'IU',
          'JobClass',
          'LEACounty',
          'LEACountyCd',
          'LEAName',
          'LEATypeDescription',
          'Last Name',
          'MiddleName',
          'PositionCd',
          'PositionDescription',
          'PublicID',
          'SY',
          'SchNum',
          'School',
          'SchoolCounty',
          'SchoolCountyCode',
          'Staff Snapshot Date',
          'Suffix',
          'YearsInEd',
          'YearsInLEA'}
```

```
In [27]: #columns in df_13_14 but not in df_14_15
         set(df_13_14.columns).difference(set(df_14_15.columns))
```

```
Out[27]: {'Active/Leave/Other', 'Pos. 2012-'}
```

```
In [28]: #columns in df_14_15 but not in df_13_14
         set(df_14_15.columns).difference(set(df_13_14.columns))
```

```
Out[28]: {'Status'}
```

```
In [ ]: #Exploring if df_13_14["Active/Leave/Other"] and df_14_15["Status"] columns show
```

```
In [29]: df_13_14["Active/Leave/Other"]
```

```
Out[29]: 0      Active
          1      Active
          2      Active
          3      Active
          4      Active
          ...
          190895  Active
          190896  Active
          190897  Active
          190898  Active
          190899  Active
          Name: Active/Leave/Other, Length: 190900, dtype: object
```

```
In [30]: df_14_15["Status"]
```

```
Out[30]: 0      Active
          1      Active
          2      Active
          3      Active
          4      Active
          ...
          188460  Active
          188461  Active
          188462  Active
          188463  Active
          188464  Active
          Name: Status, Length: 188465, dtype: object
```

```
In [ ]:
```

```
In [ ]:
```