In [2]:
```
#initial importing of data from excel files to DataFrame, can uncomment to switch Data
```

In [77]:
```
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 [78]:
```
#Examining the contents of the DataFrames, can uncomment to switch DataFrame
```

In [79]:
```
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[79]:

	SY	Staff Snapshot Date	PublicID	Last Name	First Name	MiddleName	Suffix	Gender	AnnualSalary	DegreeCode	...	FT/PT	JobCl
0	2014-15	2014-10-01 00:00:00	-2144618496	Eckert	Jennifer	NaN	NaN	F	54590.0	1054	...	full time	
1	2014-15	2014-10-01 00:00:00	-2144570305	McCann	Natalie	NaN	NaN	F	59455.0	1054	...	full time	N
2	2014-15	2014-10-01 00:00:00	-2144567542	Olisky	Margaret	NaN	NaN	F	90000.0	1054	...	full time	N
3	2014-15	2014-10-01 00:00:00	-2144509443	WEISS	LISA	M	NaN	F	59653.0	1054	...	full time	
4	2014-15	2014-10-01 00:00:00	-2144475810	VIECZOREK	MARIE	A	NaN	F	52654.0	1054	...	full time	

5 rows × 34 columns

In [80]:
```
#Examining the columns of the DataFrame
```

In [178]:
```
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[178]: 34

In [82]:
```
len(df_14_15.columns)
```

Out[82]: 34

In [83]:
```
#How many records are there in df_13_14?
```

In [84]:
```
len(df_13_14)
```

Out[84]: 190900

In [85]:
```
#How many records are there in df_14_15?
```

In [86]:
```
len(df_14_15)
```

Out[86]: 188465

In [179]:
```
#unique column names in both df_14_15 and df_21_22
set(df_14_15.columns).intersection(set(df_21_22.columns))
```

Out[179]:
```
{'AUN',
 'AnnualSalary',
 'AssignCd',
 'Assignment Description',
 'DegreeCode',
 'DegreeDescription',
 'FT/PT',
 'FTE%',
 'First Name',
 'Gender',
 'IU',
 'JobClass',
 'LEACounty',
 'LEACountyCd',
 'LEAName',
 'LEATypeDescription',
 'Last Name',
 'PublicID',
 'SY',
 'SchNum',
 'School',
 'Staff Snapshot Date',
 'Status',
 'Suffix',
 'YearsInLEA'}
```

In [180]:
```
#columns in df_14_15 but not in df_21_22
set(df_14_15.columns).difference(set(df_21_22.columns))
```

Out[180]:
```
{'CategoryDescription',
 'EDFCd',
 'EDFactsDescription',
 'MiddleName',
 'PositionCd',
 'PositionDescription',
 'SchoolCounty',
 'SchoolCountyCode',
 'YearsInEd'}
```

In [181]:
```
#columns in df_21_22 but not in df_14_15
set(df_21_22.columns).difference(set(df_14_15.columns))
```

Out[181]:
```
{'Category Description 2017-',
 'EDF Category',
 'EDFNum',
 'IUName',
 'Middle Name',
 'Pos 2017-',
 'Position Description 2017-',
 'Primary Assignment',
 'YearsInED'}
```

In [90]:
```
#Exploring if df_13_14["Active/Leave/Other"] and df_14_15["Status"] columns show the same data
```

In [91]:
```
df_13_14["Active/Leave/Other"]
```

Out[91]:
```
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 [92]:
```
df_14_15["Status"]
```

Out[92]:
```
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 [93]:
```
#Listing all Assignment Descriptions in df_13_14
```

In [94]:
```
df_13_14["Assignment Description"]
```

Out[94]:
```
0                                  Art, Secondary (7-12)
1          Developmental Reading, Secondary Classes, 7-12
2                        Middle Level Social Studies, 7-9
3                                Kindergarten, age 5 (K5)
4                            English/Communication, 10-12
                                   ...
190895                            Pre Kindergarten (PreK)
190896                           Art, Elementary (PreK-6)
190897                            Environmental Education
190898            Special Ed, Elementary Subjects, PreK-6
190899                  Elementary, Primary Grades 1-3 1/
Name: Assignment Description, Length: 190900, dtype: object
```

In [96]:
```
#Importing matplotlib

import matplotlib.pyplot as plt

#Grouping teachers by "Assignment Description" and saving the top 5 in df_top5

assignments = df_14_15.groupby("Assignment Description")['PublicID'].nunique()
df_top5 = assignments.sort_values(ascending=False)[:5]
```

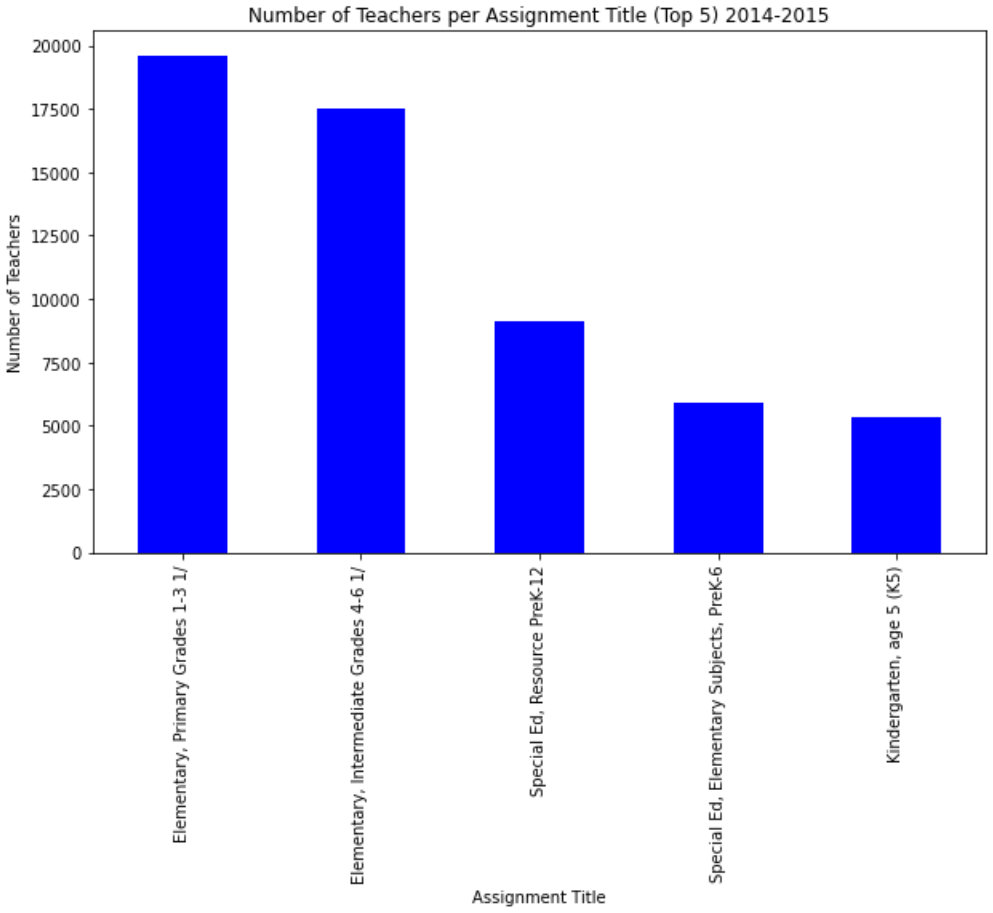
In [99]:
```
#Print test of df_top5
```

In [100]:
```
print(df_top5)


#Creating a Bar Chart to plot the top 5 Teacher Assignment Titles and their frequency
```

```
Assignment Description
Elementary, Primary Grades 1-3 1/          19603
Elementary, Intermediate Grades 4-6 1/     17521
Special Ed, Resource PreK-12                9125
Special Ed, Elementary Subjects, PreK-6     5912
Kindergarten, age 5 (K5)                    5330
Name: PublicID, dtype: int64
```

In [126]:
```
top5_chart_14_15 = df_top5.plot(kind='bar', figsize=(10, 6), color = 'blue')
top5_chart_14_15.set_title('Number of Teachers per Assignment Title (Top 5) 2014-2015')
top5_chart_14_15.set_xlabel('Assignment Title')
top5_chart_14_15.set_ylabel('Number of Teachers')
#plt.show()
```

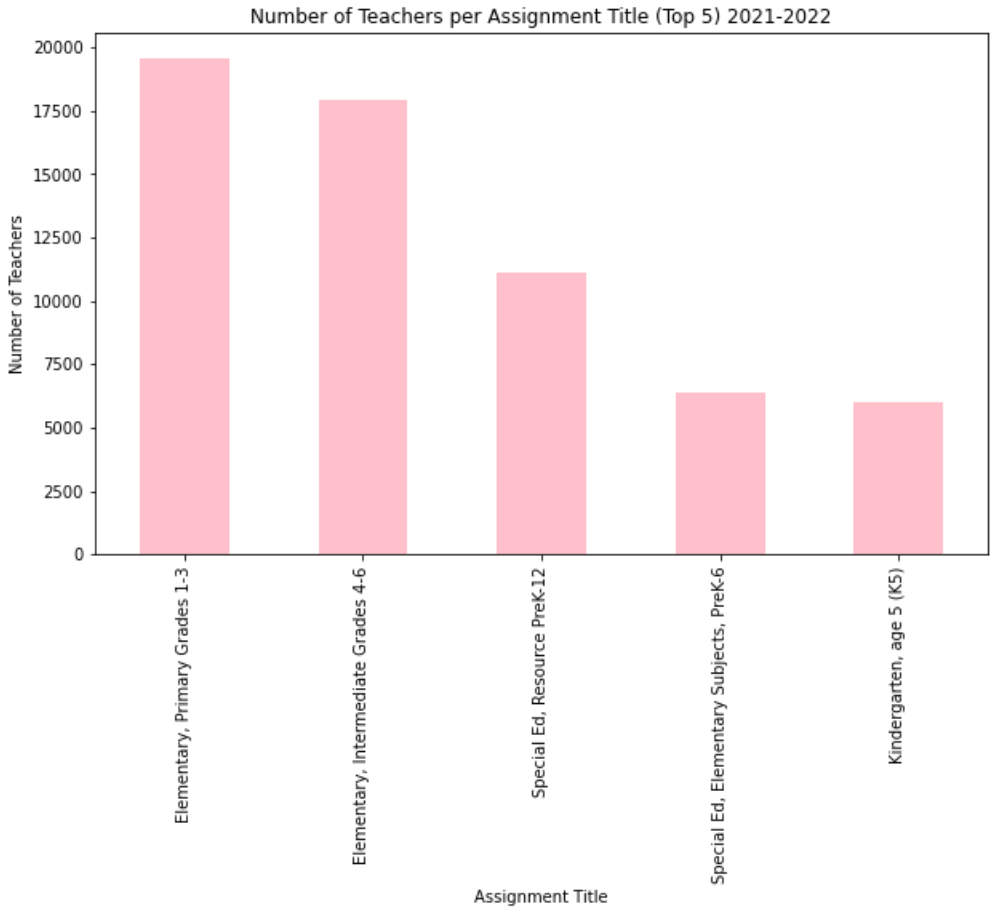
Out[126]:
```
Text(0, 0.5, 'Number of Teachers')
```



In [127]:
```
#Creating a Bar Chart to plot the top 5 teacher Assignment Titles and their frequency in 2021-202
```

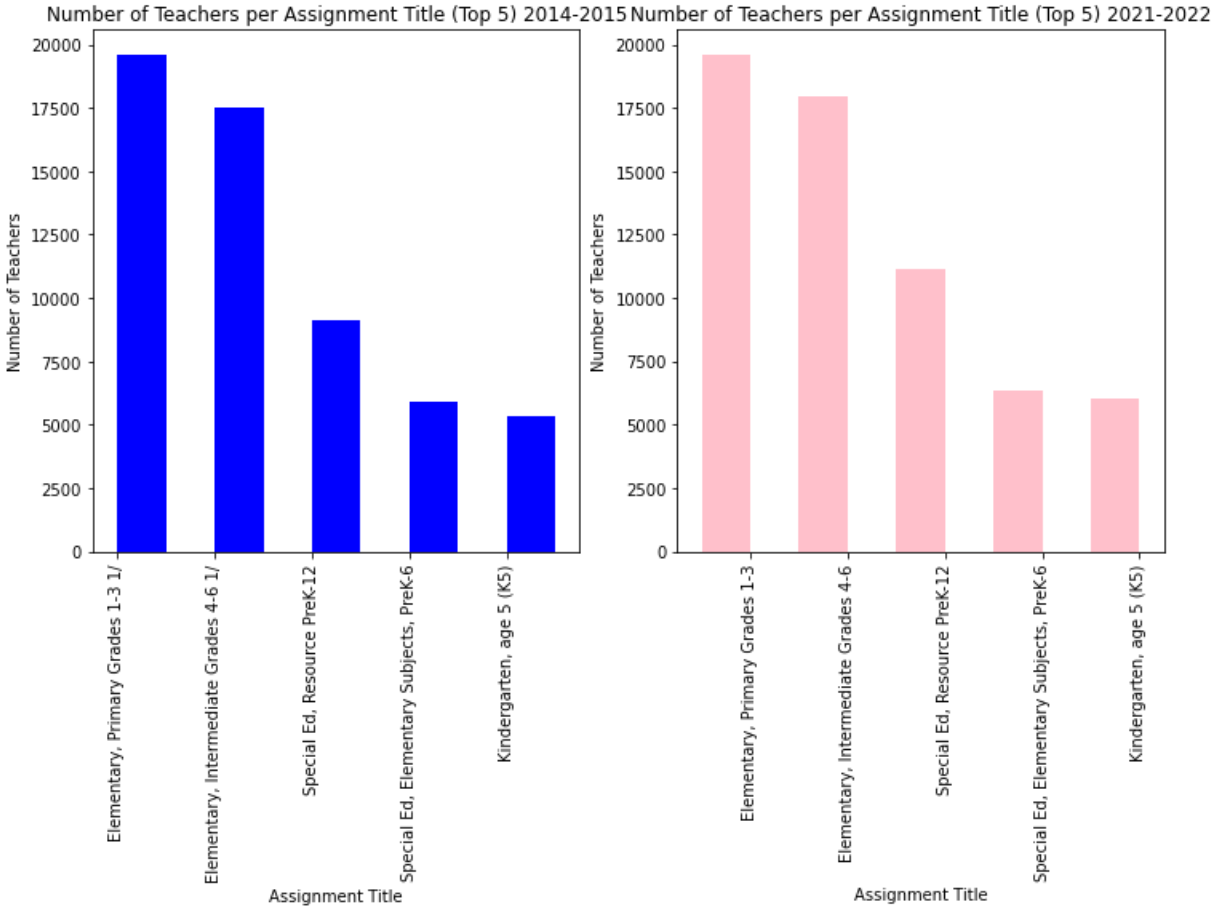
In [136]:
```
assignments_21_22 = df_21_22.groupby("Assignment Description")['PublicID'].nunique()
df_top5_21_22 = assignments_21_22.sort_values(ascending=False)[:5]
top5_chart_21_22 = df_top5_21_22.plot(kind='bar', figsize=(10, 6), color = 'pink')
top5_chart_21_22.set_title('Number of Teachers per Assignment Title (Top 5) 2021-2022')
top5_chart_21_22.set_xlabel('Assignment Title')
top5_chart_21_22.set_ylabel('Number of Teachers')
#plt.show()
```

Out[136]:
```
Text(0, 0.5, 'Number of Teachers')
```



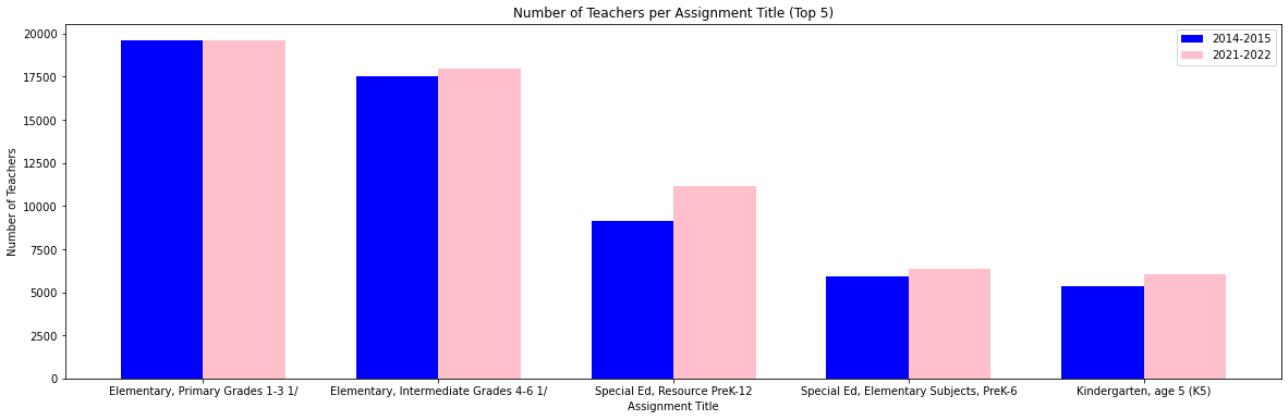
In [ ]:

In [137]:
```
# get top 5 assignments from 2014-2015 data
assignments = df_14_15.groupby("Assignment Description")['PublicID'].nunique()
df_top5 = assignments.sort_values(ascending=False)[:5]

# get top 5 assignments from 2021-2022 data
assignments_21_22 = df_21_22.groupby("Assignment Description")['PublicID'].nunique()
df_top5_21_22 = assignments_21_22.sort_values(ascending=False)[:5]

# create two axes
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 6))

# plot top 5 assignments from 2014-2015 on first axis
df_top5.plot.bar(ax=ax1, position=0, color='blue')
ax1.set_title('Number of Teachers per Assignment Title (Top 5) 2014-2015')
ax1.set_xlabel('Assignment Title')
ax1.set_ylabel('Number of Teachers')

# plot top 5 assignments from 2021-2022 on second axis
df_top5_21_22.plot.bar(ax=ax2, position=1, color='pink')
ax2.set_title('Number of Teachers per Assignment Title (Top 5) 2021-2022')
ax2.set_xlabel('Assignment Title')
ax2.set_ylabel('Number of Teachers')

plt.show()
```



In [142]:

```
# Data for chart 1
assignments_14_15 = df_14_15.groupby("Assignment Description")['PublicID'].nunique()
df_top5_14_15 = assignments_14_15.sort_values(ascending=False)[:5]
values1 = df_top5_14_15.values
labels1 = df_top5_14_15.index

# Data for chart 2
assignments_21_22 = df_21_22.groupby("Assignment Description")['PublicID'].nunique()
df_top5_21_22 = assignments_21_22.sort_values(ascending=False)[:5]
values2 = df_top5_21_22.values
labels2 = df_top5_21_22.index

# Set up the plot
fig, ax = plt.subplots(figsize=(20, 6))

# Set up the x-axis
x = np.arange(len(labels1))
width = 0.35  # the width of the bars

# Plot the bars for chart 1
rects1 = ax.bar(x - width/2, values1, width, label='2014-2015', color='blue')

# Plot the bars for chart 2
rects2 = ax.bar(x + width/2, values2, width, label='2021-2022', color='pink')

# Set up the labels and title
ax.set_ylabel('Number of Teachers')
ax.set_xlabel('Assignment Title')
ax.set_title('Number of Teachers per Assignment Title (Top 5)')
ax.set_xticks(x)
ax.set_xticklabels(labels1)
ax.legend()


plt.show()
```



In [170]:
```
"""
Printing all of the unique Assignment Descriptions from 2014 - 2015 data
"""
assignments_14_15 = df_14_15.groupby("Assignment Description")['PublicID'].nunique()
print(assignments_14_15)
for x in assignments_14_15.keys():
    print(x)
```

```
Alternate Education, Middle Level English, 7-9
Alternate Education, Middle Level Math, 7-9
Alternate Education, Middle Level Science, 7-9
Alternate Education, Middle Level Social Studies, 7-9
Alternate Education, Secondary English, 10-12
Alternate Education, Secondary Math, 10-12
Alternate Education, Secondary Science, 10-12
Alternate Education, Secondary Social Studies, 10-12
Alternative Education, K-12 Resource Room
Anthropology
Appliance Repair
Arabic
Architectural-Design Technology
Art, Elementary (PreK-6)
Art, Secondary (7-12)
Assistant Superintendent
Assistant Vocational Director
Assistant or Vice Elementary Principal
Assistant or Vice Middle School Principal
Assistant or Vice Secondary Principal
```

In [176]:
```
"""
Printing all of the unique Assignment Descriptions from 2021 - 2022 data
"""

assignments_21_22 = df_21_22.groupby("Assignment Description")['PublicID'].nunique()
print(assignments_21_22)
for x in assignments_21_22.keys():
    print(x)
```

```
Aeronautical Technology
Agricultural Mechanics
Agricultural Power and Machinery
Agriculture
Air Conditioning
Air Conditioning/Refrigeration
Allied Health Science Technology
Alternate Education Program, K-6
Alternate Education, Middle Level English, 7-9
Alternate Education, Middle Level Math, 7-9
Alternate Education, Middle Level Science, 7-9
Alternate Education, Middle Level Social Studies, 7-9
Alternate Education, Secondary English, 10-12
Alternate Education, Secondary Math, 10-12
Alternate Education, Secondary Science, 10-12
Alternate Education, Secondary Social Studies, 10-12
Alternative Education, K-12 Resource Room
Anthropology
Arabic
Arboretum, Aviary, Greenhouse
```

In [171]:
```
"""
Finding the number of teachers assigned as 'Music, Elementary, PreK-6' in 2014-2015.
"""
music_13_14 = df_13_14[df_13_14["Assignment Description"].str.contains('Music, Elementary, PreK-6
print(len(music_13_14))
```

```
5787
```

In [174]:
```
"""
Finding the number of teachers assigned as 'Music, Secondary, 7-12' in 2014-2015.
"""
music_13_14 = df_13_14[df_13_14["Assignment Description"].str.contains('Music, Secondary, 7-12')]
print(len(music_13_14))
```

```
3034
```

In [177]:
```
music_21_22 = df_21_22[df_21_22["Assignment Description"].str.contains('Music, Elementary, PreK-6
print(len(music))
```

```
8830
```

In [175]:
```
music_21_22 = df_21_22[df_21_22["Assignment Description"].str.contains('Music, Secondary, 7-12')]
print(len(music))
```

```
8830
```

In [192]:
```
"""
Finding the number of teachers assigned as 'Art, Secondary' in 2014-2015.
"""
art_13_14_sec = df_13_14[df_13_14["Assignment Description"].str.contains('Art, Secondary')]
print(len(art_13_14_sec))
```

```
2235
```

In [193]:
```
"""
Finding the number of teachers assigned as 'Art, Elementary' in 2014-2015.
"""
art_13_14_elem = df_13_14[df_13_14["Assignment Description"].str.contains('Art, Elementary')]
print(len(art_13_14_elem))
```

```
3368
```

In [194]:
```
"""
Finding the number of teachers assigned as 'Art, Secondary' in 2021-2022.
"""
art_21_22_sec = df_21_22[df_21_22["Assignment Description"].str.contains('Art, Secondary')]
print(len(art_21_22_sec))
```

```
2197
```

In [195]:
```
"""
Finding the number of teachers assigned as 'Art, Elementary' in 2021-2022.
"""
art_21_22_elem = df_21_22[df_21_22["Assignment Description"].str.contains('Art, Elementary')]
print(len(art_21_22_elem))
```

```
2125
```

In [ ]:
```
```