NY_Public_School

December 21, 2022

1 Investigating Student Demographic and Enrollment Status In NYC Public Schools

1.1 Introduction

Dataset:
I am using the dataset from NYC Open Data called "the snapshot of NYC Public School student enrollment and demographic information year 2020-2021 sorted by Borough"(https://data.cityofnewyork.us/Education/2020-2021-Demographic-Snapshot-Borough/vquv-pjuh) The dataset contains data in time periods of 2016-17, 2017-18,2018-19,2019-20 and 2020-21, instead of just 2020-21. Enrollment counts are based on the October 31 Audited Register for each school year, so the year variable does not mean starting from Janurary and end in December.

Background According to New York City Department of Education, the school's Economic Need Index(ENI) estimates the percentage of students facing economic hardship.ENI is a metric calculated used by the Department of Education to investigate and incorporate economic diversity in NYC's schools.

Basic Questions: 1.Which borough's public schools has the highest students' average Economic Need Index(ENI) in 2020-2021? 2.In that borough, what's the student demographic profile (i.e. ethnicity, gender) look like in 2020-2021?

Extra Questions: I also explored other related questions when exploring the data.

Hypothesis:
I hypothesis that Bronx has the highest students' average ENI.In that borough, black, Hispanic, and female students take up a bigger proportion.

2 Get ready for the data exploration

2.0.1 Import some useful packages - pandas, plotly, and plotly.io, which will help with the analysis and visualization.

[1]:
```
import pandas as pd
import plotly.express as px
import plotly.io as pio
import matplotlib.pyplot as plt
import numpy as np
pio.renderers.default = "notebook_connected+pdf"
```

3 Data Overview & Cleanup

3.0.1 Let's read dataframe into the notebook and get an overview.

[2]:
```
df= pd.read_csv('https://data.cityofnewyork.us/resource/vquv-pjuh.csv')
df.head(10)
```

[2]:

	borough	year	total_enrollment	grade_3k_pk_half_day_full	grade_k	grade_1	grade_2	grade_3	grade_4	grade_5	…	white_1	missing_race_ethnicity_data	missing_race_ethnicity_data_1
0	Bronx	2016-17	241776	14825	17974	19579	19611	18943	18505	17824	…	0.043	725	0.003
1	Bronx	2017-18	239955	14948	17812	18598	19008	18940	18535	18173	…	0.044	436	0.002
2	Bronx	2018-19	236267	15105	17259	18071	17962	18318	18426	18026	…	0.043	461	0.002
3	Bronx	2019-20	235448	17990	17103	17680	17598	17462	17981	18216	…	0.042	1367	0.006
4	Bronx	2020-21	227224	14769	15468	16863	17031	17016	16906	17579	…	0.042	937	0.004
5	Brooklyn	2016-17	344408	24126	25680	26405	26928	26474	25928	24746	…	0.173	1379	0.004
6	Brooklyn	2017-18	342622	23277	25578	25928	25812	26253	25851	25690	…	0.177	1316	0.004
7	Brooklyn	2018-19	339985	24104	24074	25562	25159	25060	25561	25307	…	0.180		
8	Brooklyn	2019-20	342332	28420	24224	24232	24979	24615	24533	25172	…	0.181		
9	Brooklyn	2020-21	330905	24119	22249	23447	23202	23974	23882	23747	…	0.179		

```
7                          1350                           0.004
8                          2595                           0.008
9                          2369                           0.007

   students_with_disabilities  students_with_disabilities_1  \
0                       55462                         0.229
1                       55719                         0.232
2                       55847                         0.236
3                       54899                         0.233
4                       53144                         0.234
5                       66309                         0.193
6                       67444                         0.197
7                       68086                         0.200
8                       67756                         0.198
9                       65784                         0.199

   english_language_learners  english_language_learners_1  poverty  poverty_1  \
0                      41129                        0.170   197520      0.817
1                      41316                        0.172   208334      0.868
2                      40055                        0.170   203040      0.859
3                      37509                        0.159   202976      0.862
4                      37190                        0.164   194432      0.856
5                      44335                        0.129   244956      0.711
6                      44390                        0.130   261947      0.765
7                      42678                        0.126   255263      0.751
8                      41294                        0.121   254914      0.745
9                      42661                        0.129   243307      0.735

   economic_need_index
0                0.774
1                0.857
2                0.854
3                0.868
4                0.858
5                0.632
6                0.737
7                0.728
8                0.743
9                0.725

[10 rows x 42 columns]
```

[3]:
```
df.info
```

[3]:
```
<bound method DataFrame.info of           borough     year  total_enrollment
grade_3k_pk_half_day_full  \
0           Bronx  2016-17            241776                       14825
```

1	Bronx	2017-18	239955	14948
2	Bronx	2018-19	236267	15105
3	Bronx	2019-20	235448	17990
4	Bronx	2020-21	227224	14769
5	Brooklyn	2016-17	344408	24126
6	Brooklyn	2017-18	342622	23277
7	Brooklyn	2018-19	339985	24104
8	Brooklyn	2019-20	342332	28420
9	Brooklyn	2020-21	330905	24119
10	Manhattan	2016-17	178383	8478
11	Manhattan	2017-18	177752	8365
12	Manhattan	2018-19	177512	8922
13	Manhattan	2019-20	180636	11201
14	Manhattan	2020-21	173851	9870
15	Queens	2016-17	310741	20983
16	Queens	2017-18	309302	20946
17	Queens	2018-19	307114	21609
18	Queens	2019-20	305623	23558
19	Queens	2020-21	294923	21013
20	Staten Island	2016-17	65924	4141
21	Staten Island	2017-18	65703	3992
22	Staten Island	2018-19	65623	4164
23	Staten Island	2019-20	67829	6311
24	Staten Island	2020-21	67235	6210

	grade_k	grade_1	grade_2	grade_3	grade_4	grade_5	…	white_1	\
0	17974	19579	19611	18943	18505	17824	…	0.043	
1	17812	18598	19008	18940	18535	18173	…	0.044	
2	17259	18071	17962	18318	18426	18026	…	0.043	
3	17103	17680	17598	17462	17981	18216	…	0.042	
4	15468	16863	17031	17016	16906	17579	…	0.042	
5	25680	26405	26928	26474	25928	24746	…	0.173	
6	25578	25928	25812	26253	25851	25690	…	0.177	
7	24074	25562	25159	25060	25561	25307	…	0.180	
8	24224	24232	24979	24615	24533	25172	…	0.181	
9	22249	23447	23202	23974	23882	23747	…	0.179	
10	11480	11795	11721	11446	11179	10745	…	0.161	
11	11237	11538	11454	11533	11173	11146	…	0.166	
12	10988	11241	11123	11168	11228	10986	…	0.169	
13	11015	10884	11024	10964	10931	11021	…	0.170	
14	10037	10249	10229	10413	10440	10395	…	0.163	
15	22763	23431	23233	23310	23686	22476	…	0.130	
16	22434	22767	22911	22854	22933	23384	…	0.129	
17	21936	22376	22254	22381	22540	22846	…	0.127	
18	21717	21874	22006	21774	22039	22332	…	0.125	
19	20166	20946	20961	21154	21104	21463	…	0.122	
20	4620	4798	4786	4877	4887	4772	…	0.468	

21	4527	4747	4793	4783	4875	4884	…	0.457
22	4370	4658	4707	4779	4755	4879	…	0.442
23	4528	4501	4691	4740	4803	4772	…	0.432
24	4345	4532	4481	4680	4702	4795	…	0.421

	missing_race_ethnicity_data	missing_race_ethnicity_data_1	\
0	725	0.003	
1	436	0.002	
2	461	0.002	
3	1367	0.006	
4	937	0.004	
5	1379	0.004	
6	1316	0.004	
7	1350	0.004	
8	2595	0.008	
9	2369	0.007	
10	1328	0.007	
11	1259	0.007	
12	1246	0.007	
13	2004	0.011	
14	1907	0.011	
15	1247	0.004	
16	1210	0.004	
17	1281	0.004	
18	1811	0.006	
19	1662	0.006	
20	220	0.003	
21	209	0.003	
22	323	0.005	
23	207	0.003	
24	325	0.005	

	students_with_disabilities	students_with_disabilities_1	\
0	55462	0.229	
1	55719	0.232	
2	55847	0.236	
3	54899	0.233	
4	53144	0.234	
5	66309	0.193	
6	67444	0.197	
7	68086	0.200	
8	67756	0.198	
9	65784	0.199	
10	36271	0.203	
11	37135	0.209	
12	38063	0.214	
13	38878	0.215	

```
14                        37644                          0.217
15                        53236                          0.171
16                        54174                          0.175
17                        54805                          0.178
18                        54520                          0.178
19                        52743                          0.179
20                        17069                          0.259
21                        17438                          0.265
22                        17696                          0.270
23                        18079                          0.267
24                        17790                          0.265

    english_language_learners  english_language_learners_1  poverty  \
0                       41129                         0.170   197520
1                       41316                         0.172   208334
2                       40055                         0.170   203040
3                       37509                         0.159   202976
4                       37190                         0.164   194432
5                       44335                         0.129   244956
6                       44390                         0.130   261947
7                       42678                         0.126   255263
8                       41294                         0.121   254914
9                       42661                         0.129   243307
10                      19358                         0.109   114277
11                      19089                         0.107   120957
12                      18102                         0.102   118397
13                      17095                         0.095   119314
14                      17041                         0.098   114781
15                      44828                         0.144   207432
16                      44773                         0.145   226458
17                      43773                         0.143   218554
18                      42266                         0.138   217708
19                      44123                         0.150   206978
20                       3816                         0.058    36800
21                       4093                         0.062    38787
22                       4207                         0.064    38749
23                       4256                         0.063    39745
24                       4708                         0.070    39033

    poverty_1  economic_need_index
0       0.817                0.774
1       0.868                0.857
2       0.859                0.854
3       0.862                0.868
4       0.856                0.858
5       0.711                0.632
6       0.765                0.737
```

```
7       0.751                0.728
8       0.745                0.743
9       0.735                0.725
10      0.641                0.598
11      0.680                0.673
12      0.667                0.665
13      0.661                0.679
14      0.660                0.671
15      0.668                0.502
16      0.732                0.658
17      0.712                0.652
18      0.712                0.672
19      0.702                0.655
20      0.558                0.436
21      0.590                0.533
22      0.590                0.537
23      0.586                0.557
24      0.581                0.545

[25 rows x 42 columns]>
```

4 Part II: Question I Exploration

4.1 Which borough's public schools has the higest students' average Ecnomic Need Index(ENI)?

Pivot the dataframe, then we see the value of each cell represents the average ENI of corresponding borough and year.

[4]:
```
df_ecndex= df.pivot(index="borough", columns=["year"],␣
 ↪values="economic_need_index")
df_ecndex
```

[4]:
```
year           2016-17  2017-18  2018-19  2019-20  2020-21
borough
Bronx            0.774    0.857    0.854    0.868    0.858
Brooklyn         0.632    0.737    0.728    0.743    0.725
Manhattan        0.598    0.673    0.665    0.679    0.671
Queens           0.502    0.658    0.652    0.672    0.655
Staten Island    0.436    0.533    0.537    0.557    0.545
```

[5]:
```
df_ecndex_20to21=df_ecndex['2020-21']
df_ecndex_20to21
```

[5]:
```
borough
Bronx            0.858
Brooklyn         0.725
```

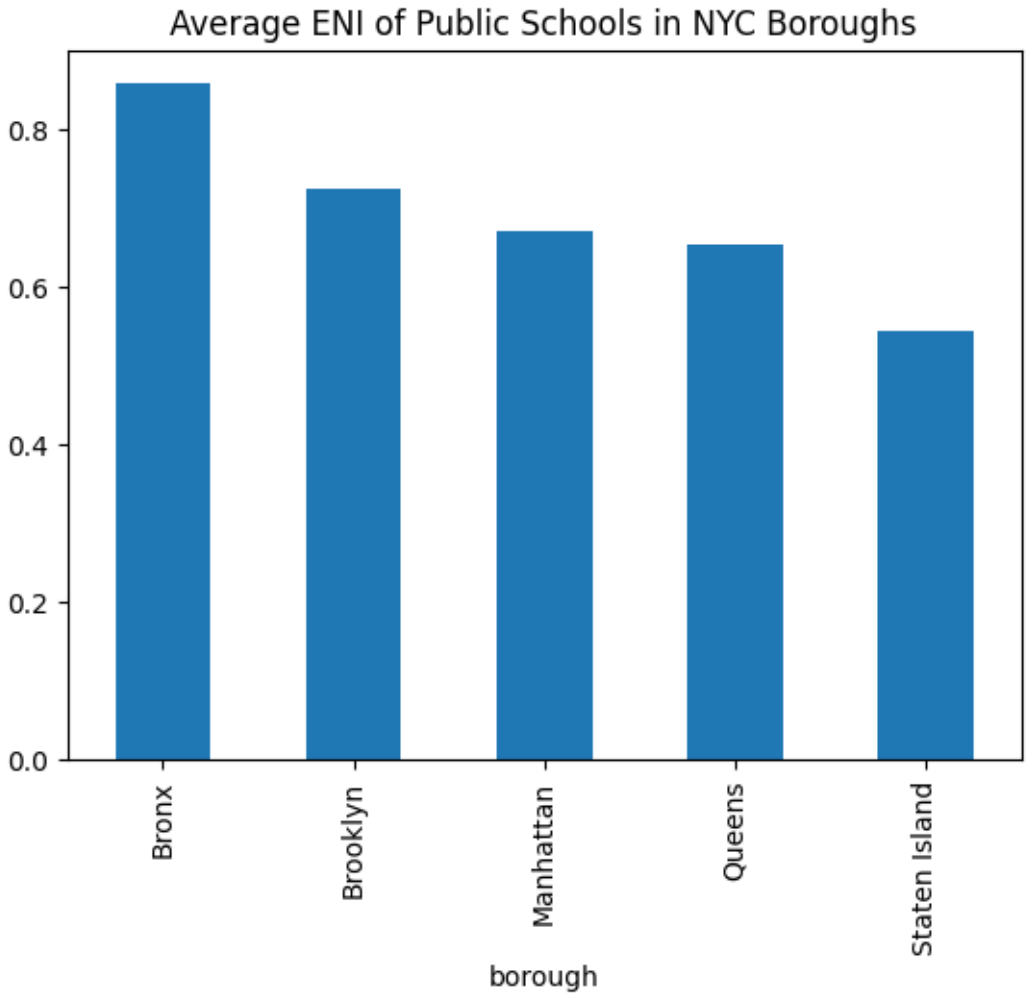
```
Manhattan        0.671
Queens           0.655
Staten Island    0.545
Name: 2020-21, dtype: float64
```

4.1.1 Zoom in - we find that in 2020-2021, public schools in Bronx has the highest average EDI, confirming the hypothesis to question 1.

The bar chart gives us the same answer.

[6]:
```
df_ecndex_20to21.plot(kind="bar",x="Borough", title="Average ENI of Public␣
↪Schools in NYC Boroughs")
```

[6]:
```
<AxesSubplot: title={'center': 'Average ENI of Public Schools in NYC Boroughs'},
xlabel='borough'>
```



4.1.2 Another prettier bar chart is available even without pivotting the dataframe.

[7]:
```
df_20to21 = df[df["year"] == "2020-21"]
df_19to20 = df[df["year"] == " 2019-20"]
df_18to19 = df[df["year"] == " 2018-19"]
df_17to18 = df[df["year"] == "2017-18"]
df_16to17 = df[df["year"] == " 2016-17"]
df_20to21
```

[7]:

	borough	year	total_enrollment	grade_3k_pk_half_day_full	\
4	Bronx	2020-21	227224	14769	
9	Brooklyn	2020-21	330905	24119	
14	Manhattan	2020-21	173851	9870	
19	Queens	2020-21	294923	21013	
24	Staten Island	2020-21	67235	6210	

	grade_k	grade_1	grade_2	grade_3	grade_4	grade_5	…	white_1	\
4	15468	16863	17031	17016	16906	17579	…	0.042	
9	22249	23447	23202	23974	23882	23747	…	0.179	
14	10037	10249	10229	10413	10440	10395	…	0.163	
19	20166	20946	20961	21154	21104	21463	…	0.122	
24	4345	4532	4481	4680	4702	4795	…	0.421	

	missing_race_ethnicity_data	missing_race_ethnicity_data_1	\
4	937	0.004	
9	2369	0.007	
14	1907	0.011	
19	1662	0.006	
24	325	0.005	

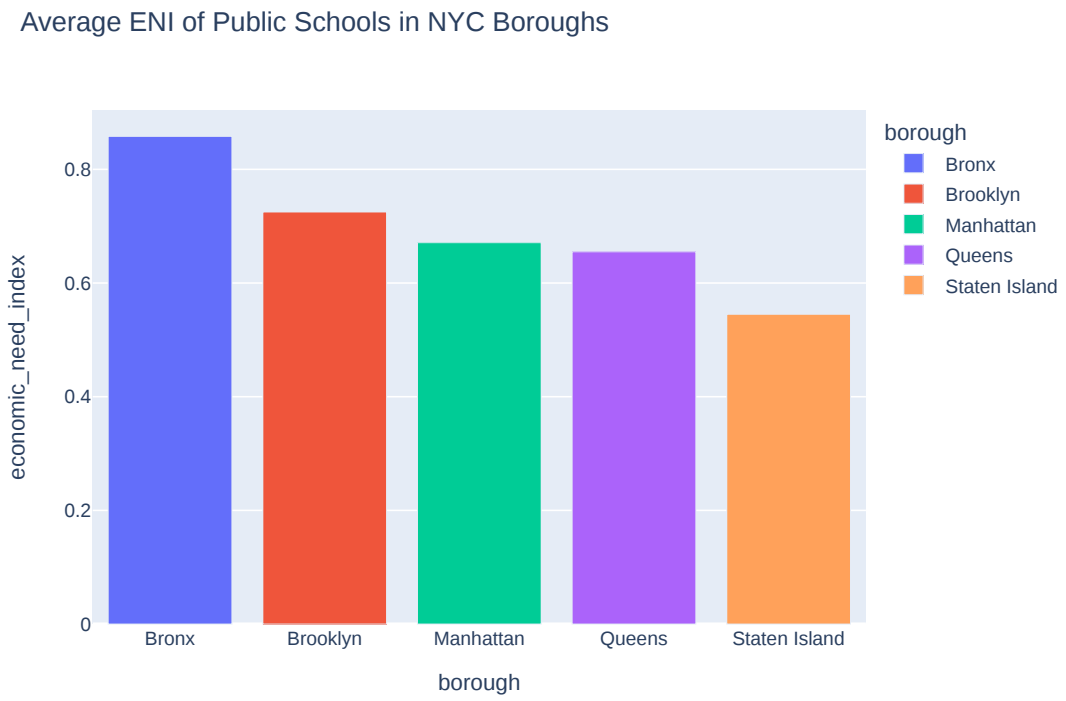
	students_with_disabilities	students_with_disabilities_1	\
4	53144	0.234	
9	65784	0.199	
14	37644	0.217	
19	52743	0.179	
24	17790	0.265	

	english_language_learners	english_language_learners_1	poverty	\
4	37190	0.164	194432	
9	42661	0.129	243307	
14	17041	0.098	114781	
19	44123	0.150	206978	
24	4708	0.070	39033	

	poverty_1	economic_need_index
4	0.856	0.858
9	0.735	0.725

```
14      0.660                0.671
19      0.702                0.655
24      0.581                0.545

[5 rows x 42 columns]
```

[8]:
```
fig=px.bar(df_20to21,x="borough",color="borough",␣
 ↪y="economic_need_index",title="Average ENI of Public Schools in NYC␣
 ↪Boroughs")
fig.show()
```



4.2 Further Exploration

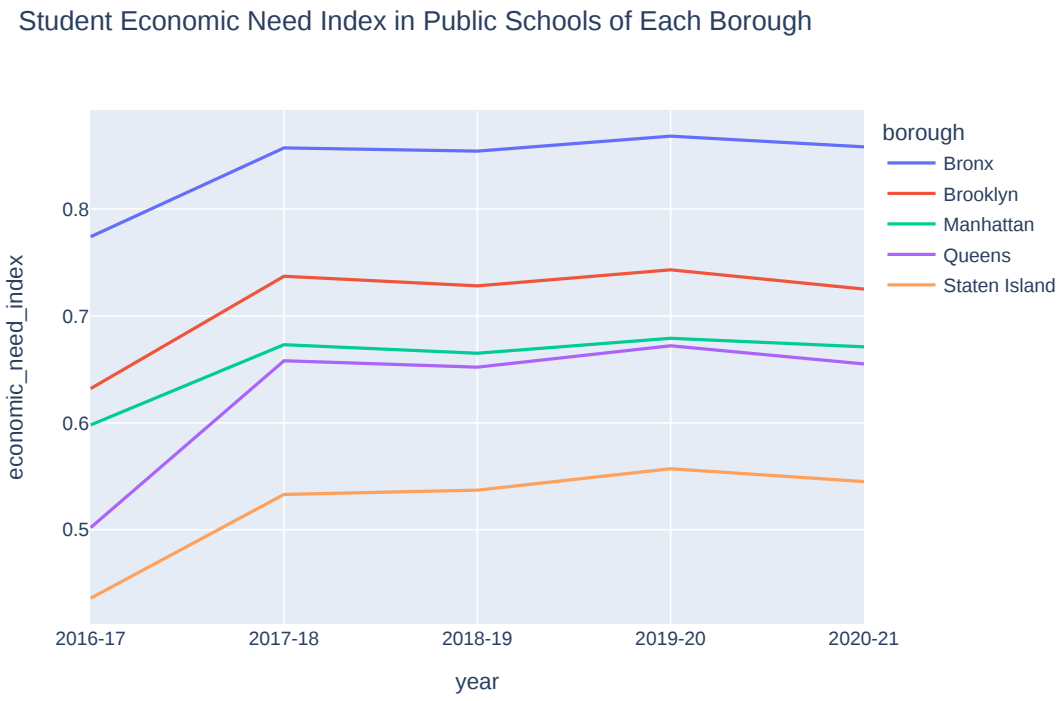
4.2.1 Q1: Which borough's public schools has the highest average ENI in years 2016-17,2017-18,2018-19,2019-20?

4.2.2 Q2: How have the average ENIs changed from year 2016-17 to year 2020-21?

4.2.3 Q1: From 2016-2021, the Bronx has the highest average ENI.

4.2.4 Q2: From year 2016-17 to 2017-18, the ENI inceased in all boroughs; From 2017-18 to 2019-20, the ENI remained static: slightly increased then slightly decreased; from 2019-20 to 2020-21, the ENI decreased.

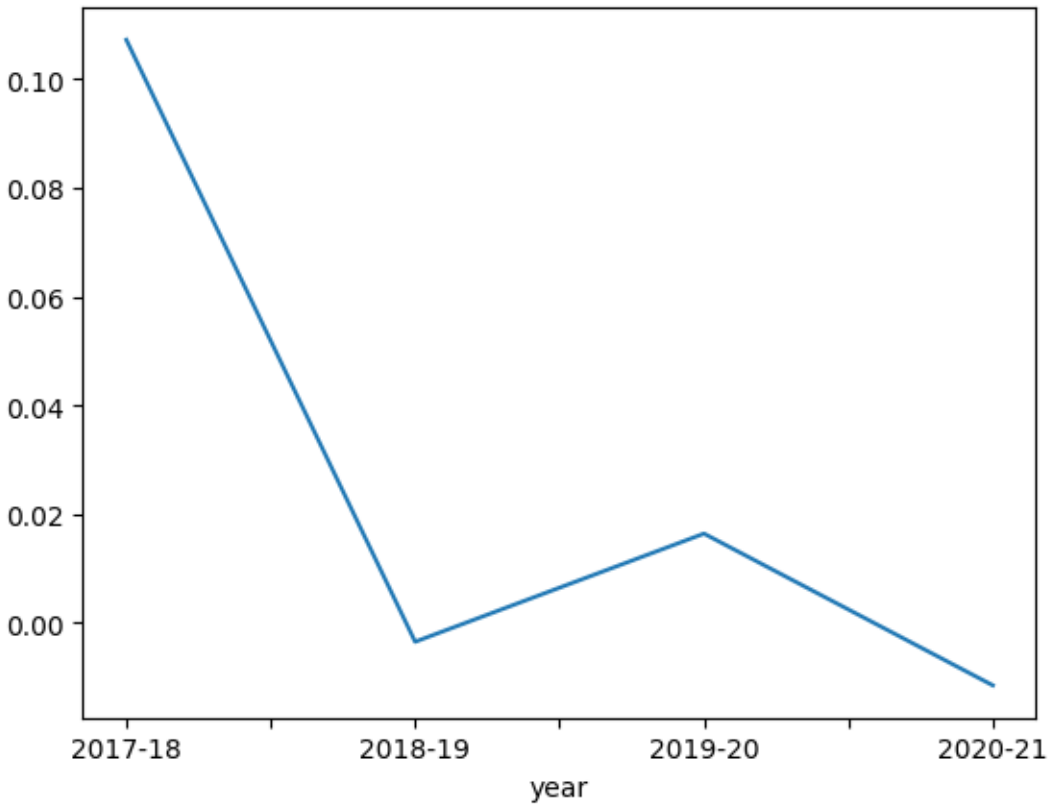
[9]:
```
fig=px.line(df,x="year", y="economic_need_index",color="borough",title="Student␣
↪Economic Need Index in Public Schools of Each Borough")
fig.show()
```



4.2.5 Zoom into the Bronx: I calculated the relative change rates in average ENI rates by dividing the difference between two years by the baseline year, we see the relative change trend. From 2017-18to 2018-19, while the EDI increased, the change rate decreased. In 2020-21, the EDI decresed, which may indicate a progress.

[10]:
```
df_ecndex.loc['Bronx',:].pct_change().plot(kind='line')
```

[10]:
```
<AxesSubplot: xlabel='year'>
```



5 Part II Q2 Exploration

5.1 In the borough with highest average EDI, what's the student demographic profile (i.e. ethnicity, gender) look like in year 2020-2021?

Filter & Drop: only keep variables that we need to answer the question.

5.1.1 Race and Ethnicity

In the data description, each cell represents the percentages that the demographic group takes up. Thus, we change them into percentage format.

[11]:
```
df_racep = df[['borough','year','asian_1',␣
 ↪'black_1','hispanic_1','native_american_1','white_1','multi_racial_1','missing_race_ethnici
df_racep.style.format({
    'asian_1':'{:.2%}'.format,
    'black_1':'{:.2%}'.format,
    'hispanic_1':'{:.2%}'.format,
    'native_american_1':'{:.2%}'.format,
    'white_1': '{:.2%}'.format,
    'multi_racial_1':'{:.2%}'.format,
    'missing_race_ethnicity_data_1':'{:.2%}'.format
})
```

[11]:
```
<pandas.io.formats.style.Styler at 0x7f4a9e655060>
```

Only keep the race and ethnicity percentages of year 2020-21.

Althernatively, we could only keep the race and ethnicity populations.

[12]:
```
df_race = df[['borough','year','asian',␣
 ↪'black','hispanic','native_american','white','multi_racial','missing_race_ethnicity_data']]
df_race_20to21=df_race[(df_race['year']=='2020-21')]
df_race_20to21
```

[12]:

	borough	year	asian	black	hispanic	native_american	white	multi_racial	missing_race_ethnicity_data
4	Bronx	2020-21	11429	60851	140947	2206	9523	1331	937
9	Brooklyn	2020-21	55957	111976	93778	2831	59204	4790	2369
14	Manhattan	2020-21	20759	39296	77613	1314	28327	4635	1907
19	Queens	2020-21	84062	49714	113584	6027	36122	3752	1662
24	Staten Island	2020-21	8405	8703	19983	328	28275	1216	325

[13]:
```
df_race_20to21Bronx=df_race_20to21.drop([9,14,19,24])
df_race_20to21Bronx
```

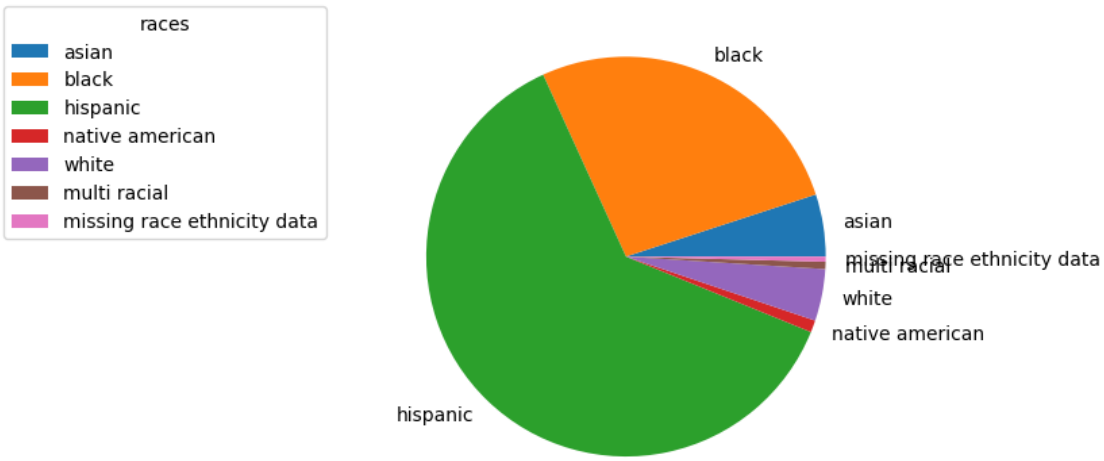
[13]:

	borough	year	asian	black	hispanic	native_american	white	multi_racial	missing_race_ethnicity_data
4	Bronx	2020-21	11429	60851	140947	2206	9523	1331	937

[14]:
```
df_race.to_csv('data.csv')
```

I created a pie chart to show the ethnicity distribution of students in the Bronx's public schools in 2020-21.

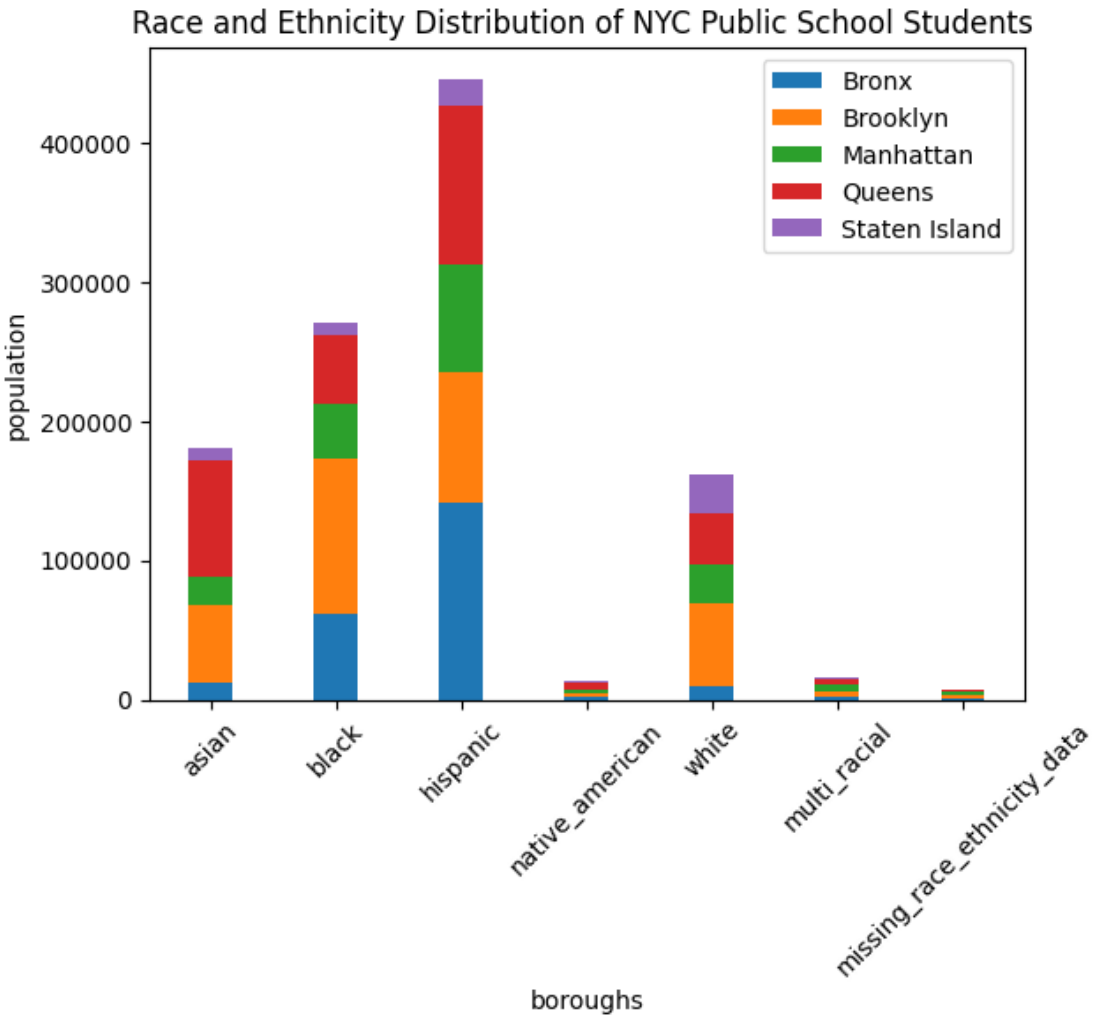
[15]:
```
import matplotlib.pyplot as plt
import numpy as np

y = np.array([11429,60851,140947,2206,9523,1331,937])
races = ["asian", "black", "hispanic", "native american","white","multi␣
 ↪racial","missing race ethnicity data"]

plt.pie(y, labels = races)
plt.legend(races, loc='best', title="races",borderaxespad=0,bbox_to_anchor=(-0.
 ↪1, 1.))
plt.show()
```



5.1.2 Hispanic and black takes up a majority proportion in public schools in Bronx. Then is Asian, White and Native American, confirming the hypothesis of question 2. There are also a small proportion of multi racial students and some missing data.

[16]:
```
df_race_20to21 = df_race_20to21.set_index('borough')
bar_width = 0.35
one = df_race_20to21.loc['Bronx','asian':'missing_race_ethnicity_data'].
 ↪to_numpy()
two = df_race_20to21.loc['Brooklyn','asian':'missing_race_ethnicity_data'].
 ↪to_numpy()
three = df_race_20to21.loc['Manhattan','asian':'missing_race_ethnicity_data'].
 ↪to_numpy()
four = df_race_20to21.loc['Queens','asian':'missing_race_ethnicity_data'].
 ↪to_numpy()
```

```
five = df_race_20to21.loc['Staten Island','asian':
 ↪'missing_race_ethnicity_data'].to_numpy()
index = np.arange(7)
name = df_race_20to21.columns[1:]
plt.bar(index, one, bar_width, label='Bronx')
plt.bar(index, two, bar_width, bottom=one, label='Brooklyn')
plt.bar(index, three, bar_width, bottom=one+two, label='Manhattan')
plt.bar(index, four, bar_width, bottom=one+two+three, label='Queens')
plt.bar(index, five, bar_width, bottom=one+two+three+four, label='Staten␣
 ↪Island')
plt.xticks(index,name,rotation=45)
plt.xlabel('boroughs')
plt.ylabel('population')
plt.title('Race and Ethnicity Distribution of NYC Public School Students ')
plt.legend()
plt.show()
```



5.1.3 Unfortunately, I didn't find sex/gender information of public school students in the Bronx and NYC. So my hypothesis that there are more female students than male students can not be confirmed or rejected in this analysis.

5.2 Conclusions

My hypothesis was conistent regarding:In 2020-2021, public schools in Bronx has the highest average EDI.Hispanic and black takes up a majority proportion in public schools in Bronx.Then is Asian and White. Such results may relate to the average household income, further analysis could reseach on their relationships.I also found that Bronx from 2016 to 2021 has the higest EDI index.Nevertheless,the relatively change rates in EDI decreased.In 2020-21, the EDI value itself decresed, which may indicate a progress. Education disparity has always been important for kids. We identify the problem, record the changes, see the progress and face the challenges.