Graduation_outcome

- id:"qk7d-gecv"
- name:"2005 - 2015 Graduation Outcomes"
- assetType:"dataset"
- attribution:"Department of Education (DOE)"
- averageRating:0
- category:"Education"
- createdAt:1457473988.

Description:
Graduation results for all students by year; cohorts of 2001 through 2011 (Classes of 2005 through 2015).
Graduation Outcomes as Calculated by the New York State Education Department.
The New York State calculation method was first adopted for the Cohort of 2001 (Class of 2005). The cohort consists of all students who first entered 9th grade in a given school year (e.g., the Cohort of 2006 entered 9th grade in the 2006-2007 school year).
Graduates are defined as those students earning either a Local or Regents diploma and exclude those earning either a special education (IEP) diploma or GED.
In order to comply with FERPA regulations on public reporting of education outcomes, rows with a cohort of 20 or fewer students are suppressed.
Due to the small number of students identified as Native American or Multi-Racial, these ethnicities are not reported on the Ethnicity tab; however, these students are included in the counts on all other tabs.
4 Year August outcomes are available for the 2004-2011 cohorts at the citywide level and for the 2005-2011 cohorts at the borough, district and school level. 5 Year August outcomes are available for the 2006-2010 cohorts. Cells are blank when data is not available.
Charter schools are not included in the NYSED calculation of graduation rates for NYC.
School level results are not presented for District 79 and District 75 schools, but their outcomes are included in citywide totals.
Starting with the 2009 cohort (Class of 2013), students with IEPs could use the compensatory score option (CSO) to meet graduation requirements, which contributed to an increase in the graduation rate.
Students are considered English Language Learners by NY State if classified as an ELL as of graduation.

In [52]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import json
import cufflinks as cf
from plotly.offline import download_plotlyjs,init_notebook_mode,plot,iplot
init_notebook_mode(connected=True)
cf.go_offline()
```

In [53]:
```
data = json.load(open("2005_-_2015_Graduation_Outcomes.json"))
```

In [54]:
```
df_meta = pd.DataFrame(data["meta"]["view"]["columns"]) # 245938028
df_meta.head(20)
```

Out[54]:

	id	name	dataTypeName	fieldName	position	renderTypeName	format	flags	tableColumnId
0	-1	sid	meta_data	:sid	0	meta_data	{}	[hidden]	NaN
1	-1	id	meta_data	:id	0	meta_data	{}	[hidden]	NaN
2	-1	position	meta_data	:position	0	meta_data	{}	[hidden]	NaN
3	-1	created_at	meta_data	:created_at	0	meta_data	{}	[hidden]	NaN
4	-1	created_meta	meta_data	:created_meta	0	meta_data	{}	[hidden]	NaN
5	-1	updated_at	meta_data	:updated_at	0	meta_data	{}	[hidden]	NaN
6	-1	updated_meta	meta_data	:updated_meta	0	meta_data	{}	[hidden]	NaN
7	-1	meta	meta_data	:meta	0	meta_data	{}	[hidden]	NaN
8	245938027	Cohort Year	number	cohort_year	1	number	{'precisionStyle': 'standard', 'noCommas': 'tr...	NaN	34866904
9	245938028	Cohort Category	text	cohort_category	2	text	{}	NaN	34866905
10	245938029	Demographic	text	demographic	3	text	{}	NaN	34866906
11	245938030	# Total Cohort	number	total_cohort	4	number	{}	NaN	34866907
12	245938031	# Total Grads	number	total_grads	5	number	{}	NaN	34866908
13	245938032	% of cohort Total Grads	number	of_cohort_total_grads_1	6	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866909
14	245938033	# of cohort Total Grads	number	of_cohort_total_grads_2	7	number	{}	NaN	34866910
15	245938034	% of cohort Total Regents	number	of_cohort_total_regents	8	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866911
16	245938035	% of grads Total Regents	number	of_grads_total_regents_1	9	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866912
17	245938036	# of grads Total Regents	number	of_grads_total_regents_2	10	number	{}	NaN	34866913
18	245938037	% of cohort Advanced Regents	number	of_cohort_advanced_regents	11	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866914
19	245938038	% of grads Advanced Regents	number	of_grads_advanced_regents_1	12	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866915

In [55]:
```
df_meta.tail(10)
```

Out[55]:

	id	name	dataTypeName	fieldName	position	renderTypeName	format	flags	tableColumnId
20	245938039	# of grads Advanced Regents	number	of_grads_advanced_regents_2	13	number	{}	NaN	34866916.0
21	245938040	% of cohort Regents w/o Advanced	number	of_cohort_regents_w_o_advanced	14	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866917.0
22	245938041	% of grads Regents w/o Advanced	number	of_grads_regents_w_o_advanced_1	15	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866918.0
23	245938042	# of grads Regents w/o Advanced	number	of_grads_regents_w_o_advanced_2	16	number	{}	NaN	34866919.0
24	245938043	% of cohort Local	number	of_cohort_local	17	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866920.0
25	245938044	% of grads Local	number	of_grads_local	18	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866921.0
26	245938045	# Still Enrolled	number	still_enrolled	19	number	{}	NaN	34866922.0
27	245938046	% of cohort Still Enrolled	number	of_cohort_still_enrolled	20	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866923.0
28	245938047	# Dropped Out	number	dropped_out	21	number	{'precisionStyle': 'standard', 'noCommas': 'fa...	NaN	34866924.0
29	245938048	% of cohort Dropped Out	number	of_cohort_dropped_out	22	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866925.0

In [56]:
```
# df["position"].value_counts(dropna=False)
```

In [57]:
```
df_meta.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 11 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   id              30 non-null     int64
 1   name            30 non-null     object
 2   dataTypeName    30 non-null     object
 3   fieldName       30 non-null     object
 4   position        30 non-null     int64
 5   renderTypeName  30 non-null     object
 6   format          30 non-null     object
 7   flags           8 non-null      object
 8   tableColumnId   22 non-null     float64
 9   width           22 non-null     float64
 10  cachedContents  22 non-null     object
dtypes: float64(2), int64(2), object(7)
memory usage: 2.7+ KB
```

data["meta"]

In [58]:
```
df_data = pd.DataFrame(data["data"])
df_data.head(10)
```

Out[58]:

	0	1	2	3	4	5	6	7	8	9	...	20	21	22	23	24	25
0	row-8eze_fvpi~9vcu	00000000-0000-0000-13BA-9C4637CE687D	0	1542139135	None	1542139135	None	{ }	2001	4 Year June	...	677	6.40	24.30	1803	17.10	64.60
1	row-t2yg.eh7e~sddq	00000000-0000-0000-CDC8-E0DED8C2FA8C	0	1542139135	None	1542139135	None	{ }	2001	5 Year June	...	862	8.20	22.00	2701	25.60	68.90
2	row-izud_8uvm_Smvi	00000000-0000-0000-8E9D-003A851D26C4	0	1542139135	None	1542139135	None	{ }	2001	6 Year	...	911	8.60	21.20	3028	28.70	70.50
3	row-6d6m.hkrp_u2g4	00000000-0000-0000-30F7-6D14D65EA77C	0	1542139135	None	1542139135	None	{ }	2002	4 Year June	...	507	6.80	30.00	1062	14.20	62.80
4	row-ugfs.egjj~p8kf	00000000-0000-0000-2659-9200A4525471	0	1542139135	None	1542139135	None	{ }	2002	5 Year June	...	618	8.30	26.30	1578	21.20	67.00
5	row-ki35.2fvi~eiju	00000000-0000-0000-ECD0-F8ADDC831614	0	1542139135	None	1542139135	None	{ }	2002	6 Year	...	671	9.00	24.60	1891	25.40	69.30
6	row-88dc_bt6p~tmdc	00000000-0000-0000-FA52-2FC059D4A4EE	0	1542139135	None	1542139135	None	{ }	2003	4 Year June	...	619	7.60	30.20	1143	14.00	55.70
7	row-wxkd~iasz_knwi	00000000-0000-0000-3C34-2F2E77EB4515	0	1542139135	None	1542139135	None	{ }	2003	5 Year June	...	801	9.80	25.80	1946	23.80	62.70
8	row-yswx~g7en.p3jm	00000000-0000-0000-81AC-0CC480CF66BB	0	1542139135	None	1542139135	None	{ }	2003	6 Year	...	853	10.60	25.10	2178	27.10	64.20
9	row-p8s2~8xaq.f8tc	00000000-0000-0000-9F23-93F2C06C8F35	0	1542139135	None	1542139135	None	{ }	2004	4 Year June	...	1017	12.00	33.50	1535	18.10	50.50

10 rows × 30 columns

In [59]:
```
df_meta.head(20)
```

Out[59]:

	id	name	dataTypeName	fieldName	position	renderTypeName	format	flags	tableColumnId
0	-1	sid	meta_data	:sid	0	meta_data	{}	[hidden]	NaN
1	-1	id	meta_data	:id	0	meta_data	{}	[hidden]	NaN
2	-1	position	meta_data	:position	0	meta_data	{}	[hidden]	NaN
3	-1	created_at	meta_data	:created_at	0	meta_data	{}	[hidden]	NaN
4	-1	created_meta	meta_data	:created_meta	0	meta_data	{}	[hidden]	NaN
5	-1	updated_at	meta_data	:updated_at	0	meta_data	{}	[hidden]	NaN
6	-1	updated_meta	meta_data	:updated_meta	0	meta_data	{}	[hidden]	NaN
7	-1	meta	meta_data	:meta	0	meta_data	{}	[hidden]	NaN
8	245938027	Cohort Year	number	cohort_year	1	number	{'precisionStyle': 'standard', 'noCommas': 'tr...	NaN	34866904
9	245938028	Cohort Category	text	cohort_category	2	text	{}	NaN	34866905
10	245938029	Demographic	text	demographic	3	text	{}	NaN	34866906
11	245938030	# Total Cohort	number	total_cohort	4	number	{}	NaN	34866907
12	245938031	# Total Grads	number	total_grads	5	number	{}	NaN	34866908
13	245938032	% of cohort Total Grads	number	of_cohort_total_grads_1	6	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866909
14	245938033	# of cohort Total Grads	number	of_cohort_total_grads_2	7	number	{}	NaN	34866910
15	245938034	% of cohort Total Regents	number	of_cohort_total_regents	8	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866911
16	245938035	% of grads Total Regents	number	of_grads_total_regents_1	9	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866912
17	245938036	# of grads Total Regents	number	of_grads_total_regents_2	10	number	{}	NaN	34866913
18	245938037	% of cohort Advanced Regents	number	of_cohort_advanced_regents	11	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866914
19	245938038	% of grads Advanced Regents	number	of_grads_advanced_regents_1	12	number	{'precisionStyle': 'percentage', 'percentScale...	NaN	34866915

In [60]:
```
df_data.columns
```

Out[60]: RangeIndex(start=0, stop=30, step=1)

In [61]:
```
df_meta.index
```

Out[61]: RangeIndex(start=0, stop=30, step=1)

the index in the df_meta are the columns in the df_data if you look in the data_meta from row 0 to 7 you will find out this is the name of the columns in the df_data

the columns name in data_meta about information you Demographic should be text and the expected postion 3 (the third columns).

stepes.

- the data types are not correct
- the columns name are int
- we have alot of columns that has duplicate values
- we have some columns that have bractes and Null values.

Task:

- removing the columns 0 to 7
- renaming the columns.

In [62]:
```
df_data.shape
```

Out[62]: (430, 30)

In [63]:
```
df_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 430 entries, 0 to 429
Data columns (total 30 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   0       430 non-null    object
 1   1       430 non-null    object
 2   2       430 non-null    int64
 3   3       430 non-null    int64
 4   4       0 non-null      object
 5   5       430 non-null    int64
 6   6       0 non-null      object
 7   7       430 non-null    object
 8   8       430 non-null    object
 9   9       430 non-null    object
 10  10      430 non-null    object
 11  11      430 non-null    object
 12  12      430 non-null    object
 13  13      430 non-null    object
 14  14      430 non-null    object
 15  15      430 non-null    object
 16  16      430 non-null    object
 17  17      420 non-null    object
 18  18      420 non-null    object
 19  19      420 non-null    object
 20  20      420 non-null    object
 21  21      420 non-null    object
 22  22      420 non-null    object
 23  23      430 non-null    object
 24  24      430 non-null    object
 25  25      430 non-null    object
 26  26      430 non-null    object
 27  27      430 non-null    object
 28  28      430 non-null    object
 29  29      430 non-null    object
dtypes: int64(3), object(27)
memory usage: 100.9+ KB
```

In [64]:
```
copy_df = df_data.copy()
```

In [65]:
```
copy_df.describe()
```

Out[65]:

	2	3	5
count	430.0	4.300000e+02	4.300000e+02
mean	0.0	1.542139e+09	1.542139e+09
std	0.0	0.000000e+00	0.000000e+00
min	0.0	1.542139e+09	1.542139e+09
25%	0.0	1.542139e+09	1.542139e+09
50%	0.0	1.542139e+09	1.542139e+09
75%	0.0	1.542139e+09	1.542139e+09
max	0.0	1.542139e+09	1.542139e+09

In [66]:
```
copy_df[3].value_counts()
```

```
1542139135    430
Name: 3, dtype: int64
```

In [67]:
```
copy_df[5].value_counts()
```

```
1542139135    430
Name: 5, dtype: int64
```

In [68]:
```
copy_df[6].value_counts()
```

```
Series([], Name: 6, dtype: int64)
```

In [69]:
```
copy_df[7].value_counts()
```

```
{ }    430
Name: 7, dtype: int64
```

In [70]:
```
copy_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 430 entries, 0 to 429
Data columns (total 30 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   0       430 non-null    object
 1   1       430 non-null    object
 2   2       430 non-null    int64
 3   3       430 non-null    int64
 4   4       0 non-null      object
 5   5       430 non-null    int64
 6   6       0 non-null      object
 7   7       430 non-null    object
 8   8       430 non-null    object
 9   9       430 non-null    object
 10  10      430 non-null    object
 11  11      430 non-null    object
 12  12      430 non-null    object
 13  13      430 non-null    object
 14  14      430 non-null    object
 15  15      430 non-null    object
 16  16      430 non-null    object
 17  17      420 non-null    object
 18  18      420 non-null    object
 19  19      420 non-null    object
 20  20      420 non-null    object
 21  21      420 non-null    object
 22  22      420 non-null    object
 23  23      430 non-null    object
 24  24      430 non-null    object
 25  25      430 non-null    object
 26  26      430 non-null    object
 27  27      430 non-null    object
 28  28      430 non-null    object
 29  29      430 non-null    object
dtypes: int64(3), object(27)
memory usage: 100.9+ KB
```

Removing the columns.

In [71]:
```
copy_df = copy_df.drop(axis=1,columns=[0,1,2,3,4,5,6,7])
```

In [72]:
```
copy_df.head()
```

Out[72]:

	8	9	10	11	12	13	14	15	16	17	...	20	21	22	23	24	25	26	27	28	29
0	2001	4 Year June	English Language Learner	10540	2791	26.50	992	9.00	35.50	315	...	677	6.40	24.30	1803	17.10	64.60	3895	37.00	3220	30.60
1	2001	5 Year June	English Language Learner	10540	3920	37.20	1233	11.70	31.50	371	...	862	8.20	22.00	2701	25.60	68.90	1516	14.40	4193	39.80
2	2001	6 Year	English Language Learner	10540	4296	40.80	1291	12.20	30.10	380	...	911	8.60	21.20	3028	28.70	70.50	655	6.20	4598	43.60
3	2002	4 Year June	English Language Learner	7454	1691	22.70	639	8.60	37.80	132	...	507	6.80	30.00	1062	14.20	62.80	3062	41.10	2330	31.30
4	2002	5 Year June	English Language Learner	7454	2354	31.60	794	10.70	33.70	176	...	618	8.30	26.30	1578	21.20	67.00	1457	19.50	3041	40.80

5 rows × 22 columns

Renaming the columns.

In [73]:
```
dict(df_meta["name"].head(30))
```

```
{0: 'sid',
 1: 'id',
 2: 'position',
 3: 'created_at',
 4: 'created_meta',
 5: 'updated_at',
 6: 'updated_meta',
 7: 'meta',
 8: 'Cohort Year',
 9: 'Cohort Category',
 10: 'Demographic',
 11: '# Total Cohort',
 12: '# Total Grads',
 13: '% of cohort Total Grads',
 14: '# of cohort Total Grads',
 15: '% of cohort Total Regents',
 16: '% of grads  Total Regents',
 17: '# of grads  Total Regents',
 18: '% of cohort  Advanced Regents',
 19: '% of grads  Advanced Regents',
 20: '# of grads  Advanced Regents',
 21: '% of cohort  Regents w/o Advanced',
 22: '% of grads  Regents w/o Advanced',
 23: '# of grads  Regents w/o Advanced',
 24: '% of cohort Local',
 25: '% of grads Local',
 26: '# Still Enrolled',
 27: '% of cohort Still Enrolled',
 28: '# Dropped Out',
 29: '% of cohort Dropped Out'}
```

In [74]:
```
column_name = {8: 'Cohort Year',
 9: 'Cohort Category',
 10: 'Demographic',
 11: '# Total Cohort',
 12: '# Total Grads',
 13: '% of cohort Total Grads',
 14: '# of cohort Total Grads',
 15: '% of cohort Total Regents',
 16: '% of grads  Total Regents',
 17: '# of grads  Total Regents',
 18: '% of cohort  Advanced Regents',
 19: '% of grads  Advanced Regents',
 20: '# of grads  Advanced Regents',
 21: '% of cohort  Regents w/o Advanced',
 22: '% of grads  Regents w/o Advanced',
 23: '# of grads  Regents w/o Advanced',
 24: '% of cohort Local',
 25: '% of grads Local',
 26: '# Still Enrolled',
 27: '% of cohort Still Enrolled',
 28: '# Dropped Out',
 29: '% of cohort Dropped Out'}
```

In [75]:
```
copy_df = copy_df.rename(columns=column_name)
copy_df.head()
```

Out[75]:

	Cohort Year	Cohort Category	Demographic	# Total Cohort	# Total Grads	% of cohort Total Grads	# of cohort Total Grads	% of cohort Total Regents	% of grads Total Regents	# of grads Total Regents	...	# of grads Advanced Regents	% of cohort Regents w/o Advanced	% of grads Regents w/o Advanced
0	2001	4 Year June	English Language Learner	10540	2791	26.50	992	9.00	35.50	315	...	677	6.40	24.30
1	2001	5 Year June	English Language Learner	10540	3920	37.20	1233	11.70	31.50	371	...	862	8.20	22.00
2	2001	6 Year	English Language Learner	10540	4296	40.80	1291	12.20	30.10	380	...	911	8.60	21.20
3	2002	4 Year June	English Language Learner	7454	1691	22.70	639	8.60	37.80	132	...	507	6.80	30.00
4	2002	5 Year June	English Language Learner	7454	2354	31.60	794	10.70	33.70	176	...	618	8.30	26.30

5 rows × 22 columns

In [76]:
```
copy_df.columns = copy_df.columns.str.replace("#","Number").str.replace("%","Percentage")
# copy_df
```

In [77]:
```
copy_df.head()
```

Out[77]:

	Cohort Year	Cohort Category	Demographic	Number Total Cohort	Number Total Grads	Percentage of cohort Total Grads	Number of cohort Total Grads	Percentage of cohort Total Regents	Percentage of grads Total Regents	Number of grads Total Regents	...	Number of grads Advanced Regents	Percentage of cohort Regents w/o Advanced
0	2001	4 Year June	English Language Learner	10540	2791	26.50	992	9.00	35.50	315	...	677	6.40
1	2001	5 Year June	English Language Learner	10540	3920	37.20	1233	11.70	31.50	371	...	862	8.20
2	2001	6 Year	English Language Learner	10540	4296	40.80	1291	12.20	30.10	380	...	911	8.60
3	2002	4 Year June	English Language Learner	7454	1691	22.70	639	8.60	37.80	132	...	507	6.80
4	2002	5 Year June	English Language Learner	7454	2354	31.60	794	10.70	33.70	176	...	618	8.30

5 rows × 22 columns

In [78]:
```
copy_df["Cohort Year"] = pd.to_datetime(copy_df["Cohort Year"]).dt.year

copy_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 430 entries, 0 to 429
Data columns (total 22 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   Cohort Year                             430 non-null    int64
 1   Cohort Category                         430 non-null    object
 2   Demographic                             430 non-null    object
 3   Number Total Cohort                     430 non-null    object
 4   Number Total Grads                      430 non-null    object
 5   Percentage of cohort Total Grads        430 non-null    object
 6   Number of cohort Total Grads            430 non-null    object
 7   Percentage of cohort Total Regents      430 non-null    object
 8   Percentage of grads  Total Regents      430 non-null    object
 9   Number of grads  Total Regents          420 non-null    object
 10  Percentage of cohort  Advanced Regents  420 non-null    object
 11  Percentage of grads  Advanced Regents   420 non-null    object
 12  Number of grads  Advanced Regents       420 non-null    object
 13  Percentage of cohort  Regents w/o Advanced  420 non-null    object
 14  Percentage of grads  Regents w/o Advanced   420 non-null    object
 15  Number of grads  Regents w/o Advanced   430 non-null    object
 16  Percentage of cohort Local              430 non-null    object
 17  Percentage of grads Local               430 non-null    object
 18  Number Still Enrolled                   430 non-null    object
 19  Percentage of cohort Still Enrolled     430 non-null    object
 20  Number Dropped Out                      430 non-null    object
 21  Percentage of cohort Dropped Out        430 non-null    object
dtypes: int64(1), object(21)
memory usage: 74.0+ KB
```

In [79]:
```
copy_df.head(20)
```

Out[79]:

	Cohort Year	Cohort Category	Demographic	Number Total Cohort	Number Total Grads	Percentage of cohort Total Grads	Number of cohort Total Grads	Percentage of cohort Total Regents	Percentage of grads Total Regents	Number of grads Total Regents	...	Number of grads Advanced Regents	Percentage of cohort Regents w/o Advanced
0	2001	4 Year June	English Language Learner	10540	2791	26.50	992	9.00	35.50	315	...	677	6.40
1	2001	5 Year June	English Language Learner	10540	3920	37.20	1233	11.70	31.50	371	...	862	8.20
2	2001	6 Year	English Language Learner	10540	4296	40.80	1291	12.20	30.10	380	...	911	8.60
3	2002	4 Year June	English Language Learner	7454	1691	22.70	639	8.60	37.80	132	...	507	6.80
4	2002	5 Year June	English Language Learner	7454	2354	31.60	794	10.70	33.70	176	...	618	8.30
5	2002	6 Year	English Language Learner	7454	2729	36.60	856	11.50	31.40	185	...	671	9.00
6	2003	4 Year June	English Language Learner	8163	2052	25.10	909	11.10	44.30	290	...	619	7.60
7	2003	5 Year June	English Language Learner	8163	3105	38.00	1159	14.20	37.30	358	...	801	9.80
8	2003	6 Year	English Language Learner	8041	3394	42.20	1216	15.10	35.80	363	...	853	10.60
9	2004	4 Year June	English Language Learner	8486	3039	35.80	1504	17.70	49.50	487	...	1017	12.00
10	2004	4 Year August	English Language Learner	8486	3404	40.10	1599	18.80	47.00	None	...	None	None
11	2004	5 Year June	English Language Learner	8320	4065	48.90	1799	21.60	44.30	558	...	1241	14.90
12	2004	6 Year	English Language Learner	8306	4394	52.90	1873	22.50	42.60	568	...	1305	15.70
13	2005	4 Year June	English Language Learner	8778	3481	39.70	1908	21.70	54.80	583	...	1325	15.10
14	2005	4 Year August	English Language Learner	8778	3899	44.40	2015	23.00	51.70	594	...	1421	16.20
15	2005	5 Year June	English Language Learner	8743	4546	52.00	2232	25.50	49.10	642	...	1590	18.20
16	2005	6 Year	English Language Learner	8673	4852	55.90	2304	26.60	47.50	647	...	1657	19.10
17	2006	4 Year June	English Language Learner	9056	3755	41.50	2437	26.90	64.90	593	...	1844	20.40
18	2006	4 Year August	English Language Learner	9056	4175	46.10	2573	28.40	61.60	608	...	1965	21.70
19	2006	5 Year June	English Language Learner	9158	4775	52.10	2783	30.40	58.30	633	...	2150	23.50

20 rows × 22 columns

Correcting the Data types

In [80]:
```
copy_df.iloc[:,3:] = copy_df.iloc[:,3:].apply(pd.to_numeric)
```

copy_df.loc[:,"Number Total Cohort":"Percentage of cohort Dropped Out"] = copy_df.loc[:,"Number Total Cohort":"Percentage of cohort Dropped Out"].astype(float)

In [81]:
```
copy_df.describe()
```

Out[81]:

	Cohort Year	Number Total Cohort	Number Total Grads	Percentage of cohort Total Grads	Number of cohort Total Grads	Percentage of cohort Total Regents	Percentage of grads Total Regents	Number of grads Total Regents	Percentage of cohort Advanced Regents	Percentage of grads Advanced Regents
count	430.000000	430.000000	430.000000	430.000000	430.000000	430.000000	430.000000	420.000000	420.000000	420.000000
mean	2006.186047	30728.830233	20082.302326	62.319767	16250.748837	49.482558	75.448140	5002.169048	16.532143	23.751190
std	2.866540	21193.889781	15423.868054	16.013714	13673.310038	20.837416	20.221693	4220.128136	12.296028	14.100374
min	2001.000000	6974.000000	1287.000000	17.100000	256.000000	2.900000	13.800000	35.000000	0.400000	1.900000
25%	2004.000000	10821.500000	7944.250000	52.975000	6456.500000	35.700000	63.875000	1626.750000	7.000000	13.000000
50%	2006.000000	26626.000000	16168.000000	65.450000	12180.500000	52.350000	80.700000	3343.000000	14.900000	24.250000
75%	2009.000000	39287.250000	27429.250000	75.200000	23951.500000	65.625000	92.575000	7120.000000	19.600000	28.225000
max	2011.000000	70545.000000	53974.000000	87.700000	52374.000000	85.500000	99.200000	13479.000000	47.100000	58.600000

In [82]:
```
copy_df.head(20)
```

Out[82]:

	Cohort Year	Cohort Category	Demographic	Number Total Cohort	Number Total Grads	Percentage of cohort Total Grads	Number of cohort Total Grads	Percentage of cohort Total Regents	Percentage of grads Total Regents	Number of grads Total Regents	...	Number of grads Advanced Regents	Percentage of cohort Regents w/o Advanced
0	2001	4 Year June	English Language Learner	10540	2791	26.5	992	9.0	35.5	315.0	...	677.0	6.4
1	2001	5 Year June	English Language Learner	10540	3920	37.2	1233	11.7	31.5	371.0	...	862.0	8.2
2	2001	6 Year	English Language Learner	10540	4296	40.8	1291	12.2	30.1	380.0	...	911.0	8.6
3	2002	4 Year June	English Language Learner	7454	1691	22.7	639	8.6	37.8	132.0	...	507.0	6.8
4	2002	5 Year June	English Language Learner	7454	2354	31.6	794	10.7	33.7	176.0	...	618.0	8.3
5	2002	6 Year	English Language Learner	7454	2729	36.6	856	11.5	31.4	185.0	...	671.0	9.0
6	2003	4 Year June	English Language Learner	8163	2052	25.1	909	11.1	44.3	290.0	...	619.0	7.6
7	2003	5 Year June	English Language Learner	8163	3105	38.0	1159	14.2	37.3	358.0	...	801.0	9.8
8	2003	6 Year	English Language Learner	8041	3394	42.2	1216	15.1	35.8	363.0	...	853.0	10.6
9	2004	4 Year June	English Language Learner	8486	3039	35.8	1504	17.7	49.5	487.0	...	1017.0	12.0
10	2004	4 Year August	English Language Learner	8486	3404	40.1	1599	18.8	47.0	NaN	...	NaN	NaN
11	2004	5 Year June	English Language Learner	8320	4065	48.9	1799	21.6	44.3	558.0	...	1241.0	14.9
12	2004	6 Year	English Language Learner	8306	4394	52.9	1873	22.5	42.6	568.0	...	1305.0	15.7
13	2005	4 Year June	English Language Learner	8778	3481	39.7	1908	21.7	54.8	583.0	...	1325.0	15.1
14	2005	4 Year August	English Language Learner	8778	3899	44.4	2015	23.0	51.7	594.0	...	1421.0	16.2
15	2005	5 Year June	English Language Learner	8743	4546	52.0	2232	25.5	49.1	642.0	...	1590.0	18.2
16	2005	6 Year	English Language Learner	8673	4852	55.9	2304	26.6	47.5	647.0	...	1657.0	19.1
17	2006	4 Year June	English Language Learner	9056	3755	41.5	2437	26.9	64.9	593.0	...	1844.0	20.4
18	2006	4 Year August	English Language Learner	9056	4175	46.1	2573	28.4	61.6	608.0	...	1965.0	21.7
19	2006	5 Year June	English Language Learner	9158	4775	52.1	2783	30.4	58.3	633.0	...	2150.0	23.5

20 rows × 22 columns

In [83]:
```
copy_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 430 entries, 0 to 429
Data columns (total 22 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   Cohort Year                             430 non-null    int64
 1   Cohort Category                         430 non-null    object
 2   Demographic                             430 non-null    object
 3   Number Total Cohort                     430 non-null    int64
 4   Number Total Grads                      430 non-null    int64
 5   Percentage of cohort Total Grads        430 non-null    float64
 6   Number of cohort Total Grads            430 non-null    int64
 7   Percentage of cohort Total Regents      430 non-null    float64
 8   Percentage of grads  Total Regents      430 non-null    float64
 9   Number of grads  Total Regents          420 non-null    float64
 10  Percentage of cohort  Advanced Regents  420 non-null    float64
 11  Percentage of grads  Advanced Regents   420 non-null    float64
 12  Number of grads  Advanced Regents       420 non-null    float64
 13  Percentage of cohort  Regents w/o Advanced  420 non-null    float64
 14  Percentage of grads  Regents w/o Advanced   420 non-null    float64
 15  Number of grads  Regents w/o Advanced   430 non-null    int64
 16  Percentage of cohort Local              430 non-null    float64
 17  Percentage of grads Local               430 non-null    float64
 18  Number Still Enrolled                   430 non-null    int64
 19  Percentage of cohort Still Enrolled     430 non-null    float64
 20  Number Dropped Out                      430 non-null    int64
 21  Percentage of cohort Dropped Out        430 non-null    float64
dtypes: float64(13), int64(7), object(2)
memory usage: 74.0+ KB
```

In [84]:
```
copy_df["Demographic"].value_counts(dropna=False)
```

Out[84]:

```
English Language Proficient    43
Student with Disability        43
Not Student with Disability    43
Male                           43
```

```
English Language Learner       43
Asian                          43
Hispanic                       43
White                          43
Black                          43
Female                         43
Name: Demographic, dtype: int64
```

In [85]:
```
copy_df.isna().sum()
```

Out[85]:
```
Cohort Year                                  0
Cohort Category                              0
Demographic                                  0
Number Total Cohort                          0
Number Total Grads                           0
Percentage of cohort Total Grads             0
Number of cohort Total Grads                 0
Percentage of cohort Total Regents           0
Percentage of grads Total Regents            0
Number of grads Total Regents               10
Percentage of cohort Advanced Regents       10
Percentage of grads Advanced Regents        10
Number of grads Advanced Regents            10
Percentage of cohort Regents w/o Advanced   10
Percentage of grads Regents w/o Advanced    10
Number of grads Regents w/o Advanced        10
Percentage of cohort Local                   0
Percentage of grads Local                    0
Number Still Enrolled                        0
Number Dropped Out                           0
Percentage of cohort Dropped Out             0
dtype: int64
```

In [86]:
```
copy_df.iloc[:,3:].describe()
```

Out[86]:

	Number Total Cohort	Number Total Grads	Percentage of cohort Total Grads	Number of cohort Total Grads	Percentage of cohort Total Regents	Percentage of grads Total Regents	Number of grads Total Regents	Percentage of cohort Advanced Regents	Percentage of grads Advanced Regents	Number of grads Advanced Regents
count	430.000000	430.000000	430.000000	430.000000	430.000000	430.000000	420.000000	420.000000	420.000000	420.000000
mean	30728.830233	20082.302326	62.319767	16250.748837	49.482558	75.448140	5002.169048	16.532143	23.751190	11331.492857
std	21193.889781	15423.868054	16.013714	13673.310038	20.837416	20.221693	4220.128136	12.296028	14.100374	9898.656051
min	6974.000000	1287.000000	17.100000	256.000000	2.900000	13.800000	35.000000	0.400000	1.900000	218.000000
25%	10821.500000	7944.250000	52.975000	6456.500000	35.700000	63.875000	1626.750000	7.000000	13.000000	3231.250000
50%	26626.000000	16168.000000	65.450000	12180.500000	52.350000	80.700000	3343.000000	14.900000	24.250000	8880.000000
75%	39287.250000	27429.250000	75.200000	23951.500000	65.625000	92.575000	7120.000000	19.600000	28.225000	16572.250000
max	70545.000000	53974.000000	87.700000	52374.000000	85.500000	99.200000	13479.000000	47.100000	58.600000	38895.000000

pivot table

In [87]:
```
table = copy_df.pivot_table(columns=["Demographic"],aggfunc="sum")
table.head(100)
# copy_df.plot(figsize=(20,20))
# plt.legend(labels)
```

Out[87]:

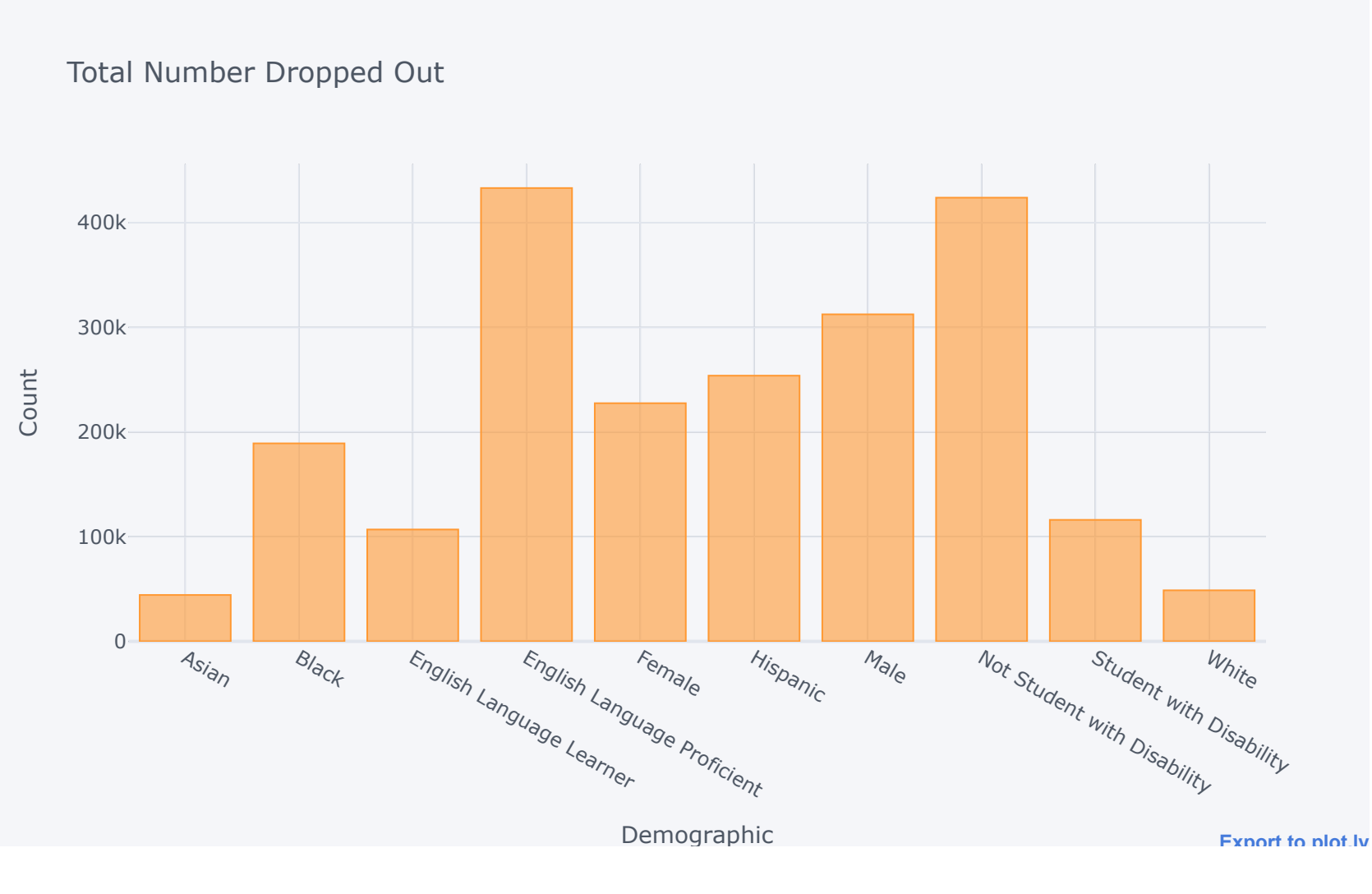
Demographic	Asian	Black	English Language Learner	English Language Proficient	Female	Hispanic	Male	Not Student with Disability	Student with Disability	White
Cohort Year	86266.0	86266.0	86266.0	86266.0	86266.0	86266.0	86266.0	86266.0	86266.0	86266.0
Number Dropped Out	44335.0	189168.0	106885.0	433099.0	227537.0	253857.0	312447.0	423909.0	116075.0	48785.0
Number Still Enrolled	41434.0	190299.0	85119.0	401721.0	203414.0	211646.0	283426.0	365173.0	121667.0	39697.0
Number Total Cohort	502300.0	1091597.0	376064.0	2932758.0	1627122.0	1247013.0	1681700.0	2842375.0	466449.0	446019.0
Number Total Grads	410124.0	660272.0	161744.0	2000454.0	1151613.0	731293.0	1010585.0	2007539.0	154659.0	347107.0
Number of cohort Total Grads	378395.0	493540.0	105182.0	1644468.0	936209.0	560186.0	813441.0	1689436.0	60214.0	306751.0
Number of grads Advanced Regents	155537.0	412526.0	82100.0	1109693.0	631088.0	450233.0	560705.0	1137008.0	54785.0	165552.0
Number of grads Regents w/o Advanced	31804.0	167038.0	56651.0	356741.0	215797.0	171455.0	197595.0	318893.0	94499.0	40469.0
Number of grads Total Regents	215444.0	72136.0	21483.0	504398.0	287549.0	100807.0	238332.0	521222.0	4659.0	134881.0
Percentage of cohort Advanced Regents	1830.7	284.5	244.3	738.0	759.1	345.6	607.9	788.6	41.8	1303.0
Percentage of cohort Regents w/o Advanced	1311.1	1642.7	938.1	1617.9	1662.4	1529.1	1424.1	1719.4	480.4	1605.1
Percentage of cohort Dropped Out	388.2	739.0	1223.5	638.4	602.6	884.3	802.4	640.5	1093.8	462.3
Percentage of cohort Local	291.0	645.5	644.1	527.7	572.8	604.1	509.8	481.8	869.3	383.8
Percentage of cohort Still Enrolled	355.1	753.0	978.1	590.2	538.9	732.2	726.3	553.9	1105.7	383.6
Percentage of cohort Total Grads	3500.3	2605.6	1844.4	2928.8	3040.7	2509.6	2579.3	3036.7	1398.1	3354.0
Percentage of cohort Total Regents	3209.5	1961.5	1200.6	2401.9	2468.5	1907.3	2070.4	2556.2	530.1	2971.5
Percentage of grads Advanced Regents	2245.9	472.2	566.6	1086.3	1074.2	592.4	1018.9	1117.7	129.1	1672.2
Percentage of grads Regents w/o Advanced	1601.1	2659.6	2157.5	2340.4	2323.7	2565.5	2327.4	2397.4	1393.5	2042.6
Percentage of grads Total Regents	3934.7	3192.9	2771.3	3498.9	3467.8	3220.1	3416.8	3588.1	1553.0	3799.1
Percentage of grads Local	366.5	1109.1	1531.9	803.1	833.7	1082.7	885.6	714.0	2749.8	502.1

Data Visualiztion.

In [88]:
```
Drop_out = table.iloc[1,:]
Drop_out
```

Out[88]:
```
Demographic
Asian                            44335.0
Black                           189168.0
English Language Learner        106885.0
English Language Proficient     433099.0
Female                          227537.0
Hispanic                        253857.0
Male                            312447.0
Not Student with Disability     423909.0
Student with Disability         116075.0
White                            48785.0
Name: Number Dropped Out, dtype: float64
```

In [89]:
```
Drop_out.iplot(kind="bar",title="Total Number Dropped Out",xTitle="Demographic",yTitle="Count")
```



Number Dropped Out:

Demographic.

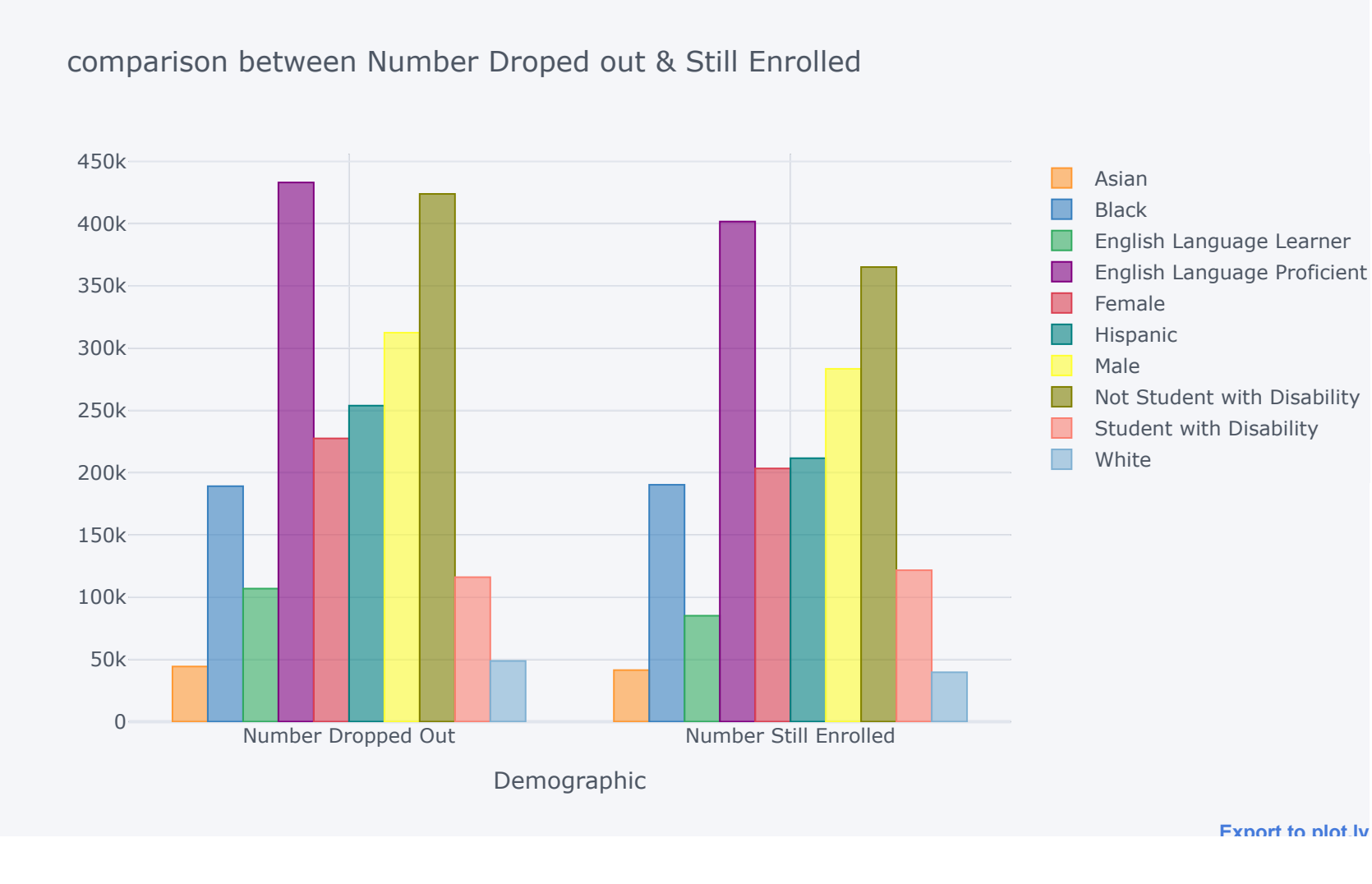
- Asian : 44335.0
- Black : 189168.0
- English Language Learner : 106885.0
- English Language Proficient : 433099.0
- Female : 227537.0
- Hispanic : 253857.0
- Male : 312447.0
- Not Student with Disability : 423909.0
- Student with Disability : 116075.0
- White : 48785.0

In [90]:
```
enrolled_dropped = table.iloc[1:3,:]
enrolled_dropped
```

Out[90]:

Demographic	Asian	Black	English Language Learner	English Language Proficient	Female	Hispanic	Male	Not Student with Disability	Student with Disability	White
Number Dropped Out	44335.0	189168.0	106885.0	433099.0	227537.0	253857.0	312447.0	423909.0	116075.0	48785.0
Number Still Enrolled	41434.0	190299.0	85119.0	401721.0	203414.0	211646.0	283426.0	365173.0	121667.0	39697.0

In [91]:
```
enrolled_dropped.iplot(kind="bar",title="comparison between Number Droped out & Still Enrolled",xTitle="Demo
```



Comarsion between Number Droped out & Still Enrolled

Demographic.

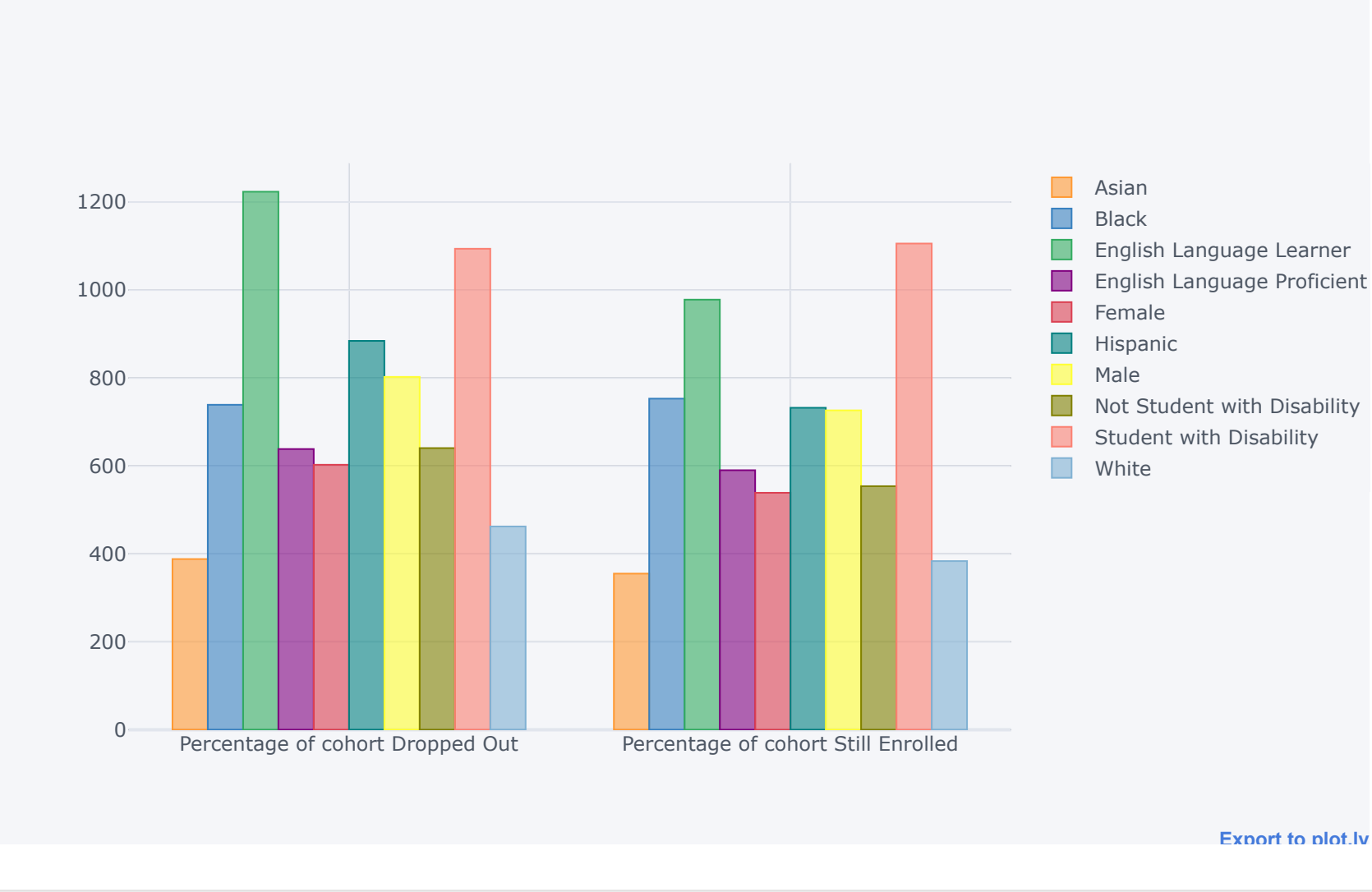
- Asian : 44335.0
- Black : 189168.0
- English Language Learner : 106885.0
- English Language Proficient : 433099.0
- Female : 227537.0
- Hispanic : 253857.0
- Male : 312447.0
- Not Student with Disability : 423909.0
- Student with Disability : 116075.0
- White : 48785.0

In [92]:
```
enrolled_droped_percent = table.loc[["Percentage of cohort Dropped Out","Percentage of cohort Still Enrolled
enrolled_droped_percent
```

Out[92]:

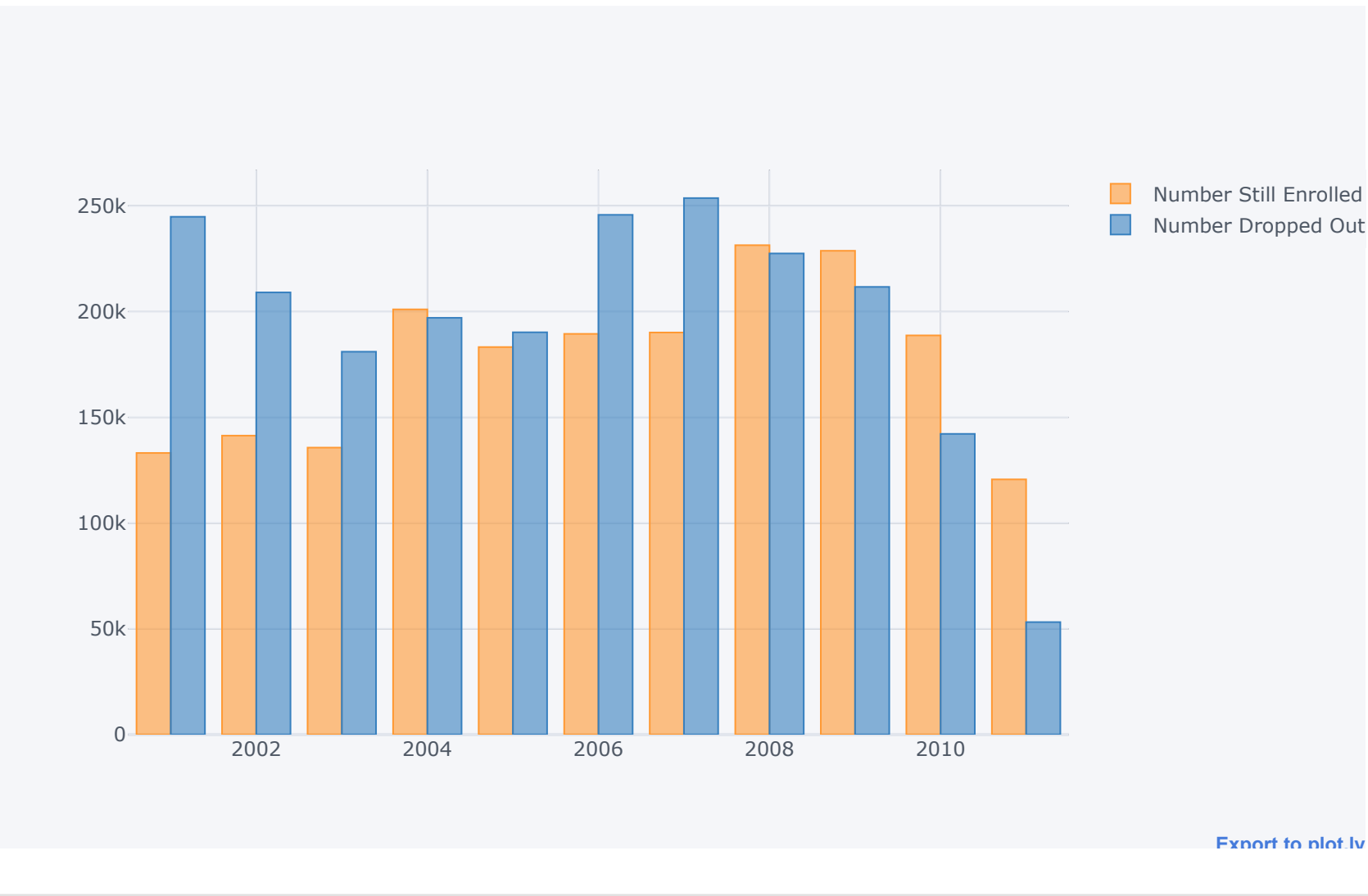
Demographic	Asian	Black	English Language Learner	English Language Proficient	Female	Hispanic	Male	Not Student with Disability	Student with Disability	White
Percentage of cohort Dropped Out	388.2	739.0	1223.5	638.4	602.6	884.3	802.4	640.5	1093.8	462.3
Percentage of cohort Still Enrolled	355.1	753.0	978.1	590.2	538.9	732.2	726.3	553.9	1105.7	383.6

In [93]:
```
enrolled_droped_percent.iplot(kind="bar")
```



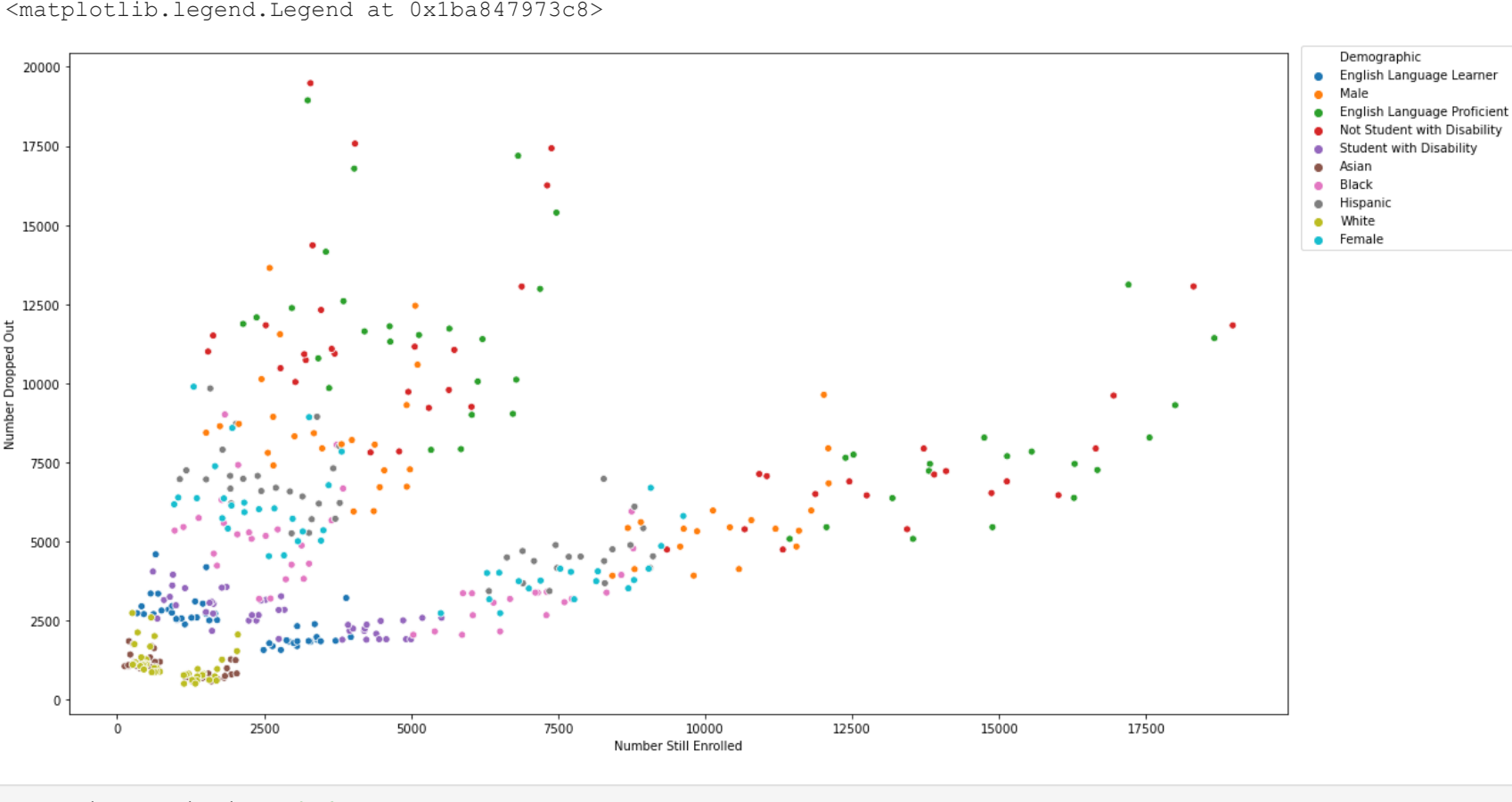
In [94]:
```
# student_dis = copy_df[copy_df["Demographic"] == "Not Student with Disability"]
# student_dis["Percentage of cohort Still Enrolled"].sort_values(ascending=False).sum()
```

In [95]:
```
copy_df.groupby("Cohort Year").sum().loc[2001:,["Number Still Enrolled","Number Dropped Out"]].iplot(kind="b
```



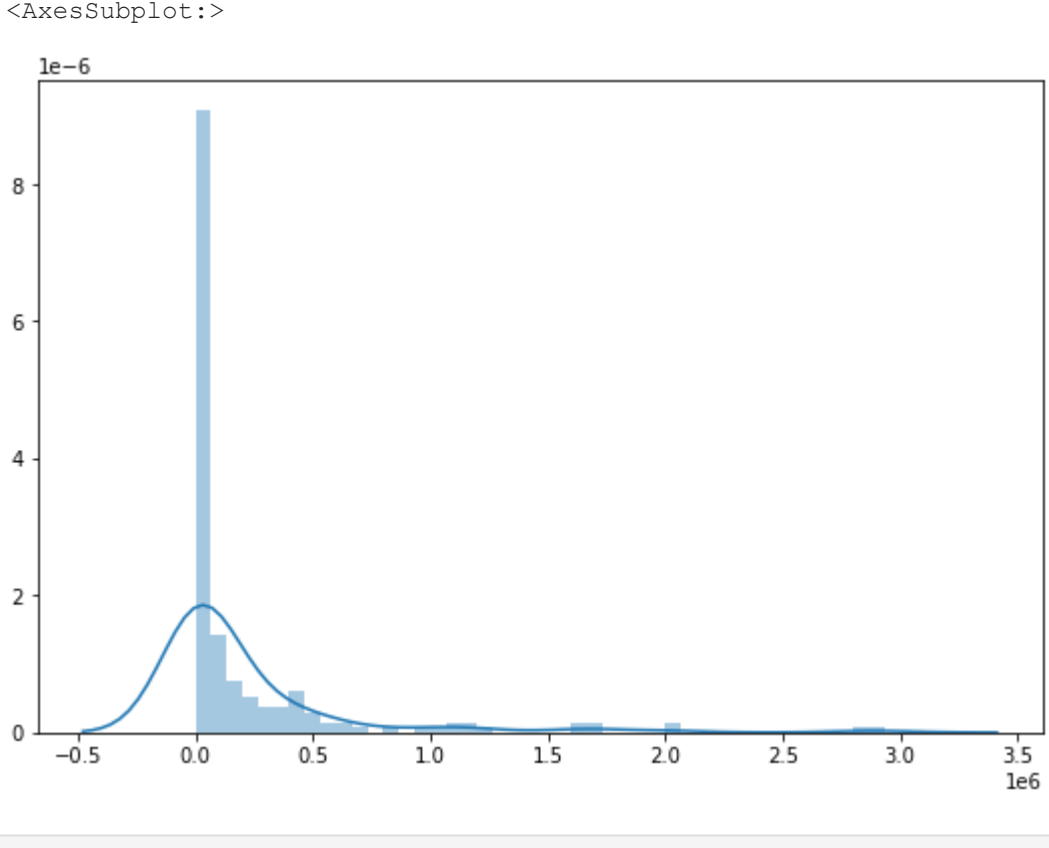
In [96]:
```
plt.figure(figsize=(18,10))
sns.scatterplot(x="Number Still Enrolled",y="Number Dropped Out",data=copy_df,hue=copy_df["Demographic"])
# ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right")
plt.legend(loc="upper right", bbox_to_anchor=(1.19, 1.02))
```

Out[96]:
```
<matplotlib.legend.Legend at 0x1ba4d79c3c8>
```



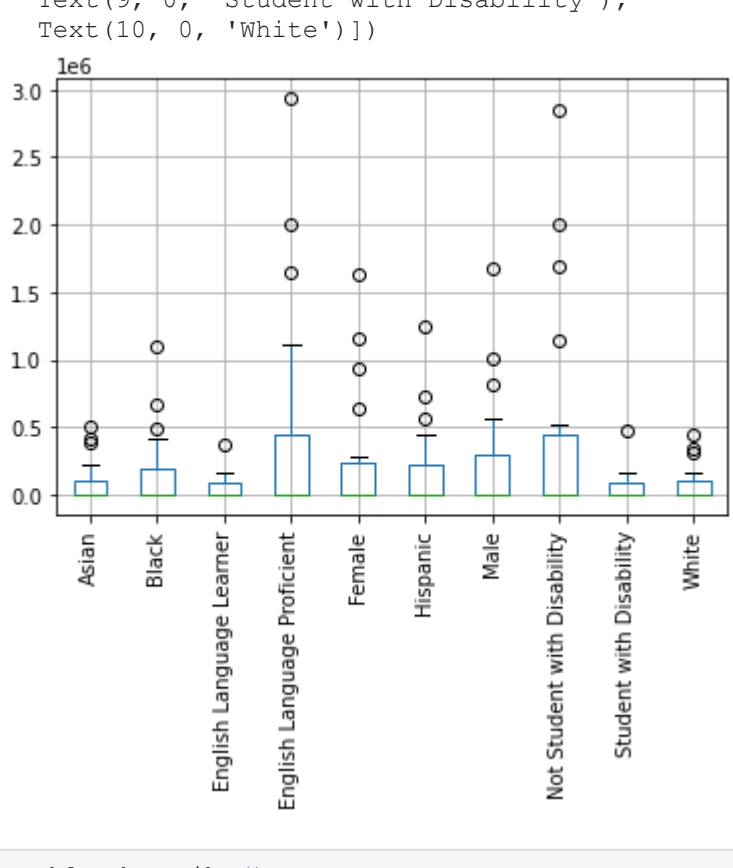
In [97]:
```
plt.figure(figsize=(9,6))
sns.distplot(table)
```

Out[97]:
```
<AxesSubplot:>
```



In [98]:
```
table.boxplot()
plt.xticks(rotation=90)
```

Out[98]:
```
(array([ 1,  2,  3,  4,  5,  6,  7,  8,  9, 10]),
 [Text(1, 0, 'Asian'),
  Text(2, 0, 'Black'),
  Text(3, 0, 'English Language Learner'),
  Text(4, 0, 'English Language Proficient'),
  Text(5, 0, 'Female'),
  Text(6, 0, 'Hispanic'),
  Text(7, 0, 'Male'),
  Text(8, 0, 'Not Student with Disability'),
  Text(9, 0, 'Student with Disability'),
  Text(10, 0, 'White')])
```



In [99]:
```
table.describe()
```

Out[99]:

Demographic	Asian	Black	English Language Learner	English Language Proficient	Female	Hispanic	Male	Not Student with Disability	Student with Disa
count	20.000000	2.000000e+01	20.000000	2.000000e+01	2.000000e+01	2.000000e+01	2.000000e+01	2.000000e+01	20.00
mean	94233.65500	1.689454e+05	54779.720000	4.743385e+05	2.691970e+05	1.914364e+05	2.600433e+05	4.704708e+05	58530.88
std	157092.81478	2.886476e+05	90373.069361	8.200371e+05	4.604941e+05	3.268398e+05	4.445162e+05	8.118638e+05	108649.93
min	291.00000	2.845000e+02	244.300000	5.277000e+02	5.389000e+02	3.456000e+02	5.098000e+02	4.818000e+02	41.80
25%	1528.60000	1.020075e+03	1144.975000	1.015500e+03	1.014075e+03	1.033100e+03	9.855750e+02	1.035425e+03	1037.67
50%	3717.50000	2.926250e+03	2464.400000	3.213850e+03	3.254250e+03	2.892800e+03	2.998050e+03	3.312400e+03	2151.40
75%	103583.75000	1.894508e+05	85405.750000	4.509238e+05	2.425400e+05	2.221988e+05	2.906812e+05	4.482372e+05	88324.25
max	502300.00000	1.091597e+06	376064.000000	2.932758e+06	1.627122e+06	1.247013e+06	1.681700e+06	2.842375e+06	466449.00

In [100]:
```
correlation = copy_df.corr()
correlation
```

Out[100]:

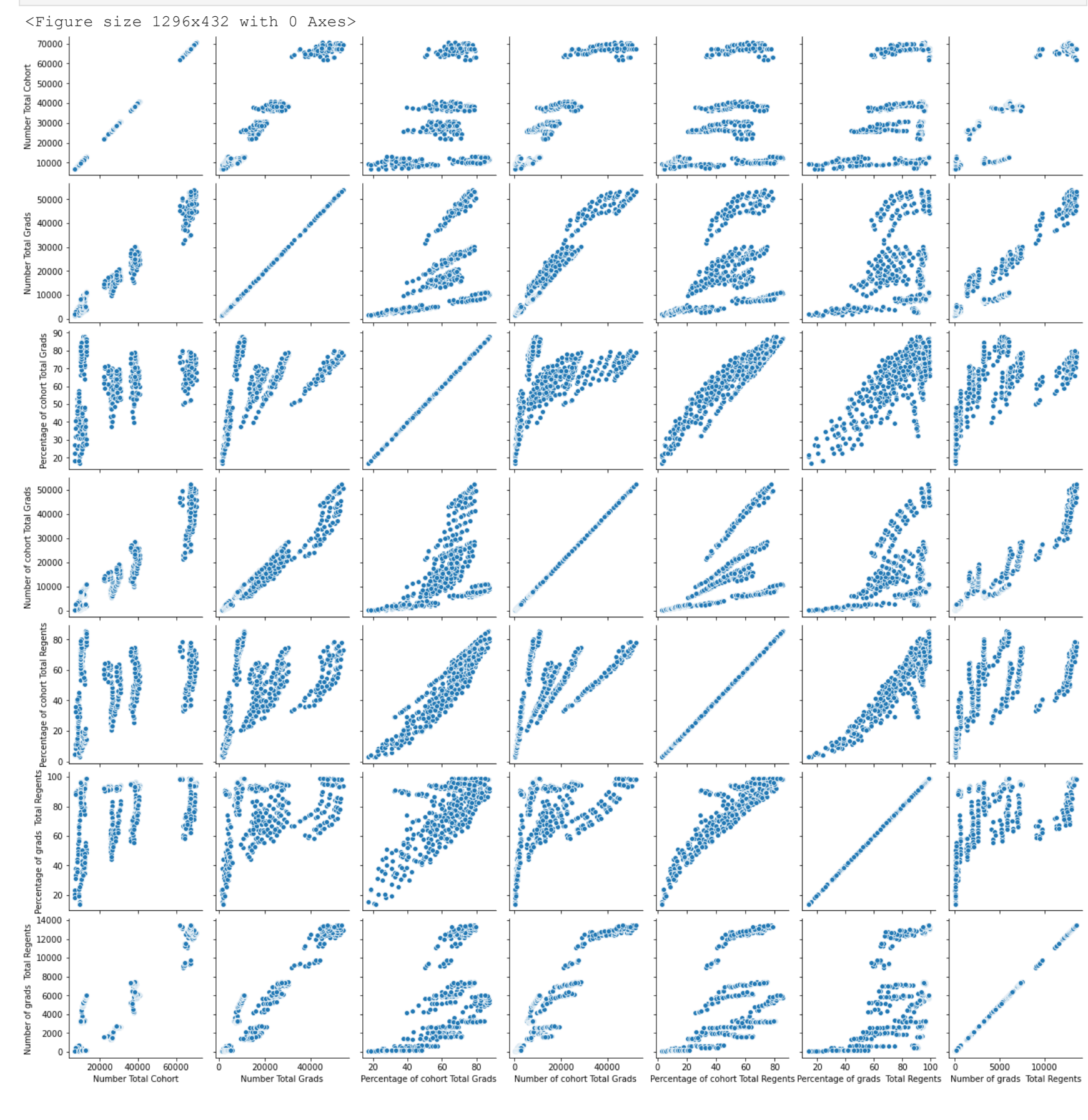
	Cohort Year	Number Total Cohort	Number Total Grads	Percentage of cohort Total Grads	Number of cohort Total Grads	Percentage of cohort Total Regents	Percentage of grads Total Regents	Number of grads Total Regents	Percentage of cohort Advanced Regents	Percentage of grads Advanced Regents	Number of grads Advanced Regents	P
Cohort Year	1.000000	0.017576	0.103372	0.273136	0.262052	0.491712	0.610734	0.108798	0.099772	0.033982	0.315935	
Number Total Cohort	0.017576	1.000000	0.978621	0.275410	0.928518	0.237270	0.254839	0.896170	-0.031183	0.008771	0.906683	
Number Total Grads	0.103372	0.978621	1.000000	0.415488	0.972033	0.375616	0.353325	0.939155	0.085005	0.102302	0.948415	
Percentage of cohort Total Grads	0.273136	0.275410	0.415488	1.000000	0.464393	0.934682	0.763104	0.542888	0.803787	0.757280	0.412329	
Number of cohort Total Grads	0.262052	0.928518	0.972033	0.464393	1.000000	0.487282	0.480533	0.935642	0.145650	0.153614	0.988614	
Percentage of cohort Total Regents	0.491712	0.237270	0.375616	0.934682	0.487282	1.000000	0.915406	0.523359	0.794150	0.756590	0.453892	
Percentage of grads Total Regents	0.610734	0.254839	0.353325	0.763104	0.480533	0.915406	1.000000	0.457438	0.608760	0.616404	0.471127	
Number of grads Total Regents	0.108798	0.896170	0.939155	0.542888	0.935642	0.523359	0.457438	1.000000	0.368430	0.391622	0.871878	
Percentage of cohort Advanced Regents	0.099772	-0.031183	0.085005	0.803787	0.145650	0.794150	0.608760	0.368430	1.000000	0.982324	0.045016	
Percentage of grads Advanced Regents	0.033982	0.008771	0.102302	0.757280	0.153614	0.756590	0.616404	0.391622	0.982324	1.000000	0.046179	
Number of grads Advanced Regents	0.315935	0.906683	0.948415	0.412329	0.988614	0.453892	0.471127	0.871878	0.045016	0.046179	1.000000	
Percentage of cohort Regents w/o Advanced	0.676096	0.400407	0.509038	0.717621	0.626897	0.829468	0.865789	0.477314	0.319286	0.276975	0.666330	
Percentage of grads Regents w/o Advanced	0.744025	0.314463	0.355700	0.292857	0.471418	0.489327	0.721318	0.233446	-0.094860	-0.100749	0.554571	
Number of grads Regents w/o Advanced	-0.517183	0.621275	0.552265	0.014689	0.341046	-0.226058	-0.291342	0.431475	-0.178450	-0.137351	0.279542	
Percentage of cohort Local	-0.719975	-0.064968	-0.143289	-0.423107	-0.332200	-0.717545	-0.838375	-0.270133	-0.448612	-0.443940	-0.348276	
Percentage of grads Local	-0.612753	-0.254243	-0.352835	-0.762642	-0.480245	-0.915126	-0.999977	-0.456950	-0.608137	-0.615654	-0.470935	
Number Still Enrolled	0.021993	0.647294	0.524744	-0.131607	0.514134	-0.039418	0.142480	0.499500	-0.176324	-0.073372	0.500195	
Percentage of cohort Still Enrolled	0.054763	-0.126054	-0.251425	-0.689511	-0.233722	-0.501049	-0.228303	-0.276993	-0.451515	-0.364662	-0.199257	
Number Dropped Out	-0.312299	0.820904	0.760422	0.045953	0.624042	-0.087786	-0.093261	0.602181	-0.239368	-0.196669	0.603760	
Percentage of cohort Dropped Out	-0.495498	-0.192640	-0.274251	-0.681072	-0.374816	-0.784561	-0.767463	-0.437409	-0.671672	-0.661795	-0.337812	

In [101]:
```
plt.figure(figsize=(30,18))
sns.heatmap(correlation,annot=True)
```

Out[101]:
```
<AxesSubplot:>
```

In [102]:
```
plt.figure(figsize=(18,6))
g = sns.PairGrid(copy_df.iloc[:,3:10])
g.map(sns.scatterplot)
sns.set_context(font_scale=1.2)
```

```
<Figure size 1296x432 with 0 Axes>
```



summery:

English Language Proficient & Not Student with Disability have the highest number of Dropped out of school. Asian and White has the Lowest number of Dropped out of school. number of droped out and number of enrolled are very similier to each others they do have the same distrbutions.

generally if you comapre between number of enrolled with dropped there is increase of number 2001 to 2002 about 6.155 in increase in percentage and for the dropped out there about 14.6 decrease in percentage

from 2003 till 2007 it increase and number of enrolled are catching up with dropped out and the dropped out start to increase in number again and from 2008 the number are deacreasing till 2011.

Scatter plot
from the chart you can as the number of enrlloed increase the number of dropped out also increase two observation that English Language Proficient & Not Student with Disability have the highest number of droped and and enrolled and not superisingly we can see Asian and white have both lowest enrolled and droped out.

correlation.

- Number Total Grads & Number of cohort Total Grads & Number of grads Total Regents & Number of grads Advanced Regents have Strong correlation with Number Total Cohort.
- Number Total Grads & Number of cohort Total Grads & Number of grads Total Regents & Number of grads Advanced Regents have Strong correlation with Number Total Grads.
- and if you look you will see that number of enrolled and dropedd they have very weak correlation between them
- number of droped out have strone strong links bwtween Number Total Cohort and Number of grads w/o Advanced.

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: