



December 13, 2021

Carlotta Amaduzzi, Abou Keita, Jenell Lewis,

Michaela Steinrueck, Marissa Tan, Tim Van Wert,

Jeneva Williams-Blackwell

Our Team!!!



Carlotta



Michaela



Tim



Abou



Jenell



Marissa



Jeneva

US Education

- H.S. Graduation rates are used to measure the effectiveness of a School District
- H.S. Graduation rates are used to set long-term goals tied to Federal Funding
- H.S. Graduation rates are tied to local decision makers

Research Question & Objective

- Can high school graduation rates be used as a proxy for effectiveness in education?
- If so, what variables affect graduation rates the most?
- Can States and Local Education Agencies use these levers to enhance their education outcomes and guarantee continuous federal funding?

Overview - Data Sources & Data Cleaning

EDGE Open Data

Public School Characteristics 2018-19



NCES Open Data
National Center for Education
Statistics

Administrative attributes for public elementary and secondary schools from the 2018-2019 Common Core of Data - CCD

https://data-nces.opendata.arcgis.com/datasets/146f28635d01435e978c6d2d5ccf6ea9_0.csv?outSR=%7B%22latestWkid%22%3A3857%2C%22wkid%22%3A102100%7D

- School Level equal High
- Male/Female/Free and Reduced lunch percentages created
- School Type map using dictionary - Regular = 1, all others = 0
- 23,243 unique NCESSCH - School ID



Adjusted Cohort Graduation Rate

Beginning in SY 2015-16, the public file data documentation and data notes have been split into separate documents.

EDFacts File Long Title	LEA Level CSV File*	School Level CSV File*	File Documentation	Supporting Materials
SY 2018-19 Adjusted Cohort Graduation Rate New!	Wide File (1.5 MB) Long File (8.3 MB) New!	Wide File (4.1 MB) Long File (26.0 MB) New!	Documentation	Data Notes

<https://www2.ed.gov/about/inits/ed/edfacts/data-files/index.html>

- Suffix removed for readability
- Category rates dropped - focus on number of students
- Nulls filled with "0" students
- State, District, and School IDs, each Cohort, Success Rate columns selected
- 22,900 unique NCESSCH - School ID



Financial activity of public elementary and secondary school systems

<https://www2.census.gov/programs-surveys/school-finances/tables/2019/secondary-education-finance/elsec19t.xls>

- NCESID (District ID) & PPCSTOT (Per Pupil Spending) columns selected
- NCESID renamed to LEAID
- 14,197 unique LEAID - District ID



Assessment Proficiency

Beginning in SY 2015-16, the Performance Level Mapping tables will be available as a separate Excel spreadsheet. In prior school years, the mapping tables are included as an appendix in the file documentation.

EDFacts File Long Title	LEA Level CSV File*	School Level CSV File*	File Documentation	Supporting Materials
SY 2018-19				

<https://www2.ed.gov/about/inits/ed/edfacts/data-files/index.html>

- Math and Reading Assessment - Students Completed and percentage scored above percentage
- Nulls dropped equal elementary and middle school
- 24,597 unique NCESSCH - School ID

Datasets & Features

District of Columbia, Maryland & Virginia

```
Data columns (total 36 columns):
# Column Non-Null Count Dtype
---
0 Unnamed: 0 571 non-null int64
1 NCESSCH 571 non-null int64
2 TITLEI 571 non-null object
3 LEAD 571 non-null int64
4 G12 571 non-null float64
5 SCHOOL_LEVEL 571 non-null object
6 TOTFRL 571 non-null int64
7 FTE 571 non-null float64
8 STUTERATIO 563 non-null float64
9 TOTMENROL 569 non-null float64
10 TOTFENROL 570 non-null float64
11 PCT_MEN 571 non-null float64
12 PCT_FEN 571 non-null float64
13 PCT_FRL 568 non-null float64
14 SCHOOL_TYPE 571 non-null int64
15 STNAM 571 non-null object
16 FIPST 571 non-null int64
17 ALL_RATE 571 non-null float64
18 ALL_COHORT 571 non-null int64
19 MAM_COHORT 571 non-null float64
20 MAS_COHORT 571 non-null float64
21 MBL_COHORT 571 non-null float64
22 MHI_COHORT 571 non-null float64
23 MTR_COHORT 571 non-null float64
24 MWH_COHORT 571 non-null float64
25 CWD_COHORT 571 non-null float64
26 ECD_COHORT 571 non-null float64
27 FCS_COHORT 571 non-null float64
28 HOM_COHORT 571 non-null float64
29 LEP_COHORT 571 non-null float64
30 Success_Rate 571 non-null int64
31 PPCSTOT 571 non-null int64
32 MATH_NUM 571 non-null float64
33 MATH_PCT 571 non-null float64
34 READ_NUM 571 non-null float64
35 READ_PCT 571 non-null float64
dtypes: float64(24), int64(9), object(3)
```

Feature	Type	Description
NCESSCH	Character	School NCES ID
TITLEI	Character	Title I eligible
LEAID	Character	District NCES ID
G12	Number	Grade 12 students
SCHOOL_LEVEL	Character	Elementary, High, Other, Not reported, Middle, Secondary, Prekindergarten, Not applicable, Ungraded, Adult Education
TOTFRL	Number	Total of free lunch and reduced-price lunch eligible
FTE	Number	Total Teachers
STUTERATIO	Number	Student teacher ratio
TOTMENROL	Number	Total Male Enrollment
TOTFENROL	Number	Total Female Enrollment
PCT_MEN	Number	Total Male Enrollment/ Total Enrollment
PCT_FEN	Number	Total Female Enrollment/ Total Enrollment
PCT_FRL	Number	Total of free lunch and reduced-price lunch eligible/ Total Enrollment
SCHOOL_TYPE	Number	Alternative/other school:0, 'Regular school':1, 'Vocational school':0, 'Special education school':0
STNAM	Character	State Name
FIPST	Character	The two-digit American National Standards Institute (ANSI) code for state
ALL_RATE	Number	Rate of students who graduated within the four-year adjusted-cohort
ALL_COHORT	Number	Total number of students within the four-year adjusted-cohort
MAM_COHORT	Number	Total number of American Indian/Alaska Native students within the four-year adjusted-cohort
MAS_COHORT	Number	Total number of Asian/Pacific Islander students within the four-year adjusted-cohort
MBL_COHORT	Number	Total number of Black students within the four-year adjusted-cohort
MHI_COHORT	Number	Total number of Hispanic students within the four-year adjusted-cohort
MTR_COHORT	Number	Total number of Multiracial students within the four-year adjusted-cohort
MWH_COHORT	Number	Total number of White students within the four-year adjusted-cohort
CWD_COHORT	Number	Total number of students with disabilities within the four-year adjusted-cohort
ECD_COHORT	Number	Total number of economically disadvantaged students within the four-year adjusted-cohort
FCS_COHORT	Number	Number of students in foster care in the graduation cohort
HOM_COHORT	Number	Number of homeless enrolled students in the graduation cohort
LEP_COHORT	Number	Total number of students with limited English proficiency within the four-year adjusted-cohort
Success_Rate	Number	ALL_RATE 67% & Above - Success = 1 ALL_RATE 66% & Below - Low Performing = 0 Every Student Succeeds Act (ESSA) (Paul, n.d.) to continue to enhance schools' performance and accountability to receive federal education funding. This act requires States to set goals for all students as well as subgroups (unlike NCLB), and to report schools with graduation rates below 67% as low performing (2018, Achieve.org, par. 5)
PPCSTOT	Number	The per pupil spending amounts are derived from current spending totals and the fall membership data. Per pupil expenditure does not include spending for nonelementary-secondary programs (community service, adult education), or spending by a school system for students not included in its fall membership counts.
MATH_NUM	Number	Total number of students that completed a MATH assessment and for whom a proficiency level was assigned
MATH_PCT	Number	Percentage of students in the school that scored at or above proficient in MATH assessment
READ_NUM	Number	Total number of students that completed a READING assessment and for whom a proficiency level was assigned
READ_PCT	Number	Percentage of students in the school that scored at or above proficient in READING assessment

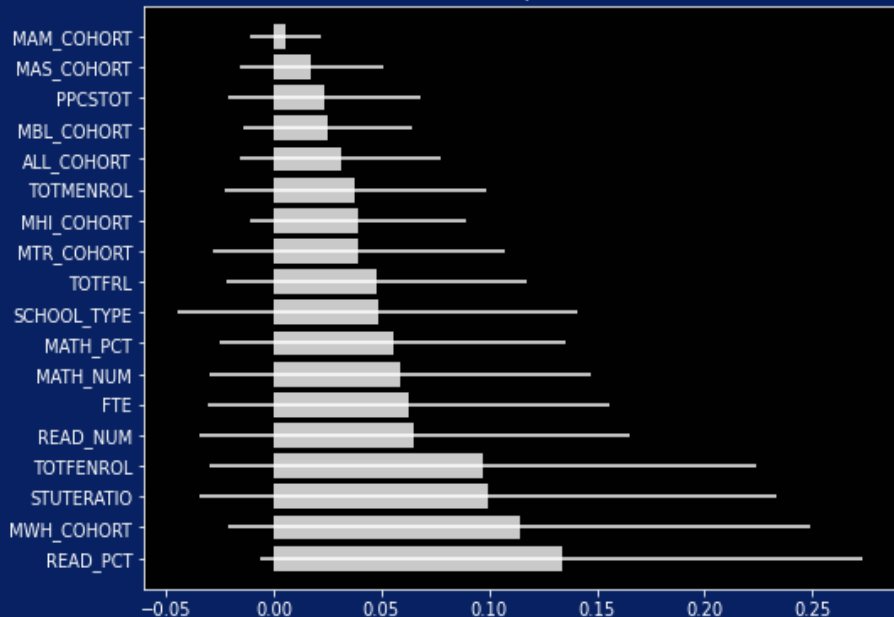
New York

```
Data columns (total 36 columns):
# Column Non-Null Count Dtype
---
0 Unnamed: 0 630 non-null int64
1 NCESSCH 630 non-null int64
2 TITLEI 630 non-null object
3 LEAD 630 non-null int64
4 G12 630 non-null float64
5 SCHOOL_LEVEL 630 non-null object
6 TOTFRL 630 non-null int64
7 FTE 630 non-null float64
8 STUTERATIO 628 non-null float64
9 TOTMENROL 630 non-null float64
10 TOTFENROL 630 non-null float64
11 PCT_MEN 630 non-null float64
12 PCT_FEN 630 non-null float64
13 PCT_FRL 630 non-null float64
14 SCHOOL_TYPE 630 non-null int64
15 STNAM 630 non-null object
16 FIPST 630 non-null int64
17 ALL_RATE 630 non-null float64
18 ALL_COHORT 630 non-null int64
19 MAM_COHORT 630 non-null float64
20 MAS_COHORT 630 non-null float64
21 MBL_COHORT 630 non-null float64
22 MHI_COHORT 630 non-null float64
23 MTR_COHORT 630 non-null float64
24 MWH_COHORT 630 non-null float64
25 CWD_COHORT 630 non-null float64
26 ECD_COHORT 630 non-null float64
27 FCS_COHORT 630 non-null float64
28 HOM_COHORT 630 non-null float64
29 LEP_COHORT 630 non-null float64
30 Success_Rate 630 non-null int64
31 PPCSTOT 630 non-null int64
32 MATH_NUM 630 non-null float64
33 MATH_PCT 630 non-null float64
34 READ_NUM 630 non-null float64
35 READ_PCT 630 non-null float64
dtypes: float64(24), int64(9), object(3)
```


Datasets & Feature Importance-Random Forest

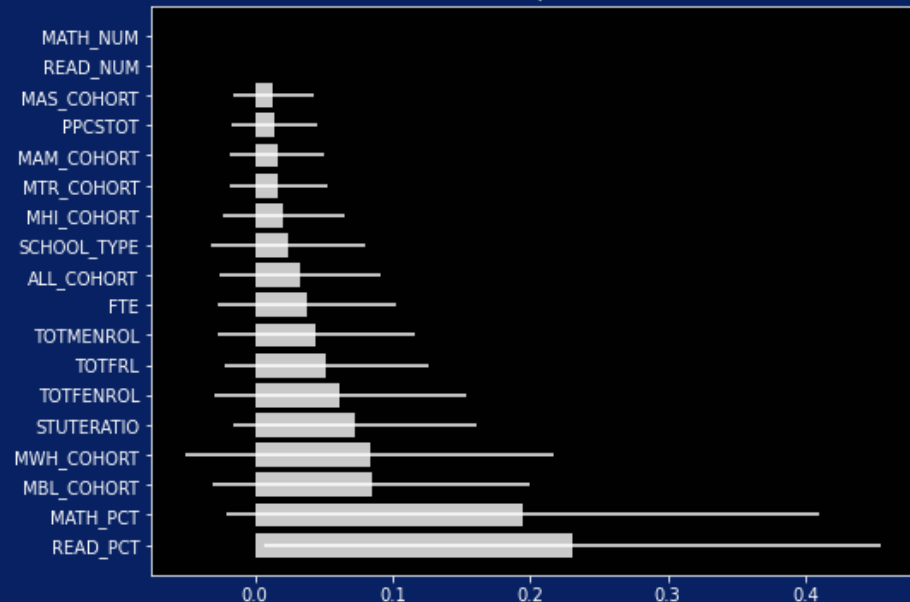
District of Columbia, Maryland & Virginia

Feature importances



New York

Feature importances



Hypothesis and Methodology

- **Focus:**
Variables affecting graduation rates and the prediction of successful H.S. completion.
Success Rate threshold = 67% (ESSA legislation marker for low-performing schools)
- **Exploratory Data Analysis**
- **Models from LazyPredict:**
 - Regression Model:**
Gradient Boosting Regressor
 - Classification Models:**
 - LinearSVC
 - Logistic Regression(with Lasso and Ridge)
 - Random Forest
- **Other Models:**
 - Neural Networks
 - Agglomerative Clustering

Expected Outcomes

Gender

Females students are more likely to graduate than their male counterparts.

Student Teacher Ratio

As the student-teacher ratio increases, the graduation rate decreases.

Adjusted Cohort Graduation Rate

Percentage of public high school students who graduate on time with a regular diploma,

First 6 years the ACGR was collected (2010–11 through 2015–16), the rate increased from 79 percent to 84 percent.

Free and Reduced Lunch

The average percentage of free and reduced lunch students is higher for schools with lower graduation rates compared to schools with higher graduation rates.

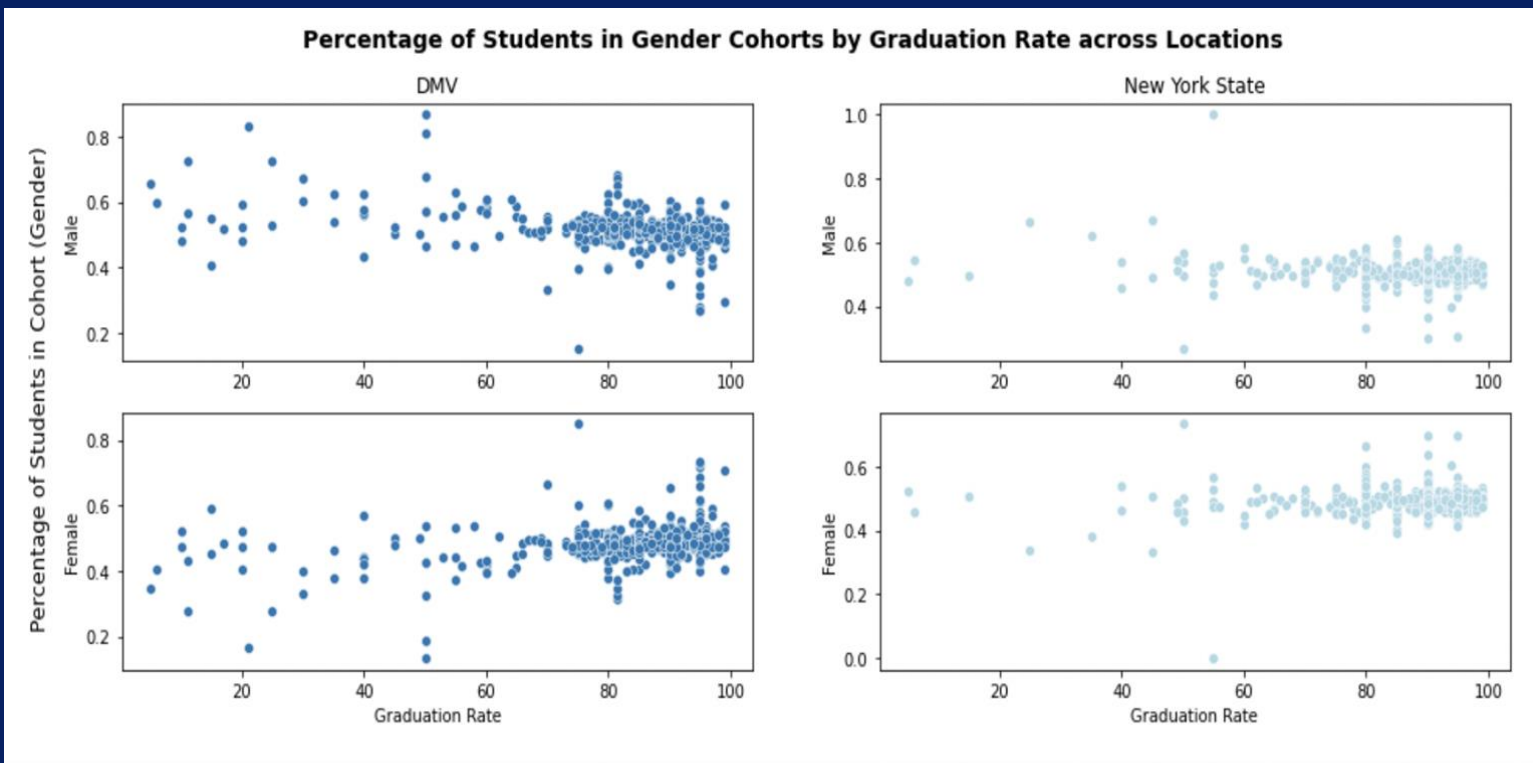
Funding

A significant portion of Title I eligible schools receive less funding than their non-eligible counterparts.

Schools with more female students, lower student-teacher ratios, a lower percentage of free and reduced lunch recipients, and non-eligible Title I schools should have higher graduation rates.

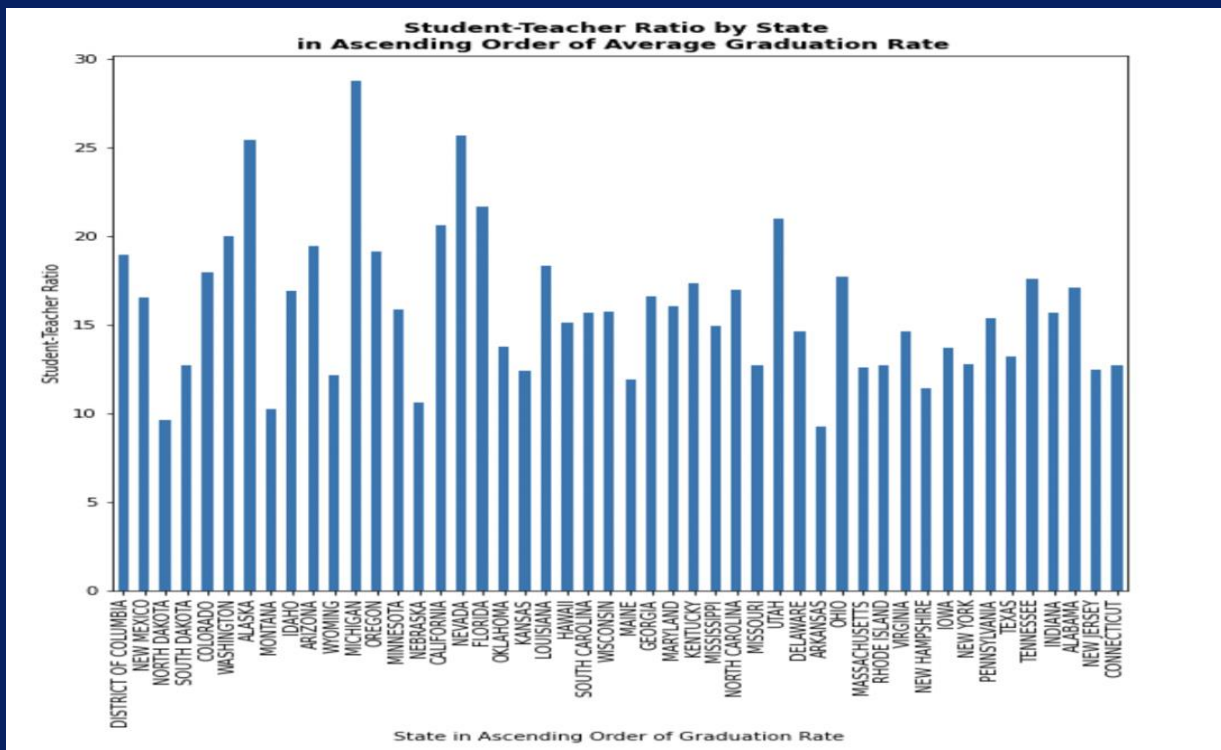
Gender - Key Findings & Results

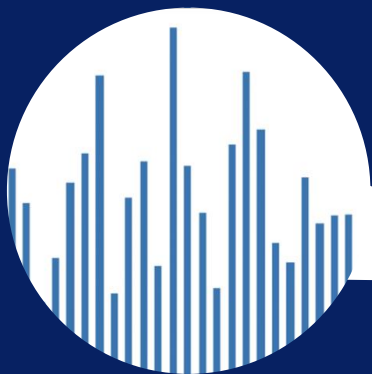
- Lower graduation rates for males



Student Teacher Ratio - Key Findings & Results

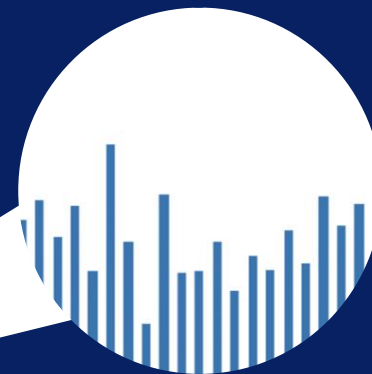
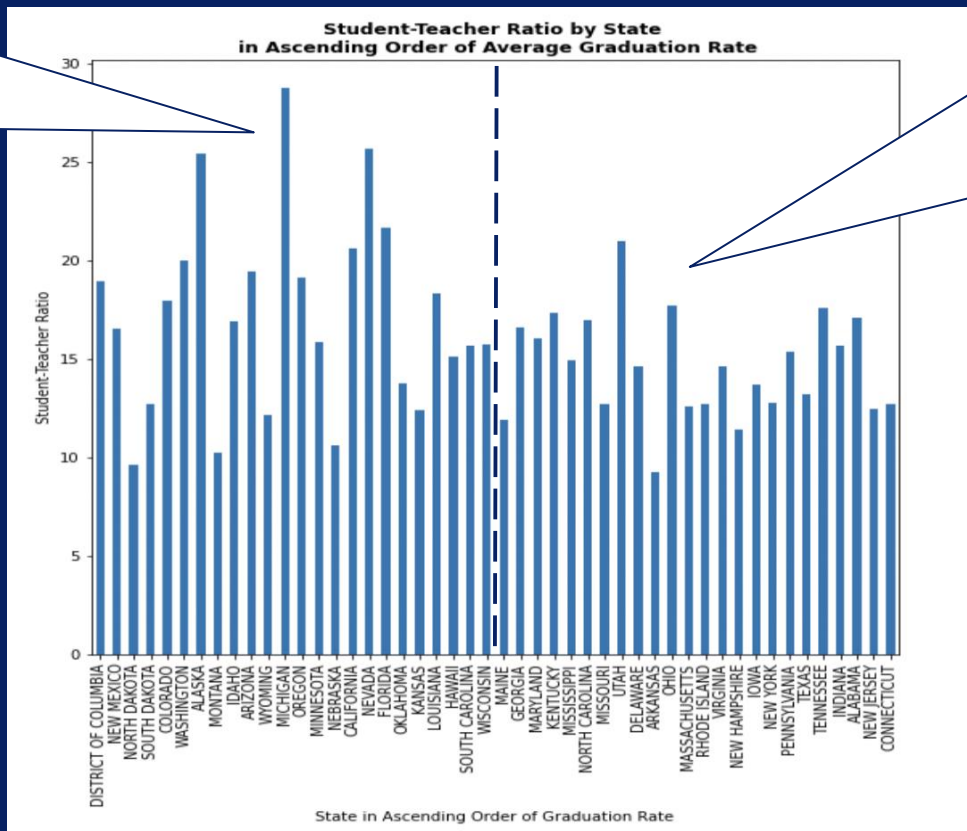
- There is a negative correlation between Student-Teacher ratio and graduation rates





Lowest Graduation Rate

District of Columbia: 62.4%

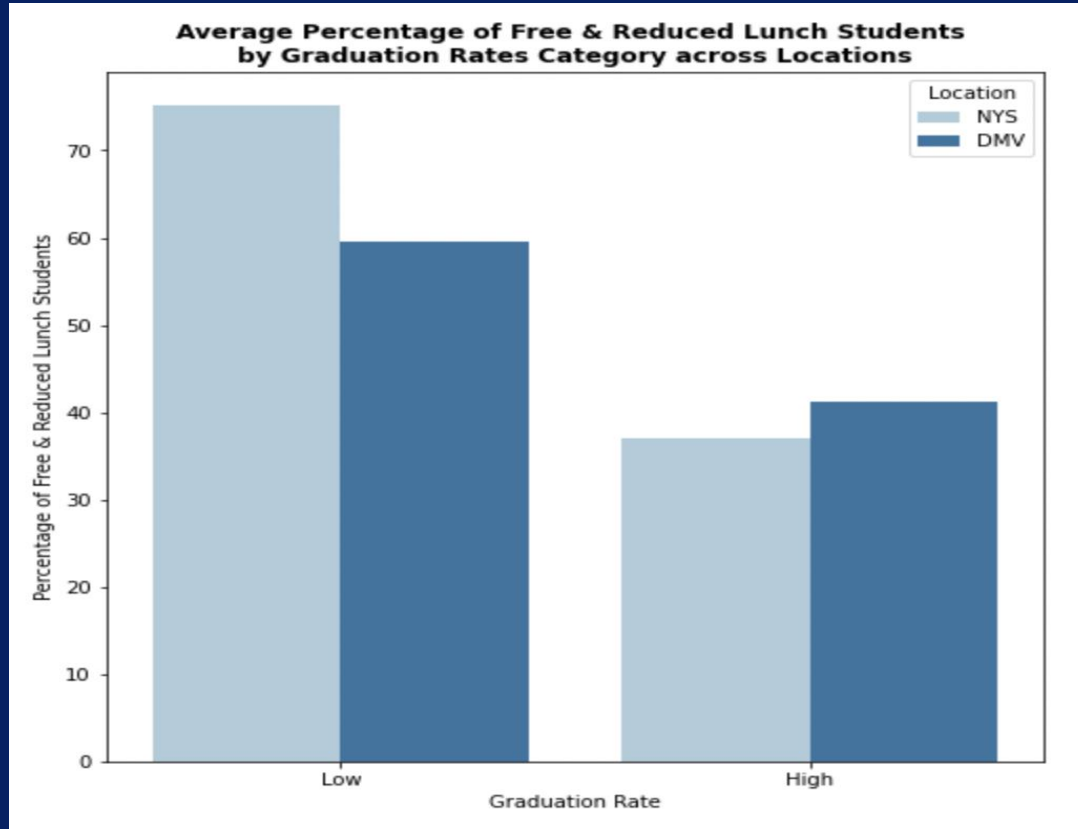


Highest Graduation Rate

Connecticut: 90.7%

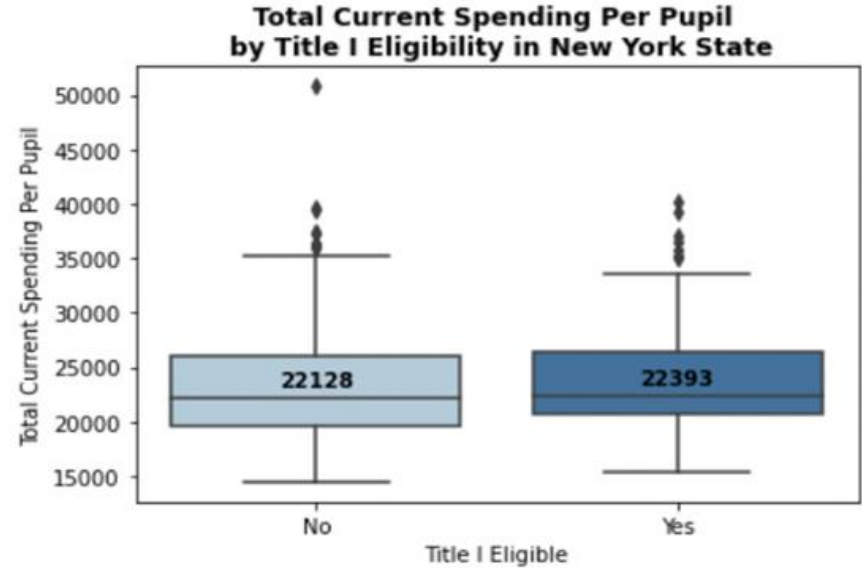
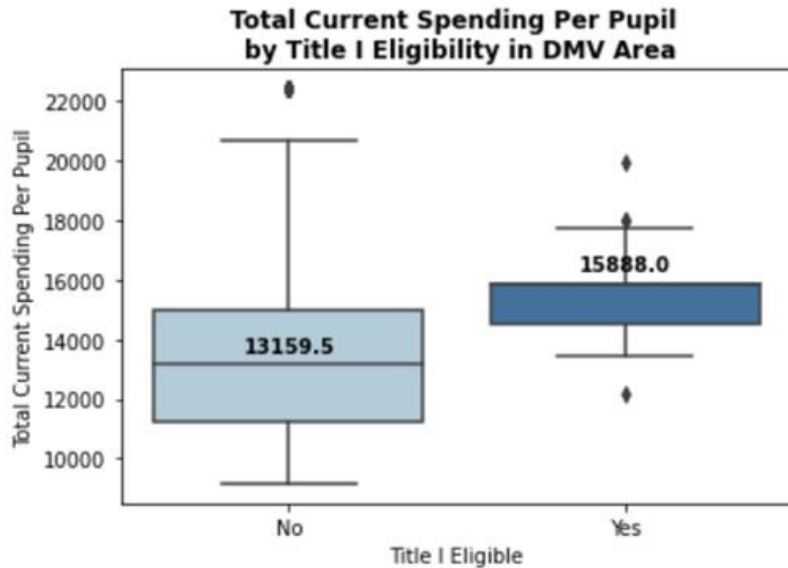
Free and Reduced Lunch - Key Findings & Results

- Inverse relationship between the graduation rate and the lunch program



Funding - Key Findings & Results

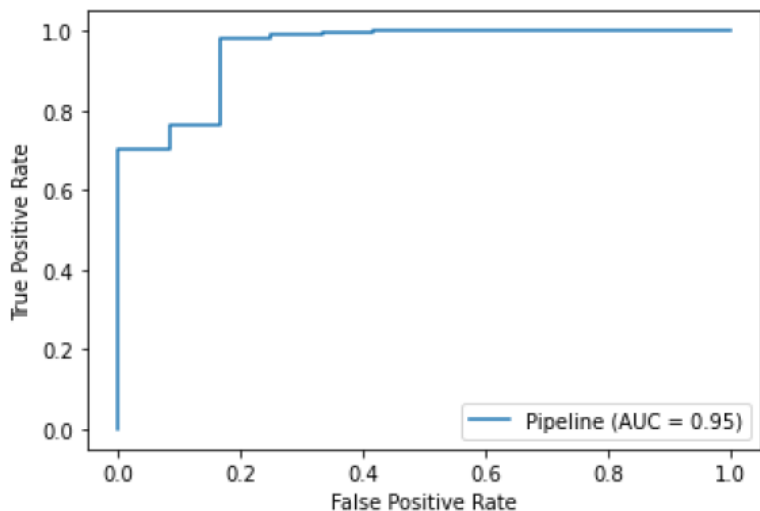
- More funding for Title I eligible schools in the DMV area
- For NY, both Title I eligible schools and schools that are not Title I eligible received approximately the same amount of funding



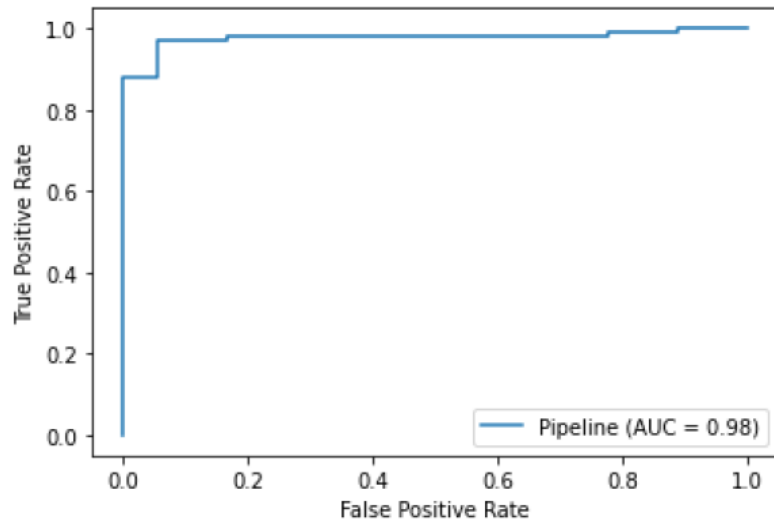
ROC Graphs of LinearSVC

- NY's AUC score was 0.95 and DMV's was 0.98 (accuracy score was 0.97 for both)

New York State

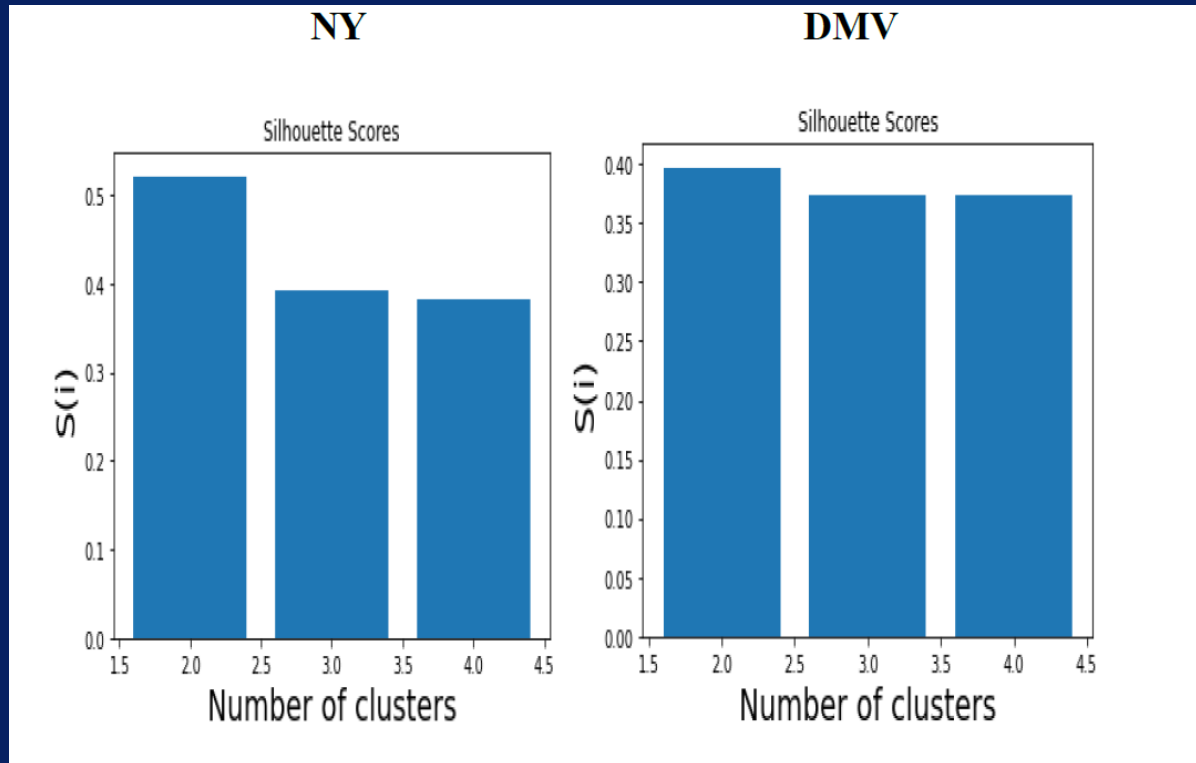


DMV Area



Clustering- Silhouette Scores

- $k=2$ has the highest silhouette score, so the number of clusters in the dataset is two.



Limitations

- The Family Educational Rights and Privacy Act (FERPA)
- Local Education Agencies (LEAs) Subdivisions and reporting standards
- Financial and Statistical data disclosure practices

Future Research

- Title I schools' focus
- Rural vs. Urban area schools
- Long Term State-level planning & goal- setting

