

Grad Rate and Outcome in New York 2022

Mai Riordan

1. Data

1a. Data Source

The primary data is publicly available open-source data:

[https://data.nysed.gov/Graduation Rate in New York 2022 | NSYED](https://data.nysed.gov/Graduation-Rate-in-New-York-2022-|NSYED)

The additional data is also publicly open-source data:

<https://mhvcommunityprofiles.org/education/high-school-cohort-graduation-rate>

1b. Data Collection

The primary database contains annual graduation, and dropout data for the New York state as well as by county, Need to Resource Capacity group, district, public school and charter school, collected from New York State Education Department. Annual graduation data is included for the current four-year cohort (June and August graduates), five-year June and August, and six-year June and August cohorts.

The additional database contains the graduation rate of NY state across all counties from 2022 to 2022.

1c. Data Contents

The primary dataset includes outcomes of designated subgroups are reported by total public school (aggregated data for all districts and charter schools), county (aggregated data for all districts and charter schools in the county), Needs-to-Resource-Capacity (N/RC) group, district, and public schools

1d. Data Profile

- The data has 221866 rows \times 36 columns. After the Data Quality check, the dataset contains 221866 rows \times 31 columns.
- Changes were made:

- Drop unnecessary columns: 'lea_beds', 'lea_name', 'nrc_code', 'boces_code', 'nyc_ind'
- Rename columns:
 - 'nrc_desc': 'needs_to_resource_capacity_description',
 - enroll_cnt': 'enroll_count',
 - 'grad_cnt': 'grad_count',
 - 'grad_pct': 'grad_percentage',
 - 'reg_cnt': 'regents_diploma_count',
 - 'reg_pct': 'regents_diploma_percentage',
 - 'local_cnt' : 'local_count',
 - 'local_pct' : 'local_percentage',
 - 'reg_adv_cnt': 'regents_advanced_count',
 - 'reg_adv_pct': 'regents_advanced_percentage',
 - 'Non_diploma_credential_cnt': 'non_diploma_credential_count',
 - 'non_diploma_credential_pct': 'non_diploma_credential_percentage',
 - 'still_enr_cnt': 'still_enroll_count',
 - 'still_enr_pct': 'still_enroll_percentage',
 - 'ged_cnt': 'ged_count',
 - 'ged_pct': 'ged_percentage',
 - 'dropout_cnt': 'dropout_count',
 - 'dropout_pct': 'dropout_percentage'}
- Convert data: aggregation_code, county_code, membership_code, membership_key, subgroup_code into string type.
- There are no duplicate values.
- Table: **GRAD_RATE_AND_OUTCOMES_2022**

Index	Column	Description	Time Variant/ Invariant	Data Type
-------	--------	-------------	----------------------------	-----------

1	REPORT_SCHOOL_YEAR	School Year in which the data was collected	Variant	Qualitative
2	AGGREGATION_INDEX	Numeric index assigned to assist in aggregating data at Statewide (0), Need/Resource Category (1), County (2), District (3) and School (4)	Invariant	Quantitative
3	AGGREGATION_TYPE	Aggregation type description	Invariant	Qualitative
4	AGGREGATION_CODE	Codes include N/RC codes and BEDS codes	Invariant	Quantitative
5	AGGREGATION_NAME	The name of the entity (district or school)	Invariant	Qualitative
6	ENTITY_INACTIVE_DATE	When an LEA is no longer active (e.g., a school has been closed, a district merged into another), this field contains the date it was made inactive	Variant	Qualitative
7	COUNTY_CODE	2-digit county code	Invariant	Quantitative
8	COUNTY_NAME	County name	Invariant	Qualitative
9	NRC_DESP	Description of Needs to Resources Capacity	Invariant	Qualitative
10	BOCES_NAME	Name of the BOCES	Invariant	Qualitative
11	MEMBERSHIP_CODE	1 or 2-digit code designating the cohort membership being reported	Invariant	Quantitative

12	MEMBERSHIP_KEY	3-digit code that is specific to the cohort and the school year being reported	Invariant	Quantitative
13	MEMBERSHIP_DESC	Description of the cohort membership	Invariant	Qualitative
14	SUBGROUP_NAME	Name of subgroup	Invariant	Qualitative
15	SUBGROUP_CODE	2-digit code identifying the various demographic subgroups	Invariant	Quantitative
16	ENROLL_CNT	Count of students in the cohort based on the last enrollment record as of June 30, 2021, with a First Date of Entry into Grade 9 during the 2017-18 school year, regardless of their current grade level	Invariant	Quantitative
17	GRAD_CNT	Number of students in the cohort who earned either a Regents or Local diploma	Invariant	Quantitative
18	LOCAL_CNT	Number of students in the cohort who earned a Local diploma	Invariant	Quantitative
19	REG_CNT	Number of students in the cohort who earned a Regents diploma	Invariant	Quantitative
20	REG_ADV_CNT	Number of students in the cohort who earned a Regents diploma with Advanced Designation	Invariant	Quantitative
21	GRAD_PCT	Percent of students in the cohort who earned either a Regents or Local diploma	Invariant	Qualitative

22	LOCAL_PCT	Percent of students in the cohort who earned a Local diploma	Invariant	Qualitative
23	REG_PCT	Percent of students in the cohort who earned a Regents diploma	Invariant	Qualitative
24	REG_ADV_PCT	Percent of students in the cohort who earned a Regents diploma with Advanced Designation	Invariant	Qualitative
25	NON_DIPLOMA_CREDENTIAL_CNT	Number of students in the cohort who earned a non-diploma commencement credential	Invariant	Quantitative
26	NON_DIPLOMA_CREDENTIAL_PCT	Percent of students in the cohort who earned a non-diploma commencement credential	Invariant	Qualitative
27	STILL_ENR_CNT	Number of students in the cohort who were still enrolled	Invariant	Qualitative
28	STILL_ENR_PCT	Percent of of students in the cohort who were still enrolled	Invariant	Qualitative
29	GED_CNT	Number of students in the cohort who entered an approved high school equivalency program	Invariant	Quantitative
30	DROPOUT_CNT	Numbers of students who dropped out	Invariant	Quantitative
31	DROPOUT_PCT	Percentage of students who dropped out	Invariant	Qualitative

2. Limitations and Ethics

Data Limitations:

- Incomplete or missing data: The data contains missing values, which can limit the accuracy and reliability of analyses.
- Sampling bias: If the data collection process involves sampling, there is a possibility of bias if the sample does not accurately represent the target population.
- Data quality and reliability: The accuracy and reliability of the data depend on the data collection methods and the competence of the data collectors.

Data Ethics:

- The data contains no personal Information so no PLA Security is required.

3. Questions to explore

- How does the graduation rate vary across different cohorts and subgroup demographics?
- Are there any significant differences in graduation rates between students who earned a Regents diploma and those who earned a Local diploma?
- What is the distribution of students who earned a Regents diploma with Advanced Designation across different cohorts?
- How does the enrollment count vary across different districts and schools?
- Are there any patterns or trends in the dropout rates among different demographic subgroups?
- How does the percentage of students still enrolled in the cohort vary across different cohorts and entities?
- Any correlations or relationships between the enrollment count and the graduation rates?