1. Can you count something interesting?

Yes, the unique count of school districts in the selected state ('MA') =301;

Unique non-null counts for selected subgroups;

Years selected unique count =5 (2011-2015)

2. Can you find trends (e.g high, low, increasing, decreasing, anomalies)

Yes, mean of graduation rate grouped by years across all local education agency (LEA) IDs is plotted to get the trending-

1. For Limited English proficient students (LEP) subgroup of students- the rate is linear with steep slope followed by a dip in the junior high school year; Overall this would indicate we had good number of cohort members graduating by the end of 2014-2015 school year; the dip factors in LEP requirements some of whom graduating early and some taking additional time.

Also to note the graduation rate is calculated for the cohort entering 9th grade for the first time in 2011-12 school year (plus students who transferring in minus who transferred out, emigrated etc during school years 2011-12, 2012-13, 2013-14 and 2014-15)

1. Mean of graduation rate grouped by years across all LEA IDs, is plotted to get the trending.

For two or more races (MTR) subgroup of students- the rate is trending downward correlating with highest dropout rates as referenced in Table 10 -Competency Determination (CD) status as

attached -



3. Can you make a bar plot or histogram?

Yes, bar plot showing the counts from all the states for 2011-15 cohort year.

4. Can you compare two related quantities?

Yes, with seaborn pairplot, (<https://seaborn.pydata.org/generated/seaborn.pairplot.html>) subset of variables of interest are plotted scatterplots are drawn for joint relationships -

Two or More Races Rate (MTR\_RATE),

Limited English proficient students Rate (LEP\_RATE),

Children with disabilities Rate (CWD\_RATE),

Economically disadvantaged students Rate (ECD\_RATE)

5. Can you make a scatter plot?

Yes. Further, kde (kernel density estimates) are used for univariate plots and to fit linear regression models to the scatter plots.

6. Can you make a time-series plot?

Yes, mean of graduation rate grouped by years across all LEAIDs is plotted to get the trending

Children with Disabilities - (aka Individuals with Disabilities Education Act (IDEA)) subgroup of students- the rate is linear with about 5% variance positive slope. Students receiving Special Education services graduates in 4 year or less.

Also to note the graduation rate is calculated for the cohort entering 9th grade for the first time in 2011-12 school year (plus students who transferring in minus who transferred out, emigrated etc during school years 2011-12, 2012-13, 2013-14 and 2014-15)

II. Looking at the plots, what are some insights you can make? Do you see any correlations?

Is there a hypothesis you’d like to investigate further? What other questions do the insights lead you to ask?

From the plots, it is apparent that there is strong correlation between the LEP, IDEA and ECD subgroups and tends to increase over time.

MTR rate indicates almost flat correlation (with minimal slope) against the other 3 variables, which indicates the performance pattern not much influenced by Race/Ethnicity subgroups.

This type of plot provides us with insights on the chosen variables with little or none reliance on domain knowledge and draws our attention to the reported variables that didn't have a pre-existing interest to start with.

III. Now that you’ve asked questions, hopefully you’ve found some interesting insights. Is there a narrative or a way of presenting the insights using text and plots that tells a compelling story? What are some other trends/relationships you think will make the story more complete?

These plots are generated to preview just a very simple use case -with one state; to analyze/predict

performance pattern(s) against reported subgroups.

The analysis can be extended to compare against all 50 states to bring out any outliers by state/LEA and school level.

in order to further evaluate and correlate with other data assets - such as financial grant information;

highly-qualified teacher (HQT) information; and its impact on graduation/dropout rates among the subgroups.

This will support and better align with the U.S Department of Education's (USDE) Edfacts initiative to promote the use of high-quality data, pre-KG through grade 12, and to support planning, policymaking and management/decision making.