BrainStation Data Science

Capstone Project – Sprint 1

Predicting Secondary School Graduation Rates in British Columbia

William Reggler – September 2024

The Problem Area

Can we use information about students and factors affecting graduation rates to predict whether a high school student in British Columbia will graduate?

- "The purpose of the British Columbia school system is to enable learners to develop their individual potential and to acquire the knowledge, skills, and attitudes needed to contribute to a healthy society and a prosperous and sustainable economy." – Mission Statement, *Mandate for the School System of British Columbia*, September 1989

- The curriculum of British Columbia's school system was redesigned in 2015 to focus on **core competencies** (*communication*, *thinking*, and *personal & social*) and **curricular competencies** (*key content*, *skills & strategies*, and *big ideas*).

Why does this matter?

- Graduating from secondary school is one of the key indicators of the success or failure of the BC Government's mission statement.
- People who graduate from secondary school have significantly better outcomes than those who do not in terms of:
  - *Work prospects*: Canadians who do not graduate earn 22% less than graduates on average.
  - *Health & well-being*: Secondary school graduates are healthier and rely less on public health and social services than non-graduates.
  - *Crime*: People who do not graduate from secondary school are more likely to be incarcerated.
  - *Economic impact*: Societies with less educated workers tend to have lower tax revenues and increased government costs.

Source: https://www.pathwaystoeducation.ca/the-impacts-of-socioeconomic-status-and-educational-attainment-on-youth-success/

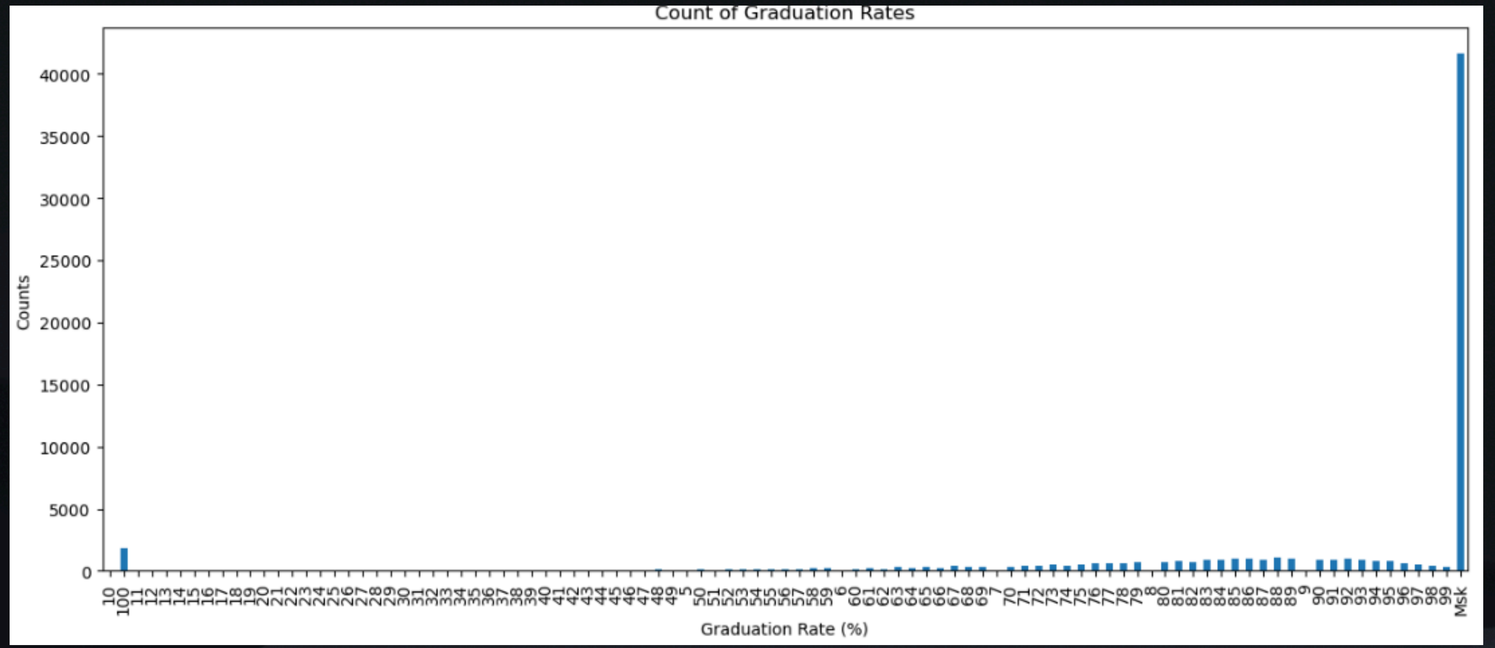
How will this project make an impact?

- A categorical model will be trained to predict if a student is ‘likely’ or ‘unlikely’ to graduate from secondary school (ideally, a percentage likelihood will be calculated).
- This model could be used to give a school more understanding of its student body and where time, staff and resources should be focused.
- Armed with this information, a teacher will be able to:
  - Tailor the focus of a class to more academic or more practical topics.
  - Adjust their approach to the big ideas and curricular competencies.
  - Differentiate for students with diverse needs, including those with Individual Education Plans (IEPs).

Insights into the data

A masked dilemma

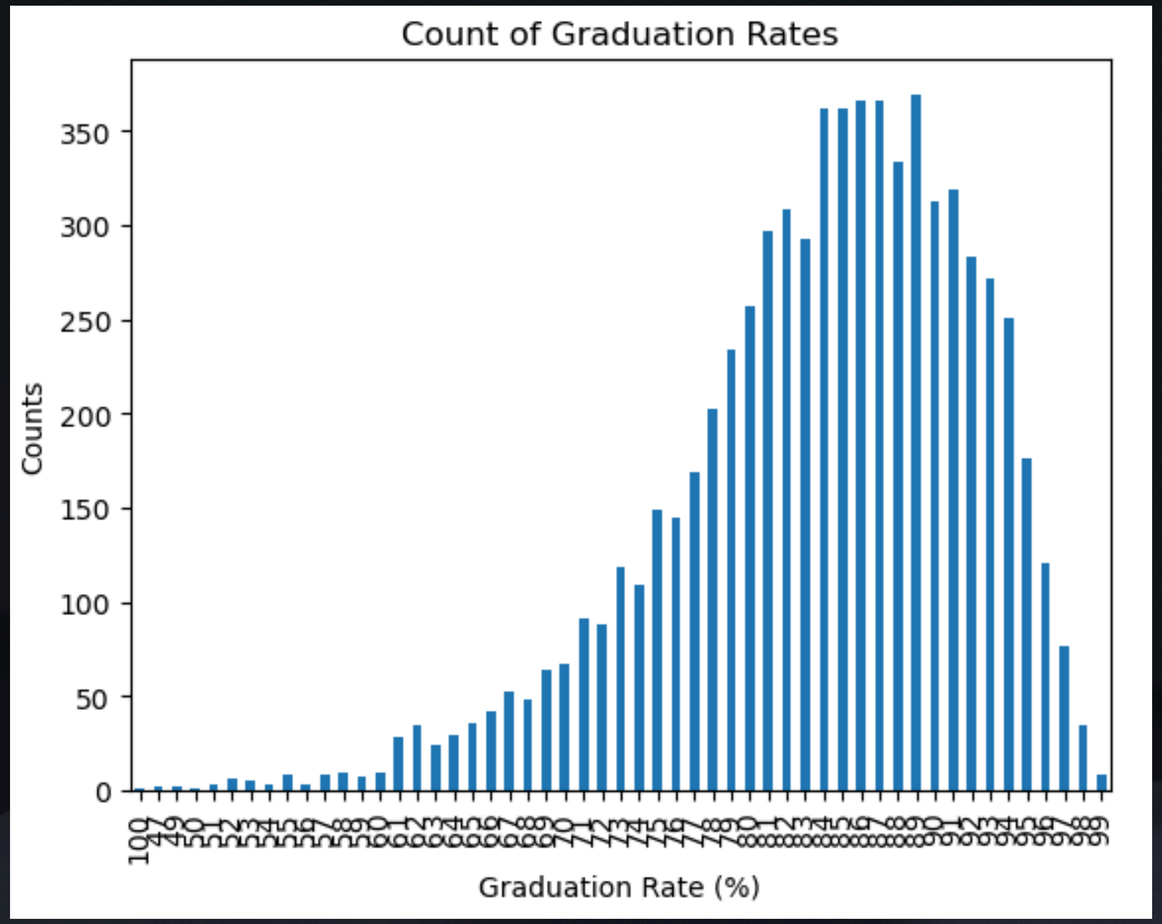
- It is the policy of the BC provincial government to protect the privacy of individuals, by suppressing (**masking**) very small population numbers.
- This means that when attempting to ascertain counts of graduation rates, this happens:



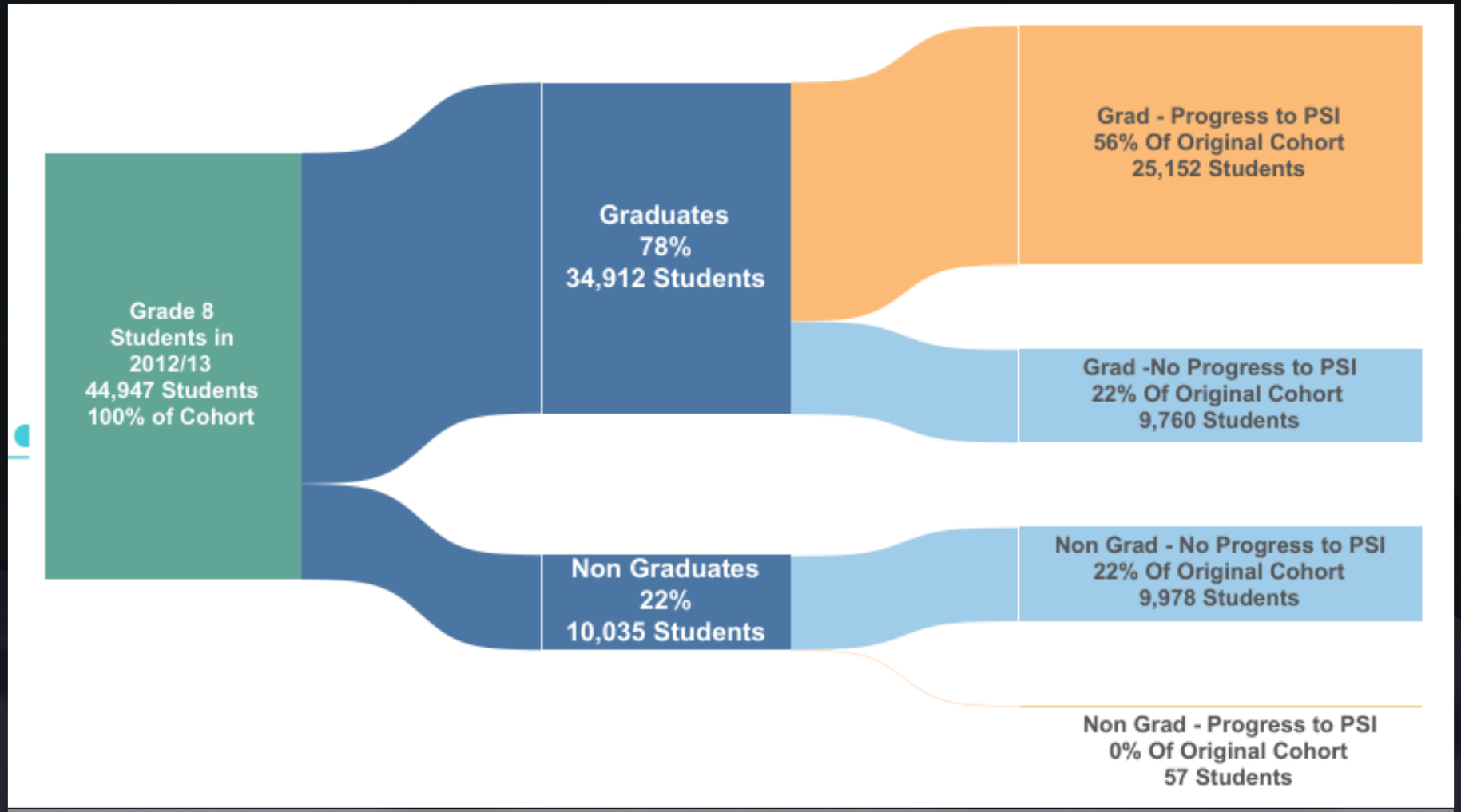
Insights into the data

A masked dilemma

- When masked data are removed, clearer (but fundamentally unreliable) patterns emerge.
- It would be inappropriate to impute so much masked data when it is the overwhelming majority.
- This kind of masked data is present in every dataset I have investigated thus far.



Secondary School Graduates 2016/7



Next Steps

- See if the masked data can be released to me (potentially via a FIPPA request).
- If the data cannot be released, the scope of the project will need to be adjusted.
  - It is possible that other jurisdictions will have different privacy laws.
- When a complete dataset is secured, explore different categorical models to ascertain the most appropriate option.
  - Options include logistic regression, XGBoost (eXtreme Gradient Boosting) and decision trees.