**Project topic**: Improve the student performance in exams

**Data source**: The dataset, Student Performance in Exams, is from <https://www.kaggle.com/spscientist/students-performance-in-exams>

**Goal**

A high school chancellor would like to know how to improve their students’ exam performance in order to make the school better. To achieve this goal, I will use the student performance data which includes the students’ past performance on three different subjects, and three factors, family, personal, and economics. Based on these information from the dataset, I am going to present graph analysis and identify the following questions,

* How do the three factors affect student’s performance?
* How effective is the test preparation course?

This dataset contains 8 columns and 1000 rows. The first question can help me identify whether those factors are related to the performance. The second question can help find out if the existing preparation course is essential for students or the course should be improved so the student will not waste their time on it.

After knowing how the factors affect, the school might change the strategy of teaching or provide another specific course for student in order to have a better learning.

**Graph data model**

(Changed: I separated 3 subjects as nodes instead of stored them as a property in test grade. This way I can do the algorithms in a better way.)

**Diagram

Description automatically generated**

**Data preprocessing**

This dataset includes the student’s test score on three subjects. Adding the letter grade columns on each score can let them easier to compare how the student’s performance in each subject.

**Neo4j database screenshot**

**Graphical user interface, text, application

Description automatically generated**

**Cypher Actions**

1. Know the student's information by id $student\_id

This query can let user(may be the student’s advisor) quickly search the student by id in order to know more details about the student with his/her grades or social status.

Code:

MATCH (s:StudentInfo{student\_id:$student\_id})

WITH s

MATCH (s)-[r1: GET\_MATH\_GRADE]-(m:MathGrade)

MATCH (s)-[r2:GET\_READING\_GRADE]-(read:ReadingGrade)

MATCH (s)-[r3:GET\_WRITING\_GRADE]-(w:WritingGrade)

MATCH (s)-[r4:HAS\_PARENTS\_BACKGROUND]-(p:ParentalLevelOfEducation)

MATCH (s)-[r5:CLASSIFIED\_AS]-(race:Race)

RETURN s,m,r1,r2,r3,r4,r5,read,w,p,race

The example below, I search student ID = 5, the graph shows all the features that related to this student. This action is good for teachers who want to know about his student.

A screenshot of a computer

Description automatically generated

1. Find student who has good math with race group search by their parent's education level $level

This action can get multiple information base on the student who has good performance on math.

Code:

MATCH (s:StudentInfo)-[r1:HAS\_PARENTS\_BACKGROUND]->(p:ParentalLevelOfEducation{level:$level})

WITH s,r1,p

MATCH (race:Race)<-[r2:CLASSIFIED\_AS]-(s)

MATCH (s)-[r3:GET\_MATH\_GRADE]-(m:MathGrade)

WHERE m.grade = "A" OR m.grade = "B"

RETURN race,s,r1,p,r2,r3,m

Diagram

Description automatically generatedFor example, when I search master’s degree, we can easily find that there is no race group A in the image. The purple node on the upper right side, group B, has only one student match with good math grade and high parental education.

**Visualization**

1. Students who share the same reading grade and writing grade. The blue color means male and the yellow means female. In the middle circle, the students share grade F, which contains more male students. Additionally, it looks like if student who failed on one of reading or writing exam, most likely that student will also fail on the other one.

On the bottom left circle, that is grade A with more female students. Therefore, female students have better performance in the reading and writing fields. The school may provide more materials that can help boys improving their reading and writing grade.

Code:  
match (s:StudentInfo)-[r2:GET\_READING\_GRADE]->(r:ReadingGrade)

match (s:StudentInfo)-[r3:GET\_WRITING\_GRADE]->(w:WritingGrade)

where r.grade = w.grade

Return \*

Background pattern

Description automatically generated

1. Students who got all A or all F in three exams, comparing with their test preparation course status. In the image below, on the left side is all A and the right side is all F, with green nodes which means the student who have completed the course. According to this result, we can know that the student who got all A took more effort on the exams. Most of them finished the preparation course. On the other hand, most of the student who failed on all exams did not complete the course. The school may reconsider should they keep this course or adjust the course content. From the right side of graph, looks like there are 1/10 students who have completed the course still failed on all 3 exams.

match (m:MathGrade)<-[r1:GET\_MATH\_GRADE]-(s:StudentInfo)

match (r:ReadingGrade)<-[r2:GET\_READING\_GRADE]-(s:StudentInfo)

match (w:WritingGrade)<-[r3:GET\_WRITING\_GRADE]-(s:StudentInfo)

where (m.grade = 'F' AND r.grade = 'F' AND w.grade = 'F') OR (m.grade = 'A' AND r.grade ='A' AND w.grade = 'A')

Return \*

A picture containing chart

Description automatically generated