Zhenrong Lao

Professor Robinson

DSBA 6520

6/14/2021

Dataset Use Case:

Nowadays, there are so many resources available to students. For example, there are teachers there to help them in and out of class, their peers can give them a hand, tutors are available for parents to hire, and there are so many free online learning materials. However, not every student can do a good job in high school such as obtaining a high score on their exams even if all of them have chances to do it. In this project, I mainly focus on understanding why they should perform great which still is not the case. To better analyze the influence of various factors on the student's performance, I collected student data with 8 variables described as follows:

1) gender: the gender of each student (male/female)

2) race: the race of each student (there are three groups: group A,group B,group C)

3) parental level of education: the highest educational qualification of any parent of each student

4) lunch_type: the type of lunch package selected for each student(standard/reduced)

5) test_prep: if the test preparation course was completed by the student or not

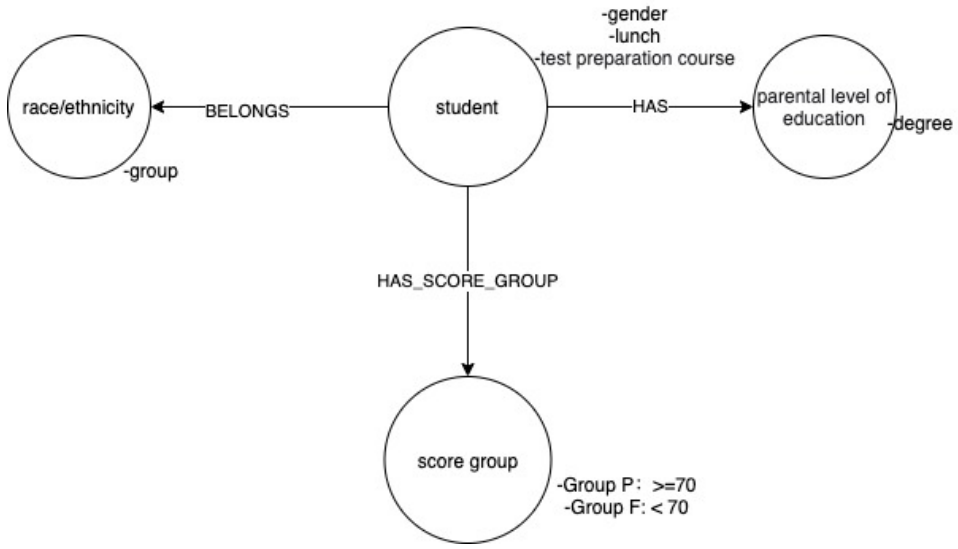
6) math_score: score in math (our target variable)

7) reading_score: score in reading

8) writing_score: score in writing

More specifically, I would like to answer several questions through graph analytics on this dataset: How many students pass the test for each test and how many students fail the test for each test? Is there any similarity for students who pass or fail exams? Is test preparation enough for exam pass? What's the main driver of good test scores? Based on analysis, could we improve the students performance in the tests?

Graph Data Model:



Graph Projections: