URL to Colab:

https://colab.research.google.com/drive/1fO53QZX5bO8uOYzZhnDJtN4Rd-0RjPoJ#scrollTo=3uyah1W0f_X2

Project Explanation:

This Student Performance Evaluation Project is to utilize student personal information including exam scores, gender, ethnicity, test preparation, parental level of education and lunch nutrition information to explore the solutions to improve student math exam performance.

Data Processing

Since most variables, except exam scores, in the dataset are categorical and contain string values, I transformed all the categorical variables to dummy variables.

Data Visualization

Facet Plot: The relationship between student math scores and gender, test preparation, lunch nutrition, race/ethnicity and parents' education level
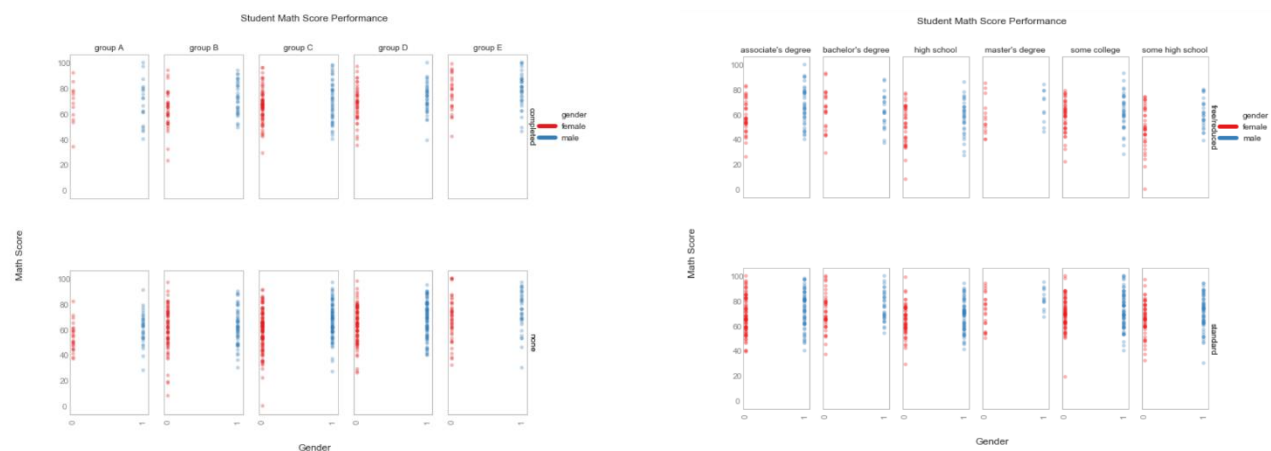
From the above two graphs, we are able to visualize the relationships between student math scores and gender, test preparation, lunch nutrition, parents' education level and race/ethnicity:

- In the sample data, males usually do better in math than females.
- There more students who didn't take the preparation exams than the students who took, while students who took the preparation exams didn't necessary do better than students who didn't.
- Race/ethnicity is not an important factor to determine student math scores.
- Students with standard lunch do better in the math exams than students with free or reduced lunch.
- Students whose parents have higher degree generally do better in math exams, but not necessarily.

Correlation Heatmap: Correlations between each variable prove the conclusions above
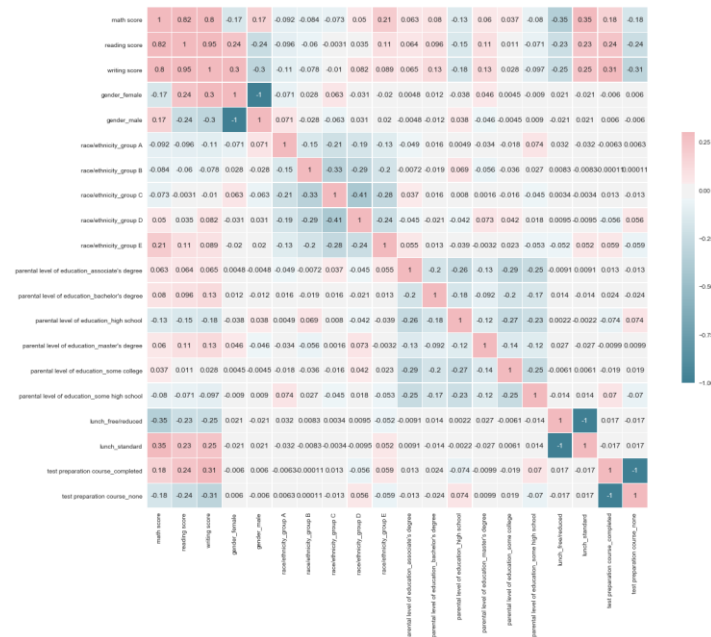
From the above graph, we can see that math score has high positive correlations with male gender, ethnicity group E, and standard lunch; while high negative correlations with female gender and reduced/free lunch.

Data Modeling

After splitting current dataset to train and test datasets, we use simple linear regression to predict students' math scores. The coefficient of standard lunch is 1.57, while the coefficient of reduced lunch is -1.57, which indicate the importance of lunch nutrition to math exam performance. In addition, KNN method is also applied to cluster students. By clustering students into two groups, we can focus on the specific group with homogeneities.
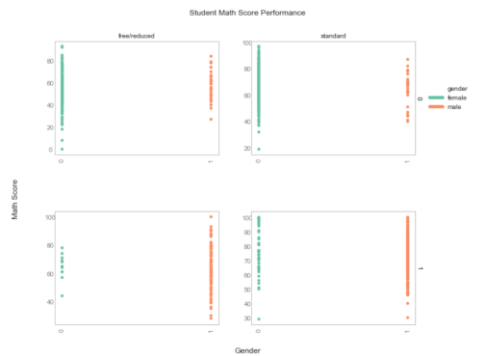
Recommendations to Management:

Since lunch nutrition is an external factor that really determines students' math scores, we suggest the management group to design nutritional lunch for female and male students respectively to improve students' math exam performance. By applying experiments within two clusters with homogeneities, we can better monitor students' math exam scores.