**Data used:**

<https://www.kaggle.com/spscientist/students-performance-in-exams>

**Statistical/Hypothetical Question**

1. Does gender of the student affect the performance of the student in the exams?
2. Does race/ethnicityof the student effect the performance of the student in the exams?
3. How does the parent’s education level effect the performance of student?
4. Does the availability of a subsidized lunch have any effect on the performance of a student?
5. Does taking the test preparation course have any effect on the performance of student in the exam?

**Outcome of your EDA.**

The outcome of the EDA was interesting.

1. For regression analysis math\_score as a function of test\_preparation\_course

we saw that can see the slope and the intercept are statistically significant, which means that they were unlikely to occur by chance, but the R2 value for this model is small. which means that "test preparation course" doesn’t account for a substantial part of the variation in math score.

1. For math\_score as a function of lunch

we saw that the slope and the intercept are statistically significant, which means that they were unlikely to occur by chance, and we see the R2 value for this model also significant. which means that "lunch” accounts for a 12% of the variation in math score.

1. For math\_score as a function of test\_preparation\_course+race\_ethnicity+lunch+parental\_level\_of\_education+gender

we saw that test\_preparation\_course+race\_ethnicity+lunch+parental\_level\_of\_education+gender accounts for 25% (R2 0.255) of the variation in math score

1. For writing\_score as a function of test\_preparation\_course+race\_ethnicity+lunch+parental\_level\_of\_education+gender

test\_preparation\_course+race\_ethnicity+lunch+parental\_level\_of\_education+gender accounts for 33.4% of the variation in writing score

1. For reading\_score as a function of test\_preparation\_course+race\_ethnicity+lunch+parental\_level\_of\_education+gender

test\_preparation\_course+race\_ethnicity+lunch+parental\_level\_of\_education+gender accounts for 22.7% of the variation in reading score

**What do you feel was missed during the analysis?**

Did not feel anything was missed during the analysis.

**Were there any variables you felt could have helped in the analysis?**

The income of the parents, Domestic health could have been other variables that would have helped the analysis more.

**Were there any assumptions made you felt were incorrect?**

One of the assumptions made was that the exam preparation course will have a great effect on the test score of the students, but though there is affect the statistical significance is very less, In fact the availability of a subsidized lunch had more effect on the test scores.

**What challenges did you face, what did you not fully understand?**

One of the challenges was that the variable names were having space and /, so later down I had to rename the variables to overcome some issues.

**Submit a link to your repository to the assignment link during the final week of class.**

<https://github.com/GitTInto/EDA>