**Verzeo Technologies**

****

**Introduction to Machine learning using Python**

**Internship (April-May 2020)**

**Internship Report**

**Name: Anirudh Sai S B**

**Batch: 3**

**Topic: Explanatory data analysis**

**Data set: Student Performance**

**Analysis**

* The Student Performance data set consists of 1000 rows and 8 columns.
* It has a total of 1000 values; that is the data set of a 1000 students.
* The eight parameters are gender, race/ethnicity, parental level of education, test preparation course, math score, reading score and writing score.
* The maximum marks scored by a student is 100 in all the three subjects and the minimum scores in math, reading and writing are 0, 17 and 10 respectively.
* The mean score in math is 66.08900, in reading is 69.169 and in writing is 68.054.
* From the relationship plots, we find that a greater number of male students have scored well in both math and reading rather than their female counter parts.
* But, the majority of male students have done well in math, whereas a majority of female students have done well in reading.
* The same pattern is found in the second relationship plot, except that the only difference here is that the female students have done well in writing instead of reading.
* The third relationship plot shows that both the genders who have performed well in both reading and writing are almost the same.
* The bar graphs show that the a majority of students have scored between 60 to 80 in math, between 65 to 80 in reading and between 65 to 80 in writing.
* Also, we find that almost equal number of students who have secured a 100 in all the three subjects.
* But, the count of students who have scored 20 or less is more in reading when compared to the other subjects.

**Parameter: Gender**

* The violin plots show that the count of female students who have scored between 60 to 80 is greater than their male counter parts in all the three subjects.
* The average score of the male students seems to be greater than that of the female students in all the three subjects.
* We also find that the subject in which the male students have scored well when compared to the other subjects is math.
* Female students have performed equally well in both reading and writing.
* But, we find that the female students are those who scored the least in all the three subjects.

**Parameter: Test preparation course**

* The count plot illustrates that the majority of both the male and female students have not taken up the test preparation course.
* The students who have completed the test preparation course have performed well in all the subjects.

**Parameter: Parental level of education**

* The count plot indicates that the parents of both the male and female students are equally well-educated.
* The children of parents who have earned a master’s degree have excelled in all the three subjects.
* On the contrary, some children of parents who have just earned a high school degree have scored the least in all the three subjects.

**Parameter: Lunch**

* On analysing the count plot, we find that a greater number of students prefer standard lunch as opposed to free/reduced lunch.
* The violin plots show that students who prefer the standard lunch have scored significantly in math than their counter parts. But, this difference is not the same when it comes to the other two subjects. Conclusively, students who take standard lunch have a good average of scores.

**Parameter: Race/ethnicity**

* Race/ethnicity does not have a relevant role in determining a student’s score.

**Inferences**

* The male students seem to be good in math.
* The female students seem to be good in reading and writing.
* The test preparation course helps the students in getting better results.
* The parental level of education does not have a significant influence on their children. Having said that, the students whose parents have earned a master’s degree seem to score well in all the subjects. The

Well-educated parents do help their children in getting better scores.

* Eating standard lunch help the students to score better.
* Race/ethnicity does not have an influence on the scores of the students.