How Outside Factors Affect Student Achievement

The purpose of the research is to examine the relationships between measures of support outside the school and student's achievement score in terms of self-esteem.

BMS (2013) states that parents are considering school standardized test scores as concrete evidence of school quality. Parents pay thousands of dollars to live near better schools (Ferguson, 2015). Rental homes are more expensive nearby A schools. Parents want to see school performance measured. They are looking for concrete evidence of school quality therefore, parents are focused on test scores (BMS, 2013).

However, school quality is not the only factor on student achievement. Many research states that outside factors such as parental involvement, parental education level affect student achievement scores. We aim to find out how the independent variables like number of visits to school by parents or mother's education level affects the dependent variables measured as self esteem score from the data set called "School Data". The data we will use is readily available through the Psych package in R.

The data set is collected by a US federally sponsored program. The program provided remedial assistance to disadvantaged primary school students across the US. The total data is collected from 70 entire school sites The data types included in this data frame "consists of results from three different kind of tests, a reading score, a math score, , and a self–esteem score, , which are considered outputs in the model, and five different independent variables: the education level of the mother, the highest occupation of a family member, number of parental

visits to school, time spent with children in school-related topics, and the number of teachers at the site" (Charnes et al. 1981)

We used MultiLinear regression and KMeans clustering techniques to further our previous analysis on the same data using the Canonical Correlations. We started our analysis by exploring the data and visualizing the data. We ran Mardia Kurtosis and Mardia Skewness tests to check for normality. We realized there was an outlier and removed it because it did not live up to the description of the data. After removing the outlier we ran new visualizations. Next, we ran Connaical Correlations to repeat our previous research to compare our findings with the new knowledge. Secondly, We used Multiple Linear Regression technique to find the best model that fits our data. For that purpose we first separated our data to "train" and "test" data through random sampling. With the training data we found out that the number of teachers in an elementary school has no effect on self esteem and the rest of the independent variables combined have almost 99% correlations through R squared data.. We confirmed our model through Stepwise Linear regression and adapted this as a training model for our data. We ran predictions for the test data and found out very close numbers to our actual data.

In the third step, we wanted to do an unsupervised data analysis to find out what else the data can tell us other than predicting output variables. We first explored the number of clusters through WSS, Silhouette and Gap Stat Methods. Based on our findings we decided on 2 clusters. Then, We ran KMeans analysis and visualized the data. However, this graph did not tell us much about our findings of the KMeans analysis. Next, we ran PCA analysis and visualized the data through the Biplot scatter graph by visualizing the variables on the graph as vectors. All

schools were visualized on the graph individually as a point with the corresponding school number.

We have created another visual to show the cluster graph and Biplot PCA Analysis graph next to each other. This way we could see clusters and vector variables together. Once we ddid that we had these findings:

- We found out that all the variable vectors on the graph fell into one cluster. All variables except teacher variables were very closely located. That means that all independent variables except the teacher variable complement each other.
- Teacher variable was almost perpendicular to all dependent variables which confirms that the teacher variable had almost no effect on the dependent variables.
- By having all dependent variables in the same cluster, we can assume that these are the most successful schools.
- We can also assume that the schools in these clusters have the most benefit of all outside factors in terms of parental visit to school, counselling, occupation and education.
- The other cluster fell into the negative values of these vectors meaning that they had less values for these outside factors along with less values of the output variables making them a cluster of less successful schools in general.

In conclusion, our prediction is that most successful schools are successful not because they provide the best education but because they benefit the most from outside environmental factors which have a direct high correlation with student success.