Report for Data Analysis Assignment – Ghaleb Dweikat

- **DataSet Used:** student-mat

- **Summary of the Data:**

The data has 395 rows and 33 columns

Data columns (total 33 columns):

school 395 Values, No Empty Fields, Categorical

sex 395 Values, No Empty Fields, Categorical

age 395 Values, No Empty Fields, Numerical

address 395 Values, No Empty Fields, Categorical

famsize 395 Values, No Empty Fields, Categorical

Pstatus 395 Values, No Empty Fields, Categorical

Medu 395 Values, No Empty Fields, Numerical

Fedu 395 Values, No Empty Fields, Numerical

Mjob 395 Values, No Empty Fields, Categorical

Fjob 395 Values, No Empty Fields, Categorical

reason 395 Values, No Empty Fields, Categorical

guardian 395 Values, No Empty Fields, Categorical

traveltime 395 Values, No Empty Fields, Numerical

studytime 395 Values, No Empty Fields, Numerical

failures 395 Values, No Empty Fields, Numerical

schoolsup 395 Values, No Empty Fields, Categorical

famsup 395 Values, No Empty Fields, Categorical

paid 395 Values, No Empty Fields, Categorical

activities 395 Values, No Empty Fields, Categorical

nursery 395 Values, No Empty Fields, Categorical

higher 395 Values, No Empty Fields, Categorical

internet 395 Values, No Empty Fields, Categorical

romantic 395 Values, No Empty Fields, Categorical

famrel 395 Values, No Empty Fields, Numerical

freetime 395 Values, No Empty Fields, Numerical

goout 395 Values, No Empty Fields, Numerical

Dalc 395 Values, No Empty Fields, Numerical

Walc 395 Values, No Empty Fields, Numerical

health 395 Values, No Empty Fields, Numerical

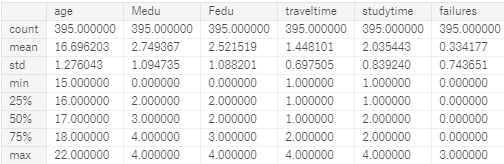
absences 395 Values, No Empty Fields, Numerical

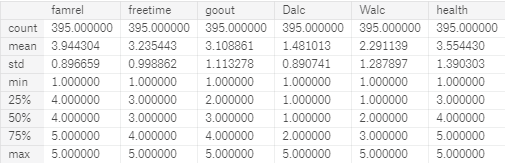
G1 395 Values, No Empty Fields, Numerical

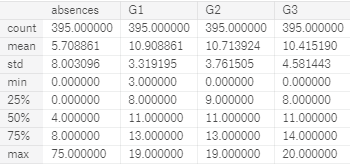
G2 395 Values, No Empty Fields, Numerical

G3 395 Values, No Empty Fields, Numerical

- **Numerical Field Description:**







- **Quality of the Data:** After analyzing the description and summary of the data fields, we can tell that the quality of the data is decent as there aren't any columns with null values and every cell has a single piece of data. This will significanlty simplify the processing stage of the data as we would not be required to compensate for null values or split dynamic data. However, there are many categorical fields in the data set and that requires some additional processing to generate better results from.

- **Most Useful Fields**: Mother Education, Father Education, Study Time, Age, Travel Time, Number of times they go out, number of Failures, Father’s Job, Higher Education Goal and Guardian.

**- Best Fitting Model Results:** The best fitting model has a variance score of 0.79 which is a decent place to start.