Inrodaction :

Data preprocessing and data visualization plays a crucial role in any data science project.  
To begin preprocessing the student grades dataset, we first search for the missing data and attempt to deal with these missing data. We then investigate which columns are irrelevant to our objective and which columns can be combined in order to streamline the process . after preprocessing which is the most essential part to move to the second part, which is visualizing the preprocessed dataset , we used various approaches for different kinds of data such as count plot and boxplot for categorical data and scatter plot for numerical data , data visualization empowers us to look deeper into the data, understand more about the relations between different columns , and interpret the relations with a more clear vision , which leads us to complete the third part on this phase of the project to explore the data and describe it by the graphs and the data statistics , we discover a lot of interesting facts form the relationship between the grades and various attribute on the data set , also we observe a lot of patterns that would help us in building the models in the following steps.

|  |  |  |
| --- | --- | --- |
| Variables | | |
| Variable name | Description | Variable type |
| school | binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira | Categorical |
| sex | binary: 'F' - female or 'M' - male |
| address | tudent's home address type (binary: 'U' - urban or 'R' – rura)l |
| famsize | family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3) |
| Pstatus | parent's cohabitation status (binary: 'T' - living together or 'A' - apart) |
| Medu | mother's education (numeric: 0 - none, 1 – primary education (4th grade), 2 â€“ 5th to 9th grade, 3 â€“ secondary education or 4 â€“ higher education) |
| Fedu |
| Mjob | job ('teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other') |
| Fjob |
| reason | reason to choose this school ( close to 'home', school 'reputation', 'course' preference or 'other') |
| guardian | student's guardian ( 'mother', 'father' or 'other') |
| traveltime | home to school travel time ( 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour) |
| studytime | weekly study time ( 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours) |
| failures | number of past class failures (n if 1<=n<3, else 4) |
| schoolsup | extra educational support (binary: yes or no) |
| famsup | family educational support (binary: yes or no) |
| paid | extra paid classes within the course subject (Math or Portuguese) (binary: yes or no |
| activities | extra-curricular activities (binary: yes or no) |
| nursery | attended nursery school (binary: yes or no) |
| higher | wants to take higher education (binary: yes or no |
| Variables | | |
| Variable name | Description | Variable type |
| internet | Internet access at home (binary: yes or no) Internet access at home (binary: yes or no) | Categorical |
| romantic | with a romantic relationship (binary: yes or no) |
| famrel | quality of family relationships ( from 1 - very bad to 5 - excellent) |
| freetime | free time after school (from 1 - very low to 5 - very high) |
| goout | going out with friends (from 1 - very low to 5 - very high) |
| Dalc | workday alcohol consumption ( from 1 - very low to 5 - very high) |
| Walc | weekend alcohol consumption ( from 1 - very low to 5 - very high) |
| health | current health status (from 1 - very bad to 5 - very good) |
| age | student's age (from 15 to 22) | Numeric |
| absences | number of school absences (from 0 to 93) |
| G1 | first period grade (from 0 to 20) |
| G2 | second period grade (from 0 to 20) |
| G3 | final grade ( from 0 to 20,) |
| avg | Average of the three periods grades |