--------------------------------------------------------------------------------------------------------------------------

Student name: Navjot Singh Virk, Student Number: x13112406

--------------------------------------------------------------------------------------------------------------------------

REPORT OF ANALYSIS

**Students Alcohol Consumption**

*DATASET 1: student-mat.csv*

(Mathematics Students)

*DATASET 2: student-por.csv*

(Portuguese Language Students)

Currently available at:

<https://archive.ics.uci.edu/ml/datasets/STUDENT+ALCOHOL+CONSUMPTION>

--------------------------------------------------------------------------------------------------------------------------

**Introduction**

The aim this report is present analysis results based on a series of analyses carried out on Dataset 1 and Dataset 2 using R language and the programming environment RStudio.

**Formal Description of the Datasets**

**DATASET 1**

The dataset 1 (student-mat.csv) used in this analysis is about Mathematics students in secondary school and contains 33 attributes which was composed by P.Cortez and A. Silva, at University of Minho in Portugal.

**DATASET 2**

The dataset 2 (student-por.csv) used in this analysis is about Portuguese subject students in secondary school and it also contains 33 attributes like Dataset 1 and is composed by the same author’s as Dataset 1.

Below are the attributes of datasets (both the datasets contains same attributes which will be helpful during the analysis and comparing attributes results and draw meaningful results and predictions based on analysis) -

|  |  |
| --- | --- |
| Attribute | Description |
| school | Students School (binary: "GP" - Gabriel Pereira or "MS" - Mousinho da Silveira) |
| sex | Student’s Sex (binary: "F" - female or "M" - male) |
| age | Student’s Age (numeric: from 15 to 22) |
| address | Student's home address type (binary: "U" - urban or "R" - rural) |
| 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 – 5th to 9th grade, 3 – secondary education or 4 – higher education) |
| Fedu | Father’s education (numeric: 0 – 4) |
| Mjob | mother's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at home" or "other") |
| Fjob | father's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at home" or "other") |
| reason | reason to choose this school (nominal: close to "home", school "reputation", "course" preference or "other") |
| guardian | student's guardian (nominal: "mother", "father" or "other") |
| traveltime | home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour) |
| studytime | weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours) |
| failures | number of past class failures (numeric: 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) |
| internet | Internet access at home (binary: yes or no) |
| romantic | with a romantic relationship (binary: yes or no) |
| famrel | quality of family relationships (numeric: from 1 - very bad to 5 - excellent) |
| freetime | free time after school (numeric: from 1 - very low to 5 - very high) |
| goout | going out with friends (numeric: from 1 - very low to 5 - very high) |
| Dalc | workday alcohol consumption (numeric: from 1 - very low to 5 - very high) |
| Walc | weekend alcohol consumption (numeric: from 1 - very low to 5 - very high) |
| health | current health status (numeric: from 1 - very bad to 5 - very good) |
| absences | number of school absences (numeric: from 0 to 93) |
| G1 | first period grade (numeric: from 0 to 20) |
| G2 | second period grade (numeric: from 0 to 20) |
| G3 | final grade (numeric: from 0 to 20) |

**Objective of Analysis**

The objective of the analysis of these 2 datasets is to draw meaningful results and predictions based on the analysis that will be carried out on these datasets. The datasets contains information like students age, study time, travel time, failures and more which will help us understand –

* What is the age group that consumes alcohol the most. (First analysis based on each dataset and then the comparison based on subject)
* Alcohol consumption based on the subject of the student. (Maths or Portuguese)
* Which schools maths students consume more alcohol and also on which days
* Which schools Portuguese students consume more alcohol and also on which days
* Comparison about the alcohol consumption between maths and Portuguese students regardless of school.
* Which schools students consume more alcohol regardless of the subjects they take (as this may help understand problems or issues or reasons why that students of particular school consumed more alcohol then the students of other school.
* Which gender is consuming more alcohol if that is boys or the girls. (As, this will help us understand alcohol consumption based on the gender of the student and their drinking habits).
* What difference does students whose guardians are their parents have from students whose legal guardian is not their biological parents(others).
* Does good or bad family relations have any impact on alcohol consumption.
* Does students that have happy romantic relationships consume less or more alcohol and also in which subject.
* Does going out constitutes to more alcohol consumption if yes then with student of which subject.
* Health is an important factor during study times students can get stressed, depressed and may suffer anxiety and lead to alcohol consumption. It will be interesting to know if health issues are contributing to more consumption this fluid.
* The amount of hours spent on studies by students of each subject which can help understand which subject takes more time and practice and require more study time.
* And the overall, analysis results will help us understand if the maths students are consuming more alcohol then Portuguese class or the other way around based on the difficulty of the subject (as mathematics is much complex subject for an average student than a language subject as seen from the hours spent on average by the students of each subject).

**Data Processing Activities Carried out**

**Final Analysis Results ~~(along with notes and supplementary information)~~**