**Data Science Workshop**

**Group Project**

**# Attributes for both student-mat.csv (Math course) and student-por.csv (Portuguese language course) datasets:**

1 school - student's school (binary: "GP" - Gabriel Pereira or "MS" - Mousinho da Silveira)

2 sex - student's sex (binary: "F" - female or "M" - male)

3 age - student's age (numeric: from 15 to 22)

4 address - student's home address type (binary: "U" - urban or "R" - rural)

5 famsize - family size (binary: "LE3" - less or equal to 3 or "GT3" - greater than 3)

6 Pstatus - parent's cohabitation status (binary: "T" - living together or "A" - apart)

7 Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)

8 Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)

9 Mjob - mother's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at\_home" or "other")

10 Fjob - father's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at\_home" or "other")

11 reason - reason to choose this school (nominal: close to "home", school "reputation", "course" preference or "other")

12 guardian - student's guardian (nominal: "mother", "father" or "other")

13 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)

14 studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)

15 failures - number of past class failures (numeric: n if 1<=n<3, else 4)

16 schoolsup - extra educational support (binary: yes or no)

17 famsup - family educational support (binary: yes or no)

18 paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)

19 activities - extra-curricular activities (binary: yes or no)

20 nursery - attended nursery school (binary: yes or no)

21 higher - wants to take higher education (binary: yes or no)

22 internet - Internet access at home (binary: yes or no)

23 romantic - with a romantic relationship (binary: yes or no)

24 famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)

25 freetime - free time after school (numeric: from 1 - very low to 5 - very high)

26 goout - going out with friends (numeric: from 1 - very low to 5 - very high)

27 Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)

28 Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)

29 health - current health status (numeric: from 1 - very bad to 5 - very good)

30 absences - number of school absences (numeric: from 0 to 93)

# these grades are related with the course subject, Math or Portuguese:

31 G1 - first period grade (numeric: from 0 to 20)

31 G2 - second period grade (numeric: from 0 to 20)

32 G3 - final grade (numeric: from 0 to 20, output target)

**# Dataset**

Portuguese = 650 records

Maths = 396 records

Total = 650 + 396 = 1046 records

Common records = 382

Therefore, actual records = 1046 - 382 = 664

No of fields in both (Portuguese & Maths) = 33 fields

**# Possible Questions**

1. How many girls students mother is having higher education like teacher?

Girls = 187,

Higher: yes= 375, no = 20

1. How many boys mother's job is at home?
2. How many girls are interested in higher education?
3. How many categorize the boys students staying in nearby area of school? (near by time 15 - 30 mins)
4. Find girl students where their parents are in civil services.
5. How many students have selected this school because it is very close to their home?
6. How many students are busy in extra curricular activities? How many girls and boys?
7. How these students are managing their study time? Do cluster as per weekly study time.
8. Number of students having internet facility at home?
9. Students having internet, family’s support in education.
10. How many girls and boys are from rural area where father and mother is educated up to secondary education.
11. For GP school
    1. How many students between 15 to 18 years of age of either mother or father in teaching profession are successful
    2. Compare with number of students above 18 years of parents in teaching who are successful.

13. In above replace teaching profession with mother’s or father’s education = none or up to fourth grade.

14. Impact of internet access, family relation, study time, travel time.

15. How is the performance of the students who want to pursue studies?

16. How does having ample or less free time affect the performance of the student?

17. List schools having students whose address is Rural with their travel time.

List schools having students whose address is Urban with their travel time.

Which school is having maximum count of students under address category and maximum travel time.

18. Based on travel time of Urban and Rural students, display study time and failure count.

What are your observations.

19. List students with absences greater than 75 %, their travel time, health status. What are your observations.

20. Plot graph of number of male and number of female students.

21. Plot graph for students whose health is in range 1 & 2 and compare with students whose health range is between 3 to 5.

22. Plot graph for students who want to pursue higher education with those who do not want to.