**DATA VISUATISATION**

*Support : http://uu.diva-portal.org/smash/get/diva2:1464106/FULLTEXT01.pdf*

Problématique : Peut-on réellement accomplir ce que l'on souhaite ? L'expression "Qui veut peut" résonne souvent dans nos vies, que ce soit en tant que mantra personnel pour se motiver ou en tant que conseil de nos enseignants et parents, soulignant que la persévérance et le travail acharné mènent à la réussite. Vous avez certainement déjà eu des conversations similaires :

Professeur : Tu sais, si tu te donnais plus de mal, tu aurais obtenu une meilleure note ! Qui veut peut !

Pourtant, se demande-t-on toujours si cette maxime reste universellement valable ? C'est précisément la question que nous nous efforcerons d'explorer dans cette étude. Nous nous appuierons sur base de données comprenant 396 étudiants issus de deux universités portugaises, tous étudiant les mathématiques. Notre objectif est d'analyser les multiples variables socio-démographiqeus telles que liées à leur identité, leur parcours scolaire, leur contexte familial, leur environnement et leurs activités extracurriculaires. Nous chercherons à déterminer dans quelle mesure la motivation et le travail acharné sont véritablement des facteurs déterminants pour atteindre le succès, en particulier académique.

Nous ferons donc d’abord une étude de data visualisation qui nous permettre d’apporter des réponses à cette problématique puis régression logistique permettant de prédire la note en fonction des différentes variables ( les + significatives ).

Description base de données

* school - student's 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 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
* 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, output target)

Variables lié à l’étudiant  
situation familiale   
Scolarité avant la fac   
FAC   
Activités extrascolaire   
environnement de l'étudiant

Faire d’abord comme les pages 7 à 10 ( étude général ), regarder aussi page 22

Attention on veut regarder si les variables sociodémographiques ont un impact sur les notes ( ne pas perdre de vu cet objectif)

Plan :

Introduction ( amélie 2 semaines )

* Présentation du projet
* Introduction
* Problématique -> objectif ( voir si variable socio demo ont une influence sur les étudiants)
* Prestation base de données avec graphiques

Visualisation des données ( lundi)

* Utiliser la matrice de corrélation pour regarder les liens et faire des graphiques
* Et faire de manière thématique (l’école / métier de la mere / enviornnement familiale support / cours particulier / going out et comparer à ceux qui ont des bonnes et mauavises notes )

Modèle de regression logistic pour choisir les variables les plus explicatives ( yassmine )

INTRODUCTION

On peut regarder les proportions pour chacune des variables. Quelques commentaires :

The mother is usually the guardian and in the majority of families the family relations are good or excellent.

• The majority of the parents have an education level above 9 th grade. The proportion of higher education is larger among mothers. It's more common among mothers to work as a teacher, to stay at home or to work in the health category compared to the fathers. Other and services are the most popular working categories for both mothers and fathers.

• Approximately 50% of the students study between 2 and 5 hours per week. Approximately 25% of the student study less than 2 hours. The rest study more than 5 hours per week. Approximately 80% of the students don't have any failures.

• Most of the students have a low travel time to school. The most popular reason for choosing one of the schools is for the courses the school oers.

• A majority of the students consider their free time to be at a medium to high level. The mode for going out with friends is at a medium level.

• The weekend alcohol consumption is higher than the weekday alcohol consumption. During the weekdays a majority of the students have very low alcohol consumption but at the weekend the very low alcohol consumption is a minority.

• The majority of the students have medium to very good health.

The numbers of absences are the number of lessons the student have missed. Most of the students have a low number of absences but there are also some outliers with a very high number of absences. The majority of the students were 15 to 18 years old (mainly secondary students but with some repeating students of 19 to 22 years of age).