

Using Python to Analyze Secondary School Student Alcohol Consumption and Their Academic Performance

¹Poonam Kumari and ²Aditya Pratap

¹Research Scholar, Department of Computer Science, IITM Janakpuri, New Delhi, India

²Research Scholar, Department of Computer Science, IITM Janakpuri, New Delhi, India
poonam.kumari561999@gmail.com, aditya1998pratap@gmail.com

Abstract

In this project, we use a data set about Portuguese student on two courses (Mathematics and Portuguese) which was collected in University of Minho, Portugal. Our work intends to approach student addiction on alcohol in secondary level using Python. The result shows that a good predictive accuracy can be achieved, provided that addiction of alcohol can impact to the student performance. In addition, the result also provides the correlation between alcohol usage and the social, gender and study time attributes for each student. As a direct outcome of our project, more efficient solutions can be developed in order to pay more attention to the student and share how the alcohol impact so badly in their life.

Keywords: Python, Portuguese, NGO

1. Introduction

Alcohol had lots of bad impact in our life. Drinking too much can take a serious toll on our health. Alcohol has many short-term and long-term health effects. Taking alcohol as teenager age, reduce a child's mental and physical abilities, affecting judgment and co-ordination which can lead to trouble. The level of alcohol gets so high that the brain's vital functions, which include breathing control, are blocked. Alcoholics were more likely to get injured or have accidents than non-drinkers. When children drink, their decision making skills are affected and they're more likely to take big risks. While excessive drinking by adolescents is a problem in its own right, it is at times linked to other harmful behaviors like taking illicit drugs. Underage drinkers are more likely to suffer from a range of health issues including major weight gain or weight loss, bad skin, disturbed sleep, headaches. During childhood and teenage years, the brain is still developing. Adding alcohol to that process is asking for trouble. It can affect memory function, reactions, learning ability and attention span all especially important during their school years. Drinking could affect child's

performance at school and prevent them from reaching their full potential. Every parent wants their child to make the best of themselves and performing well at school plays a big part in that. The stats show underage drinking makes that less likely. Children who start to drink by age 13 are more likely to go on to have worse grades, to skip school and, in the worst case scenario, to be excluded from school. They have less self-control and their brains struggle to recognize warning signs. This can lead to aggression and fights. Their risk of being involved in violence and serious vandalism increases directly in line with alcohol consumption, which could lead to arrest and a criminal record.

Objective

- Understand different parameters impacting performance of students and how students are progressing while dealing with these parameters in their life.
- Correlation between features
- How alcohol consumption is correlated to academic performance.

- How family influences students' alcohol consumption
- How your desire for higher education correlates with alcohol consumption
- Does alcohol have an effect on success?

This will be helpful for:

1. NGOs or other bodies working for the betterment of the troubled youths. They can accordingly arrange awareness campaigns, different programmes, etc.
2. Victim children by letting them know about this dangerous trend and its adverse effects.
3. Parents or guardians – “Families are the most central and have enduring influence on children’s lives regardless of their education, composition, income, or values”. We would like to analyse how family shapes student’s behaviour and what steps they can take for the betterment of their children.
4. Government Officials can start programmes, schemes to prevent the country’s treasures. With the help of Government, we would be better able to know about the status of the problem at hand.

2. Dataset

Data collection is the process of gathering and measuring information on targeted variables in an established systematic fashion, which then enables one to answer relevant questions and evaluate outcomes.

Here we use a data set about Portuguese student which was composed by Paulo Cortez and Alice Silva, University of Minho, Portugal. This study will consider data collected during the 2005-2006 school year from two public schools, from the Alentejo region of Portugal. Hence, the database was built from two sources: school reports, based on paper sheets and including few attributes and questionnaires, used to complement the previous information. They designed the latter with closed questions related to several demographic (e.g. mother’s education, family income), social/emotional (e.g. alcohol consumption) and school related (e.g. number of past class failures) variables that were expected to affect student performance. Finally, the data was integrated into two data-sets related to

Mathematics (with 395 examples) and the Portuguese language (649 records) classes. During the preprocessing stage, some features were discarded due to the lack of discriminative value. For instance, few respondents did not answer about their family income (probably due to privacy issues), while almost 100 % of the students live with their parents and have a personal computer at home. The remaining attributes are shown in Table 1, where the last four rows denote the variables taken from the school reports.

3. Data Preprocessing

Real-world databases are highly susceptible to noisy, missing and inconsistent data due to their typically huge size (often several gigabytes or more) and their likely origin from multiple, heterogeneous sources. Low-quality data will lead to low-quality results. We use several data pre-processing techniques: Perform data cleaning to remove noise and correct inconsistencies in data and merge two different data-set by using data integration. There are many possible reasons for inaccurate data (i.e., having incorrect attribute values). The data collection instruments used may be faulty. There may have been human or computer errors occurring while entering data. Users may purposely submit incorrect data values for mandatory fields when they do not wish to submit personal information (e.g., by choosing the default value "January 1" displayed for birthday). This is known as disguised missing data. Our goal is finding alcohol consumption by secondary school student. In this data-set, there are two different attribute on alcohol. The former is alcohol taking in work day (Dalc) and the latter is alcohol taking in weekend (walc). Only one index can be predicted, so we made an attribute that represents the total alcohol taking by a specific student in a whole week. So we merge those two attribute using the following equation 1

$$\mathbf{final['Alc']} = \mathbf{final['Dalc']} + \mathbf{final['Walc']} \quad (1)$$

The new attribute changes between one and ten (from 1-very less to 10-very high).

Similarly for attributes mother’s education (Medu) and father’s education (Fedu) we made a new attribute parent’s education (maxparent_edu) using the following method.

def max_parenteducation(dataframe):

```
    return(max(dataframe['Medu'],
                dataframe['Fedu']))
```

`final['maxparent_edu']=final.apply(lambda row: max_parenteducation(row), axis = 1)`

Attribute	Description (Domain)
Sex	student's sex (binary: 'F' - female or 'M' - male)
Age	student's age (numeric: from 15 to 22)
School	student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
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: from 0 to 4)
Fedu	father's education (numeric: from 0 to 4)
Mjob	mother's job (nominal: 'teacher', 'health' care related, civil 'services', 'at_home' or 'other')
Fjob	father's job (nominal: 'teacher', 'health' care related, civil 'services', 'at_home' or 'other')
reason	reason to choose this school (close to home, school reputation, course preference or other)
guardian	student's guardian (categorical: mother, father or 'other')
traveltime	home to school travel time (0-< 15 mins., 1-15 to 30 mins, 2-30 mins to 1 hour or 3 -> 1 hour)
studytime	weekly study time (ordinal: 0-< 2 hours, 1- 2 to 5 hours, 2-5 to 10 hours or 3 -> 10 hours)
failures	number of past class failures (ordinal: n if $1 \leq n < 3$, else 4)
schoolsup	extra educational school support (binary: yes or no)
famsup	family educational support (binary: yes or no)
activities	extra-curricular activities (binary: yes or no)
paid	extra paid classes (binary: yes or no)
Internet	Internet access at home (binary: yes or no)
nursery	attended nursery school (binary: yes or no)
higher	wants to take higher education (binary: yes or no)
romantic	with a romantic relationship (binary: yes or no)
famrel	quality of family relationships (ordinal: from 1 - very bad to 5 - excellent)
freetime	free time after school (ordinal: from 1- very low to 5- very high)
goout	going out with friends (ordinal: from 1- very low to 5- very high)
Dalc	weekday alcohol consumption (ordinal: from 1 - very low to 5 - very high)
Walc	weekend alcohol consumption (ordinal: from 1 - very low to 5 - very high)
health	current health status
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 period grade (numeric: from 0 to 20)

Table 1: Attributes of Dataset

The new attribute changes between one and five (from 0-'None', 1-'Primary education', 2-'5th to 9th grade', 3-'secondary education', 4-'higher education').

For grade attributes G1, G2 and G3 we made a new single attribute average grade (AvgGrade) using the following source code written in Python.

```
final['AvgGrade'] = final[['G1', 'G2', 'G3']].mean(axis=1)(2)
```

4. Analysis Results

After completing the data preprocessing, we perform some analysis and here are some results

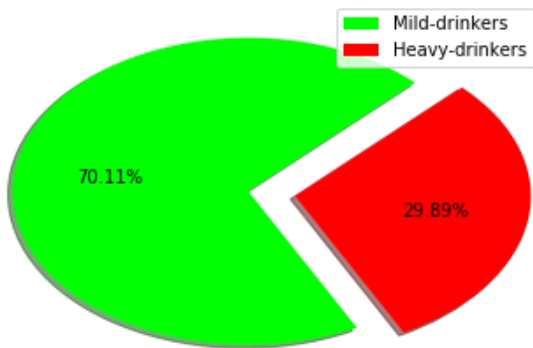


Figure 1: Percentage of Mild-drinkers and Heavy-drinkers

Figure 1 tells us about the percentage of Mild-drinkers and Heavy-drinkers, i.e., those who ingest small quantity and those who ingest large quantity of alcohol. It is clear that the number of people who do not consume much alcohol is way larger. Nearly 70.11% of the students are mild drinkers whereas only 29.89% of the students are heavy drinkers.

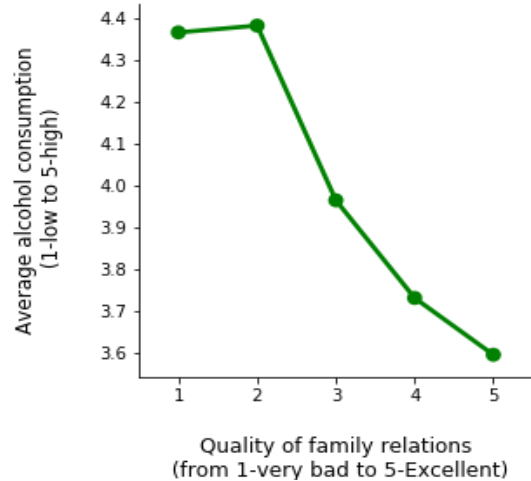


Figure 2: Quality of family relations v/s Average alcohol consumption

From Figure 2, it is clear that quality of family relations is a key factor in determining the average rate of alcohol consumption. It can be said that those families where the quality of family relations is good, students' tendency to indulge in alcoholic activities is less compared to those families where the relations are not very good.

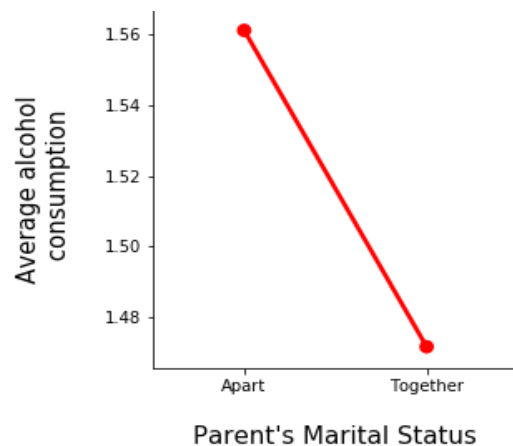
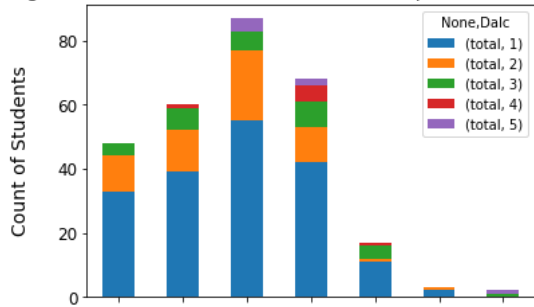


Figure 3: Parent's Marital Status v/s Average alcohol consumption

Figure 3 represents that average consumption of alcohol is very high where parents are not together or are divorced compared to where parents' marriage is a success and they are together. Hence, it is a key factor in determining the indulgence of a child in alcoholic activities.

age wise distribution of alcohol consumption in rural areas



age wise distribution of alcohol consumption in urban areas

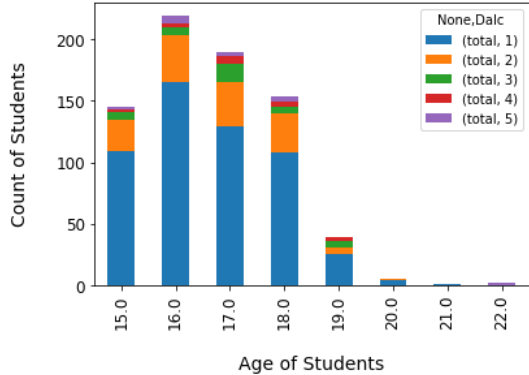


Figure 4: Age wise distribution of alcohol consumption in rural and urban areas

These graphs show the age of students' v/s count of students who are active drinkers in rural and urban areas. For example, it is clear from Figure that, most drinkers are of the age 16 and 17 in the rural and urban areas. Hence, it is the most crucial time to keep an eye on the behavior of students and to help them out of their crisis. It also shows that there is a drastic decline in the number of drinkers when they turn out to be 19. Thus, it can be safely concluded that as the children grow to be 19, they become much more mature.

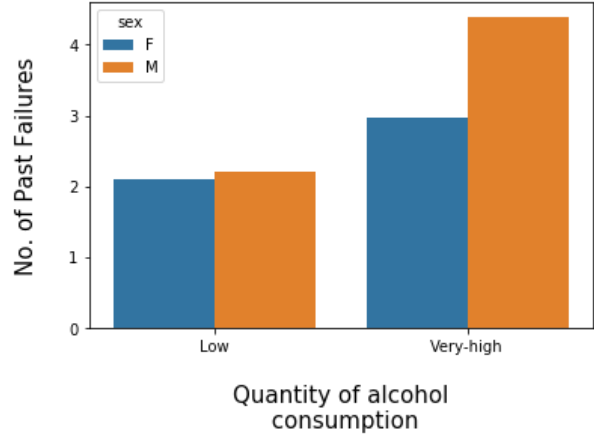


Figure 5: Amount of alcohol consumption v/s Failures

In figure 5 we can clearly see that, those who consume high quantity of alcohol have faced more failures in their life than those who consume less quantity. Among the alcoholics too, it seems that males have suffered failures more than females. So, it can be safely concluded that alcohol consumption leads to more failures in life. And therefore, alcohol should be avoided in order to succeed in life.

Above all visualization represents the contributing factors of indulgence of teen agers in the alcoholic activities. Here are some results which show us the factors which affect students' academic performance.

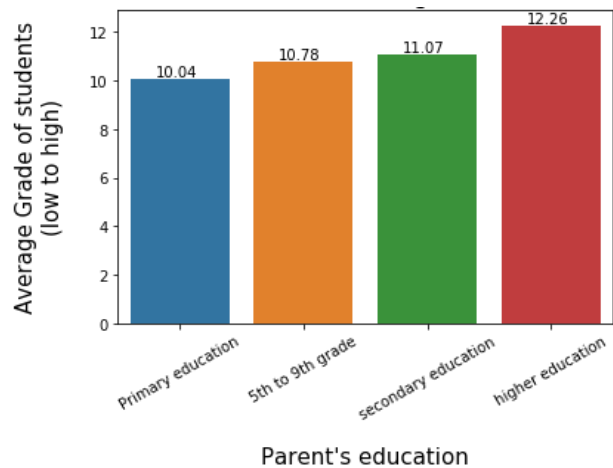


Figure 6: Parent's Education v/s Average Grade of students

From the figure 6, it is clear that this factor "Maximum Education of Parents" is a crucial factor in determining the average grades obtained by the students. It can be said that those families where the

parents are highly educated, average grades of students are higher compared to those families where the parents are not very educated.

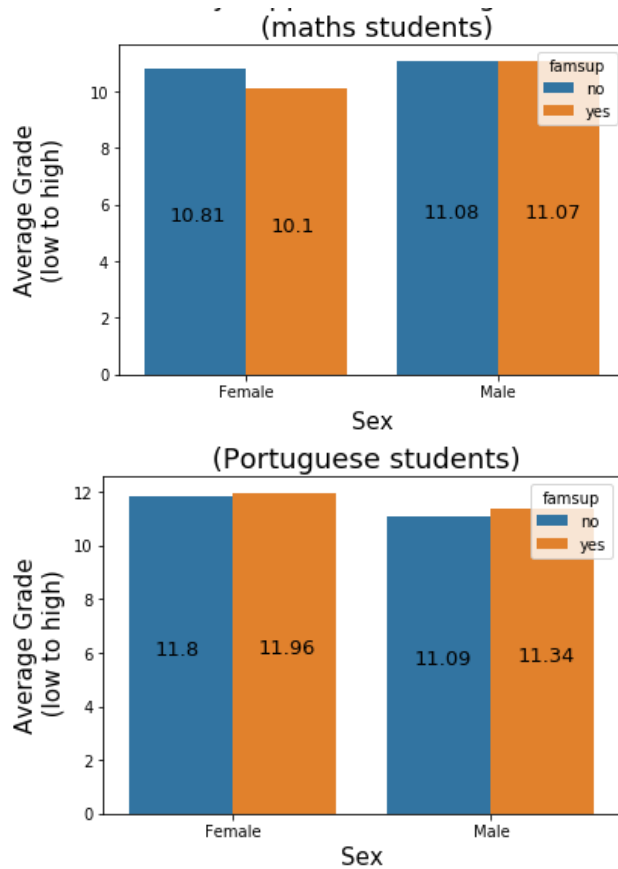


Figure 7: Family support v/s Average Grade

It is clear from Figure 7 that, in case of male students of Portuguese language, if there is lack of availability of family support, they tend to perform poorer compare to where there is enough family support. Hence, it is a crucial factor determining the grades and hence, the future of children and should be kept in mind. It also shows that females of Portuguese language are comparatively less affected by the lack of availability of family support.

However, this result is reversed in case of the students of mathematics. It shows that male students respond better than female students if there is less family support. Furthermore, the male students are not much affected by the availability or unavailability of their family's support.

5. Conclusion and Future Work

Through this study we are acknowledged with the significant factors that contribute in the indulgence of teenagers in alcoholic activities and affect their academic performance in secondary schools.

Age, sex and home address was identified as the most significant factors in consuming alcohol by students. At the age of 15-21 they want to prove to their friends and to the society that they are no longer kids. But as the children grow little more, they become much more mature and are likely to perform better in their lives too. The home environment is also a primary socialization agent, which affects students' life outside the school, the interest in school as well as the aspirations for the future. Home environment includes parental socio-economic status, parental education background, parental marital status and the quality of family relations, etc. Quality of family relations is one of a key factor. Students from broken homes suffer psychological effects while in school and this affects their academic performance. Therefore, family support is a crucial factor in determining the grades and hence, the future of children and should be kept in mind.

It is also identified that those who consume high quantity of alcohol have faced more failures in their life than those who consume less quantity. So, it can be safely concluded that alcohol consumption leads to more failures in life. And therefore, alcohol should be avoided in order to succeed in life and it is high time to handle all these which can be done with the effort of school, family and students themselves.

In future studies, the dataset can be collected from different types of secondary schools (i.e. private/public on multiple locations (i.e. urban/suburban/rural). The questionnaire can be designed to be more detailed and subject oriented. Future research can focus on the analysis of the factors that influence student alcohol consumption and intention of higher education. Rather than selecting random factors, the investigation can focus on the impact of student's psychology, emotion and mental health.

References

[1]2018. [Online]. Available: <https://www.kaggle.com/uciml/student-alcohol-consumption>. [Accessed: 09- Jul- 2018].

[2]2018. [Online]. Available: <https://www.udemy.com/learning-python-for-data-analysis-and-visualization/>. [Accessed: 10- Jul- 2018].

[3]"Data analysis", En.wikipedia.org, 2018. [Online]. Available: https://en.wikipedia.org/wiki/Data_analysis. [Accessed: 12- Jul- 2018].

[4]"Understand why children drink alcohol", *Drinkaware.co.uk*, 2018. [Online]. Available: <https://www.drinkaware.co.uk/advice/underage-drinking/understand-why-children-drink-alcohol/>. [Accessed: 13- Jul- 2018].

[5]"New Python Tutorial: Diagnose data for cleaning", YouTube, 2017. [Online]. Available: <https://www.youtube.com/watch?v=hDtQZWL-3uk>. [Accessed: 14- Jul- 2018].

[6]"Exploring the Factors that Affect Secondary Student's Mathematics and Portuguese Performance in Portugal", *Arrow.dit.ie*, 2017. [Online]. Available: <https://arrow.dit.ie/cgi/viewcontent.cgi?article=1103&context=scschcomdis>. [Accessed: 12- Jul- 2018].