Smoking, Drinking and Drug Use in adolescence in the UK: A Visual Exploration

Baibhav Datta
Department of Computer Science
City, University of London

Abstract—This is a report analysing the data from the surveys carried out by NHS, aimed to provide national estimates of the proportion of secondary school children aged 11-15 who smoke, drink or use drugs and to describe their substance use behaviour. The survey data represents certain behavioural patterns of smoking, drinking or using drugs among the children, sources that illegally provide these substances to children, popular influences to children, causes for smoking/drinking/using drugs as the children think, whether tried to quit, family attitudes, influences from surroundings, and many other relevant information appropriate to analyse the behavioural pattern and influences of young people.

1. Problem Statement

Regular smoking, drinking and drug use among young people has both immediate and long-term health consequences. People who establish smoking habits in early adolescence increase their risk of having respiratory illness or cancer. They are also more likely to experiment other drugs. The use of illegal substances by under age people are increasing day by day, which is a major problem for the generations to come. In order to address the issue, we aim to expose the severity of the ongoing process of intoxication among young people.

We analyse the data looking to uncover any trends influencing young children to smoke, drink or use drugs, and various other factors, such as parental attitude, knowledge and peer influences. Next we shift our attention to how well the children are informed about the negative aspects of smoking and drinking habits. We also aim to get an estimate of what proportions of the younger generation are influenced to drink or smoke, the causes behind them, and what proportion of them are addicted to it, or are beginners. What are the most popular sources which gets illegal substances to children.Throughout the paper, we will highlight how the visualisation techniques has helped refine our findings.

2. State of the Art

The three relatable papers I have analysed are-

- Mark G. Myers and John F. Kelly on "Ciggarette Smoking Among Adolescents With Alcohol and Other Drug Use Problems"
- Emla Fitzsimons and Aase Villadsen on "Substance Use and antisocial behaviour in adolescence"
- Patrick McC Miller and Martin Plant on "Drinking, smoking, and illicit drug use among 15 and 15 year olds in the United Kingdom"

In the first paper, the authors begin the report with the concern that cigarette and alcohol use often develop concurrently, and smoking is especially common among youth treated for alcohol and other drug use disorders. They also point out that studies examining the origins of alcohol and other drug use problems, consistently find that cigarette smoking is closely related with other drug use. Although researchers have identified a broad range of factors that influence the initiation and progression of tobacco and drug use, peer influences are particularly salient. Ethnic and regional variations exist in the development of psychoactive substance use, yet adolescents typically start using cigarettes or alcohol prior to other drugs (Ellickson et al. 1992). In the paper, they have visually represented bar graphs which indicated that approximately one-third of current drinkers smoked, whereas approximately 95% of current smokers used alcohol.

In the second paper, the authors have reported about the relation between substance use and antisocial behaviour in adolescence, analysing survey data from nearly 10,000 adolescent individuals across the UK. They have used visualisation techniques to show bar graphs depicting proportions of people's experiences with different substances among different age groups, compared substance use by gender and ethnicity. Some of the prime findings stated in the report suggest that males were more likely than females to binge drink, use cannabis and take harder drugs. Stark differences were observed in rates of substance use between ethnic groups, with young people of white ethnicity reporting much higher rates of experimentation and more regular use than ethnic minorities. The findings show a sharp increase in the use of substances compared to earlier in adolescence, but little change in antisocial behaviours. Results highlight that sex is a significant factor with males engaging more in risky behaviours.

In the third paper, the authors have reported about examined patterns of self reported drinking, smoking and illicit drug use among 7722 pupils aged 15 and 16. They have found that almost all the pupils had drunk alcohol, 36% had smoked cigarettes in the past 30 days, and 42.3% had at some time used illicit drugs, mainly cannabis. Higher levels of smoking were associated with poorer school performance with below average performance. Levels of drug use in 15 and 16 year olds in 1995 were higher in Scotland than in England, Wales, or Northern Ireland. Their conclusion stated that drug experimentation was high among 15 and 16 year olds, and use of cannabis was particularly high among smokers. Cigarette smoking was more common among girls than boys.

3. Properties of the Data

I have accessed the data from UK Data Service, sponsored by NHS Digital and collected by Ipsos MORI.

The data was collected in the mode of a survey of secondary school children in the UK. The survey was conducted in 2021 in schools by asking pre-selected group of pupils to complete a confidential questionnaire. This was done in a classroom setting, supervised by interviewers from Ipsos MORI. The survey population is pupils in school years 7 to 11 in the UK. Therefore, those taking part are mainly aged between 11 and 15. Any children who were already 16 by the time they were surveyed have been recoded as age 15 as have the handful of 17 year olds who took part. Similarly, the few 10 year olds who took part have been recoded as age 11.

The raw survey data was taken directly from their usual data processing system and because it was possible for any question to have more than one response, each response option from each question was output as a set of binary variables. The variables were named to reflect the naming used in previous year's survey and consisted of the variable name as the root of each variable and a numeric suffix sequentially identifying each response option plus an additional variable to indicate no response.

The resultant dataset consisted of 739 variables, corresponding to different features reflected from the survey questions. In this analysis, we will be using only 70 variables, which were meticulously selected relevant to the purpose of reflecting on our problem statement. The selected variables included some basic features like the pupil's age, gender, region, etc., sets of features extracted from questions related to smoking, alcohol, and drug use. The values of the variables are numeric, with varying ranges like 11-15 for age, 1-9 for region, etc. where each value has labels depicting the meaning of the value, which are provided along with the dataset in a separate data dictionary file. After initial investigations of the dataset, I've found no missing values or any other errors in the dataset.

On inspecting the distribution of the dataset among genders, region and ethnic groups, it is found that the dataset is quite normally distributed according to gender and region but not so for ethnicity. For gender, it contains survey data from 4587 boys, 4531 girls, and 170 of those who hasn't answered. The distribution plots for region and ethnic groups are shown below, which were aided by python function distplot of the seaborn library, which is particularly a data visualization library based on matplotlib.

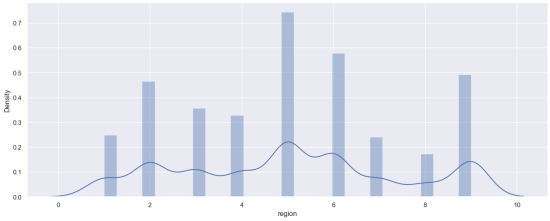

Figure 1: Distribution by region

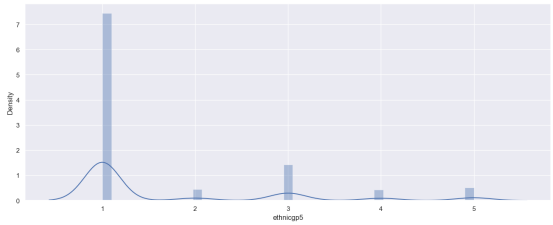

Figure 2: Distribution by ethnic group

4. ANALYSIS

4.1. Approach

In this section, the steps carried out to answer the research questions are discussed. I have used Python for visualisations as it offers several plotting libraries, namely Matplotlib, Seaborn and many other such data visualisation packages with different features for creating informative, customised, and appealing plots to present data in the most simple and effective way.

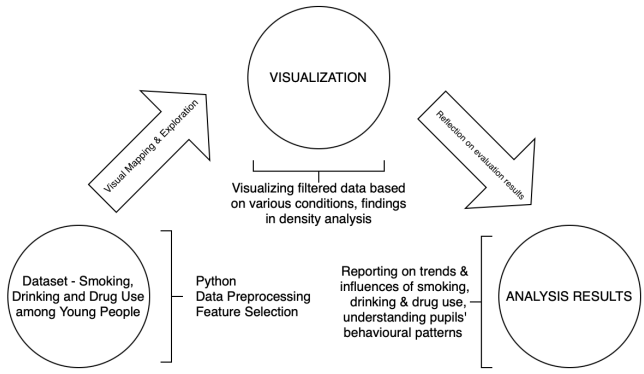

Figure 3: Analysis Workflow Diagram

For density analysis, I have compared the bar graphs of aggregate level results- total number of pupils belonging to certain groups such as smokers, alcohol users or drugs users, groups of people getting alcohol for different sources, groups of people stating their thought for reasons to drink, how many of them have tried to quit or not, how many have friends, parents or siblings around them who smoke or drink (for analysing influence), number of users of drugs among 12 types, number of pupils who have ever tried a drug or not.

Next, I have depicted the influence of alcohol or cigarettes towards using drugs by plotting histograms showing proportions of pupils who use drugs or not among drinkers and non drinkers, and similarly for smokers and non-smokers.

Comprehensively, the visual analysis is done based on numerous factors, as follows-

- Gender- density of each gender with respect to their habits of smoking, drinking or drug use.
- Ethnic Group- density of each ethnic groups involving in smoking, drinking or drug use.
- Age- density of pupils for each age ranging from 11 to 15 involving in smoking, drinking or drug use. The ages when they start taking these substances.
- Source of alcohol- density of each source of alcohol from where the pupils under age are acquiring alcohol.
- Reason for drinking- Among various reasons, pupils have selected the ones that aligns with their point of view. We'll visually observe the most popular choice of reason among all the different reasons.
- Reason for smoking- density of each option of reason for smoking, to observe the most popular reason.
- Types of drugs- density of drug users among 12 different kinds of drugs considered in the survey.

• Proportions of drug users among pupils who drink and who do not.

• Proportions of drug users among pupils who smoke and who do not.

4.2. Process

In this section we carry out all the analysis tasks mentioned in the previous section.

Firstly, for comparing the ages of the pupils when they first started smoking, drinking and using drugs, I have used the distplot function of python library Seaborn, to get the frequency distribution plot, which depicts how the data density is distributed among the ages. In Figure 4, the graph shows distributions of starting age for each category, namely alcohol, cigarette and drugs. Based on the graph, it can be inferred that 13 is the most popular age for starting to use alcohol and cigarette as the curve reaches its peak, while the curve reaches its peak at 14 for drugs, meaning mostly pupils start to use drugs around the age of 14. Another striking detail of this graph is that even children of age starting from 5 have been exposed to alcohol, smoking and drugs, be it very few in numbers.

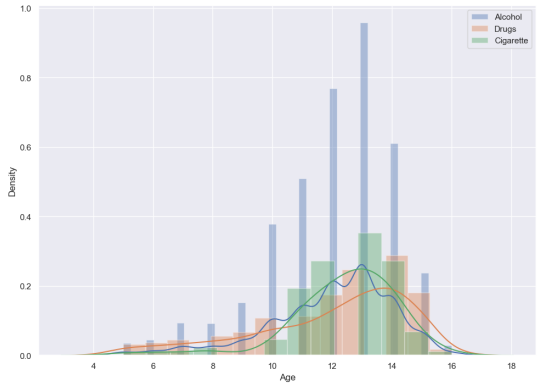

Figure 4

In Figure 5, I've shown a bar plot depicting the density for each source of alcohol from where the under age pupils get it from. Based on the graph, we can clearly see how the bar for parents as source of alcohol is leading, followed by friends.While figure 6 shows density for each source of cigarettes, where we can see that the most popular source of cigarettes are friends.

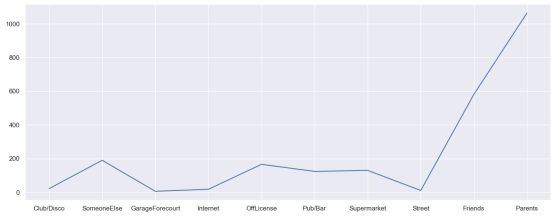

Figure 5

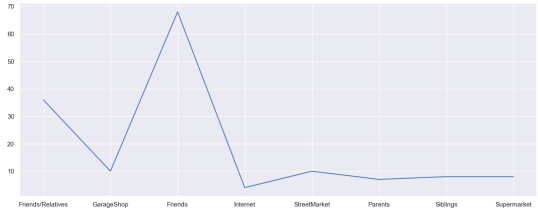

Figure 6

The survey questionnaire had some interesting questions which reflects about the pupils' mindset, asking them about their point of view of why do people smoke or drink. Figure 7 & 8 shows the frequency of chosen reasons for drinking and smoking respectively. From the graphs, we can infer that pupils were most likely to think that people of their age drink to look cool in front of their friends, because their friends pressure them into it, to be more sociable with friends.

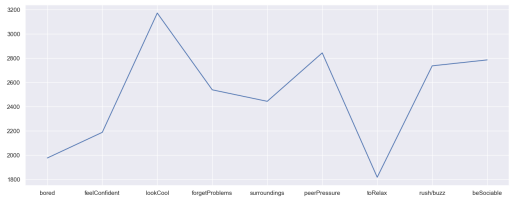

Figure 7

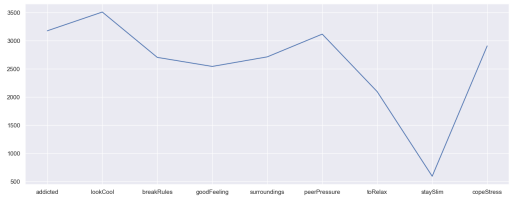

Figure 8

Next, coming to drugs, we will be looking at the number of pupils who have tried drugs among the various kinds of drugs that were included in the survey questionnaire. Figure 9 shows the frequencies of each drugs used by the pupils. We can clearly see that the curve attains three peeks, for Cannabis being the most used, followed by Volatile substances and Psychoactive drugs.

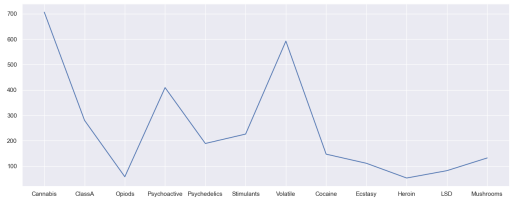

Figure 9

Figure 10 below shows how much of the proportions of pupils have ever used drugs among pupils who drink alcohol and who don't drink alcohol. From the graph, we can see around 50% of the people who drink alcohol, have used drugs, which is quite low compared to pupils who do not drink.

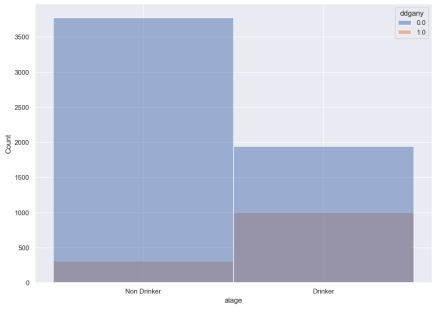

Figure 10

For the drug users among smokers vs non-smokers, the histogram for it wasn't much informative since our dataset has a huge number of non-smokers (80%) compared to smokers (20%).

Also, there are three questions in the survey, that tells us whether pupils think it's okay or not okay to drink alcohol once a week, take cannabis once a week, take cocaine once a week. On evaluating the proportions of pupils' votes for "It's Okay" and "It's not Okay" to use alcohol, cannabis, and cocaine respectively, it was found that majority have voted for "It's not Okay" for three of the substances, but for alcohol, it is concerning that there is quite a considerable number of pupils who think that it's okay to drink alcohol once a week.

Lastly, we have used K-Means Clustering to group the pupils according to their survey responses to check if the clusters have any significant feature, like for example two clusters may be formed, one for who use drugs and the other for who doesn't use drugs. The K-Means algorithm was given the data after transforming the high dimensional data into 2 dimensions using Principal Component Analysis transform. We have also used elbow method to find the optimal K-value for the K-Means Clustering algorithm. In order to check for feature importances, we fitted our data to a classification model, Random Forest Classifier, with all the features as independent variables, and the cluster label, which can be acquired from the trained K-Means model, as the target variable. After training the model, we got the feature importances. On comparing the feature importances, it was found that three features had remarkable high scores of feature_importance, which means these three features were the deciding factors for forming the clusters. The important features were- Age when pupil started drinking, frequency of drinking alcohol, when is the last time pupil had alcohol. So it is quite clear that the clusters were formed based on pupils' responses related to alcohol. On further exploring the cluster features, it was found that one of the clusters had most of the pupils who don't drink alcohol at all, and the other was for pupils who do.

4.3. Results

In this section, we will discuss the results we've got from analysing the data. Certain trends of pupils using addictive substances were found. In our dataset, we've got survey response data of pupils aged 11 to 15, and on comparing the ages when a pupil starts using one of the substances, namely alcohol, cigarettes or drugs, it was found that majority of the survey population started smoking and drinking when they were 13, and the popular starting age of using drugs is 14. The fact that the minimum age when pupils have started using these addictive substances is 5, is concerning. We've also discovered from that in majority of the cases, parents are the usual source from where under aged pupils get alcohol and cigarettes. Friends have been a source of cigarettes as well, but the friends might have got it from their parents only.

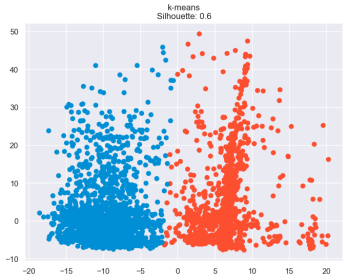

Figure 11

Figure 11 above, shows the scatterplot for the clusters formed using the K-Means algorithm. As we can see, the clusters are clearly not overlapping at all, the clusters have quite distinct boundaries. As inferred from the feature importances, the two clusters are depicting pupils who drink alcohol, and who have never tried alcohol.

5. Critical Reflection

The overall understanding after the analysis carried out, points out some significant trends. Firstly, that parents are the most common source of cigarettes and alcohol to under age children. Most of the under age children start using any of the addictive substances when they're around 13 or 14. The type of drug majority of the pupils try at first drug use is either cannabis or volatile substances, which suggests how easily accessible these are to children. Based on gender, it was found that both boys and girls are equally involved with habits of smoking, drinking or using drugs.

It's worth mentioning how the youth's mentality about smoking and drinking are so badly naive. The majority of the pupils have stated the reason for drinking or smoking is to look cool, which suggests the lack of proper information about the ill-effects of addictive substances among the pupils, as they are clearly not understanding the gravity of the effects of using those substances, compared to the cause being "to look cool".

Among the different kinds of drugs, it was found that cannabis is the one which majority of the pupils use, followed by volatile substances and psychedelics. With this information, now we know investigating which type of drug users would help to impact more. We have also found that many of the pupils who drink alcohol, also use drugs.

Table of word counts

Problem statement	200
State of the art	484
Properties of the data	429
Analysis: Approach	366
Analysis: Process	800
Analysis: Results	200
Critical reflection	227

References

1. OECD (2010), "Smoking and Alcohol Consumption among Children" in Health at a Glance: Europe, OECD Publishing

2. Mark G. Myers and John F. Kelly on "Ciggarette Smoking Among Adolescents With Alcohol and Other Drug Use Problems"

3. Emla Fitzsimons and Aase Villadsen on "Substance Use and antisocial behaviour in adolescence"

4. Patrick McC Miller and Martin Plant on "Drinking, smoking, and illicit drug use among 15 and 15 year olds in the United Kingdom"

5.Emi Nakaseko, Sayaka Kotera and Minato Nakazawa on "Factors Associated with Smoking and Drinking among early adolescents in Vanuatu: A Cross- Sectional Study of Adolescents and their Parents"