



Survey on Consumption of Alcohol Among College Students

IBM-306 : Marketing Research

Under the guidance of :

Dr. Joginder Kumar Nayak

Assistant Professor
Department of Management Studies
Indian Institute of Technology, Roorkee

Presented by :

Swapnesh Kumar

Reg. No.: 16116070
Phone No.: +91-9835813177
Email: skumar2@ec.iitr.ac.in

Yash Jain

Reg. No.: 16116080
Phone No.: +91-8192007320
Email: yjain@ec.iitr.ac.in

Shivanshu Raj Shrivastava

Reg.No.: 16116063
Phone No.: +91-9425646127
Email: sshrivastava@ec.iitr.ac.in

Vishal Sharma

Reg.No.: 16116076
Phone No.: +91-8755517136
Email: vsharma@ec.iitr.ac.in

CONTENT

1. Acknowledgement	<u>2</u>
2. Case study background information	<u>3</u>
3. Methodology and Research Design	<u>4</u>
4. Data visualization	<u>9</u>
5. Hypothesis development	<u>16</u>
6. Factor Analysis (using SPSS Software package)	<u>17</u>
7. Machine Learning Model	<u>20</u>
8. Conclusion	<u>23</u>
9. References	<u>24</u>

ACKNOWLEDGEMENT

We have taken our efforts in completing this project. However, it would not be possible without our mentor Dr. Joginder Kumar Nayak. His kind support and guidance helped us, clearing our doubts about this project.

We would like to express our gratitude towards our peers as they provided us data for my research by filling the questionnaire for this project. Their responses helped us getting a better knowledge of the outside world about our research statement.

Thank you.

CASE STUDY BACKGROUND INFORMATION

It has been a fairly long time that we have spent in the college. Amongst the various activities that happen around everyday, one of the most hyped, interesting and disputed discussion is about alcohol consumption for sure. Whether it's a fest, college group parties, weekends, hanging out with seniors/juniors and many more, a huge pool of students can be easily spotted with the drinks around. There are many who don't consume it initially but eventually things change irrespective of whatever belief they follow or the family background they belong to. There are people who consume it very frequently; their life and fun revolves around it and they try to find out reasons to just go and have a shot & there are many who barely drink or don't drink at all.

Talking specific to IIT Roorkee, a tremendous rise has been observed in the amount of alcohol consumption among the student's community in recent times. Many unwanted incidents have been reported in our campus of students being drinking in campus and roaming around all hungover. We have found many such cases of misbehaviour of stoned students in fests and other occasions. Many wardens have performed checking for such things like alcohol and found these in the rooms.

This is a survey to find out some useful data which can be processed further to extract some facts which might help in decreasing Alcohol consumption in campus. We sincerely request you to fill the form using best of your knowledge. Feel free to answer sensitive questions, we are not tracking the respondent hence confidentiality will be maintained.

METHODOLOGY AND RESEARCH DESIGN

Research Questions

Classification Questions:

1. In which institute are you ?
2. What is your programme, branch and year of study ?

Judgemental questions:

1. What is the approximate percentage of people do you think are frequent alcohol consumer in your college?
2. How do you rate an addicted consumer?
3. What percentage of alcohol consumers do you think are addicted ?
4. In which year do you think a non-drinker has a high chance of trying or getting addicted to alcohol consumption?
5. What percentage of drinkers do you think are binge drinkers (drink more than 5 units in one sitting) ?

Reasons-based Questions:

1. Why does one continue alcoholism?
2. What can be the reason of one consuming alcohol for the first time ?

Assumption-based Questions:

1. How likely do you think alcohol injurious to health ?

Personal Questions:

1. Do you consume alcohol ?

** The responder remains anonymous.

Methodology

This research was conducted over a 30 day period in the month of September - October 2018 by us. The information contained in this report will not be made available to others without any convincing reasons.

Primary source of information was the online survey that we conducted through Google forms. The survey was extensively shared on various social platforms. To ensure the diversity of students taking part in the survey the questionnaire was shared in student groups of specific year, branch, campus groups etc.

Other sources of information include:

- Online research including third party references
- Data collected by others on the same topic

Every effort is made to provide accurate information. The information provided is “best available” on the day the data was gathered.

Target Audience

The target audience will be the people who will answer our research questions and the people whose responses matter to our research. We are mainly focused to educational population for this research.

There are many factors which contribute to alcoholism in students and weather in different institutions, these factors are more or less same. But some responses of our research questions differed for different study programmes. So, we decided to target the entire studying audience and then classify them according to their field of study or their study programme.

We are classifying the audience in their study programmes because the responses/views of people of a study programmes were seemingly related.

Questionnaire Design: ([Questionnaire](#))

We decided to perform an associative research because this is a research question which when asked to someone in first person, he/she will probably lie. So we made a questionnaire in which the responder has to answer the question in the context of third person.

It is a psychological fact when some bizzare question is asked to a person directly or in first person context, the responder probably lies. But if the same question is asked indirectly or

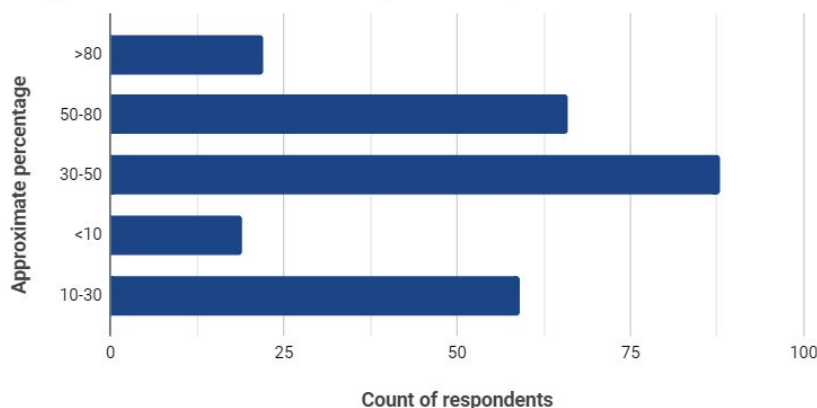
say in third person context, he/she will most probably answer what he does or his actual views on that research question.

The questionnaire has to be easy going for the responder, which he can answer quickly and should be appealing to him/her. So we decided to add multiple choice questions and scaling based questions instead of subjective ones as people only like to devote 2 to 3 minutes to fill some research questionnaire. This can produce less or limited data but if we use subjective questions, the responses of the targets will be different hence it won't create a classified data. Classified data is necessary over a diverse data.

The first section of the questionnaire was to gather information of the responder to classify the responses. The second section is based on the judgements or assumptions of the responder. In this section, the responder has to judge the percentage of people on his/her institution according to the questions. The third section was the probable reasons for a research question. The responder has to rate each reason on a likert scale. The fourth section was assumption based question which the responder is supposed to rate his/her views on a likert scale. The fifth section was a personal question which the responder has to answer in 'YES' or 'NO'.

Basic Data Analysis [[Response Sheet](#) [Response Graph](#)]

What is the approximate percentage of people do you think are frequent alcohol consumer in your college?



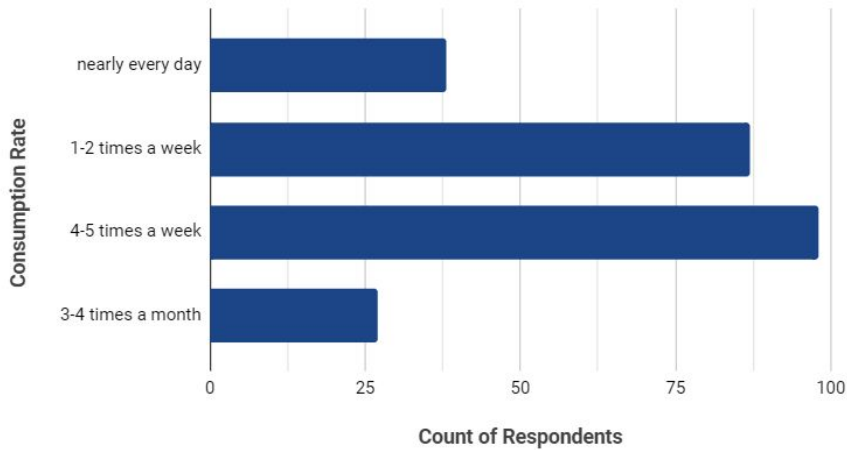
Each of the respondents of this survey are surrounded by an immediate circle of friends. They may be their hostel neighbours, campus group seniors/juniors etc. So, the idea of this question was to get a rough

estimate of the alcohol consumption based on the response of this random sample space. Talking about it with a response error of 10%, both the extremes i.e. <10 and >80 has not been upvoted much as expected. Maximum no. of people have responded neutrally nearing 50%. Almost an equal response can be observed for the range above and below it. Inferring anything out of it might not be advisable but this definitely clarifies the variation of respondents starting from the one who absolutely deny alcoholism to the one who are completely addicted to it.

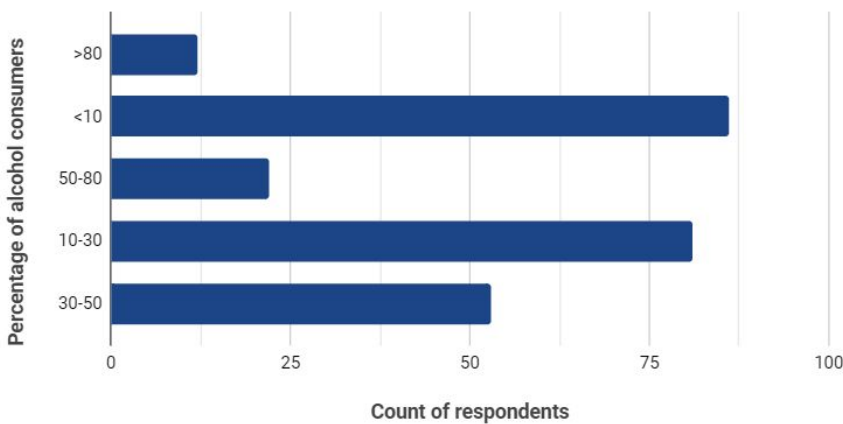
The most generic hypothesis to this question is that the one who are frequent alcohol consumer, their criteria to judge an addicted consumer will be probably high. Everyone has a defined ability to consume alcohol, as you start pushing yourself your ability increases. So, for one who doesn't consume at all, they might find 1-2

times a week addiction. We can safely assume an error of 10% due to some fake and incorrect responses. But that still gives a quite fair idea of the two thought process of the respondents as described beforehand. 1-2 times a week and 3-4 times a week has received the maximum and nearly response while the mid value i.e. 2-3 times a week is quite less. This also leads us to an approximate 50% alcohol consumer as evident from the previous graph as well.

How do you rate an addicted consumer?



What percentage of alcohol consumers do you think are addicted ?



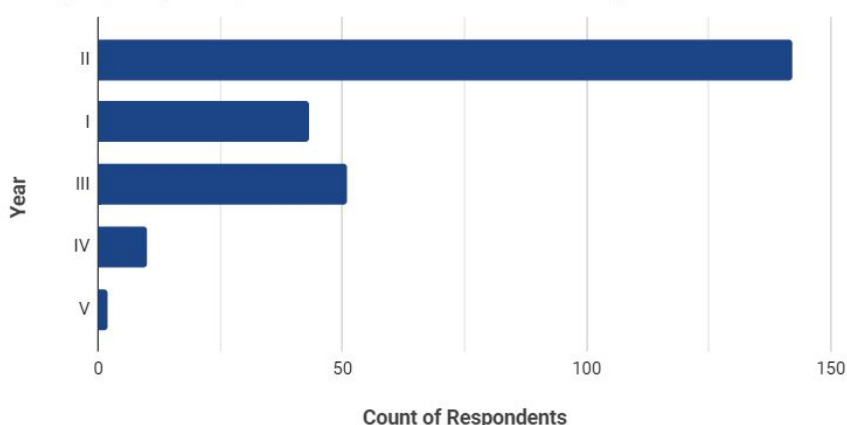
Once we have defined the addiction criteria of our set of respondents, this question gives a fair idea of the amount of people who might be addicted amongst those who are into alcoholism. Assuming a 50-50 scenario of alcohol consumption, maximum no. of people think that only

less than 10% people are addicted to it. If we average the responses, it will go around 30% which shows that out of those who drink, only 30% of them are addicted. This is a fairly good response.

This is a the most justified response and definitely the most expected one due to various reasons. The 1st yearites usually come to the college to stay alone for the first time in their

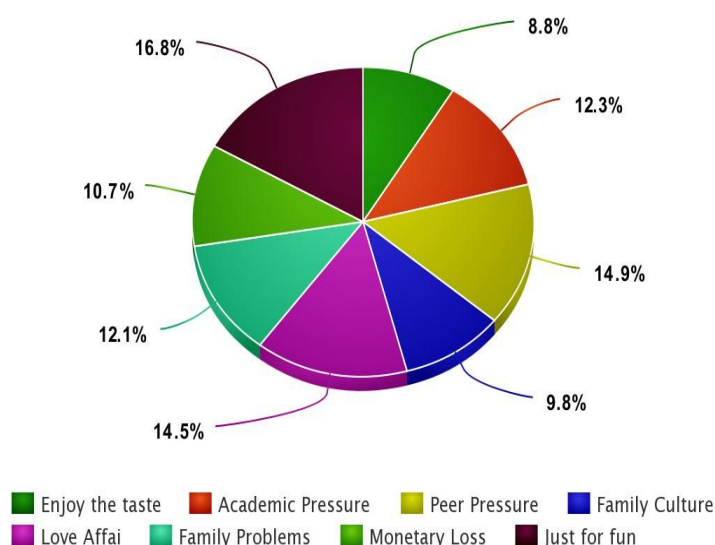
life. Before that they are under their parents supervision so nearly 80% of them don't know about it. Moreover, talking about the college culture, the students community don't insist the newcomers to drink maybe because it's ethically wrong guiding them in the wrong direction and definitely their bonding is less. But this might not be true in all the cases as displayed by the value shown

In which year do you think a non-drinker has the highest chance of trying or getting addicted to alcohol consumption? [Combin...



above. By the end of 1st year, you know almost all the activities that happen around in the college. So, it's pretty obvious that you may start this up or get addicted due to various reasons as written discussed in the following sections. Few of the one who survive their 2nd year lose their resistance in 3rd year. 4th yearites being the most senior year turn out of the most mature people in the crowd. So, even the discussions related to alcoholism fades away. This is the most apt graph and if the administration wants to decrease alcoholism then the most important target shall be the 2nd yearites.

WHAT DO YOU THINK CAN BE THE POSSIBLE REASONS FOR ONE TO CONTINUE ALCOHOLISM?



This pi-chart has been generated by averaging the responses on the scale of 5 for all the variables respectively and then calculating the net contribution of each variable among the total eight. Further sections presents a more detailed analysis among these variables using the, their interdependencies and combined effect using the correlation matrix. This

chart gives a brief idea of the most and the least important reason for alcoholism and doesn't absolutely define the severity of a particular case. It can be clearly seen that the highest contributed factor is fun. The other two important reasons are again influenced by people i.e. peer pressure and love affair. This means that it's eventually the people who are passing on this legacy of alcoholism in college. This is possibly due to the inefficient monitoring of alcoholism in the college and availability of these products around the college premises. Following that we have academic pressure and family problems significantly important. This is a matter of concern which might go unnoticed, so this needs to be strictly looked upon both by the administration and the parents.

DATA VISUALIZATION

(Detailed Data Analysis)

Link: <https://drive.google.com/open?id=194DL1HsDNKfW0AGBjDkdqlsjTTVn1gpo>

In this section collected data has been plotted to analyze the trend and to extract the important factors, various Data visualisation techniques are used to deduce results from collected data

The following graph shows distribution of students consuming alcohol in IITR.

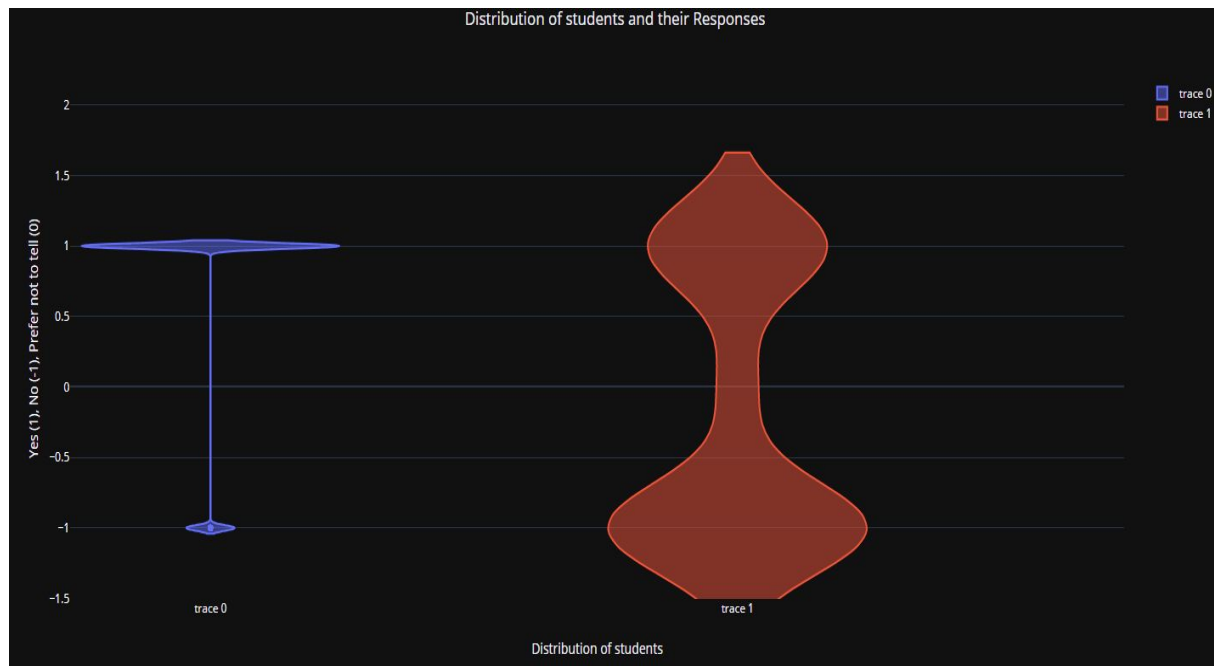


fig.1

Trace 0: College of student (1: IIT T, 2: Other)

Trace 1: Consumption of alcohol by students of IITR.

The following graph represents the distribution of our respondents

Your Program and current year

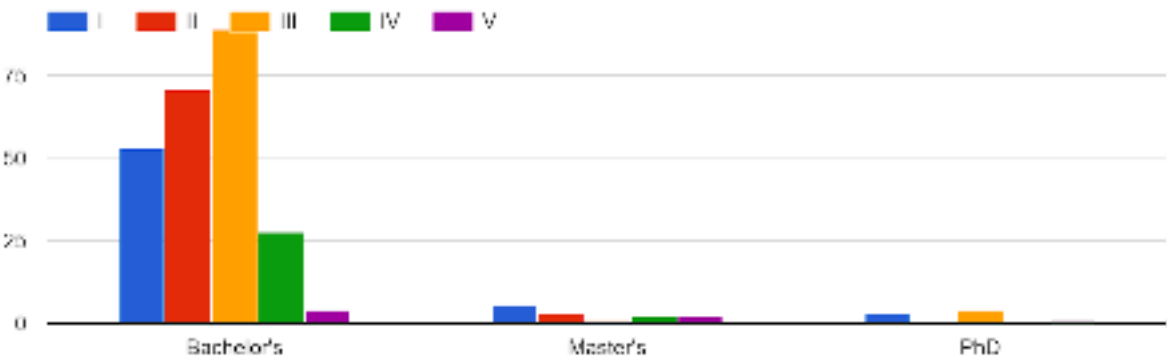


fig.2

This following Box Plot represents the distribution of students in different years and respective rating of consumption of alcohol. The first three plots represents B.Tech, M.Tech and PhD students consuming alcohol according to respondents (from range 1-5) and the next three plots represents B.Tech, M.Tech and PhD students consuming and then continuing alcohol according to respondents.

The speciality in these box plots is that the both two sections (first three and second three) are similar in nature i.e. if the chance of consuming and chance of consuming and then continuing is almost equal.

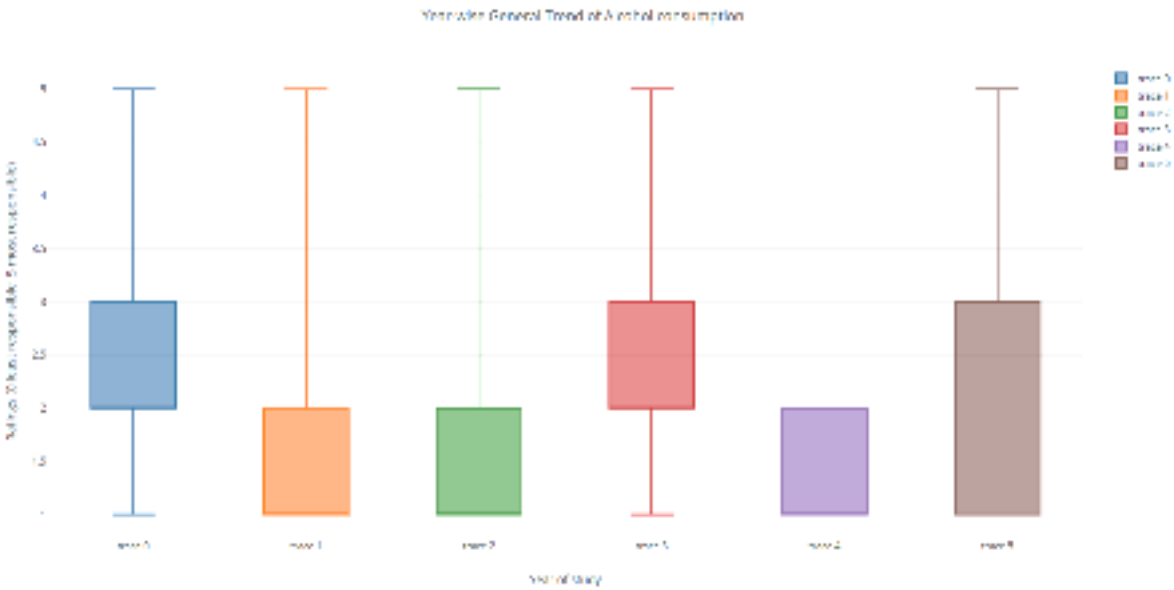


fig.3

Trace 0,3: B.Tech Students

Trace 1,4: M.Tech Students

Trace 2,5: PhD Students

According to our study the respondent thinks that the students who drinks in the fashion mentioned in the figure are addicted. This factor acts as an important factor to train our Machine learning model as those who are addicted to alcohol thinks that it is fine to drink alcohol 4-5 times a week, whereas those who do not drinks thinks that people consuming alcohol 1-2 times a week are addicted.

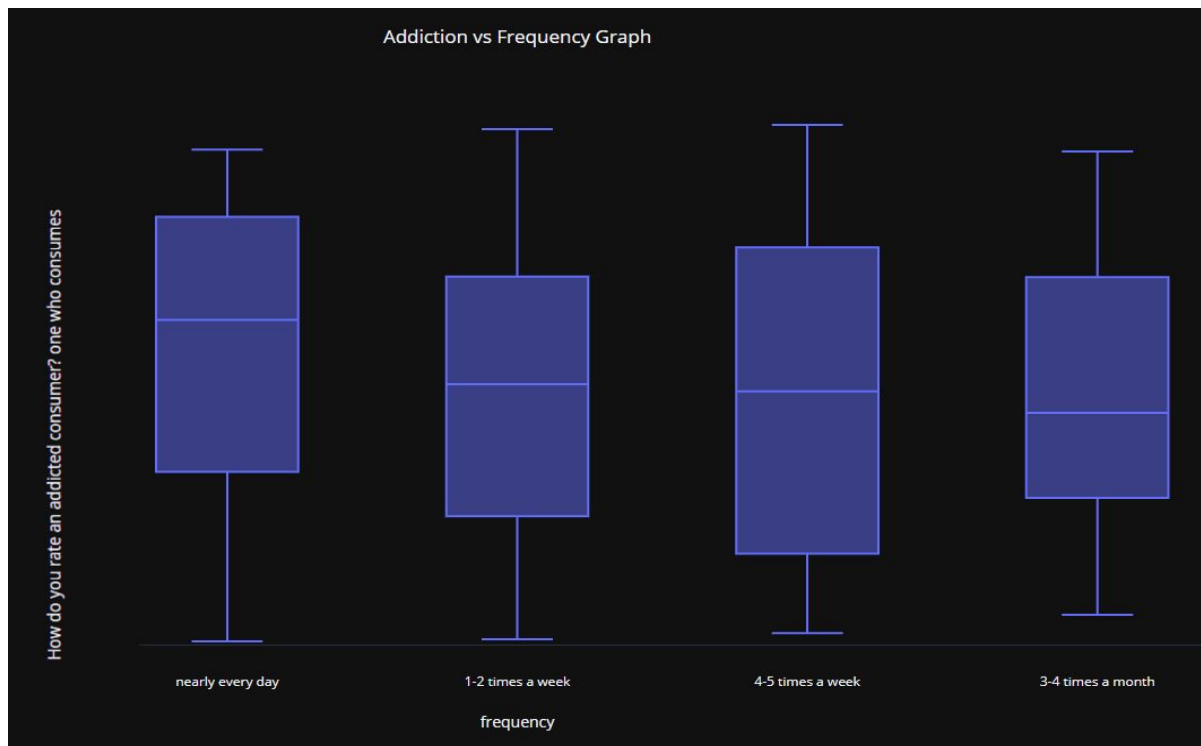


fig.4

The following Violin plots represents the density of people answering various questions represented by mark "Trace". We tracked these details and also tracked the respondent nature that is whether he/she is alcohol consumer or not. These data helped us in analysing independence i.e. correlation between different variables.



fig.5

Trace 0: What is the approximate percentage of people do you think are frequent alcohol consumer in your college?

Trace 1: What percentage of alcohol consumers do you think are addicted ?

Trace 2: What percentage of drinkers do you think are binge drinkers (drink more than 5 units in one sitting) ?

Trace 3: How likely do you think alcohol consumption is injurious to health ?

The major factor of any student to try Alcohol and to continue Alcoholism are highly correlated. In order to find their mutual dependence their data must analysed to deduce facets

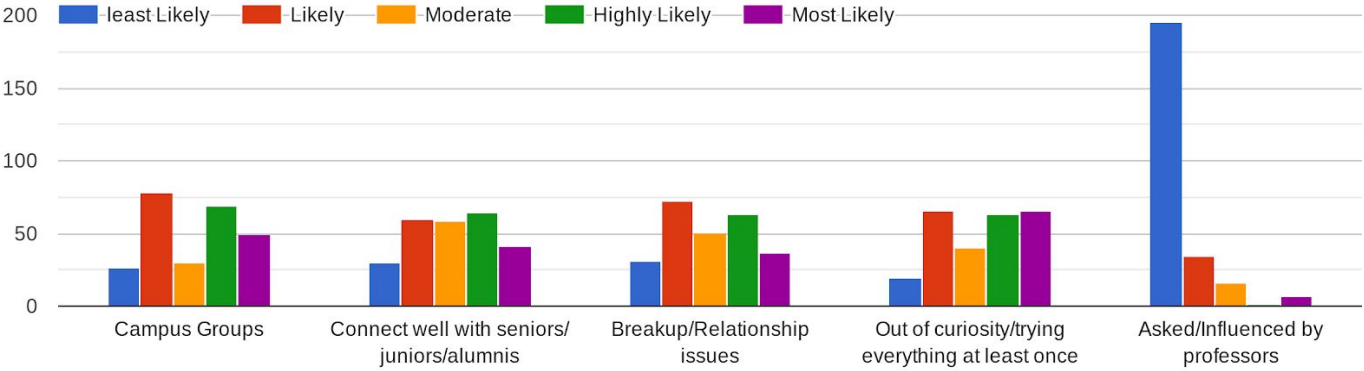


Fig.6.

The variables of above bar graph are interrelated and their interdependence can be find out using Heatmap or Correlation matrix. The following Diagram is heat map of correlated values of reasons for one to Try Alcohol.

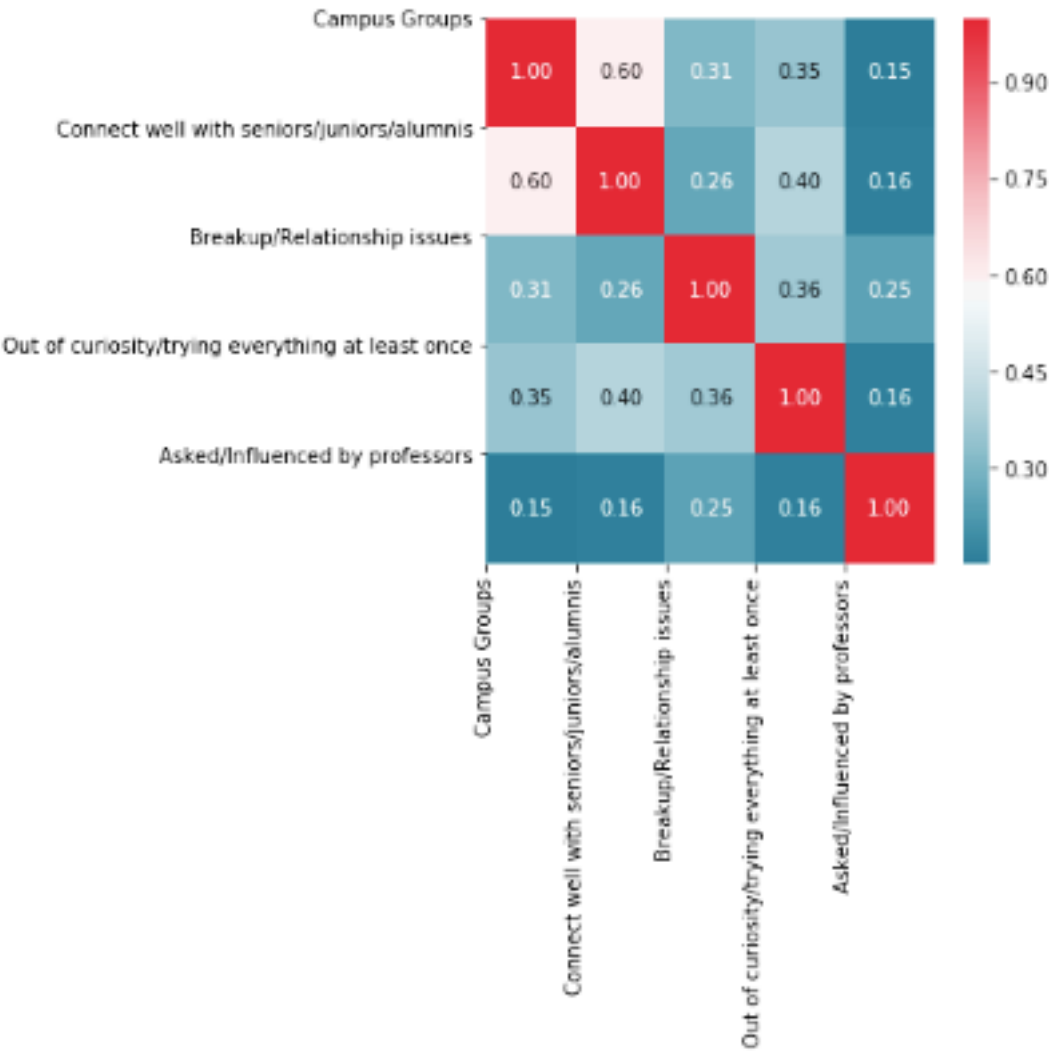


fig.7

On Analysing the data collected from respondent (detailed data given in above link), We find that the reasons for one to continue Alcohol consumption are also highly dependent on each other i.e. if a person is experiencing peer pressure and also experiencing academic pressure than there is a high chance (0.44) for he/she to become a addicted alcohol drinker if he/she had already tried drinking alcohol because of these factors.



In order to use factor analysis and to combine factors and also to find correlated variables. A final correlated matrix is needed to be generated which can depict the correlated values of Reasons responsible for one to try drinking and reasons for one to continue drinking.

	Enjoy the taste.	Academic pressure.	Peer pressure	Family culture.	Love affairs.	Family problems.	Monetary loss.	Just for fun.	Campus Groups	Connect well with seniors/juniors/alumnis
Enjoy the taste.	1.000000	0.183452	0.088507	0.276030	0.074657	0.084124	0.211138	0.173001	-0.027222	-0.042088
Academic pressure.	0.183452	1.000000	0.437257	0.221190	0.413808	0.500040	0.338839	-0.171123	0.010061	-0.070729
Peer pressure	0.088507	0.437257	1.000000	0.188517	0.282960	0.276891	0.243370	-0.018407	0.177947	0.228224
Family culture.	0.276030	0.221190	0.188517	1.000000	0.215019	0.414230	0.251287	0.022831	-0.057606	0.037733
Love affairs.	0.074657	0.413808	0.282960	0.215019	1.000000	0.454919	0.414511	0.094538	-0.038735	0.015985
Family problems.	0.084124	0.500040	0.276891	0.414230	0.454919	1.000000	0.579198	-0.162829	0.000706	0.038817
Monetary loss.	0.211138	0.338839	0.243370	0.251287	0.414511	0.579198	1.000000	0.002238	-0.035697	-0.008635
Just for fun.	0.173001	-0.171123	-0.018407	0.022831	0.094538	-0.162829	0.002238	1.000000	0.220577	0.198260
Campus Groups	-0.027222	0.010061	0.177947	-0.057606	-0.038735	0.000706	-0.035697	0.220577	1.000000	0.597826
Connect well with seniors/juniors/alumnis	-0.042088	-0.070729	0.228224	0.037733	0.015985	0.038817	-0.008635	0.198260	0.597826	1.000000
Breakup/Relationship issues	0.022613	0.299417	0.226647	0.018058	0.377926	0.305384	0.272022	-0.064195	0.307565	0.260231
Out of curiosity/trying everything at least once	0.055799	0.010066	0.104082	0.044046	0.004843	-0.074068	-0.030999	0.269197	0.345948	0.399623
Asked/Influenced by professors	0.280755	0.055765	0.108901	0.070983	0.043576	0.066887	0.174348	0.055192	0.152772	0.161797

Fig.9

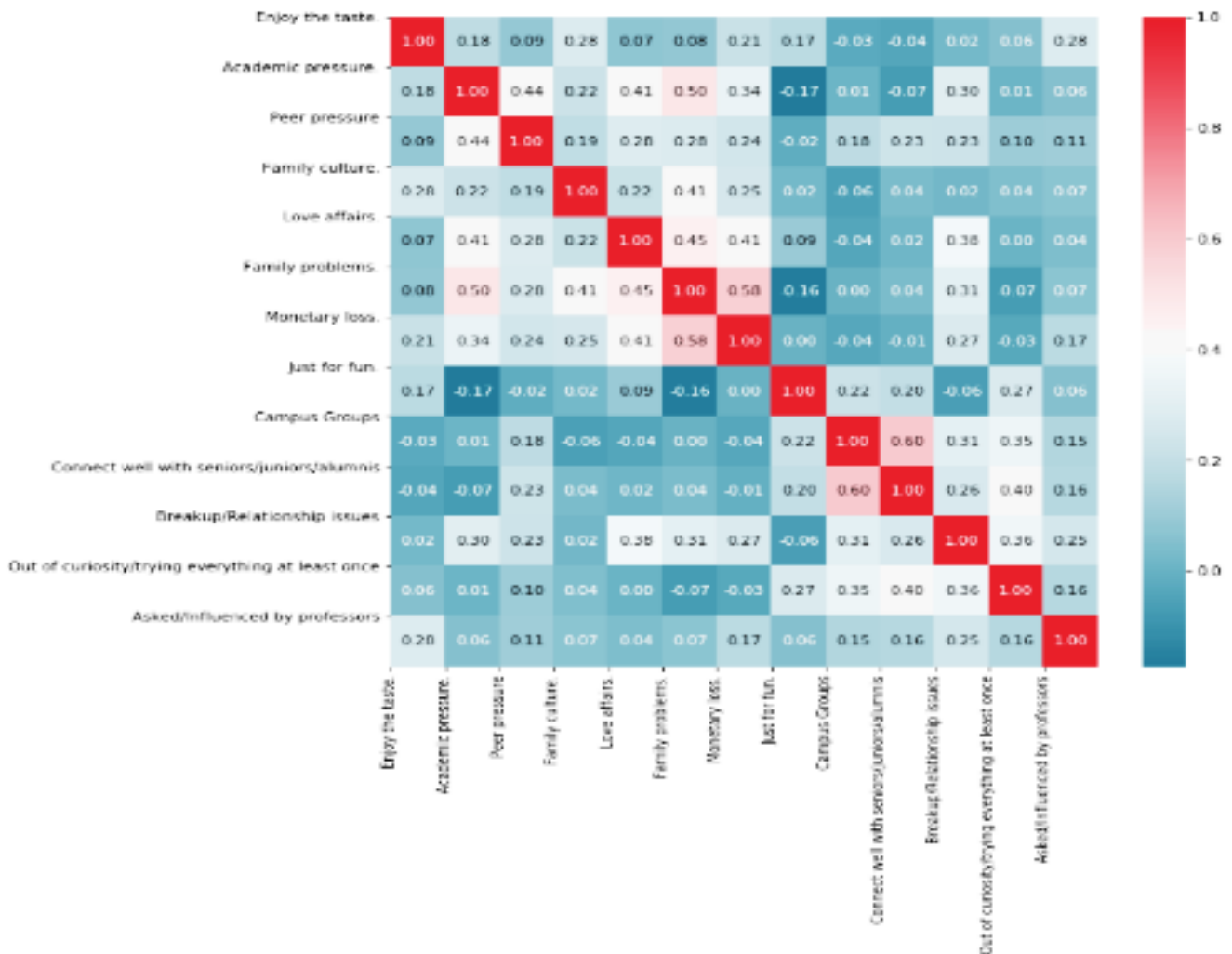


fig.10

HYPOTHESIS DEVELOPMENT

Relation between the awareness of people and their Drinking habits

To find whether the drinking habits of students and their awareness about health effects of alcohol are related or not. We decided to test a hypothesis on our obtained samples.

Hypothesis testing lets us identify that. It lets a sample statistic to be checked against a population statistic or statistic of another sample to study any intervention etc.

Hypothesis testing is defined in two terms – **Null Hypothesis** and **Alternate Hypothesis**.

- **Null Hypothesis** being the sample statistic to be equal to the population statistic. For eg: The Null Hypothesis for the above example would be that the average marks after extra class are same as that before the classes.
- **Alternate Hypothesis** for this example would be that the marks after extra class are significantly different from that before the class.

In our research we developed a hypothesis that the students who do not consume alcohol, consider that alcohol consumption is injurious to health and they have given a response of '4' or '5' on the likert scale. And those who consume alcohol do not consider it as injurious to health. That means they should have rated that question as '3' or less on the likert scale.

Hence we can develop a NULL hypothesis that those who do not consume alcohol are aware of its serious effects to one's health.

To test this hypothesis we will perform a 'Z' test. Our test will be performed in a confidence interval of 90% since the data we collected is small and these are someone's thoughts, which may vary.

First we will change our data to binary, responses ≥ 4 will be set as '1' and the responses < 3 will be set as '0'. In our data after thresholding, there are 139 responders who do not consume alcohol and 115 responders who consume alcohol. Thus, according to the value assigned after thresholding, the population mean for the NULL hypothesis will be ' μ_0 ' = **139/254**.

In our sample data we have **108** respondents who consider that alcohol is not injurious to health and **146** respondents think alcohol is injurious to health. Hence by using the assigned

values, our sample mean will be equal to ' \bar{X} ' = **146/254**. In our sample, we got a standard deviation of ' s ' = **0.4953488816**. So, using a confidence interval of **90%**,

$$Z = (\bar{X} - \mu_0) / s$$

$$Z = 0.89126$$

For 90% confidence interval, Z should be out of the range **(-1.64, 1.64)** to reject the NULL hypothesis. Now, with **Z = 0.89126**, we are failed to reject the NULL hypothesis.

So, we can come to a conclusion that the students who do not consume alcohol are aware of its effects on one's health. Hence continued consumption of alcohol by a student is a major cause of his/her lack of awareness towards the health effects of alcohol.

FACTOR ANALYSIS

(Using SPSS Software Package)

Relation between starting alcohol consumption and continuation of alcohol consumption.

We have asked the views of the respondents on the possible reasons for one to start consuming alcohol and also for the possible reasons for one continuing its consumption.

We can develop a hypothesis that the probable reasons of one starting alcohol consumption and the reasons for one continuing its consumption are related. If this hypothesis comes true, one can be prevented from being a frequent consumer of alcohol by knowing the reasons for starting it.

In our questionnaire we have asked respondents views on a likert scale, with 5 being the most likely and 1 being the least likely.

We can consider the reasons for starting alcohol consumption as a factor 'X' and reasons for continuing alcohol consumption as a factor 'Y'. Both X and Y will be having reasons A, B, C, etc as variables which will be weighted according to the values assigned to the likert scale.

There will be loadings for each factor for each variable. The factor loading of a variable for a factor will be equal to the standard error for the variable.

Factor Analysis

We analysed the factors : continuation of alcohol consumption and starting alcohol consumption and their variables in the questionnaire. We fed our surveyed data to **IBM SPSS** and performed factor reduction technique.

The analysis of factors is :

Communalities		
	Initial	Extraction
Justforfun	1.000	.671
Loveaffairs	1.000	.528
Peerpressure	1.000	.419
Familyproblems	1.000	.685
Connectwellwithseniorsjunior salumnis	1.000	.674
AskedInfluencedbyprofes sors	1.000	.767
Outofcuriositytryingeveryth ing atleastonce	1.000	.519
BreakupRelationshipissu es	1.000	.686
CampusGroups	1.000	.654
Monetaryloss	1.000	.547
Familyculture	1.000	.534
Academicpressure	1.000	.573
Enjoythetaste	1.000	.700
Extraction Method: Principal Component Analysis.		

fig.16

By factor analysis we obtained communalities of the variables.

Total Variance Explained						
Component	Total	Initial Eigenvalues		Extraction Sums of Squared Loadings		
		% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.335	25.656	25.656	3.335	25.656	25.656
2	2.247	17.282	42.938	2.247	17.282	42.938
3	1.352	10.402	53.340	1.352	10.402	53.340
4	1.023	7.872	61.213	1.023	7.872	61.213
5	.890	6.849	68.062			
6	.828	6.371	74.433			
7	.746	5.739	80.173			
8	.603	4.635	84.807			
9	.541	4.159	88.966			
10	.470	3.613	92.579			
11	.381	2.929	95.508			
12	.309	2.373	97.881			
13	.275	2.119	100.000			

Extraction Method: Principal Component Analysis.

fig.17

To make our dataset more reliable we obtained the variance analysis. By this we were able to remove the outliers and ill-responses.

Further we applied factor and variable reduction on our data and we got the following matrix of factor matrix.

Variable \ Factor	Continuation of alcohol	Starting alcohol
For Fun	0.700	0.671
Love Affairs	0.686	0.528
Peer pressure	0.671	0.419
Monetary Loss	0.547	
Academic Pressure	0.573	
Influenced by professors		0.767

We have used only common variables in the two factors for our factor analysis.

Variable \ Factor	Continuation of alcohol	Starting alcohol	Sigma(h ²)
For Fun	0.7	0.671	0.924
Love	0.686	0.528	0.749

Affairs			
Peer pressure	0.671	0.419	0.626

There were some other variables in the two factors but those variables were different, hence to make data clear and easy, we used only common variables for factor analysis.

The factor loadings are decided by **pearson's correlation coefficient**. Pearson's correlation coefficient is given by :

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

The communality value(h^2) is as in the above table. If communality is < 0.6 then the two factors for that variable are are dependent.

The results of the factor analysis gives **communality** > 0.6 . So, we can reject the NULL hypothesis and hence both the factors X and Y for a variable are independent.

By this result, we are able to find that the reasons for one starting consumption of alcohol and the reasons for one continuing consuming it are not dependent. Hence we can come to a conclusion that reasons for one starting consuming alcohol and one continuing it are independent. So, we can arrive to a conclusion that if a person starts consuming alcohol, he will find reasons to drink it often and often and become a continuous drinker.

MACHINE LEARNING MODEL

ML Model: <https://drive.google.com/open?id=18T3wqHNdW66ziYvBdbaEEUVtaXbGf0of>

Source Code: <https://github.com/VishalCR7/drinkr>

Web App: <https://mydrinkr.herokuapp.com/>

To get more insight from the response collected and get some real life application of the overall project, we brought one of the most trending fields of this decade “Machine Learning and Data Science” in the picture by training a Machine Learning model on the response dataset to predict how likely is a person involved in alcohol consumption and convert the parameters and functions obtained into a webapp that can be shared with institute bodies like

wellness center and student mentorship program to find out about the same from indirect questions.

The following steps were followed to get required trained model and consequently the webapp:

(i) Data Preprocessing :

The raw response data had a lot of inconsistencies like missing responses to some questions by the respondent. Also the google form data obtained was not suitable for any machine learning model to be applied.

Therefore the raw data was preprocessed by following steps:

1. The question missing about 30% of the data were dropped from the analysis to prevent delusional results.
2. The questions to which only a small number of responses were missing were filled by their **mode** values which seems convincing that results will be close enough to population using the same values as mode rather than mean or median.
3. Finally, all the values which were in the form of strings were mapped to numbers so that various mathematical models like linear regression and xgboost can be applied.

(ii) Selection of Dependent and Independent Variables :

The independent variables were all the form fields that had a considerable amount of responses available for analysis and also had sufficient correlation with dependent variable.

The aim was to classify people who were drinkers and non drinkers, therefore dependent variable was chosen to be the question “Do you consume alcohol” for dataset training.

(iii) Application of Machine Learning Models :

After data preprocessing and identification of dependent and independent variables, we finally applied two machine learning models to data to generate the required parameters.

For a simple approach we began by applying logistic regression the most basic statistical model for binary data. It involves estimating the parameters of a logistic model. The results obtained were quite satisfactory as the obtained accuracy on rough model was around 75% .

To move further ahead in terms of accuracy and analysis we used “XgBoost”. XGBoost is short for eXtreme gradient boosting. It is a library designed and optimized for boosted tree algorithms. Its main goal is to push the extreme of the computation limits of machines to provide a *scalable*, *portable* and *accurate* for large scale tree boosting. The results obtained

from this model were quite accurate and satisfactory enough to be used for the intended webapp design.

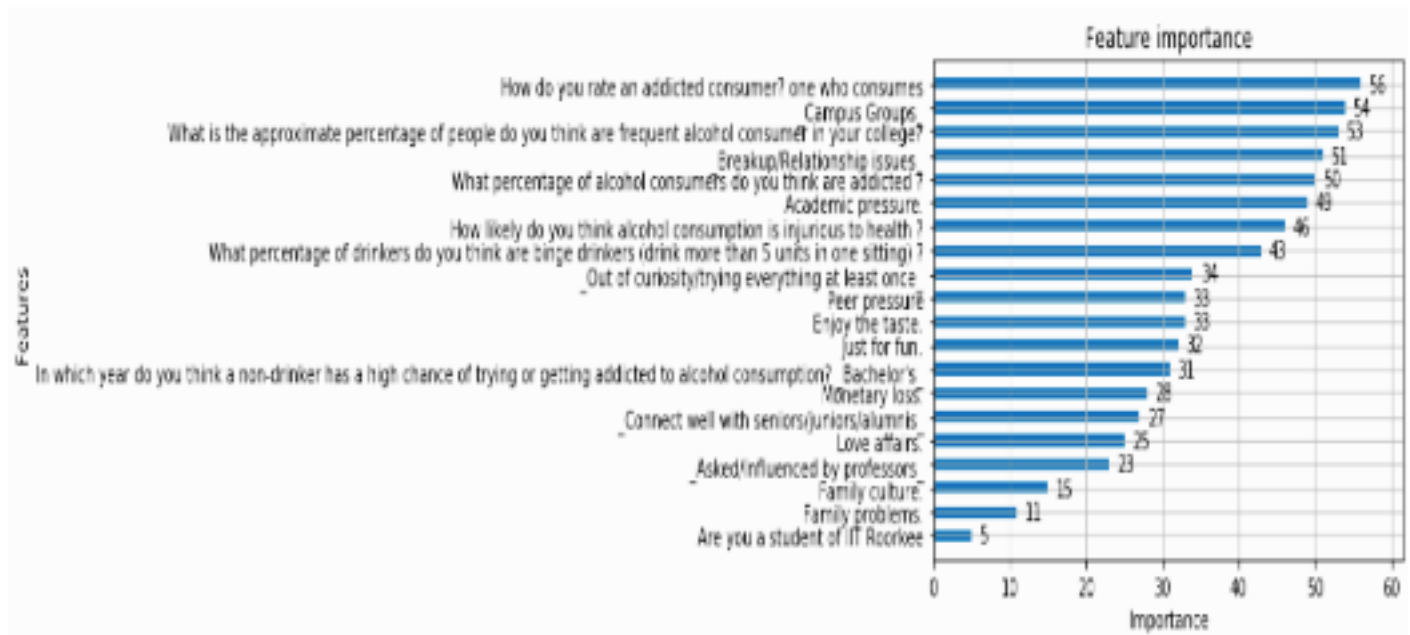


Fig.18

CONCLUSION

In this survey we have collected our data by conveying the questionnaire. Here we figured out correlation between different variables and then implemented those variables to create a proper dataset.

We created a hypothesis that the people's responses whether Alcohol is injurious to health depends on whether they consume alcohol or not. People who consume alcohol think it is not a much of a risk to one's health and the people who do not consume alcohol think that it is much of a risk to one's health. Hence it can be deduced that the drinking habits of students depend on their awareness about the health risks of alcohol. This can be prevented by proper education about alcohol and its risks to students at appropriate age.

Secondly, we checked correlation between reasons of students of starting consuming alcohol and the reasons for continued consuming of alcohol. Here we found that these two factors are not dependent. So, we can deduced that one can start alcohol consumption because of a reason, then he/she can find some other strong reasons to continue consuming alcohol.

Further, we developed a dataset and trained model for machine learning. We have developed an app which can predict whether a person drinks or not in accordance to the questions he answers on that app.

REFERENCES

1. [Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. \(1999\). Evaluating the use of exploratory factor analysis in psychological research.](#)
2. [Factor Analysis by Mark Rahn.](#)
3. [Interpreting regression coefficients by Karen Grace-Martin](#)
4. [Comprehensive & Practical Inferential Statistics Guide for data science](#)
5. [A Simple Introduction to ANOVA](#)
6. [Box plot and their use to visualize discrete data](#)
7. [Research on Violin plot and their uniqueness to represent discrete value data](#)
8. [Correlation coefficient calculation using Pearson correlation](#)
9. [Heat maps and their importance to find correlated values](#)