Earlier known as
B. V. B. College of Engineering & Technology

KLE Technological University
Creating Value
Leveraging Knowledge

School of Computer Science and Engineering

KLE TECH

PROJECT REPORT

Course: Exploratory Data Analysis LAB

(21ECSC210)

Title : Quality of life data analysis

Team Details: IVth Sem - 'C' Div – Team 05

1.) Neeraj N K - 01FE20BCS134

2.) Shreyas Joshi- 01FE20BCS145

3.) Kiran Patil - 01FE20BCS151

4.) Anoop Patil – 01FE20BCS138

Earlier known as
B. V. B. College of Engineering & Technology

KLE Technological University
Creating Value
Leveraging Knowledge

School of Computer Science and Engineering

Problem statement

- **Assesing the quality of life of each patient from Rehabilation Centre by Analyzing their Responses for WHOQOL- BREF Questionnaire .**
- **Then Concluding the QOL Status of Each Patient by Analyzing Each one of their Attributes.**

Earlier known as
B. V. B. College of Engineering & Technology

KLE Technological University
Creating Value
Leveraging Knowledge

School of Computer Science and Engineering

Data Description

- The data set contains of 86 patients details collected from a rehabilitation center.
- To analyze QOL of Indviduals, WHO Prepared a Questionnaire Comprising 100 questions to get all information that is required to calculate QOL,Usually Questions are related to Physical health, social Relationship, Psychological ,and environmental Conditions of Each Indvidual.
- BREF is a Shorter Counterpart version ,Which contains 26 questions.
- Responses to this 26 questions are on a (1-5) Likert scale.
- These 26 questions are further divided into 4 domains :

  1) Physical health(Q-3,4,10,15,16,17,18)

  2) Psychological (Q-5,6,7,11,19,26)

  3) Social relationship (Q-20,21,22)

  4) Environment(Q-8,9,12,13,14,23,24,25)

- Questions 1 and 2 don't belong to any of the domains.

Earlier known as
B. V. B. College of Engineering & Technology

KLE Technological University
Creating Value
Leveraging Knowledge

School of Computer Science and Engineering

UNDERSTANDING THE DOMAINS :

1.Physical Health – 7 Questions

Includes Questions Related mobility, daily activities, functional capacity, energy, pain, and sleep

2. Psychological – 6 Questions

The psychological domain measures include self-image, negative thoughts, positive attitudes, self-esteem, mentality, learning ability, memory concentration, religion, and the mental status

3.Social relationship– 3 Questions

The social relationships domain contains questions on personal relationships, social support, and sex life.

4.Environmental– 8 Questions

Questions related to financial resources, safety, living physical environment, opportunities to acquire new skills and knowledge, recreation, general environment (noise, air pollution, etc.), and transportation.

Earlier known as
B. V. B. College of Engineering & Technology

KLE TECH.
KLE Technological University
Creating Value
Leveraging Knowledge

School of Computer Science and Engineering

Patient Attributes

Description	Related attribites
Basic information about patient	Age,Sex,Education,Occupation
Usage of alcohol	Age when started drinking alcohol,Duration of execessive use of alcohol,Average units used in last 30 days,Alcohol type
Medical Information	Sugar(mg), Other Issues, Withdrawal symptoms experienced when the patient stopped, Medical Problem experienced(Past), Medical Problem experienced(Present), Chronic health problem,Psychiatric complication_Past, Psychiatric complication_Present, History of previous head injureies, if any , Knowledge of allergy to specific drugs(if known)
Impression of the camp officer regarding the patient	impression of Camp officer about the patient, Denial of substance use-related problems, Motivation factor, Willingness for treatment, Action Taken
Work related status of patient	Specify nature of current work, At what age did you start working, How long have you been working, Have you received any special award

Description of patient	Weight while admission (In Kg), Weight while discharge (In Kg), Height(In Ft)
Other habits	Use of Nicotine,Smoking,Other issues
Medical history of patients family	Family history of alcoholism / drug abuse, if any (who and which type of drug), If any SpecV, Family history of psychiatric illness in family, if any specify (who),
Past treatment for alcoholism	Past deaddiction treatments received/not, Name of the Hospital/ Deaddiction Center, year of Treated, Period of Treatment, Period of sober, Remarks
Family members (father, mother, sisters, brothers, wife, children) status	Living: Age, State Of Health Dead: Year of Death, Cause Of Death, Patient age at death

KLE TECH.
KLE Technological University
Creating Value
Leveraging Knowledge

School of Computer Science and Engineering

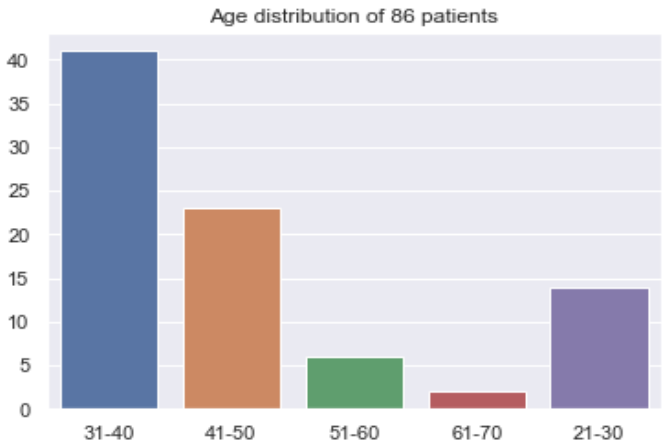
Description	Related attributes
Legal details of patient	Details of Debts to be cleared, Legal complications
Childhood impact affecting currently	Behaviour problem identified in Childhood,Family history, Describe your childhood / teenage year, Religious beliefs
Educational Information of patient	Achievements in the Education, year of Education, High achiever in extracurricular activities
Sexual status of the patient	If unmarried, premarital sexual encounters, Record extra marital experiences, Have you involved in any high risk sexual activities, At present do you have any sexual problem
Marital history	Name, Age, religion, education, occupation, monthly income, Arranged Marriage/Love marriage, Number of year of marriage, multiple marriages, Have you been separated from your spouse due to your addiction, Are you suspicious of your wife, any instance of family violence

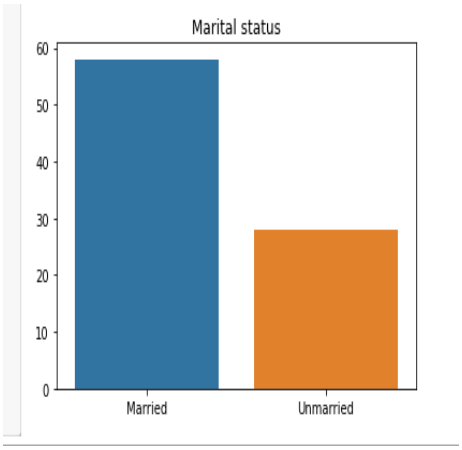
Few Insights About Patient's Data :

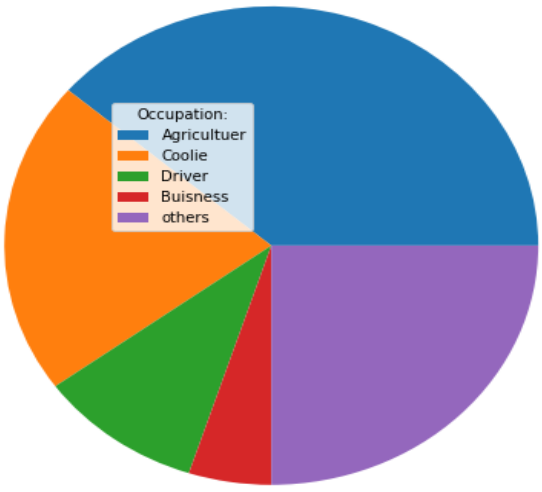
- **67% of patients are Married and 33% are unmarried.**
- **Many of them Started Drinking at Early Age.**
- **Majority of them belong are of 31 to 40.**
- **Majority of patients are in agriculture.**
- **Calculating risk level based on ethanol content and type of alcohol we got to know that 12 are in high**

and 42 are under low risk and 32 are in medium risk level.

- **Most of Them Have Debts in Banks.**







3rd	1
5th	4
6th	1
7th	7
8th	3
9th	3
B.A	3
B.Com	2
B.E, B.B.A	1
B.Tech	1
MBA	1
MSC	1
Not Mentioned	4
PUC	16
PUC\nITI	1
S.S.L.C	37

Earlier known as
B. V. B. College of Engineering & Technology

KLE Technological University
Creating Value
Leveraging Knowledge

School of Computer Science and Engineering

PRE- PROCESSING of DATA

The Original Dataset we Obtained Earlier had 108 Attributes.

But After Pre-Processing We reduced down to 48 Attributes by

Eliminating Columns which Containing Mostly Null Values

& DSM-5 Columns Deleted , they are Not Required for Analysis.

Ex: ***Rligious beliefs and behaviourQn 1***

- **Columns Containing Fewer Null Values ,Got Replaced by Mode or "Not Mentioned " Status .**

Ex: ***Quantity/day of Smoking.***

- **Few Columns are Ignored ,Because Mostly the Input Values are Same by All Patients,Hence they are Not Much Useful for our Analysis.**

Ex : ***Willingness for Treatment*** **etc.**

- **Columns with Irrelevant Attributes are also Eliminated.**

Ex: ***Assessment Date , Sister's Living age*** **etc**

- **Few Columns have Values of type 'Object', So We Converted Such Columns to String.**
- **Certain Values Indicates Same Meaning,But Still Words are Misplaced Slightly**
- **So We Pre-Processed such Values by Verifying Manually.**
  Bringing Values of Same column to Single Unit.

Ex : duration of excessive use of alcohol(in years)

Converted Days ,Months into Years.

Earlier known as
B. V. B. College of Engineering & Technology

KLE Technological University
Creating Value
Leveraging Knowledge
KLE TECH.

School of Computer Science and Engineering

Adding required rows to dataframe

The Obtained Data Set Containing Scores of Each Patient for 24 Questions Belonging to 4 Different Domains.

A Raw Score Column is Added to Dataframe by Summation of their Scores.

But There is Need of Transformed Score to add meaning to the scores and allows some kind of interpretation of the scores.

It allows direct comparison of two scores. For example, a score of 33 on the first test might not mean the same thing as a score of 33 on the second test. So we Added Tranformed Score Table (Ranging 0-100)

to Dataframe by Using Following Formula.

$$\text{Transformed Scale} = \left[\frac{(\text{Actual raw score} - \text{lowest possible raw score})}{\text{Possible raw score range}}\right] \times 100$$

Target Variable : Verdict

A Verdict Containing Categorical Data , Decides the QOL Status of Each Patient by the Help of their Transformed Scores. Which will be Added to Dataframe for Further Analysis.

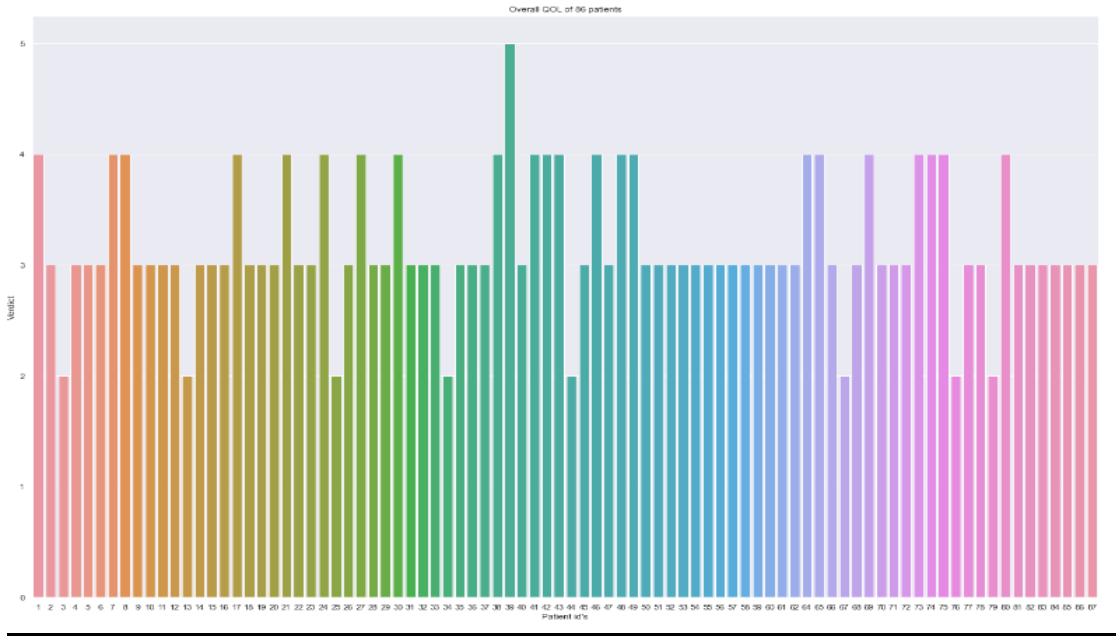
Transformed Score	Verdict
0 - 20	Very Poor
21 - 40	Poor
41 - 60	Neither Poor Nor Good
61 - 80	Good
81 - 100	Very Good

Earlier known as
B. V. B. College of Engineering & Technology

KLE Technological University
Creating Value
Leveraging Knowledge

School of Computer Science and Engineering

Overall QOL of 86 Patients



QOL Status :

From the above graph we can observe that majority of patients lie in neither poor nor good QOL status and hence the overall QOL of all 86 patients is also the same.

There is one outlier with QOL status 5(very good)

The average QOL of all patients is 53.94(which is neither poor nor good).

Analysing QOL of 86 patients considering all domains in which QOL status they fall?

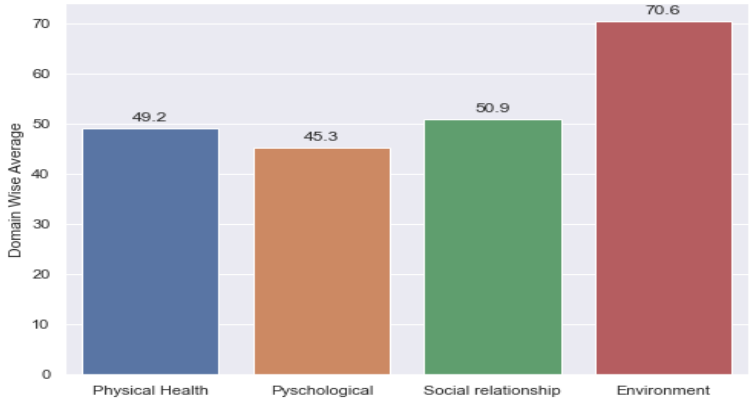
Good	22
Neither poor nor good	55
Poor	8
Very good	1

Earlier known as
B. V. B. College of Engineering & Technology

KLE Technological University
Creating Value
Leveraging Knowledge

School of Computer Science and Engineering

Domain wise average of all patients :



Correlation between domains and verdict(class label)

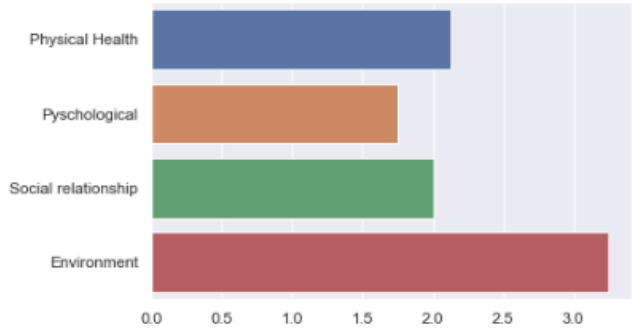
	PID	DS1	DS2	DS3	DS4	SUM_DS	ODS	Verdict	V1	V2	V3	V4
PID	1.000000	0.105329	-0.055144	0.098214	-0.081879	0.029334	0.029334	-0.016507	0.141504	-0.033554	0.072819	-0.108398
DS1	0.105329	1.000000	0.416839	0.364777	0.319729	0.682955	0.682955	0.617992	0.896008	0.384196	0.305992	0.239603
DS2	-0.055144	0.416839	1.000000	0.286905	0.373528	0.744435	0.744435	0.649213	0.433966	0.953329	0.267098	0.396324
DS3	0.098214	0.364777	0.286905	1.000000	0.392039	0.764780	0.764780	0.655339	0.352368	0.300056	0.953138	0.332605
DS4	-0.081879	0.319729	0.373528	0.392039	1.000000	0.666402	0.666402	0.577363	0.343842	0.369306	0.370571	0.887580
SUM_DS	0.029334	0.682955	0.744435	0.764780	0.666402	1.000000	1.000000	0.871878	0.662631	0.721460	0.714414	0.599295
ODS	0.029334	0.682955	0.744435	0.764780	0.666402	1.000000	1.000000	0.871878	0.662631	0.721460	0.714414	0.599295
Verdict	-0.016507	0.617992	0.649213	0.655339	0.577363	0.871878	0.871878	1.000000	0.598020	0.632419	0.608354	0.534837
V1	0.141504	0.896008	0.433966	0.352368	0.343842	0.662631	0.662631	0.598020	1.000000	0.439702	0.287667	0.290782
V2	-0.033554	0.384196	0.953329	0.300056	0.369306	0.721460	0.721460	0.632419	0.439702	1.000000	0.282859	0.404363
V3	0.072819	0.305992	0.267098	0.953138	0.370571	0.714414	0.714414	0.608354	0.287667	0.282859	1.000000	0.325180
V4	-0.108398	0.239603	0.396324	0.332605	0.887580	0.599295	0.599295	0.534837	0.290782	0.404363	0.325180	1.000000

Earlier known as
B. V. B. College of Engineering & Technology

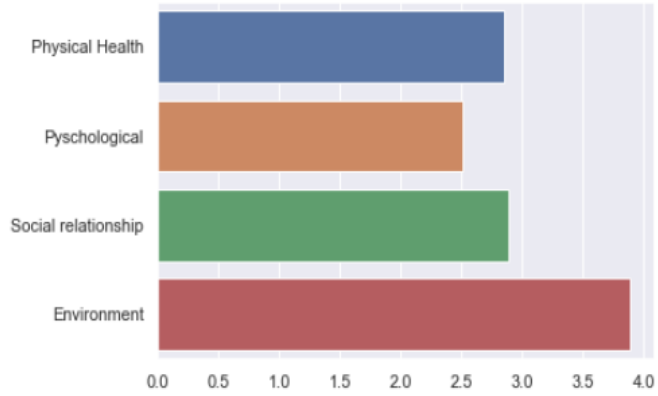
KLE Technological University
Creating Value
Leveraging Knowledge

School of Computer Science and Engineering

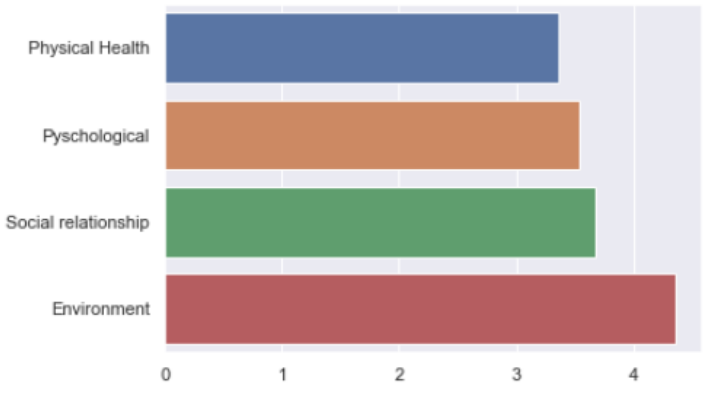
Domain-Wise QOL Analysis



(POOR)



(Neither Good Nor poor)



(Good)

KLE TECH.
KLE Technological University
Creating Value
Leveraging Knowledge

School of Computer Science and Engineering

Patients analysis with QOL status poor

- There are 8 patients with QOL status poor.
- Out of which 7 are married and 1 is unmarried.
- Many of the patients started drinking at a young age fact that their present age is above 31 and hence the period for which they consumed alcohol is more i.e 15.375 years on average.
- 75% of them started working in young age.
- Out of 8 patients, we can observe that all of them are above age of 31.

The majority of them were influenced by their family

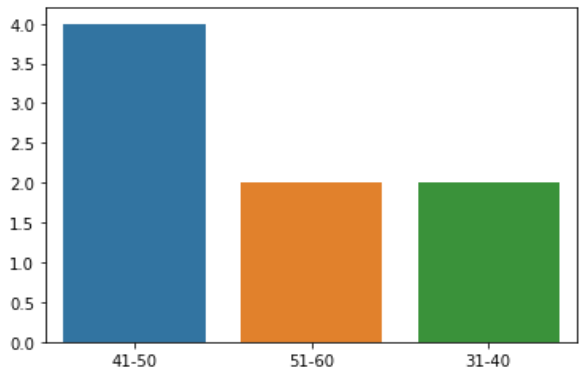
Reason for starting alcohol	
Someone in family or friends were using	6
Someone in family or friends were using, To feel better and confident or happy	1
To avoid problems and sadness	1

Earlier known as
B. V. B. College of Engineering & Technology

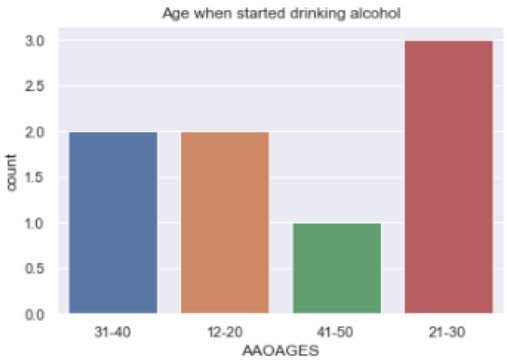
KLE Technological University
Creating Value
Leveraging Knowledge

School of Computer Science and Engineering

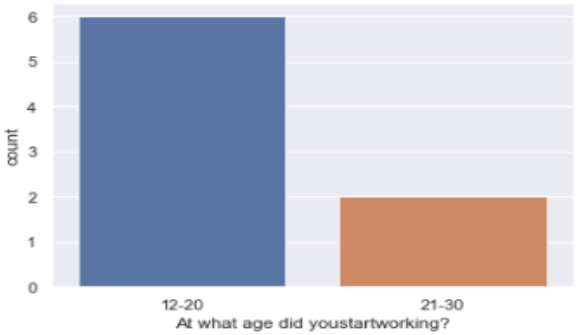
Age distribution of 8 patients.



Age When Started Drinking



Age Started Working



Education in year	
5th	2
B.Com	1
Not Mentioned	2
PUC	1
S.S.L.C	2

Occupation	
Agricultuer	5
Coolie	2
Driver	1

2	None
12	Poverty or severe debts
24	None
33	None
43	None
66	None
75	None
78	Poverty or severe debts

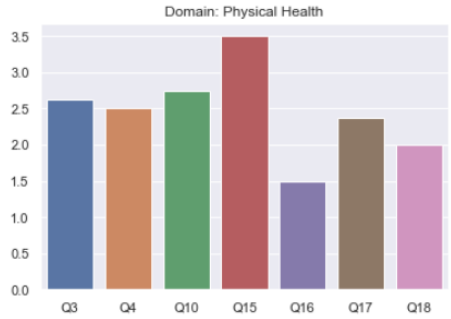
Earlier known as
B. V. B. College of Engineering & Technology

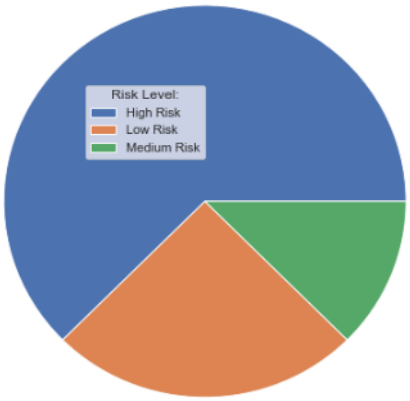
KLE Technological University
Creating Value
Leveraging Knowledge

School of Computer Science and Engineering

Domain Wise Analysis of POOR QOL Patients :

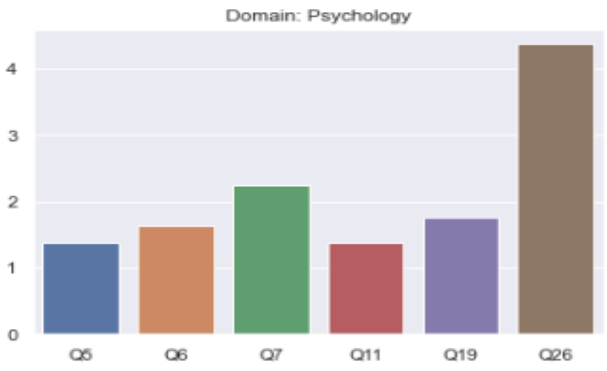
1)Physical Health





38% of them have met with accidents earlier

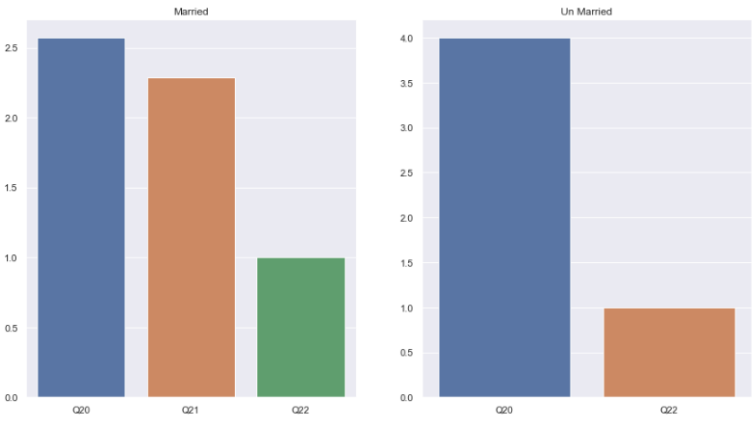
2)Psychological



It is observed that some of the patients have childhood problems of running away from home and scholastic backwardness.

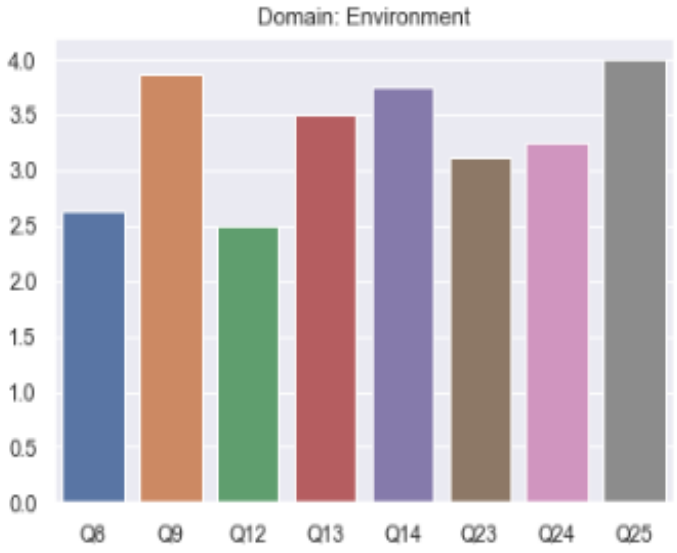
Also 38% of them have problem of confussion sggressive outburst and depression.

3)Social Relationship



Out of 8 patients 7 are married and 1 is unmarried and both don't get support from friend and hence Q22 has least response.

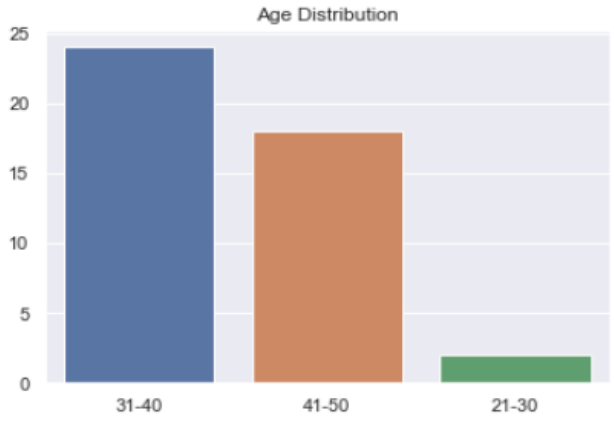
4)Environment

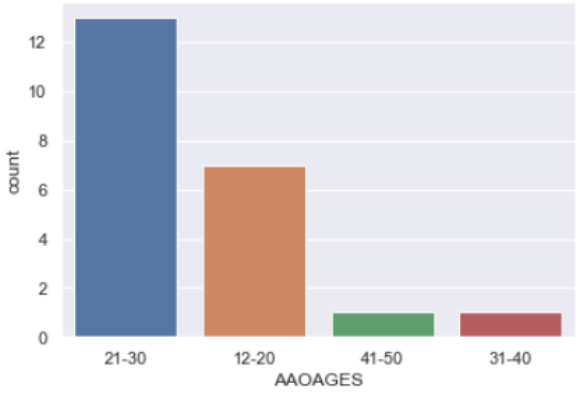


Q12 has less response which about does they have enough money to meet their needs as all of them have debts.

KLE Technological University
Creating Value
Leveraging Knowledge
School of Computer Science and Engineering

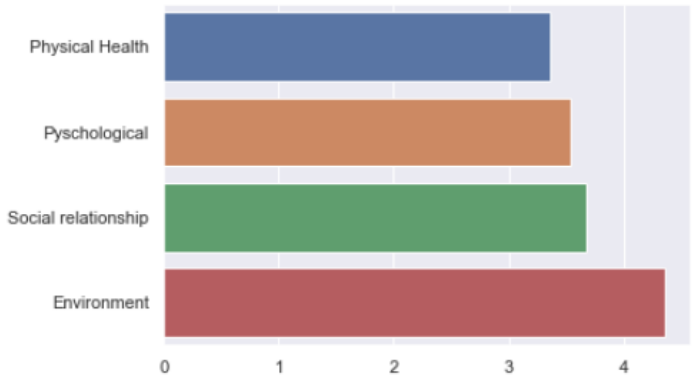
Patients analysis with QOL status Good :





- **Out of 22 patients, we can observe that all of them are above age of 21 and under the age 40.**
- **Of which 50% of patients have education only till 10th grade.**
- **All of the patients started drinking in middle age hence consumed alcohol quantity is less than patients with poor QOL and NPG i.e 13.31 years on average.**

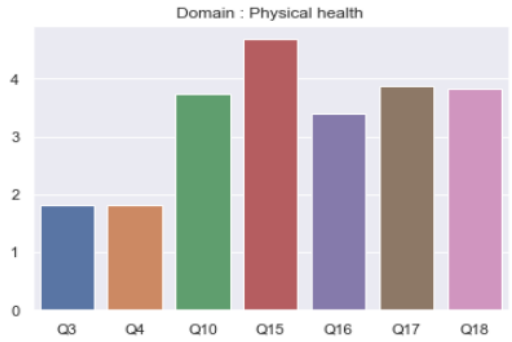
Domain Wise Analysis of GOOD QOL Patients :

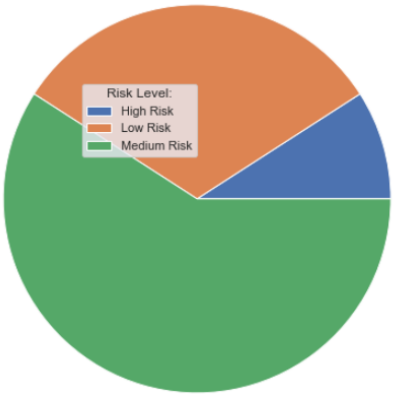


Earlier known as
B. V. B. College of Engineering & Technology

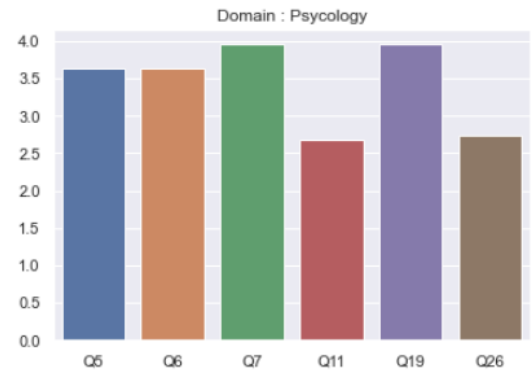
KLE Technological University
Creating Value
Leveraging Knowledge
School of Computer Science and Engineering

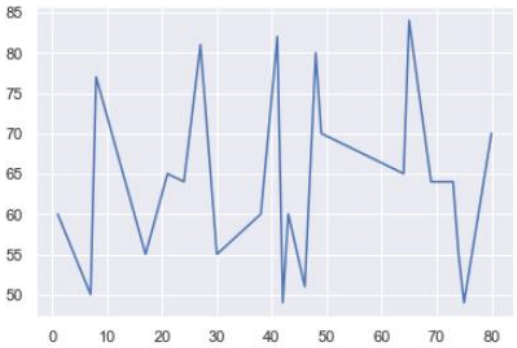
1)Physical Health



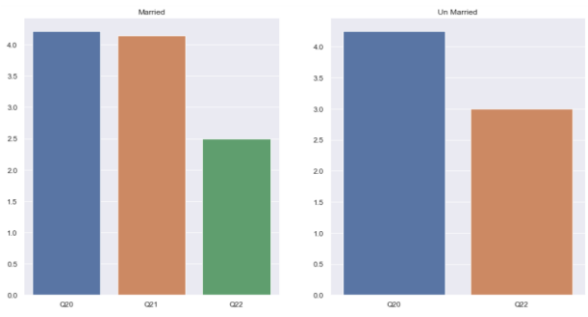


2)Psychological



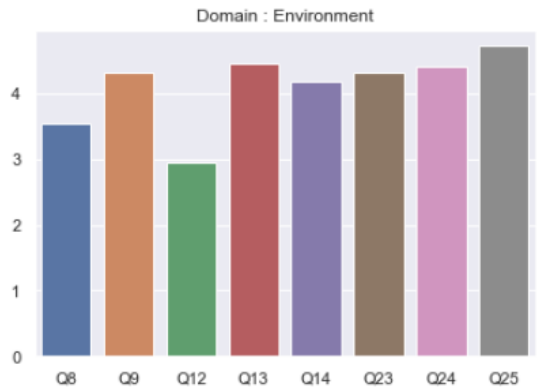


3)Social Relationship



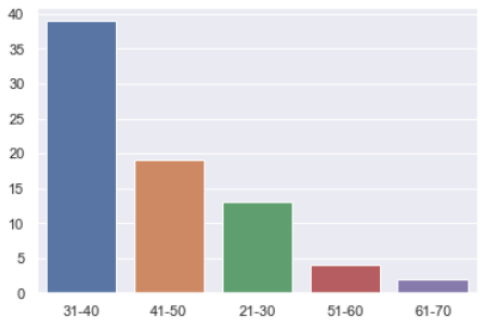
Out of 22 patients are 14 married and 8 is unmarried and both don't get support from friend hence Q22 has the least response.

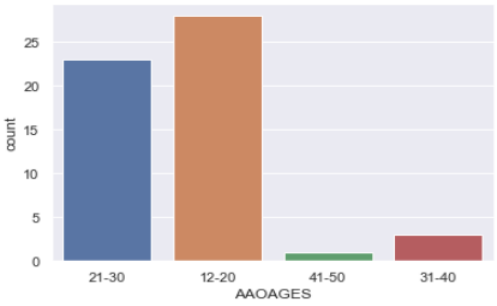
4)Environment.



Patients analysis with QOL status Neither Poor Nor Good :

Age distribution of 55 patients.





The majority of the patients started drinking in middle age.

Out of 55 patients, we can observe that all of them are above age of 21 and under the age 70.

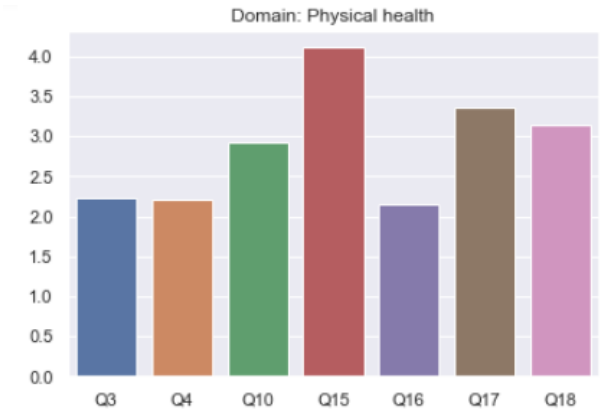
And 70% of them have only education till 10^{th}.

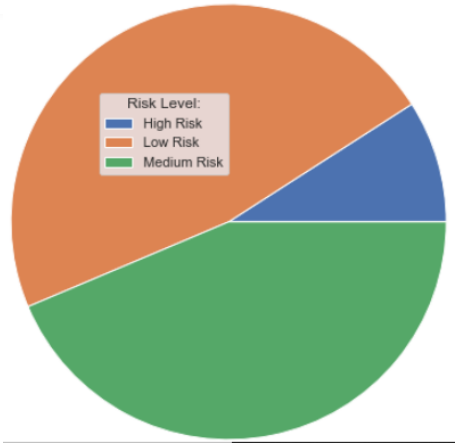
All of the patients started drinking in young and middle age hence consumed alcohol quantity is less than patients with poor QOL and more NPG i.e 13.74 years on average.

KLE Technological University
Creating Value
Leveraging Knowledge
School of Computer Science and Engineering

Domain Wise Analysis of Neither Poor Nor Good QOL Patients :

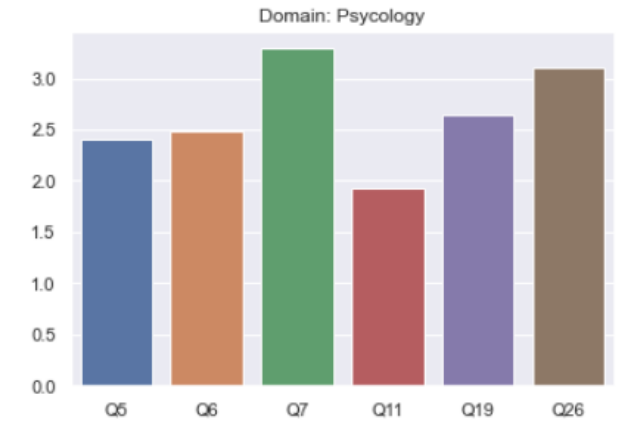
1)Physical Health

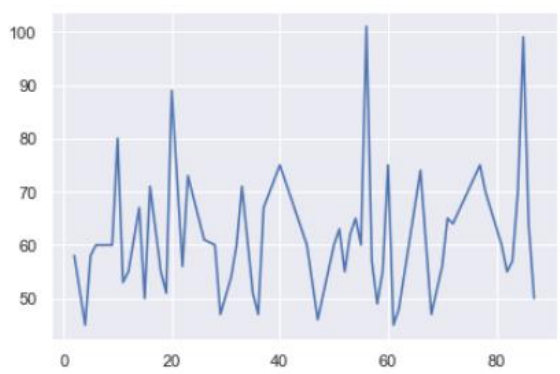




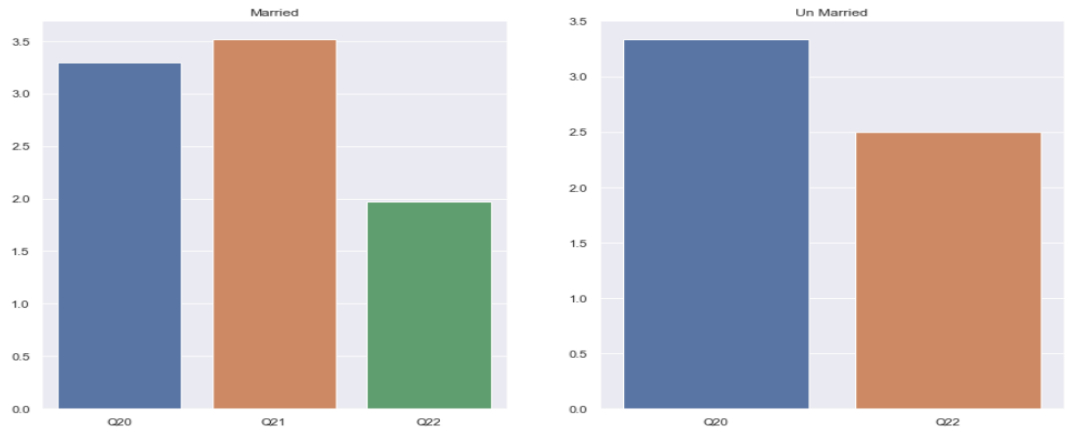
Many of the patients have medium risk.

2)Psychological.





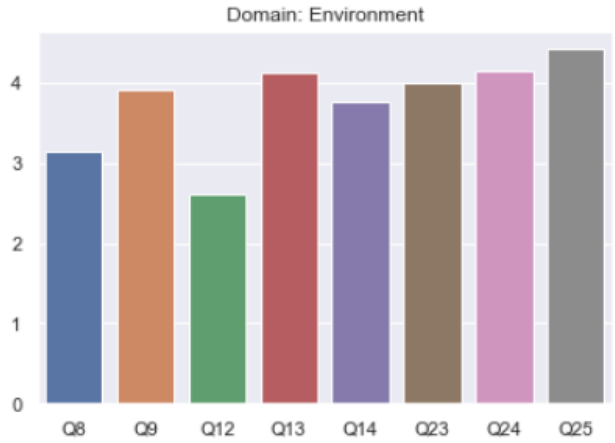
3)Social Relationship



Out of 55 patients are 37 married and 18 is unmarried and both don't get support from friend hence Q22 has the least response.

4)Environment.

50% of patients have debts.



Conclusion

Class label	Psychiatric complications	Stressors	Average alcohol consumption(in years	Withdrawal symptoms	Medical Problem experienced	If any SpecV
Poor	38%	50%	15.37	75%	12.5%	25%
Neither poor nor good	18%	38.18%	13.74	80%	20%	41%
Good	15%	36.36%	13.31	63%	10%	50%

- **Most of them have debts.**
- **More is psychiatric complications, more is stress and more is average consumption hence less is QOL.**
- **On average QOL of married patients is 53 whereas that of unmarried is 56**
- **Psychological Health is Highly Co-Related with Overall Quality of Life.**
- **Overall Quality of Life for 86 Patients falls Under Neither Poor Nor good.**