CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

VISUALIZATION ON STUDENT ACADEMIC DATASET

By Ankit Saraf

(Member at CnA Club IITG)

Viz.it

OVERVIEW

This is a flagship project by Consulting & Analytics Club of IIT Guwahati (India) to understand the relation between the grades of a first year student at IIT Guwahati with their previous background and their activities at the campus.

The project was aimed to help the next batch of incoming freshers to better prioritize their activities at campus.

The data was collected via door-to-door survey collection in all hostels by representatives of C&A Club evenly to allow for data collection in all hostels.

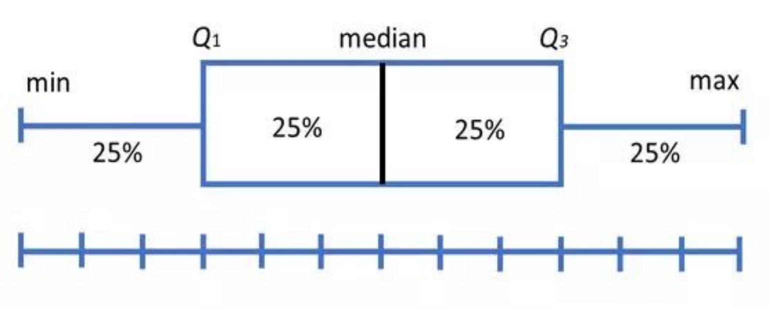
CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

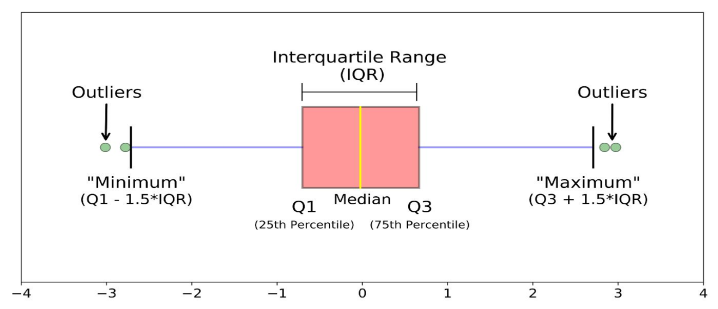
For Better Understanding

BOXPLOT:

A boxplot is a standardized way of displaying the distribution of data based on a five number summary ("minimum", first quartile (Q1), median, third quartile (Q3), and "maximum"). It can tell you about your outliers and what their values are. It can also tell you if your data is symmetrical, how tightly your data is grouped, and if and how your data is skewed.

In a box plot, we draw a box from the first quartile to the third quartile. A vertical line goes through the box at the median. The whiskers go from each quartile to the minimum or maximum.





Viz.it

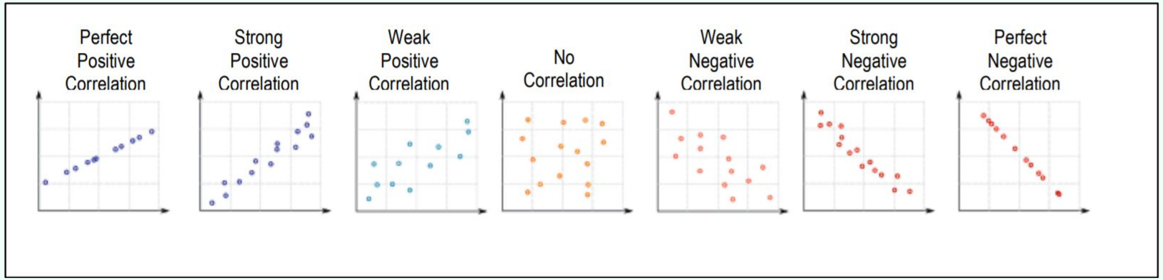
For Better Understanding

Scatterplot and correlation coefficient:

A scatterplot is a type of data display that shows the relationship between two numerical variables. Each member of the dataset gets plotted as a point whose (x, y) coordinates relates to its values for the two variables.

There are three types of correlation: positive, negative, and none (no correlation).

- Correlation is positive when the values increase together.
- Correlation is negative when one value decreases as the other increases



The correlation coefficient is a statistical measure of the strength of the relationship between the relative movements of two variables. The values range between -1.0 and 1.0.

A correlation of -1.0 shows a perfect negative correlation, while a correlation of 1.0 shows a perfect positive correlation. A correlation of 0.0 shows no linear relationship between the movement of the two variables.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

For Better Understanding

Significance Level (alpha):

The significance level, also denoted as alpha or α, is the probability of rejecting the null hypothesis when it is true.

P-value:

P-values are the probability of obtaining an effect at least as extreme as the one in your sample data, assuming the truth of the null hypothesis.

Null Hypothesis:

A null hypothesis is a type of hypothesis used in statistics that proposes that there is no difference between certain characteristics of a population.

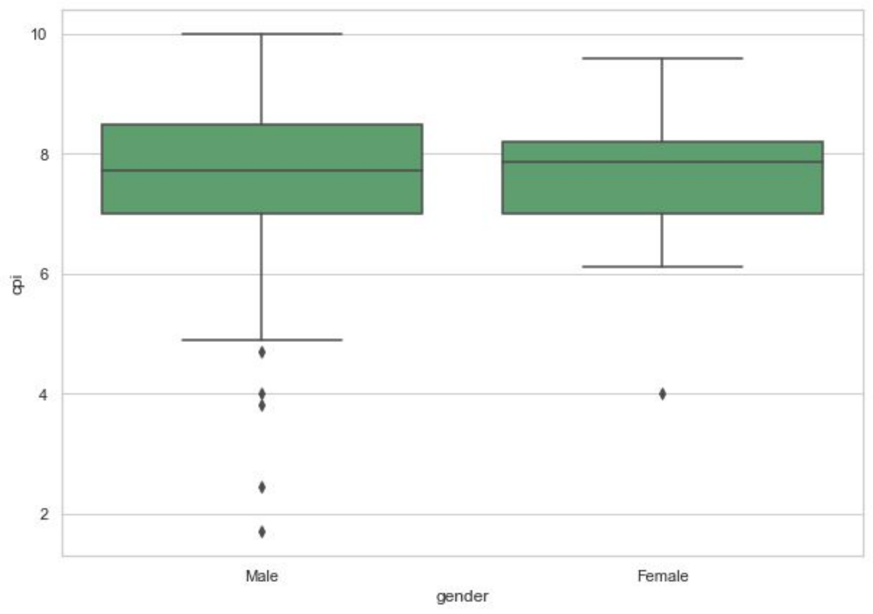
*When a P-value is less than or equal to the significance level, you reject the null hypothesis.

*Mean(i)=Mean(j) : Here i and j refers to the category of given feature of our dataset and i is not equal j.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Gender



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.968234111202742

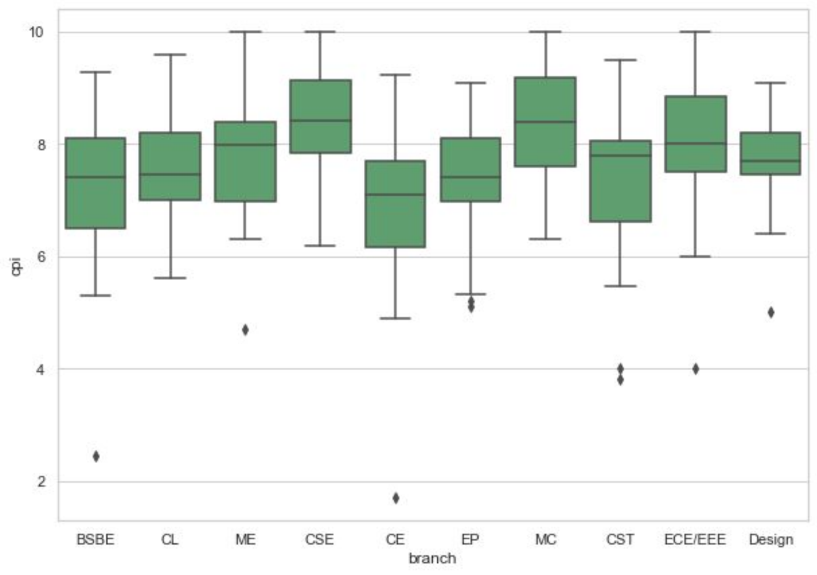
Conclusion: We accept null hypothesis stated.

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Branch



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

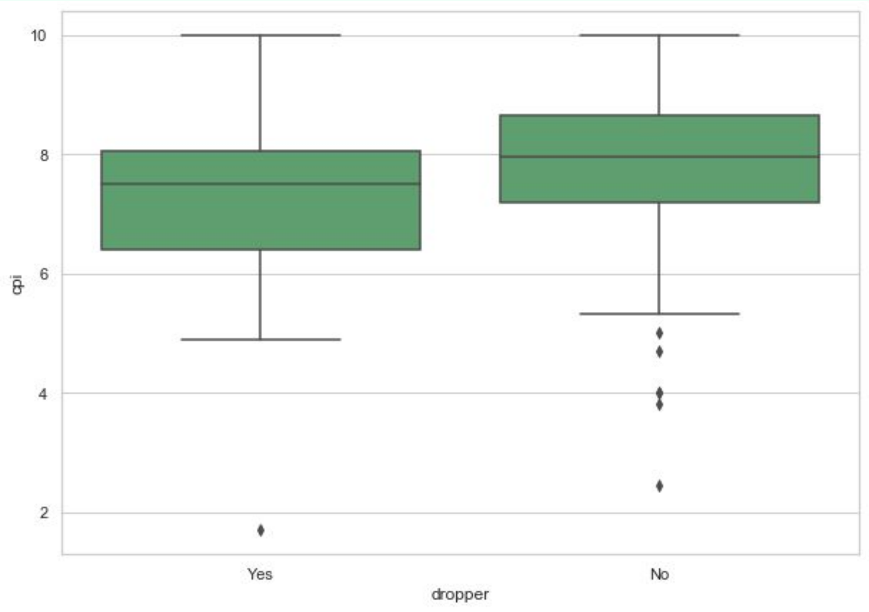
Observed p-value: 1.3773593380632256e-10

Conclusion: We reject null hypothesis stated.

Viz.it

UNDERSTANDING RELATION

CPI vs Dropper



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

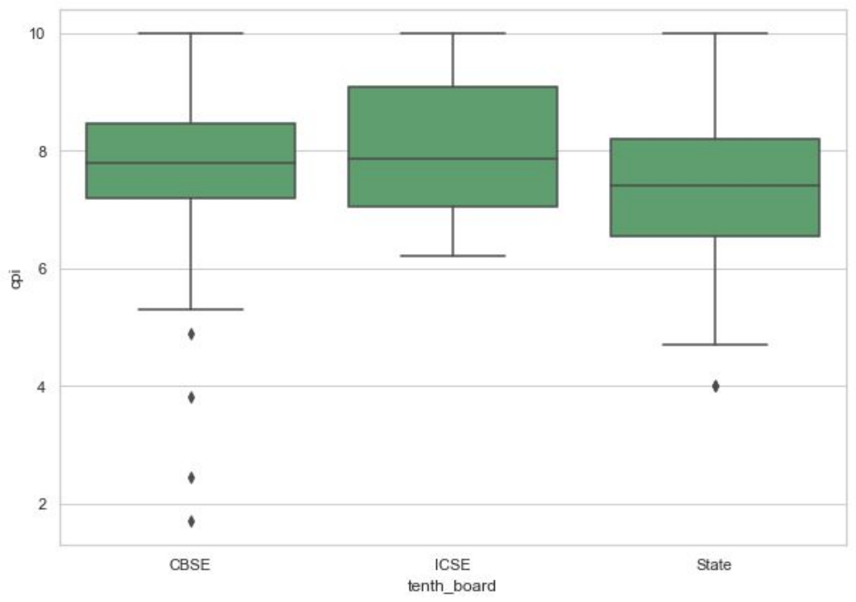
Observed p-value: 4.714740397283642e-06

Conclusion: We reject null hypothesis stated.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs 10th Board



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.008727675769859323

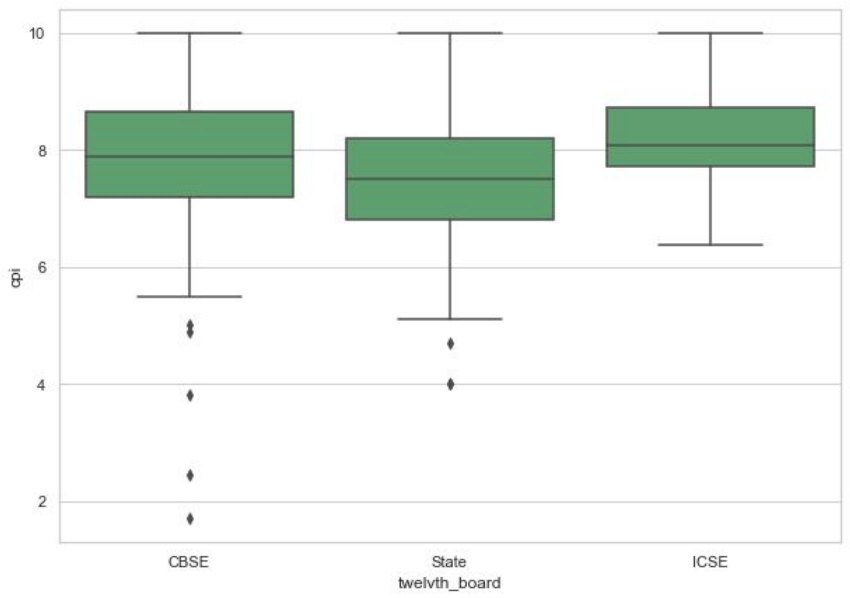
Conclusion: We reject null hypothesis stated.

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs 12th Board



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

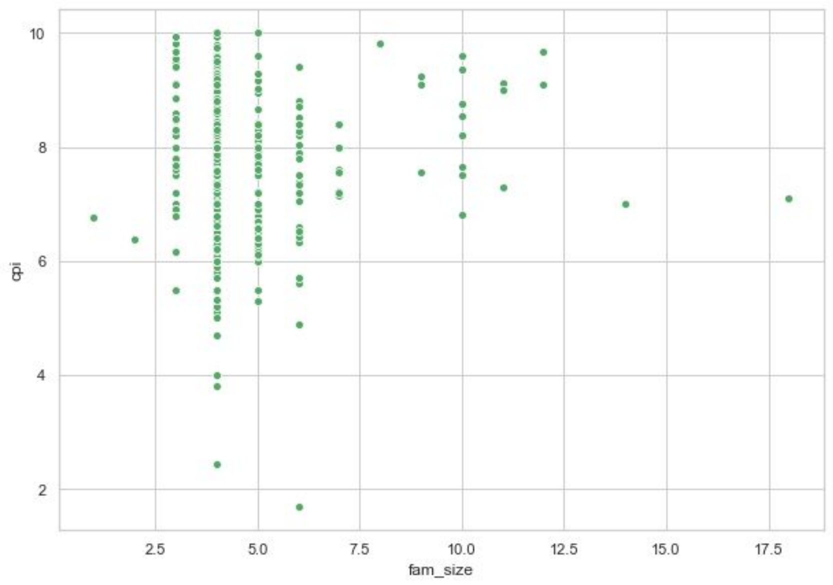
Observed p-value: 0.026278147829353983

Conclusion: We reject null hypothesis stated.

Viz.it

UNDERSTANDING RELATION

CPI vs Family Size



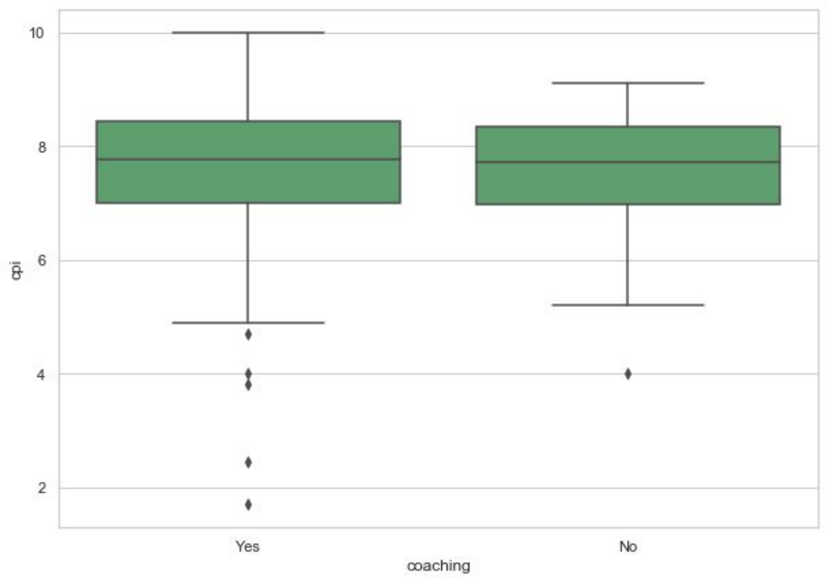
Correlation coefficient : 0.0496

Conclusion : Since correlation coefficient is very close to 0 in comparison to 1, we can conclude there is no considerable correlation between CPI and family size.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Coaching status



Null Hypothesis: Mean(i) = Mean(j)

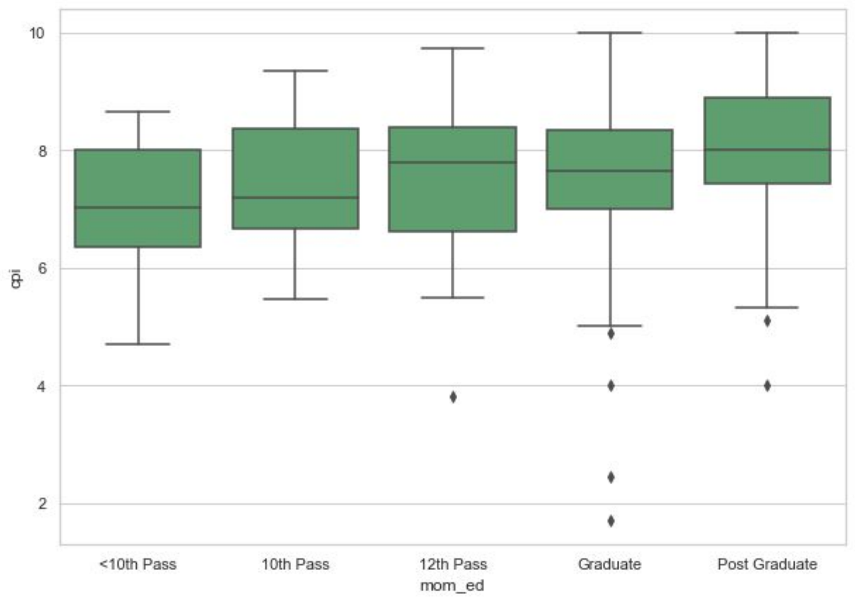
Significance level (alpha): 0.05

Observed p-value: 0.33370607329430335

Conclusion: We accept null hypothesis stated.

UNDERSTANDING RELATION

CPI vs Mom's Education



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

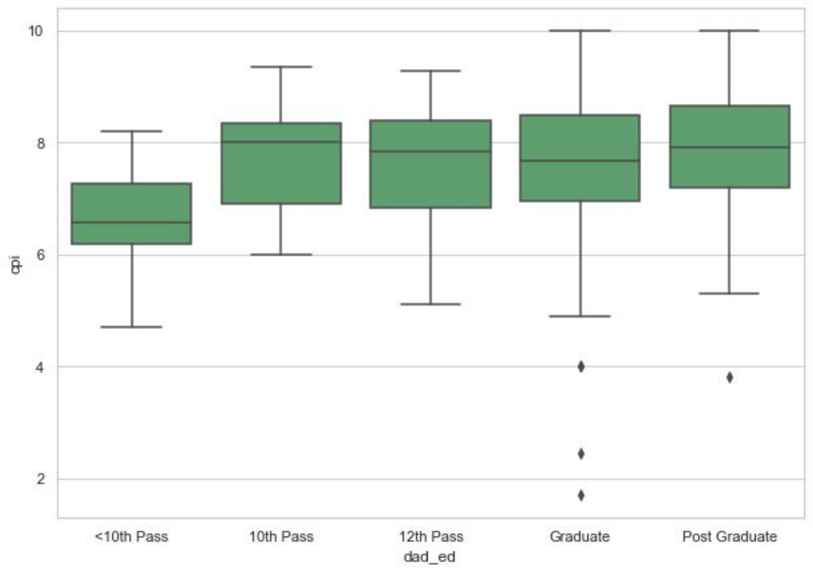
Observed p-value: 0.00024189556554582718

Conclusion: We reject null hypothesis stated.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Dad’s education



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.001966715164184643

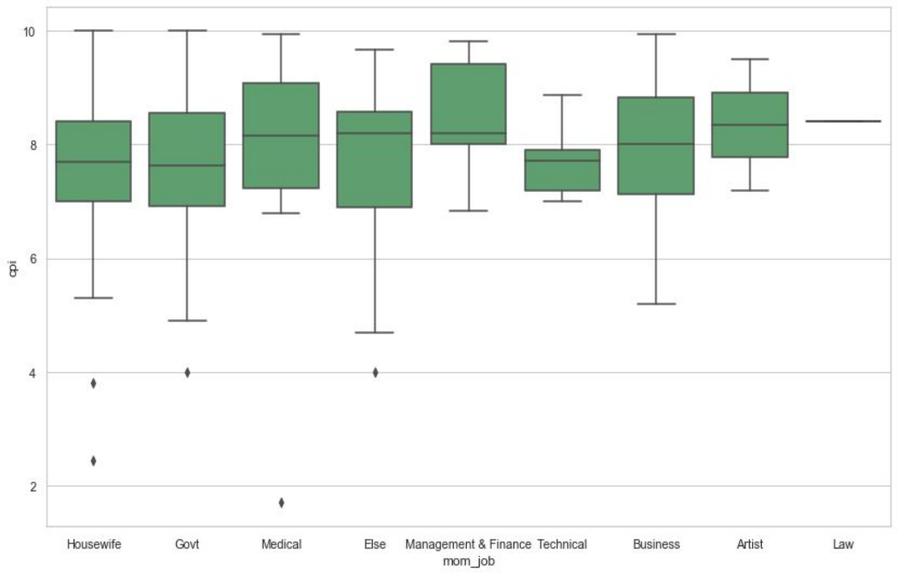
Conclusion: We reject null hypothesis stated.

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Mom’s job



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.8272254632917686

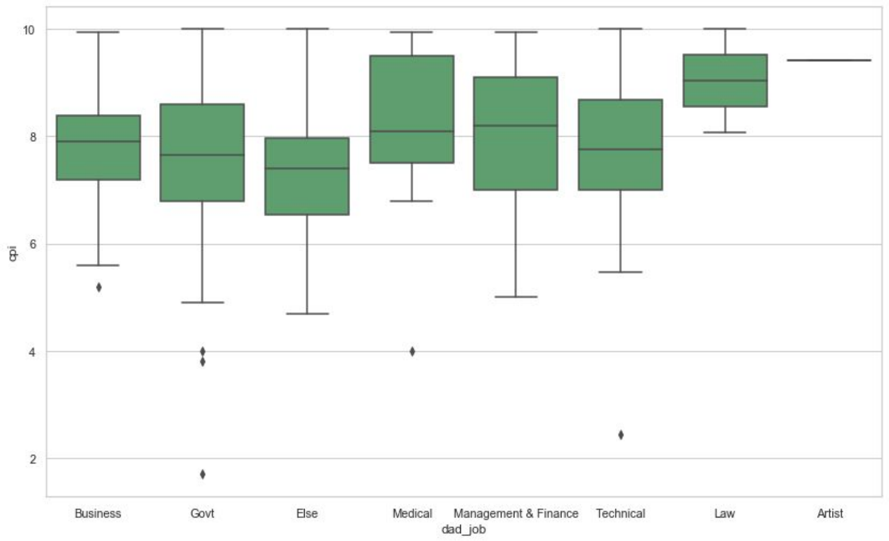
Conclusion: We accept null hypothesis stated.

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Dad’s job



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.0100583928301829

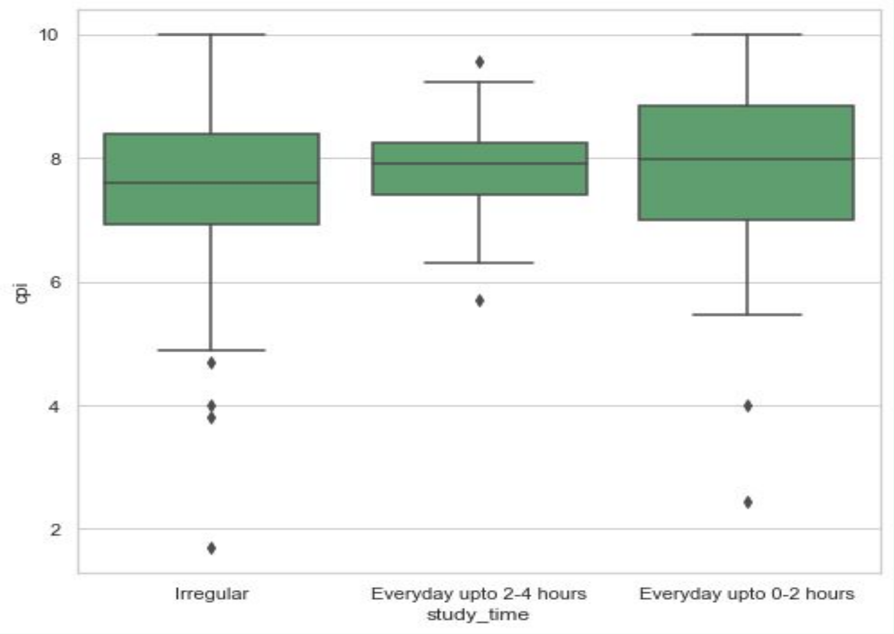
Conclusion: We reject null hypothesis stated.

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Study time



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.13599960399194846

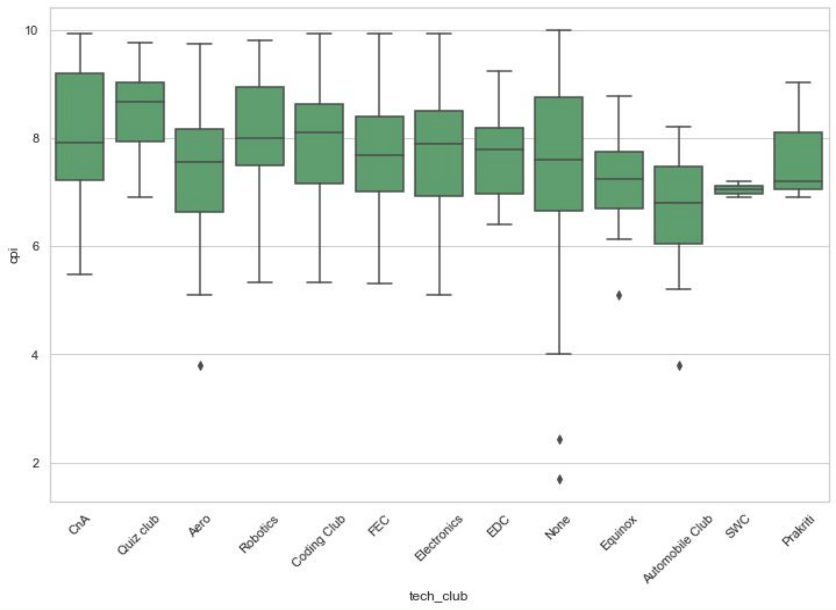
Conclusion: We accept null hypothesis stated.

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Technical Club Joined



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

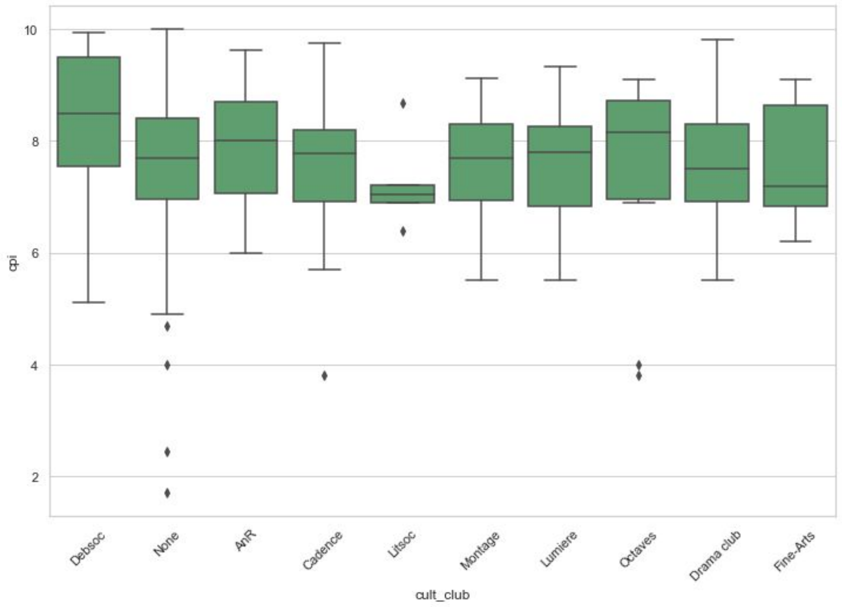
Observed p-value: 0.0009018612847080804

Conclusion: We reject null hypothesis stated.

Viz.it

UNDERSTANDING RELATION

CPI vs Cultural Club Joined



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

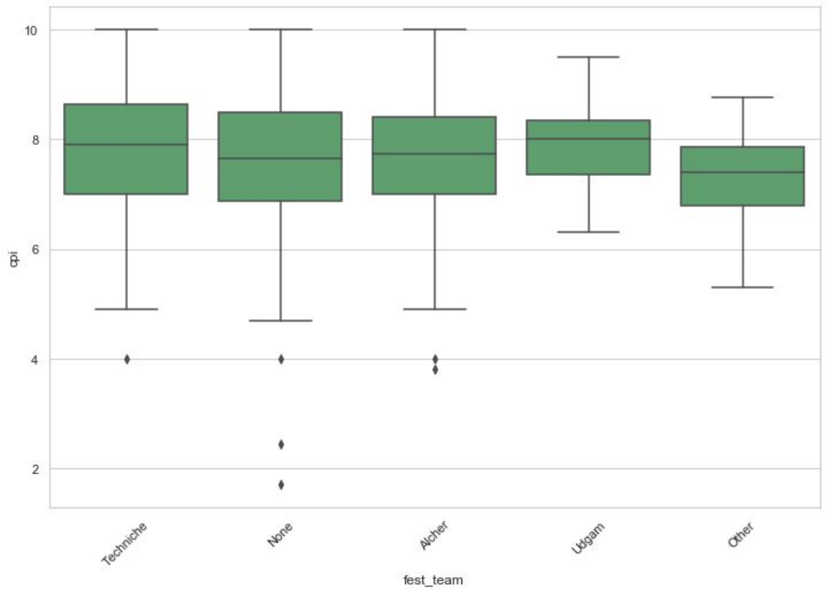
Observed p-value: 0.20803444379055766

Conclusion: We accept null hypothesis stated.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Fest Team Joined



Null Hypothesis: Mean(i) = Mean(j)

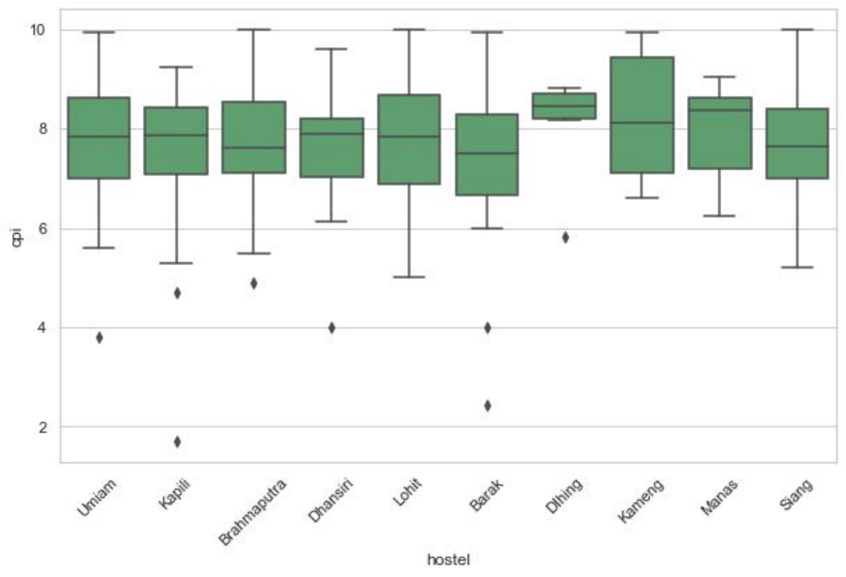
Significance level (alpha): 0.05

Observed p-value: 0.21803184286550825

Conclusion: We accept null hypothesis stated.

UNDERSTANDING RELATION

CPI vs Hostel



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.6749315180898611

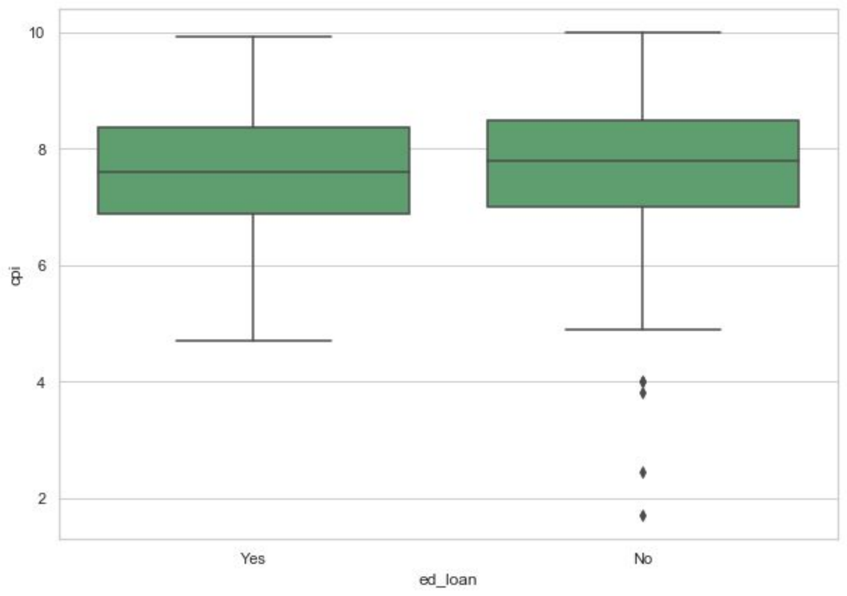
Conclusion: We accept null hypothesis stated.

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Education Loan Opted (Y/N)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.5651610858429756

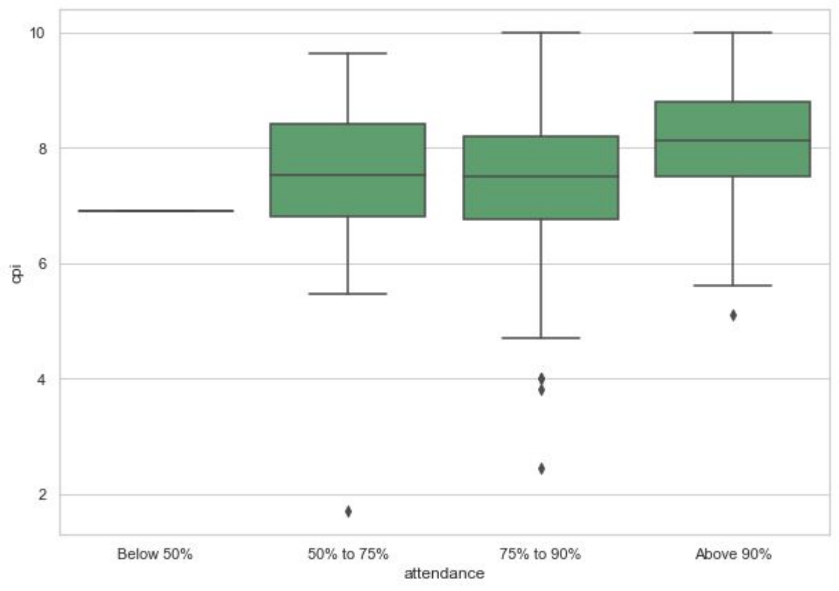
Conclusion: We accept null hypothesis stated.

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Attendance



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

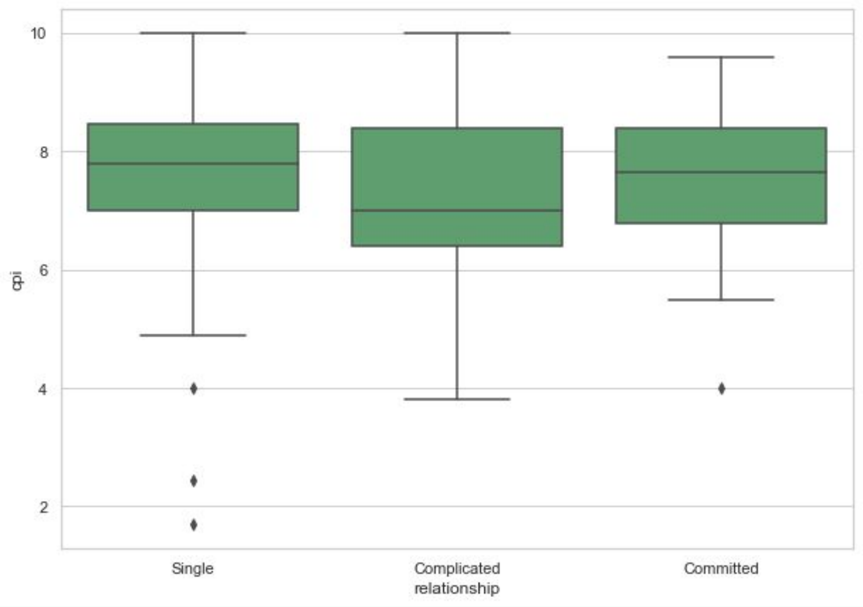
Observed p-value: 7.039252412297359e-06

Conclusion: We reject null hypothesis stated.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Relationship Status



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.1280755821745034

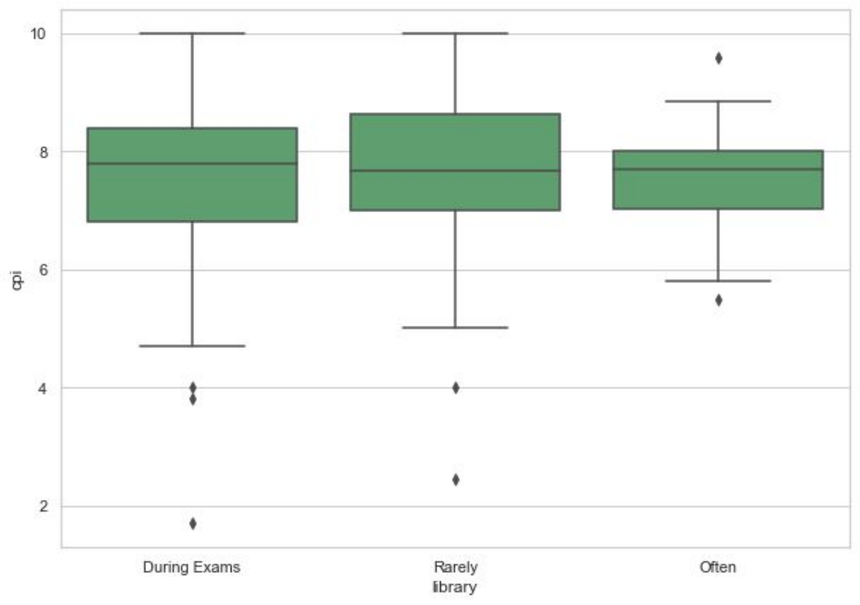
Conclusion: We accept null hypothesis stated.

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Frequency of visiting library



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.5059632258696345

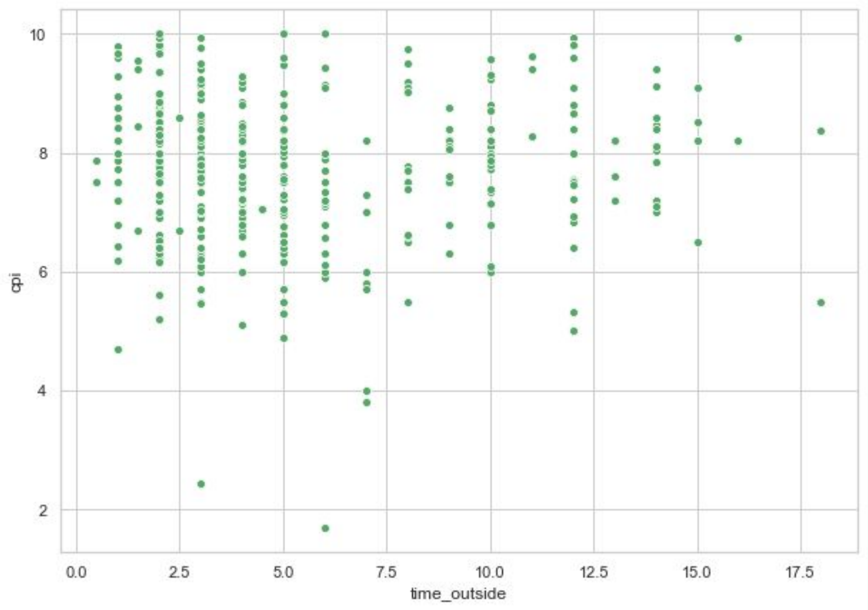
Conclusion: We accept null hypothesis stated.

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Time spent outside room (In hrs.)



Correlation coefficient : 0.0205

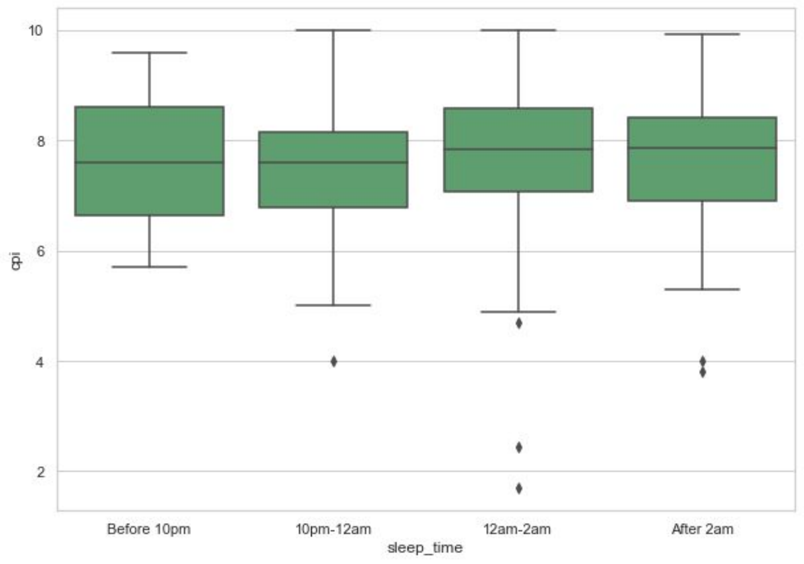
Conclusion: Since correlation coefficient is very close to 0 in comparison to 1, we can conclude there is no considerable correlation between CPI and time spent outside room.

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Time of sleep



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

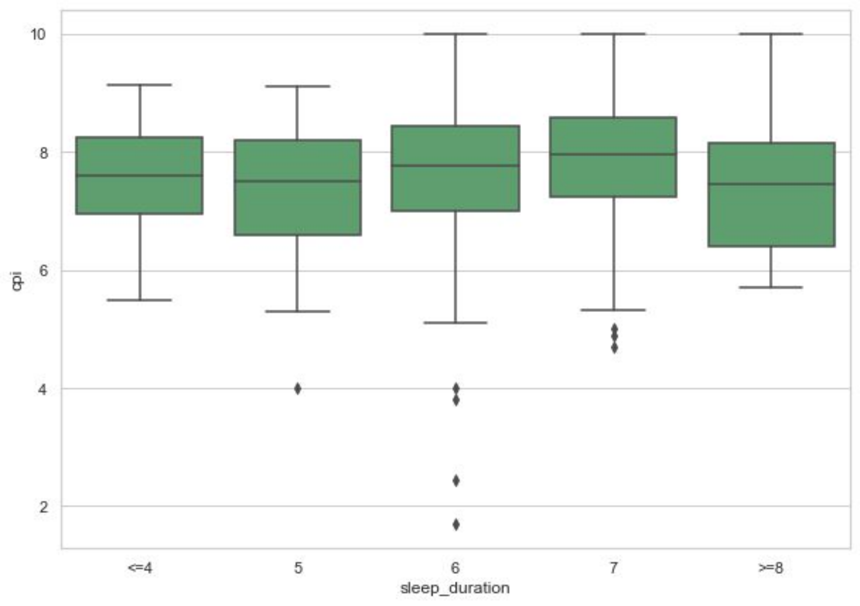
Observed p-value: 0.2849108987006208

Conclusion: We accept null hypothesis stated.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Sleep Duration (hrs)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.11043696612894544

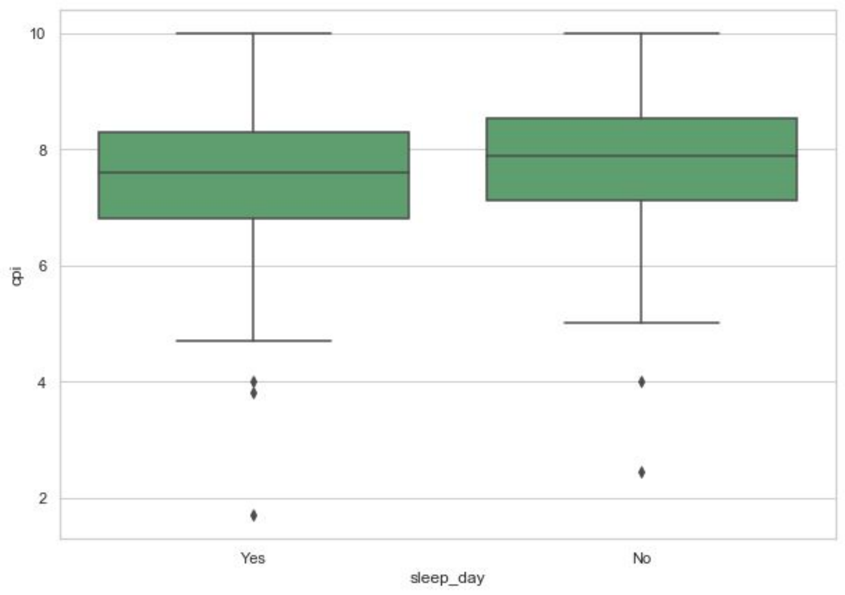
Conclusion: We accept null hypothesis stated.

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Day sleep habit (Y/N)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

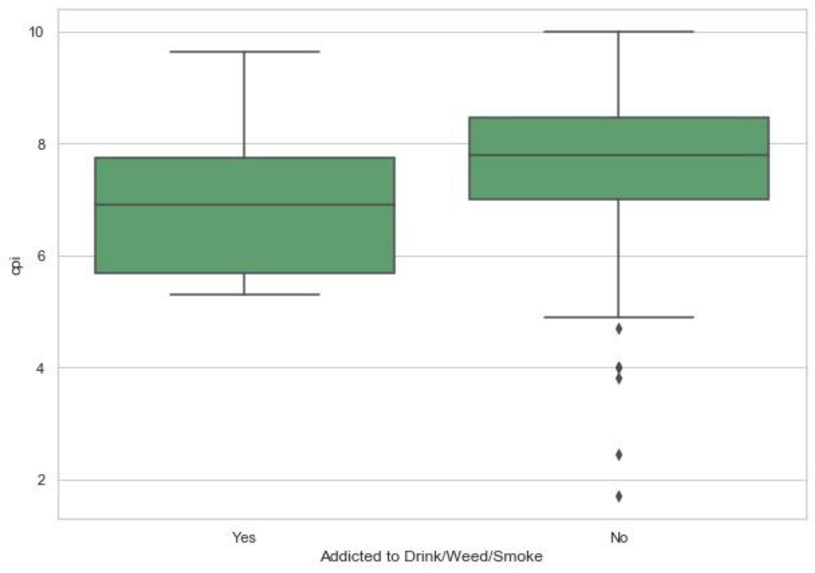
Observed p-value: 0.027688686645234557

Conclusion: We reject null hypothesis stated.

Viz.it

UNDERSTANDING RELATION

CPI vs Addiction (Y/N)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.007966653316786974

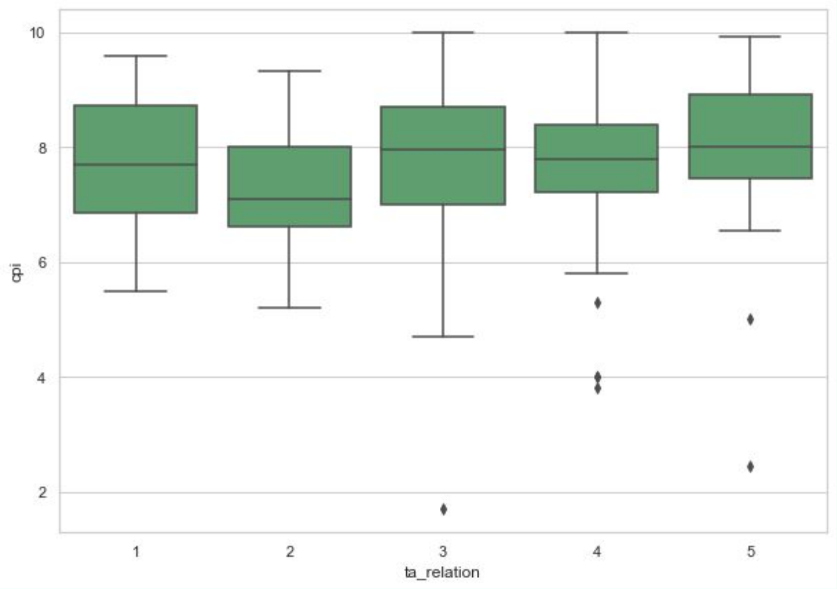
Conclusion: We reject null hypothesis stated.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Relation with TA

(1 - Very bad | 5 - Excellent)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.025002433721025813

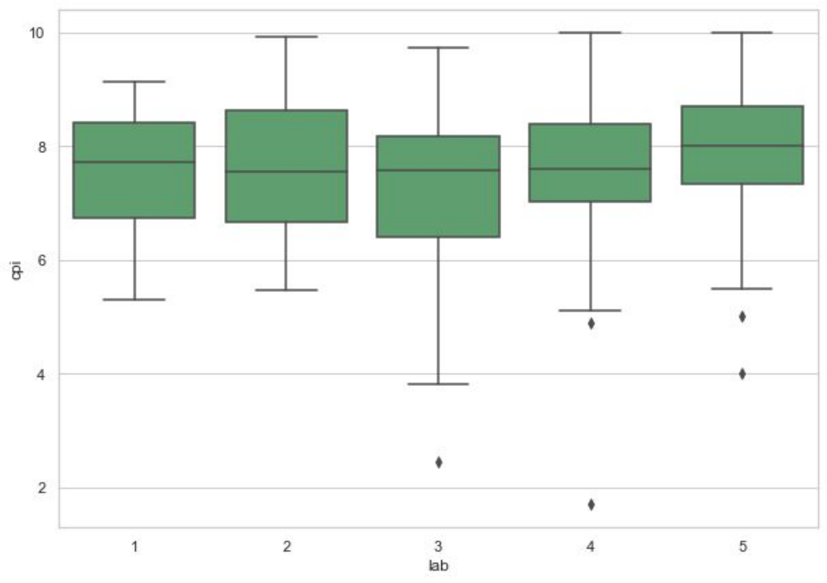
Conclusion: We reject null hypothesis stated.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Seriousness towards lab

(1 - Not at all serious | 5 - Very serious)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.0037498899884395023

Conclusion: We reject null hypothesis stated.

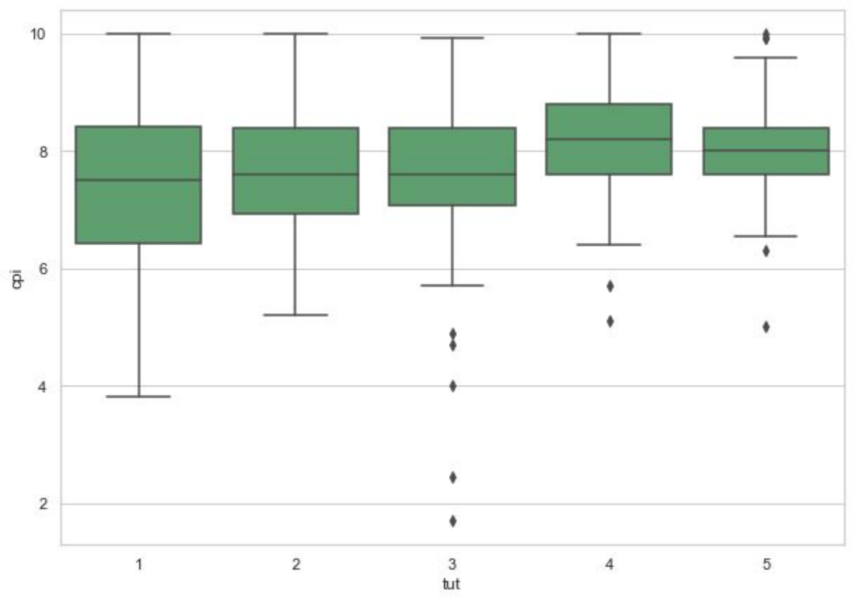
Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Seriousness towards tutorials

(1 - Not at all serious | 5 - Very serious)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.0016441563380684317

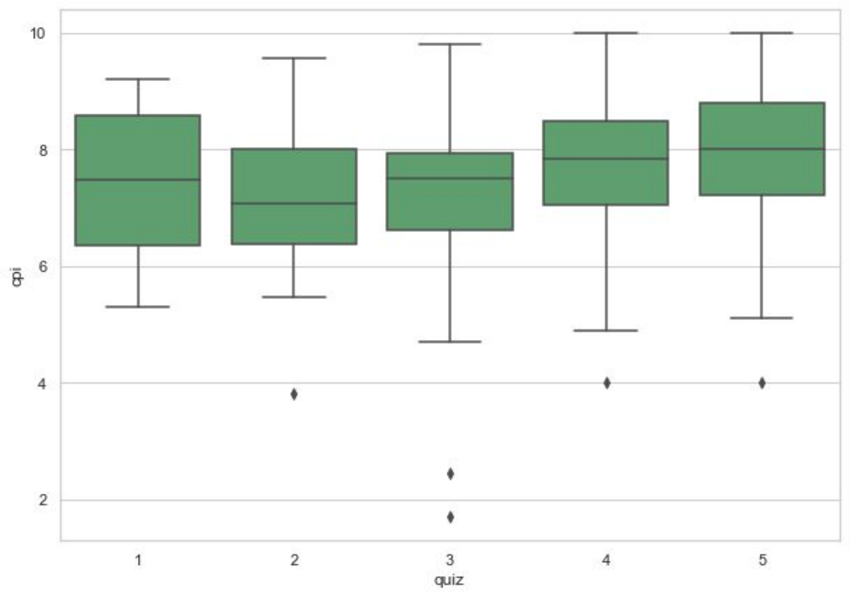
Conclusion: We reject null hypothesis stated.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Seriousness towards Quiz

(1 - Not at all serious | 5 - Very serious)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 8.153870239333924e-05

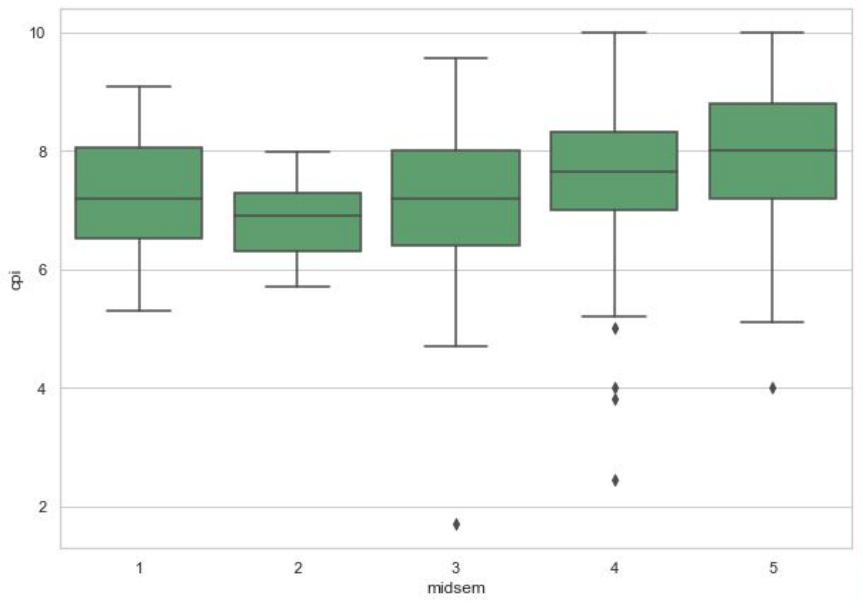
Conclusion: We reject null hypothesis stated.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Seriousness towards Midsem

(1 - Not at all serious | 5 - Very serious)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 2.2121348813022917e-05

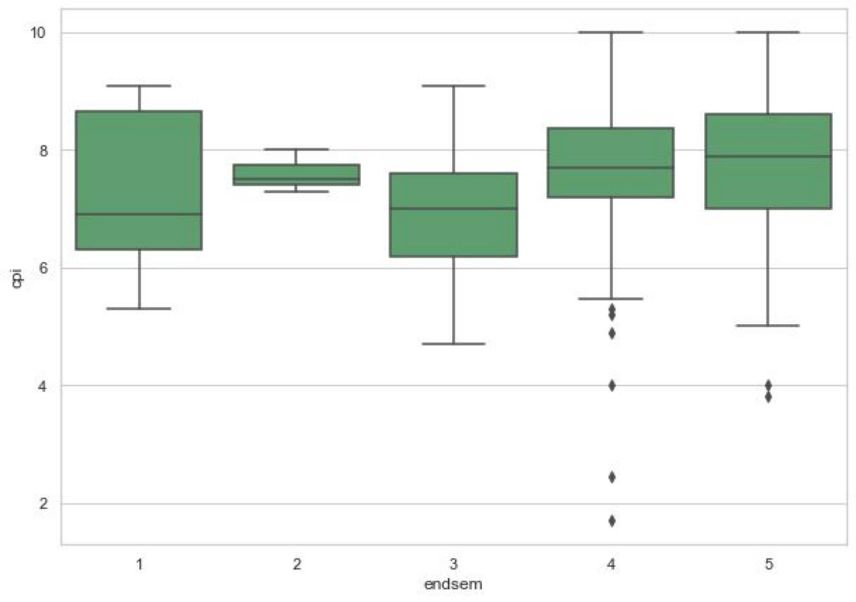
Conclusion: We reject null hypothesis stated.

Viz.it

UNDERSTANDING RELATION

CPI vs Seriousness towards End-sem

(1 - Not at all serious | 5 - Very serious)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

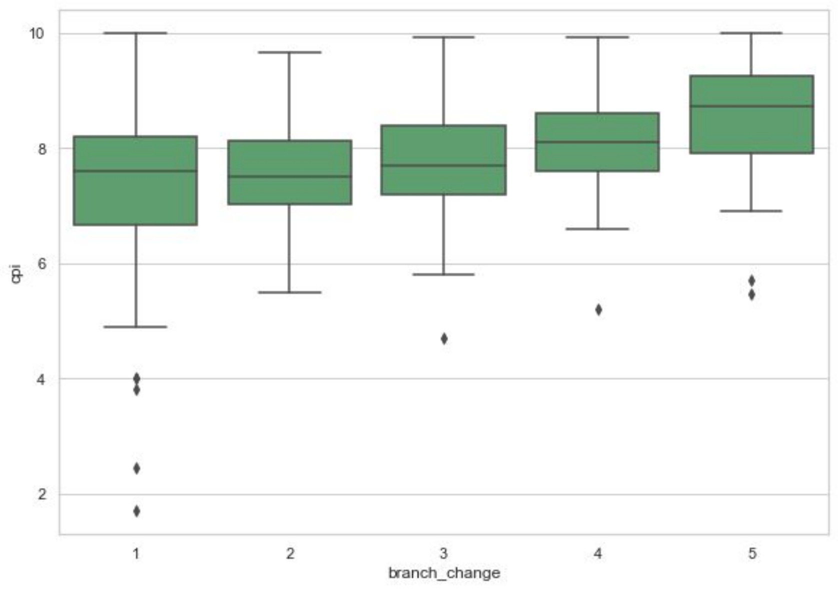
Observed p-value: 0.001958529074287955

Conclusion: We reject null hypothesis stated.

UNDERSTANDING RELATION

CPI vs Seriousness towards Branch Change

(1 - Not at all serious | 5 - Very serious)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 1.7358915137558518e-06

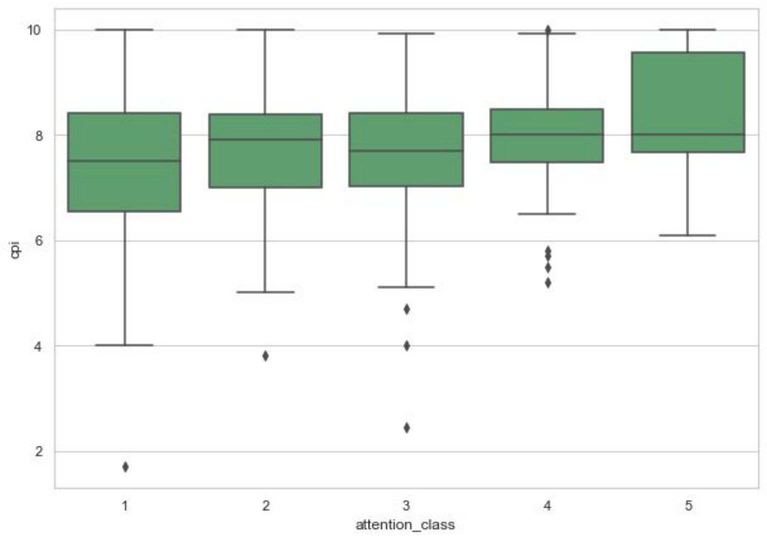
Conclusion: We reject null hypothesis stated.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Attention in class

(1 - Least attention | 5 - Complete attention)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

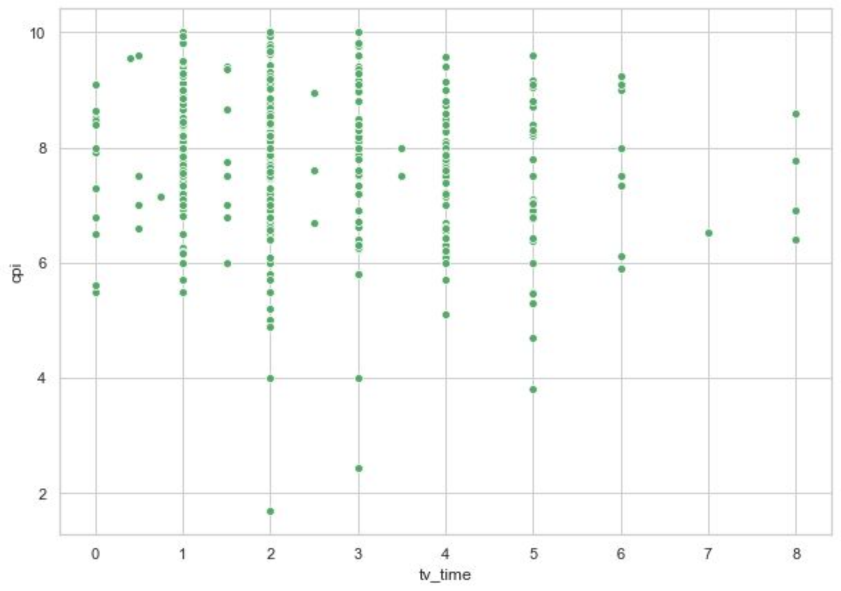
Observed p-value: 0.07748973952915558

Conclusion: We accept null hypothesis stated.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Time devoted in TV



Correlation coefficient : -0.103

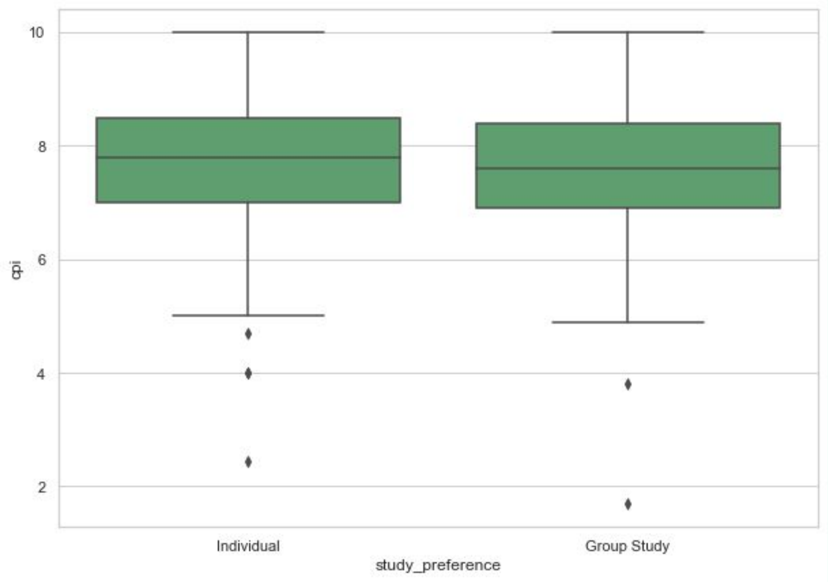
Conclusion : Since correlation coefficient is very close to 0 in comparison to -1, we can conclude there is no considerable correlation between CPI and time spent watching TV(includes movies, webseries and other entertainment media.).

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Study preference



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.1694947551679758

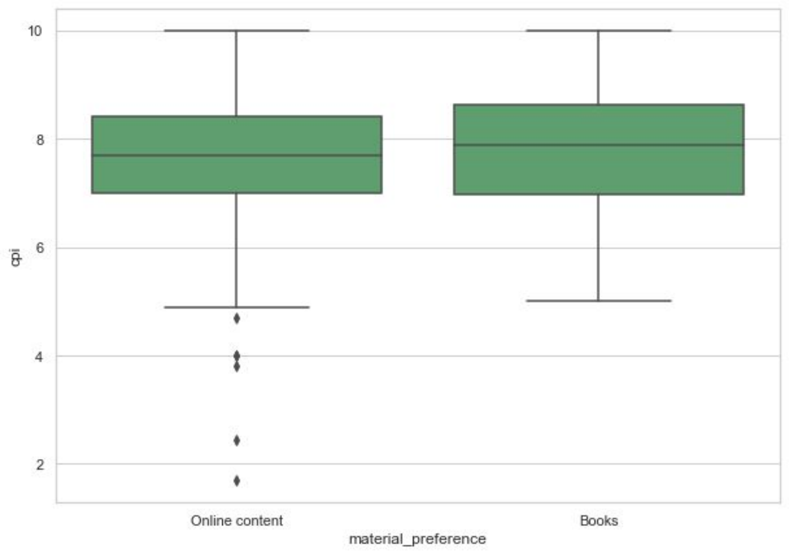
Conclusion: We accept null hypothesis stated.

Viz.it

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Material Preference



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.48268194269187004

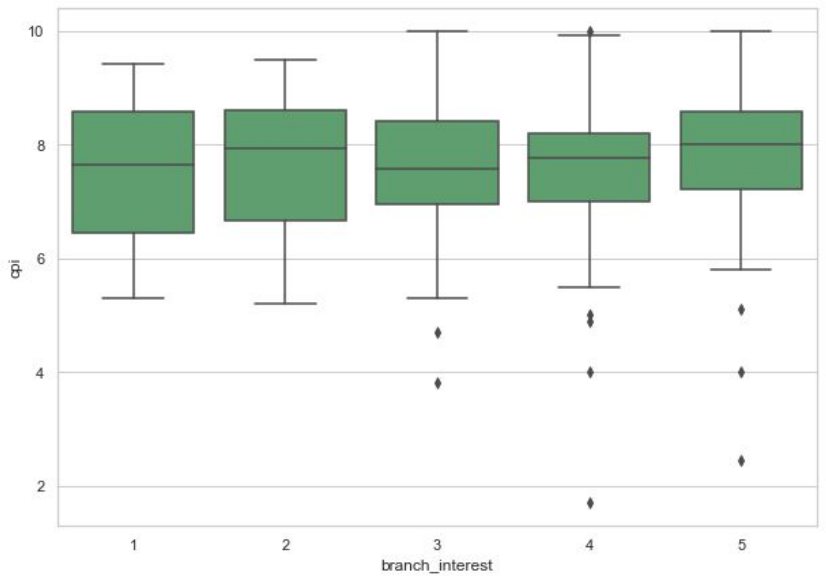
Conclusion: We accept null hypothesis stated.

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

UNDERSTANDING RELATION

CPI vs Interest in own Branch

(1 - No interest | 5 - Very much interest)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

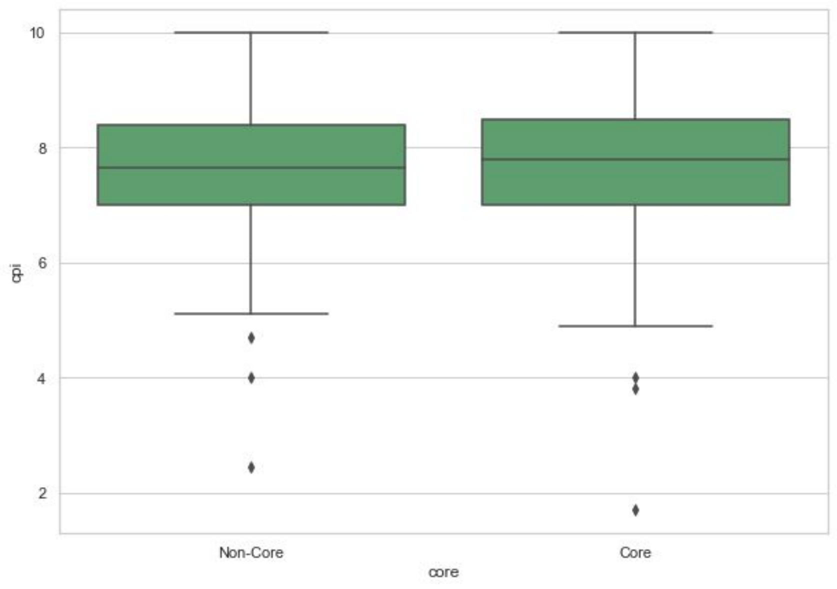
Observed p-value: 0.6618033614399144

Conclusion: We accept null hypothesis stated.

Viz.it

UNDERSTANDING RELATION

CPI vs Branch (Core/non-Core)



Null Hypothesis: Mean(i) = Mean(j)

Significance level (alpha): 0.05

Observed p-value: 0.8185533127425892

Conclusion: We accept null hypothesis stated.

PREDICTIVE MODELLING WITH MACHINE LEARNING

Keeping in mind the hypothesis test, we included following features in our model:

- Branch
- Dropper
- 10th Board
- 12th Board
- Mom_ed
- Dad_ed
- Dad_job
- Attendance
- Day sleep habit
- Addiction
- Relation with TA
- Seriousness towards lab
- Seriousness towards tutorials
- Seriousness towards quiz
- Seriousness towards midsem
- Seriousness towards end-sem
- Interest in own branch

PREDICTIVE MODELLING WITH MACHINE LEARNING

After trying various regression models the regression model that fits best was the Support Vector Regression.

Evaluation of model performance:

- Mean Square Error = 0.854
- Root Mean Square Error = 0.924

The Support Vector Regression Model can be accessed at :

https://drive.google.com/open?id=10LNHmW2RTnD3E1BnyY_H8O0ZN384AHry

CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

THANK YOU!