

|  |  |
| --- | --- |
| ***Module Title:*** | *DATA VISUALIZATION* |
| ***Module Code:*** | *B9DA106* |
| ***Module Leader:*** | *Basel Magableh* |
| ***Assessment Title:*** | *Exploratory Data Analysis* |
| ***Assessment Type:*** | *Group (3)* |
| ***Student Names:*** | *Devadirajan Thirumalai* |
| ***Student ID:*** |  |
| ***Student Names:*** | *Baisha B* |
| ***Student ID:*** | *10385217* |
| ***Student Names:*** | *GaneshKumar Lingappan* |
| ***Student ID:*** | *10394521* |

**Contents**

[Introduction 4](#_Toc7825722)

[DATASET SELECTION 4](#_Toc7825723)

[Visualization tool 4](#_Toc7825724)

[INITIAL ANALYSIS QUESTION 4](#_Toc7825725)

[DISCOVERY AND INSIGHTS 5](#_Toc7825726)

[DESCRIPTION for each visualization 5](#_Toc7825727)

[**PIEChart analysis for Schools under Gender category** 5](#_Toc7825728)

[**Stackbar PLOT Analysis under grade count under gender category** 6](#_Toc7825729)

[**bar plot analysis about the student alcohol consumption on the weekend** 6](#_Toc7825730)

[**Bubble graph analysis about the comparsion of STUDY TIME VS GRADE** 8](#_Toc7825731)

[**DENSITY PLOT analysis about the HEALTH BY DAILY ALCOHOL** 8](#_Toc7825732)

[**scatter Plot analysis on the GRADE VARIATION BASED ON THE STUDY TIME, TRAVEL TIME, ABSENCE** 9](#_Toc7825733)

[**Tabular analysis about the details of the STUDENTS (School name, gender, age) with grade condition** 10](#_Toc7825734)

[**Bar Graph analysis about the student Alcohol consumption during the weekdays** 11](#_Toc7825735)

[**Dashboard of all the Graph** 12](#_Toc7825736)

[CONCLUSION 12](#_Toc7825737)

# Introduction

This data approach student achievement in secondary education of two Portuguese schools. The data attributes include student grades, demographic, social and school related features) and it was collected by using school reports and questionnaires. Two datasets are provided regarding the performance in two distinct subjects: Mathematics (mat) and Portuguese language (por).

In [Cortez and Silva, 2008], the two datasets were modeled under binary/five-level classification and regression tasks. Important note: the target attribute G3 has a strong correlation with attributes G2 and G1. This occurs because G3 is the final year grade (issued at the 3rd period), while G1 and G2 correspond to the 1st and 2nd period grades. It is more difficult to predict G3 without G2 and G1, but such prediction is much more useful (see paper source for more details).

# DATASET SELECTION

This data approach student achievement in secondary education of two Portuguese schools. The data attributes are student grades, demographic, social and school related features) and it was collected by using school reports and questionnaires.

Dataset link: [https://archive.ics.uci.edu/ml/datasets/student+performance#](https://archive.ics.uci.edu/ml/datasets/student+performance)

Attributes: school, sex, age, address, famsize, Pstatus, Medu, Fedu, Mjob, Fjob, reason, guardian, traveltime, studytime, failures, schoolsup, famsup, paid, activities, nursery, higher, internet, romantic, famrel, freetime, goout, Dalc, Walc, health, absences.

# Visualization tool

After collected the data, we analysis the dataset and come up the tool which will most effective and interactive visualization. The tool which we choose is Jupyter notebook and we have used Altair, Plotly, matplotlib, seaborn, folium, cufflinks packages in the python to achieve the result which we are expecting.

**Github Link:** <https://github.com/10504925/StudentPerformanceAnalysis>

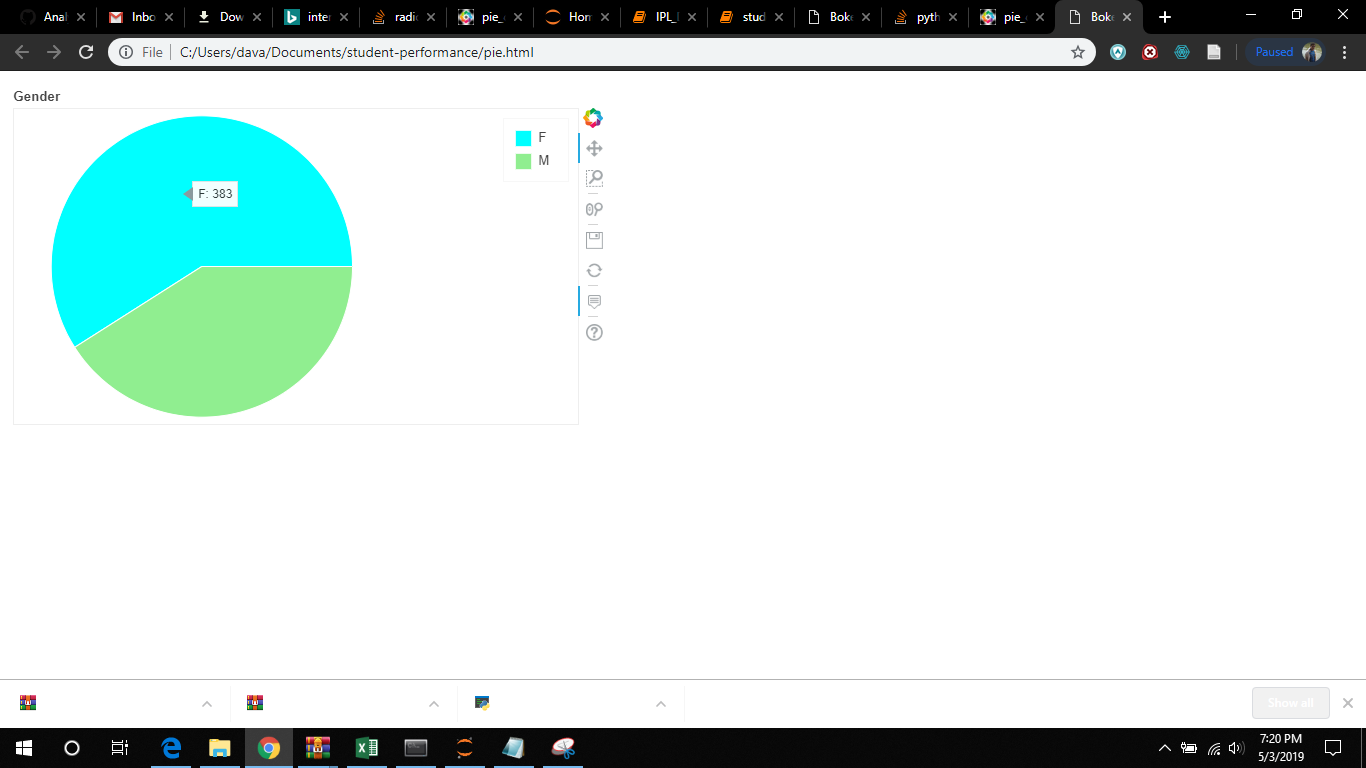
# INITIAL ANALYSIS QUESTION

1. Who is getting more grade in the exam under Gender category?
2. When during the weekday the alcohol consumption is more for students?
3. How much of study time they are spending?
4. How many days students are absent for the class?

# DISCOVERY AND INSIGHTS

# DESCRIPTION for each visualization

## **PIEChart analysis for Schools under Gender category**



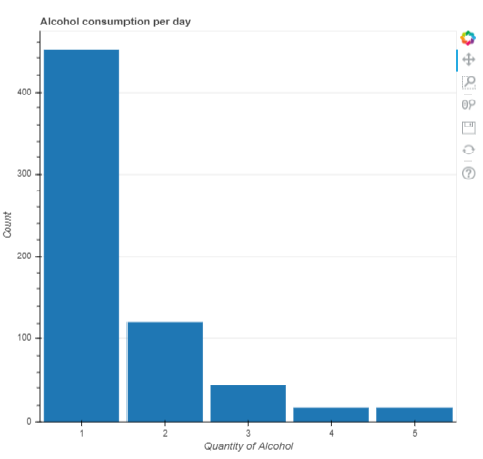
This plot gives the Overview/Summary of the difference in gender ratio of the total number of male and female students in a Portuguese school. This pie chart shows us the female rate is 383 and the male rate is 234

## **Stackbar PLOT Analysis under grade count under gender category**



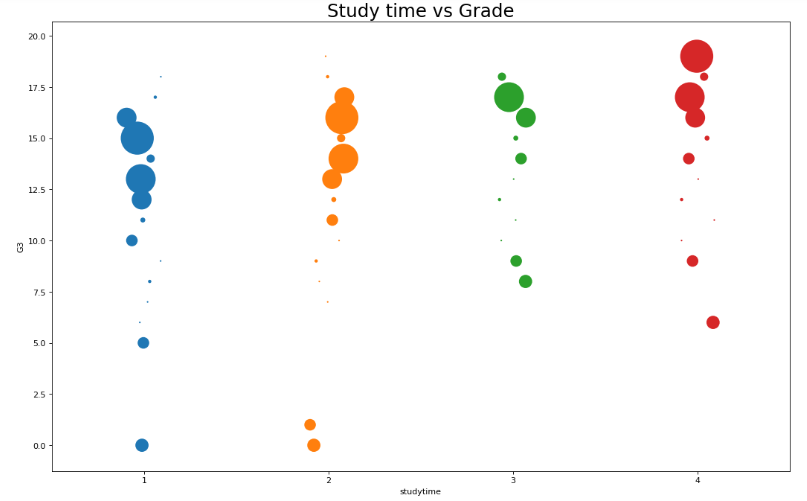
This bar graph shown below represents the grade counts by sex is in the series of 0 to 19 and describes the number of male and female percentage is measured while in the range of 11 and 13 grade the female measure is around 52 percentage which is shaded in the color blue and the male section is around 40 to 50 percentage compared with the grades of 10 and 11.

## **bar plot analysis about the student alcohol consumption on the weekday**



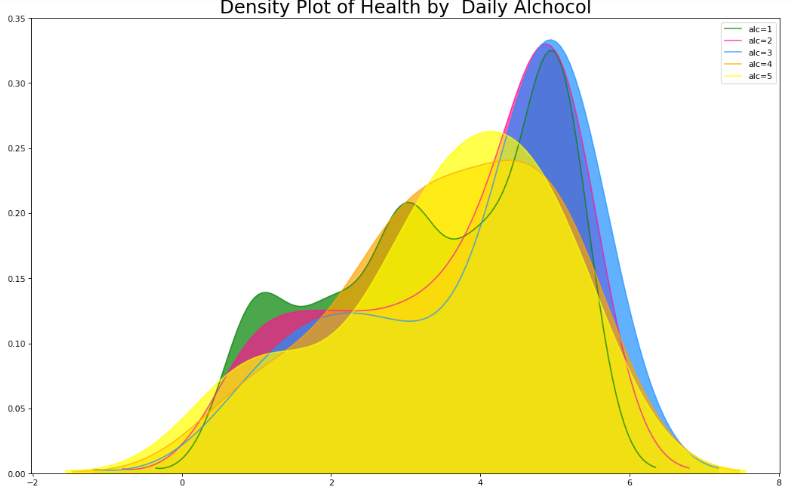
This bar graph represents the alcohol consumption of students per day in a week while it is compared with the range of series that is on Monday to Friday we discuss that there is a huge increase of alcohol consumption up to 450 on Monday and gradually decreases to 50 percentage on Friday. In the above bar chart, x- axis gives the details of the days – (1 – Monday, 2 – Tuesday, 3 – Wednesday, 4 – Thursday, 5 – Friday) and y-axis gives the details of the number of student consuming alcohol during the weekdays.

## **Bubble graph analysis about the comparsion of STUDY TIME VS GRADE**



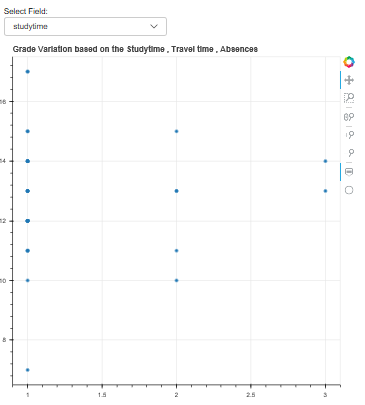
This graph status that the study time vs Grade discuss about how much the students spend their time in studies with a good grade of percentage is used defined in the below graph as in the 4th range the grade of red is very high in accuracy.

## **DENSITY PLOT ANALYSIS ABOUT the HEALTH BY DAILY ALCOHOL**



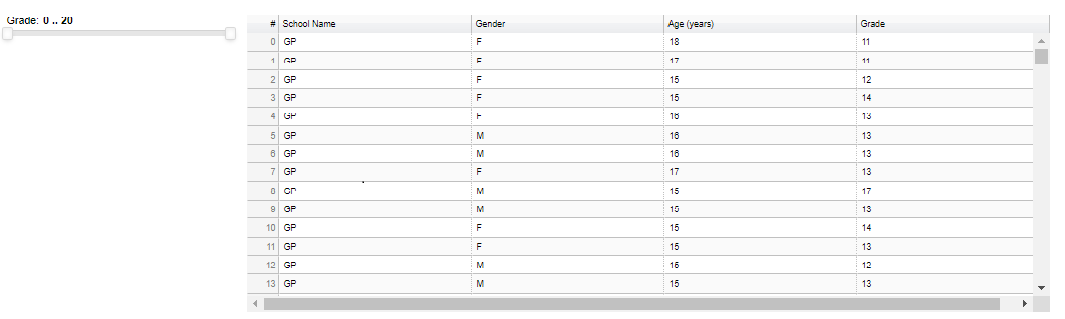
The visualization in this graph shown in different colors represent the ratio of the alcohol consumed by the density plot ranging from 0 to 8 and with the highest value of 0.33 percentage high the students are been addicted to the also in the Portuguese school.

## **scatter Plot analysis on the GRADE VARIATION BASED ON THE STUDY TIME, TRAVEL TIME, ABSENCE**



The above plotted graph represents the grade variation based on the study time, travel time, absence as the students are graded with the time taken to study for an overall calculation is been done.

## **Tabular analysis about the details of the STUDENTS (School name, gender, age) with grade condition**



The tabular analysis will give the details of the school name, gender and Age of the student with the Grade filter. The slider will filter the grade accordingly, it will display the result of it. From, the slider we can see the highest grade is 20 and the grade is within 0 – 20.

## **Dashboard of all the Graph**

A close up of a map

Description automatically generated

The above screenshot is the dashboard which contain all the interactive graphs at one place.

# CONCLUSION

It is clear to interpret the information about the analysis in the interactive Visualization. We have used various charts and graphs in an interactive way to grab audience attendance and interest to travel along the side of it. From the visualization, we are more concentrated on the Grade, Alcohol consuming, Study time and absent of class of the students. To sum up all, we can clearly see the highest number students are enrolled in the GP school in Portuguese because alcohol consumption of the student is low and study time is high in it.