

An aerial photograph of a modern, multi-story building with a complex, stepped design. The building features extensive green roofs with various plants and trees. A central courtyard area is visible, surrounded by the building's structure. The overall aesthetic is modern and eco-friendly.

# Student Performance Data Analysis

Final Project (Lim Swee Ming)  
Dataset from Kaggle

# Project Overview

## Student Performance Dataset

649 students participated in this survey and a set of data values featuring 31 columns/

Across left to right, the value consist of student's parent education and occupation, student's social activities and their academic grades.

This project scope aims to reveal specific inequalities of the student's background. Teachers may see them in better light to guide students who performed poorly and help them succeed in academics

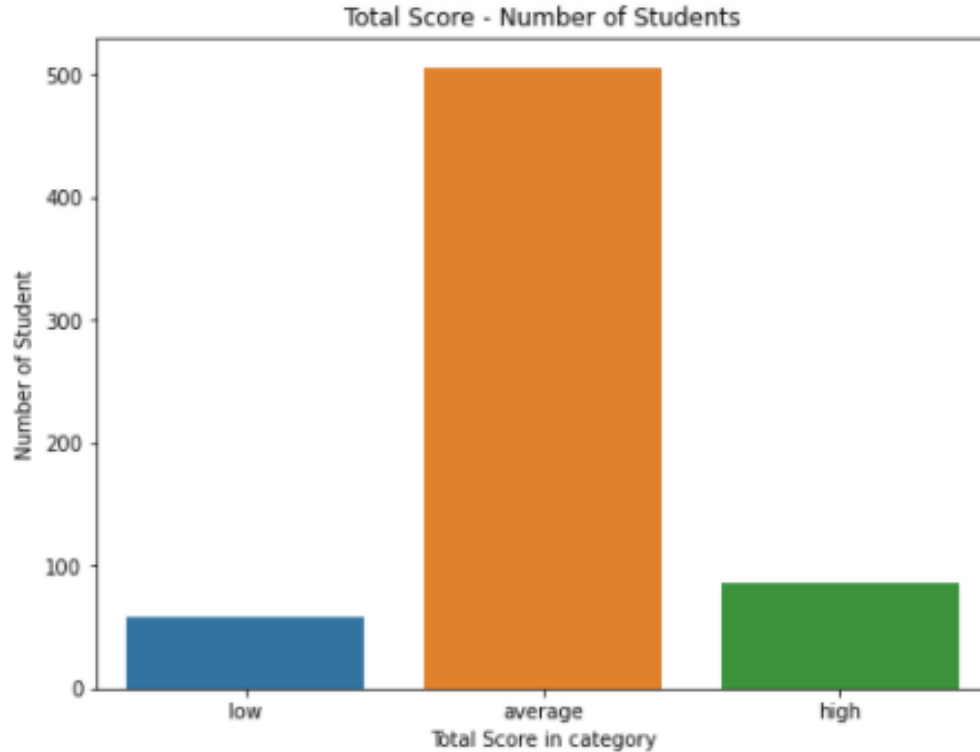
# Data Cleaning

Check for missing values

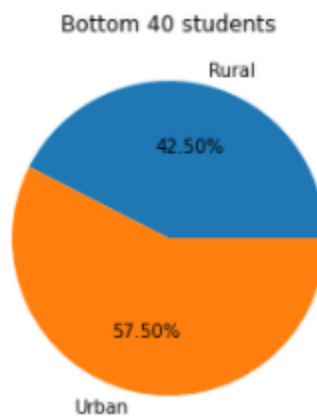
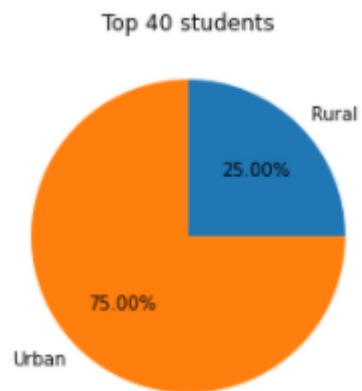
Creating new columns and assigning new values for correlation function at later report

For instance, the dataset comes with three grade periods. Sum up altogether and categorize into low, average and high score. Other variables from scale 1 – 5 is also done in the same manner.

## Identify the top/bottom 40 students based on their grades



## Analysis top/bottom 40 student's living in rural/urban



## Analysis the top/bottom 40 student's parents education using mean

```
: 1 top_40.groupby('grades')[["Medu", "Fedu"]].agg([np.mean])
```

```
:
```

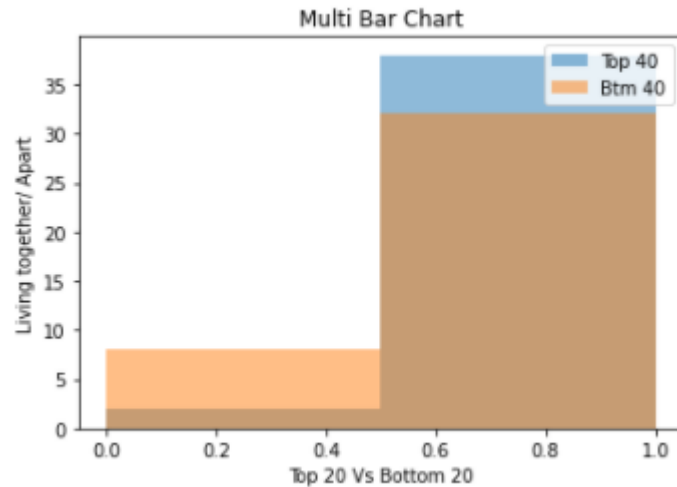
	Medu	Fedu
	mean	mean
grades		
high	3.025	2.575

```
: 1 btm_40.groupby('grades')[["Medu", "Fedu"]].agg([np.mean])
```

```
:
```

	Medu	Fedu
	mean	mean
grades		
low	2.25	1.9

## Analysis the top/bottom 40 student's living with/apart parents



# Performance by gender

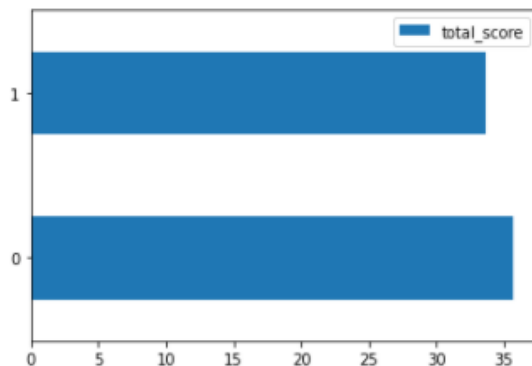
```
] 1 student_score_sex = student_data.groupby("sex")['total_score'].mean().reset_index()
2 student_score_sex
```

```
]:
```

	sex	total_score
0	F	35.712794
1	M	33.669173

```
] 1 student_score_sex.plot(kind = "barh")
```

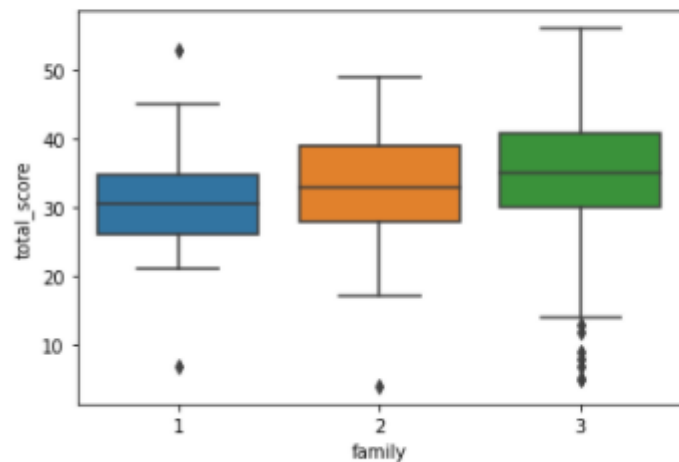
```
] <matplotlib.axes._subplots.AxesSubplot at 0x21bb7bd8d30>
```





# Does family support have impact towards achievements?

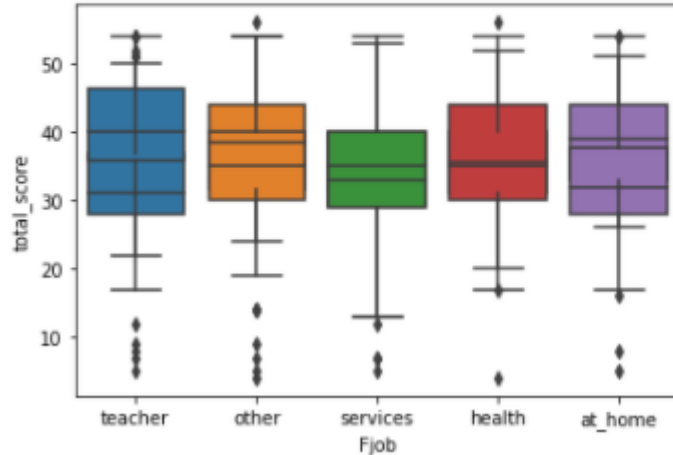
```
: 1 sns.boxplot(data=student_data, x="family", y="total_score")  
: <matplotlib.axes._subplots.AxesSubplot at 0x21bb7c44a90>
```



## What type of job parent's hold will have impact towards student achievements?

```
1 sns.boxplot(data=student_data, x="Mjob", y="total_score")  
2 sns.boxplot(data=student_data, x="Fjob", y="total_score")
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x21bb7cb2dc0>

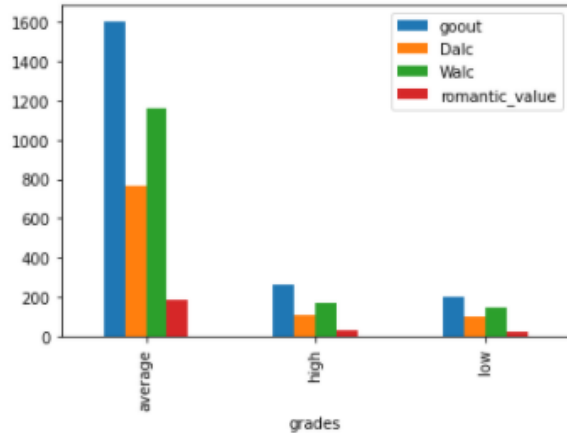


# Which social factors has the most influence on academic achievement?

```
1 Social_factors = student_data.groupby('grades')[["goout", 'Dalc', 'Walc', 'romantic_value']].sum()
```

```
1 Social_factors.plot(kind = "bar")
```

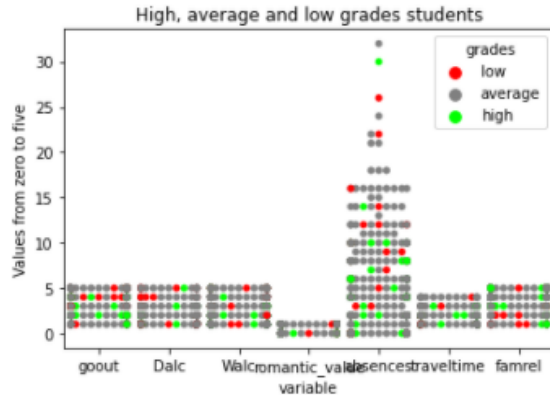
<matplotlib.axes.\_subplots.AxesSubplot at 0x21bb7dc98b0>



# Which social factors has the most influence on academic achievement?

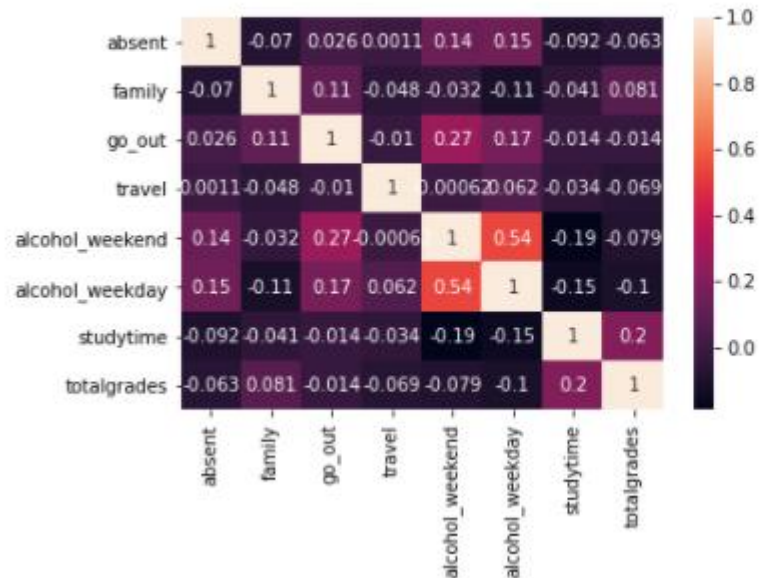
```
1 sns.swarmplot(x='variable',y='value',hue='grades' , data=melt,palette={'high':'lime','average':'grey','low':'red'})
2 plt.ylabel('Values from zero to five')
3 plt.title('High, average and low grades students')
4 plt.show
```

```
<function matplotlib.pyplot.show(*args, **kw)>
```



# Social factors correlation with student performance

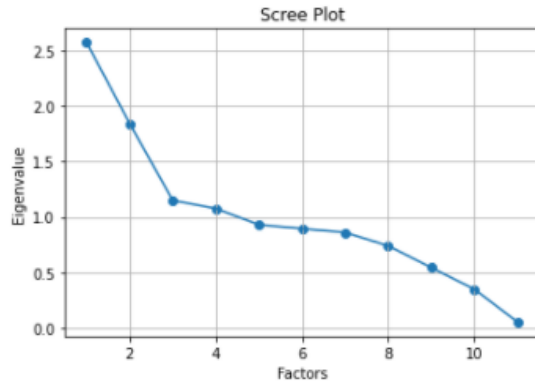
```
: 1 sns.heatmap(a.corr(), annot=True)  
  2 plt.show()
```



## Import Factor Analyzer – sklearn function

To identify which factor is the most important for academics to know.

```
1 plt.scatter(xvals, ev)  
2 plt.plot(xvals, ev)  
3 plt.title('Scree Plot')  
4 plt.xlabel('Factors')  
5 plt.ylabel('Eigenvalue')  
6 plt.grid()  
7 plt.show()
```



# Conclusion and Recommendation

Insights we have identified based on the report

- Students who stay in urban areas are more likely to have higher score
- Students living together or apart with parents has no contribution to their grade
- Parents education level has effects on the student's grade
- Female academic performance is slightly higher than male
- Higher family support can achieve greater grades
- Parents job has an impact as well as the grades.
- Study, family and parent's education factors are the most important influence for greater achievements while alcohol, absent have adverse effects. Travelling can also be seen affecting the grades. Going out not so much impact.

# Conclusion and Recommendation

Clearly, teachers are not able to alter the student's family background and can only educate on their social activities and keep an eye on student's who scored poorly. The school should promote anti consumption of alcohol and assist to find out why students go absent especially for those who always do that.



# Thanks!

<https://github.com/limsweeming/Mages.git>

