

Project: Student Alcohol Consumption Analysis

Python Libraries

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Data Collection

```
In [2]: df=pd.read_csv('E:/Stats survey.csv')
```

Data Cleaning and Preprocessing

```
In [3]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 406 entries, 0 to 405
Data columns (total 17 columns):
 #   Column                                     Non-Null
 1  Count  Dtype                                -----
 ---  ---
-----
 0   Timestamp                                406 non
-null   object
 1   Your Sex?                                404 non
-null   object
 2   Your Matric (grade 12) Average/ GPA (in %) 399 non
-null   float64
 3   What year were you in last year (2023) ? 333 non
-null   object
 4   What faculty does your degree fall under? 399 non
-null   object
 5   Your 2023 academic year average/GPA in % (Ignore if you are 2024 1st year student) 320 non
-null   float64
 6   Your Accommodation Status Last Year (2023) 383 non
-null   object
 7   Monthly Allowance in 2023                375 non
-null   object
 8   Were you on scholarship/bursary in 2023? 398 non
-null   object
 9   Additional amount of studying (in hrs) per week 406 non
-null   object
10   How often do you go out partying/socialising during the week? 404 non
-null   object
11   On a night out, how many alcoholic drinks do you consume? 406 non
-null   object
12   How many classes do you miss per week due to alcohol reasons, (i.e: being hungover or too tired?) 403 non
-null   object
13   How many modules have you failed thus far into your studies? 403 non
-null   object
14   Are you currently in a romantic relationship? 403 non
-null   object
15   Do your parents approve alcohol consumption? 402 non
-null   object
16   How strong is your relationship with your parent/s? 403 non
-null   object
dtypes: float64(2), object(15)
memory usage: 54.0+ KB
```

```
In [4]: df.shape
```

```
Out[4]: (406, 17)
```

```
In [6]: df.duplicated().sum()
```

```
Out[6]: 0
```

```
In [7]: df.isna().sum()
```

```
Out[7]: Timestamp 0
Your Sex? 2
Your Matric (grade 12) Average/ GPA (in %) 7
What year were you in last year (2023) ? 73
What faculty does your degree fall under? 7
Your 2023 academic year average/GPA in % (Ignore if you are 2024 1st year student) 86
Your Accommodation Status Last Year (2023) 23
Monthly Allowance in 2023 31
Were you on scholarship/bursary in 2023? 8
Additional amount of studying (in hrs) per week 0
How often do you go out partying/socialising during the week? 2
On a night out, how many alcoholic drinks do you consume? 0
How many classes do you miss per week due to alcohol reasons, (i.e: being hungover or too tired?) 3
How many modules have you failed thus far into your studies? 3
Are you currently in a romantic relationship? 3
Do your parents approve alcohol consumption? 4
How strong is your relationship with your parent/s? 3
dtype: int64
```

```
In [8]: df.dropna(inplace=True)
```

```
In [11]: df.isna().sum()
```

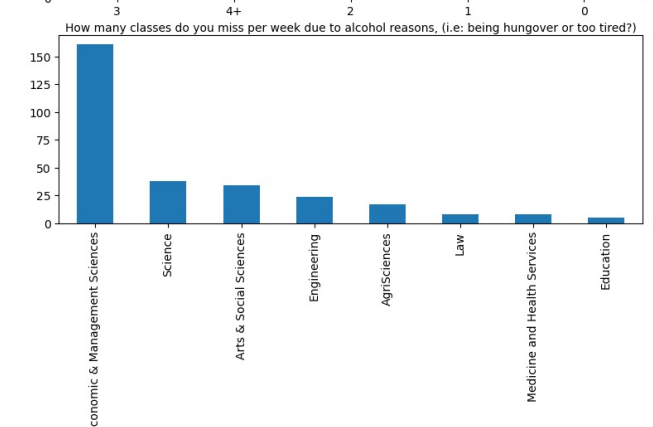
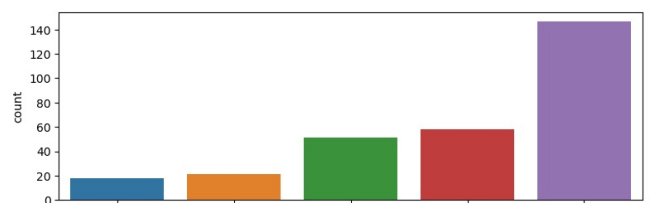
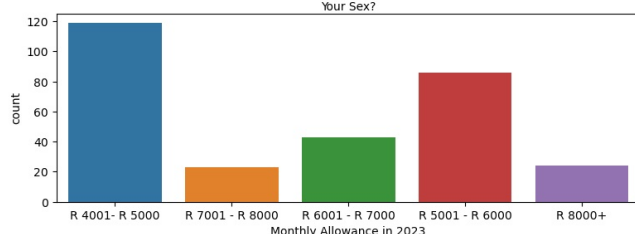
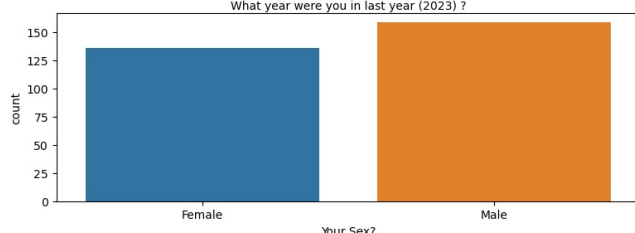
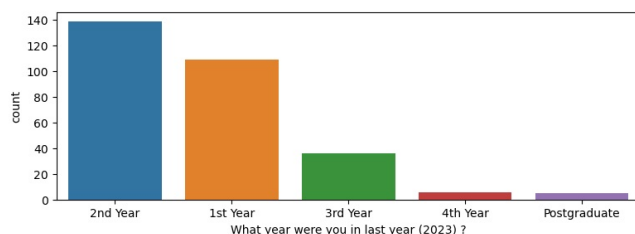
```
Out[11]: Timestamp 0
Your Sex? 0
Your Matric (grade 12) Average/ GPA (in %) 0
What year were you in last year (2023) ? 0
What faculty does your degree fall under? 0
Your 2023 academic year average/GPA in % (Ignore if you are 2024 1st year student) 0
Your Accommodation Status Last Year (2023) 0
Monthly Allowance in 2023 0
Were you on scholarship/bursary in 2023? 0
Additional amount of studying (in hrs) per week 0
How often do you go out partying/socialising during the week? 0
On a night out, how many alcoholic drinks do you consume? 0
How many classes do you miss per week due to alcohol reasons, (i.e: being hungover or too tired?) 0
How many modules have you failed thus far into your studies? 0
Are you currently in a romantic relationship? 0
Do your parents approve alcohol consumption? 0
How strong is your relationship with your parent/s? 0
dtype: int64
```

Exploratory Data Analysis (EDA)

Univariate Analysis

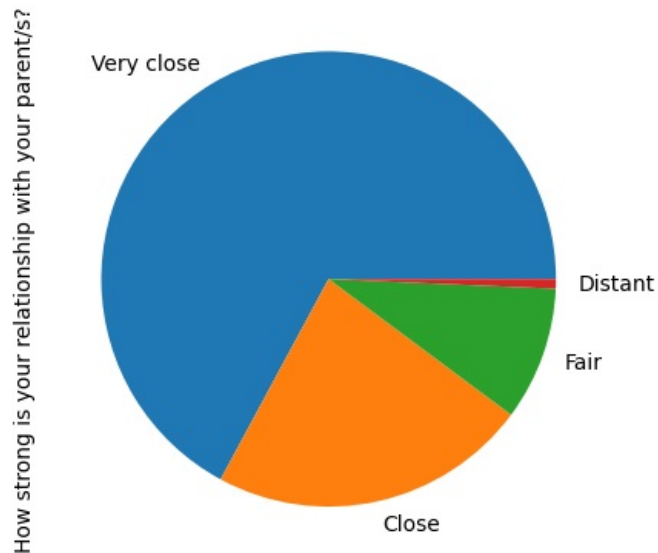
```
In [13]: figure=plt.figure(figsize=(20,10))
plt.subplot(3,2,1)
sns.countplot(x=df['What year were you in last year (2023) ?'])
plt.subplot(3,2,2)
sns.countplot(x=df['How many classes do you miss per week due to alcohol reasons, (i.e: being hungover or too t
plt.subplot(3,2,3)
sns.countplot(x=df['Your Sex?'])
plt.subplot(3,2,4)
df['What faculty does your degree fall under?'].value_counts().plot(kind='bar')
plt.subplot(3,2,5)
sns.countplot(x=df['Monthly Allowance in 2023'])
```

```
Out[13]: <Axes: xlabel='Monthly Allowance in 2023', ylabel='count'>
```



```
In [14]: df['How strong is your relationship with your parent/s?'].value_counts().plot(kind='pie')
```

```
Out[14]: <Axes: ylabel='How strong is your relationship with your parent/s?'>
```



```
In [15]: df['Your Sex?'].value_counts()
```

```
Out[15]: Male      159
Female    136
Name: Your Sex?, dtype: int64
```

```
In [17]: df['What faculty does your degree fall under?'].value_counts()
```

```
Out[17]: Economic & Management Sciences    161
Science                                   38
Arts & Social Sciences                     34
Engineering                               24
AgriSciences                             17
Law                                        8
Medicine and Health Services              8
Education                                5
Name: What faculty does your degree fall under?, dtype: int64
```

```
In [18]: df['What year were you in last year (2023) ?'].value_counts()
```

```
Out[18]: 2nd Year      139
1st Year      109
3rd Year       36
4th Year        6
Postgraduate   5
Name: What year were you in last year (2023) ?, dtype: int64
```

```
In [19]: df['Are you currently in a romantic relationship?'].value_counts()
```

```
Out[19]: No      167
Yes      128
Name: Are you currently in a romantic relationship?, dtype: int64
```

```
In [37]: df['Do your parents approve alcohol consumption?'].value_counts()
```

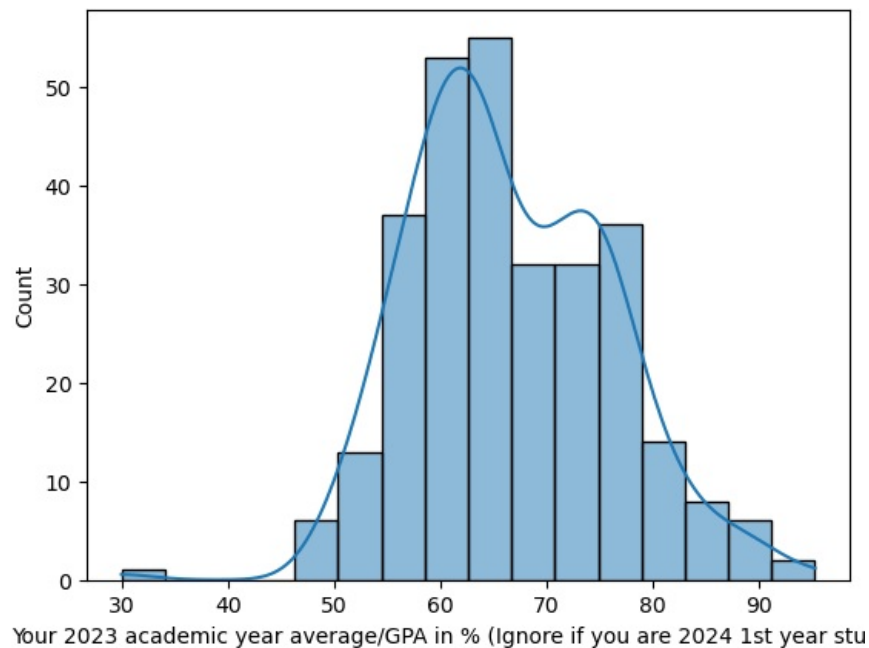
```
Out[37]: Yes      260
No        35
Name: Do your parents approve alcohol consumption?, dtype: int64
```

```
In [38]: df['How strong is your relationship with your parent/s?'].value_counts()
```

```
Out[38]: Very close    198
Close        67
Fair         28
Distant       2
Name: How strong is your relationship with your parent/s?, dtype: int64
```

```
In [39]: sns.histplot(x=df['Your 2023 academic year average/GPA in % (Ignore if you are 2024 1st year student)'],kde=True)
```

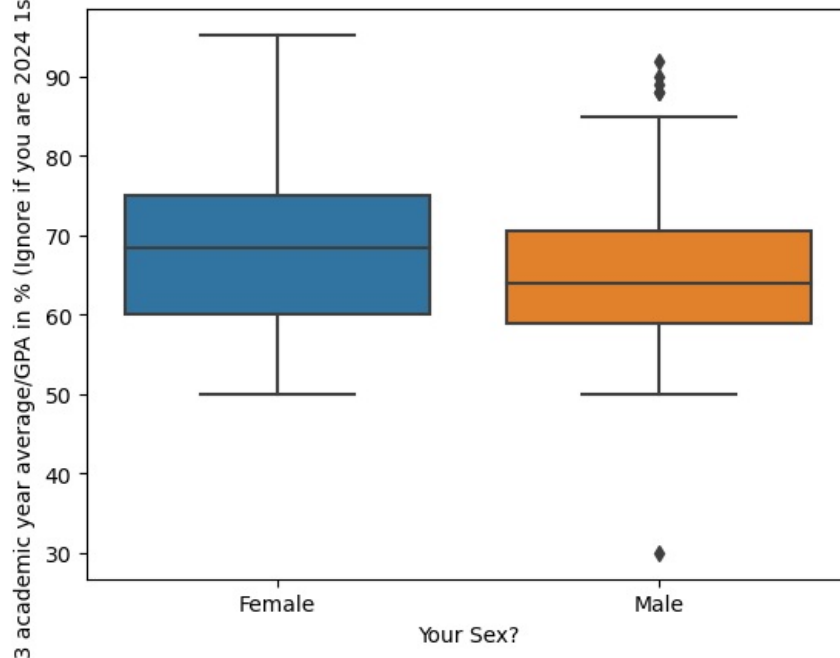
```
Out[39]: <Axes: xlabel='Your 2023 academic year average/GPA in % (Ignore if you are 2024 1st year student)', ylabel='Count'>
```



Bivariate Analysis

```
In [42]: sns.boxplot(x=df['Your Sex?'],y=df['Your 2023 academic year average/GPA in % (Ignore if you are 2024 1st year student)'])
```

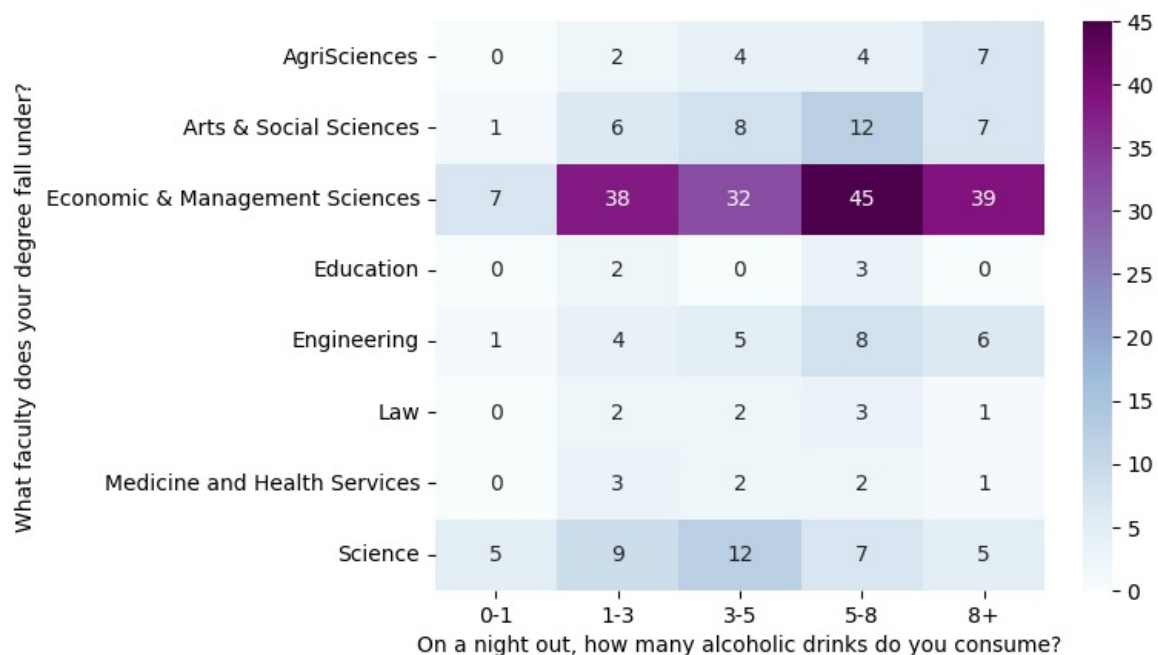
```
Out[42]: <Axes: xlabel='Your Sex?', ylabel='Your 2023 academic year average/GPA in % (Ignore if you are 2024 1st year student)'\>
```



Above Boxplot shows that there are outliers in male percentage field

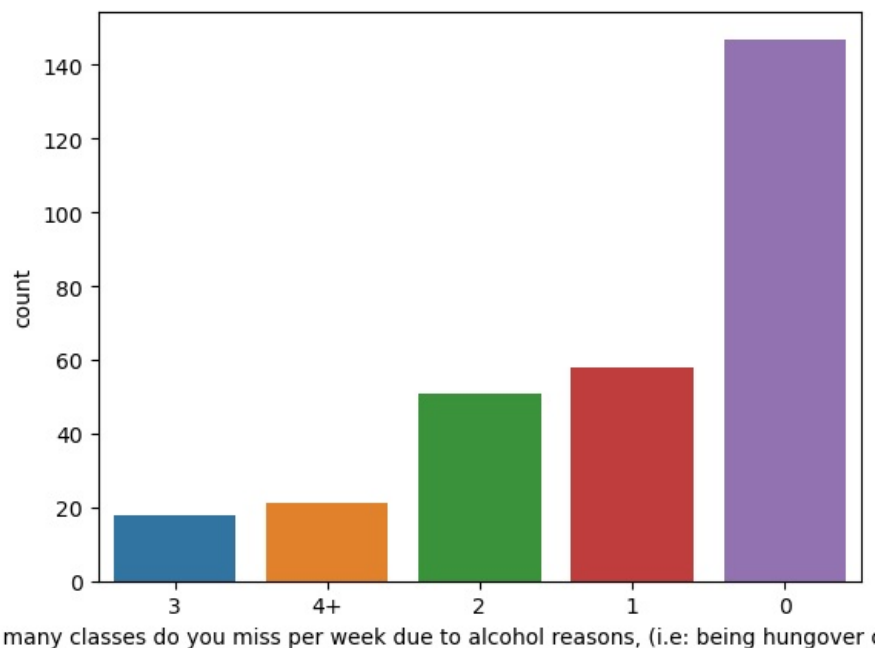
```
In [49]: ctab=pd.crosstab(df['What faculty does your degree fall under?'],df['On a night out, how many alcoholic drinks do you consume?'])
sns.heatmap(ctab,annot=True,cmap='BuPu')
```

```
Out[49]: <Axes: xlabel='On a night out, how many alcoholic drinks do you consume?', ylabel='What faculty does your degree fall under?'\>
```



It shows almost all the students drink and Economic & Management Sciences students consume most no of drinks.

```
In [44]: sns.countplot(x=df['How many classes do you miss per week due to alcohol reasons, (i.e: being hungover or too t
Out[44]: <Axes: xlabel='How many classes do you miss per week due to alcohol reasons, (i.e: being hungover or too tired?
)', ylabel='count'>
```



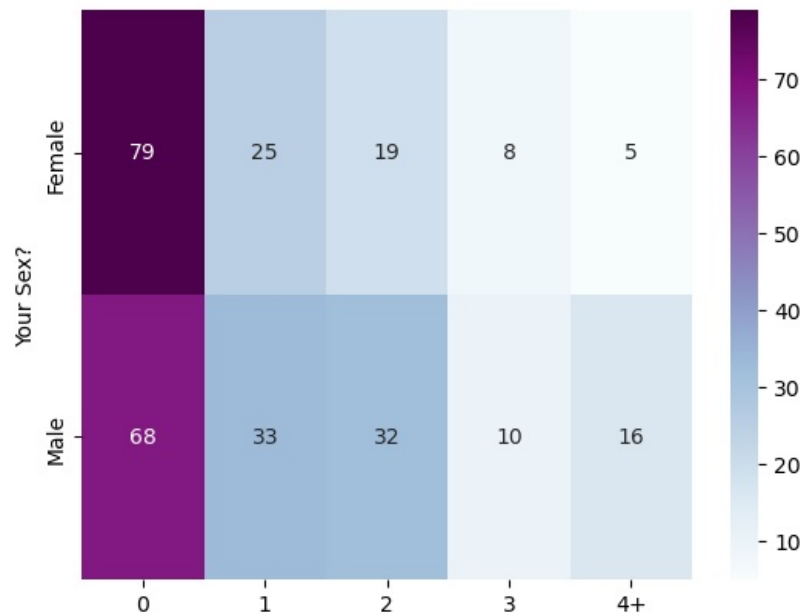
many classes do you miss per week due to alcohol reasons, (i.e: being hungover or too tired?)

```
In [45]: df['How many classes do you miss per week due to alcohol reasons, (i.e: being hungover or too tired?').value_c
Out[45]: 0      147
         1      58
         2      51
         4+     21
         3      18
Name: How many classes do you miss per week due to alcohol reasons, (i.e: being hungover or too tired?), dtype:
int64
```

It shows almost 50% students are missing classes per week due to alcohol (hungover)

```
In [43]: ctab=pd.crosstab(df['Your Sex?'],df['How many classes do you miss per week due to alcohol reasons, (i.e: being
sns.heatmap(ctab,annot=True,cmap='BuPu')
```

```
Out[43]: <Axes: xlabel='How many classes do you miss per week due to alcohol reasons, (i.e: being hungover or too tired?)', ylabel='Your Sex? '>
```

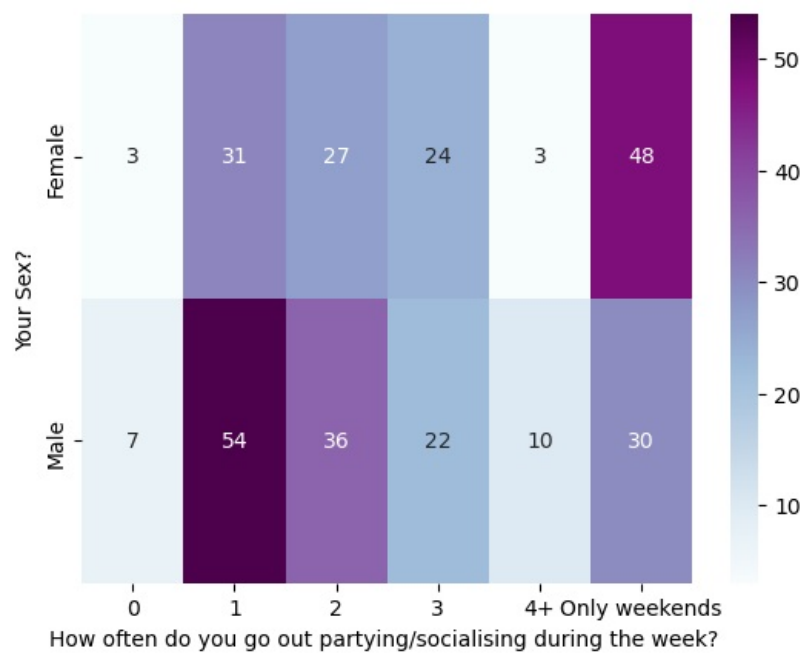


asses do you miss per week due to alcohol reasons, (i.e: being hungover or t

Males are missing more number of classes due to alcohol reasons than females.

```
In [46]: ctab=pd.crosstab(df['Your Sex?'],df['How often do you go out partying/socialising during the week? '])
sns.heatmap(ctab,annot=True,cmap='BuPu')
```

```
Out[46]: <Axes: xlabel='How often do you go out partying/socialising during the week? ', ylabel='Your Sex? '>
```



```
In [47]: ctab=pd.crosstab(df['Your Sex?'],df['On a night out, how many alcoholic drinks do you consume?'])
sns.heatmap(ctab,annot=True,cmap='BuPu')
```

```
Out[47]: <Axes: xlabel='On a night out, how many alcoholic drinks do you consume?', ylabel='Your Sex? '>
```


Out[53]:

	Timestamp	Your Sex?	Your Matric (grade 12) Average/ GPA (in %)	What year were you in last year (2023) ?	What faculty does your degree fall under?	Your 2023 academic year average/GPA in % (Ignore if you are 2024 1st year student)	Your Accommodation Status Last Year (2023)	Monthly Allowance in 2023	Were you on scholarship/bursary in 2023?	Additional amount of studying (in hrs) per week	How oft partying/s during t
73	2024/03/07 5:39:33 pm EET	Female	98.33	2nd Year	Economic & Management Sciences	95.22	Non-private accommodation ie. Res	R 4001- R 5000	No	8+	
237	2024/03/07 10:08:29 pm EET	Male	84.00	1st Year	Science	92.00	Non-private accommodation ie. Res	R 5001 - R 6000	No	3-5	Only
38	2024/03/07 5:20:31 pm EET	Male	99.00	2nd Year	Economic & Management Sciences	90.00	Private accommodation/ stay with family/friends	R 4001- R 5000	No	5-8	Only
54	2024/03/07 5:25:13 pm EET	Male	98.00	1st Year	Engineering	89.00	Private accommodation/ stay with family/friends	R 4001- R 5000	Yes (NSFAS, etc...)	1-3	Only
310	2024/03/08 10:54:38 am EET	Female	75.00	1st Year	Economic & Management Sciences	89.00	Private accommodation/ stay with family/friends	R 6001 - R 7000	No	5-8	

In [54]: df2.tail(5)

Out[54]:

	Timestamp	Your Sex?	Your Matric (grade 12) Average/ GPA (in %)	What year were you in last year (2023) ?	What faculty does your degree fall under?	Your 2023 academic year average/GPA in % (Ignore if you are 2024 1st year student)	Your Accommodation Status Last Year (2023)	Monthly Allowance in 2023	Were you on scholarship/bursary in 2023?	Additional amount of studying (in hrs) per week	How oft partying/s during t
177	2024/03/07 7:55:47 pm EET	Female	75.0	3rd Year	Economic & Management Sciences	50.0	Private accommodation/ stay with family/friends	R 7001 - R 8000	No	0-1	Only
192	2024/03/07 8:25:17 pm EET	Male	75.0	2nd Year	Science	50.0	Private accommodation/ stay with family/friends	R 6001 - R 7000	No	1-3	Only
361	2024/03/11 2:07:05 pm EET	Male	80.0	2nd Year	Science	50.0	Private accommodation/ stay with family/friends	R 5001 - R 6000	No	8+	
358	2024/03/11 2:06:41 pm EET	Male	60.0	2nd Year	Arts & Social Sciences	50.0	Private accommodation/ stay with family/friends	R 4001- R 5000	No	3-5	
303	2024/03/08 9:50:48 am EET	Male	67.0	1st Year	Economic & Management Sciences	30.0	Private accommodation/ stay with family/friends	R 4001- R 5000	No	1-3	

Average percentage

In [55]: df['Your 2023 academic year average/GPA in % (Ignore if you are 2024 1st year student)'].mean()

Out[55]: 66.56277966101695

In [57]: df3=df[(df['What faculty does your degree fall under?']=='Economic & Management Sciences') & (df['On a night ou

In [58]: df3.value_counts()


```
Out[58]: On a night out, how many alcoholic drinks do you consume? Your Sex? What faculty does your degree fall under?
8+ Male Economic & Management Sciences
35 Female Economic & Management Sciences
4
dtype: int64

In [60]: df4=df[df['On a night out, how many alcoholic drinks do you consume?']=='8+'][['Your Sex?','Your Matric (grade

In [61]: df4['Your Sex?'].value_counts()

Out[61]: Male      56
Female    10
Name: Your Sex?, dtype: int64

In [62]: df4['How many classes do you miss per week due to alcohol reasons, (i.e: being hungover or too tired?)'].value_

Out[62]: 2      19
0      18
1      14
4+     10
3       5
Name: How many classes do you miss per week due to alcohol reasons, (i.e: being hungover or too tired?), dtype:
int64

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js