Alcohol effect -Project

October 11, 2022

```python
[3]: import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
     %matplotlib inline
     #plt.style.use('seaborn-whitegrid')
```

1 Alcohol Effects On Studies

2 Math Dataset

```python
[13]: maths=pd.read_csv("Maths.csv")
```

```python
[19]: #Math language student dataset downloaded and displaded below.
      maths
```

```
[19]:      school sex  age address famsize Pstatus  Medu  Fedu     Mjob      Fjob  \
     0        GP   F   18       U     GT3       A     4     4  at_home   teacher
     1        GP   F   17       U     GT3       T     1     1  at_home     other
     2        GP   F   15       U     LE3       T     1     1  at_home     other
     3        GP   F   15       U     GT3       T     4     2   health  services
     4        GP   F   16       U     GT3       T     3     3    other     other
     ..      ...  ..  ...     ...     ...     ...   ...   ...      ...       ...
     390      MS   M   20       U     LE3       A     2     2 services  services
     391      MS   M   17       U     LE3       T     3     1 services  services
     392      MS   M   21       R     GT3       T     1     1    other     other
     393      MS   M   18       R     LE3       T     3     2 services     other
     394      MS   M   19       U     LE3       T     1     1    other   at_home

          … famrel freetime  goout  Dalc  Walc health absences  G1  G2  G3
     0    …      4        3      4     1     1      3        6   5   6   6
     1    …      5        3      3     1     1      3        4   5   5   6
     2    …      4        3      2     2     3      3       10   7   8  10
     3    …      3        2      2     1     1      5        2  15  14  15
     4    …      4        3      2     1     2      5        4   6  10  10
     ..   …    ...      ...    ...   ...   ...    ...      ...  ..  ..  ..
     390  …      5        5      4     4     5      4       11   9   9   9
```

```
391  …    2    4    5    3    4    2    3  14  16  16
392  …    5    5    3    3    3    3    3  10   8   7
393  …    4    4    1    3    4    5    0  11  12  10
394  …    3    2    3    3    3    5    5   8   9   9

[395 rows x 33 columns]
```

[20]: `maths.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 395 entries, 0 to 394
Data columns (total 33 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   school      395 non-null    object
 1   sex         395 non-null    object
 2   age         395 non-null    int64
 3   address     395 non-null    object
 4   famsize     395 non-null    object
 5   Pstatus     395 non-null    object
 6   Medu        395 non-null    int64
 7   Fedu        395 non-null    int64
 8   Mjob        395 non-null    object
 9   Fjob        395 non-null    object
 10  reason      395 non-null    object
 11  guardian    395 non-null    object
 12  traveltime  395 non-null    int64
 13  studytime   395 non-null    int64
 14  failures    395 non-null    int64
 15  schoolsup   395 non-null    object
 16  famsup      395 non-null    object
 17  paid        395 non-null    object
 18  activities  395 non-null    object
 19  nursery     395 non-null    object
 20  higher      395 non-null    object
 21  internet    395 non-null    object
 22  romantic    395 non-null    object
 23  famrel      395 non-null    int64
 24  freetime    395 non-null    int64
 25  goout       395 non-null    int64
 26  Dalc        395 non-null    int64
 27  Walc        395 non-null    int64
 28  health      395 non-null    int64
 29  absences    395 non-null    int64
 30  G1          395 non-null    int64
 31  G2          395 non-null    int64
 32  G3          395 non-null    int64
dtypes: int64(16), object(17)
```

```
memory usage: 102.0+ KB
```

```
[21]: maths.describe().T
```

```
[21]:                count        mean        std   min    25%    50%    75%    max
      age            395.0   16.696203   1.276043  15.0   16.0   17.0   18.0   22.0
      Medu           395.0    2.749367   1.094735   0.0    2.0    3.0    4.0    4.0
      Fedu           395.0    2.521519   1.088201   0.0    2.0    2.0    3.0    4.0
      traveltime     395.0    1.448101   0.697505   1.0    1.0    1.0    2.0    4.0
      studytime      395.0    2.035443   0.839240   1.0    1.0    2.0    2.0    4.0
      failures       395.0    0.334177   0.743651   0.0    0.0    0.0    0.0    3.0
      famrel         395.0    3.944304   0.896659   1.0    4.0    4.0    5.0    5.0
      freetime       395.0    3.235443   0.998862   1.0    3.0    3.0    4.0    5.0
      goout          395.0    3.108861   1.113278   1.0    2.0    3.0    4.0    5.0
      Dalc           395.0    1.481013   0.890741   1.0    1.0    1.0    2.0    5.0
      Walc           395.0    2.291139   1.287897   1.0    1.0    2.0    3.0    5.0
      health         395.0    3.554430   1.390303   1.0    3.0    4.0    5.0    5.0
      absences       395.0    5.708861   8.003096   0.0    0.0    4.0    8.0   75.0
      G1             395.0   10.908861   3.319195   3.0    8.0   11.0   13.0   19.0
      G2             395.0   10.713924   3.761505   0.0    9.0   11.0   13.0   19.0
      G3             395.0   10.415190   4.581443   0.0    8.0   11.0   14.0   20.0
```

3 Total number of females students studying

```
[23]: maths[maths["sex"]=="F"]["sex"].count()
```

```
[23]: 208
```

4 Total number of males students studying

```
[24]: maths[maths["sex"]=="M"]["sex"].count()
```
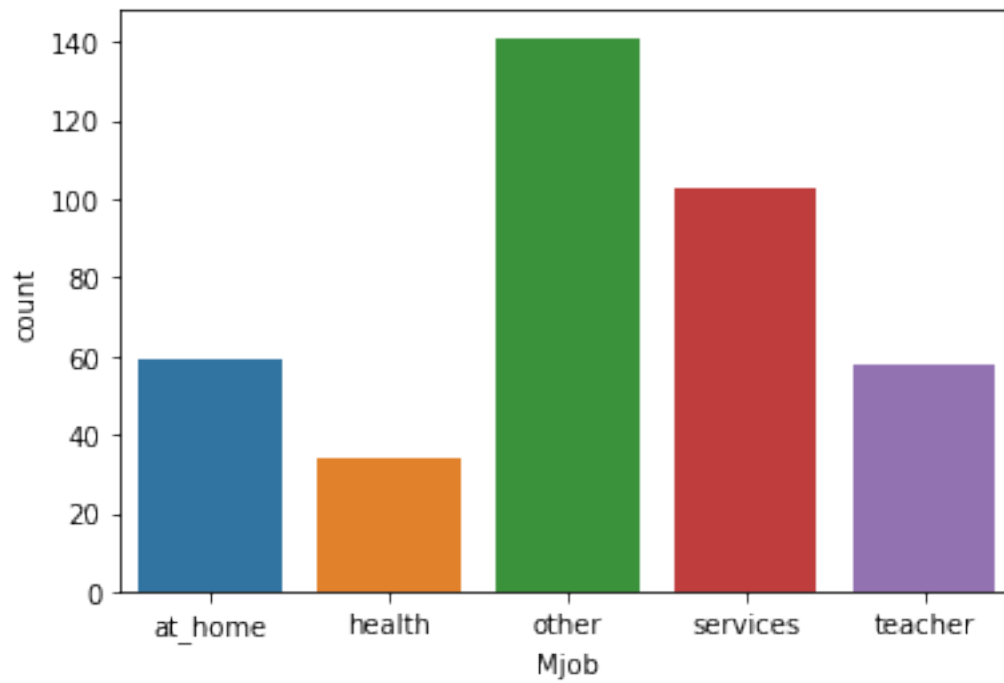
```
[24]: 187
```

5 Parents working, of the students studying

6 Mothers working

```
[26]: maths[maths["Mjob"]!="at_home"]["Mjob"].count()
```

```
[26]: 336
```

```
[27]: sns.countplot(data=maths,x='Mjob');
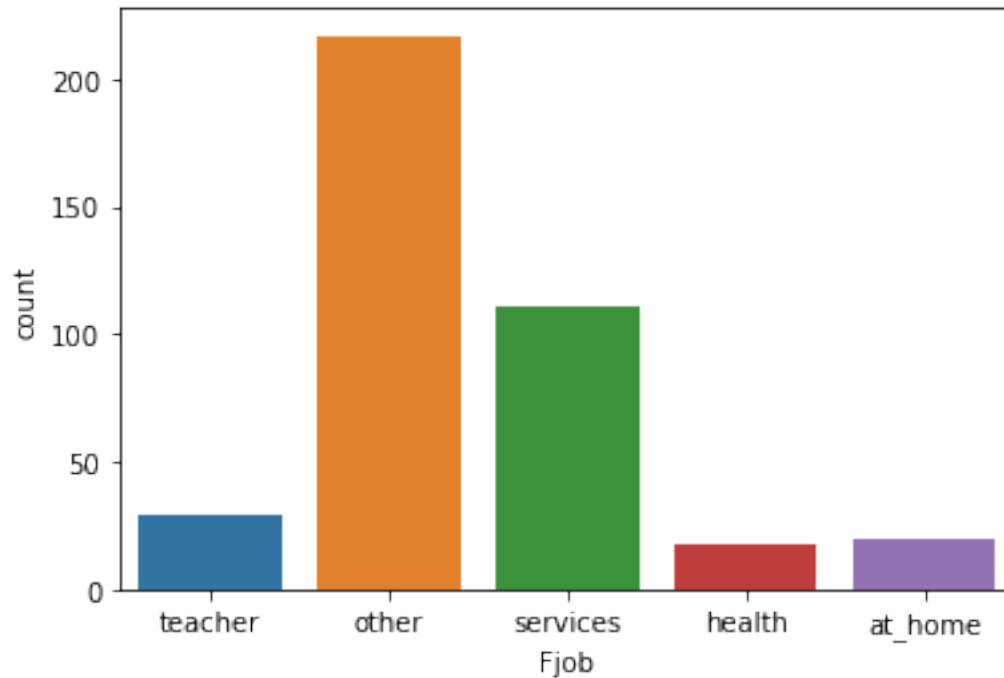```

# 7 Fathers working

```
[29]: maths[maths["Fjob"]!="at_home"]["Fjob"].count()
```

```
[29]: 375
```

```
[30]: sns.countplot(data=maths,x='Fjob')
      plt.show()
```

8 Average of Student Grades from I,II,III years from both the school

```python
[40]: gp_1 = round(maths[maths.school == 'GP']['G1'].mean(), 2)
      gp_2 = round(maths[maths.school == 'GP']['G2'].mean(), 2)
      gp_3 = round(maths[maths.school == 'GP']['G3'].mean(), 2)

      ms_1 = round(maths[maths.school == 'MS']['G1'].mean(), 2)
      ms_2 = round(maths[maths.school == 'MS']['G2'].mean(), 2)
      ms_3 = round(maths[maths.school == 'MS']['G3'].mean(), 2)

      print(f'the mean of GP students are: {gp_1}, {gp_2}, and {gp_3}')
      print(f'the mean of MS students are: {ms_1}, {ms_2}, and {ms_3}')
```

```
the mean of GP students are: 10.94, 10.78, and 10.49
the mean of MS students are: 10.67, 10.2, and 9.85
```

9 Bar Graph denoting the Average Grades of III,II & I year- GP shcool

```
[71]: GP=np.array([gp_3,gp_2,gp_1])
      GP
```

```
[71]: array([10.49, 10.78, 10.94])
```

```
[66]: sns.barplot(x=GP,y=[10,11,12])
      plt.show()
```
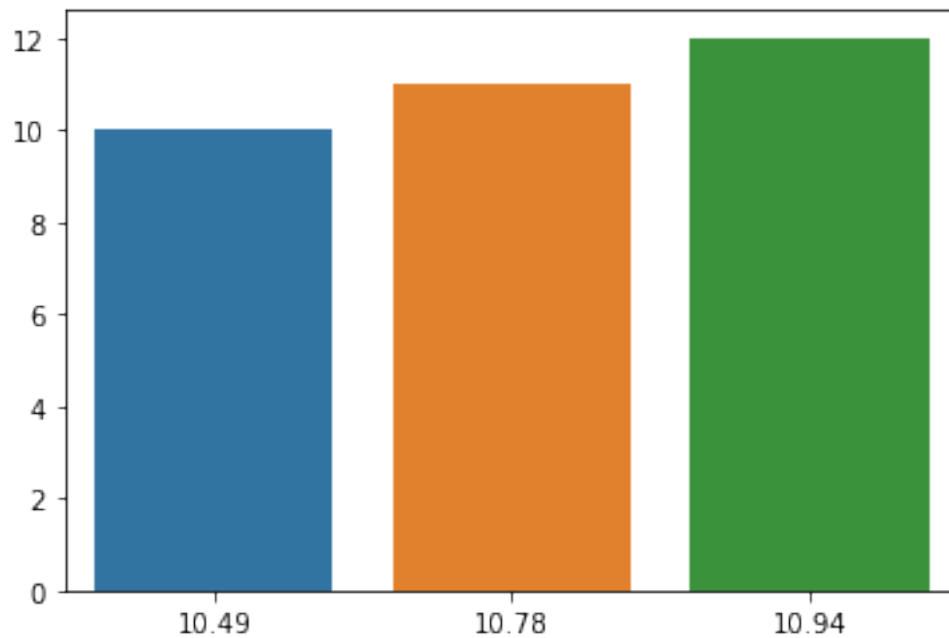

10 Bar Graph denoting the Average Grades of III,II & I year- MS shcool

```
[72]: MS=np.array([ms_3,ms_2,ms_1])
      MS
```

```
[72]: array([ 9.85, 10.2 , 10.67])
```

```
[70]: sns.barplot(x=MS,y=[10,11,12])
      plt.show()
```
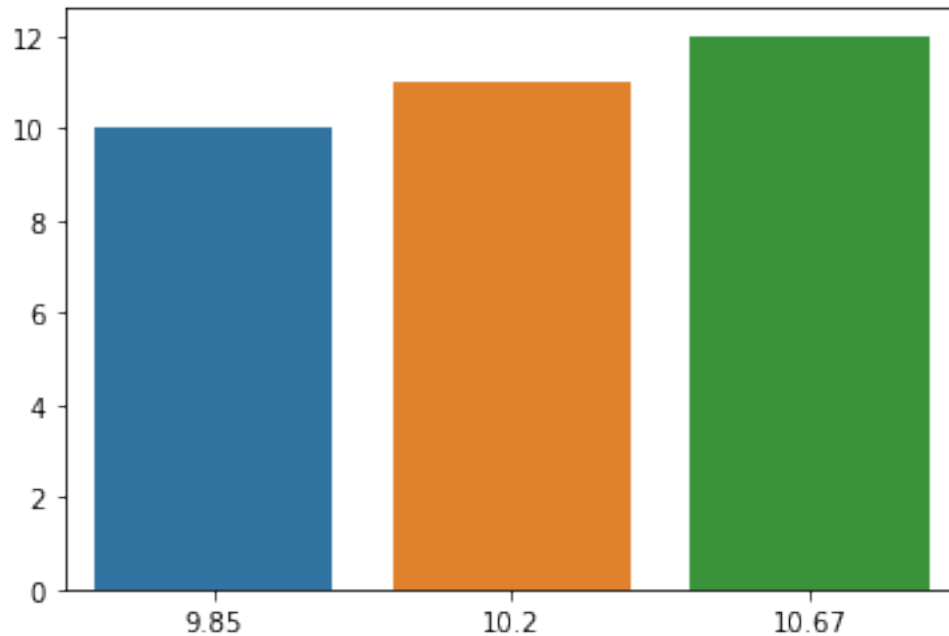
The Grades are uniformly distributed in GP school when compared to MS shcool

The average of the grades droped till 9, in the MS school.

Basis the observation while comparing (G1 and G3) there is a drop in the average grades. Lets take a look the factor affecting the same, Whether its due to alcohol or other factors.

#Identifying the Grade change over the period of time from I year to III year(G1 and G3).

```python
[73]: # Creating the column of change in the dataset
maths['change'] = maths['G3'] - maths['G1']
```

```python
[ ]: As it is shown in the above graph, the grades of the studend from both schol␣
     ↪tend to fall during the I year to III year.
Lets take a look at the factor that influence the most the grade change between␣
     ↪the first and third exam
```

```python
[ ]:
```

```python
[74]: fig, ax = plt.subplots(1,2, figsize=(12,6))

sns.barplot(ax=ax[0], data=maths, x='Dalc', y='change')
sns.barplot(ax=ax[1], data=maths, x='Walc', y='change')

plt.show()
```
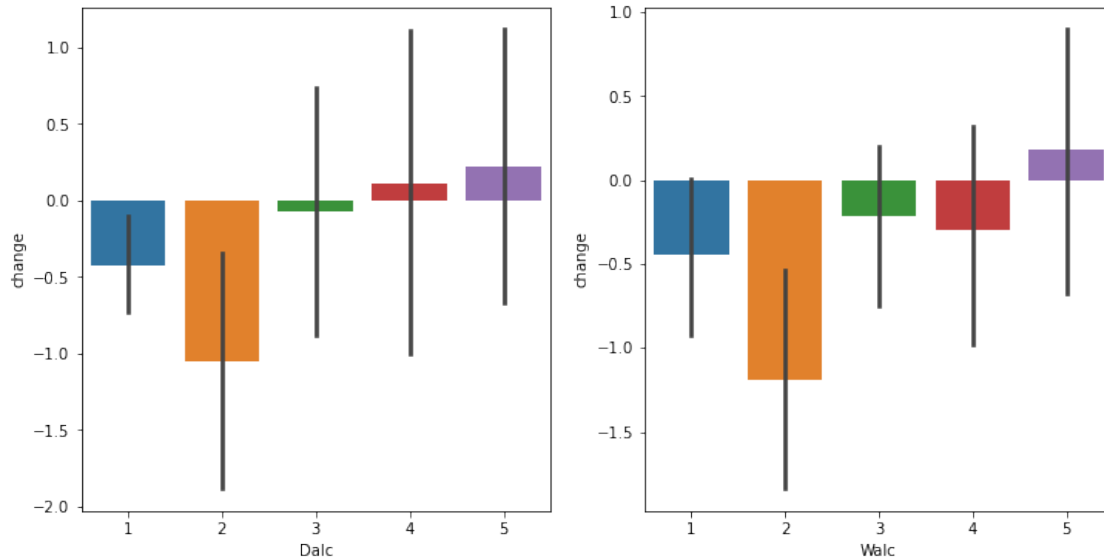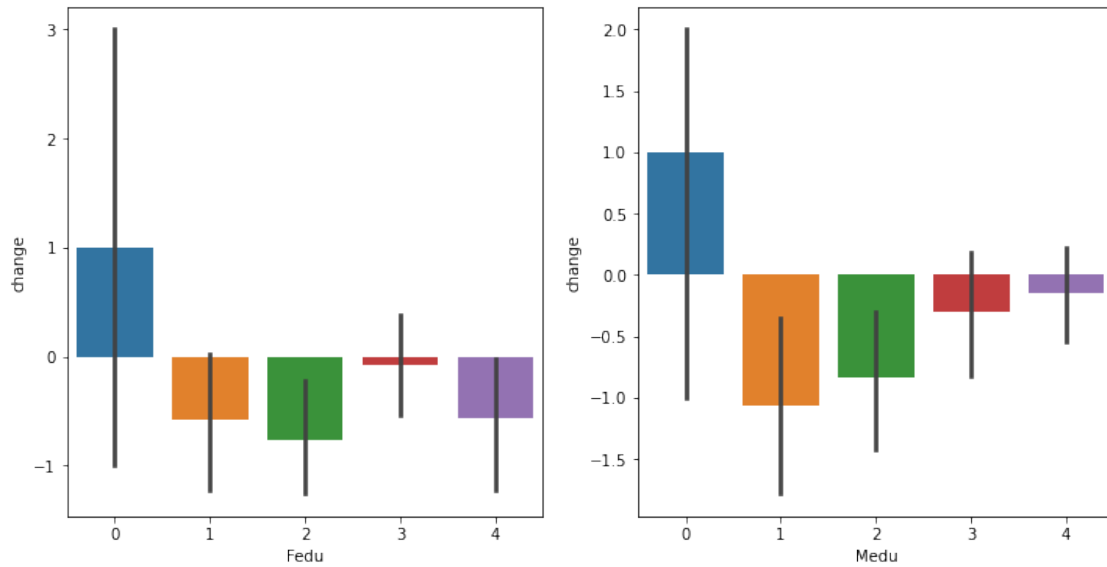
```
[ ]:  #Observation:

      As per the above graph the lowest alcohol consumption has the drop on the␣
       ↪grades.
      The highest alcohol consumption has the increase on the grades
```

```
[76]:  fig, ax = plt.subplots(1,2, figsize=(12,6))

       sns.barplot(ax=ax[0], data=maths, x='Fedu', y='change')
       sns.barplot(ax=ax[1], data=maths, x='Medu', y='change')

       plt.show()
```
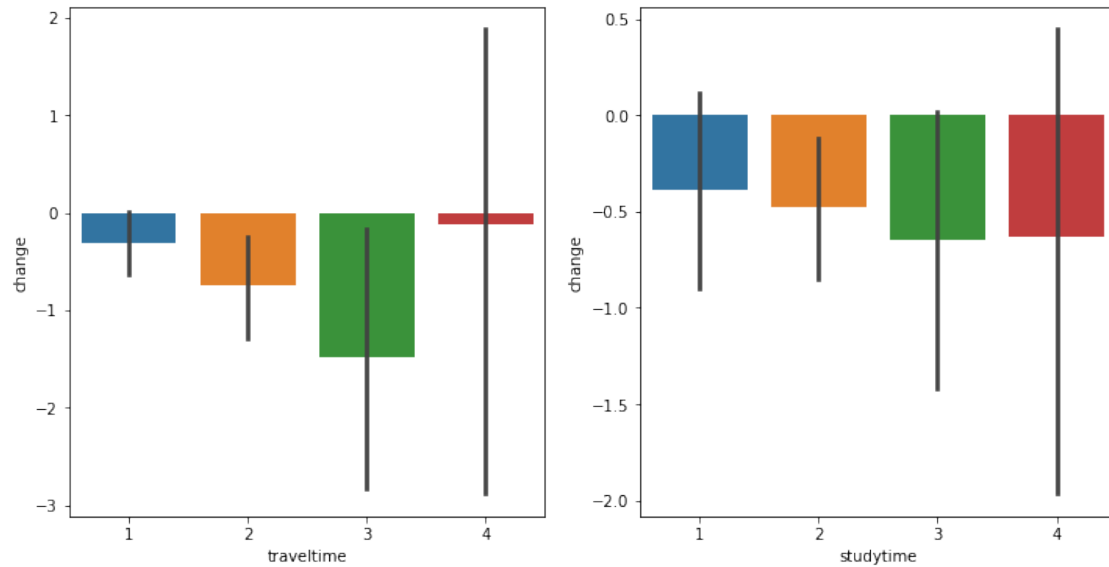
```
[ ]: #Observation:

     As per the graph uneducated parents childrens has good grades when compare to␣
     ↪parents with education background.
```

```
[ ]:
```

```
[77]: fig, ax = plt.subplots(1,2, figsize=(12,6))

      sns.barplot(ax=ax[0], data=maths, x='traveltime', y='change')
      sns.barplot(ax=ax[1], data=maths, x='studytime', y='change')

      plt.show()
```

```
[ ]:  #Observation:

      -> Student who travel more then 1 hour from home to school good grades while␣
       ↪comparing with student who travel <=15min.
      -> Student who studied weekly <2 hours has scored good grades while comparing␣
       ↪with student who studied for 10 hours.
```
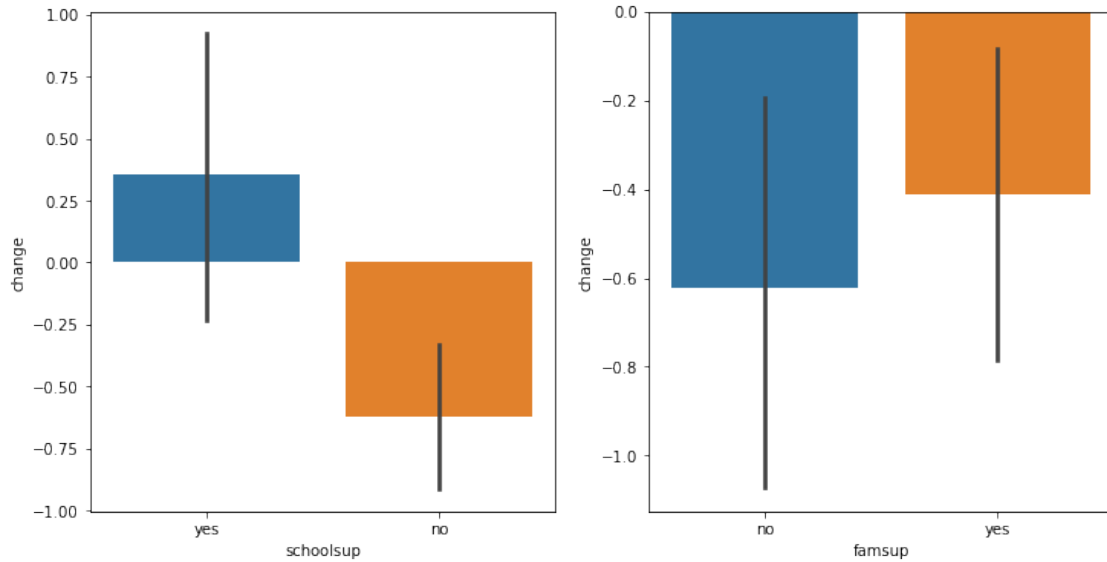
```
[ ]:
```

```
[78]: fig, ax = plt.subplots(1,2, figsize=(12,6))

      sns.barplot(ax=ax[0], data=maths, x='schoolsup', y='change')
      sns.barplot(ax=ax[1], data=maths, x='famsup', y='change')

      plt.show()
```
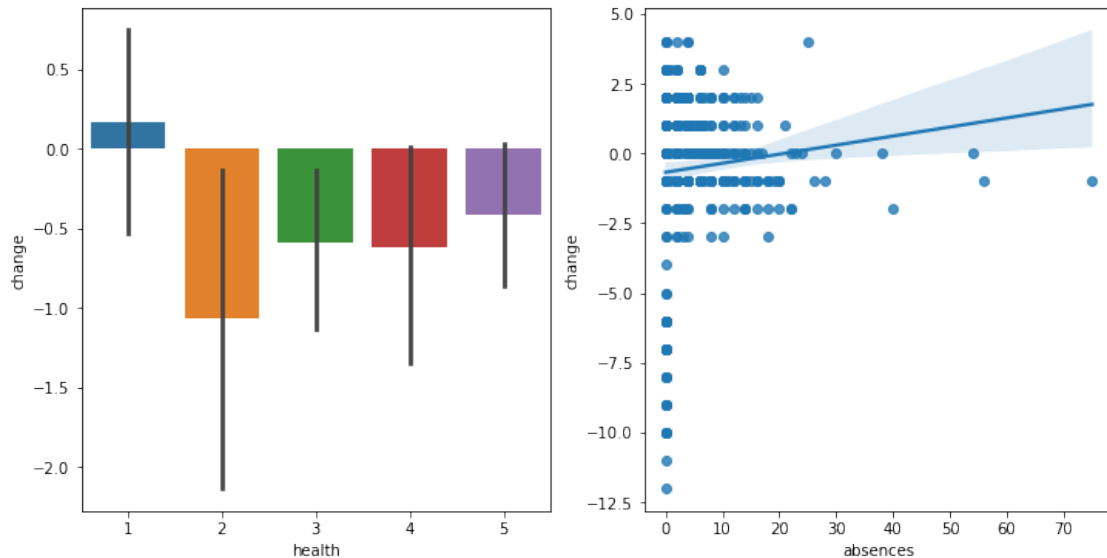
[79]: #Observation:

```
-> School support factore increased the grades over the time.
-> With no family support drop the grades over the period of time.
```

[84]:
```python
fig, ax = plt.subplots(1,2, figsize=(12,6))

sns.barplot(ax=ax[0], data=maths, x='health', y='change')
sns.regplot(ax=ax[1], data=maths, x='absences', y='change')

plt.show()
```
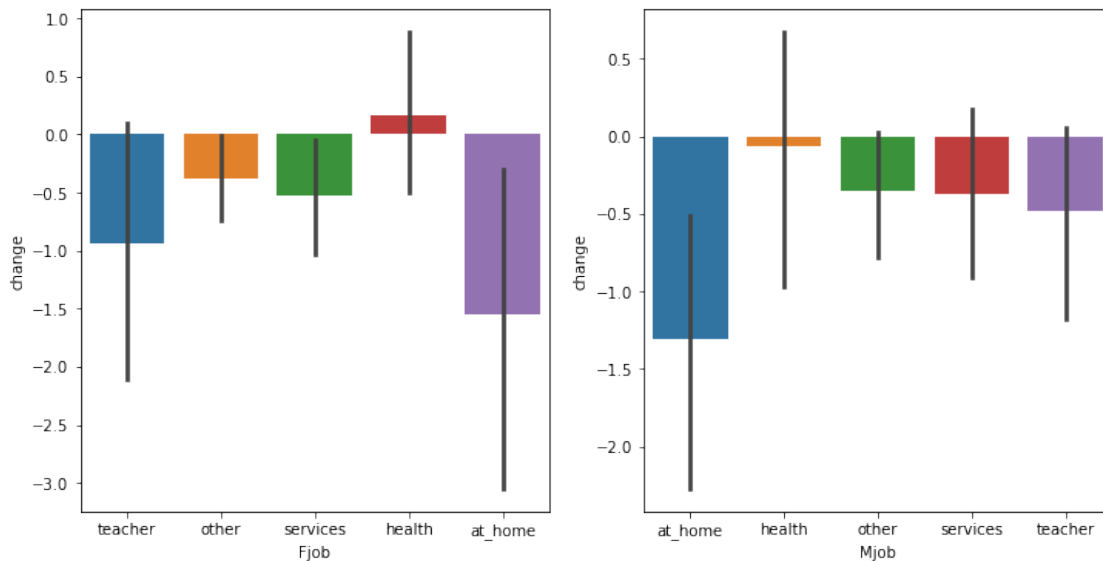
```
[ ]: #Observation:

     ->Student with health issues has maitain the good grades, while compare with␣
       ↪student with no health issues.
     ->As per the liner plot above unable to determine whether absentisum affect the␣
       ↪grades of the student.
```

```
[85]: fig, ax = plt.subplots(1,2, figsize=(12,6))

      sns.barplot(ax=ax[0], data=maths, x='Fjob', y='change')
      sns.barplot(ax=ax[1], data=maths, x='Mjob', y='change')

      plt.show()
```
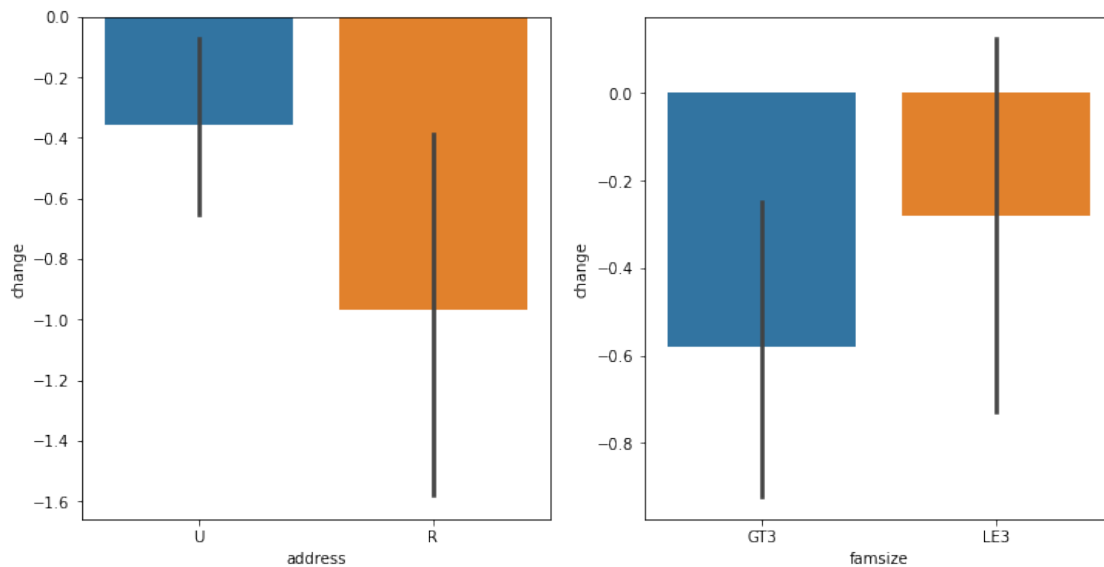


```
[ ]: #Observation:

     ->Students whose parent are working has mantained good grades then parents who␣
       ↪are staying at home.
     ->Parents working in health sector student has maitained good grades than␣
       ↪others.
```

```
[86]: fig, ax = plt.subplots(1,2, figsize=(12,6))

      sns.barplot(ax=ax[0], data=maths, x='address', y='change')
      sns.barplot(ax=ax[1], data=maths, x='famsize', y='change')
```

```
plt.show()
```



```
[ ]: #Observation:

     ->Student staying in urban has maintained decent score when comparing with␣
      ↪student staying in rural area.
     ->Student with less than 3 family members has maintained better score than␣
      ↪student having family members >3.
```

11 Summary

Many factor influence the grades but alcohol does not seem to have the influence that the title suggest

```
[ ]:
```