# HW02-320210207

December 9, 2023

#

Naive Bayes' Classifier

##

Student Performance Analysis

### 0.0.1 Student Credentials:

- Name: Yousef Ibrahim Gomaa Mahmoud
- ID: 320210207
- Section: AID 3
- Group: 1
- E-mail: yousef.gomaa@ejust.edu.eg

[ ]:

```
# Libraries needed:
# Pre-processing:
# Pandas
import pandas as pd

# Data Visualization:
# Standard Visualization Packages
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style='whitegrid')

# Machine Learning:
# Machine Learning Tools
from sklearn.metrics import accuracy_score, recall_score, precision_score,␣
 ↪confusion_matrix
from sklearn.model_selection import train_test_split
# Naive Bayes' Gaussian
from sklearn.naive_bayes import GaussianNB
```

### 0.0.2 Importing the Dataset into a Dataframe:

[ ]:

```
df = pd.read_csv("Student_Performance_Analysis - Sheet1.csv")
```

### 0.0.3 Data Preprocessing:

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 8 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   gender                       1000 non-null   object
 1   race/ethnicity               1000 non-null   object
 2   parental level of education  1000 non-null   object
 3   lunch                        1000 non-null   object
 4   test preparation course      1000 non-null   object
 5   math score                   1000 non-null   int64
 6   reading score                1000 non-null   int64
 7   writing score                1000 non-null   int64
dtypes: int64(3), object(5)
memory usage: 62.6+ KB
```

[ ]:
```
df.describe().T
```

[ ]:
```
                count    mean        std   min    25%   50%   75%    max
math score     1000.0  66.089  15.163080   0.0  57.00  66.0  77.0  100.0
reading score  1000.0  69.169  14.600192  17.0  59.00  70.0  79.0  100.0
writing score  1000.0  68.054  15.195657  10.0  57.75  69.0  79.0  100.0
```

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

[ ]:
```
gender                         0
race/ethnicity                 0
parental level of education    0
lunch                          0
test preparation course        0
math score                     0
reading score                  0
writing score                  0
dtype: int64
```

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

[ ]:
```
0
```

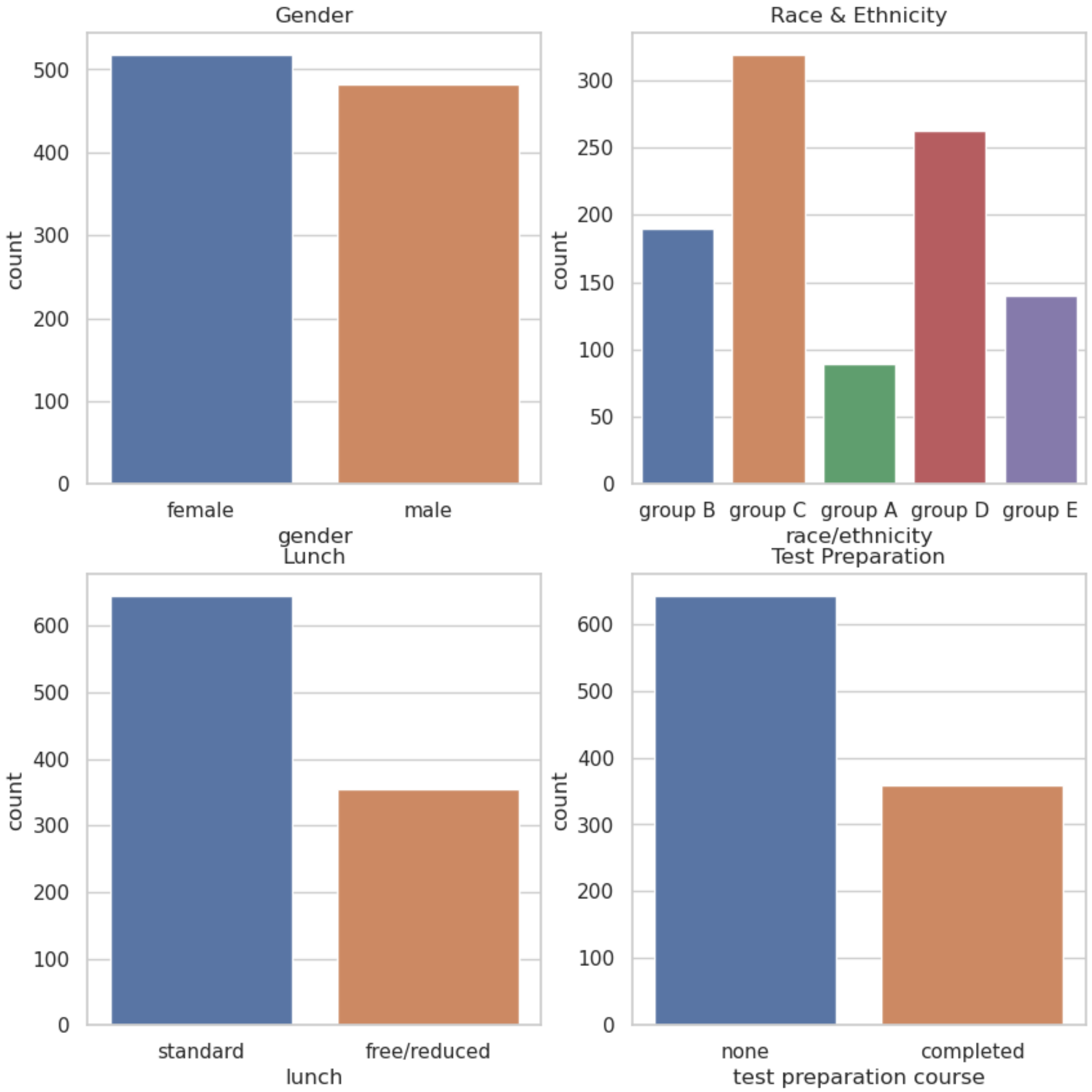
**Conclusion:**

- Data types are correctly casted.
- No missing values found in the data given.
- No duplicated records.

### 0.0.4 Data Visualization:

[ ]:

```
plt.figure(figsize = (10,10))
plt.subplot(2,2,1)
sns.countplot(x='gender', data=df)
plt.title('Gender')
plt.subplot(2,2,2)
sns.countplot(x='race/ethnicity', data=df)
plt.title('Race & Ethnicity')
plt.subplot(2,2,3)
sns.countplot(x='parental level of education', data=df)
plt.title('Parental Level of Education')
plt.subplot(2,2,3)
sns.countplot(x='lunch', data=df)
plt.title('Lunch')
plt.subplot(2,2,4)
sns.countplot(x='test preparation course', data=df)
plt.title('Test Preparation')
plt.show()
```



**Insights:**

- Higher frequency of females than males in the dataset.
- Groups C and D dominate the dataset by a margain.
- Lunch is standard most of the time.
- Most of the records have not finished the test preparation course.

```
[ ]: df.describe().T
```

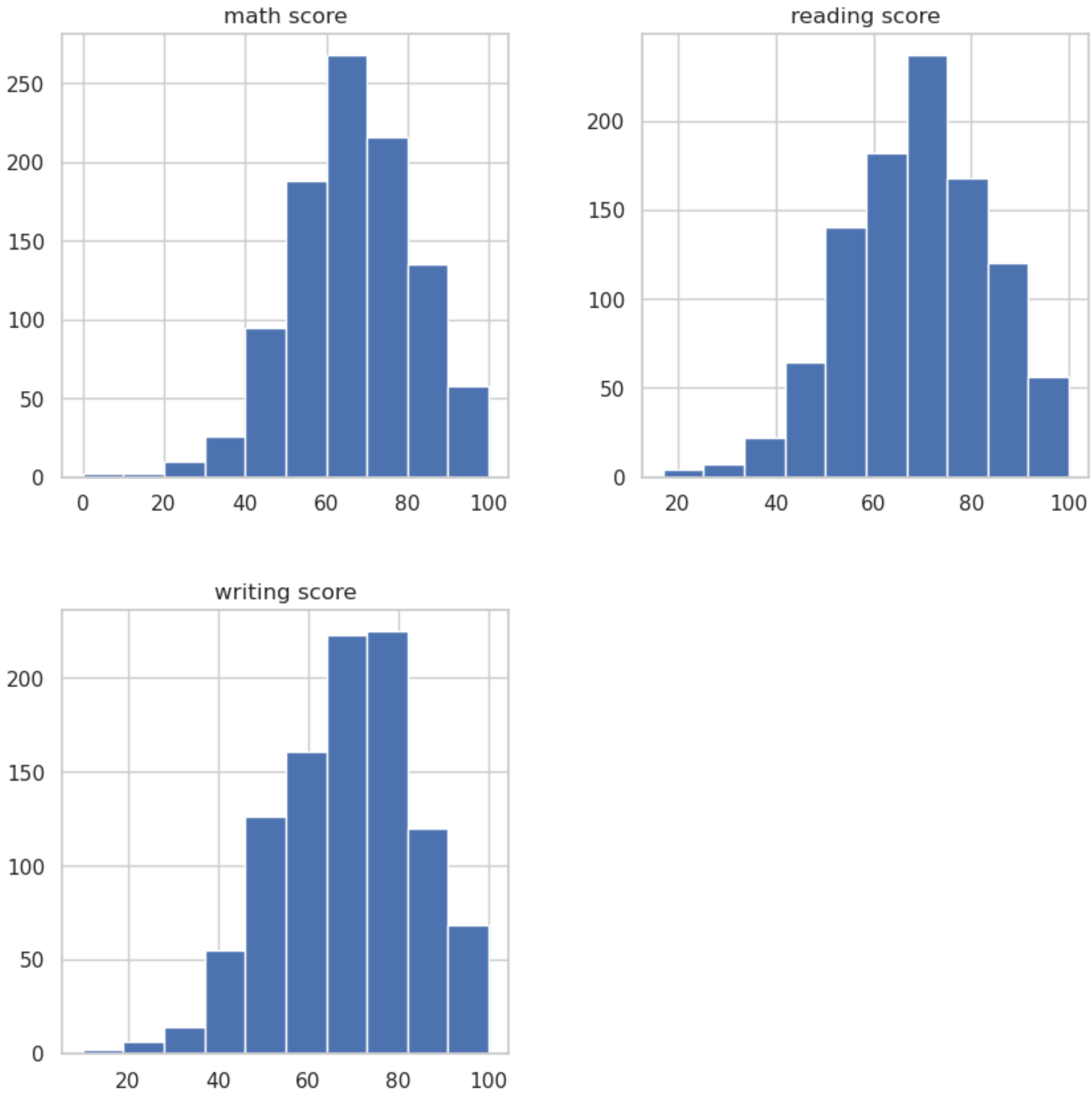
```
[ ]:                count    mean        std   min    25%   50%   75%    max
     math score     1000.0  66.089  15.163080   0.0  57.00  66.0  77.0  100.0
     reading score  1000.0  69.169  14.600192  17.0  59.00  70.0  79.0  100.0
     writing score  1000.0  68.054  15.195657  10.0  57.75  69.0  79.0  100.0
```

[ ]:

```
df.hist(figsize=(10,10))
```

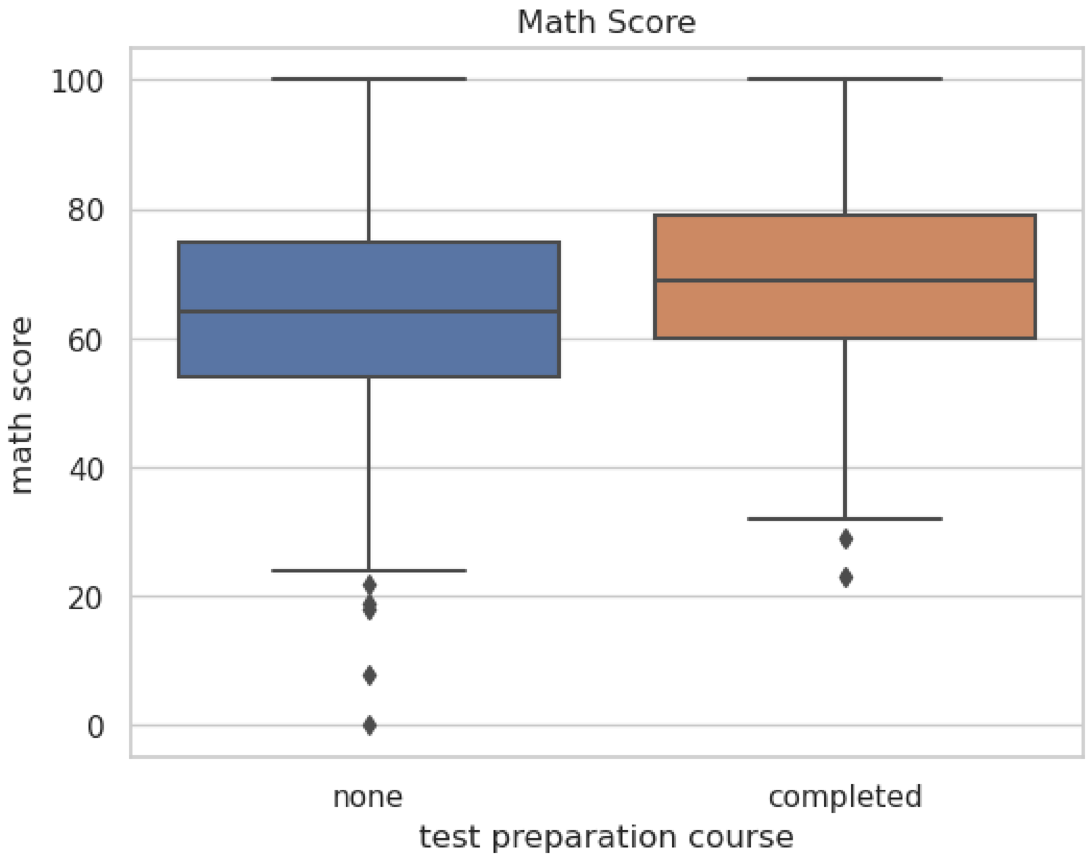
[ ]:

```
array([[<Axes: title={'center': 'math score'}>,
        <Axes: title={'center': 'reading score'}>],
       [<Axes: title={'center': 'writing score'}>, <Axes: >]],
      dtype=object)
```



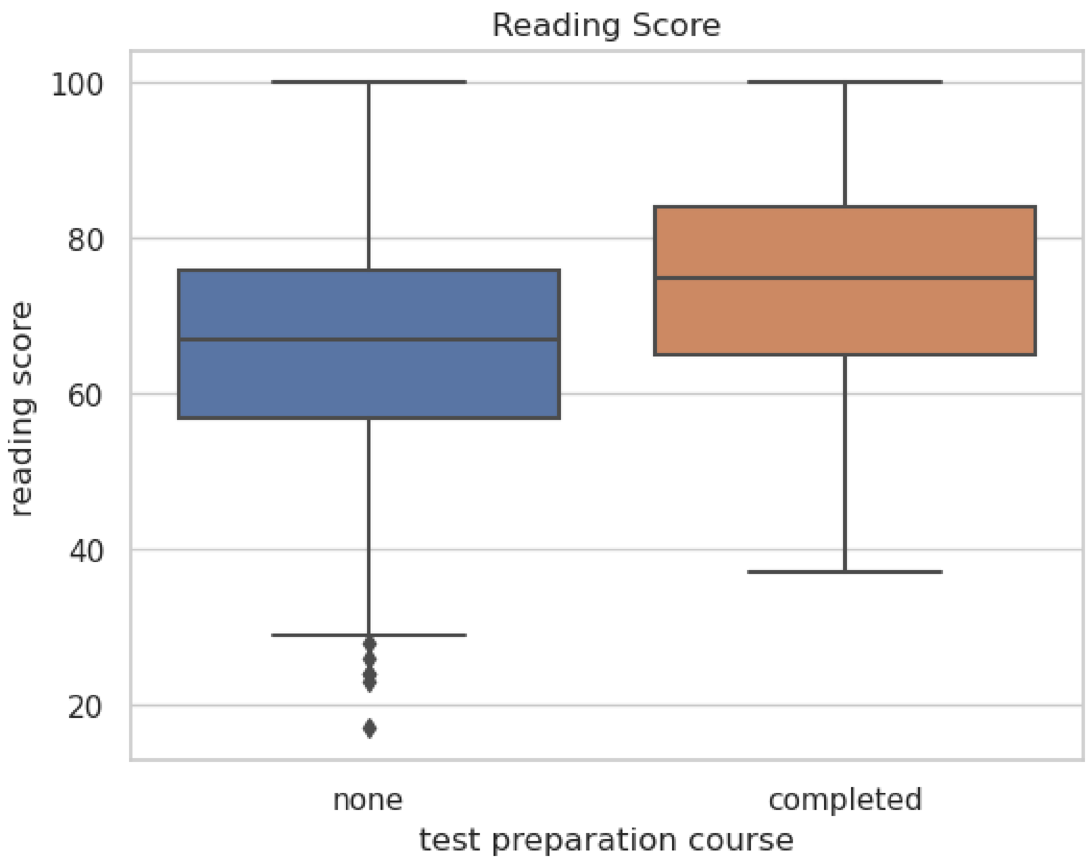
[ ]:

```
sns.boxplot(y='math score', x='test preparation course', data= df)
plt.title('Math Score')
plt.show()
```



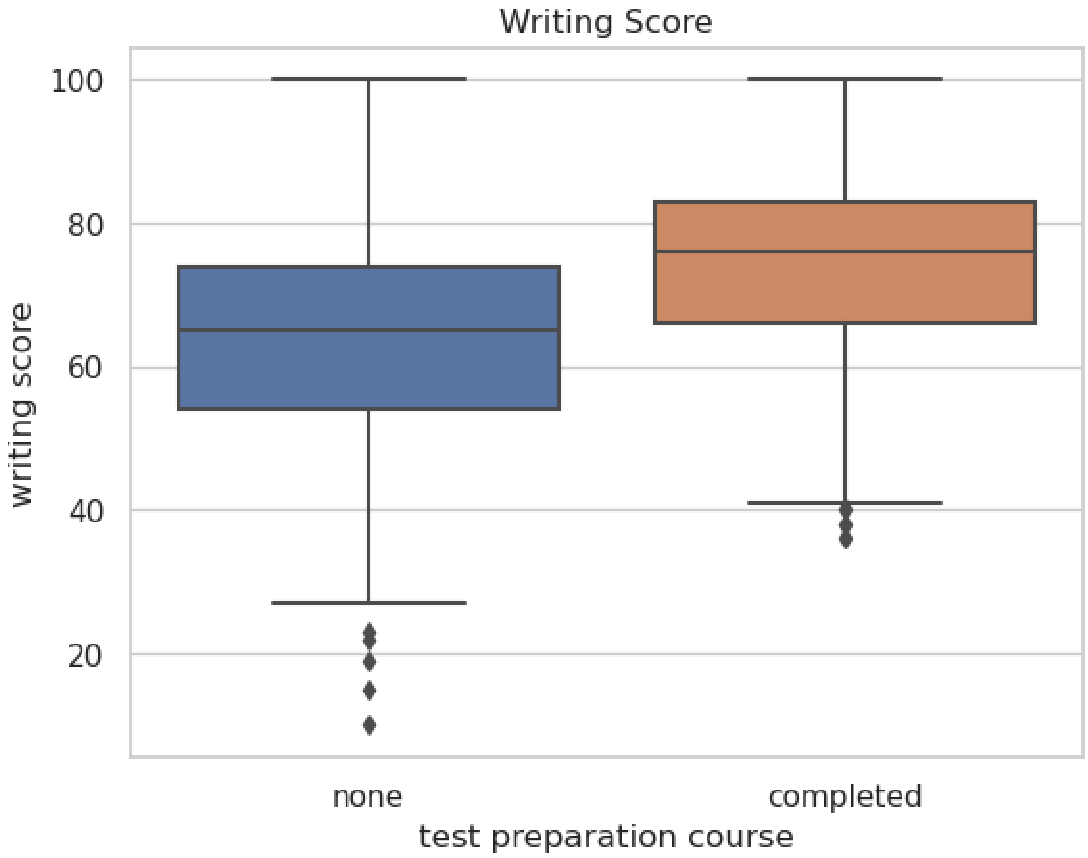
[ ]:

```
sns.boxplot(y='reading score', x='test preparation course', data= df)
plt.title('Reading Score')
plt.show()
```



[ ]:

```
sns.boxplot(y='writing score', x='test preparation course', data= df)
plt.title('Writing Score')
plt.show()
```

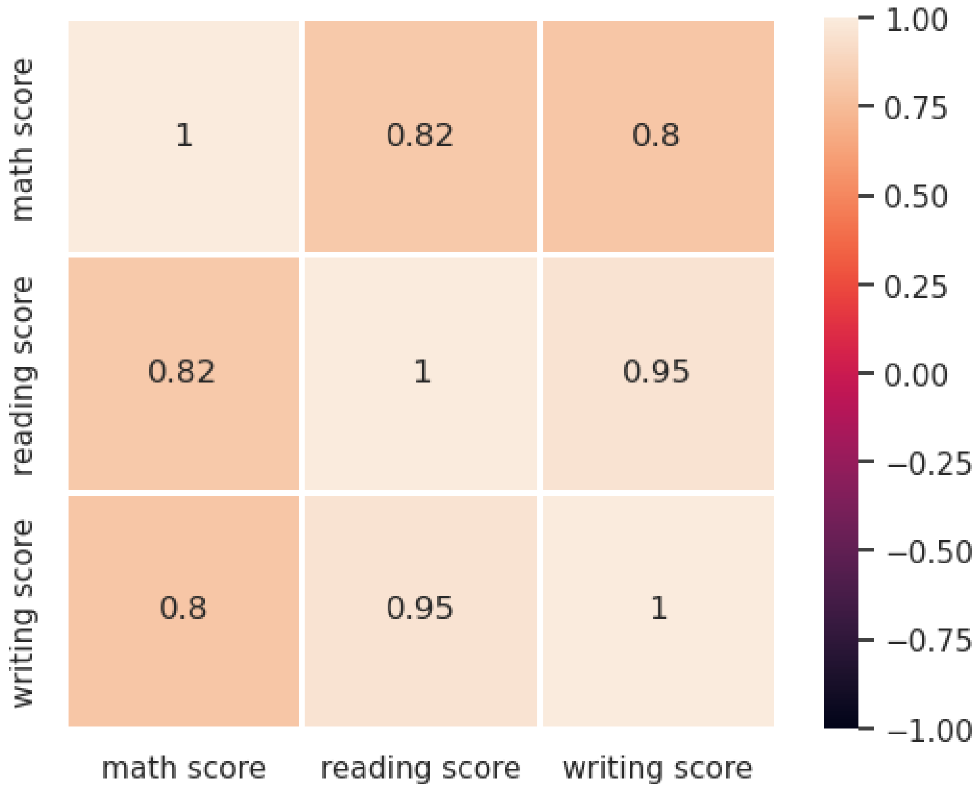


[ ]:

```
sns.heatmap(df.corr(numeric_only=True), linewidths=1, vmin=-1, vmax=1, annot=True, square=True)
```

[ ]:

```
<Axes: >
```



- Numeric only! ^

```
[ ]: df_num = df.copy()
```

```
[ ]: # Converting/encoding data to numerical values to re-calculate correlation.
     from sklearn.preprocessing import LabelEncoder

     laben = LabelEncoder()
     obj = df_num.select_dtypes(include='object')
     non_obj = df_num.select_dtypes(exclude='object')
     for i in range (0, obj.shape[1]) :
         obj.iloc[:, i] = laben.fit_transform(obj.iloc[:, i])
     df_num = pd.concat([obj, non_obj], axis = 1)
     df_num
```

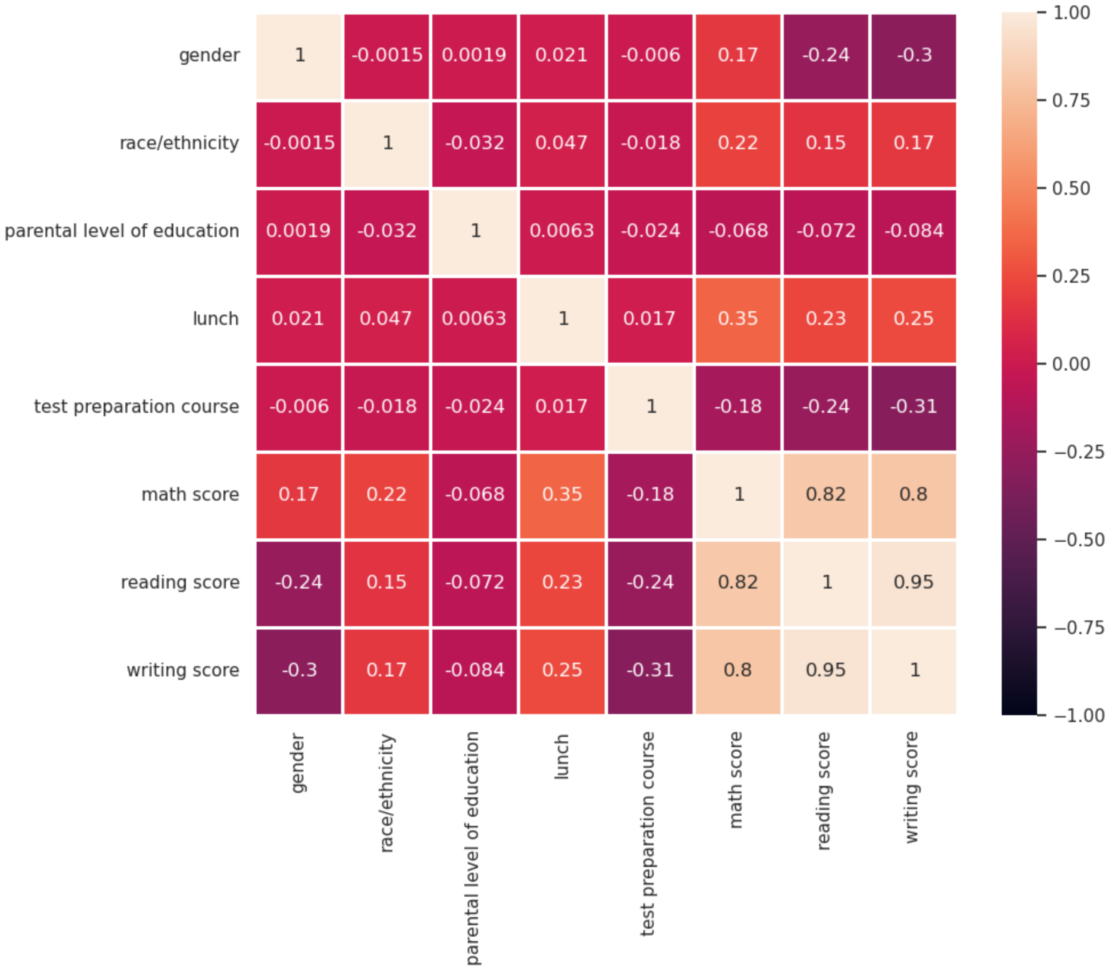
```
/tmp/ipykernel_110283/3754586768.py:8: DeprecationWarning: In a future version,
`df.iloc[:, i] = newvals` will attempt to set the values inplace instead of
always setting a new array. To retain the old behavior, use either
`df[df.columns[i]] = newvals` or, if columns are non-unique, `df.isetitem(i,
newvals)`
  obj.iloc[:, i] = laben.fit_transform(obj.iloc[:, i])
/tmp/ipykernel_110283/3754586768.py:8: DeprecationWarning: In a future version,
`df.iloc[:, i] = newvals` will attempt to set the values inplace instead of
always setting a new array. To retain the old behavior, use either
`df[df.columns[i]] = newvals` or, if columns are non-unique, `df.isetitem(i,
newvals)`
  obj.iloc[:, i] = laben.fit_transform(obj.iloc[:, i])
/tmp/ipykernel_110283/3754586768.py:8: DeprecationWarning: In a future version,
`df.iloc[:, i] = newvals` will attempt to set the values inplace instead of
always setting a new array. To retain the old behavior, use either
`df[df.columns[i]] = newvals` or, if columns are non-unique, `df.isetitem(i,
newvals)`
  obj.iloc[:, i] = laben.fit_transform(obj.iloc[:, i])
/tmp/ipykernel_110283/3754586768.py:8: DeprecationWarning: In a future version,
`df.iloc[:, i] = newvals` will attempt to set the values inplace instead of
always setting a new array. To retain the old behavior, use either
`df[df.columns[i]] = newvals` or, if columns are non-unique, `df.isetitem(i,
newvals)`
  obj.iloc[:, i] = laben.fit_transform(obj.iloc[:, i])
/tmp/ipykernel_110283/3754586768.py:8: DeprecationWarning: In a future version,
`df.iloc[:, i] = newvals` will attempt to set the values inplace instead of
always setting a new array. To retain the old behavior, use either
`df[df.columns[i]] = newvals` or, if columns are non-unique, `df.isetitem(i,
newvals)`
  obj.iloc[:, i] = laben.fit_transform(obj.iloc[:, i])
```

```
[ ]:      gender  race/ethnicity  parental level of education  lunch  \
     0         0               1                            1      1
     1         0               2                            4      1
     2         0               1                            3      1
     3         1               0                            0      0
     4         1               2                            4      1
     ..      …             …                           …    …
     995       0               4                            3      1
     996       1               2                            2      0
     997       0               2                            2      0
     998       0               3                            4      1
     999       0               3                            4      0

          test preparation course  math score  reading score  writing score
     0                          1          72             72             74
     1                          0          69             90             88
     2                          1          90             95             93
     3                          1          47             57             44
     4                          1          76             78             75
     ..                       …         …             …             …
     995                        0          88             99             95
     996                        1          62             55             55
     997                        0          59             71             65
```

```
998                                  0          68            78            77
999                                  1          77            86            86

[1000 rows x 8 columns]
```

[ ]:
```
plt.figure(figsize=(10,8))
sns.heatmap(df_num.corr(numeric_only=True), linewidths=1, vmin=-1, vmax=1, annot=True, square=True)
```

[ ]:
```
<Axes: >
```



### 0.0.5 Machine Learning:

**Naive Bayes' Algorithm:**

- predicting "Test Preparation Course" target.

**Using standard python code:**

[ ]:

```
counts = df['test preparation course'].value_counts()
prob = []
for i in range(len(counts)):
    prob.append(counts[i]/len(df))

columns = df.drop('test preparation course', axis=1)
columns
```

[ ]:

```
     gender race/ethnicity parental level of education         lunch  \
0    female        group B           bachelor's degree      standard
1    female        group C                some college      standard
2    female        group B             master's degree      standard
3      male        group A          associate's degree  free/reduced
4      male        group C                some college      standard
..      …            …                         …           …
995  female        group E             master's degree      standard
996    male        group C                 high school  free/reduced
997  female        group C                 high school  free/reduced
998  female        group D                some college      standard
999  female        group D                some college  free/reduced

     math score  reading score  writing score
0            72             72             74
1            69             90             88
2            90             95             93
3            47             57             44
4            76             78             75
..          …            …            …
995          88             99             95
996          62             55             55
997          59             71             65
998          68             78             77
999          77             86             86

[1000 rows x 7 columns]
```

[ ]:

```
prob = {}

def func(col_n, col_item):
    completed = len(df[(df['test preparation␣
 ↪course']=="completed")&(df[col_n]== col_item)])/counts[0]
    none = len(df[(df['test preparation course']=="none")&(df[col_n]==␣
 ↪col_item)])/counts[1]
    return {str(col_n)+" | "+str(col_item)+" | completed": completed,␣
 ↪str(col_n)+" | "+str(col_item)+" | none": none}

for i in columns:
    items = df[i].unique()
    for j in items:
        new_prob = func(i,j)
        prob.update(new_prob)

prob
```

```
[ ]: {'gender | female | completed': 0.2866043613707165,
      'gender | female | none': 0.9329608938547486,
      'gender | male | completed': 0.27102803738317754,
      'gender | male | none': 0.8603351955307262,
      'race/ethnicity | group B | completed': 0.1059190031152648,
      'race/ethnicity | group B | none': 0.3407821229050279,
      'race/ethnicity | group C | completed': 0.1822429906542056,
      'race/ethnicity | group C | none': 0.5642458100558659,
      'race/ethnicity | group A | completed': 0.048286604361370715,
      'race/ethnicity | group A | none': 0.16201117318435754,
      'race/ethnicity | group D | completed': 0.1277258566978193,
      'race/ethnicity | group D | none': 0.5027932960893855,
      'race/ethnicity | group E | completed': 0.09345794392523364,
      'race/ethnicity | group E | none': 0.22346368715083798,
      "parental level of education | bachelor's degree | completed":
     0.07165109034267912,
      "parental level of education | bachelor's degree | none": 0.2011173184357542,
      'parental level of education | some college | completed': 0.11993769470404984,
      'parental level of education | some college | none': 0.41620111731843573,
      "parental level of education | master's degree | completed":
     0.03115264797507788,
      "parental level of education | master's degree | none": 0.10893854748603352,
      "parental level of education | associate's degree | completed":
     0.1277258566978193,
      "parental level of education | associate's degree | none": 0.39106145251396646,
      'parental level of education | high school | completed': 0.08722741433021806,
      'parental level of education | high school | none': 0.39106145251396646,
      'parental level of education | some high school | completed':
     0.11993769470404984,
      'parental level of education | some high school | none': 0.2849162011173184,
      'lunch | standard | completed': 0.35358255451713394,
      'lunch | standard | none': 1.1675977653631284,
      'lunch | free/reduced | completed': 0.20404984423676012,
      'lunch | free/reduced | none': 0.6256983240223464,
      'math score | 72 | completed': 0.009345794392523364,
      'math score | 72 | none': 0.0335195530726257,
      'math score | 69 | completed': 0.01557632398753894,
      'math score | 69 | none': 0.061452513966480445,
      'math score | 90 | completed': 0.001557632398753894,
      'math score | 90 | none': 0.019553072625698324,
      'math score | 47 | completed': 0.004672897196261682,
      'math score | 47 | none': 0.0223463687150838,
      'math score | 76 | completed': 0.012461059190031152,
      'math score | 76 | none': 0.036312849162011177,
      'math score | 71 | completed': 0.01557632398753894,
      'math score | 71 | none': 0.0446927374301676,
      'math score | 88 | completed': 0.012461059190031152,
      'math score | 88 | none': 0.019553072625698324,
      'math score | 40 | completed': 0.004672897196261682,
      'math score | 40 | none': 0.019553072625698324,
      'math score | 64 | completed': 0.009345794392523364,
      'math score | 64 | none': 0.03910614525139665,
      'math score | 38 | completed': 0.0,
      'math score | 38 | none': 0.008379888268156424,
      'math score | 58 | completed': 0.012461059190031152,
      'math score | 58 | none': 0.04748603351955307,
      'math score | 65 | completed': 0.024922118380062305,
      'math score | 65 | none': 0.055865921787709494,
      'math score | 78 | completed': 0.014018691588785047,
      'math score | 78 | none': 0.013966480446927373,
      'math score | 50 | completed': 0.00778816199376947,
      'math score | 50 | none': 0.027932960893854747,
      'math score | 18 | completed': 0.0,
      'math score | 18 | none': 0.002793296089385475,
      'math score | 46 | completed': 0.006230529595015576,
      'math score | 46 | none': 0.019553072625698324,
      'math score | 54 | completed': 0.003115264797507788,
      'math score | 54 | none': 0.0446927374301676,
      'math score | 66 | completed': 0.012461059190031152,
      'math score | 66 | none': 0.0446927374301676,
      'math score | 44 | completed': 0.001557632398753894,
      'math score | 44 | none': 0.0223463687150838,
      'math score | 74 | completed': 0.017133956386292833,
      'math score | 74 | none': 0.03910614525139665,
      'math score | 73 | completed': 0.010903426791277258,
      'math score | 73 | none': 0.055865921787709494,
      'math score | 67 | completed': 0.021806853582554516,
      'math score | 67 | none': 0.0335195530726257,
      'math score | 70 | completed': 0.012461059190031152,
      'math score | 70 | none': 0.027932960893854747,
      'math score | 62 | completed': 0.010903426791277258,
      'math score | 62 | none': 0.0782122905027933,
      'math score | 63 | completed': 0.014018691588785047,
      'math score | 63 | none': 0.04748603351955307,
      'math score | 56 | completed': 0.006230529595015576,
      'math score | 56 | none': 0.013966480446927373,
      'math score | 97 | completed': 0.004672897196261682,
      'math score | 97 | none': 0.008379888268156424,
      'math score | 81 | completed': 0.012461059190031152,
      'math score | 81 | none': 0.03910614525139665,
      'math score | 75 | completed': 0.012461059190031152,
      'math score | 75 | none': 0.036312849162011177,
      'math score | 57 | completed': 0.012461059190031152,
      'math score | 57 | none': 0.027932960893854747,
      'math score | 55 | completed': 0.006230529595015576,
      'math score | 55 | none': 0.03910614525139665,
      'math score | 53 | completed': 0.00778816199376947,
      'math score | 53 | none': 0.05307262569832402,
      'math score | 59 | completed': 0.017133956386292833,
      'math score | 59 | none': 0.05865921787709497,
      'math score | 82 | completed': 0.012461059190031152,
      'math score | 82 | none': 0.027932960893854747,
      'math score | 77 | completed': 0.017133956386292833,
      'math score | 77 | none': 0.036312849162011177,
      'math score | 33 | completed': 0.0,
      'math score | 33 | none': 0.002793296089385475,
      'math score | 52 | completed': 0.012461059190031152,
      'math score | 52 | none': 0.027932960893854747,
      'math score | 0 | completed': 0.0,
      'math score | 0 | none': 0.002793296089385475,
      'math score | 79 | completed': 0.01557632398753894,
      'math score | 79 | none': 0.0335195530726257,
      'math score | 39 | completed': 0.001557632398753894,
      'math score | 39 | none': 0.008379888268156424,
      'math score | 45 | completed': 0.003115264797507788,
      'math score | 45 | none': 0.019553072625698324,
      'math score | 60 | completed': 0.009345794392523364,
      'math score | 60 | none': 0.027932960893854747,
      'math score | 61 | completed': 0.012461059190031152,
      'math score | 61 | none': 0.05307262569832402,
      'math score | 41 | completed': 0.0,
      'math score | 41 | none': 0.01675977653631285,
      'math score | 49 | completed': 0.004672897196261682,
      'math score | 49 | none': 0.03910614525139665,
      'math score | 30 | completed': 0.0,
      'math score | 30 | none': 0.00558659217877095,
      'math score | 80 | completed': 0.006230529595015576,
      'math score | 80 | none': 0.036312849162011177,
      'math score | 42 | completed': 0.004672897196261682,
      'math score | 42 | none': 0.008379888268156424,
      'math score | 27 | completed': 0.0,
      'math score | 27 | none': 0.00558659217877095,
      'math score | 43 | completed': 0.003115264797507788,
      'math score | 43 | none': 0.008379888268156424,
      'math score | 68 | completed': 0.017133956386292833,
      'math score | 68 | none': 0.04189944134078212,
      'math score | 85 | completed': 0.012461059190031152,
      'math score | 85 | none': 0.01675977653631285,
      'math score | 98 | completed': 0.003115264797507788,
      'math score | 98 | none': 0.002793296089385475,
      'math score | 87 | completed': 0.014018691588785047,
      'math score | 87 | none': 0.019553072625698324,
      'math score | 51 | completed': 0.006230529595015576,
      'math score | 51 | none': 0.019553072625698324,
      'math score | 99 | completed': 0.003115264797507788,
      'math score | 99 | none': 0.002793296089385475,
      'math score | 84 | completed': 0.004672897196261682,
      'math score | 84 | none': 0.0223463687150838,
      'math score | 91 | completed': 0.00778816199376947,
      'math score | 91 | none': 0.0111731843575419,
      'math score | 83 | completed': 0.004672897196261682,
      'math score | 83 | none': 0.013966480446927373,
      'math score | 89 | completed': 0.003115264797507788,
      'math score | 89 | none': 0.0111731843575419,
      'math score | 22 | completed': 0.0,
      'math score | 22 | none': 0.002793296089385475,
      'math score | 100 | completed': 0.006230529595015576,
      'math score | 100 | none': 0.008379888268156424,
      'math score | 96 | completed': 0.004672897196261682,
      'math score | 96 | none': 0.0,
      'math score | 94 | completed': 0.00778816199376947,
      'math score | 94 | none': 0.00558659217877095,
      'math score | 48 | completed': 0.001557632398753894,
      'math score | 48 | none': 0.027932960893854747,
      'math score | 35 | completed': 0.001557632398753894,
      'math score | 35 | none': 0.0111731843575419,
      'math score | 34 | completed': 0.001557632398753894,
      'math score | 34 | none': 0.002793296089385475,
      'math score | 86 | completed': 0.004672897196261682,
      'math score | 86 | none': 0.013966480446927373,
      'math score | 92 | completed': 0.004672897196261682,
      'math score | 92 | none': 0.008379888268156424,
      'math score | 37 | completed': 0.0,
      'math score | 37 | none': 0.0111731843575419,
      'math score | 28 | completed': 0.0,
      'math score | 28 | none': 0.002793296089385475,
      'math score | 24 | completed': 0.0,
      'math score | 24 | none': 0.002793296089385475,
      'math score | 26 | completed': 0.0,
      'math score | 26 | none': 0.002793296089385475,
      'math score | 95 | completed': 0.001557632398753894,
      'math score | 95 | none': 0.002793296089385475,
      'math score | 36 | completed': 0.0,
      'math score | 36 | none': 0.00558659217877095,
      'math score | 29 | completed': 0.001557632398753894,
      'math score | 29 | none': 0.00558659217877095,
      'math score | 32 | completed': 0.001557632398753894,
      'math score | 32 | none': 0.00558659217877095,
      'math score | 93 | completed': 0.004672897196261682,
      'math score | 93 | none': 0.002793296089385475,
      'math score | 19 | completed': 0.0,
      'math score | 19 | none': 0.002793296089385475,
      'math score | 23 | completed': 0.001557632398753894,
      'math score | 23 | none': 0.0,
      'math score | 8 | completed': 0.0,
      'math score | 8 | none': 0.002793296089385475,
      'reading score | 72 | completed': 0.012461059190031152,
      'reading score | 72 | none': 0.07262569832402235,
      'reading score | 90 | completed': 0.014018691588785047,
      'reading score | 90 | none': 0.0223463687150838,
      'reading score | 95 | completed': 0.004672897196261682,
      'reading score | 95 | none': 0.013966480446927373,
      'reading score | 57 | completed': 0.003115264797507788,
      'reading score | 57 | none': 0.04189944134078212,
      'reading score | 78 | completed': 0.021806853582554516,
      'reading score | 78 | none': 0.036312849162011177,
      'reading score | 83 | completed': 0.009345794392523364,
      'reading score | 83 | none': 0.0223463687150838,
      'reading score | 43 | completed': 0.001557632398753894,
      'reading score | 43 | none': 0.0223463687150838,
      'reading score | 64 | completed': 0.014018691588785047,
      'reading score | 64 | none': 0.06424581005586592,
      'reading score | 60 | completed': 0.00778816199376947,
      'reading score | 60 | none': 0.0446927374301676,
      'reading score | 54 | completed': 0.00778816199376947,
      'reading score | 54 | none': 0.036312849162011177,
      'reading score | 52 | completed': 0.006230529595015576,
      'reading score | 52 | none': 0.0335195530726257,
      'reading score | 81 | completed': 0.018691588785046728,
      'reading score | 81 | none': 0.036312849162011177,
      'reading score | 53 | completed': 0.001557632398753894,
      'reading score | 53 | none': 0.0335195530726257,
      'reading score | 75 | completed': 0.021806853582554516,
      'reading score | 75 | none': 0.0335195530726257,
      'reading score | 89 | completed': 0.009345794392523364,
      'reading score | 89 | none': 0.01675977653631285,
      'reading score | 32 | completed': 0.0,
      'reading score | 32 | none': 0.002793296089385475,
      'reading score | 42 | completed': 0.003115264797507788,
      'reading score | 42 | none': 0.013966480446927373,
      'reading score | 58 | completed': 0.00778816199376947,
      'reading score | 58 | none': 0.06424581005586592,
      'reading score | 69 | completed': 0.00778816199376947,
      'reading score | 69 | none': 0.027932960893854747,
      'reading score | 73 | completed': 0.010903426791277258,
      'reading score | 73 | none': 0.06424581005586592,
      'reading score | 71 | completed': 0.012461059190031152,
      'reading score | 71 | none': 0.0335195530726257,
      'reading score | 74 | completed': 0.0264797507788162,
      'reading score | 74 | none': 0.0446927374301676,
      'reading score | 70 | completed': 0.01557632398753894,
      'reading score | 70 | none': 0.0446927374301676,
      'reading score | 65 | completed': 0.00778816199376947,
      'reading score | 65 | none': 0.03910614525139665,
      'reading score | 87 | completed': 0.00778816199376947,
      'reading score | 87 | none': 0.0223463687150838,
      'reading score | 56 | completed': 0.00778816199376947,
      'reading score | 56 | none': 0.030726256983240222,
      'reading score | 61 | completed': 0.012461059190031152,
      'reading score | 61 | none': 0.0446927374301676,
      'reading score | 84 | completed': 0.02336448598130841,
      'reading score | 84 | none': 0.030726256983240222,
      'reading score | 55 | completed': 0.006230529595015576,
      'reading score | 55 | none': 0.0335195530726257,
      'reading score | 44 | completed': 0.003115264797507788,
      'reading score | 44 | none': 0.00558659217877095,
      'reading score | 41 | completed': 0.003115264797507788,
      'reading score | 41 | none': 0.0111731843575419,
      'reading score | 85 | completed': 0.017133956386292833,
      'reading score | 85 | none': 0.019553072625698324,
      'reading score | 59 | completed': 0.006230529595015576,
      'reading score | 59 | none': 0.036312849162011177,
      'reading score | 17 | completed': 0.0,
      'reading score | 17 | none': 0.002793296089385475,
      'reading score | 39 | completed': 0.0,
      'reading score | 39 | none': 0.01675977653631285,
      'reading score | 80 | completed': 0.00778816199376947,
      'reading score | 80 | none': 0.025139664804469275,
      'reading score | 37 | completed': 0.001557632398753894,
      'reading score | 37 | none': 0.00558659217877095,
      'reading score | 63 | completed': 0.006230529595015576,
      'reading score | 63 | none': 0.0446927374301676,
      'reading score | 51 | completed': 0.012461059190031152,
      'reading score | 51 | none': 0.0223463687150838,
      'reading score | 49 | completed': 0.003115264797507788,
      'reading score | 49 | none': 0.0223463687150838,
      'reading score | 26 | completed': 0.0,
      'reading score | 26 | none': 0.002793296089385475,
      'reading score | 68 | completed': 0.012461059190031152,
      'reading score | 68 | none': 0.04189944134078212,
      'reading score | 45 | completed': 0.001557632398753894,
      'reading score | 45 | none': 0.01675977653631285,
      'reading score | 47 | completed': 0.001557632398753894,
      'reading score | 47 | none': 0.008379888268156424,
      'reading score | 86 | completed': 0.014018691588785047,
      'reading score | 86 | none': 0.027932960893854747,
      'reading score | 34 | completed': 0.0,
      'reading score | 34 | none': 0.0111731843575419,
      'reading score | 79 | completed': 0.014018691588785047,
      'reading score | 79 | none': 0.027932960893854747,
      'reading score | 66 | completed': 0.014018691588785047,
      'reading score | 66 | none': 0.05027932960893855,
      'reading score | 67 | completed': 0.014018691588785047,
      'reading score | 67 | none': 0.05865921787709497,
      'reading score | 91 | completed': 0.003115264797507788,
      'reading score | 91 | none': 0.0111731843575419,
      'reading score | 100 | completed': 0.017133956386292833,
      'reading score | 100 | none': 0.01675977653631285,
      'reading score | 76 | completed': 0.009345794392523364,
      'reading score | 76 | none': 0.05307262569832402,
      'reading score | 77 | completed': 0.017133956386292833,
      'reading score | 77 | none': 0.036312849162011177,
      'reading score | 82 | completed': 0.020249221183800622,
      'reading score | 82 | none': 0.019553072625698324,
      'reading score | 92 | completed': 0.00778816199376947,
      'reading score | 92 | none': 0.013966480446927373,
      'reading score | 93 | completed': 0.003115264797507788,
      'reading score | 93 | none': 0.0111731843575419,
      'reading score | 62 | completed': 0.009345794392523364,
      'reading score | 62 | none': 0.0446927374301676,
      'reading score | 88 | completed': 0.00778816199376947,
      'reading score | 88 | none': 0.0111731843575419,
      'reading score | 50 | completed': 0.003115264797507788,
      'reading score | 50 | none': 0.013966480446927373,
      'reading score | 28 | completed': 0.0,
      'reading score | 28 | none': 0.002793296089385475,
      'reading score | 48 | completed': 0.003115264797507788,
      'reading score | 48 | none': 0.0223463687150838,
      'reading score | 46 | completed': 0.001557632398753894,
      'reading score | 46 | none': 0.013966480446927373,
      'reading score | 23 | completed': 0.0,
      'reading score | 23 | none': 0.002793296089385475,
      'reading score | 38 | completed': 0.0,
      'reading score | 38 | none': 0.00558659217877095,
      'reading score | 94 | completed': 0.003115264797507788,
      'reading score | 94 | none': 0.002793296089385475,
      'reading score | 97 | completed': 0.004672897196261682,
      'reading score | 97 | none': 0.00558659217877095,
      'reading score | 99 | completed': 0.004672897196261682,
      'reading score | 99 | none': 0.0,
      'reading score | 31 | completed': 0.0,
      'reading score | 31 | none': 0.00558659217877095,
      'reading score | 96 | completed': 0.004672897196261682,
      'reading score | 96 | none': 0.002793296089385475,
      'reading score | 24 | completed': 0.0,
      'reading score | 24 | none': 0.00558659217877095,
      'reading score | 29 | completed': 0.0,
      'reading score | 29 | none': 0.00558659217877095,
      'reading score | 40 | completed': 0.001557632398753894,
      'reading score | 40 | none': 0.0,
      'writing score | 74 | completed': 0.018691588785046728,
      'writing score | 74 | none': 0.06424581005586592,
      'writing score | 88 | completed': 0.014018691588785047,
      'writing score | 88 | none': 0.00558659217877095,
      'writing score | 93 | completed': 0.006230529595015576,
      'writing score | 93 | none': 0.0111731843575419,
      'writing score | 44 | completed': 0.004672897196261682,
      'writing score | 44 | none': 0.013966480446927373,
      'writing score | 75 | completed': 0.014018691588785047,
      'writing score | 75 | none': 0.036312849162011177,
      'writing score | 78 | completed': 0.020249221183800622,
      'writing score | 78 | none': 0.0335195530726257,
      'writing score | 92 | completed': 0.009345794392523364,
      'writing score | 92 | none': 0.008379888268156424,
      'writing score | 39 | completed': 0.0,
      'writing score | 39 | none': 0.00558659217877095,
      'writing score | 67 | completed': 0.009345794392523364,
      'writing score | 67 | none': 0.05307262569832402,
      'writing score | 50 | completed': 0.003115264797507788,
      'writing score | 50 | none': 0.0223463687150838,
      'writing score | 52 | completed': 0.004672897196261682,
      'writing score | 52 | none': 0.04748603351955307,
      'writing score | 43 | completed': 0.001557632398753894,
      'writing score | 43 | none': 0.030726256983240222,
      'writing score | 73 | completed': 0.014018691588785047,
      'writing score | 73 | none': 0.05307262569832402,
      'writing score | 70 | completed': 0.018691588785046728,
      'writing score | 70 | none': 0.05865921787709497,
      'writing score | 58 | completed': 0.009345794392523364,
      'writing score | 58 | none': 0.0335195530726257,
      'writing score | 86 | completed': 0.010903426791277258,
      'writing score | 86 | none': 0.013966480446927373,
      'writing score | 28 | completed': 0.0,
      'writing score | 28 | none': 0.002793296089385475,
      'writing score | 46 | completed': 0.003115264797507788,
      'writing score | 46 | none': 0.0223463687150838,
      'writing score | 61 | completed': 0.001557632398753894,
      'writing score | 61 | none': 0.055865921787709494,
      'writing score | 63 | completed': 0.00778816199376947,
      'writing score | 63 | none': 0.027932960893854747,
      'writing score | 53 | completed': 0.009345794392523364,
      'writing score | 53 | none': 0.030726256983240222,
      'writing score | 80 | completed': 0.02336448598130841,
      'writing score | 80 | none': 0.0335195530726257,
      'writing score | 72 | completed': 0.014018691588785047,
      'writing score | 72 | none': 0.0446927374301676,
      'writing score | 55 | completed': 0.001557632398753894,
      'writing score | 55 | none': 0.036312849162011177,
      'writing score | 65 | completed': 0.00778816199376947,
      'writing score | 65 | none': 0.04748603351955307,
      'writing score | 38 | completed': 0.001557632398753894,
      'writing score | 38 | none': 0.013966480446927373,
      'writing score | 82 | completed': 0.01557632398753894,
      'writing score | 82 | none': 0.03910614525139665,
      'writing score | 79 | completed': 0.018691588785046728,
      'writing score | 79 | none': 0.019553072625698324,
      'writing score | 83 | completed': 0.00778816199376947,
      'writing score | 83 | none': 0.0223463687150838,
      'writing score | 59 | completed': 0.003115264797507788,
      'writing score | 59 | none': 0.030726256983240222,
      'writing score | 57 | completed': 0.004672897196261682,
      'writing score | 57 | none': 0.0446927374301676,
      'writing score | 54 | completed': 0.00778816199376947,
      'writing score | 54 | none': 0.055865921787709494,
      'writing score | 68 | completed': 0.02336448598130841,
      'writing score | 68 | none': 0.0446927374301676,
      'writing score | 66 | completed': 0.014018691588785047,
      'writing score | 66 | none': 0.04189944134078212,
      'writing score | 62 | completed': 0.010903426791277258,
      'writing score | 62 | none': 0.055865921787709494,
      'writing score | 76 | completed': 0.018691588785046728,
      'writing score | 76 | none': 0.036312849162011177,
      'writing score | 48 | completed': 0.001557632398753894,
      'writing score | 48 | none': 0.019553072625698324,
      'writing score | 42 | completed': 0.0,
      'writing score | 42 | none': 0.019553072625698324,
      'writing score | 87 | completed': 0.009345794392523364,
      'writing score | 87 | none': 0.013966480446927373,
      'writing score | 49 | completed': 0.004672897196261682,
      'writing score | 49 | none': 0.01675977653631285,
      'writing score | 10 | completed': 0.0,
      'writing score | 10 | none': 0.002793296089385475,
      'writing score | 34 | completed': 0.0,
      'writing score | 34 | none': 0.008379888268156424,
      'writing score | 71 | completed': 0.006230529595015576,
      'writing score | 71 | none': 0.03910614525139665,
      'writing score | 37 | completed': 0.0,
      'writing score | 37 | none': 0.00558659217877095,
      'writing score | 56 | completed': 0.006230529595015576,
      'writing score | 56 | none': 0.027932960893854747,
      'writing score | 41 | completed': 0.003115264797507788,
      'writing score | 41 | none': 0.01675977653631285,
      'writing score | 22 | completed': 0.0,
      'writing score | 22 | none': 0.002793296089385475,
      'writing score | 81 | completed': 0.020249221183800622,
      'writing score | 81 | none': 0.019553072625698324,
      'writing score | 45 | completed': 0.0,
      'writing score | 45 | none': 0.019553072625698324,
      'writing score | 36 | completed': 0.001557632398753894,
      'writing score | 36 | none': 0.008379888268156424,
      'writing score | 89 | completed': 0.004672897196261682,
      'writing score | 89 | none': 0.01675977653631285,
      'writing score | 47 | completed': 0.001557632398753894,
      'writing score | 47 | none': 0.025139664804469275,
      'writing score | 90 | completed': 0.010903426791277258,
      'writing score | 90 | none': 0.008379888268156424,
      'writing score | 100 | completed': 0.014018691588785047,
      'writing score | 100 | none': 0.013966480446927373,
      'writing score | 64 | completed': 0.010903426791277258,
      'writing score | 64 | none': 0.04189944134078212,
      'writing score | 98 | completed': 0.003115264797507788,
      'writing score | 98 | none': 0.0,
      'writing score | 51 | completed': 0.003115264797507788,
      'writing score | 51 | none': 0.04189944134078212,
      'writing score | 40 | completed': 0.001557632398753894,
      'writing score | 40 | none': 0.00558659217877095,
      'writing score | 84 | completed': 0.00778816199376947,
      'writing score | 84 | none': 0.030726256983240222,
      'writing score | 69 | completed': 0.010903426791277258,
      'writing score | 69 | none': 0.0446927374301676,
      'writing score | 33 | completed': 0.0,
      'writing score | 33 | none': 0.00558659217877095,
      'writing score | 60 | completed': 0.012461059190031152,
      'writing score | 60 | none': 0.0335195530726257,
      'writing score | 85 | completed': 0.01557632398753894,
      'writing score | 85 | none': 0.0111731843575419,
      'writing score | 91 | completed': 0.010903426791277258,
      'writing score | 91 | none': 0.0111731843575419,
      'writing score | 77 | completed': 0.021806853582554516,
      'writing score | 77 | none': 0.027932960893854747,
      'writing score | 27 | completed': 0.0,
      'writing score | 27 | none': 0.008379888268156424,
      'writing score | 94 | completed': 0.006230529595015576,
      'writing score | 94 | none': 0.00558659217877095,
      'writing score | 95 | completed': 0.00778816199376947,
      'writing score | 95 | none': 0.008379888268156424,
      'writing score | 19 | completed': 0.0,
      'writing score | 19 | none': 0.002793296089385475,
      'writing score | 35 | completed': 0.0,
      'writing score | 35 | none': 0.002793296089385475,
      'writing score | 32 | completed': 0.0,
      'writing score | 32 | none': 0.00558659217877095,
      'writing score | 96 | completed': 0.001557632398753894,
      'writing score | 96 | none': 0.008379888268156424,
      'writing score | 97 | completed': 0.001557632398753894,
      'writing score | 97 | none': 0.002793296089385475,
      'writing score | 99 | completed': 0.004672897196261682,
      'writing score | 99 | none': 0.002793296089385475,
      'writing score | 15 | completed': 0.0,
      'writing score | 15 | none': 0.002793296089385475,
      'writing score | 30 | completed': 0.0,
      'writing score | 30 | none': 0.002793296089385475,
      'writing score | 23 | completed': 0.0,
      'writing score | 23 | none': 0.002793296089385475}
```

**Using the NB Gaussian module from "Scikit Learn" library:**

```
[ ]: # Using the numeric encoded dataframe
     X = df_num.drop('test preparation course', axis=1)
     Y = df_num['test preparation course']
```

```
[ ]: x_train, x_test, y_train, y_test = train_test_split(X,Y, test_size=0.3,␣
      ↪random_state=1)
```

```
[ ]: classifier = GaussianNB()
     classifier.fit(x_train, y_train)
```

```
[ ]: GaussianNB()
```

```
[ ]: predictions = classifier.predict(x_test)
```

```
predictions[:10]
```

```
[ ]: array([1, 1, 0, 1, 1, 1, 1, 1, 1, 0])
```

### NB Gaussian Accuracy:

```
[ ]: accuracy = print('Accuracy Score: ', format(accuracy_score(y_test,␣
      ↪predictions)))
```

```
Accuracy Score:  0.67
```

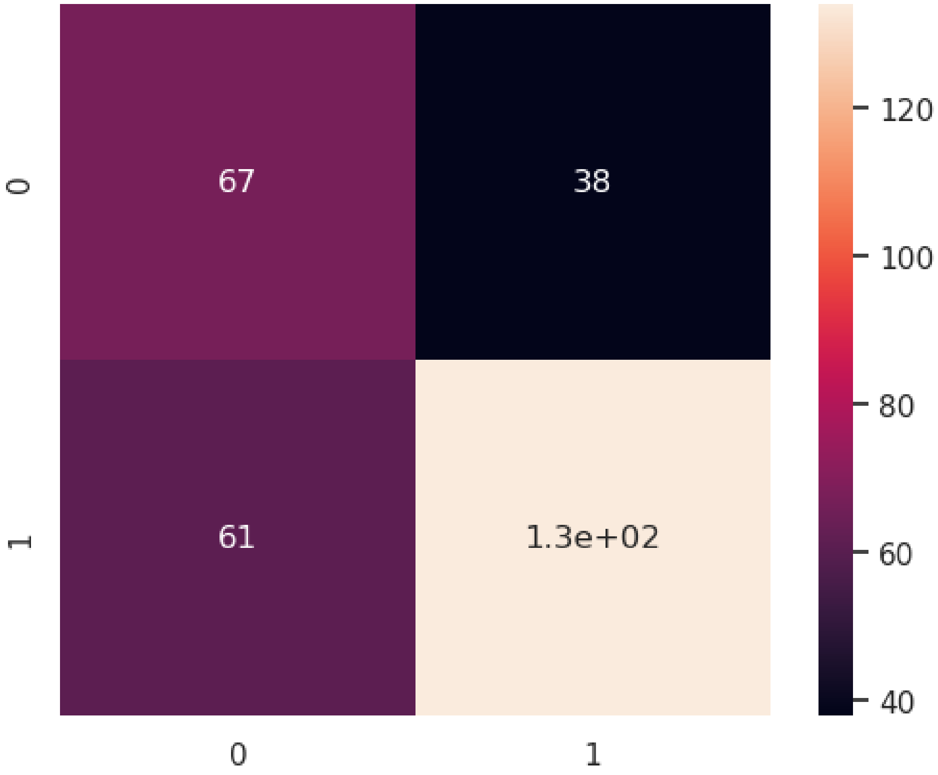
### Accuracy Report, Confusion Matrix:

```
[ ]: df_cf = confusion_matrix(y_test, predictions)
     df_cf
```

```
[ ]: array([[ 67,  38],
            [ 61, 134]])
```

```
[ ]: sns.heatmap(df_cf, annot=True, square=True)
```

```
[ ]: <Axes: >
```



```
[ ]: from sklearn.metrics import accuracy_score, precision_score, recall_score,␣
      ↪f1_score

     # Accuracy
     accuracy = print('Accuracy Score: ', format(accuracy_score(y_test,␣
      ↪predictions)))
     # Precision
     precision = print('Precision Score: ', format(precision_score(y_test,␣
      ↪predictions)))
     # Recall
     recall = print('Sensitivity/Recall Score: ', format(recall_score(y_test,␣
      ↪predictions)))
     # F1-score
     f1_score = print('F1-Measure/F1-Score: ', format(f1_score(y_test, predictions)))
```

```
Accuracy Score:  0.67
Precision Score:  0.7790697674418605
Sensitivity/Recall Score:  0.6871794871794872
F1-Measure/F1-Score:  0.7302452316076294
```