

PDS Assignment – 1 (Question – 2)

Name: Sai Pavan Pratapagiri
ID: 16343743

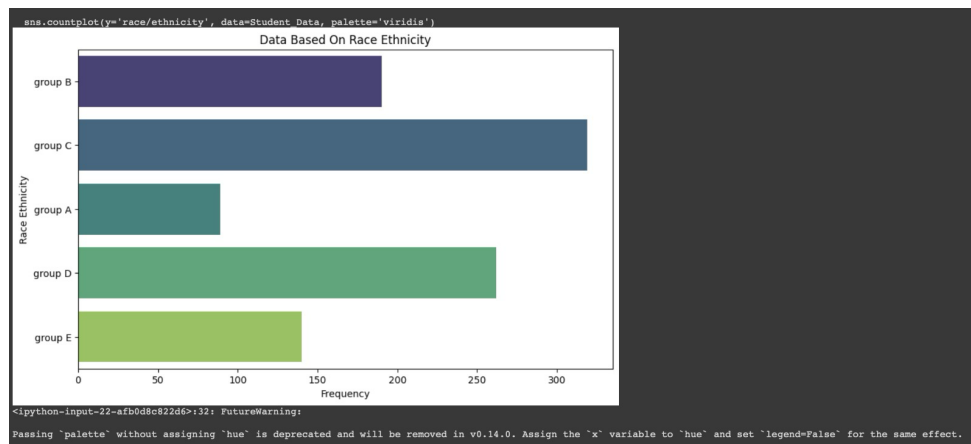
- Data Import and Inspection, DataFrame Creation, Data Exploration, the head () function gives a preview of the data, allowing for an initial assessment of its format and content. Summary statistics (describe ()) provide insights into the central tendency, dispersion, and distribution of numerical data.

```
<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   int64  
 1   race/ethnicity       1000 non-null   object  
 2   parental level of education 1000 non-null   object  
 3   test preparation course 1000 non-null   int64  
 4   math score          1000 non-null   int64  
 5   reading score       1000 non-null   int64  
 6   writing score        1000 non-null   int64  
 7   Total score         1000 non-null   int64  
dtypes: int64(6), object(2)
memory usage: 62.6+ KB

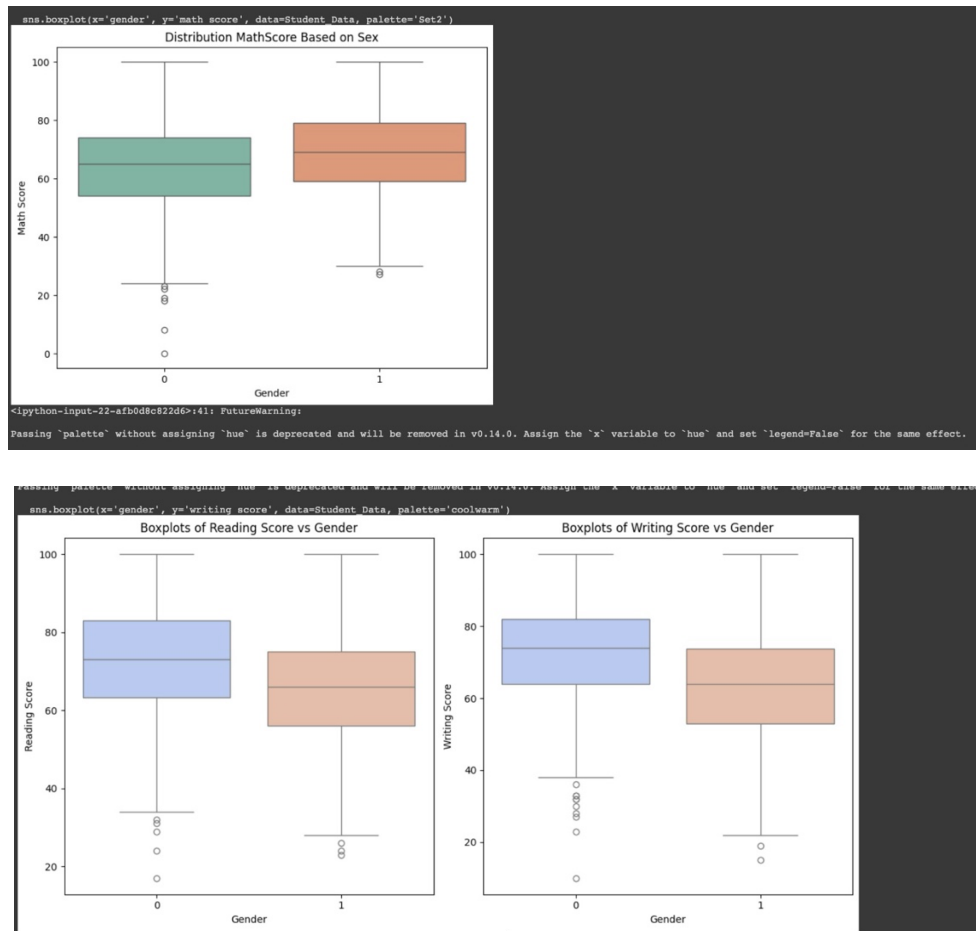
Index(['gender', 'race/ethnicity', 'parental level of education',
      'test preparation course', 'math score', 'reading score',
      'writing score', 'Total score'],
      dtype='object')
(1000, 8)
   gender  race/ethnicity  parental level of education  test preparation course \
0      0      group B      bachelor's degree          0                    \
1      0      group C      some college              1                    \
2      0      group B      master's degree            0                    \
3      1      group A      associate's degree          0                    \
4      1      group C      some college                0                    \
   math score  reading score  writing score  Total score
0      72      72      74      218
1      69      90      88      247
2      90      95      93      278
3      47      57      44      148
4      76      78      75      229
   gender  test preparation course  math score  reading score \
count  1000.000000      1000.000000      1000.000000      1000.000000
mean    0.492000          0.358000      66.089000      69.169000
std     0.499926          0.479432      15.14308      14.460192
min     0.000000          0.000000      0.000000      17.000000
25%     0.000000          0.000000      57.000000      59.000000
50%     0.000000          0.000000      66.000000      69.000000
75%     1.000000          1.000000      77.000000      79.000000
max     1.000000          1.000000      100.000000     100.000000
   writing score  Total score
count  1000.000000      1000.000000
mean    68.054000      203.312000
std     15.195457      42.771978
min     10.000000      77.000000
25%     57.750000      175.000000
50%     69.000000      205.000000
75%     79.000000      233.000000
max     100.000000     300.000000
C:\python-input-22-af0d8c822d6> FutureWarning:
Passing 'palette' without assigning 'hue' is deprecated and will be removed in v0.14.0. Assign the 'y' variable to 'hue' and set 'legend=False' for the same effect.
```

Each visualization technique serves a specific purpose and aids in understanding different aspects of the dataset.

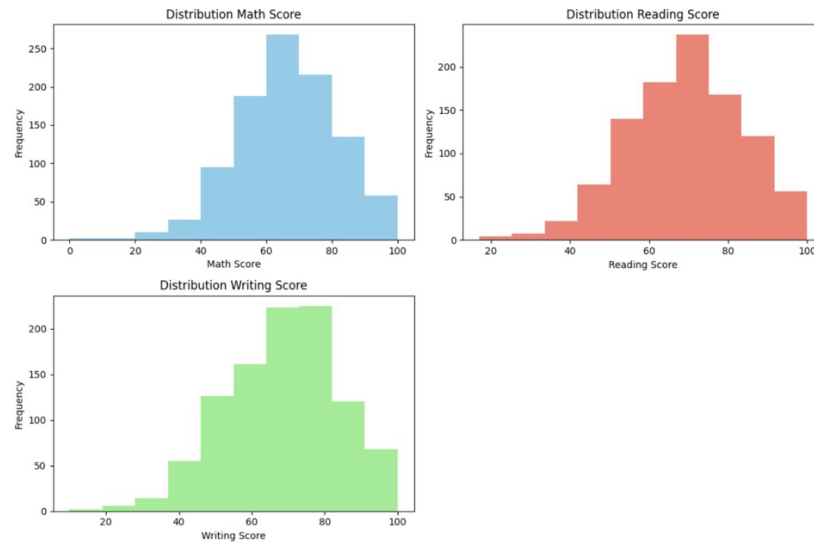
Visualization 1: Count Plot - Plotting the frequency of each category in the "race/ethnicity" feature. It becomes easier to compare the representation of various racial/ethnic groups within the dataset. This could be useful for understanding diversity within the student population and for identifying potential disparities or trends related to race/ethnicity.



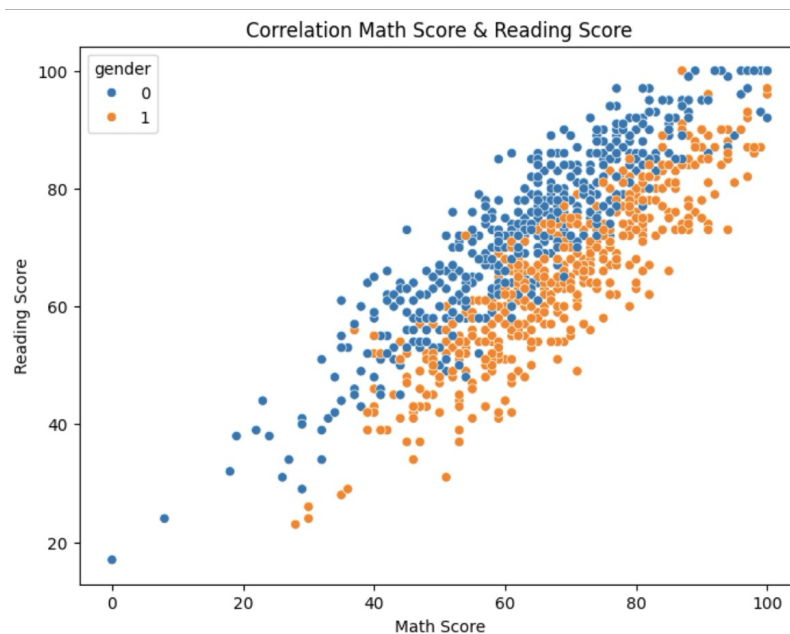
Visualization 2: Boxplot - Distribution Math Score Based on Sex, Boxplots of Reading, and Writing Score vs Gender. With boxplots, it's easier to compare the distribution of math, reading, and writing scores between male and female students. This facilitates the identification of differences in performance between genders and the presence of outliers within each group.



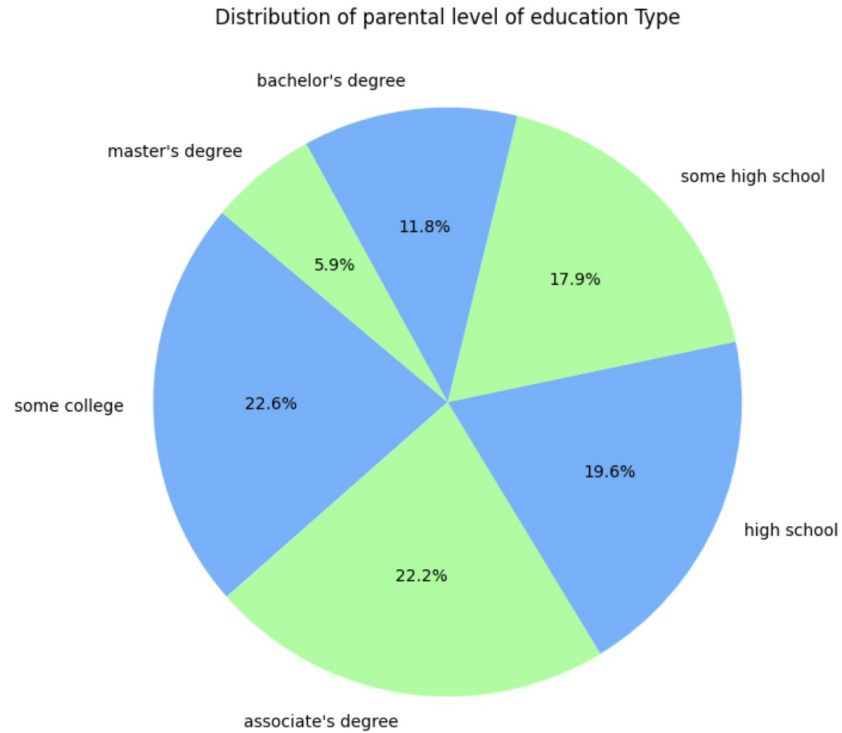
Visualization 3: Histograms - Distribution of Math, Reading, and Writing Scores.



Visualization 4: Scatter plot - Correlation Math Score & Reading Score. By visualizing the correlation between math and reading scores, it becomes easier to assess if there's a linear relationship between these two variables. Additionally, by incorporating gender as a hue, any potential differences in the correlation pattern between genders can be observed.



Visualization 5: Pie Plot - Distribution of parental level of education Type. Pie plots offer a clear visual representation of the distribution of parental levels of education. This makes it easier to grasp the relative prevalence of each category, aiding in understanding the educational backgrounds of the students' parents within the dataset.



Visualization 6: Pair Plot of Math, Reading, Writing and Total Scores with Gender. Pair plots allow for a comprehensive exploration of the relationships between multiple variables simultaneously. With the inclusion of gender as a hue, it becomes easier to observe how the scores in math, reading, and writing are related to each other across genders, along with their individual distributions.

