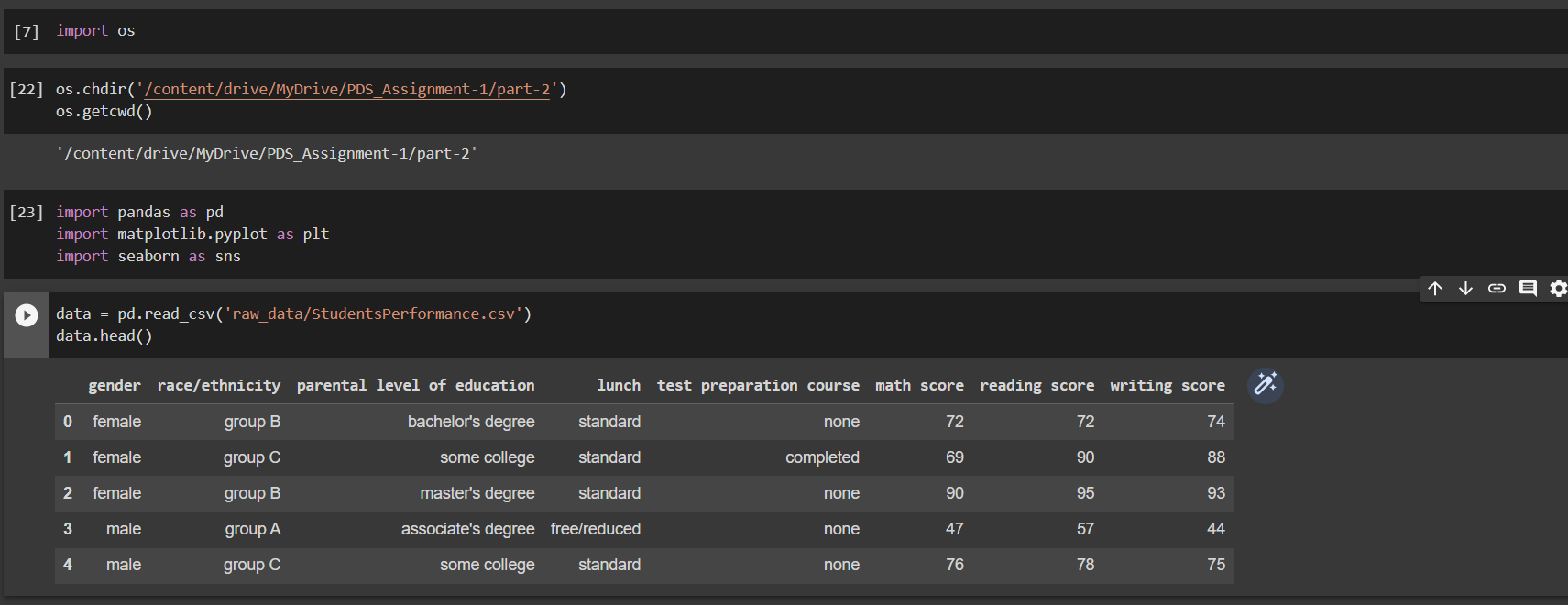
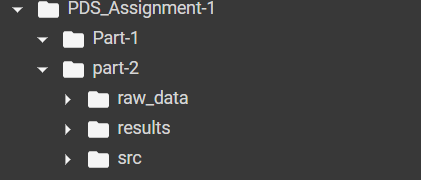
**1.Data input/collection**

****

loaded the StudentsPerformance.csv file in a python environment using ***pandas*** library.

****

**2. Data processing**

**Graphical user interface, text

Description automatically generated**

Since the data doesn’t have any null values and data types are correct didn’t make any refinements to the data.

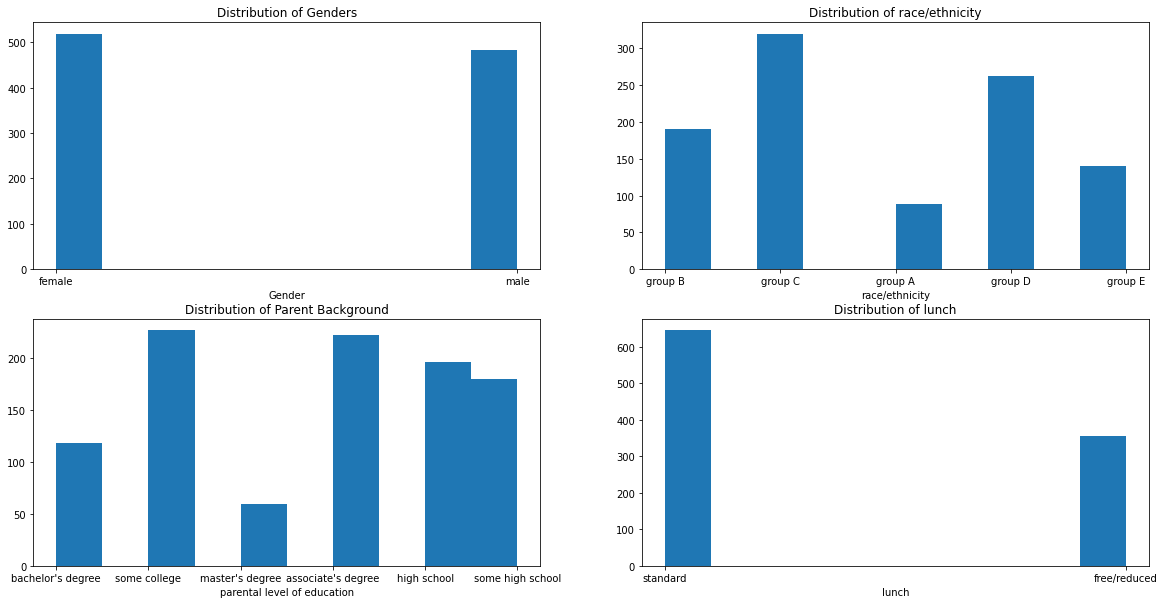
**3. Data analysis**

**a)**

Text

Description automatically generated

This visualization shows the distribution of each of the predictors in the data set in the form of histogram.



b)

Text

Description automatically generated

Chart, histogram

Description automatically generated

c)

Text

Description automatically generated

This bar chart used visualize the number of students who completed the test preparation course vs those who did not.

Logo

Description automatically generated with medium confidence

d)

Text

Description automatically generated

This visualization can help identify whether there is a correlation between math scores and reading scores.

Chart, scatter chart

Description automatically generated

e)

Text

Description automatically generated

This visualization shows box plots of different scores in the data set to give more insight on their distributions.

Chart, box and whisker chart

Description automatically generated

f)

Text

Description automatically generated

This visualization can help identify whether there is a difference in test scores between students whose parents have different levels of education.

Chart, box and whisker chart

Description automatically generated

Text

Description automatically generated

the folder structure looks like this after ***data analysis*** stage.