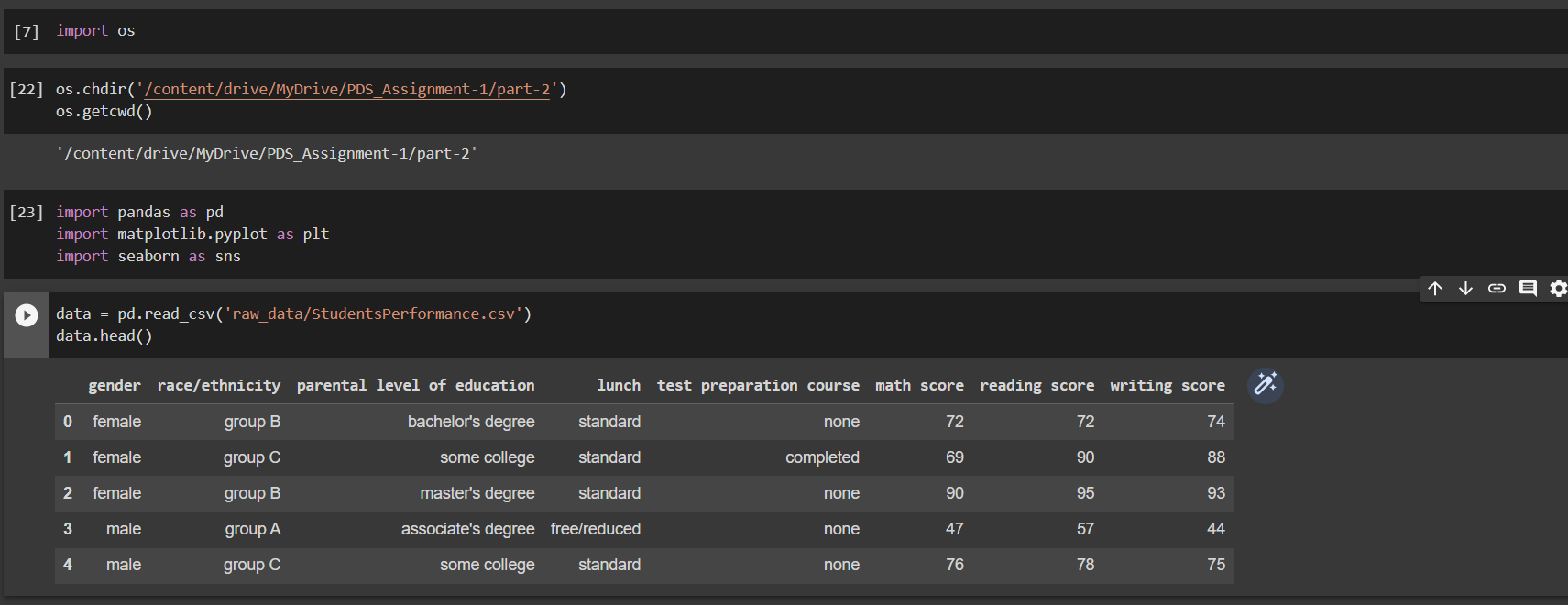
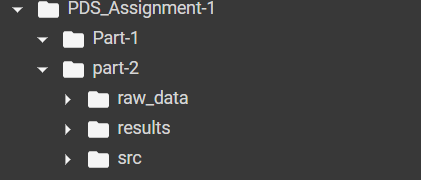
**1.Data input/collection**

****

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)**

Graphical user interface, text, application

Description automatically generated

Chart, scatter chart

Description automatically generated

With this visualization, we can see that there is a positive correlation between math score and reading score. This means that students who perform well in math are likely to perform well in reading, and vice versa.

b)

Text

Description automatically generated

Chart, bar chart

Description automatically generated

With this visualization, we can see that most of the students in the dataset belong to the group 'group C', followed by 'group D'. There are very few students in the 'group A' category.

c)

Text

Description automatically generated

Chart, histogram

Description automatically generated

With this visualization, we can see that the distribution of writing score scores is roughly normal, with a mean around 68.

d)

Chart, scatter chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

With this visualization, we can see that the median math\_score for males is slightly higher than females, but the distributions are largely similar.

e)

Graphical user interface, text, application

Description automatically generated

Chart, box and whisker chart

Description automatically generated

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

Text

Description automatically generated

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