Data Science Infrastructures – Exercise 07

**Assignment I – Getting data into Databricks**

Answer: After uploading the .csv file into the Databricks and creating the table, this is the result that I have got into Microsoft Azure.

A screenshot of a computer

Description automatically generated

Figure 1 Table Created from 'diamonds.csv'

Also, the screen shots for the Data Explorer representing the ‘Columns’ are shown below.

A screenshot of a computer

Description automatically generated

Figure 2 Columns of table created from 'diamonds.csv'

**Assignment II: Working with Data**

Answer: This is the output of the feature ‘cut’ instead of ‘color’ and correponding visualization of ‘diamonds.csv’.

A screenshot of a computer

Description automatically generated

Figure 3 Changed code for visualizing average price of different cuts

A graph of blue rectangular bars

Description automatically generated with medium confidence

Figure 4 Correspoing output of average price of different cuts of diamonds

After having a look at the Spark UI in my cluster, I obtained the following outputs. It clearly shows the timeline and the last 8 jobs that has been executed.

**A screenshot of a computer

Description automatically generated**