מסדי נתונים מטלה 4   
נגיב עבדאללה 316436328  
יהודית בריקנר 328601018

Question 1

We used PySpark in PyCharm using SQL queries to answer the questions  
In the screen shots there is explanation to the code.

CODE:

Text

Description automatically generated

Text

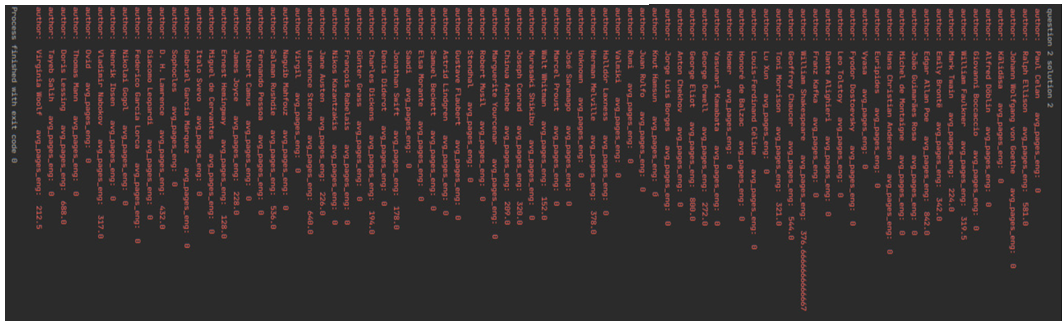
Description automatically generated

Text

Description automatically generatedANSWERS:

Graphical user interface, text, application

Description automatically generated



Question 2

We used the code for linear regression and logistic regression that is in the slides from class. (Pdf 10 slides numbers 25,42)  
we imported the data into a data frame and shuffled the data frame.  
we then divided the data into the data frame we would need for the regression and classification.  
we then change the data frames into NumPy arrays. And split them into train and test by the sizes given to us.

We then trained the models and tested the models. The results are printed below the test.

Graphical user interface

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated with medium confidence