Regression Analysis 1:

Apply regression analysis to the auto-mpg dataset. Independent variables can be selected among cyl, displ, hp, mass and acc.

The dependent varaible is mpg.

Start with a single variable model based on for example scatterplots or correlations. After that add more independent variables as long as you can increase the adjusted coefficient of determination. You can also try to add powers and/or products of variables to you model, if that leads to increase of the coefficient.

Regression Analysis 2:

The datafile has three columns. The first two columns are independent variables (x-variables) and the third column is the dependent variable (y-variable).

Perform regression modelling for the system. Use the 80/20 split for training/testing

Use first one independent variable (x - variable). Add the second variable and accept the two variable model if the adjusted coefficient of determination increases when the second variable is added.

In each step plot the residual plot.

Regression Analysis 3:

The dataset is on advertisements on social network. The columns in a the data correspond to **User Id, Sex, Age** and **EstimatedSalary** and **Purchased.**

The task is to create a logistic regression model with the **Purchased** as the dependent variable.

Use 80%/20% separation for training/testing, The threshold value can be 0.5.

Use **Balanced accuracy** as test metric.

Task 1: Use only **EstimatedSalary** as the **in**dependent variable in your logistic regression model.

Task 2: Add more variables so that **Balanced accuracy** is maximized on test data.