Mini Project for Data Mining and Adv. Statistical Modeling

Find 2 datasets, one to run linear regression on and the other for classification using logistic regression.

Prepare a report to discuss the properties of the datasets, your results and inferences.

You also must prepare a short presentation to discuss and demonstrate your findings.

Project Requirements:

No. of rows >=1000

No. variables > 2

No. of classes for dependent variable must be more than 2 for classification

Do K-fold cross validation for both.

For regression show: R2, Adjusted R2, RMSE, correlation matrix, p-values of independent variables (codes 10)

For classification show: Accuracy, confusion matrix, (Macro recall and precision for multiclass Classification) (codes 10)

Report should discuss the properties of the datasets, your results and inferences/conclusions. (10)

Presentation: Describe dataset, demonstrate execution of code, show results and discuss conclusions. (10)