TP 1

Data set : boston.csv, cambridge.csv

1 Dropping columns that are of no use:

Irrelevant dates, text data, names, repetitive information

2 delete identical rows

3 Separate data types: Numerical, Categorical

Numerical data stored as Categorical (strings)

Binary data needs to be binarized into 1's and 0's

Date/Time features need to be engineered

Nominal data to be unpacked, then later one hot encoded