

Data Wrangling [20 Marks]

You are provided with a dataset that contains the information about the housing prices. Use the given dataset to perform the following tasks:

- Load the dataset into a Pandas dataframe and display the first few rows.
- Discretize the "age" variable into three bins: 'Young', 'Middle-aged', and 'Old'.
- Create a binary variable "is_charles_river" based on the "chas" column.
- Detect and remove outliers for each numerical column in the dataset using the Interquartile Range (IQR) method. **(Don't use any in-built library for this part.)** • Identify and remove noisy data points from the dataset.
- Apply smoothing to the "rm" column and create a new smoothed column.
- Normalize the "tax" and "lstat" columns using Min-Max normalization.
- Perform a simple linear regression to predict the median value of "medv" based on the "rm" variable.
- After the regression analysis, explain if you observe any relationship between "medv" and "rm," providing interpretations based on the regression results.

NOTE: You can find the description of the dataset from the link below for better understanding, <https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html>.