

DATASET

Data Cleaning
Data Processing

Data Frame

Independent variables	y
80% TRAIN SET	
20% TEST SET	

"Test" Data Frame

Independent variables	y
80% TRAIN SET	
20% TEST SET	

Predictions

TEST Data Frame

Independent variables	y_pred	y

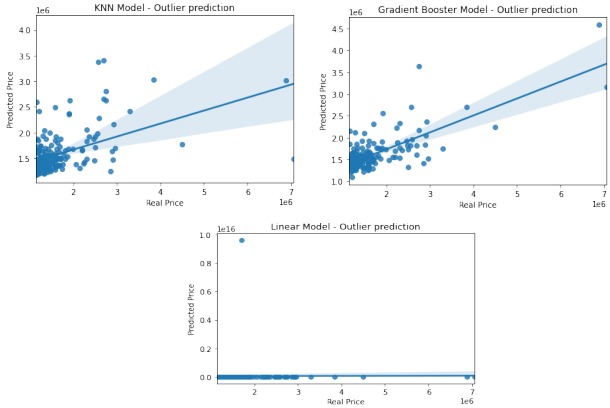
"Outliers" Data Frame

Independent variables	y
80% TRAIN SET	
20% TEST SET	

"Normal" Data Frame

Independent variables	y
80% TRAIN SET	
20% TEST SET	

Predictions



"Outliers" Data Frame with predictions

Independent variables	y	y_pred

+

"Normal" Data Frame with predictions

Independent variables	y_pred	y

TRAIN Data Frame

Independent variables	y_pred	y

MODELS

KNN
LINEAR
GRADIENT BOOSTER

MODEL PERFORMANCE