PREDICTIVE ESTIMATION RISK REPORT

Project: House Price Prediction - Date: 23/06/2022 01:08:51

Objective of analysis: Estimate the value of target variable (**House_Price**) with target value (**30**) with tolerance of minimum (**28.5**%) and maximum (**31.5**%)

Target variable: **House_Price**

Target value: **30**

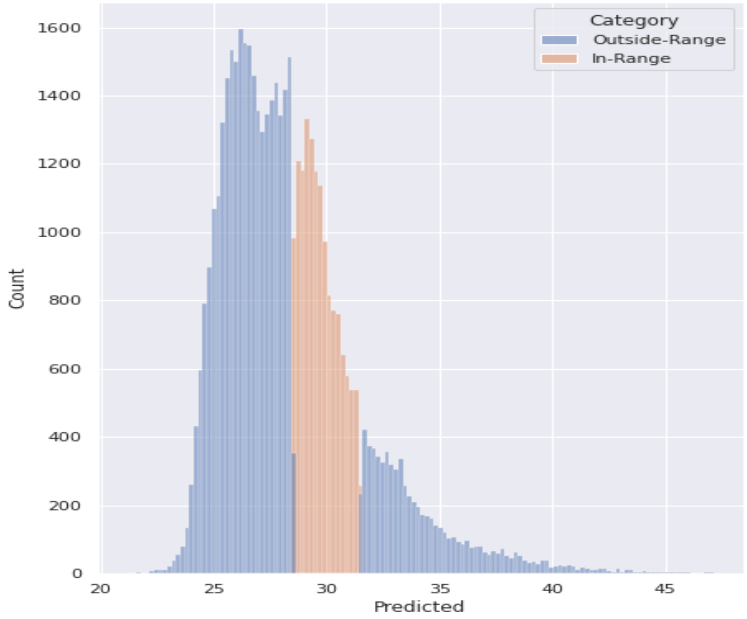
Synthetic data size: **50000** rows

No business rules applied during data creation

Algorithms used in the synthetic data generation pipeline: Gaussian Copula Model, CTGAN Model, Copula GAN Model, TVAE

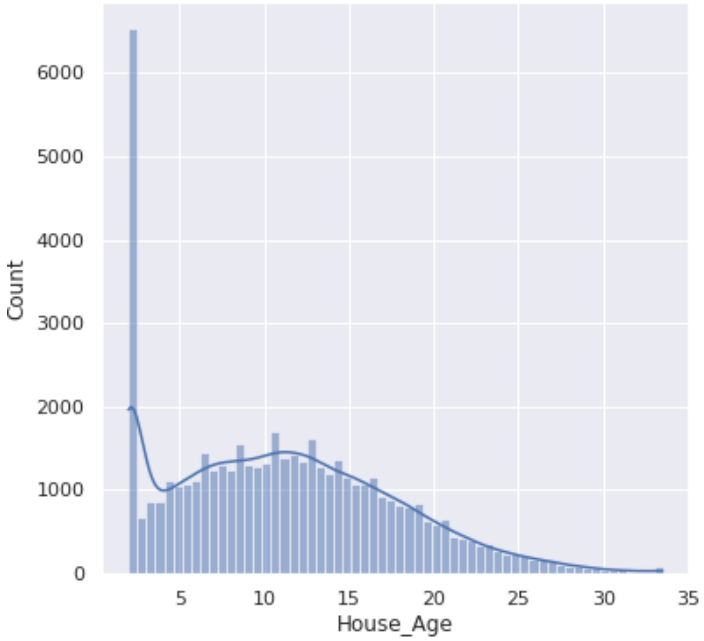
Algorithm which showed best closeness between real and synthetic dataset: **GaussianCopula**

Percentage of closeness, using two-sample Kolmogorov Smirnov test: **80%**



Model type: **Regression**
Total number of predictions made: **50000**
Predictions in line with target range: **23.86**%
Predictions out of range: **76.14**%
The time taken to run the program is **37.59** seconds with a RAM usage of **21846297** bytes.

The distribution of House_Age evaluated for 50000 synthetic data

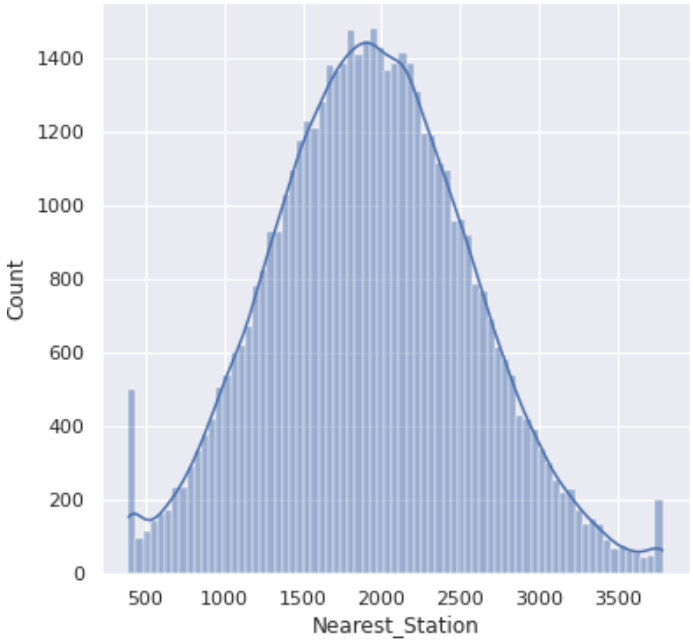


Min: 2.0

Max: 33.5

Average: 11.14

The distribution of Nearest_Station evaluated for 50000 synthetic data

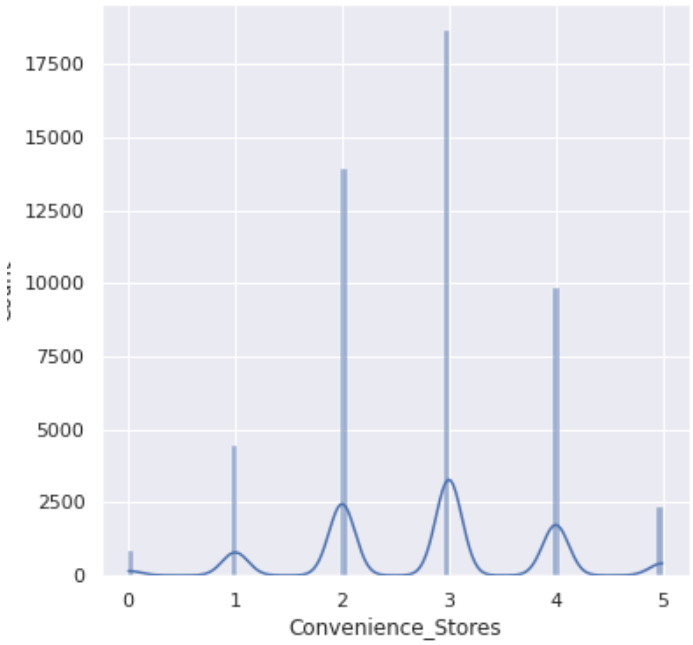


Min: 395.6747

Max: 3780.59

Average: 1933.02

The distribution of Convenience_Stores evaluated for 50000 synthetic data

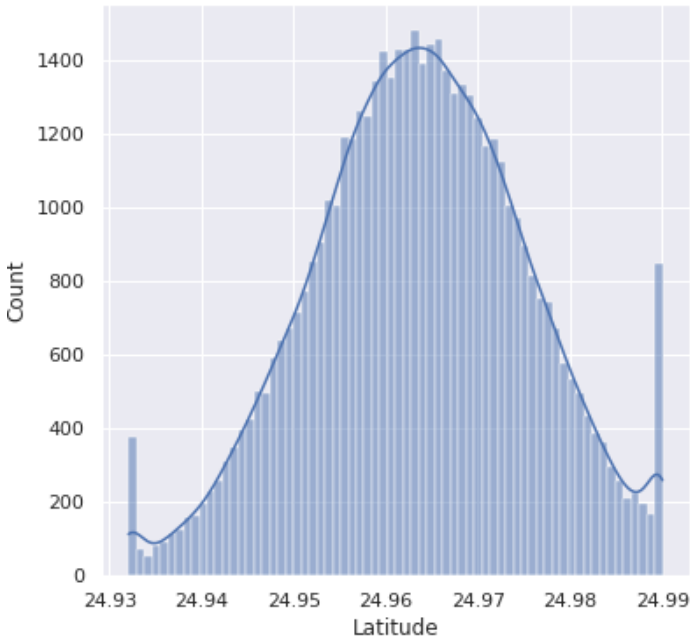


Min: 0

Max: 5

Average: 2.7800000000000002

The distribution of Latitude evaluated for 50000 synthetic data

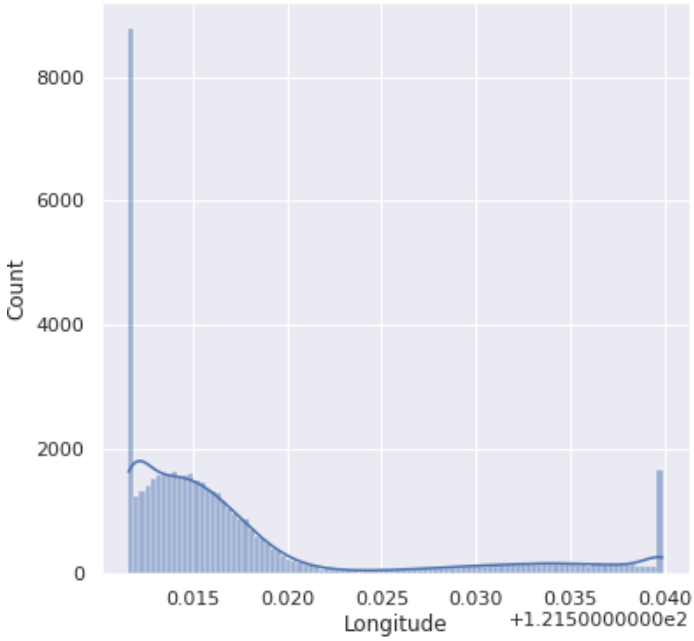


Min: 24.93207

Max: 24.99006

Average: 24.96

The distribution of Longitude evaluated for 50000 synthetic data



Min: 121.51158

Max: 121.5399

Average: 121.52

Correlation among variables using Phi_K for 50000 data synthetic data: