Predicting Transaction Price of Apartments in Tokyo

BA810 GROUP PROJECT TEAM 4

Harry Hu
Husain Ghadiali
Raj Patel
Yiwen Fan

Business Problems

Our aim with this project is to predict the transaction price apartments in Tokyo, Japan, given a number of attributes.

Boston University BA810
Team 4

Features

TRADE PRICE

Response Variable

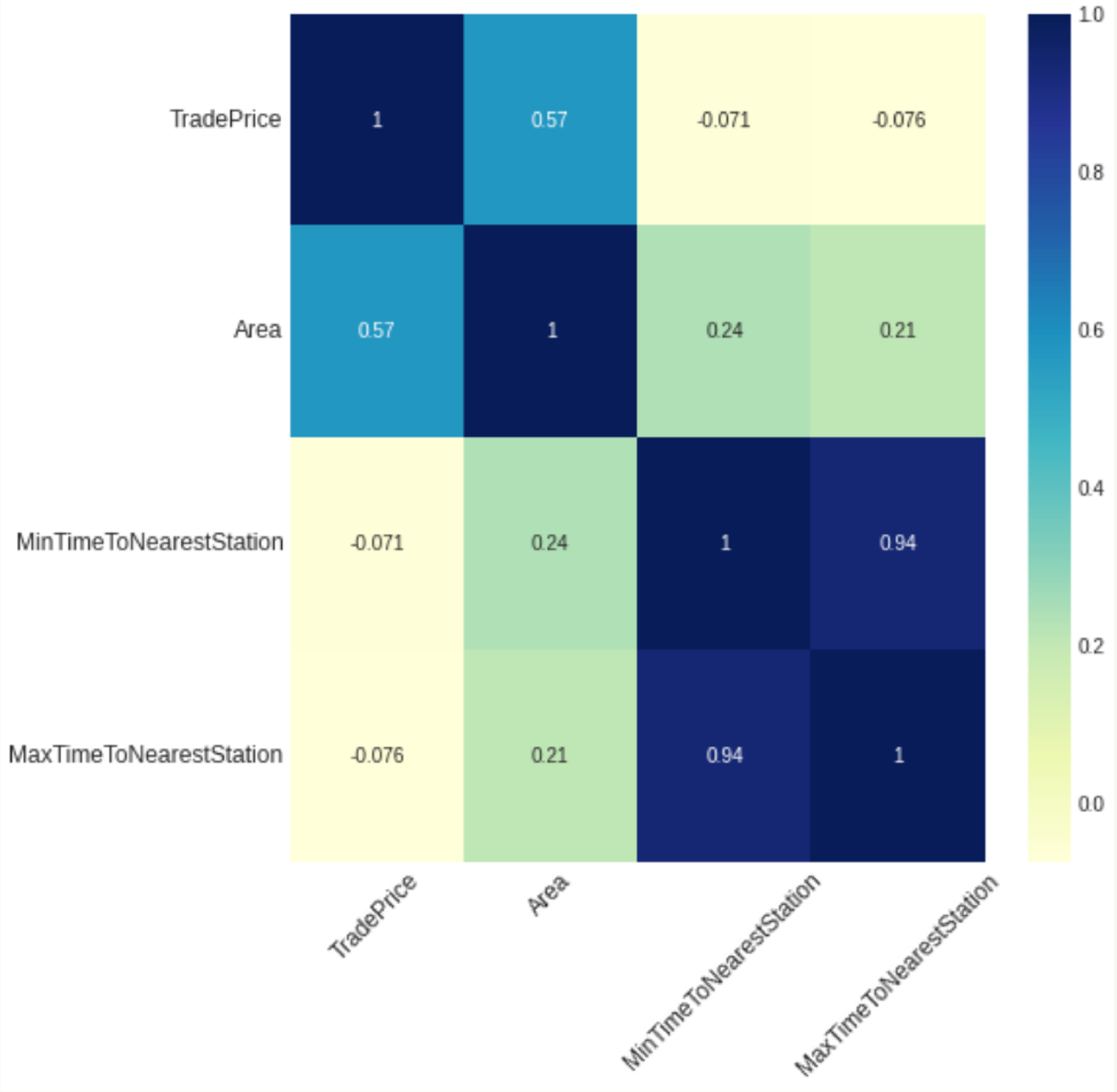
Numeric	Categorical
Area	Structure
BuildingYear	Use
Year	CityPlanning
AvgTimeToNearestStation	Location
RoomCount	
Age	
ConstructionYear	

Data Source:
Japan's Ministry of Land, Infrastructure, Transport, and Tourism (MLIT)

Correlation Analysis

Heatmap

- A relatively strong correlation between *Area* and *TradePrice*
- Collinearity occurs between *MinTimeToNearestStation* and *MaxTimeToNearestStation*



Key Features

01 City Zone

The use districts, where the property is located in, designated by the City Planning Act

02 Apartment Size

The floor area (m^2) of the exclusively owned area registered in a register (the area measured inside walls or other partitions).

03 Property Age

The age of a given apartment. Calculated by subtracting the construction year from the transaction year of the property.

INDUSTRIAL ZONE

37.7M

mean price (in JPY)

COMMERCIAL ZONE

34M

mean price (in JPY)

RESIDENTIAL ZONE

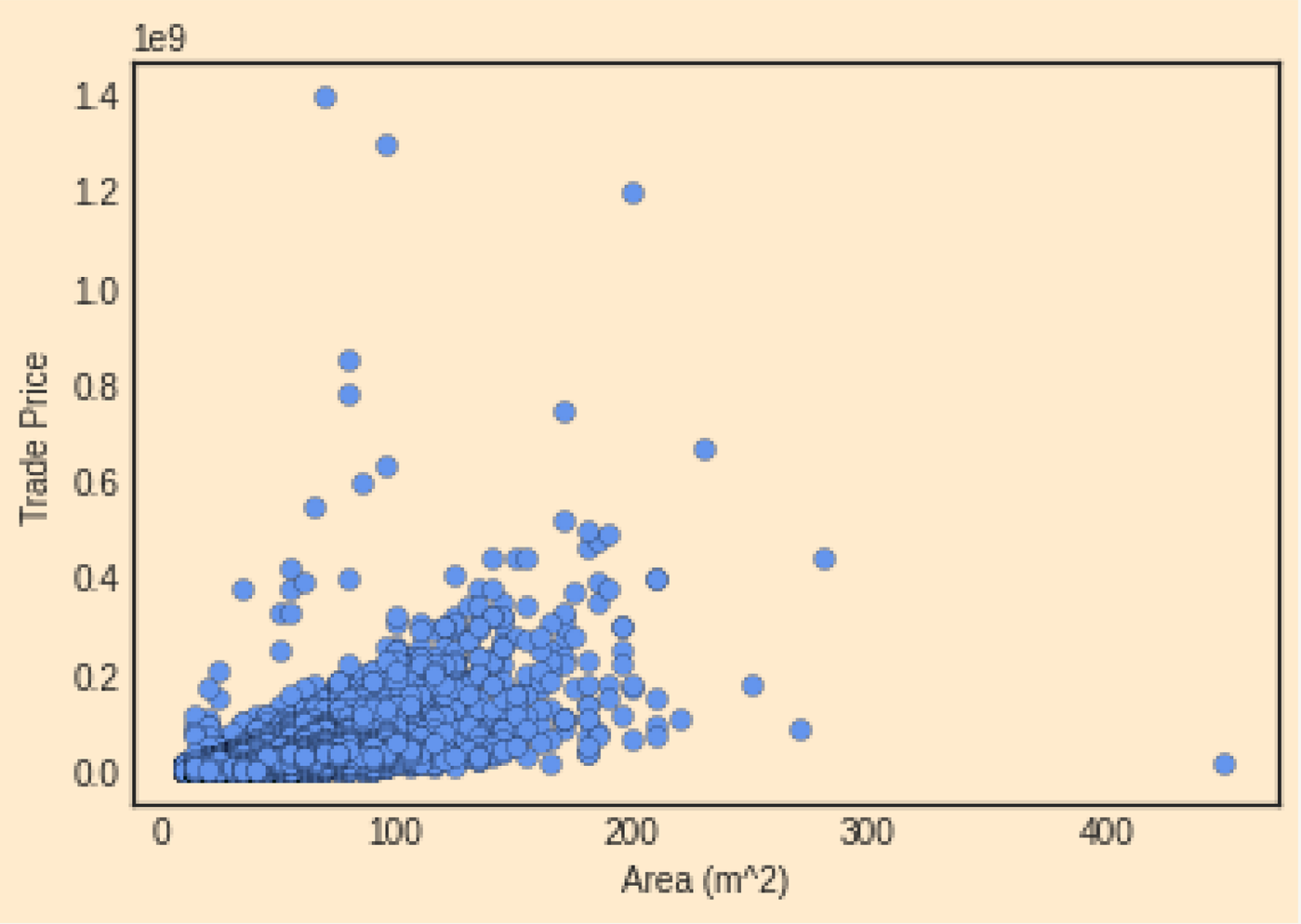
31.7M

mean price (in JPY)

Boston University BA810
Team 4

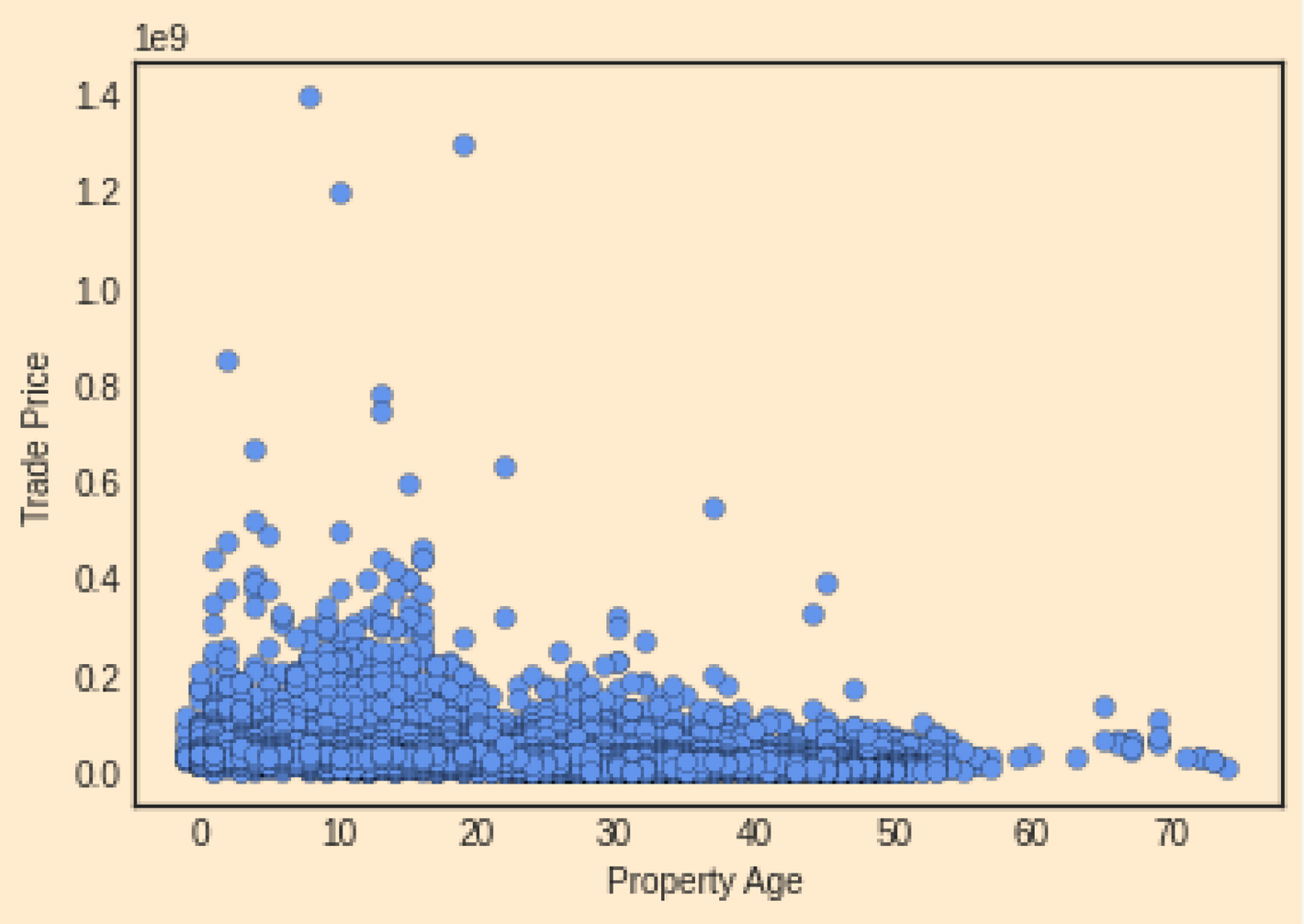
Trade Price vs Area

- The Area of the apartment in m^2
- Area ranges from $10m^2$ to $450m^2$
- Slight positive correlation

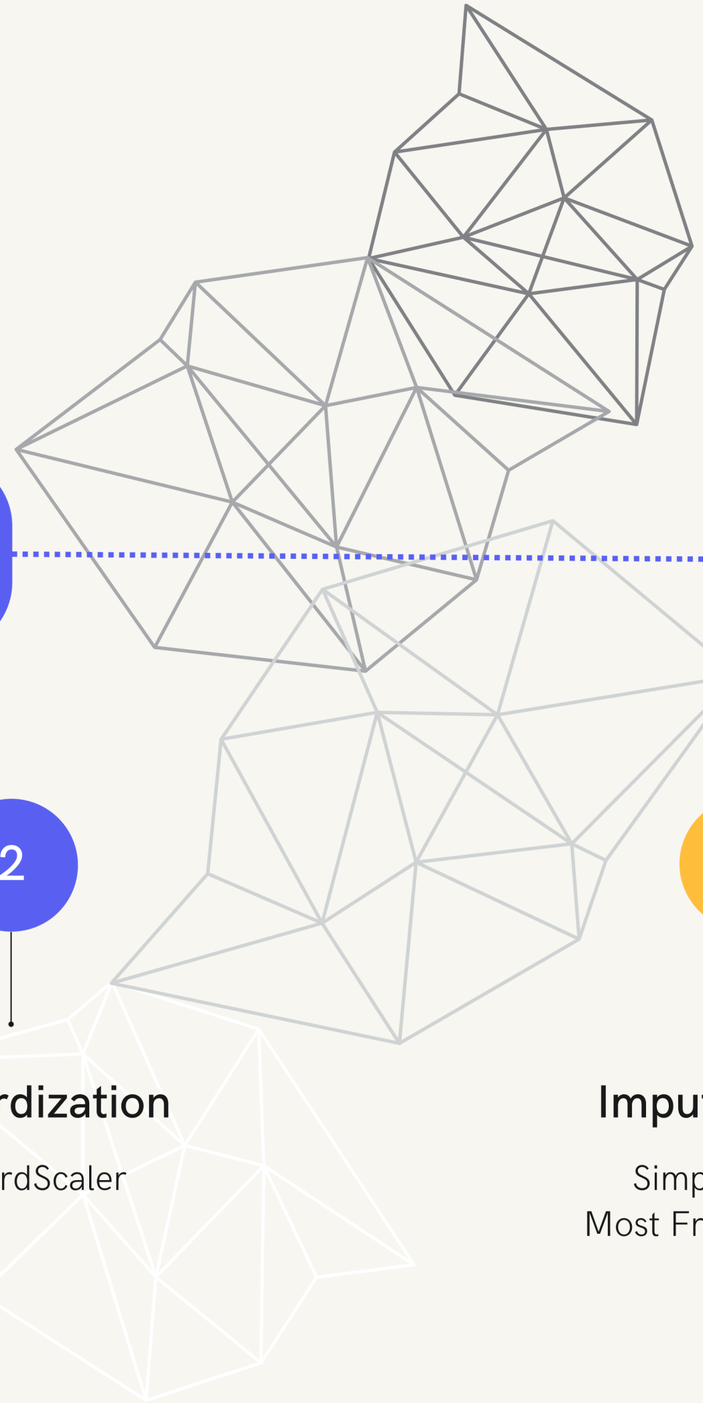


Trade Price vs Age

- Derived column
- Construction year - transaction year
- Outliers
- Possible trend



Pipeline



NUMERIC

1

Imputing Nulls

IterativeImputer

2

Standardization

StandardScaler

CATEGORICAL

1

Imputing Nulls

SimpleImputer
Most Frequent Value

2

Creating Dummy Variables

OneHotEncoder
Drop the First Category

Model Evaluation

01 Linear Regression

Models the relationship between the independent variables and the dependent variable

R-squared Score: 0.5674

02 Lasso Regression

Uses a shrinkage method wherein it shrinks the data values towards a central point as in the mean

R-squared Score: 0.5613

03 Ridge Regression

Analyzes data that may contain multicollinearity using L2 regularization

R-squared Score: 0.5614

04 SGD Regressor

Attempts to find the model parameters that correspond to the best fit between predicted and actual outputs

R-squared Score: 0.5667

05 Linear Support Vector Regression

Allows models to choose how much error is acceptable and will determine a parameter that best fits the model

R-squared Score: -1.6432

06 Decision Tree Regressor

Creates subsets of the dataset by splitting the data into a tree-like pattern in smaller subsets

R-squared Score: 0.4446

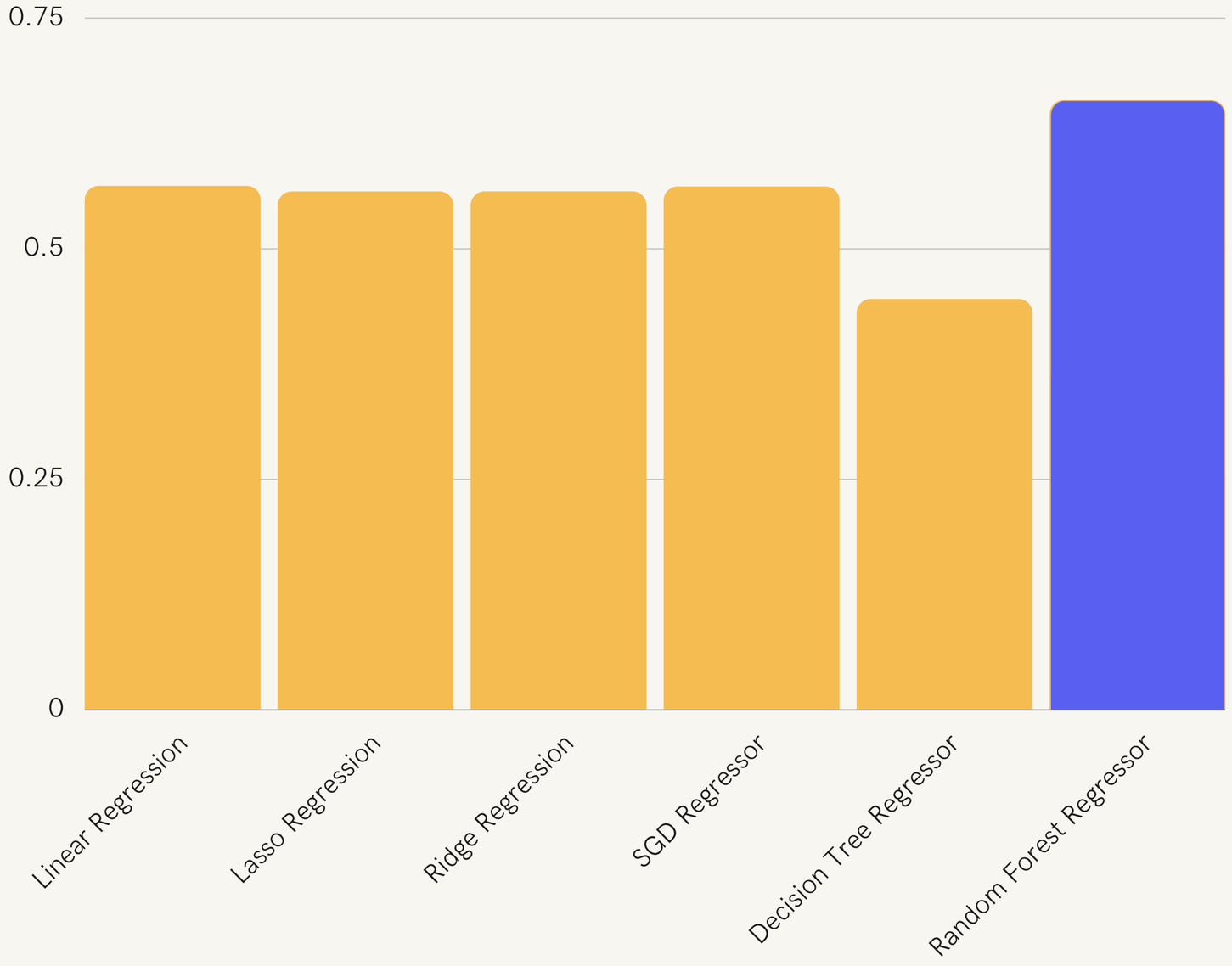
07 Random Forest Regressor

Fits multiple decision trees on several sub-samples and control over-fitting in the models

R-squared Score: 0.6604

R-squared Scores

model scores comparison



Boston University BA810
Team 4

Tuning & Deployment

Hyperparameter Tuning

- Tuning *RandomForestRegressor* using the *HalvingRandomSearch* method
- Selecting *max_depth* as a tuning parameter
- The optimal value of *max_depth* is 11.55, and the *mean_test _score* is 0.6674

Model Deployment

- Evaluating the model performance on the testing data
- *R-squared score* on the testing data 0.5536

Boston University BA810
Team 4

Real Life Implications

for Real Estate Agencies:

- Get a better understanding of the housing market
- Predict transaction price of apartments based on the predictive model

for Government:

- Implement policy-driven city planning for the affordable living zone for residents

Boston University BA810
Team 4

Summary

Boston University BA810
Team 4

1

Business Problem Definition

Our aim with this project is to predict the transaction price apartments in Tokyo, Japan, given a number of attributes

2

Exploratory Data Analysis

3

Pipeline Construction

- Numeric variables: IterativeImputer & StandardScaler
- Categorical variables: SimpleImputer & OneHotEncoder

4

Model Evaluation & Fine Tuning

- Random Forest Regressor has the best R-squared score
- Fine tuning using Halving Random Search

5

Real Life Implications

Analysis such as this can help real estate agencies and local government

Resources

- Original data: https://www.land.mlit.go.jp/webland_english/servlet/MainServlet
- Descriptions of each variable: https://www.land.mlit.go.jp/webland_english/note.html#syurui
- Graph(Cover): https://www.japan-guide.com/e/e3009.html
- Graph(P4): https://www.japan-guide.com/e/e2164.html
- Graph(P9):https://olympics.com/en/olympic-games/tokyo-2020
- Graph(P11):https://www.japanrailpass.com.au/best-hotels-tokyo-2019/

Notebook Link

- https://colab.research.google.com/drive/1HAjPILBDN2X4FG0xKyMfTfoRqATC1x6C?usp=share_link

Thank you!

Any questions?

Boston University BA810
Team 4