

House price*

Yufei Liu

November 24, 2023

1 Introduction

2 Data

```
summary(data)
```

Total_Price	District	Area	Furnished
Min. : 12.0	Length:33291	Min. : 19.53	Length:33291
1st Qu.: 115.0	Class :character	1st Qu.: 76.64	Class :character
Median : 204.0	Mode :character	Median : 92.91	Mode :character
Mean : 279.9		Mean :102.32	
3rd Qu.: 362.0		3rd Qu.:121.19	
Max. :4170.0		Max. :728.25	
Bedroom	Living_Room	Total_Floors	Detailed_Floor
Min. : 0.000	Min. :0.000	Min. : 1.0	Min. : -1.000
1st Qu.: 2.000	1st Qu.:1.000	1st Qu.: 6.0	1st Qu.: 4.000
Median : 3.000	Median :2.000	Median :11.0	Median : 5.000
Mean : 2.871	Mean :1.584	Mean :15.5	Mean : 7.729
3rd Qu.: 3.000	3rd Qu.:2.000	3rd Qu.:25.0	3rd Qu.:10.000
Max. :10.000	Max. :6.000	Max. :60.0	Max. :40.000
Facing_South			
Min. :0.0000			
1st Qu.:0.0000			
Median :1.0000			

*Code and data are available at: <https://github.com/Florence-Liu/house-price>

```
Mean    :0.5385
3rd Qu.:1.0000
Max.    :1.0000
```

```
# Install and load the required packages
library(leaflet)

# Nanjing coordinates
nanjing_coords <- c(32.0603, 118.7969) # Latitude, Longitude

# Create a map centered around Nanjing
nanjing_map <- leaflet() %>%
  setView(lng = nanjing_coords[2], lat = nanjing_coords[1], zoom = 12) %>%
  addTiles() # Add default map tiles

# Print the map
print(nanjing_map)
```

3 Model

```
model_mlg <- lm(Total_Price ~., data = train)
summary(model_mlg)
```

Call:

```
lm(formula = Total_Price ~ ., data = train)
```

Residuals:

Min	1Q	Median	3Q	Max
-817.88	-69.02	-4.87	59.85	1768.08

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-375.5725	6.8574	-54.769	< 2e-16 ***
DistrictGulou	419.3107	6.1132	68.591	< 2e-16 ***
DistrictJiangning	190.4130	6.0952	31.240	< 2e-16 ***
DistrictJianye	446.8370	6.1049	73.193	< 2e-16 ***
DistrictJurong	9.2239	6.2234	1.482	0.13831
DistrictLishui	45.3028	6.0129	7.534	5.06e-14 ***

DistrictLiuhe	135.5136	6.0554	22.379	< 2e-16	***
DistrictPukou	175.1478	6.0438	28.980	< 2e-16	***
DistrictQinhuai	351.4587	6.1342	57.295	< 2e-16	***
DistrictQixia	272.9941	6.0623	45.031	< 2e-16	***
DistrictXuanwu	357.2146	6.1256	58.315	< 2e-16	***
DistrictYuhuatai	290.2115	6.0782	47.746	< 2e-16	***
Area	4.0656	0.0280	145.186	< 2e-16	***
FurnishedNot Furnished	-16.7826	3.0296	-5.539	3.06e-08	***
FurnishedOther	-5.7113	1.8493	-3.088	0.00201	**
FurnishedPart Furnished	-30.5391	2.4264	-12.586	< 2e-16	***
Bedroom	-12.9521	1.3747	-9.422	< 2e-16	***
Living_Room	16.9012	1.7497	9.659	< 2e-16	***
Total_Floors	0.9660	0.1198	8.064	7.69e-16	***
Detailed_Floor	0.3920	0.2011	1.949	0.05126	.
Facing_South	-0.7014	1.5821	-0.443	0.65752	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 123.7 on 26611 degrees of freedom

Multiple R-squared: 0.7507, Adjusted R-squared: 0.7505

F-statistic: 4006 on 20 and 26611 DF, p-value: < 2.2e-16

```
predict_mlg <- predict(model_mlg, newdata = test)
mean((test$Total_Price-predict_mlg)^2)
```

[1] 14899.58

```
model_rf <- randomForest(Total_Price ~., data=train)
model_rf
```

Call:

```
randomForest(formula = Total_Price ~ ., data = train)
```

Type of random forest: regression

Number of trees: 500

No. of variables tried at each split: 2

Mean of squared residuals: 10521.1

% Var explained: 82.86

```
predict_rf <- predict(model_rf, newdata = test)
mean((test$Total_Price-predict_rf)^2)
```

```
[1] 9259.356
```

4 Result

5 Discussion

6 Reference