Improving Rental Price in the United States

BUAN6357
Professor: Sourav Chatterjee, PhD
Student: Chun-Li Hou

Introduction

Background, Question, Problem, Need

- The client is the **management in rental industry**.
- More of Americans than the other countries people are willing to rent for living but not buy for living. Statistically speaking, it is more than **65%** of Americans under 35 years old who decide to rent for living.
- The management of rental industry is curious to know if there is anyway to **maximize profits**.
- Our goal for this project is to analyze the data about website crawling from rental websites from the year 2019 to produce recommendation to help improving the pricing of various properties.
- We will firstly use different machine learning regression models to decide the best algorithm. Then, we will establish causality by interpreting the estimate coefficients.

Data

Source, Dataset, Collection, Dimension, Feature Selection

- Source: machines learning repository at UCI
- Dataset: apartment for rent classified dataset (Fredrick Nilsson)
- Collection: crawling from rental websites
- Dimension: **9,958** rows and **15** columns
- Feature Selection:

```
Rows: 9,958
Columns: 15
$ price          <int> 645, 660, 875, 995,
$ square_feet    <int> 430, 535, 800, 875,
$ bedrooms       <int> 1, 1, 2, 2, 2, 2, 2,
$ bathrooms      <int> 1, 1, 1, 1, 2, 2, 2,
$ clubhouse      <fct> 0, 0, 0, 0, 0, 0, 0,
$ fireplace      <fct> 0, 0, 0, 0, 0, 0, 0,
$ balcony        <fct> 0, 0, 0, 0, 0, 0, 0,
$ sport          <fct> 0, 0, 0, 0, 0, 0, 0,
$ gym            <fct> 0, 0, 0, 0, 0, 0, 0,
$ pool           <fct> 0, 0, 0, 0, 0, 0, 0,
$ parking        <fct> 0, 0, 0, 0, 0, 0, 0,
$ laundry        <fct> 0, 0, 0, 0, 0, 0, 0,
$ elevator       <fct> 0, 0, 0, 0, 0, 0, 0,
$ pets_allowed   <fct> 1, 1, 0, 1, 0, 0, 1,
$ metro_area     <fct> 0, 0, 0, 1, 0, 0, 0,
```

Machine Learning Algorithm

Regression model

- PNLR: Polynomial Non Linear Regression
- MLR: Multiple Linear Regression
- RF: Random Forest Regression
- KSVR: Kernel Support Vector Regression
- SVR: Support Vector Regression
- DT: Decision Tree Regression
- ANN: Artificial Neural Network Regression

	RMSE <dbl>
PNLR	705
MLR	705
RF	706
KSVR	717
SVR	742
DT	784
ANN	973

Log-Linear Regression Model

Regression model

- **log(price)** = **6.6388** + **0.0004*square_feet** + 0.0021*bedrooms + **0.1152*bathrooms** - 0.0368*fireplace - **0.1351*sport** + **0.1167*gym** - **0.0437*parking** + **0.2231*elevator** + 0.0301*pets_allowed + **0.2304*metro_area** - **0.0001*square_feet:pets_allowed** + **0.0001*square_feet:metro_area**
- Multiple R-squared: 0.332; Adjusted R-squared: 0.331
- Bedrooms, Fireplace, Pets_allowed are not statistically different from zero.
- Besides that, other coefficients are statistically from zero.
- All assumptions are satisfied.

Variable <chr>	estimate <dbl>	p-value <dbl>
(Intercept)	6.6388	0.0000
square_feet	0.0004	0.0000
bedrooms	0.0021	0.7673
bathrooms	0.1152	0.0000
fireplace1	-0.0368	0.0115
sport1	-0.1351	0.0000
gym1	0.1167	0.0000
parking1	-0.0437	0.0000
elevator1	0.2231	0.0000
pets_allowed1	0.0301	0.1042
metro_area1	0.2304	0.0000
square_feet:pets_allowed1	-0.0001	0.0001
square_feet:metro_area1	0.0001	0.0000

Business Question 1

As considering exercise facility, which exercise facility does increase the price more?

- Having the **outdoor** sport facility, such as basketball field or tennis court, the price **decreases** by **13.51**%.
- Having the **indoor** sport facility, such as gym or fitness center, the price **increases** by **11.67**%.
- Thus, indoor exercise facility is more profitable.

Business Question 2

What apartment attribute affected the price is the most interesting?

- The top highest coefficient is metro_area, which means whether the apartment locat ion in a metro area or not. This is not interesting due to a common knowledge.
- The second highest coefficient is elevator, which means whether the apartment have elevator or not. This is interesting due to some insights.
- As for the apartment having the **elevator**, the price **increases** by **22.31**%.
- Although living in which floors has its pros and cons, having an elevator still significantly positive affect the price.

Business Question 3

How much should the new releasing apartment set the price for some requirements?

- Example apartment:
  - 808 square feet apartment in Richardson TX
  - 1 bedroom, 1 bathroom
  - With gym, parking, and pet allowed
  - $1340 (actual)
  - https://www.sublet.com/property/2284864#

- log(price) = 6.6388 + 0.0004*808 + 0.0021*1 + 0.1152*1 - 0.0368*0 - 0.1351*0 + 0.1167*1 - 0.0437*1 + 0.2231*0 + 0.0301*1 + 0.2304*0 - 0.0001*808*1 + 0.0001*808*0

- Price = **$1225** (**predict**)

Thank You