

Beijing City Feature-Based airbnb Price Prediction

Team GeekHub

Brian Feng

Mike Liu

Tony Li

Logan Luo

Outline

Motivation

Problem Definition

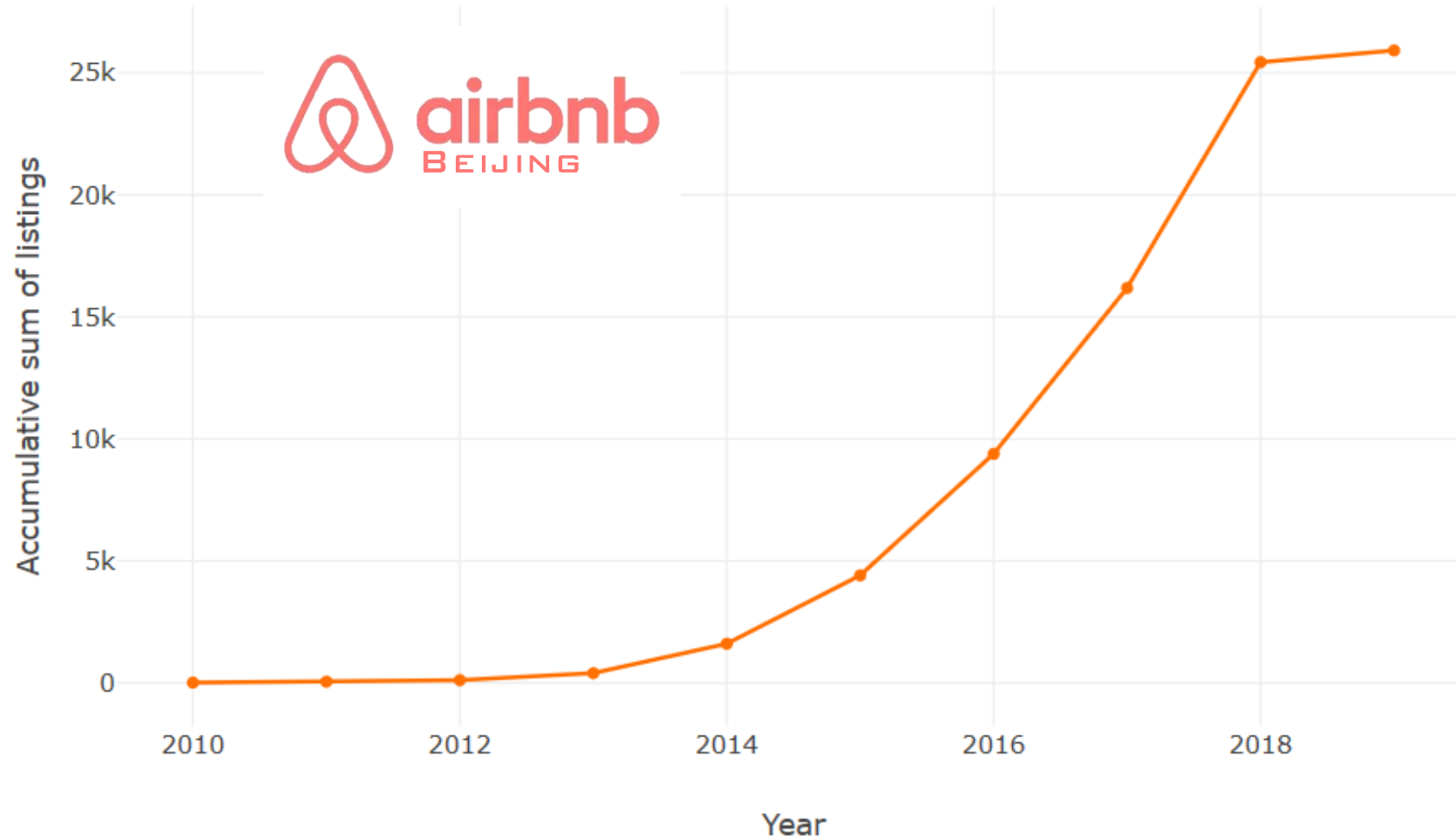
State-of-the-Art

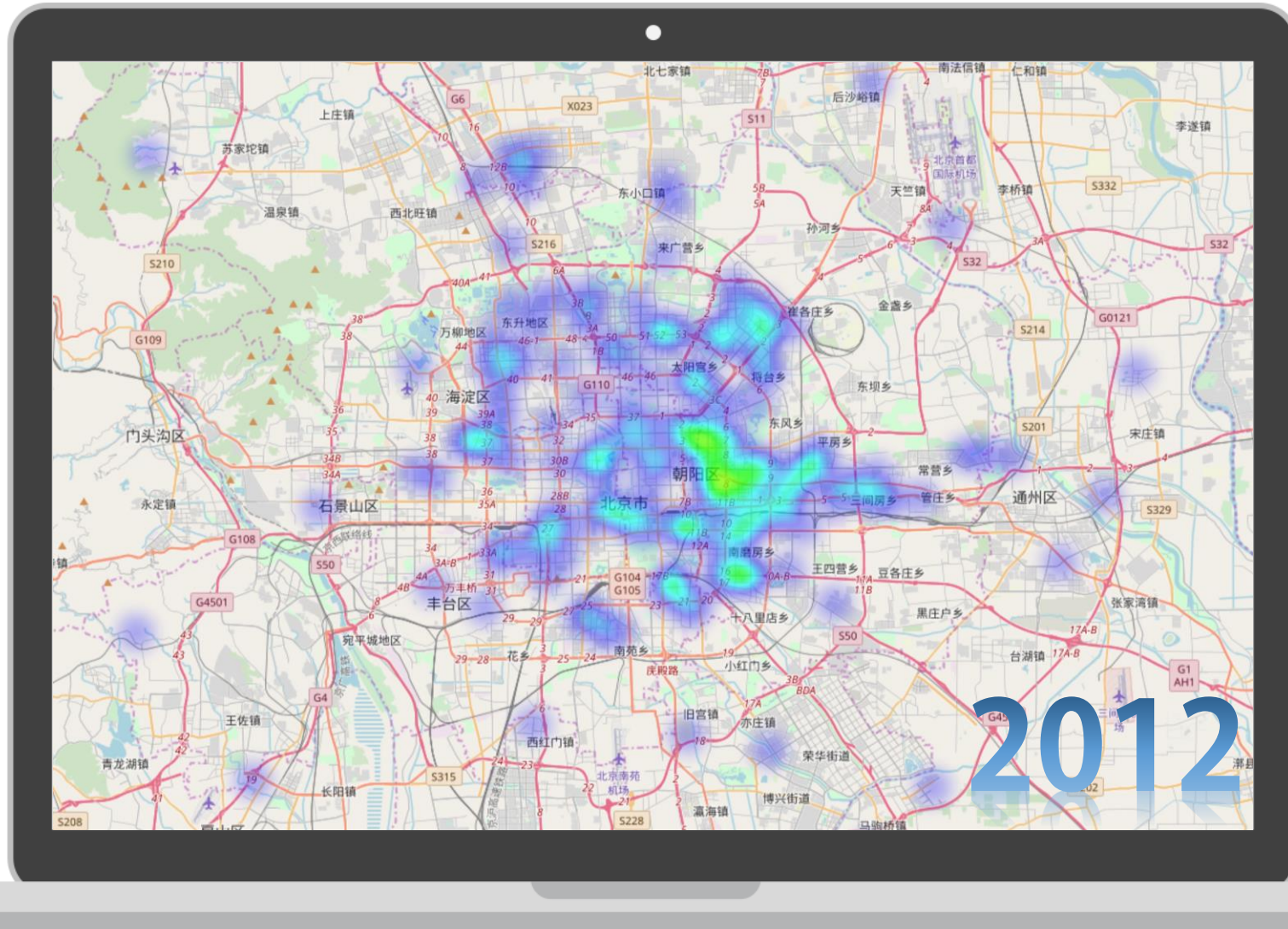
Methodology

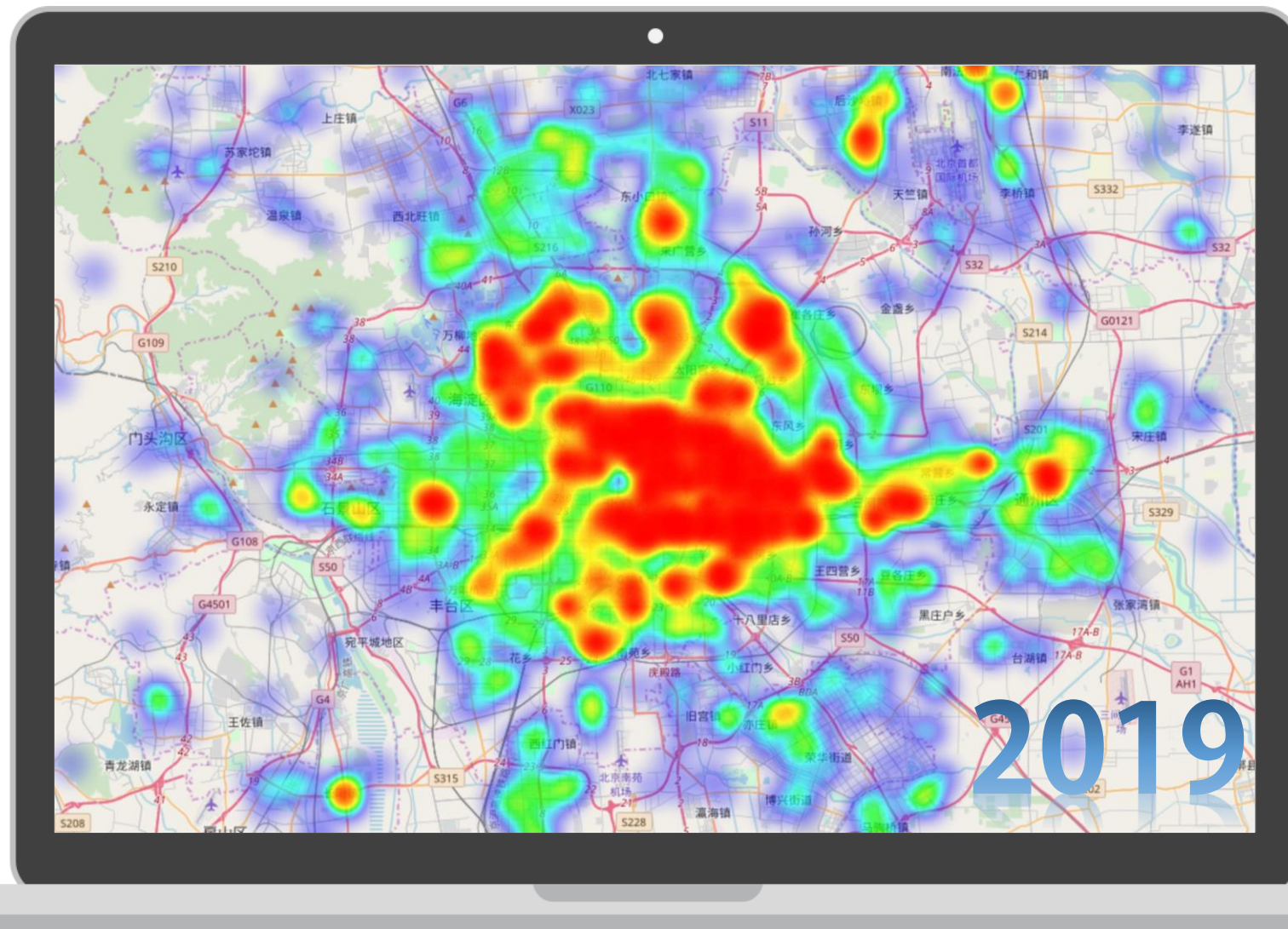
Evaluation

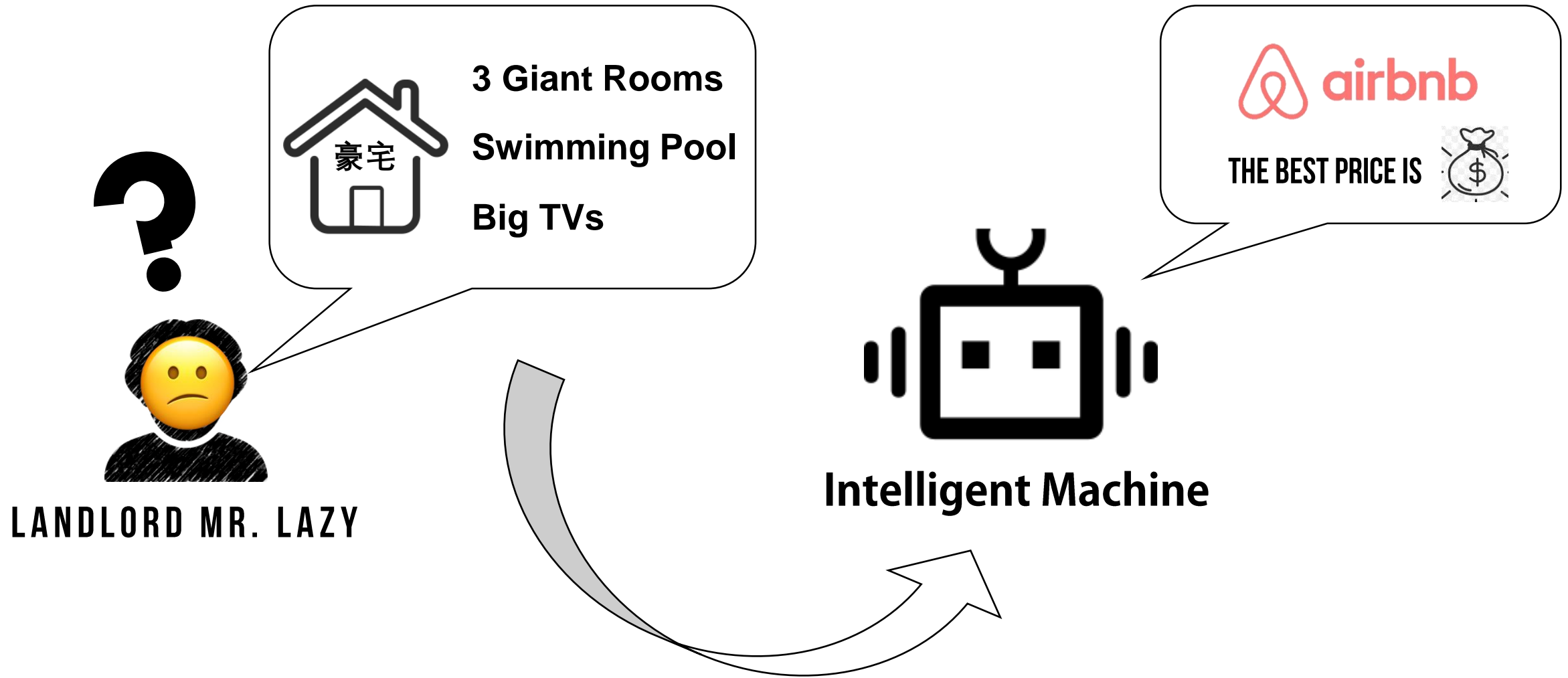
Motivation

Accumulative number of listings year by year since 2010 in Airbnb Beijing







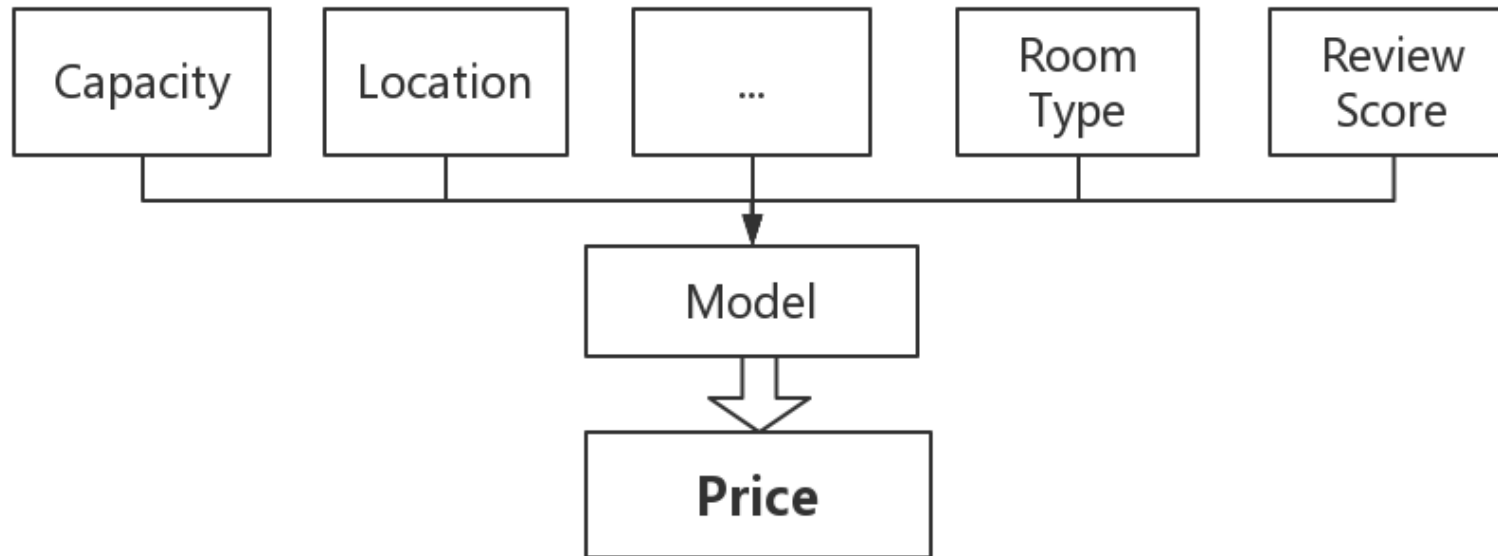


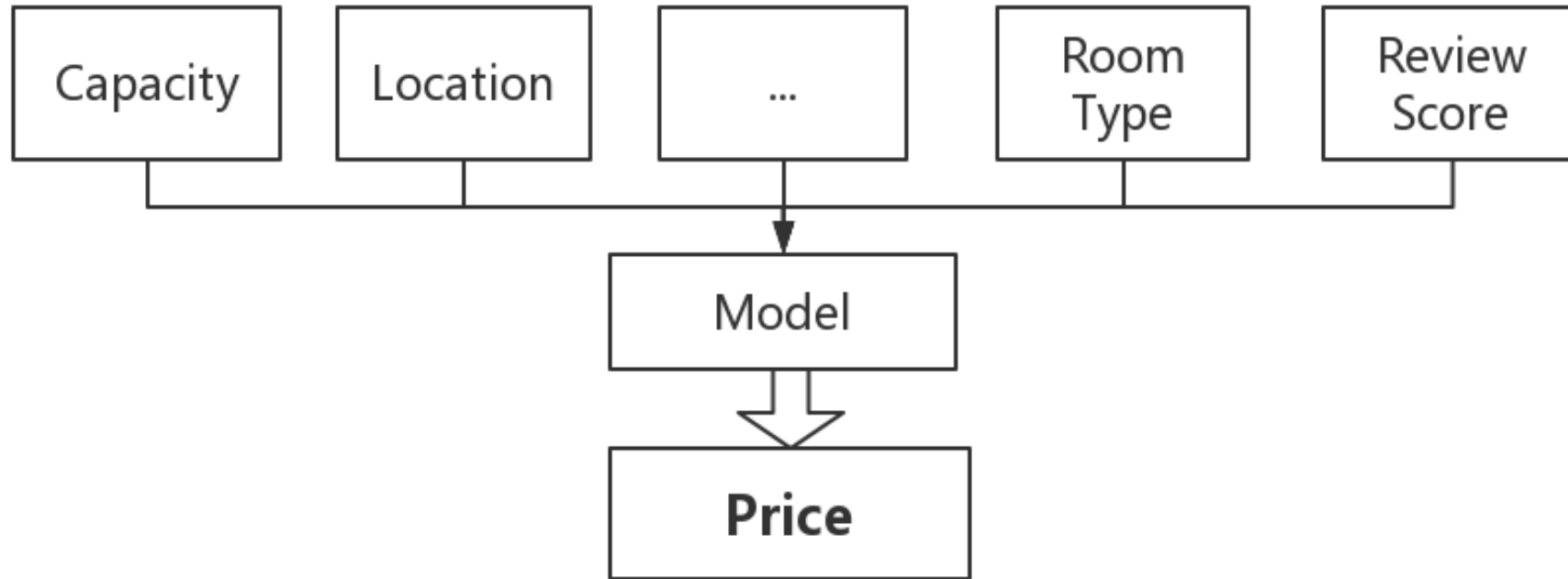
Problem Definition

Problem Statement

Given a series of data describes the property's **features**.

Output the reasonable/best **price** point for the host.



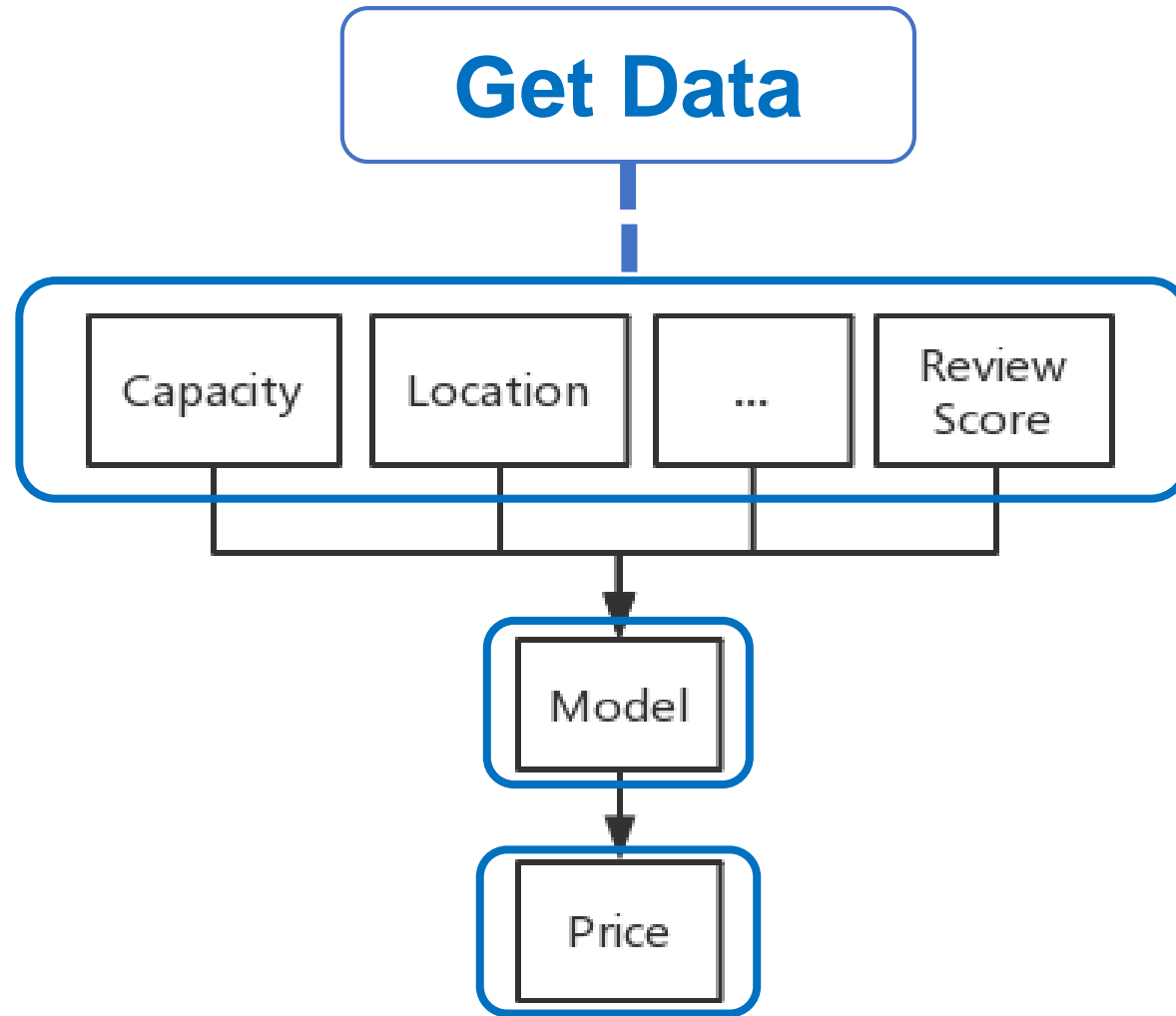


$$f(\textit{Capacity}, \textit{Location}, \dots, \textit{RoomType}, \textit{ReviewScore}) + \beta_0 \Rightarrow \textit{Price}$$

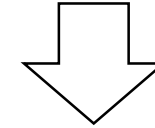
State-of-the-Art

A brief review of the literature on price influencing factors

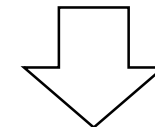
| Dimension | Factors | Effects | Author |
|-----------|---|----------|--|
| External | Price in cities | N/A | Stephen and Andrew(2016) Balaguer and Pernias(2013), Becerra(2013) Bull(1994), Wang & Nicolau(2017), Li (2016), Huang(2010) Wang & Nicolau(2017), Zhang(2017) |
| | Number of owners Distance between owners | Pos/Neg | |
| Location | Distance | Negative | |
| House | Room type | Negative | |
| | Room type | Positive | |
| Room | Bedroom, bathroom, bed | Positive | |
| Rule | Instant bookable | Negative | |
| | Cancellation policy | Positive | |
| Sociality | Number of reviews per year | Negative | |
| | | Positive | |
| | Host identity verified | Pos/Neg | |
| | | Positive | |



**Feature
Engineering**



**Tuning
Models**



Outputs

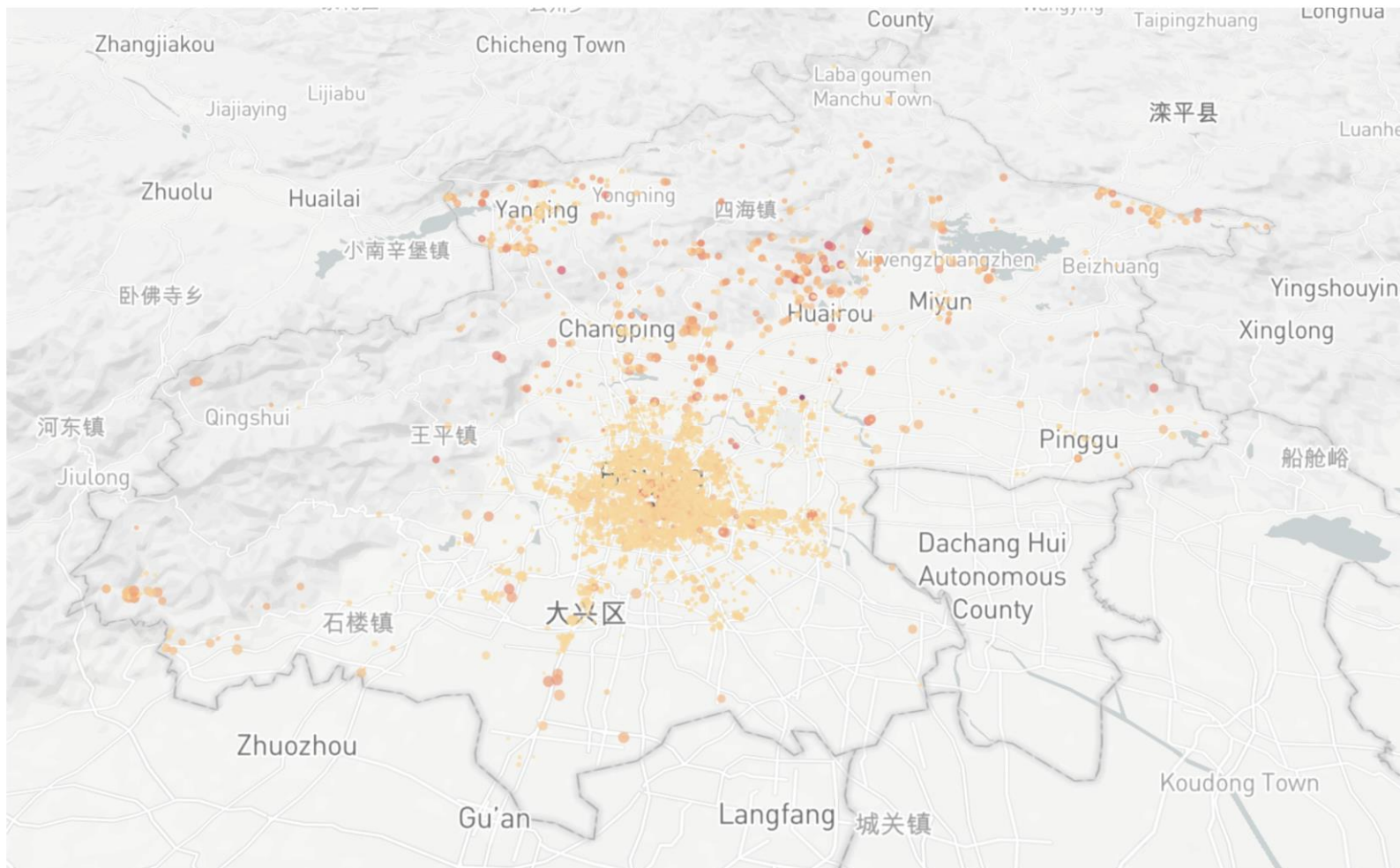
Methodology

Get Data



Raw_Data-Airbnb-Beijing.csv

**Total:
19360 Listings**



All available Data in Airbnb

Data Pre-Processing

- Drop Properties that **lacks too many features** (less than 100)
- Drop Properties with **unreasonable prices** (i.e. the host didn't actually want to Airbnb it)

~~[Price < (¥ 50 * accommodates) OR (¥ 50 * beds)]~~

~~[Price > ¥ 20,000 OR Price = ¥ 9999]~~

~~[Beds=50 OR Bathrooms=101 OR min_nights > 3]~~

- Fill up a few fixable nulls.

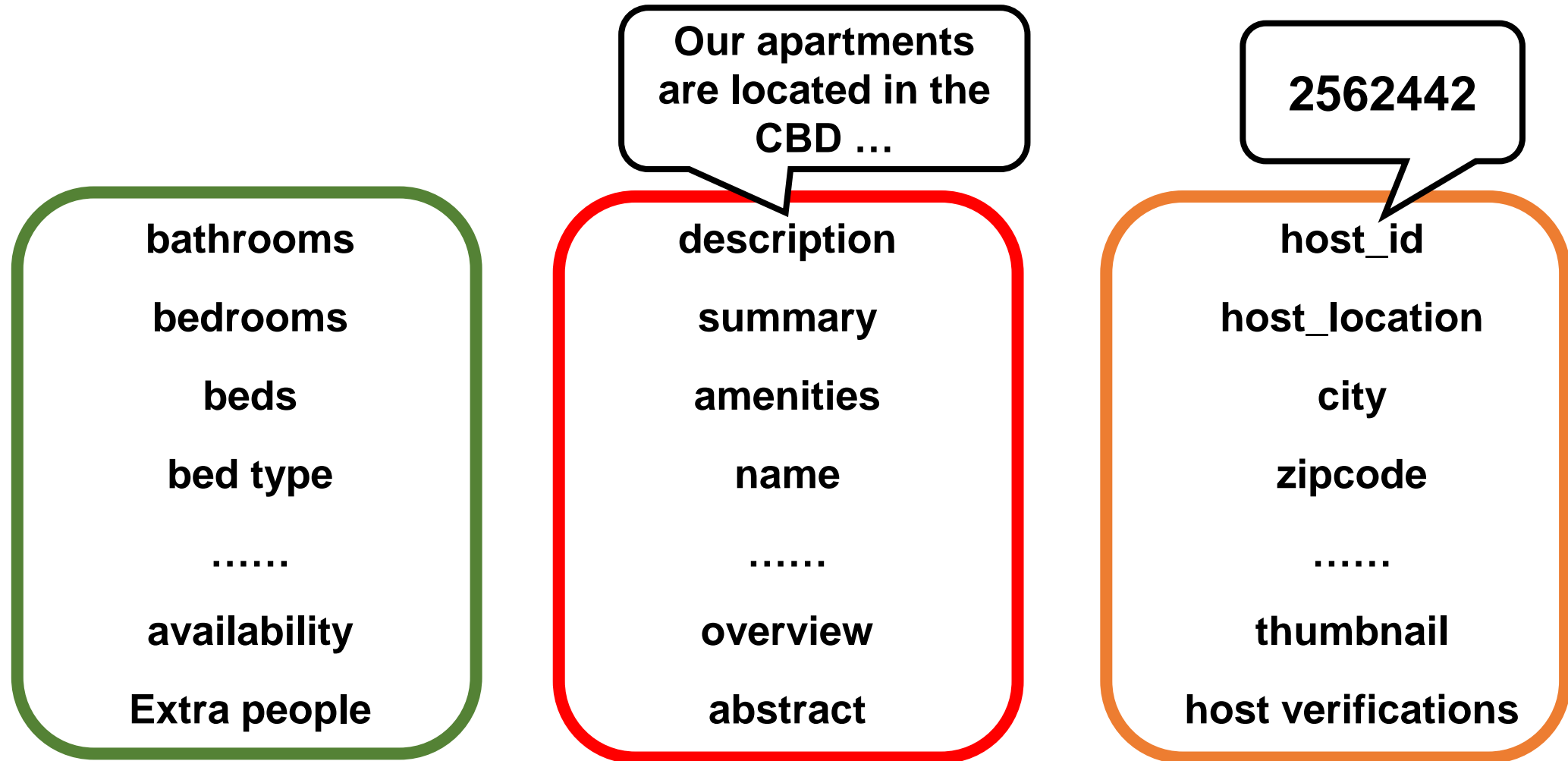


Airbnb-listings-Beijing.csv

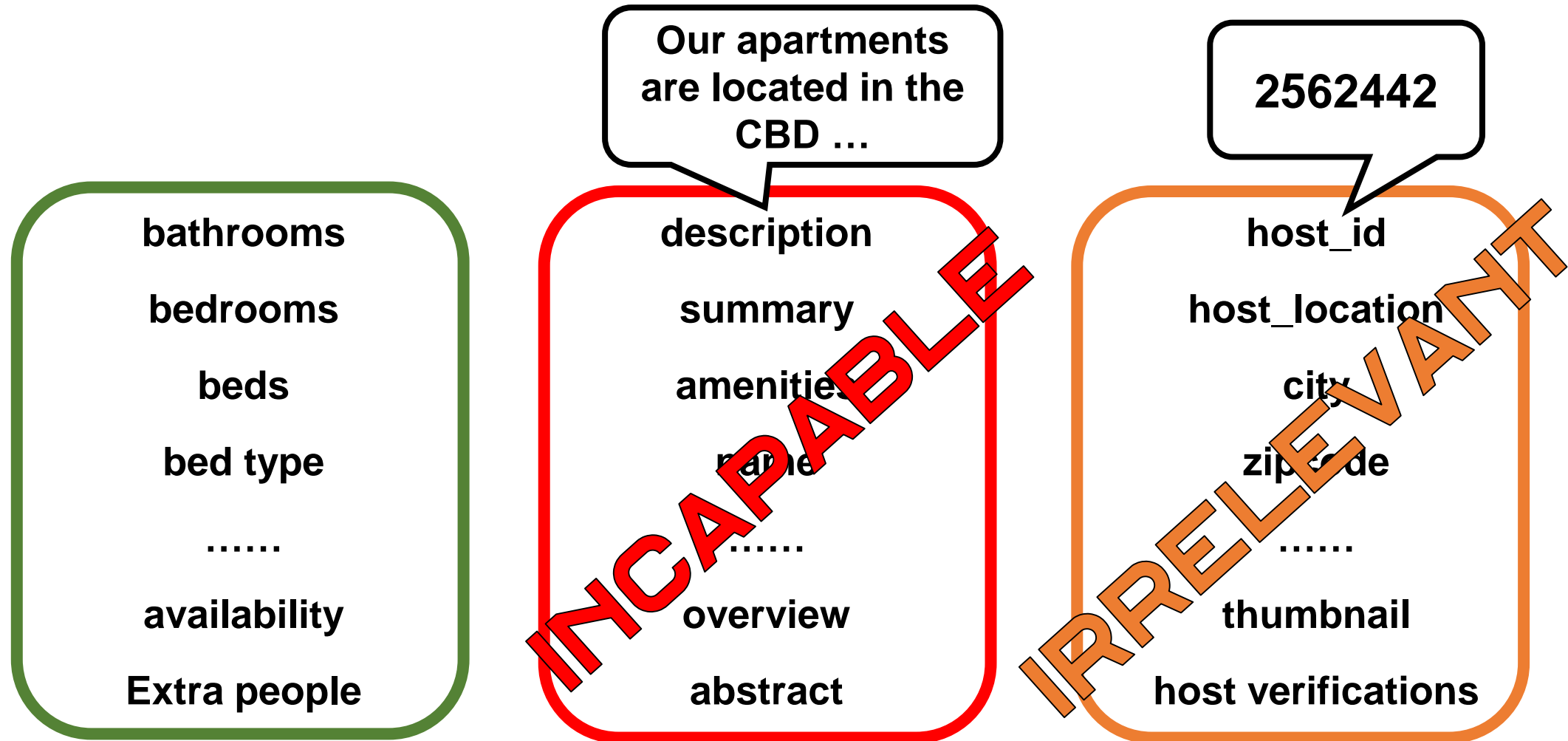
| ID | Capacity | Room types | Price | ... |
|--------|----------|--------------|-------|-----|
| 44054 | 9 | Entire home | 816 | ... |
| 100213 | 2 | Entire home | 1203 | ... |
| 128496 | 3 | Private room | 401 | ... |
| 161902 | 2 | Private room | 387 | ... |
| 162144 | 4 | Private room | 553 | ... |
| 279078 | 2 | Entire home | 401 | ... |
| 282825 | 4 | Entire home | 657 | ... |
| 287026 | 3 | Entire home | 415 | ... |
| 287511 | 3 | Entire home | 415 | ... |
| 317195 | 2 | Private room | 546 | ... |
| 322292 | 2 | Entire home | 436 | ... |
| ... | ... | ... | ... | ... |

19360 Listings**17224 Listings**

Feature Engineering



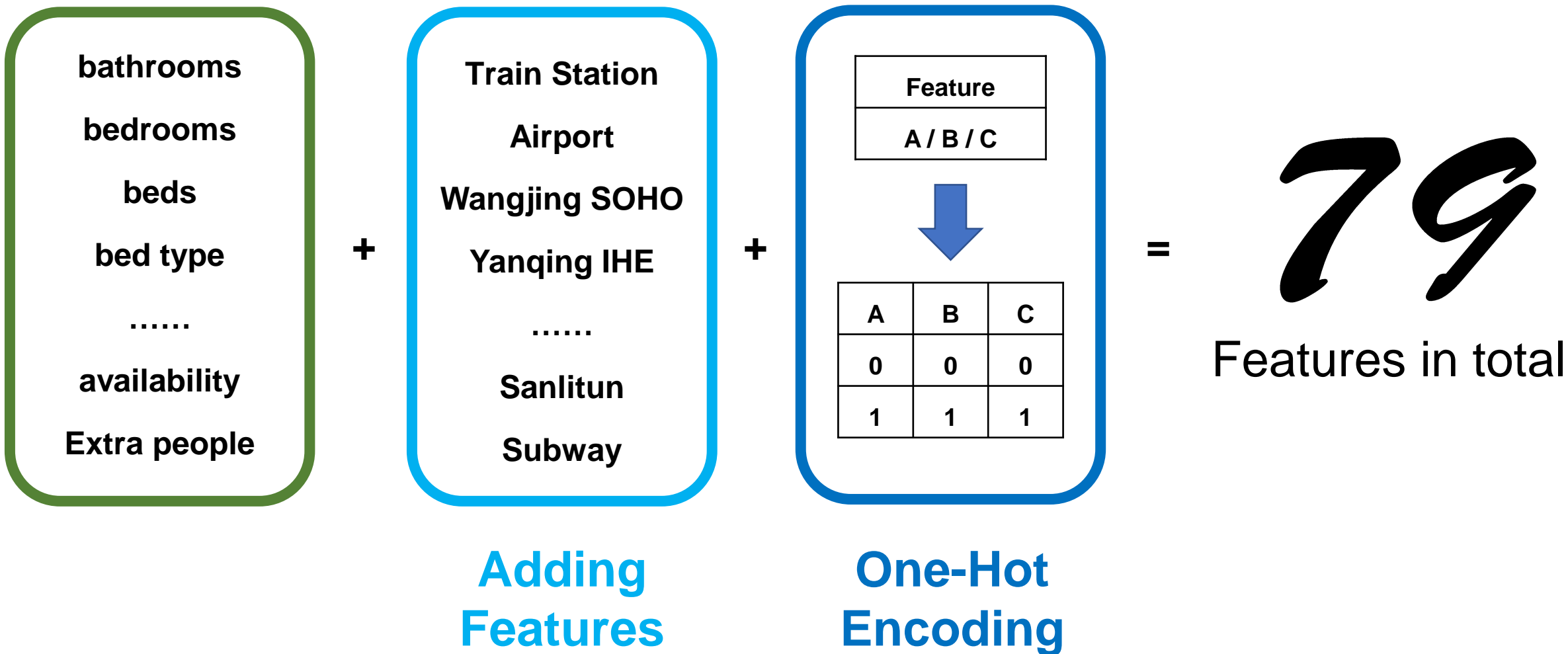
Feature Engineering

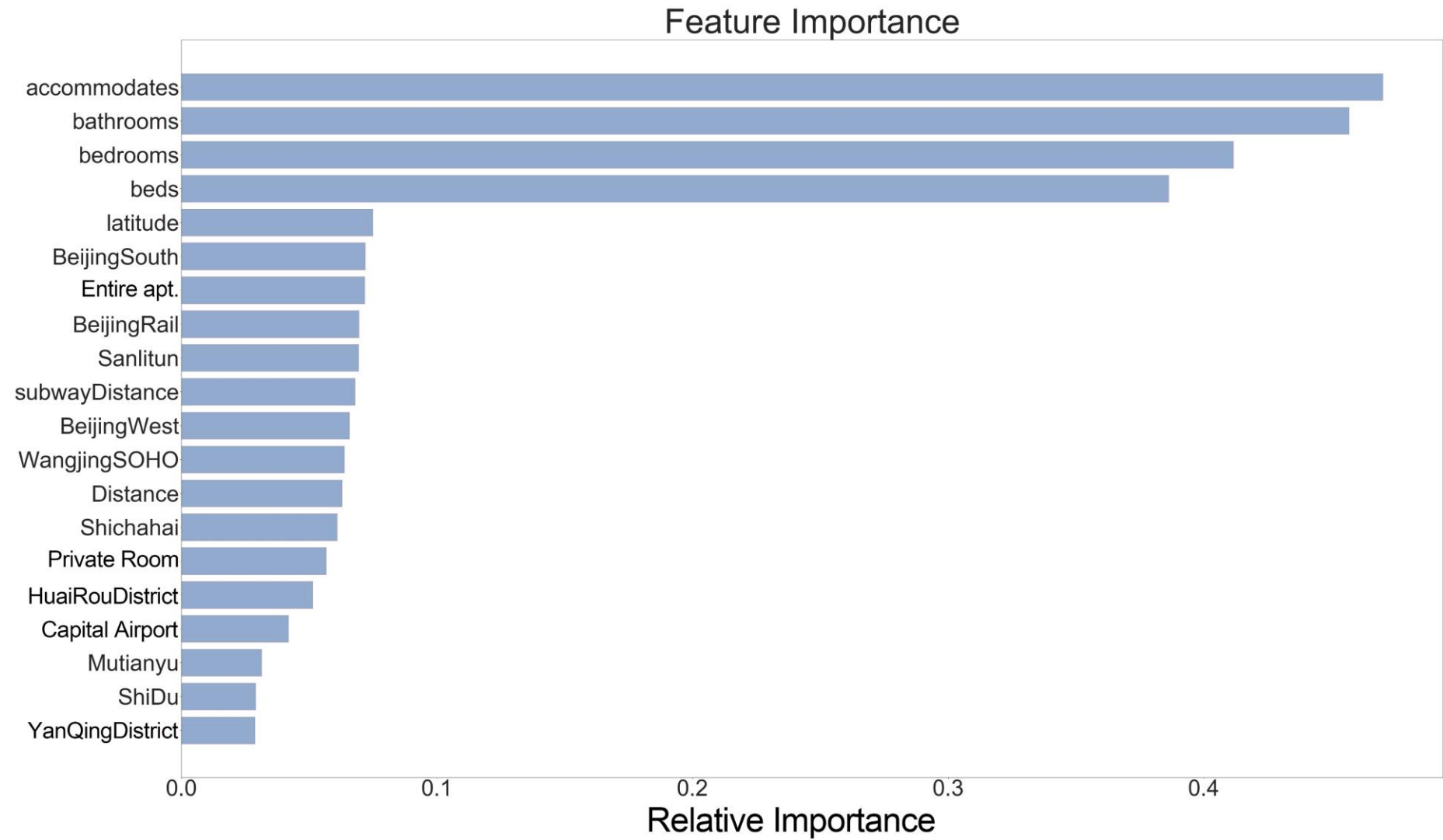


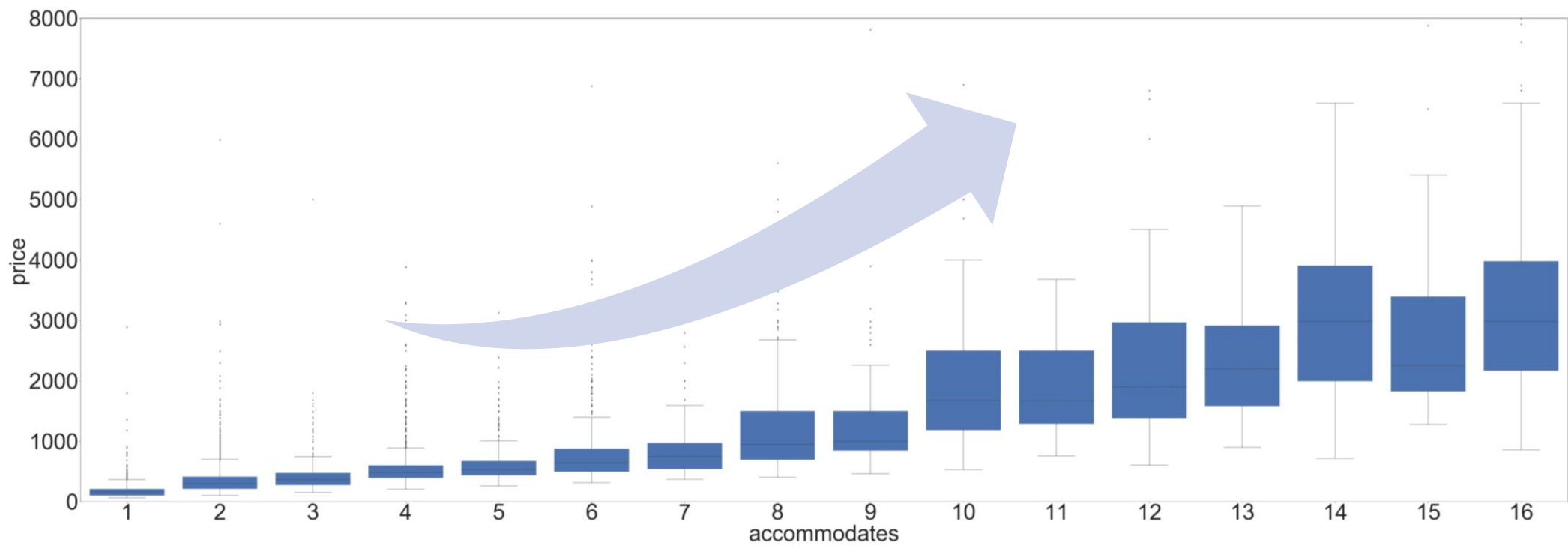
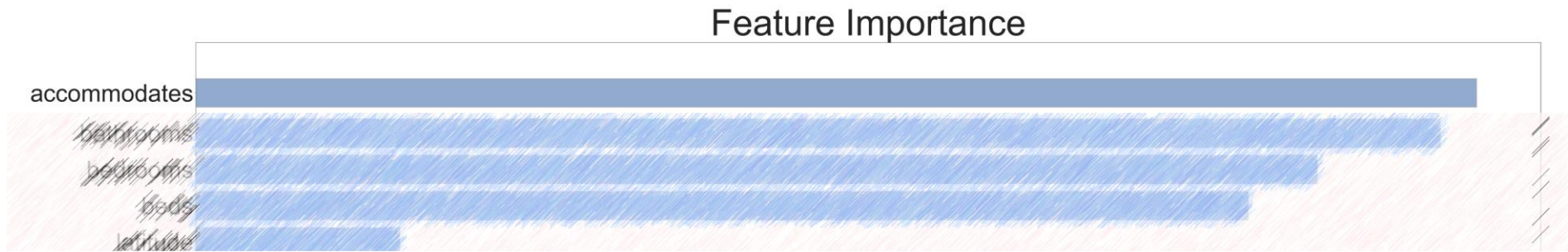
Feature Engineering

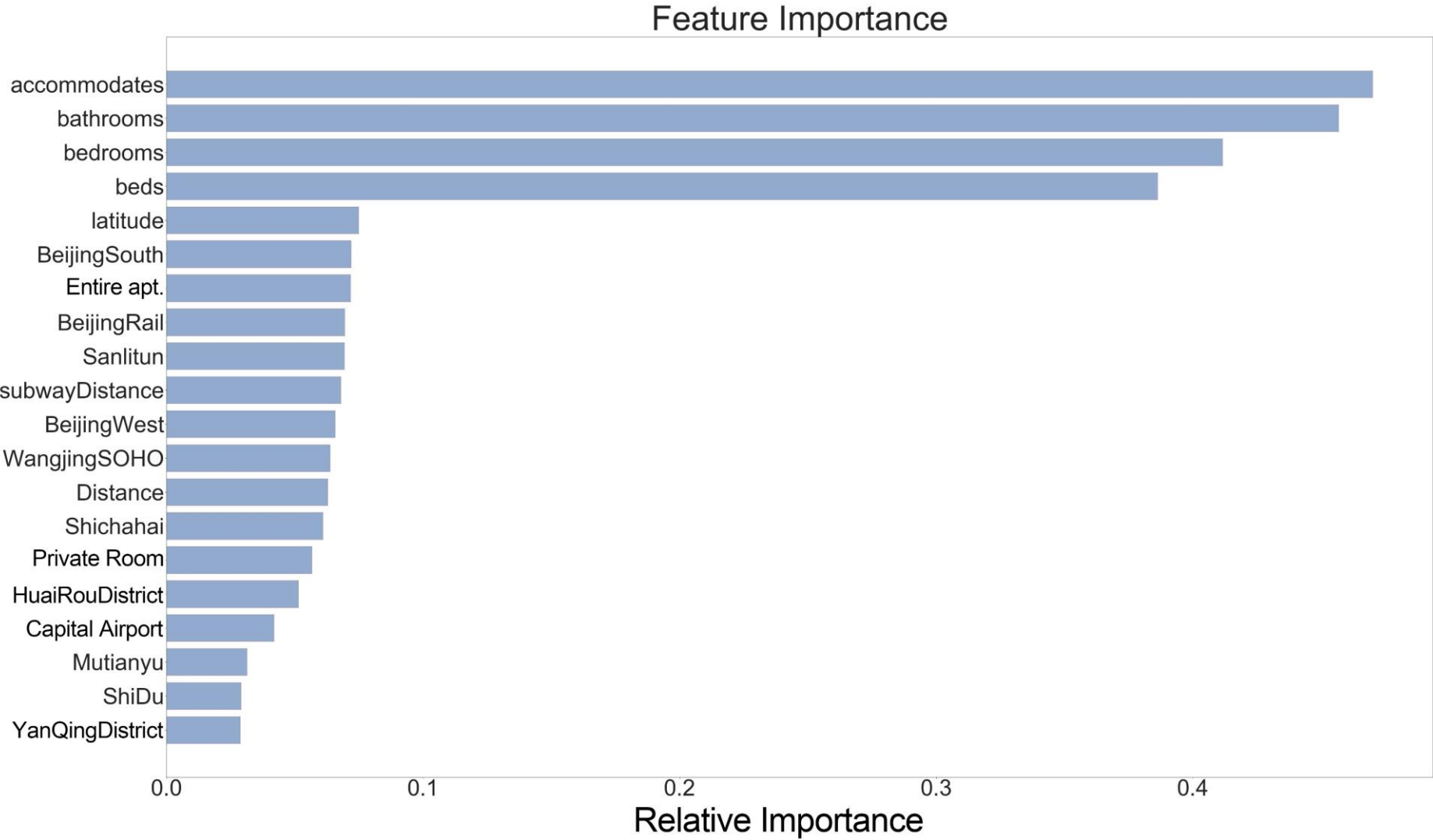


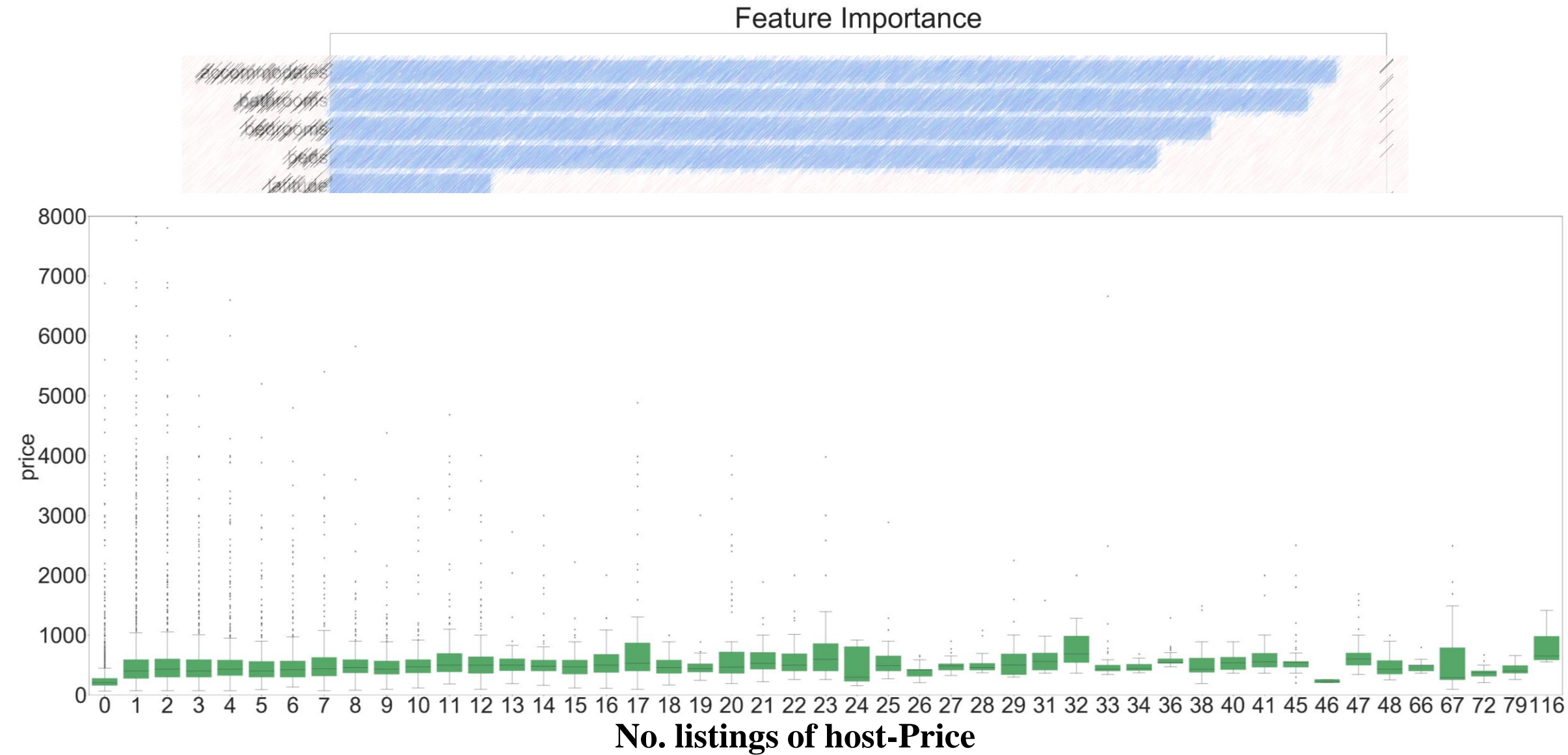
Feature Engineering



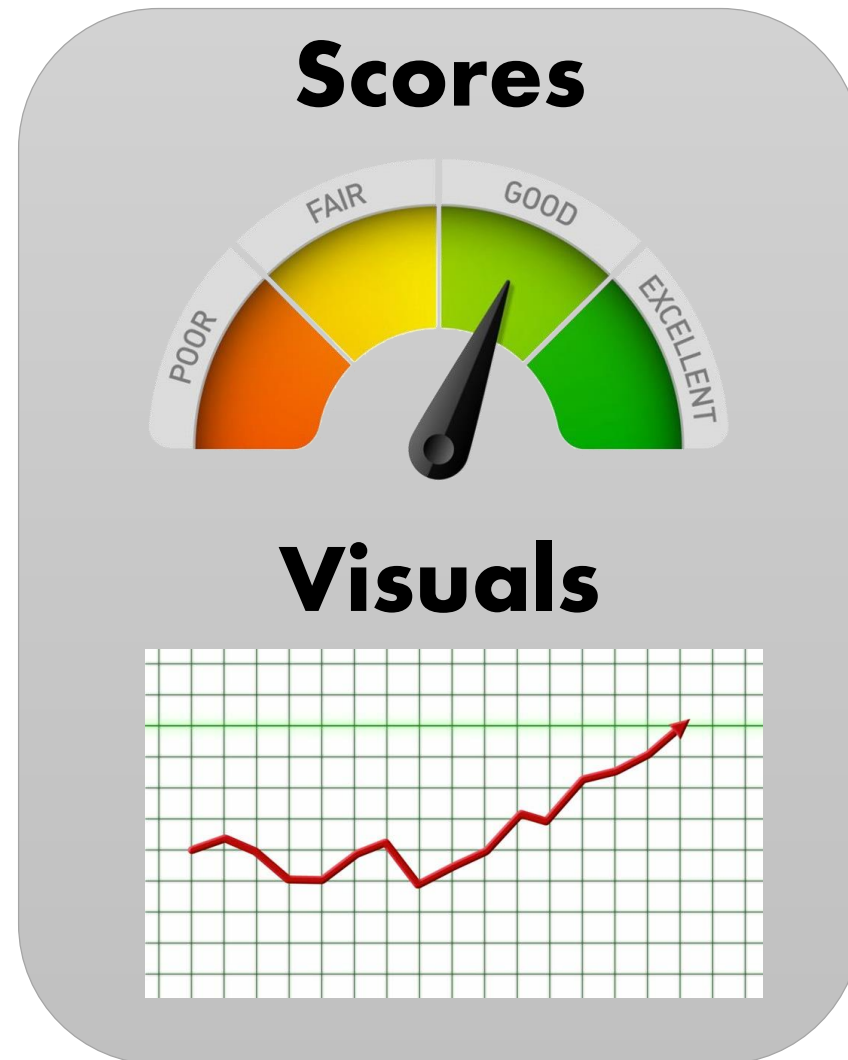
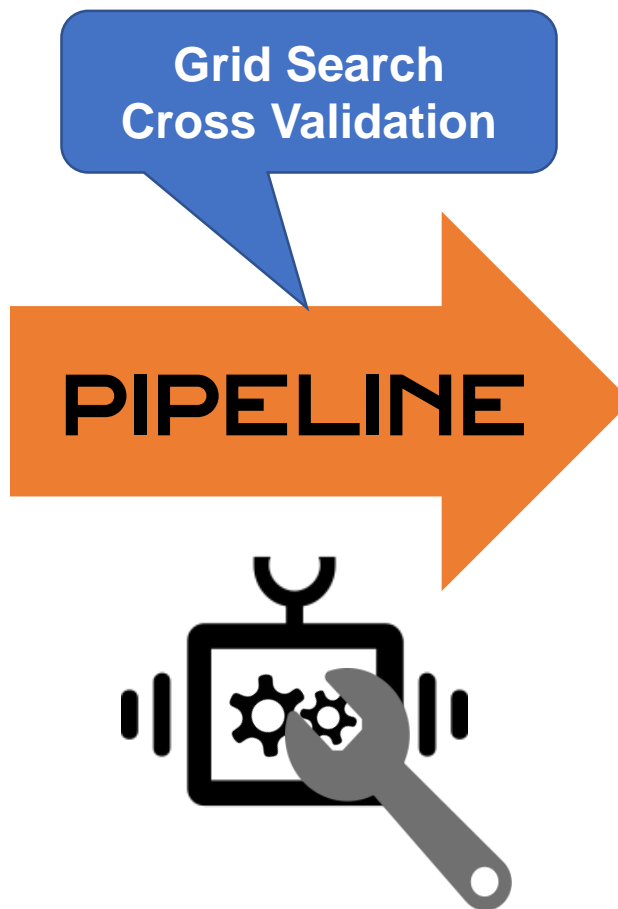
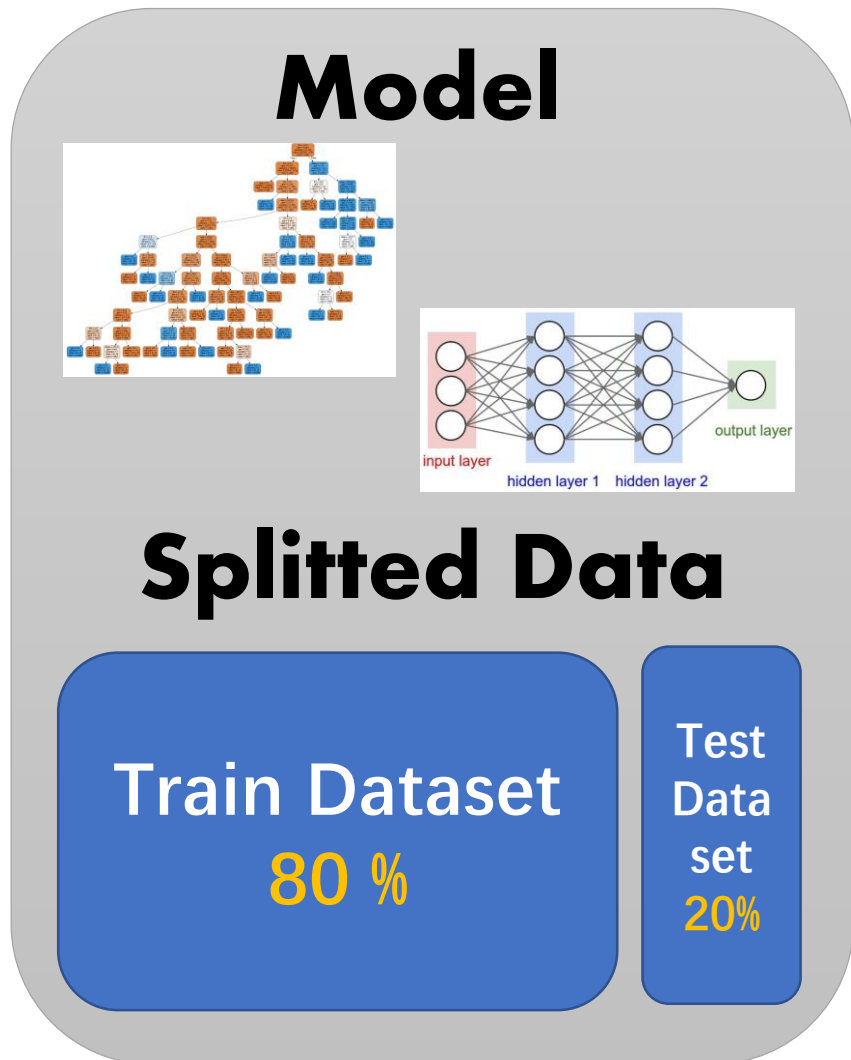




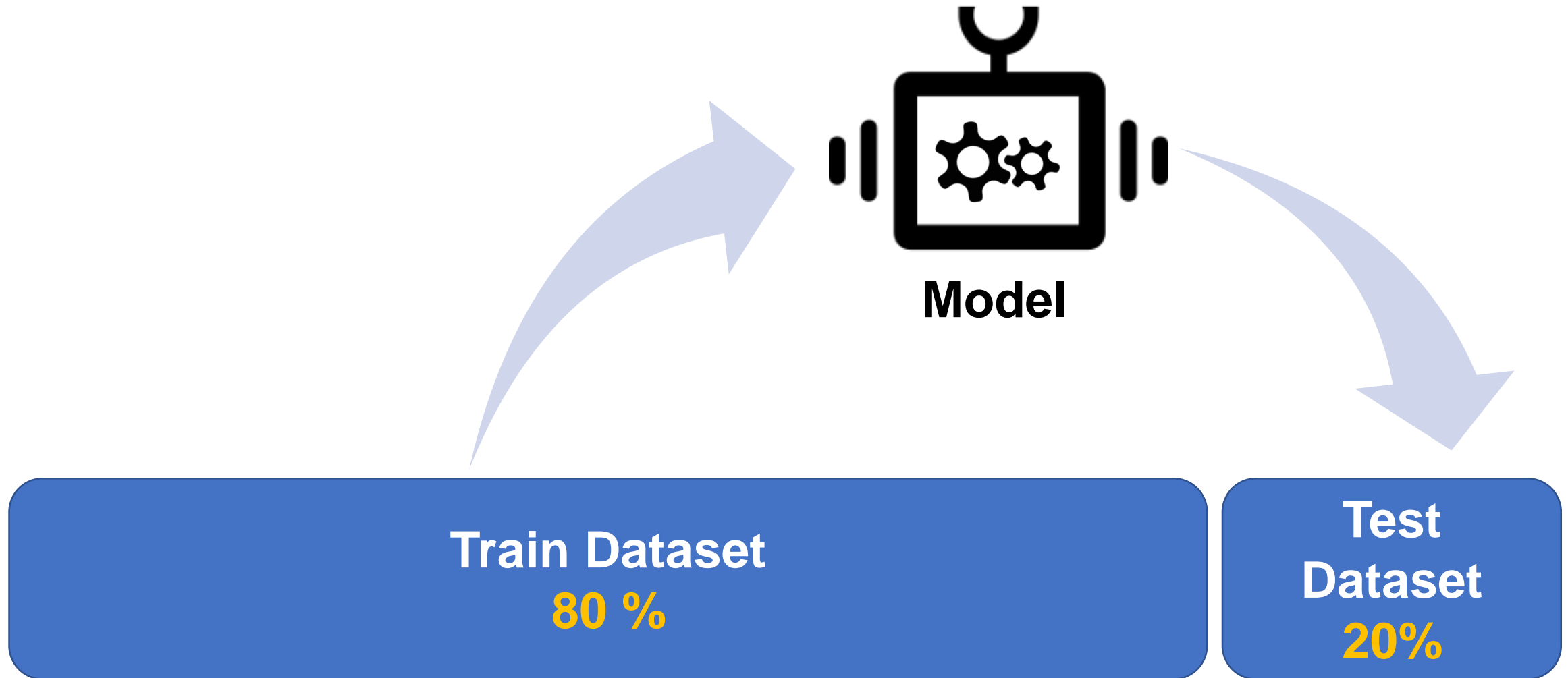




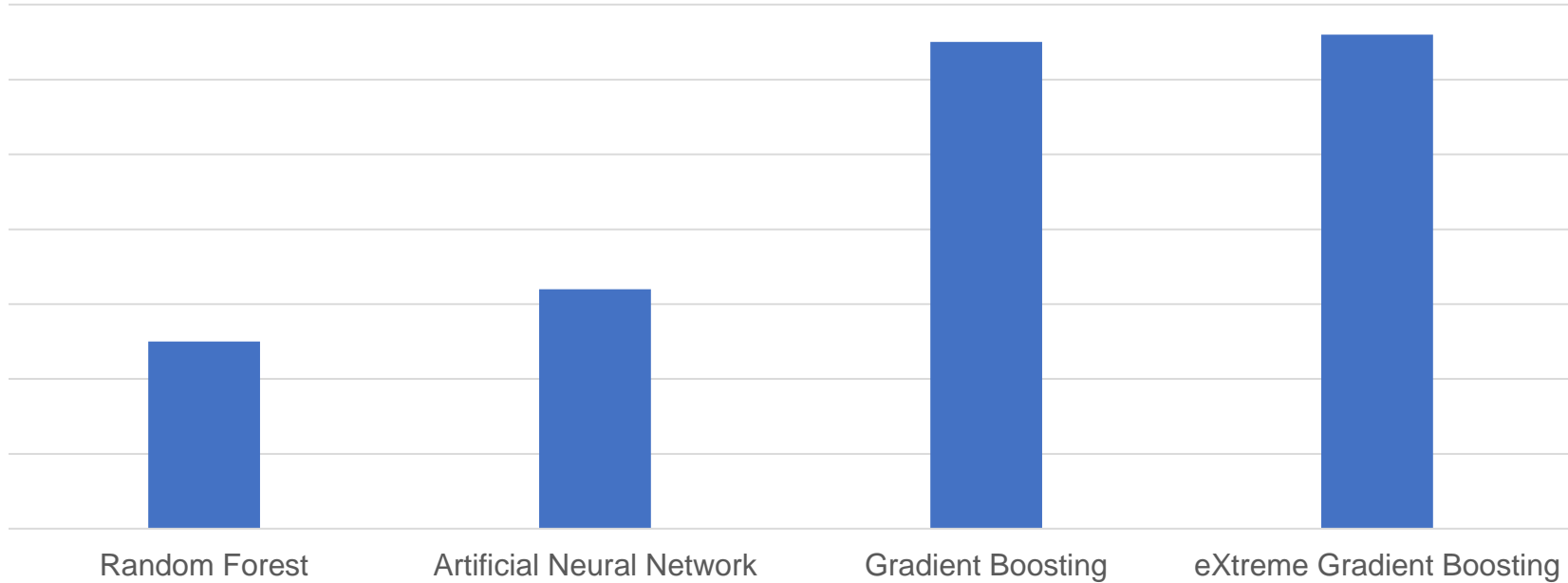
Model Building



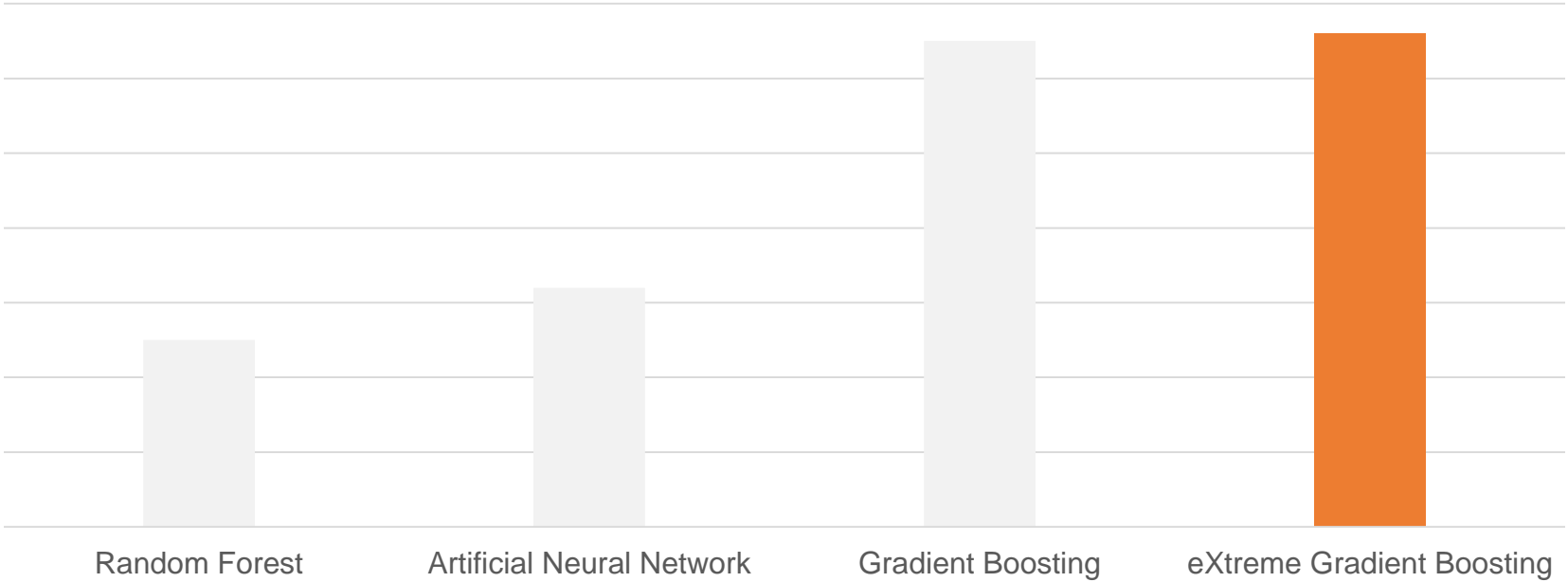
Evaluation

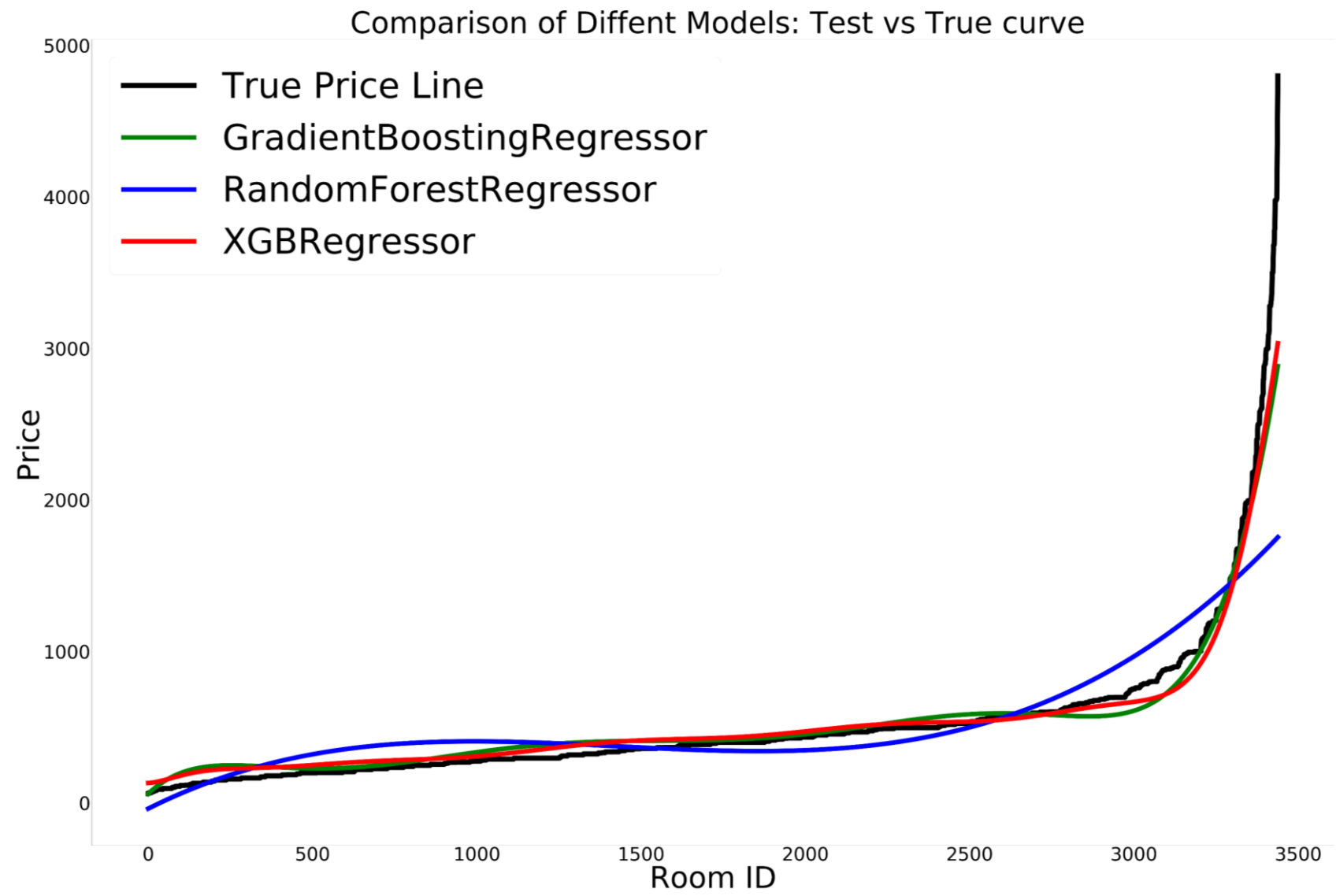


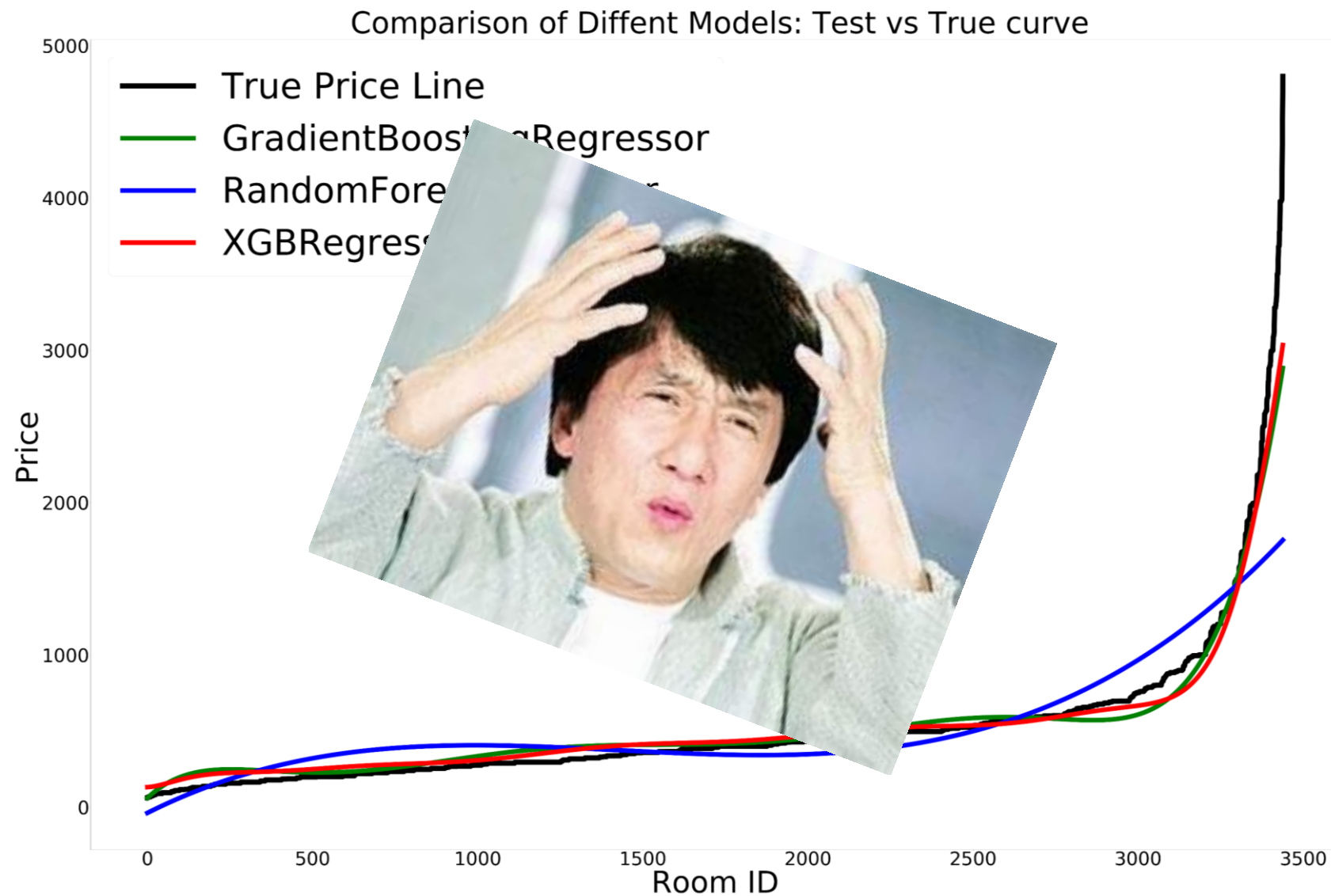
| Model Name | R ² Score |
|---------------------------|----------------------|
| Random Forest | 0.705 |
| Artificial Neural Network | 0.712 |
| Gradient Boosting | 0.745 |
| eXtreme Gradient Boosting | 0.746 |



| Model Name | R ² Score |
|----------------------------------|----------------------|
| Random Forest | 0.705 |
| Artificial Neural Network | 0.712 |
| Gradient Boosting | 0.745 |
| eXtreme Gradient Boosting | 0.746 |





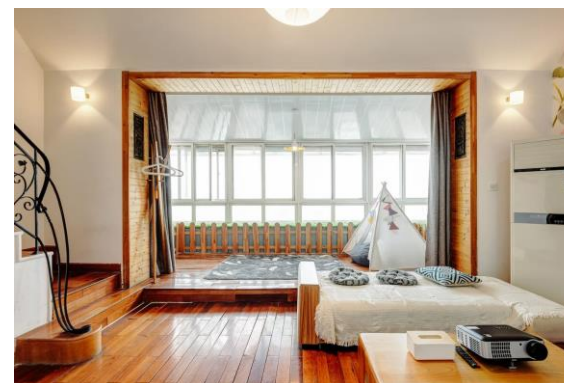


| Model Name | R ² Score | Prediction |
|----------------------------------|----------------------|------------------|
| Random Forest | 0.705 | ¥ 1798.06 |
| Artificial Neural Network | 0.712 | ¥ 1820.95 |
| Gradient Boosting | 0.745 | ¥ 1882.24 |
| eXtreme Gradient Boosting | 0.746 | ¥ 1887.61 |

Beijing

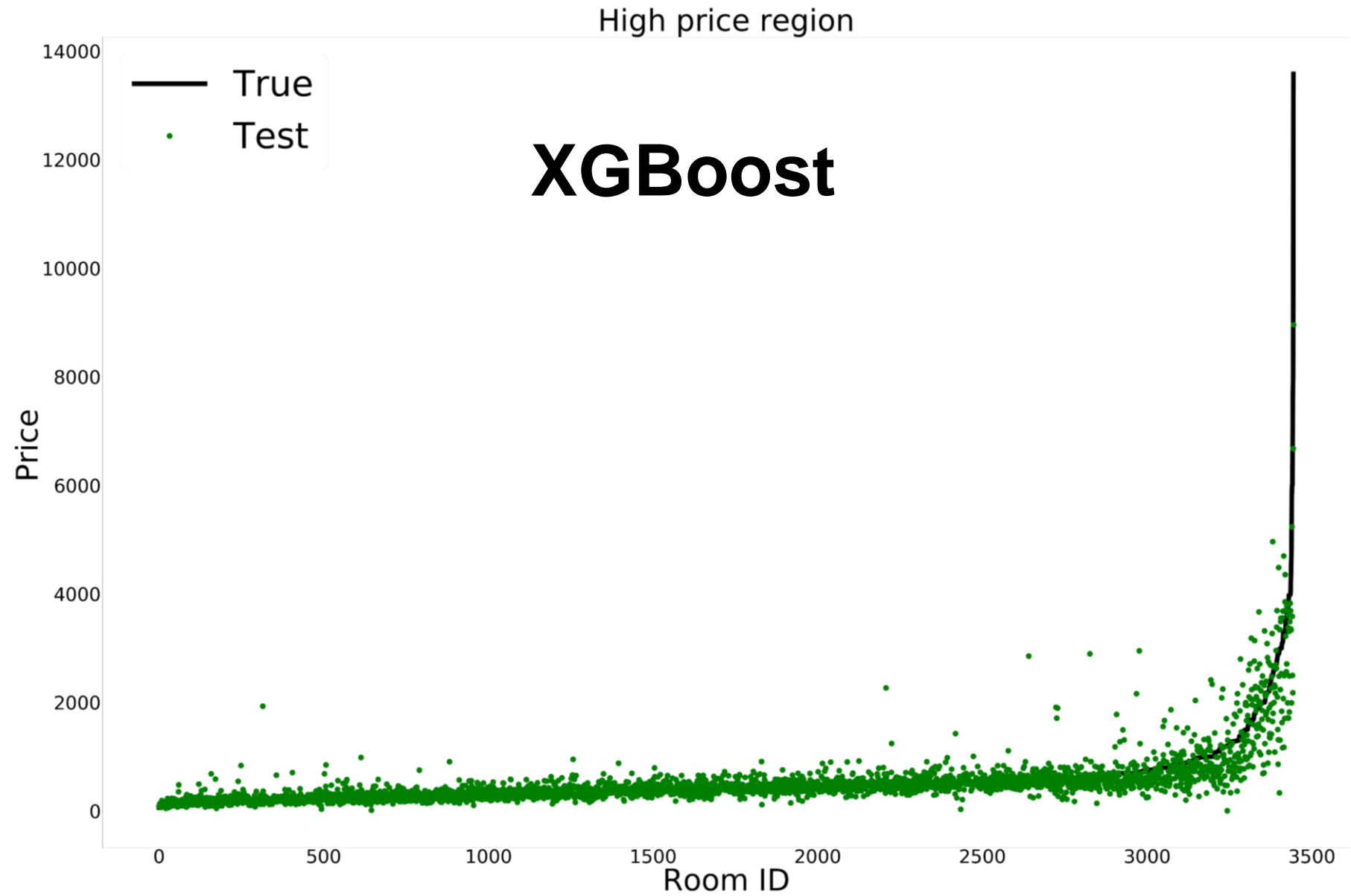
🏠 Entire house

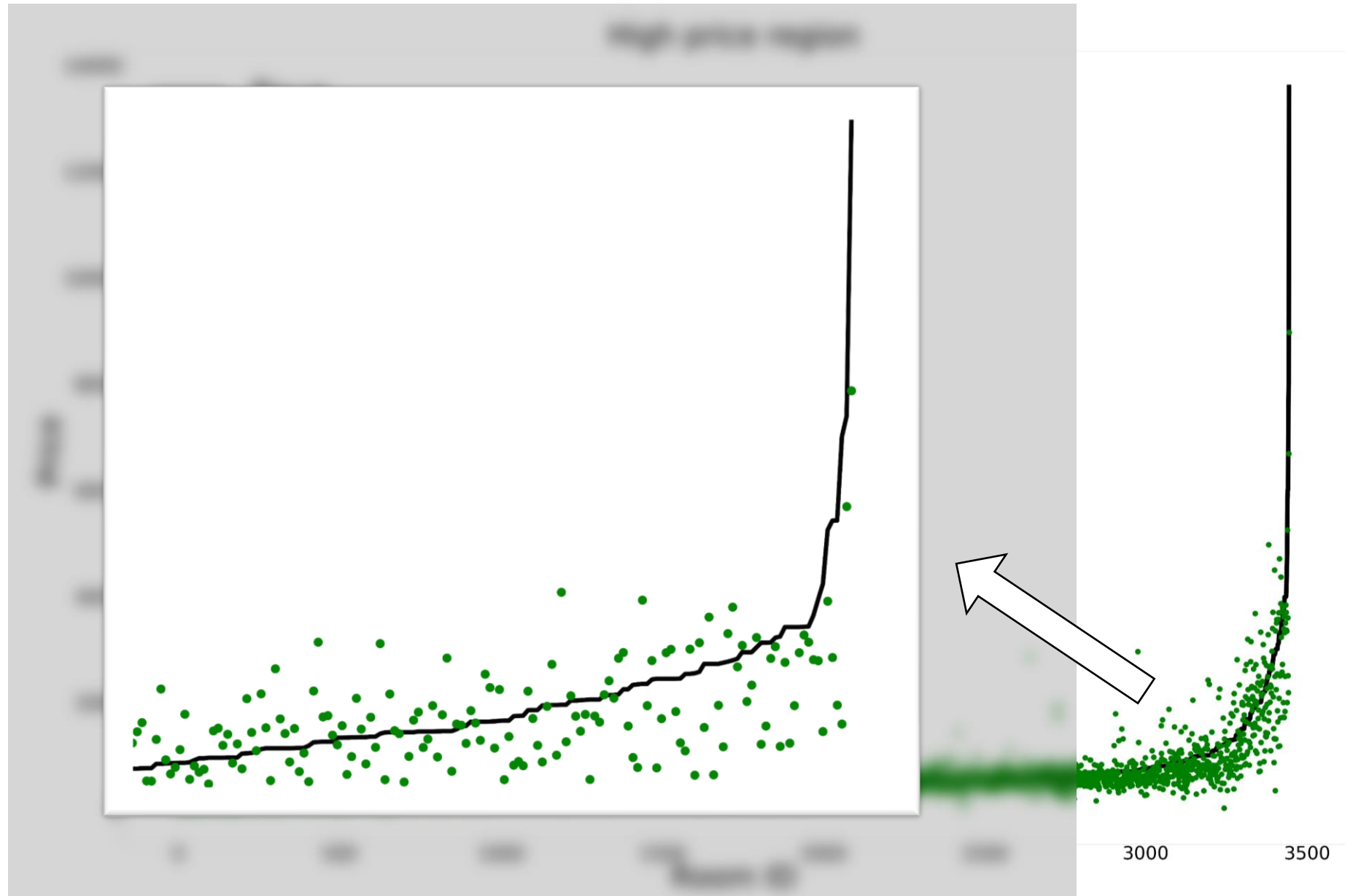
12 guests 6 bedrooms 6 beds 2 baths



¥ 1,888 per night







Error Analysis

- Expensive listings have **less** data compared to inexpensive listings

Inexpensive Listings

Expensive Listings

- Feature isn't everything for price.**

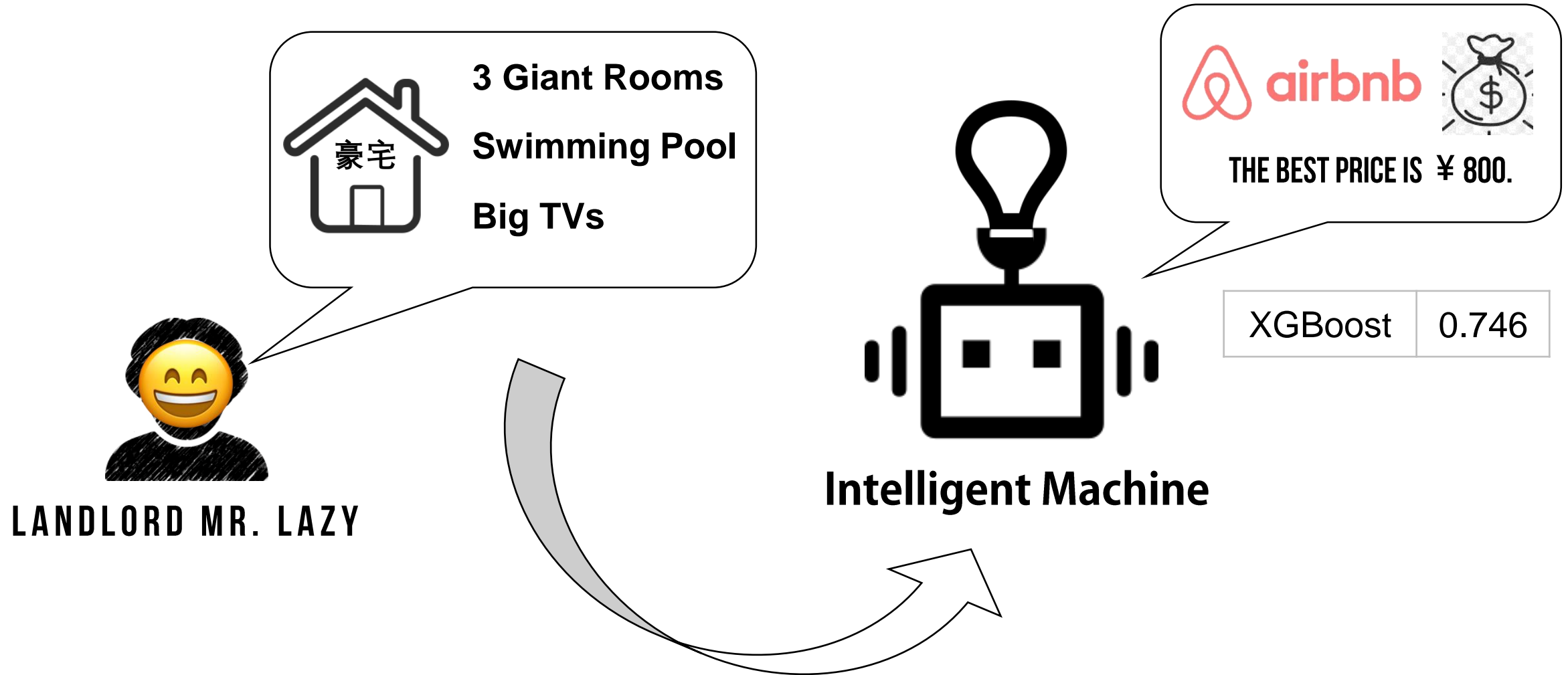
¥ 200



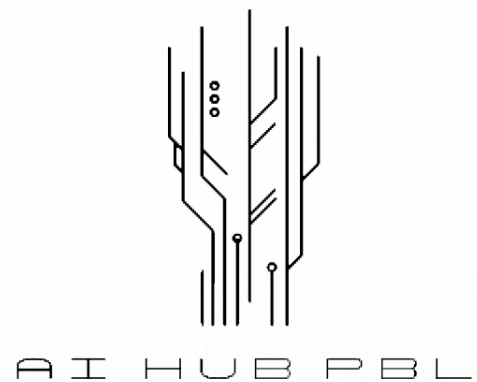
¥ 2348



To wrap up,



Thank you!



Thank you!
AI HUB 🥰



Q & A

