



中國人民大學

RENMIN UNIVERSITY OF CHINA

# Price & rent prediction model

The second group

實事求是

小组成员：赵芳誉 王子卿 陈玮灿 李冰滢

日期：2025年10月30日



➤ 数据预处理:

**房价模型:** 用训练集的中位数、众数等按板块、城市填充测试集的缺失值从源头避免了数据泄露。

**租金模型:** 将多种不同的租期描述（如“1年以上”、“季付价”）解析为统一的平均月数

➤ 特征工程

**房价模型:** 通过 K-Means 对经纬度聚类生成 location\_cluster，并对log\_建筑面积和房龄等关键连续变量进行分位数分箱，以捕捉到复杂的“地段溢价”和“非线性效应”。

**租金模型:** 面对区县和板块有较多类别的特征，用 TargetEncoder将该地段的“平均租金”信息编码为特征，并进行交叉验证

MAE	Price/Rent	In sample	out of sample	Cross-validation	Kaggle Score
OLS	Price	453,136.21	451,941.78	454,604.57	67.08
	Rent	118,912.75	119,972.06	119,841.25	
Lasso	Price	454,916.36	452,899.78	456,015.32	67.42
	Rent	119,219.66	120,216.47	120,143.04	
Ridge	Price	453,952.22	452,020.99	455,186.89	67.19
	Rent	122,038.54	123,175.82	123,206.70	