

Rent Estimations in Indianapolis, IN

Team Members: Hongyi
Chen, Sahinde Dogruer,
Tyler Ellis, Stefanie Wang

[https://github.com/stefaniewangphd/
Erdos_S22_IceBridge](https://github.com/stefaniewangphd/Erdos_S22_IceBridge)



The Erdős Institute
2022 Summer Data Science Boot Camp



Introduction

- Motivation: We want to instill confidence that a rental price is fair for the neighborhood and the listed property.
- Target Audience: Landlords and tenants in Indianapolis
- Main Question: Can we predict the rental price based on the properties' features?
- Our approaches: Cat Boost, Gradient Boost, Linear Regression, and Random Forest

Data Scraping

Apartments.com main page

8506 Westfield Blvd
Indianapolis, IN 46240
\$1,855
1 Bed
Apartment for Rent
(415) 607-4776
Email

434 Canal Ct S Dr
Indianapolis, IN 46202
\$1,733
1 Bed
Apartment for Rent
(415) 304-8805
Email

152 E New York St
Indianapolis, IN 46204
\$1,676
Studio
Apartment for Rent
(341) 766-0860
Email

4401 Eagle Creek Pkwy
Indianapolis, IN 46254
\$1,550
1 Bed

Individual Listings

Townhomes / Indiana / Indianapolis / 2372 Beckwith Dr
Today

2372 Beckwith Dr

2372 Beckwith Dr, Indianapolis, IN 46218
Martindale-Brightwood

Monthly Rent: \$751
Bedrooms: 2 bd
Bathrooms: 1 ba
Square Feet: 679 sq ft

Raw Rental Properties

url	listingid	ZIP	Monthly Rent	Bedrooms	Bathrooms	Square Feet	walkScore	transitScore	Deposit
https://www.apartments.com/2372-beckwith-dr-in...	ze0jqwy	46218.000	\$751	2 bd	1 ba	679 sq ft	25	33	\$705 deposit
https://www.apartments.com/the-marott-apartmen...	cqv6bmf	46208.000	1,185 - 1,645	1 - 2 bd	1 - 2 ba	609 - 1,092 sq ft	54	52	NaN
https://www.apartments.com/trails-at-lakeside-...	p38me3t	46220.000	1,193 - 2,805	1 - 3 bd	1 - 2 ba	594 - 1,246 sq ft	20	12	NaN
https://www.apartments.com/7491-n-shadeland-av...	s34dq54	46259.000	\$1,795	3 bd	2 ba	2,500 sq ft	54	33	NaN
https://www.apartments.com/nice-3-bedroom-ranc...	n669z9m	46237.000	\$1,095	3 bd	1 ba	1,439 sq ft	29	32	NaN

Data Cleaning

- Splitting multi-unit listings into multiple rows
- Extracting only numerical values from the numerical columns
- Handling with missing values
- Dropping unnecessary columns
- Dropping extreme values on the monthly_rent column

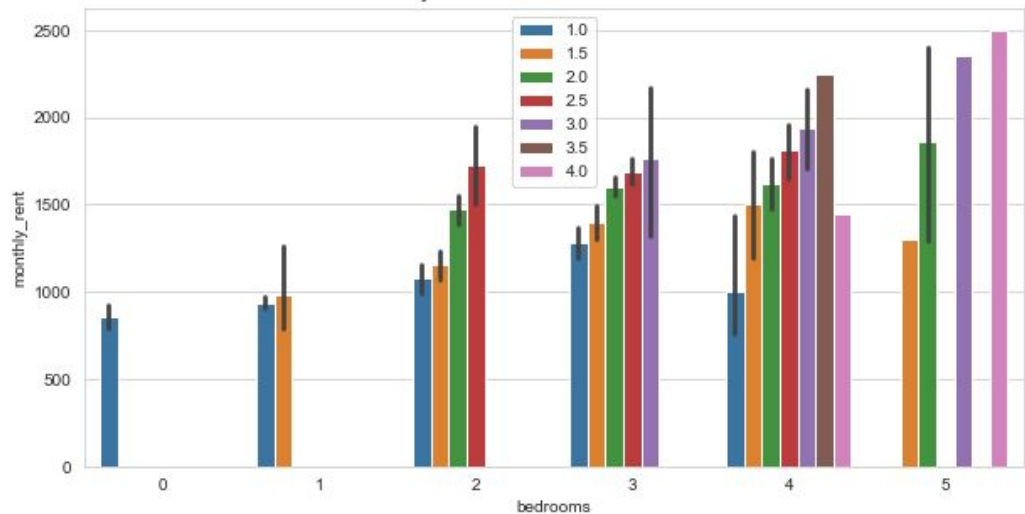
Raw Data vs Cleaned Data

1	df.head()										
	url	listingid	ZIP	Monthly Rent	Bedrooms	Bathrooms	Square Feet	walkScore	transitScore	Deposit	
	https://www.apartments.com/2372-beckwith-dr-in...	ze0jqwy	46218.000	\$751	2 bd	1 ba	679 sq ft	25	33	\$705 deposit	
	https://www.apartments.com/the-marott-apartmen...	cqv6bmf	46208.000	1,185 - 1,645	1 - 2 bd	1 - 2 ba	609 - 1,092 sq ft	54	52	NaN	
	https://www.apartments.com/trails-at-lakeside-...	p38me3t	46220.000	1,193 - 2,805	1 - 3 bd	1 - 2 ba	594 - 1,246 sq ft	20	12	NaN	
	https://www.apartments.com/7491-n-shadeland-av...	s34dq64	46250.000	\$1,795	3 bd	2 ba	2,500 sq ft	54	33	NaN	
	https://www.apartments.com/nice-3-bedroom-ranc...	n669z9m	46237.000	\$1,095	3 bd	1 ba	1,439 sq ft	29	32	NaN	

1	df_final.head()										
	url	listing_id	zipcode	monthly_rent	bedrooms	bathrooms	square_feet	walk_score	transit_score	deposit	
	https://www.apartments.com/2372-beckwith-dr-in...	ze0jqwy	46218	751	2	1.000	679.000	25	33	705	
	https://www.apartments.com/7491-n-shadeland-av...	s34dq64	46250	1795	3	2.000	2500.000	54	33	1795	
	https://www.apartments.com/nice-3-bedroom-ranc...	n669z9m	46237	1095	3	1.000	1439.000	29	32	1095	
	https://www.apartments.com/1102-n-oakland-ave-...	c4kr5zf	46201	1300	3	1.500	1500.000	67	38	800	
	https://www.apartments.com/634-e-10th-st-india...	r12dfp0	46202	1600	2	2.000	1400.000	74	53	1600	

EDA

Monthly Rent vs Bedrooms and Bathrooms

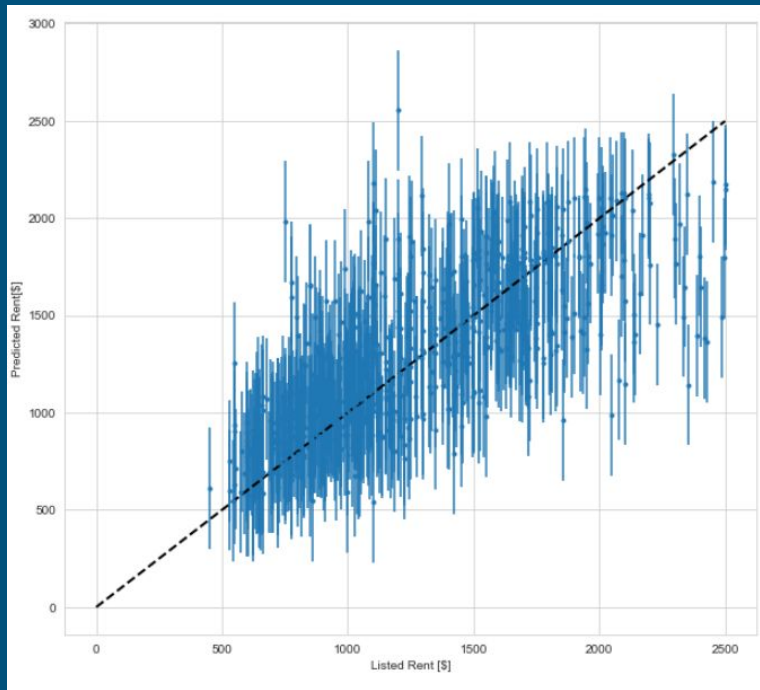


Correlation



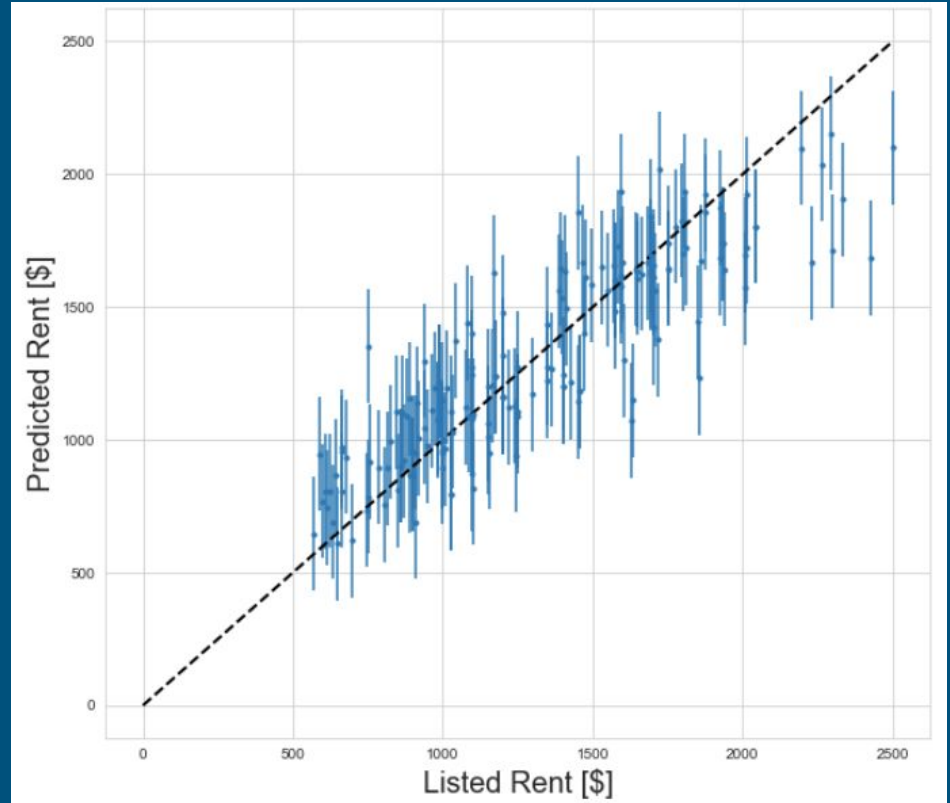
Gradient Boosting

- Least accurate by MSE
- Considered bedrooms, bathrooms, square feet, transit score, walk score



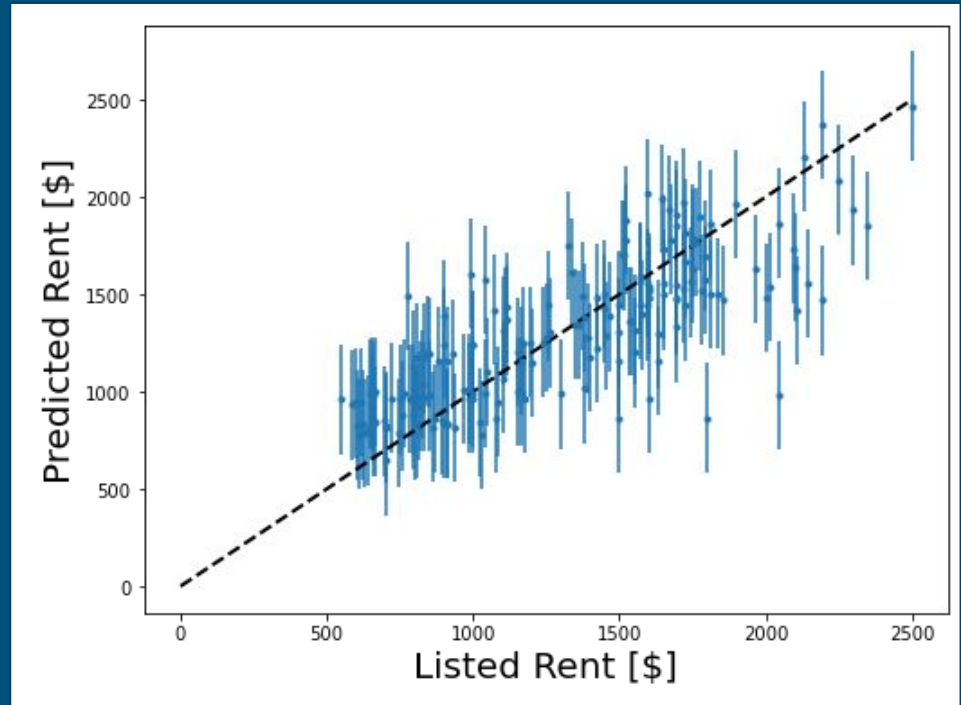
Cat Boost

- Most accurate of models employed
- Average error of \$214
- Robust handling of the categorical features (e.g. neighborhood)
- Struggled with more expensive rentals



Multiple Linear Regression

- Best performing MLR model consisted of features
 - Bedrooms, bathrooms, walk_score, bed*bath, sqft_bath, nbhd_top
- Struggled with mid-range and upper-range rental listings



Random Forest Regression Model

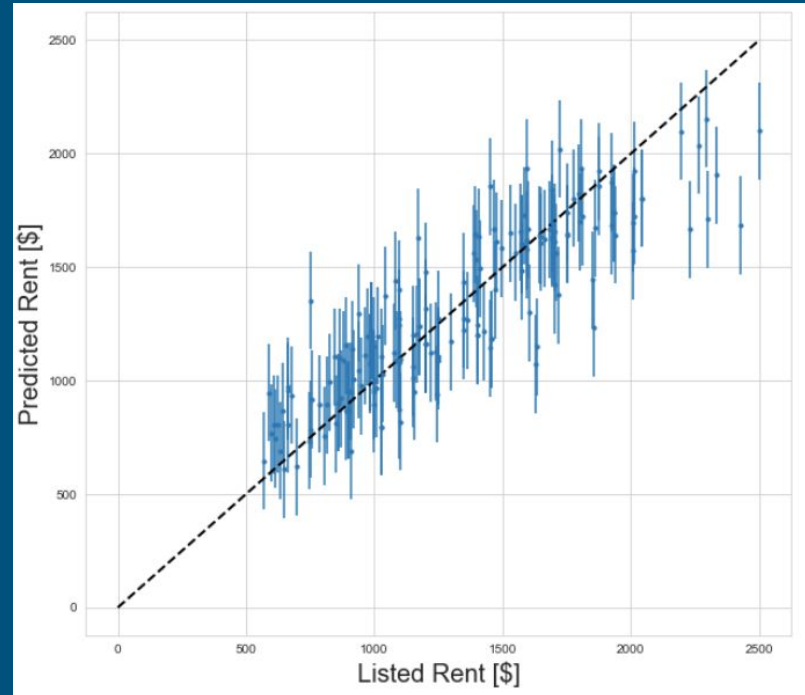
- Considering features zipcode, bedrooms, bathrooms, square root, top 10 neighbors as dummy columns.

- R^2 _score is 0.69

MAE is 205.71

MSE is 67572.06

RMSE is 259.94



Model comparison

Model	MSE	RMSE
Cat Boost	47388.74	217.68
Gradient Boosting	96963.55	311.39
Multiple Linear Regression	78401.04	280.00
Random Forest	67572.06	259.94

Key Takeaways

- All models outperforms the baseline (average rent over whole city)
- Cat Boost performs best; all models have room for improvements
- Currently, we can establish if a rent is approximately fair (within ~10%)

Future Directions

- Gather historical data to look at trends of rent prices in Indianapolis
- Extract more features from listing pages
- Better handling of multi-unit listings
- Tuning hyper-parameters of Cat Boost and Gradient Boost