Optimal Airbnb Pricing: evidence from 3 cities

Data Driven pricing considering geography, location, size and sazonability

Problem:

- There are over 6.6 million listing in Airbnb;
- How to find the optimal price per listing;
- Which features to consider;
- Are feature importance dependant of location

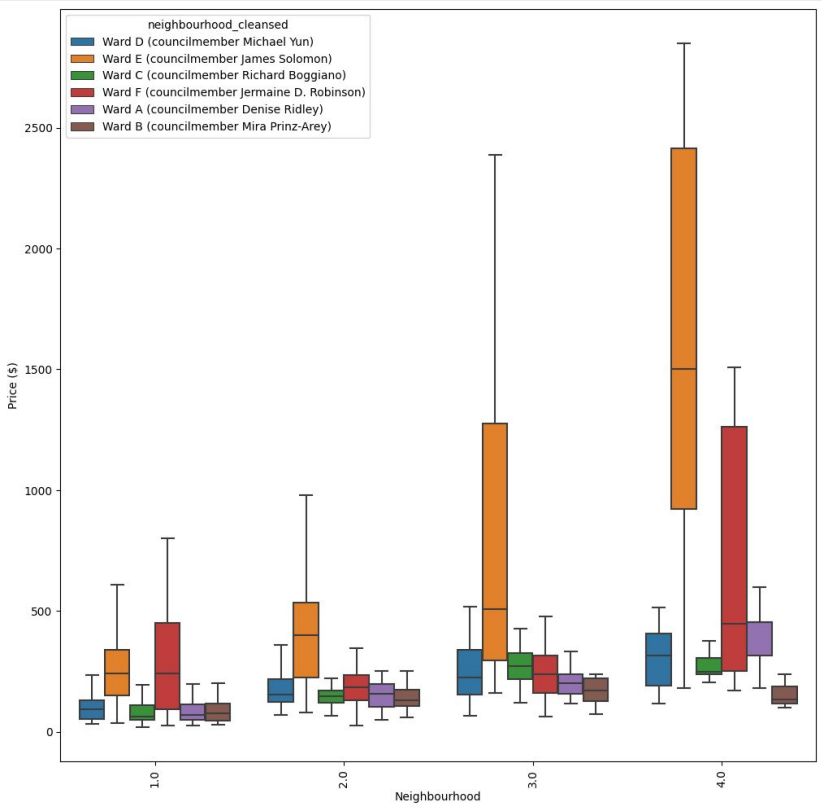
Data of 12 month, 3 cities across from 2 countries.

New York City, USA - 8,255,562 records

New Jersey, USA - 739,519 records

Rio de Janeiro, Brazil - 8,485,425 records

Jersey City

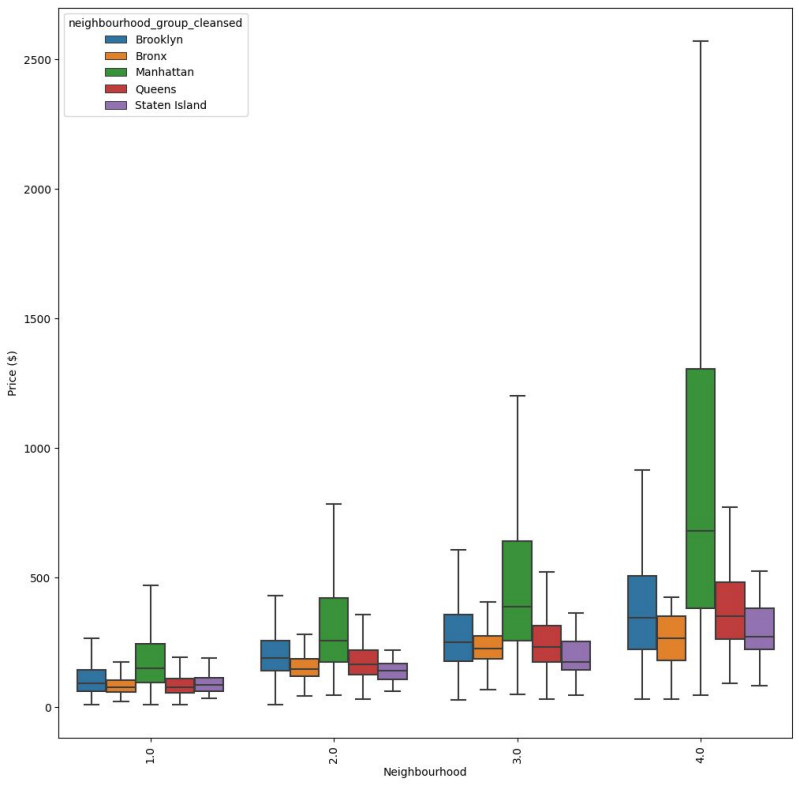



```
Ward E (councilmember James Solomon)      1076
Ward D (councilmember Michael Yun)         615
Ward F (councilmember Jermaine D. Robinson)  416
Ward C (councilmember Richard Boggiano)    215
Ward A (councilmember Denise Ridley)       134
Ward B (councilmember Mira Prinz-Arey)     105
Name: neighbourhood_cleansed, dtype: int64
```

Jersey City

bedrooms	count	mean	std	min	25%	50%	75%	max
1.0	1151.0	214.549088	212.947021	20.0	86.00	151.0	300.00	3300.0
2.0	454.0	294.370044	257.177908	41.0	130.00	200.0	388.25	2000.0
3.0	177.0	433.717514	579.000539	64.0	180.00	269.0	398.00	4929.0
4.0	56.0	942.625000	902.160146	100.0	264.00	448.0	1502.25	2850.0
5.0	13.0	685.461538	666.245527	170.0	253.00	442.0	885.00	2671.0
6.0	3.0	748.666667	652.441824	325.0	373.00	421.0	960.50	1500.0
7.0	2.0	650.500000	492.853426	302.0	476.25	650.5	824.75	999.0
8.0	1.0	467.000000	NaN	467.0	467.00	467.0	467.00	467.0
9.0	1.0	1000.000000	NaN	1000.0	1000.00	1000.0	1000.00	1000.0

New York

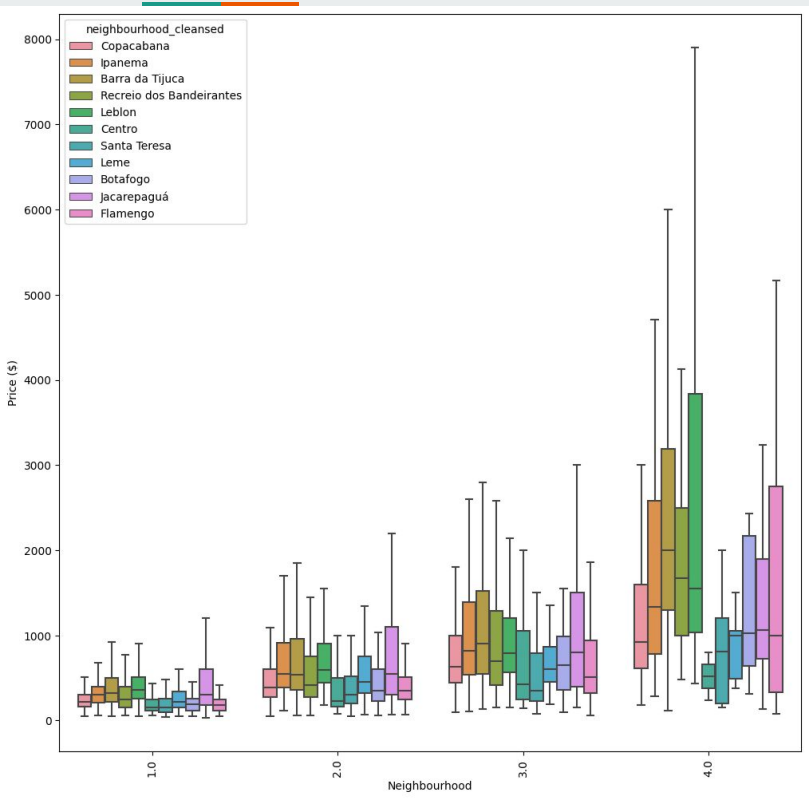


Manhattan	16748
Brooklyn	14817
Queens	6170
Bronx	1566
Staten Island	446

New York

	count	mean	std	min	25%	50%	75%	max
bedrooms								
1.0	14410.0	167.030951	317.985245	21.0	76.00	115.0	188.00	10000.0
2.0	4036.0	279.810951	271.157878	10.0	150.00	201.0	305.25	4143.0
3.0	1594.0	375.380803	390.624732	28.0	195.00	272.5	400.00	5000.0
4.0	388.0	686.481959	1067.399609	30.0	293.25	412.5	737.00	16500.0
5.0	92.0	939.130435	1179.478794	65.0	425.00	619.5	973.00	10000.0
6.0	30.0	1088.233333	917.258535	200.0	551.00	741.0	1260.75	3557.0
7.0	10.0	1012.800000	683.670616	545.0	643.00	739.0	1072.50	2800.0
8.0	7.0	1107.428571	524.106496	745.0	824.50	879.0	1118.00	2243.0
9.0	2.0	821.500000	111.015765	743.0	782.25	821.5	860.75	900.0
13.0	1.0	410.000000	NaN	410.0	410.00	410.0	410.00	410.0

Rio de Janeiro



Copacabana	7525
Barra da Tijuca	2723
Ipanema	2603
Jacarepaguá	1457
Recreio dos Bandeirantes	1327
Leblon	1262
Botafogo	1115
Santa Teresa	918
Centro	689
Flamengo	568
Leme	481
Laranjeiras	396
Tijuca	392
Camorim	303
Lagoa	218
Glória	208
São Conrado	195
Gávea	189
Catete	180
Jardim Botânico	172
Vidigal	153
Humaitá	148
Taquara	138
Itanhangá	132
Freguesia (Jacarepaguá)	130
Vila Isabel	128

Rio de Janeiro

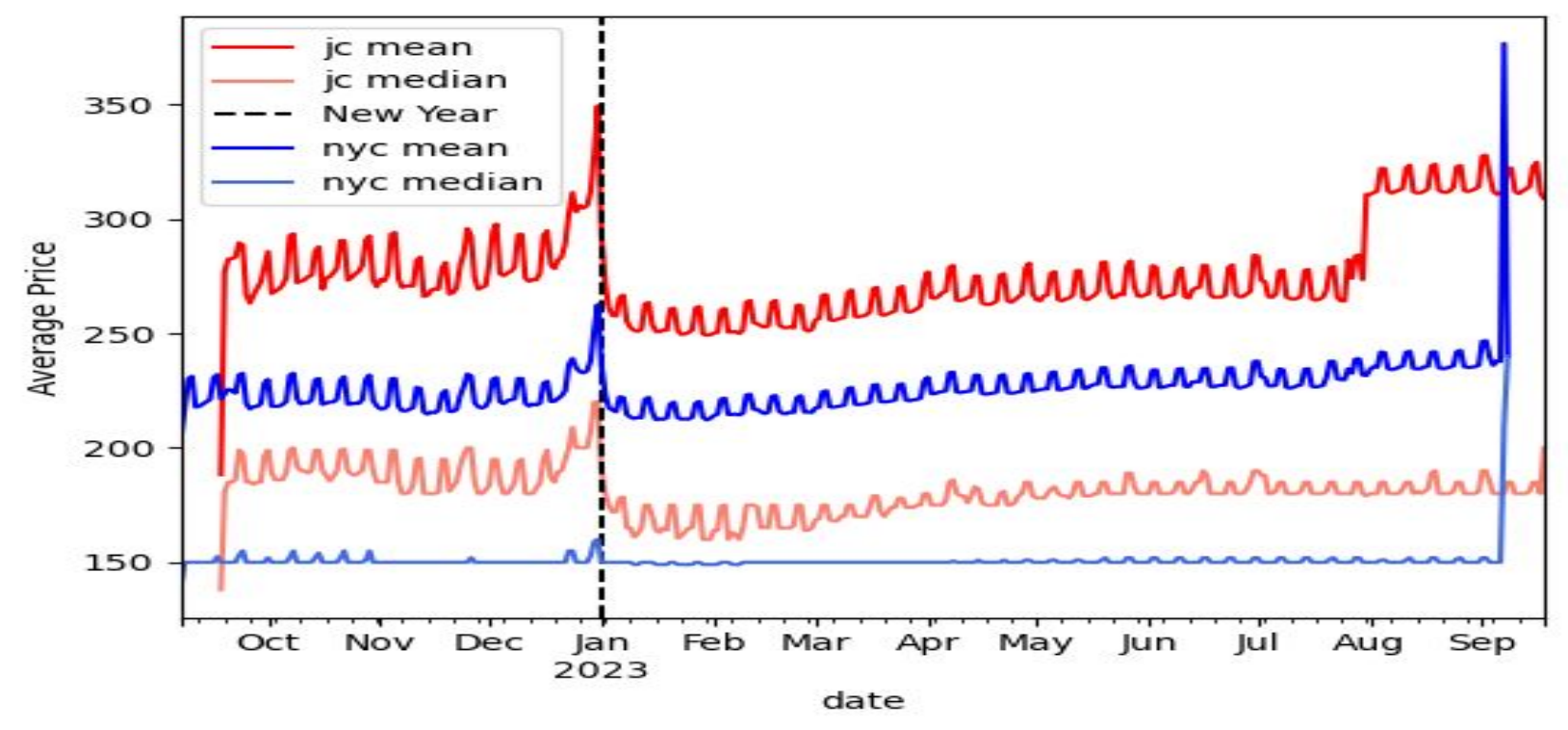
bedrooms	count	mean	std	min	25%	50%	75%	max
1.0	12246.0	464.652540	4106.403068	40.0	150.0	230.0	358.75	335103.0
2.0	6106.0	737.162299	1891.434099	48.0	290.0	450.0	800.00	100000.0
3.0	2730.0	1217.154579	3133.680076	60.0	452.5	707.5	1241.25	100000.0
4.0	665.0	2596.759398	4876.963122	70.0	800.0	1500.0	3000.00	100000.0
5.0	207.0	4696.521739	8501.356812	80.0	1182.0	2200.0	4704.00	79255.0
6.0	108.0	3939.694444	4610.451178	70.0	1200.0	2650.0	5500.00	40000.0
7.0	34.0	3609.029412	3991.243882	163.0	1164.0	2200.5	4710.25	15343.0
8.0	13.0	5065.000000	4079.808125	906.0	2329.0	4000.0	7357.00	16000.0
9.0	11.0	5655.363636	3412.055752	70.0	3363.0	6250.0	7506.00	11100.0
10.0	11.0	5324.363636	7553.355907	120.0	210.0	1800.0	7050.00	25000.0
11.0	4.0	2481.250000	3613.889537	329.0	683.0	853.5	2651.75	7889.0
12.0	3.0	3939.000000	5682.293815	600.0	658.5	717.0	5608.50	10500.0
13.0	1.0	350.000000	NaN	350.0	350.0	350.0	350.00	350.0
15.0	3.0	2000.000000	1562.049935	200.0	1500.0	2800.0	2900.00	3000.0

Percentage and absolute number of type of listings for each city

	Jersey City	%	NYC	%	Rio	%
Entire rental unit	1270.0	0.495900	16830.0	0.423428	14811.0	0.562920
Private room in rental unit	222.0	0.086685	10443.0	0.262737	3141.0	0.119380
Entire condo	210.0	0.081999	1625.0	0.040884	1984.0	0.075406
Entire serviced apartment	204.0	0.079656	614.0	0.015448	864.0	0.032838
Private room in home	150.0	0.058571	2129.0	0.053564	996.0	0.037855
Room in hotel	134.0	0.052323	658.0	0.016555	147.0	0.005587
Entire home	76.0	0.029676	1468.0	0.036934	1174.0	0.044620
Entire loft	74.0	0.028895	746.0	0.018769	811.0	0.030824
Private room in condo	52.0	0.020305	598.0	0.015045	472.0	0.017939
Entire townhouse	38.0	0.014838	659.0	0.016580	22.0	0.000836
Room in boutique hotel	37.0	0.014447	465.0	0.011699	46.0	0.001748
Private room in townhouse	26.0	0.010152	1065.0	0.026794	44.0	0.001672
Entire guest suite	18.0	0.007029	391.0	0.009837	63.0	0.002394

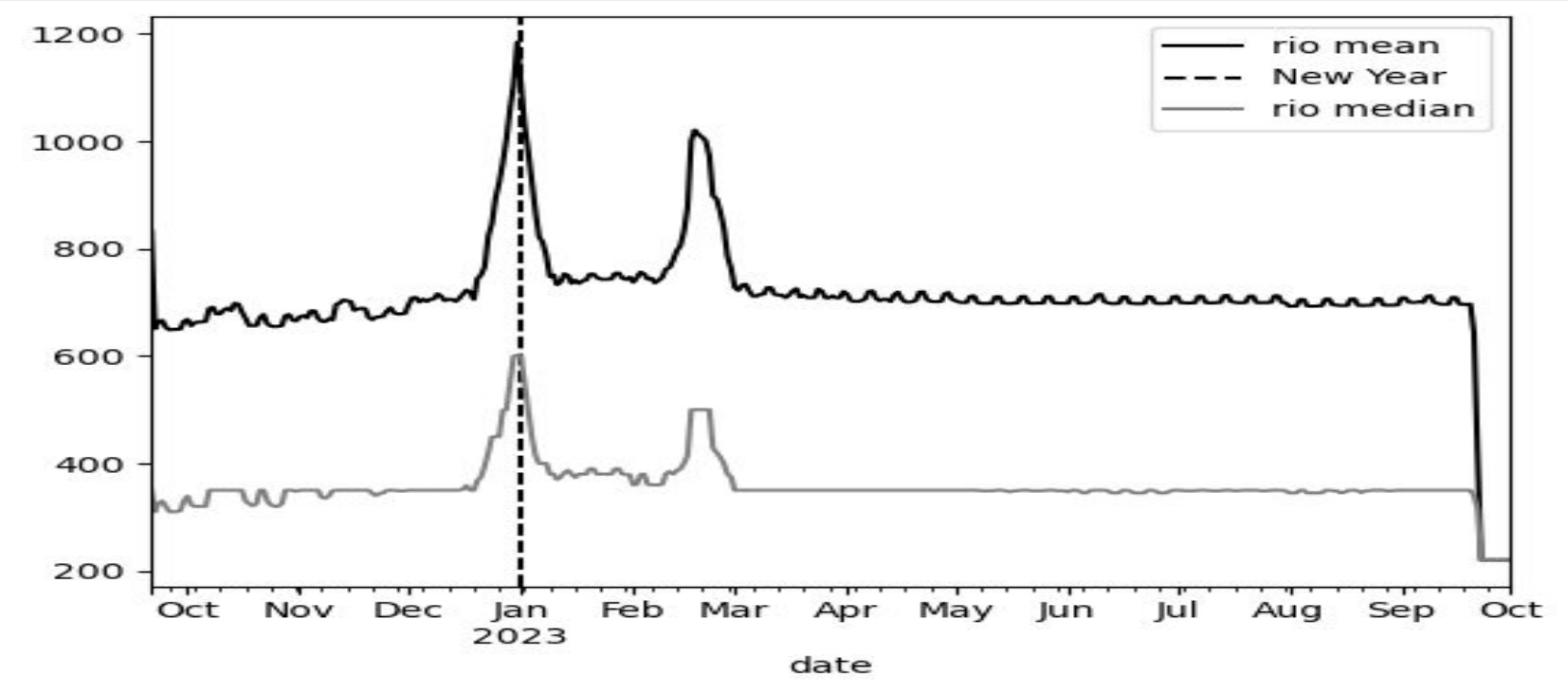
Jersey City and New York

Monthly Seasonality Study



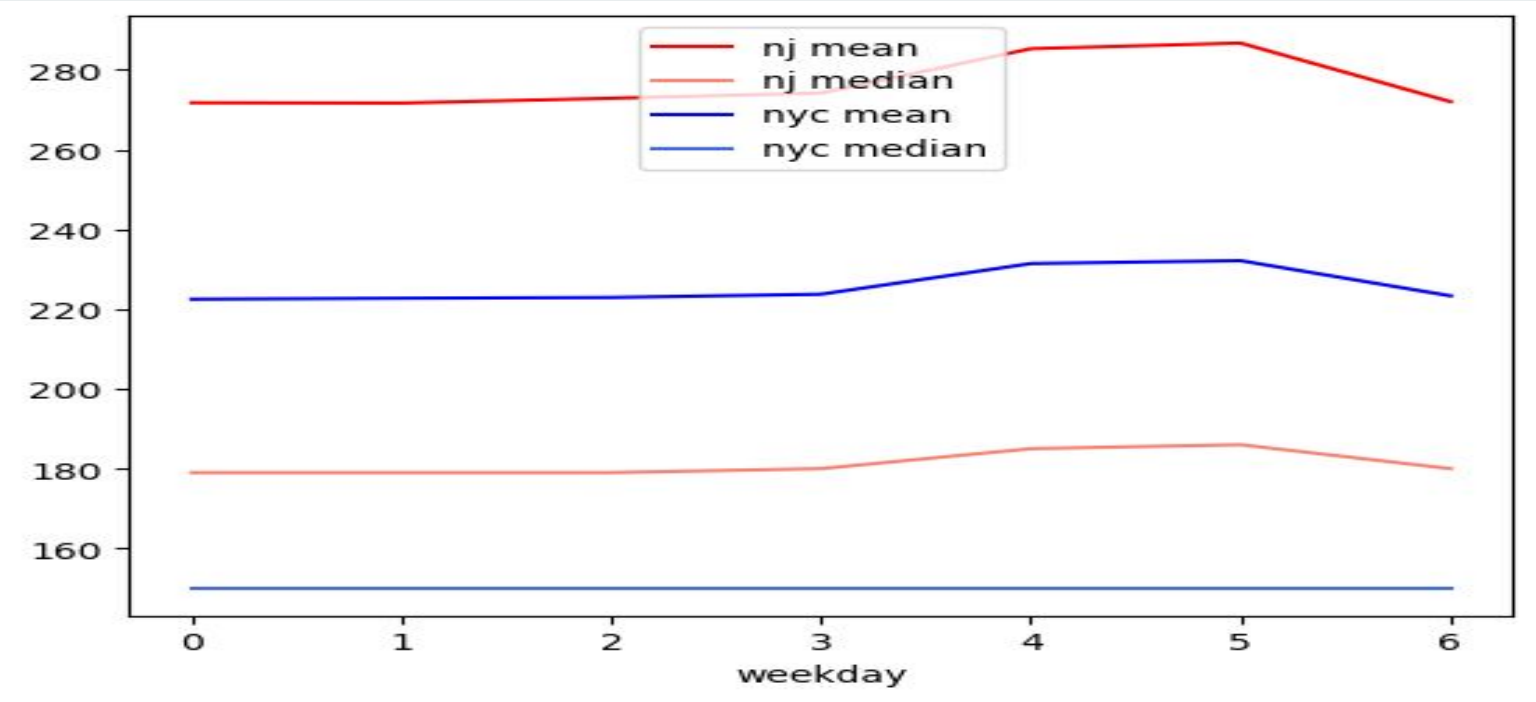
Rio de Janeiro

Monthly Seasonality Study



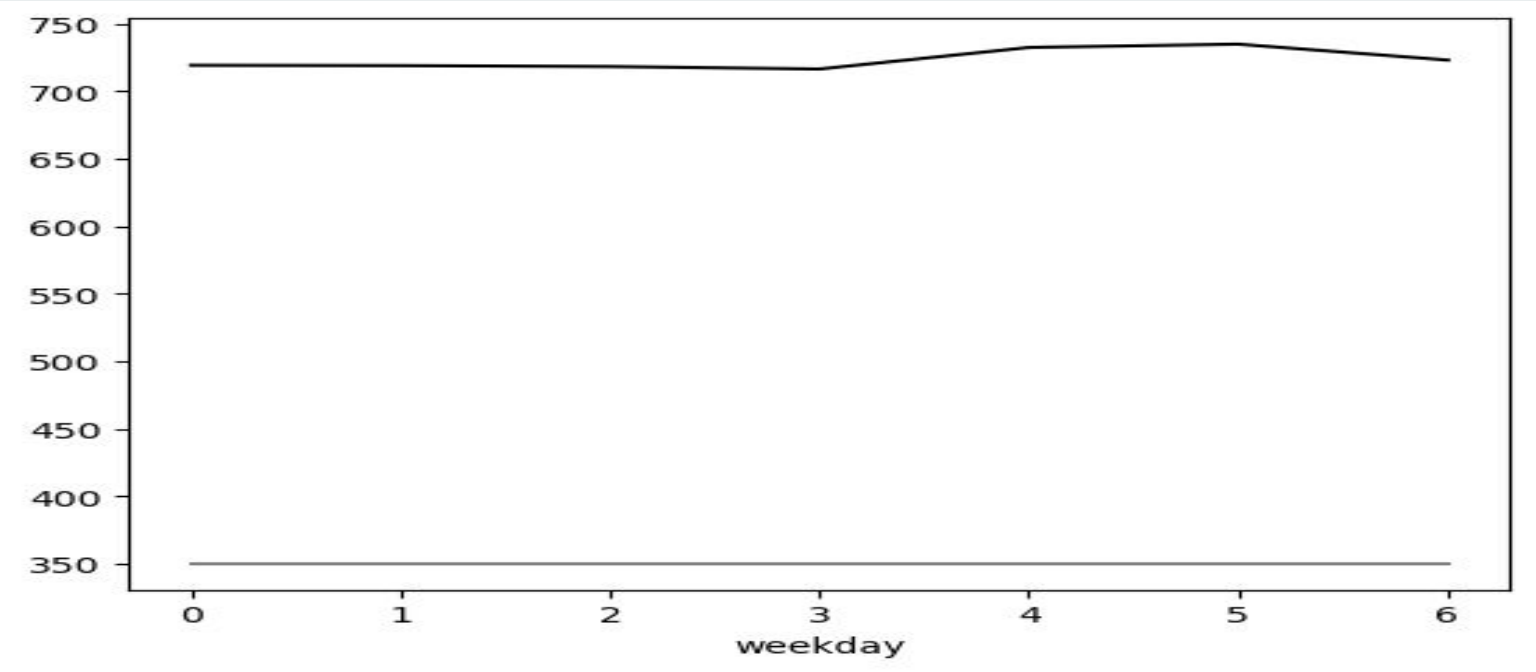
Jersey City and New York

Daily Seasonality Study



Rio de Janeiro

Daily Seasonality Study



Seasonality Study of the Cities

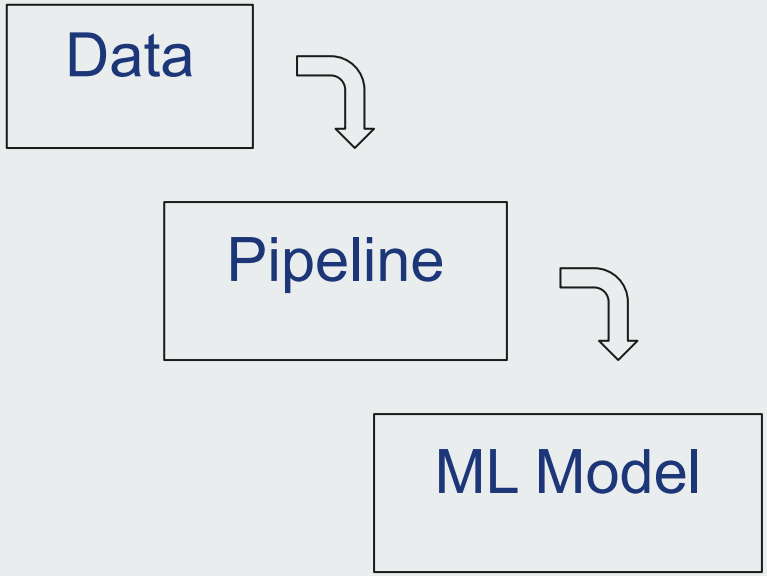
Monthly Seasonality

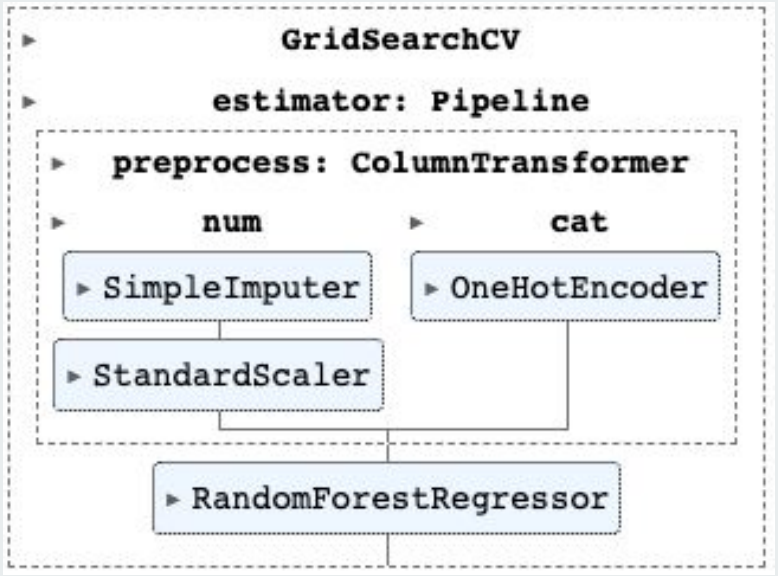
bedrooms	Jersey City	NYC	Rio	Rio_carnaval
1	0.222046	0.163604	0.762932	0.376349
2	0.250668	0.160109	0.632019	0.360687
3	0.294405	0.244866	0.777436	0.498302

Daily Seasonality

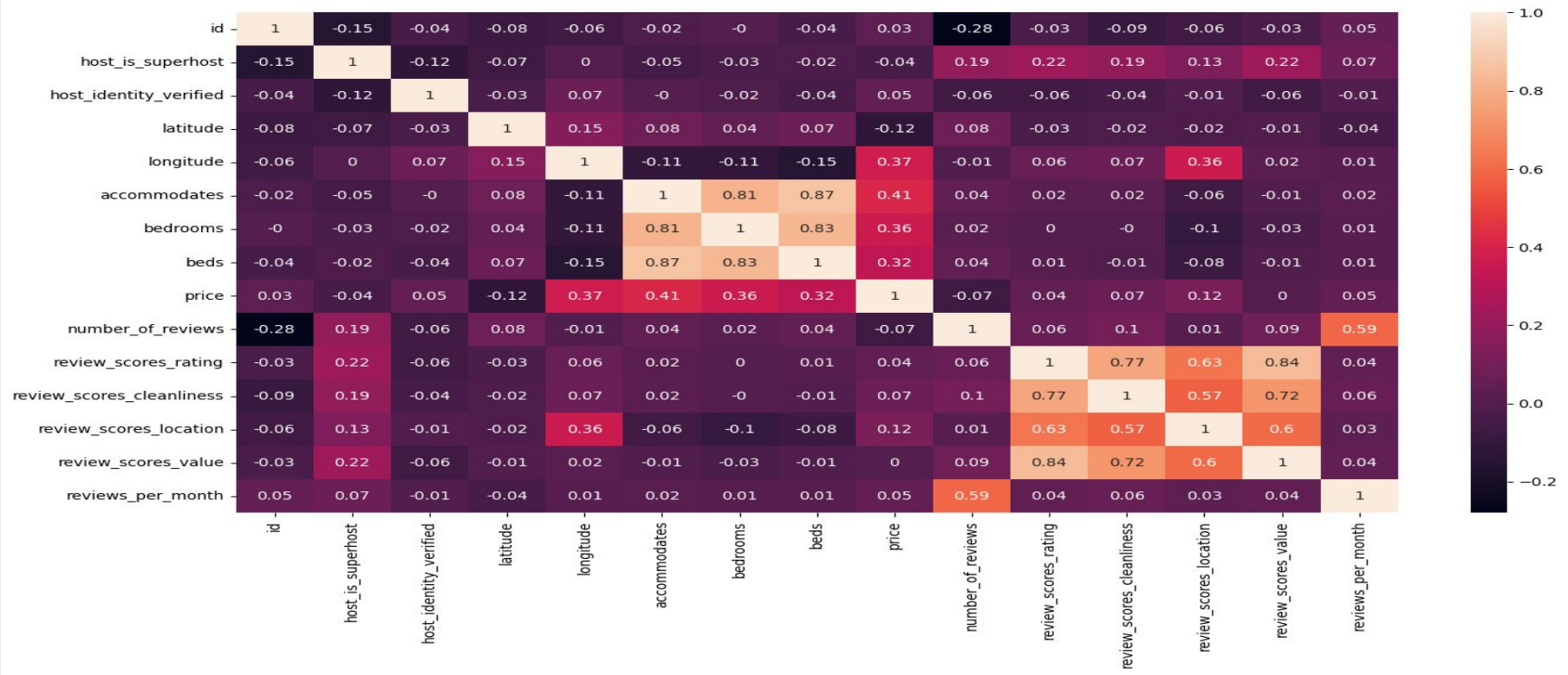
	Jersey City	NYC	Rio
price	0.054186	0.034464	0.028127
price	0.046688	0.040488	0.016723
price	0.086090	0.062983	0.028815

Model Structure





Feature Engineering - Jersey City

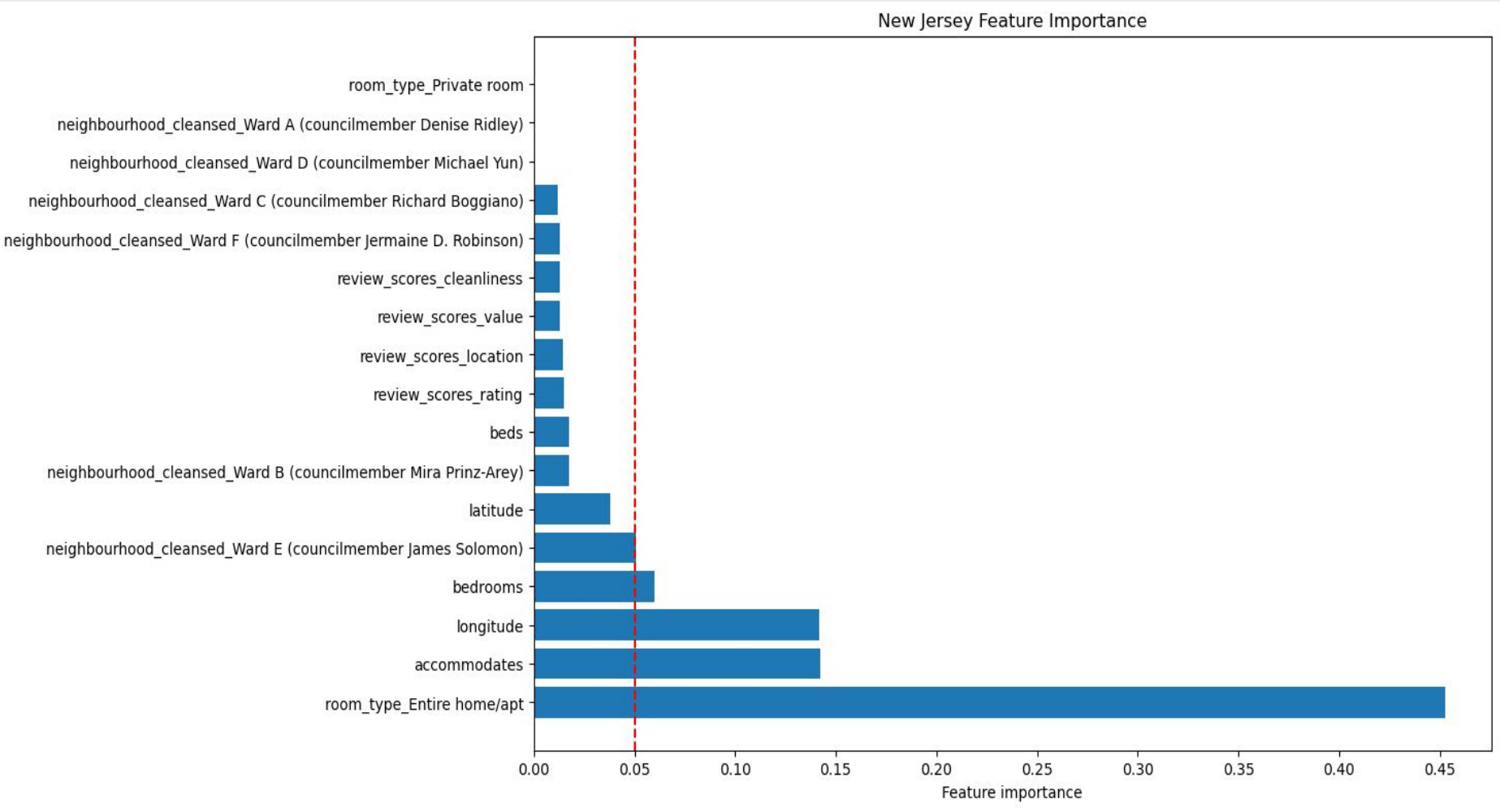


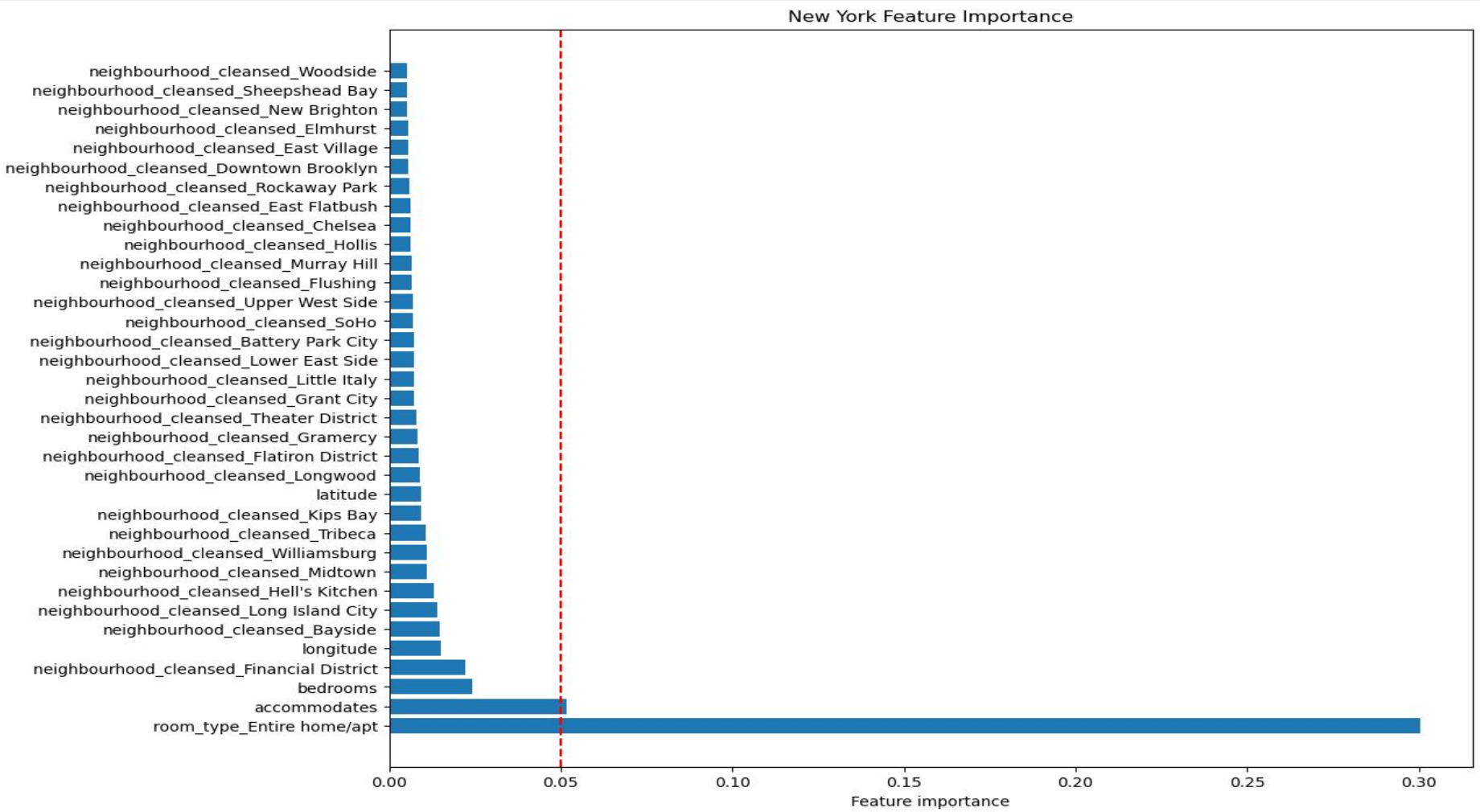
Feature Engineering - New York City

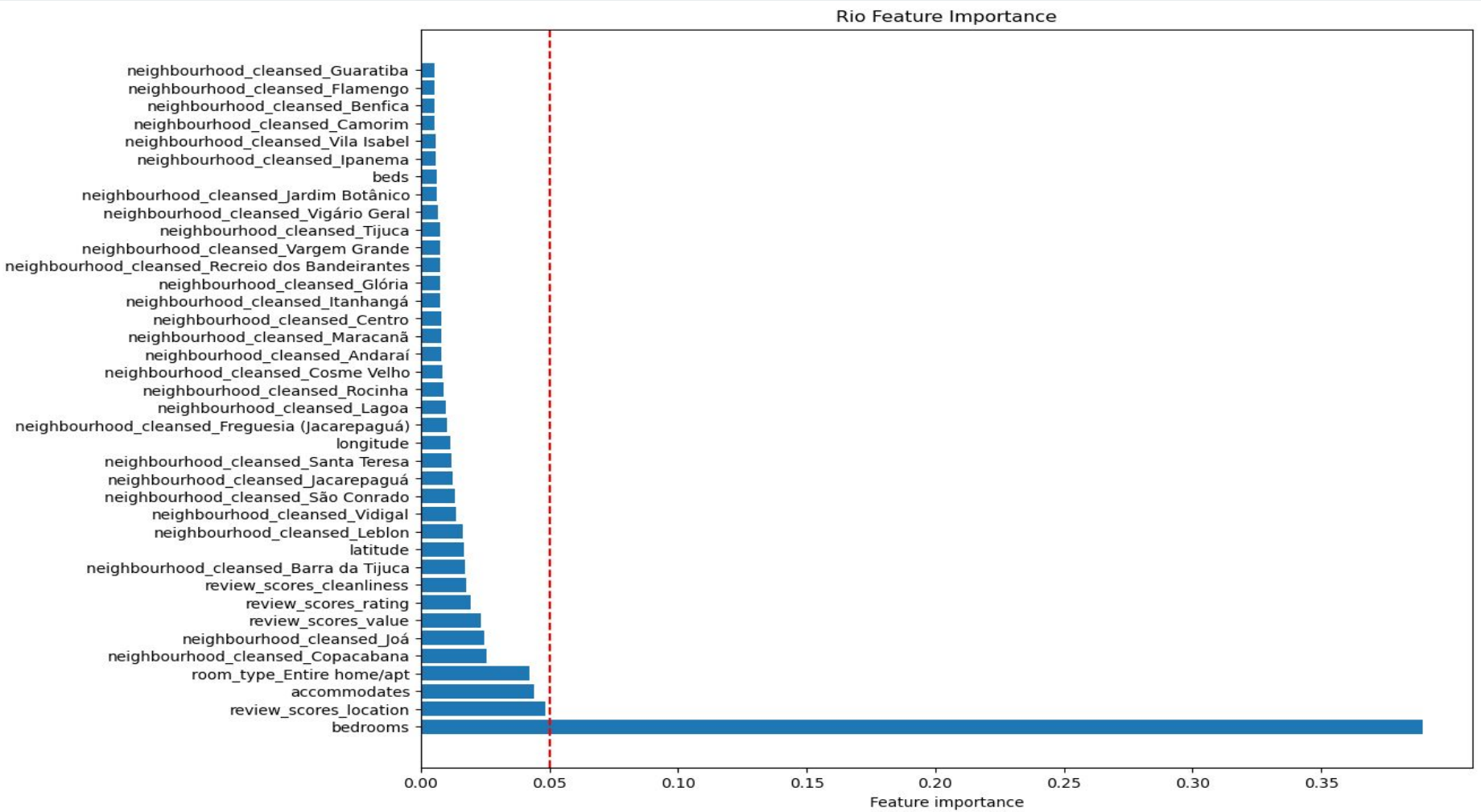
Feature Engineering - Rio

Model Evaluation Results

	R^2	rmse
JC	75.37	220.627080
Rio	55.68	912.965461
NYC	69.16	337.548466


New Jersey Feature Importance
room_type_Private room
neighbourhood_cleansed_Ward A (councilmember Denise Ridley)
neighbourhood_cleansed_Ward D (councilmember Michael Yun)
neighbourhood_cleansed_Ward C (councilmember Richard Boggiano)
neighbourhood_cleansed_Ward F (councilmember Jermaine D. Robinson)
review_scores_cleanliness
review_scores_value
review_scores_location
review_scores_rating
beds
neighbourhood_cleansed_Ward B (councilmember Mira Prinz-Arey)
latitude
neighbourhood_cleansed_Ward E (councilmember James Solomon)
bedrooms
longitude
accommodates
room_type_Entire home/apt
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
Feature importance


New York Feature Importance
neighbourhood_cleansed_Woodside
neighbourhood_cleansed_Sheepshead Bay
neighbourhood_cleansed_New Brighton
neighbourhood_cleansed_Elmhurst
neighbourhood_cleansed_East Village
neighbourhood_cleansed_Downtown Brooklyn
neighbourhood_cleansed_Rockaway Park
neighbourhood_cleansed_East Flatbush
neighbourhood_cleansed_Chelsea
neighbourhood_cleansed_Hollis
neighbourhood_cleansed_Murray Hill
neighbourhood_cleansed_Flushing
neighbourhood_cleansed_Upper West Side
neighbourhood_cleansed_SoHo
neighbourhood_cleansed_Battery Park City
neighbourhood_cleansed_Lower East Side
neighbourhood_cleansed_Little Italy
neighbourhood_cleansed_Grant City
neighbourhood_cleansed_Theater District
neighbourhood_cleansed_Gramercy
neighbourhood_cleansed_Flatiron District
neighbourhood_cleansed_Longwood
latitude
neighbourhood_cleansed_Kips Bay
neighbourhood_cleansed_Tribeca
neighbourhood_cleansed_Williamsburg
neighbourhood_cleansed_Midtown
neighbourhood_cleansed_Hell's Kitchen
neighbourhood_cleansed_Long Island City
neighbourhood_cleansed_Bayside
longitude
neighbourhood_cleansed_Financial District
bedrooms
accommodates
room_type_Entire home/apt
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Feature importance


Rio Feature Importance
neighbourhood_cleansed_Guaratiba
neighbourhood_cleansed_Flamengo
neighbourhood_cleansed_Benfica
neighbourhood_cleansed_Camorim
neighbourhood_cleansed_Vila Isabel
neighbourhood_cleansed_Ipanema
beds
neighbourhood_cleansed_Jardim Botânico
neighbourhood_cleansed_Vigário Geral
neighbourhood_cleansed_Tijuca
neighbourhood_cleansed_Vargem Grande
neighbourhood_cleansed_Recreio dos Bandeirantes
neighbourhood_cleansed_Glória
neighbourhood_cleansed_Itanhangá
neighbourhood_cleansed_Centro
neighbourhood_cleansed_Maracanã
neighbourhood_cleansed_Andaraí
neighbourhood_cleansed_Cosme Velho
neighbourhood_cleansed_Rocinha
neighbourhood_cleansed_Lagoa
neighbourhood_cleansed_Freguesia (Jacarepaguá)
longitude
neighbourhood_cleansed_Santa Teresa
neighbourhood_cleansed_Jacarepaguá
neighbourhood_cleansed_São Conrado
neighbourhood_cleansed_Vidigal
neighbourhood_cleansed_Leblon
latitude
neighbourhood_cleansed_Barra da Tijuca
review_scores_cleanliness
review_scores_rating
review_scores_value
neighbourhood_cleansed_Joá
neighbourhood_cleansed_Copacabana
room_type_Entire home/apt
accommodates
review_scores_location
bedrooms
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Feature importance

Model

XGBoosting Regression consistently outperform Random Forest Regression for the three case studies.

Grid Search hyperparameters optimization improved all XBoost models coefficient of determination. Jersey City root mean squared error (rmse) was marginally improved while Rio and New York City were marginally worse. These results increase our confidence that the models aren't overfitting the current data. Note Rio rmse is larger than the rest because it is quoted in Brazilian Reais.

Conclusion

In summary, the pricing model has demonstrated its efficacy by providing accurate predictions for both Jersey City and New York City. However, when applied to the Rio market, the model's predictive performance shows room for improvement. To harness the full potential of the model in the Rio context, it becomes imperative to undertake a more in-depth analysis, possibly involving additional data sources and localized variables unique to the region.

Looking ahead, the logical progression involves the integration of a Tableau platform, offering property owners in Jersey City and New York City an accessible tool to derive optimal listing prices for their properties. This step not only streamlines the pricing process but also empowers property owners with valuable insights based on the model's predictions, enhancing their decision-making and potentially fostering a more dynamic real estate market. The model's adaptability and potential for wider adoption remain promising, contingent upon continuous refinement and an expanded scope of study.