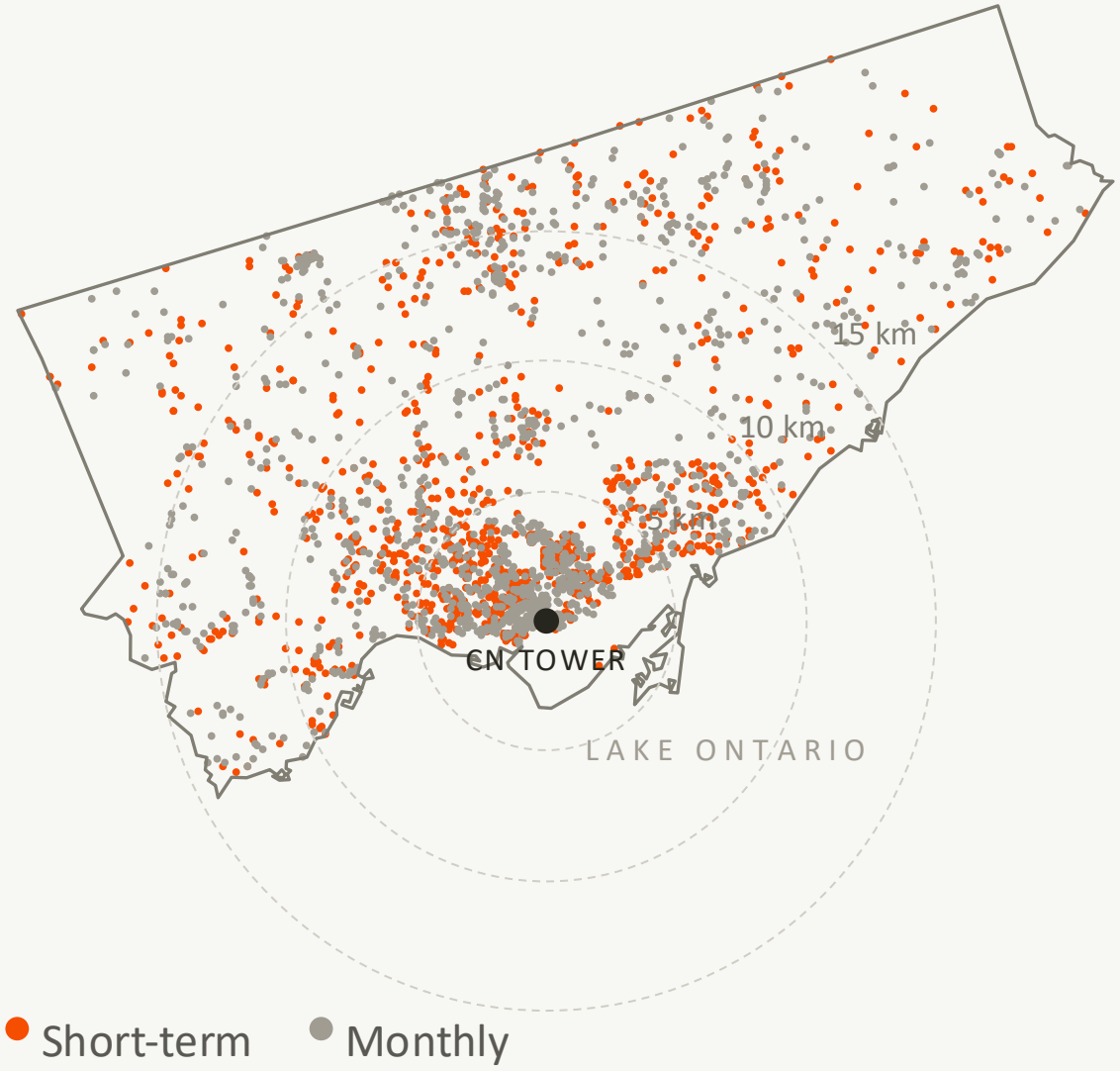
KOREA UNIVERSITY

BUSINESS ANALYTICS II · PROJECT I · TEAM 13

What Drives Airbnb Prices in Toronto?

Park Minjun



One city, two markets → two models

SHORT-TERM · 1–27 NIGHTS

$320

median per night · 7,759 listings

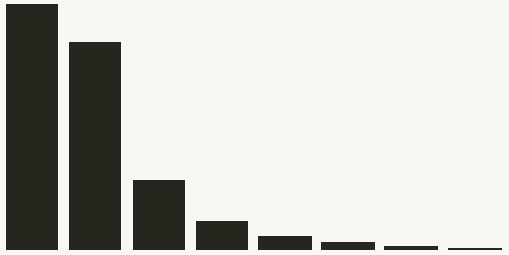


$0 $1,500+

MONTHLY · 28+ NIGHTS

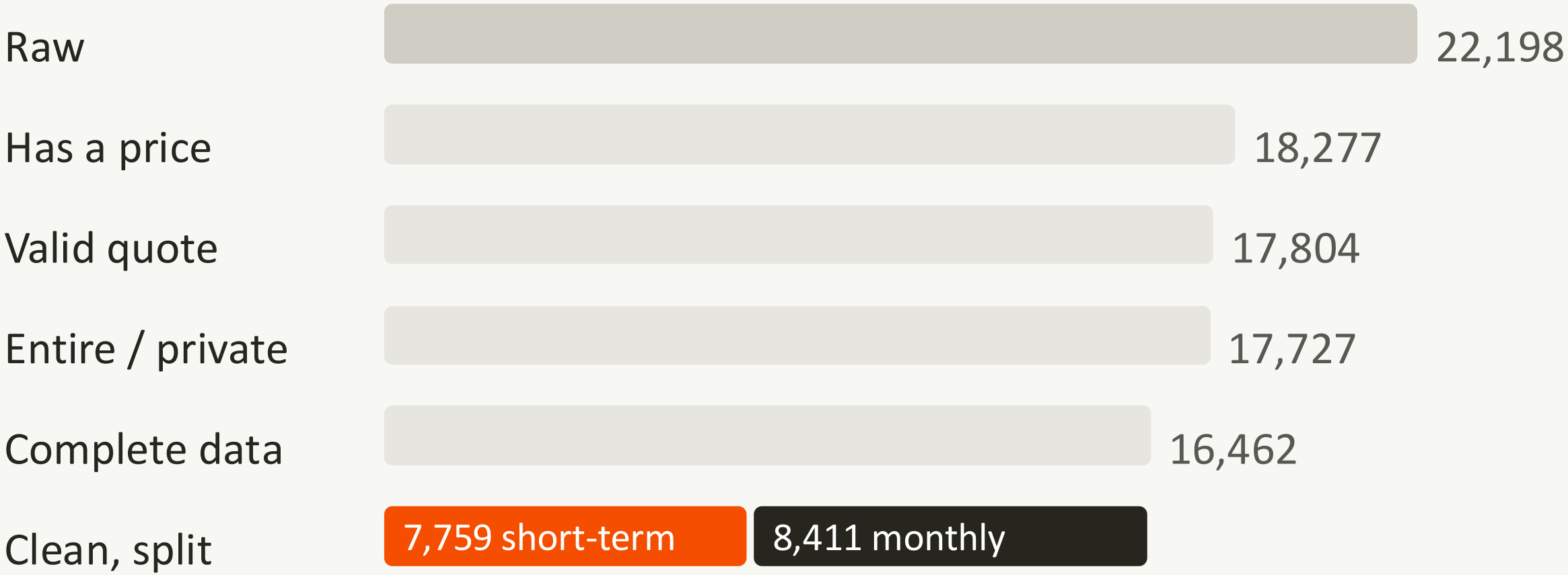
$118

median per night · 8,411 listings



$0 $1,500+

22,198 listings → 16,170 clean



Seven groups of predictors

4
Space
guests · beds · baths

2
Privacy
room · shared bath

7
Location
distance · region

10
Amenities
count + 9 flags

4
Reputation
rating · superhost

4
Host
years · listings

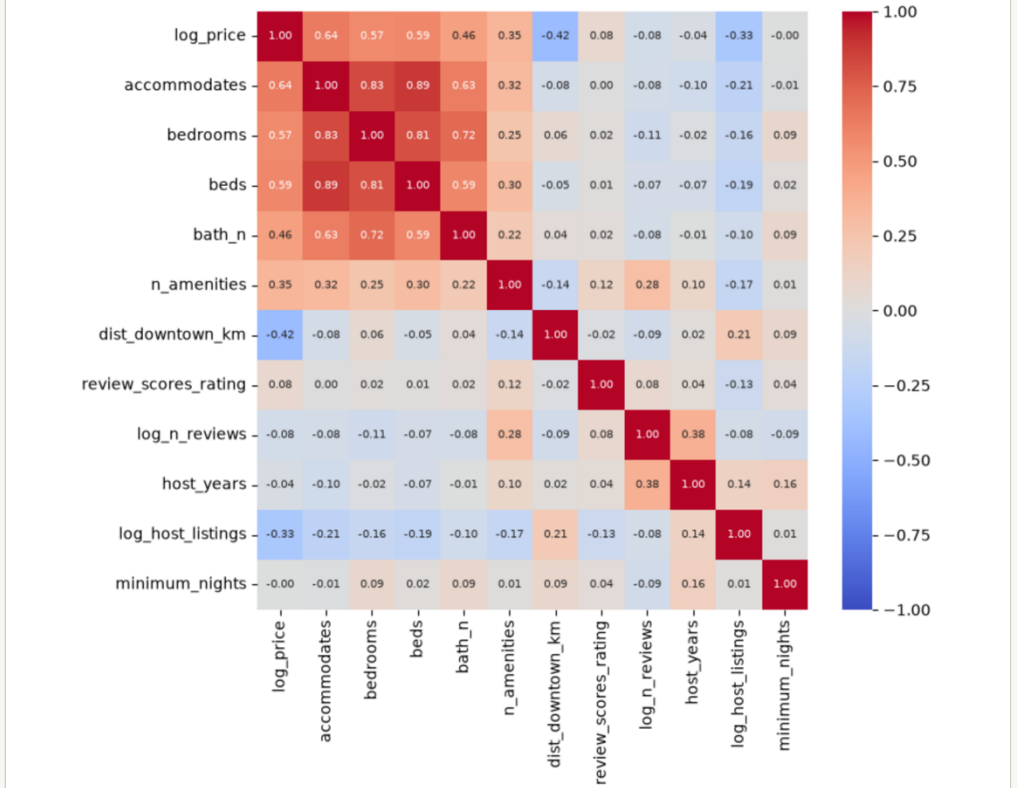
6
Property
type · nights · licence

y
log(price)
effects = % change

Size variables move together

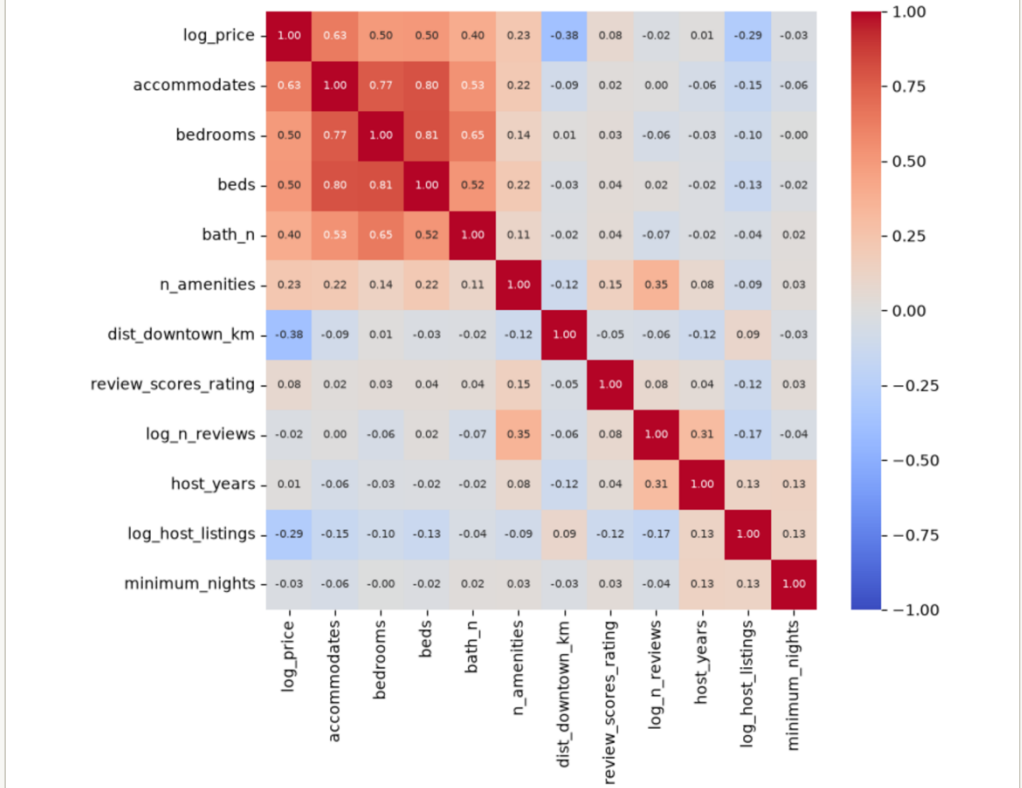
SHORT-TERM

```
[9]: num_cols = ["log_price", "accommodates", "bedrooms", "beds", "bath_n", "n_amenities", "dist_downtown_km",
                 "review_scores_rating", "log_n_reviews", "host_years", "log_host_listings", "minimum_nights"]
     corrs = short[num_cols].corr()
     …
```



MONTHLY

```
[27]: num_cols = ["log_price", "accommodates", "bedrooms", "beds", "bath_n", "n_amenities", "dist_downtown_km",
                  "review_scores_rating", "log_n_reviews", "host_years", "log_host_listings", "minimum_nights"]
      corrs = monthly[num_cols].corr()
      …
```



Five checks to the final model

1. **OLS** — all predictors
2. **VIF** — max 6.7 ✓
3. **BP test** — → robust SEs
4. **Residuals** — linear ✓
5. **Distance²** — curve ✓

Final: log(price) + distance² · HC3 robust SEs · one model per market

VIF below 10 · variance not constant

SHORT-TERM

```
[14]: vif_values = [float(vif(x_s_c.values, i)) for i in range(1, x_s_c.shape[1])]
      vif_labels = x_s_c.columns[1:]
      vif_table_s = pd.DataFrame({"Variables": vif_labels, "VIF Factor": vif_values})
      …
```

[14]:

	Variables	VIF Factor
15	dist_downtown_km	6.696293
0	accommodates	6.387904
2	beds	5.144601
1	bedrooms	4.936762
34	region_Scarborough	3.666906
32	region_North_York	3.513616
18	log_n_reviews	2.719790
25	room_type_Private_room	2.639923
27	prop_group_house	2.393366
3	bath_n	2.194029

```
[15]: residual_s = reg1_s.resid
      bp_test = sm.stats.het_breuschpagan(residual_s, x_s_c)
      bp_test2 = [float(b) for b in bp_test]
      …
```

[15]:

	Values
LM Statistics	2.978406e+02
LM p-value	1.978921e-43
F statistics	8.808369e+00
F p-value	2.039169e-44

MONTHLY

```
[32]: vif_values = [float(vif(x_m_c.values, i)) for i in range(1, x_m_c.shape[1])]
      vif_labels = x_m_c.columns[1:]
      vif_table_m = pd.DataFrame({"Variables": vif_labels, "VIF Factor": vif_values})
      …
```

[32]:

	Variables	VIF Factor
15	dist_downtown_km	6.430334
35	region_Scarborough	4.303413
1	bedrooms	4.191167
33	region_North_York	4.163284
0	accommodates	3.911557
2	beds	3.836303
18	log_n_reviews	2.938601
26	room_type_Private_room	2.761876
5	n_amenities	2.377418
17	no_reviews	2.208243

```
[33]: residual_m = reg1_m.resid
      bp_test = sm.stats.het_breuschpagan(residual_m, x_m_c)
      bp_test2 = [float(b) for b in bp_test]
      …
```

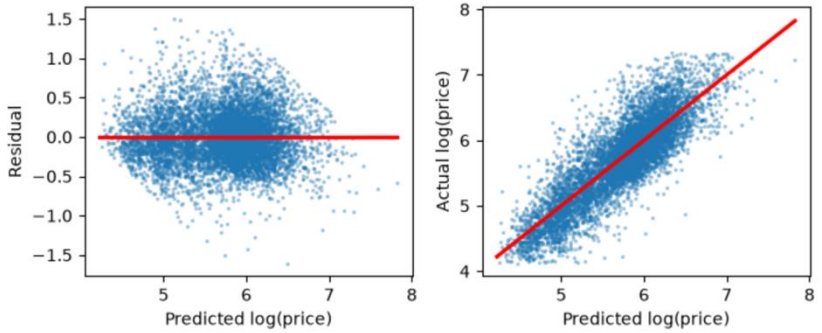
[33]:

	Values
LM Statistics	2.526803e+02
LM p-value	2.335977e-34
F statistics	7.204454e+00
F p-value	5.676156e-35

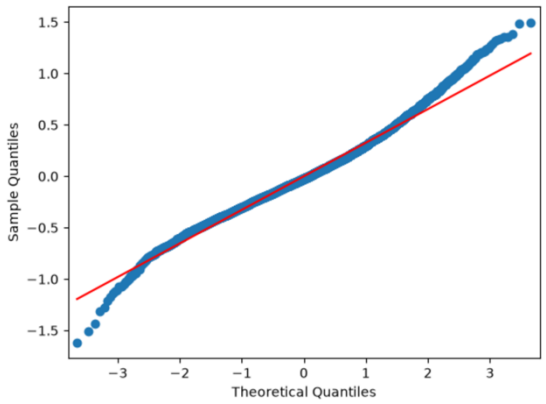
Residuals: no pattern, heavier tails

- SHORT-TERM

```
[17]: y_pred_s = reg2_s.predict()
      residual_s = reg2_s.resid
      plt.figure(figsize=(7,3))
      …
```

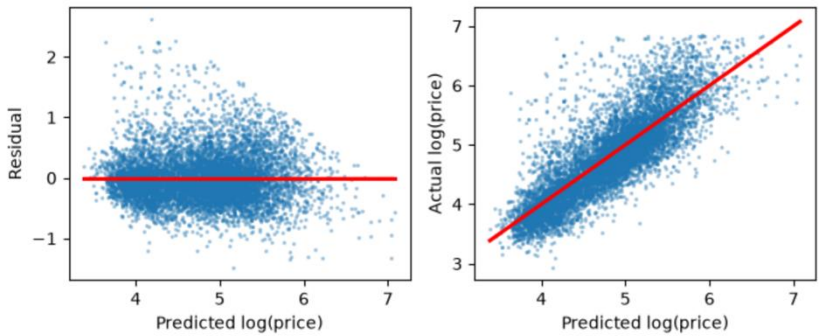



```
[18]: sm.qqplot(residual_s, line="s")
      plt.show()
```

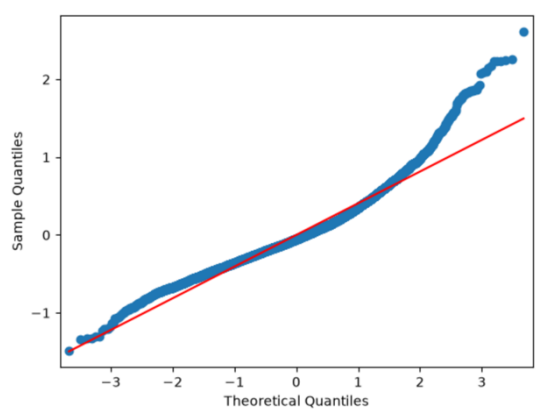


- MONTHLY

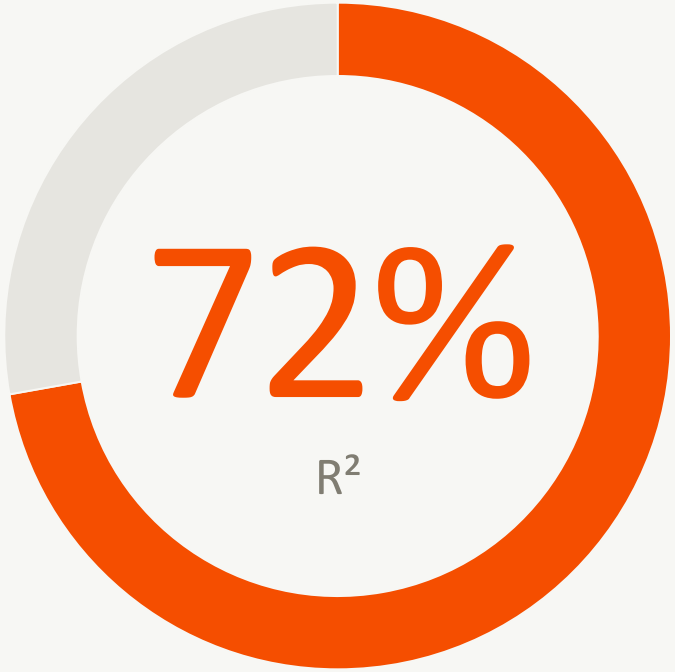
```
[35]: y_pred_m = reg2_m.predict()
      residual_m = reg2_m.resid
      plt.figure(figsize=(7,3))
      …
```




```
[36]: sm.qqplot(residual_m, line="s")
      plt.show()
```

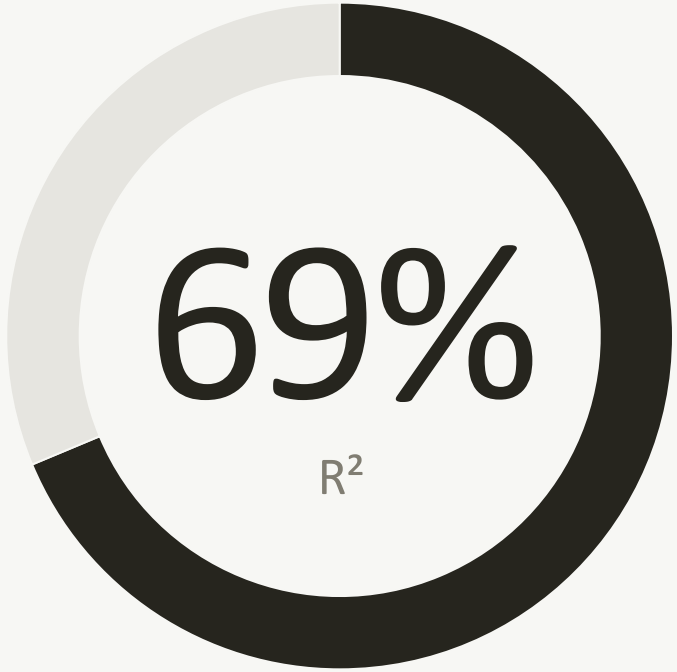


Our models explain ~70% of price



Short-term

n = 7,759



Monthly

n = 8,411

The final model, as statsmodels prints it

● SHORT-TERM

```
[19]: reg3_s = smf.ols("log_price ~ accommodates + bedrooms + … + np.power(dist_downtown_km, 2)",
                       data=short).fit(cov_type="HC3")
      print(reg3_s.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              log_price   R-squared:                       0.722
Model:                            OLS   Adj. R-squared:                  0.720
Method:                 Least Squares   F-statistic:                     581.3
Date:                Mon, 12 Oct 2026   Prob (F-statistic):               0.00
Time:                        09:10:50   Log-Likelihood:                -2328.5
No. Observations:                7759   AIC:                             4731.
Df Residuals:                    7722   BIC:                             4988.
Df Model:                          36
Covariance Type:                  HC3
================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept        4.8991      0.108     45.169      0.000       4.687       5.112
accommodates     0.0389      0.005      7.627      0.000       0.029       0.049
bedrooms         0.1489      0.013     11.323      0.000       0.123       0.175
beds            -0.0002      0.008     -0.020      0.984      -0.015       0.015
bath_n           0.1014      0.010      9.705      0.000       0.081       0.122
bath_shared     -0.3305      0.015    -21.671      0.000      -0.360      -0.301
n_amenities      0.0016      0.000      3.586      0.000       0.001       0.002
…
```

● MONTHLY

0.687
R^2

0.686
Adj. R^2

8,411
observations

HC3
covariance

PART 2 · FINDINGS

What drives the price?

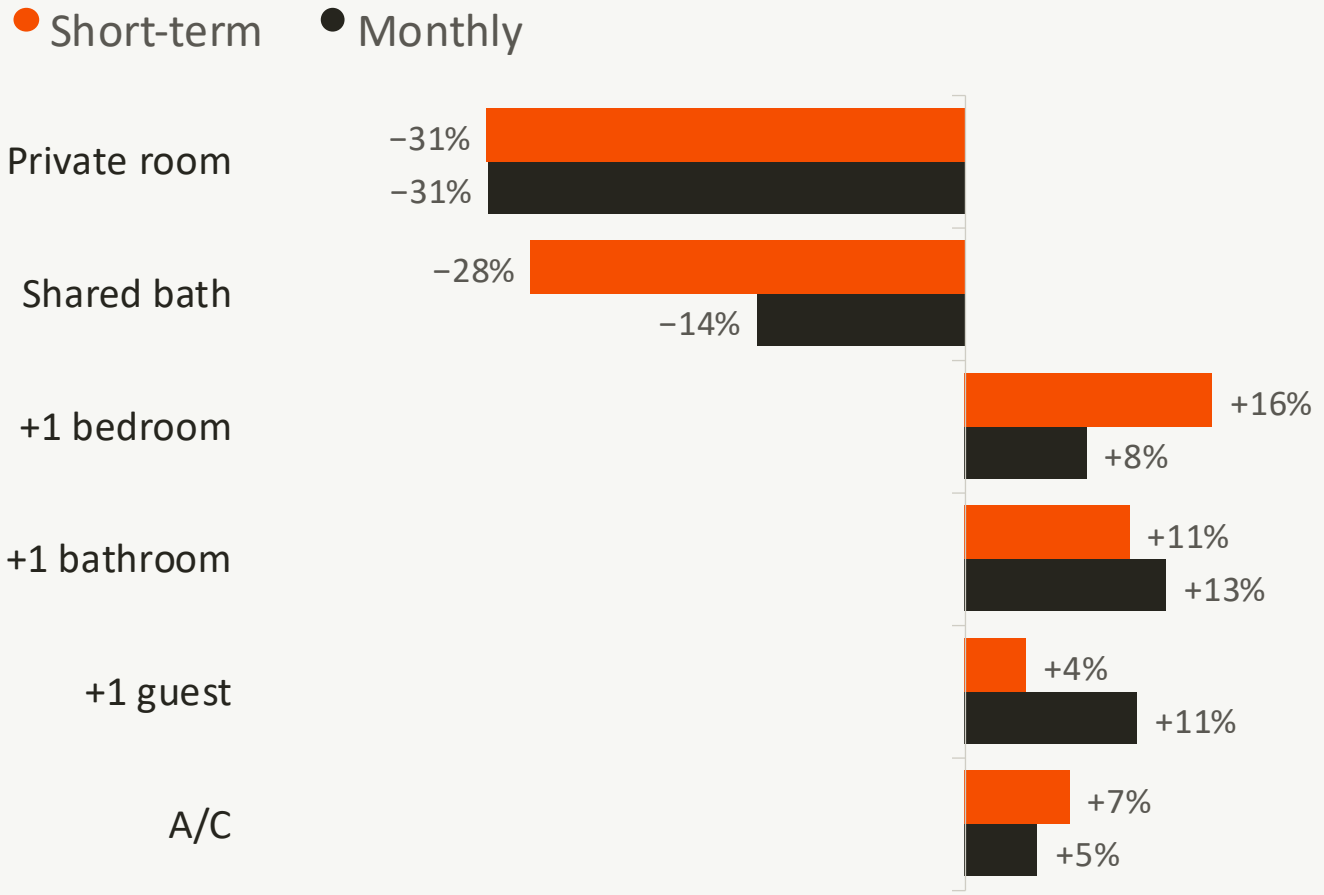
1 Common drivers 2 Market differences 3 Why

Privacy is the biggest price driver

−31%

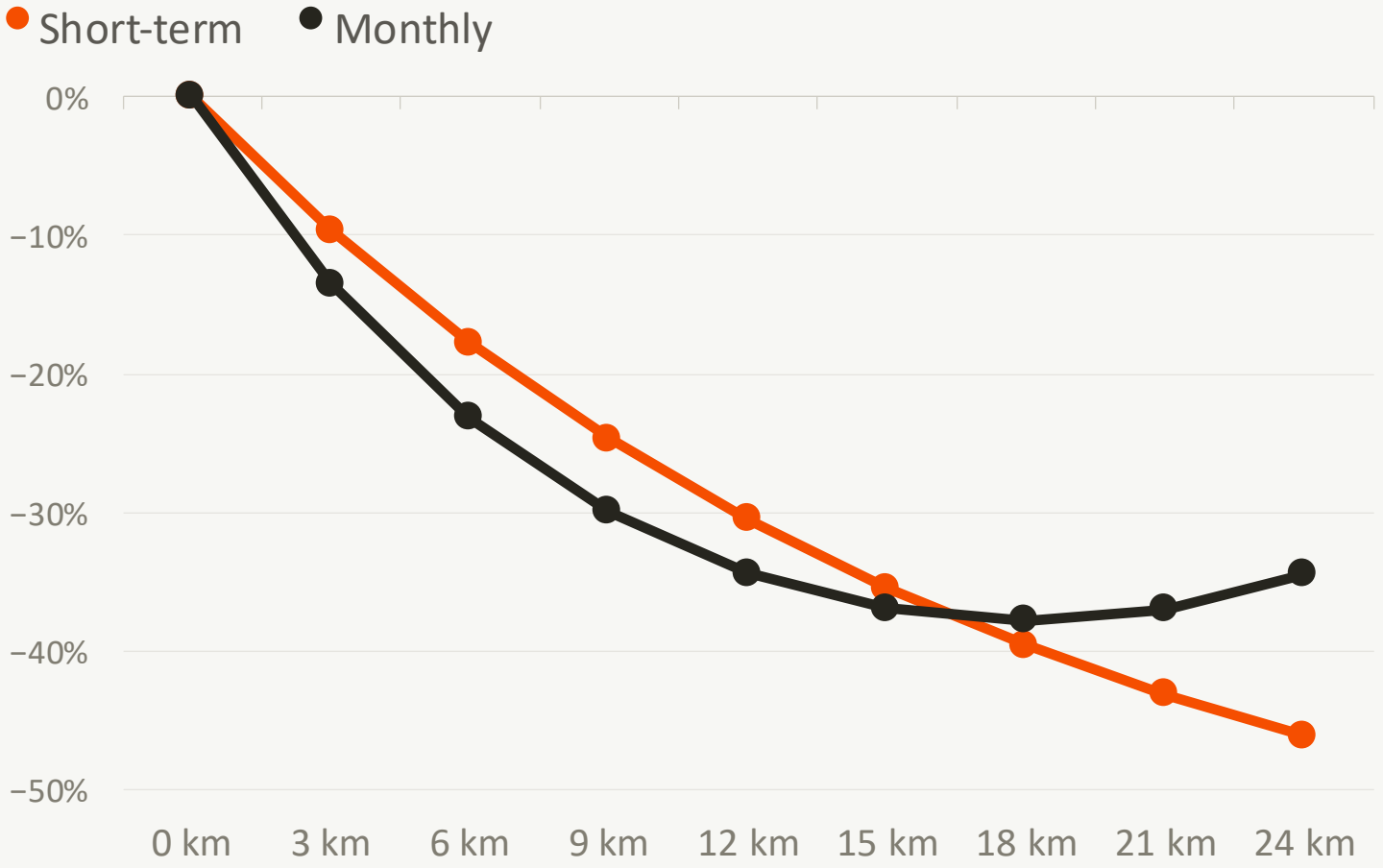
private room vs entire home

same in both markets



FINDING 1 · LOCATION

Price falls fastest near downtown



−15%

at 5 km

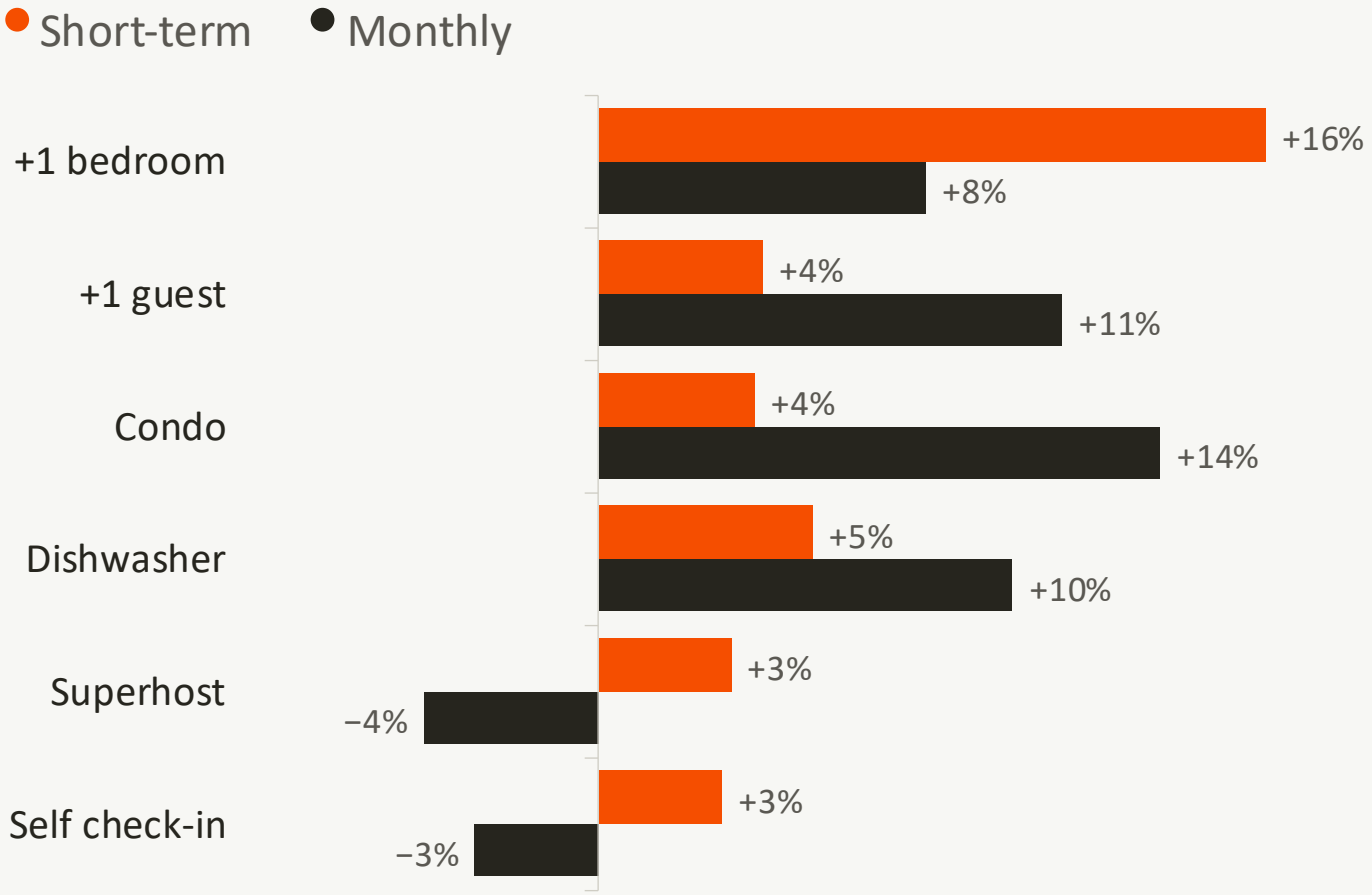
short-term

18 km

monthly low point

≈ −38%

Same feature, different value



Short stays pay for
bedroom & trust

Monthly stays pay for
space & comfort

FINDING 3 · WHY

Different hosts, different guests

SHORT-TERM

67%

solo hosts

Guests pay for trust

Superhost + 3% · Self check-in + 3%

MONTHLY

38%

solo hosts — operators dominate

Guests pay for comfort

Dishwasher +10% · Licence +27%

Four moves for hosts

−31%

Rent the whole place

private room vs entire home

2×

Match home to market

bedrooms pay short · · guests pay monthly

+27%

Get licensed

monthly · only 28% have one

+10%

Add a dishwasher

monthly · only 45% have one

Two markets, two pricing playbooks

- **Short-term** — Reviews · Superhost · Self check-in · Bedrooms
- **Monthly** — Licence · Dishwasher · Parking · Guest capacity

One price guide for both markets would misprice both.

What we cannot claim

01

Association,
not causation

02

Asking price,
not revenue

03

One snapshot,
June 2026

04

~30%
unexplained

Short-term model, with real coefficients

● SHORT-TERM

$$\log(\text{price}) = 4.899$$

$$- 0.371 \times \text{Private room}$$
$$+ 0.149 \times \text{Bedrooms}$$
$$+ 0.039 \times \text{Guests}$$
$$+ 0.0004 \times \text{Distance}^2$$
$$+ 0.031 \times \text{Superhost}$$
$$- 0.330 \times \text{Shared bath}$$
$$+ 0.101 \times \text{Bathrooms}$$
$$- 0.035 \times \text{Distance}$$
$$+ 0.066 \times \text{A/C}$$
$$+ 0.037 \times \text{Condo}$$

+ 26 more terms (amenities · reviews · host · region · property type)

Adj. R^2 0.720 · n = 7,759 · HC3 robust SEs · distance = km to CN Tower · % change = e^β − 1

Monthly model, with real coefficients

● MONTHLY

$$\log(\text{price}) = 4.870$$

$$- 0.370 \times \text{Private room} \qquad + 0.235 \times \text{Licence}$$
$$- 0.145 \times \text{Shared bath} \qquad + 0.127 \times \text{Condo}$$
$$+ 0.123 \times \text{Bathrooms} \qquad + 0.106 \times \text{Guests}$$
$$+ 0.076 \times \text{Bedrooms} \qquad + 0.095 \times \text{Dishwasher}$$
$$- 0.053 \times \text{Distance} \qquad + 0.0015 \times \text{Distance}^2$$

+ 27 more terms (amenities · reviews · host · region · property type)

Adj. R^2 0.686 · n = 8,411 · HC3 robust SEs · distance = km to CN Tower · % change = e^β − 1

Thank you

Questions?

- Privacy first
- Two markets, two values
- Act by segment

Park Minjun

