Urgency messaging

Exploratory analyses and next steps

October 9 2020

Moctar Aboubacar

Objectives and methodology

The **assignment** is to “understand the movement of the price as the day approaches the check-in date” with the **goal** of supporting price-based urgency messaging implementation

Objectives

1 Build Framework to analyze bookings

- What assumptions underly price-based urgency messaging?
- Which bookings are propitious to what level of urgency messaging?

3 Analyze data according to framework

- How can our framework inform where to focus urgency messaging?

4 Design testing for urgency messaging

- How can we test our framework's assumptions and move forward with an urgency messaging test plan?

Methodology

Timeline from October 2020 to early-Nov 2020

Methods

- Desk research on price-based urgency messaging
- Exploratory framework based on desk research
- Data exploration and modelling

Next steps

- Meeting to present and discuss first results (10/9)
- Additional data collection and re-analysis
- Update informal meeting with PO (‘are we on the right track?’)
- A/B testing urgency messages according to framework
- Results presentation and decisions (‘enough net value created to integrate as a feature?’)

Findings and recommendations

Main findings

- There is a slightly **negative relationship** between booking-to-check-in time and ADR at property level
- **ADR decrease steepens** as booking-to-check-in days near 0
- **23% of all bookings** are made the day before or day-of check-in
- **1 in 4 bookings** were made by 'deal-seeking' customers

Recommendations

1. **Understand booking decision drivers**
   - Explore additional datasets/variables
   - Update framework assumptions

3. **Test urgency messaging**
   - Type
   - Level

4. **Calculate net benefit of Urgency Messaging**
   - Cost-Benefit analysis.
   - 'Are gains from this messaging hurting us elsewhere?'

Table of contents

Price-based urgency messaging: definitions and assumptions

Definition

- Applies to a booking
- A booking is composed of:
  - Person
  - Property/ room/ date/ duration combination
- Types of urgency messaging
  - Inventory-based
  - Price-based

Assumptions

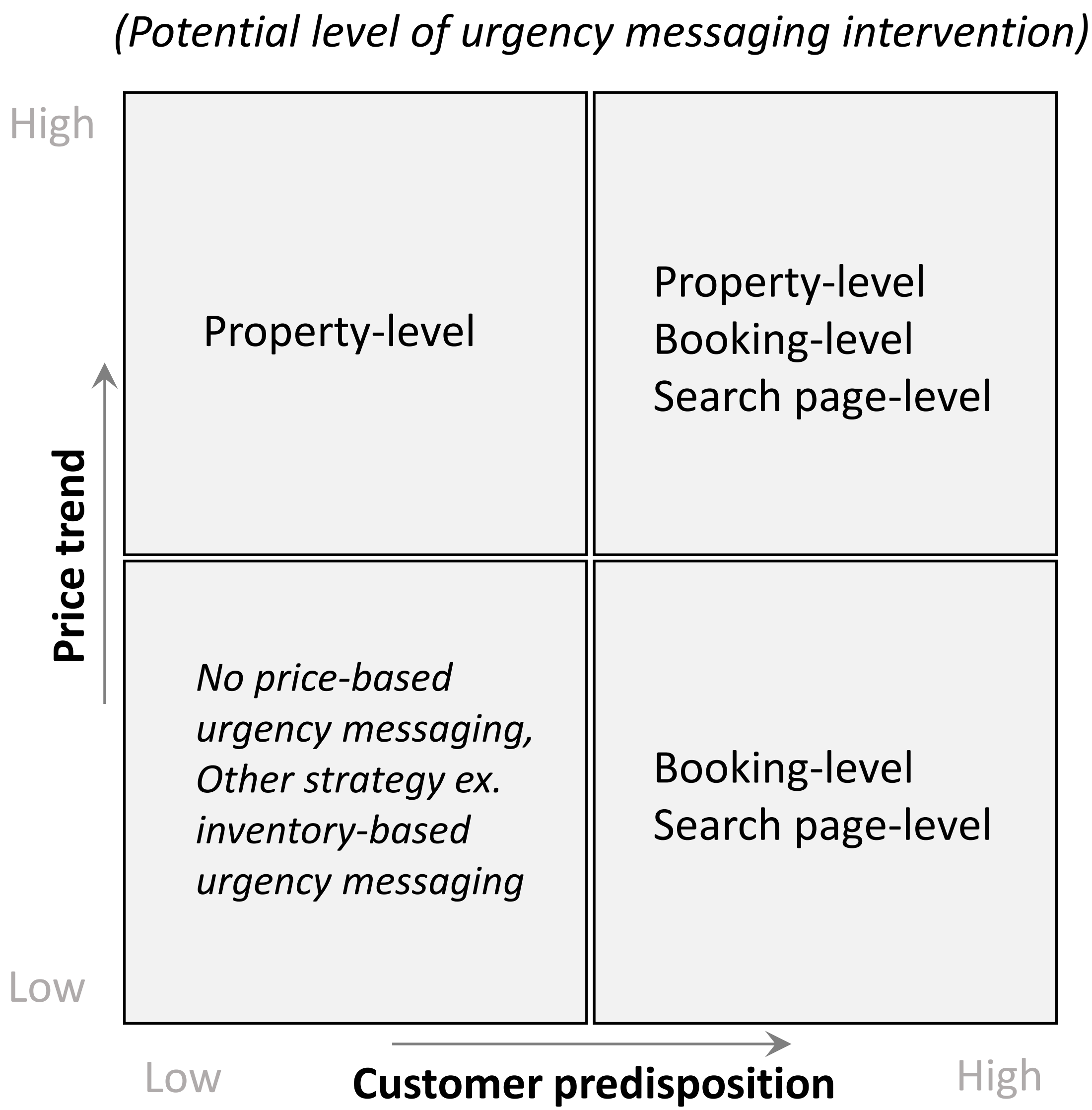
- **Truth**: Used when there is actual probable rise in price
- **Continuous use**: effect is not dampened by repeated use
- **Time-invariant effects**: far-away price rises can be successfully messaged
- **Versatile**: can be applied to different stages of sales process/funnel

Price-based urgency messaging can be introduced along two axes: the predisposition of the customer, and the trend in price

	1 **Booking Price trend** *Will prices actually rise?*		2 **Customer Predisposition** *Does the customer care?*	
Key question	Is there seasonal price fluctuation?	Is there a rising price trend?	Is the customer a 'deal-seeker'?	Is the customer making a 'late' booking?
Level of relevance	Property/room views		Home, Search, Booking pages	

This framework yields an initial breakdown of our bookings data into useful, testable segments, but <u>does not address the efficacy</u> of different urgency messaging strategies.

A simple framework to identify bookings which may benefit from price-based urgency messaging, and at what level

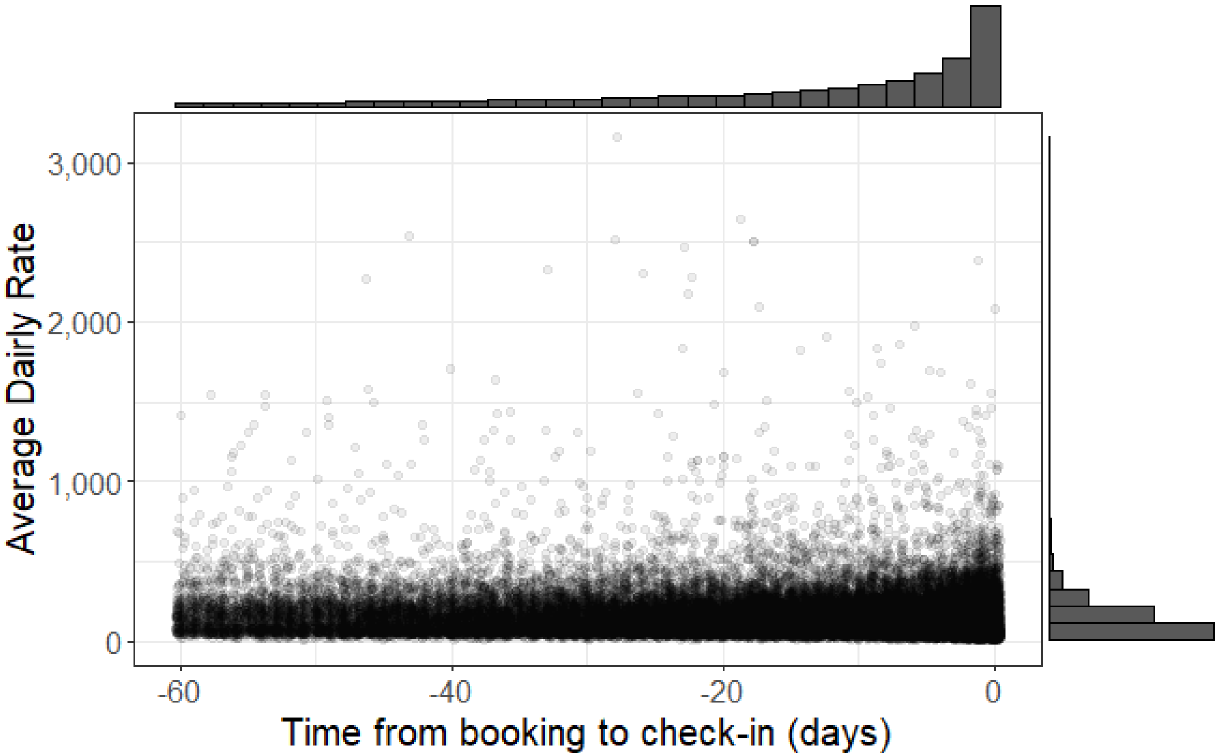


Data constraints mean that imperfect methods must be chosen to address each component of the framework

	1 **Booking Price trend** *Will prices actually rise?*		2 **Customer Predisposition** *Does the customer care?*	
Key question	Is there seasonal price fluctuation?	Is there a rising price trend?	Is the customer a 'deal-seeker'?	Is the customer making a 'late' booking?
Level of relevance	Property/room views		Home, Search, Booking pages	
Ideal method	Time series analysis at room level: model room ADR as booking/check-in gap narrows		Customer clustering (ex. Hierarchical)	Calculate exact time from booking to check-in
Method chosen	[No good alternative]	Mixed effects/ random slope model at the property level	Predicted vs. actual ADR controlling for relevant factors	Days from booking to check-in

Table of contents

ADR goes down slightly as booking-to-check-in days decrease



Most people book cheaper

- 15.7% of all bookings are under an average of 50 dollars a night and 44.6% of all bookings are under 100$ a night

People tend to book later

- 22.9% bookings made one day in advance or on the day-of
- Only 17.3% of bookings made more than 1 month out

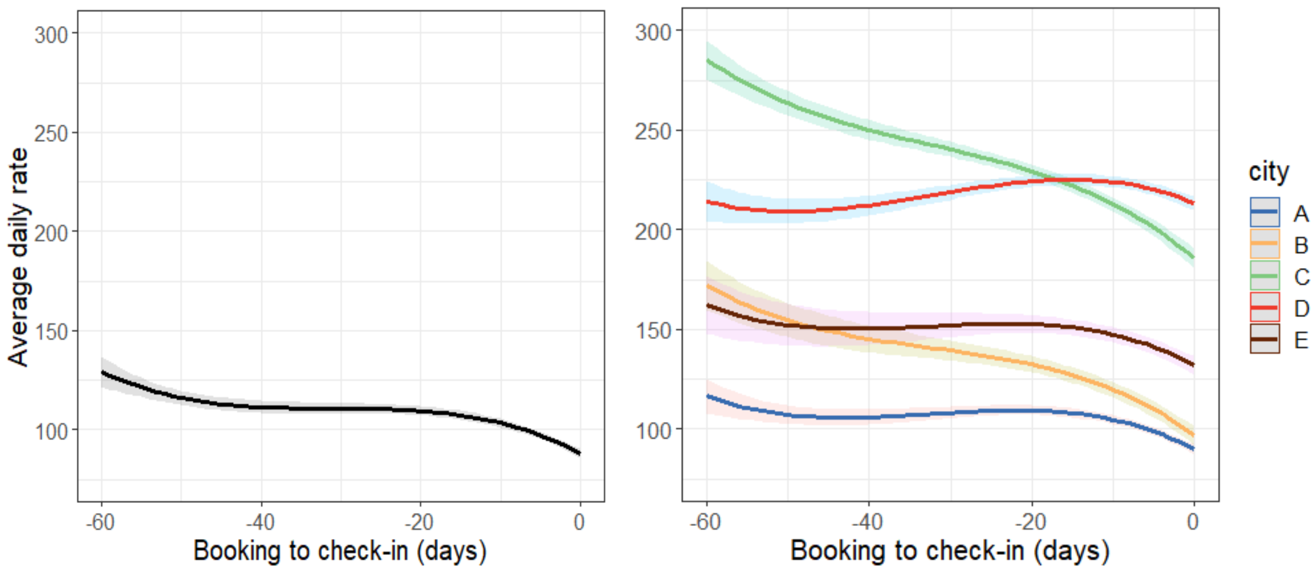
A weak negative relationship

- On average, a 1-day nearer 'B2C' is associated with a 99 cent lower ADR

Almost 1 in 5 bookings are both cheaper (under 90$) and later (5 days and sooner away)

The decline in ADR is increasing under ~20 days to check-in date, at different levels across cities

Polynomial regression of ADR on booking to check-in | controlling for city



Note: 95% confidence interval displayed on estimates

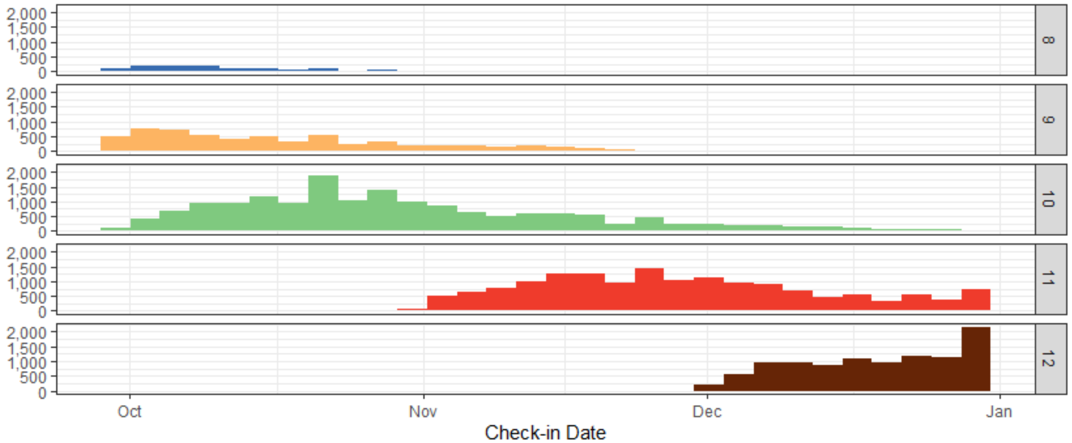
- Disparity of **bookings**: City A accounts for 46% of all bookings
- Disparity of **level**: average price levels in cities C and D are higher than A, B, E
- Disparity of **change**: price levels drop faster overall in cities B and C as B2C declines

Variation in ADR between cities explains only 16.25% of total variation in ADR

Bookings made in Sept-Oct explain some of the high ADR/high 'booking to check-in gap' observed

Month	Median B-to-C days	Median ADR	% Booking City C, D	% Booking City A
August	50	114.9	46	36
Sept	26	133.1	42	41
Oct	8	113.4	35	45
Nov	8	108.4	34	47
Dec	3	109.2	30	48

Histogram of booking check-in date by booking month



- August and September bookings are **higher-value**, made **further out**
- They are more likely to be in **City C and D** (more expensive overall)
- However they are not more likely to be in a different **type of property**

Some higher-ADR properties (Resorts) are still booked earlier on average

Booking value

Hotel bookings are higher-value…

Apartment bookings are lower-value…

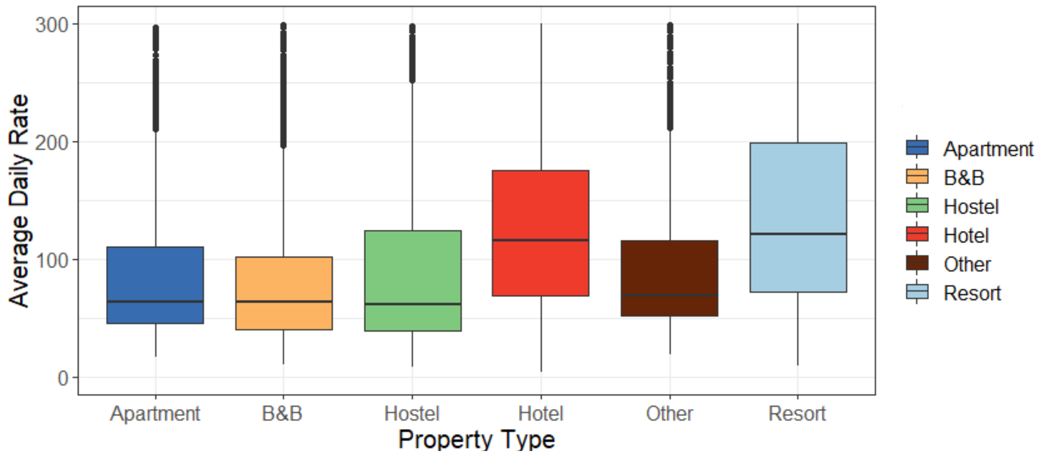
Resort bookings are higher-value…

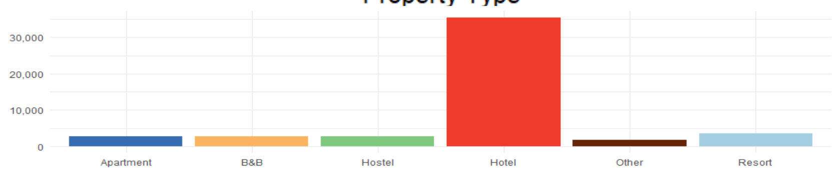
Booking timeline

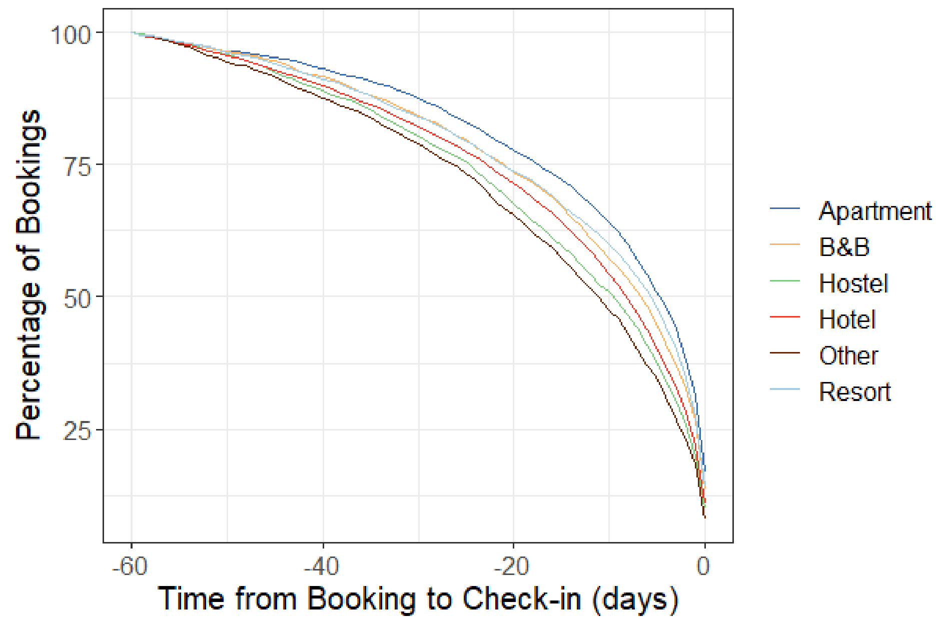
…and are booked earlier

…and are booked later than other types

…and are booked **later than hotels**





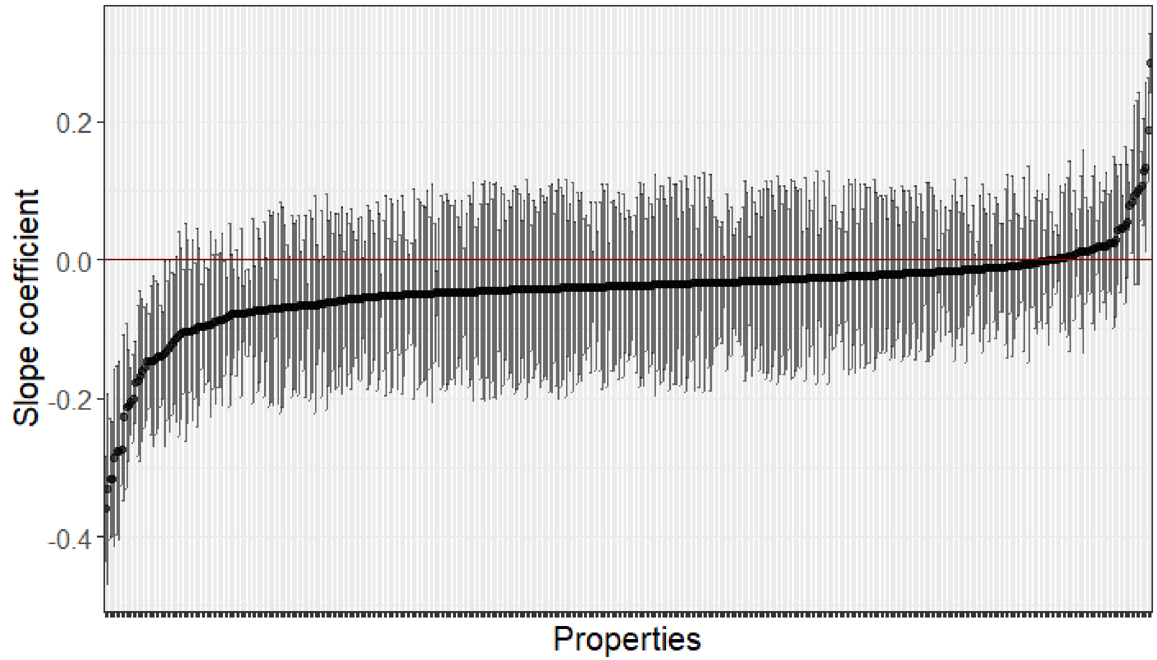


Note: boxplot outliers were excluded above ADR = 300 (total 10% of the sample)

A property level analysis of ADR on 'booking to check-in gap' must account for variation seen in city, booking type and booking date[1]

- Overall, a weak/null trend in ADR over booking-to-check-in by property
- 13% of slopes (61 properties)[1] are statistically significant, mainly negative
- The slopes from this model are adopted as a measure of 'booking price trend'

Mixed-effects model booking-to-check-in: slope coefficients



Why look at the property level?
- Practical reasons (data)
- Price differences between properties account for over 77% of the variation in ADR

Confounding variables missed here
- Patterns in differentiated room quality / price over booking to check-in times

[1]Only properties with 10 or more bookings were included in this process (455 total)

Table of contents

Later and last-minute bookings are different from bookings made more in advance

Overall[1]

Later bookings are more likely to be…

…in **chain hotels** (0.5 days more buffer on average than non-chains)

…made on **Friday/Saturday**

…for **fewer days of stay** (2.7 days more buffer for each additional stay day).

A 100$ more expensive booking is expected to have 1 day's additional buffer on average.

Last minute bookings[2]

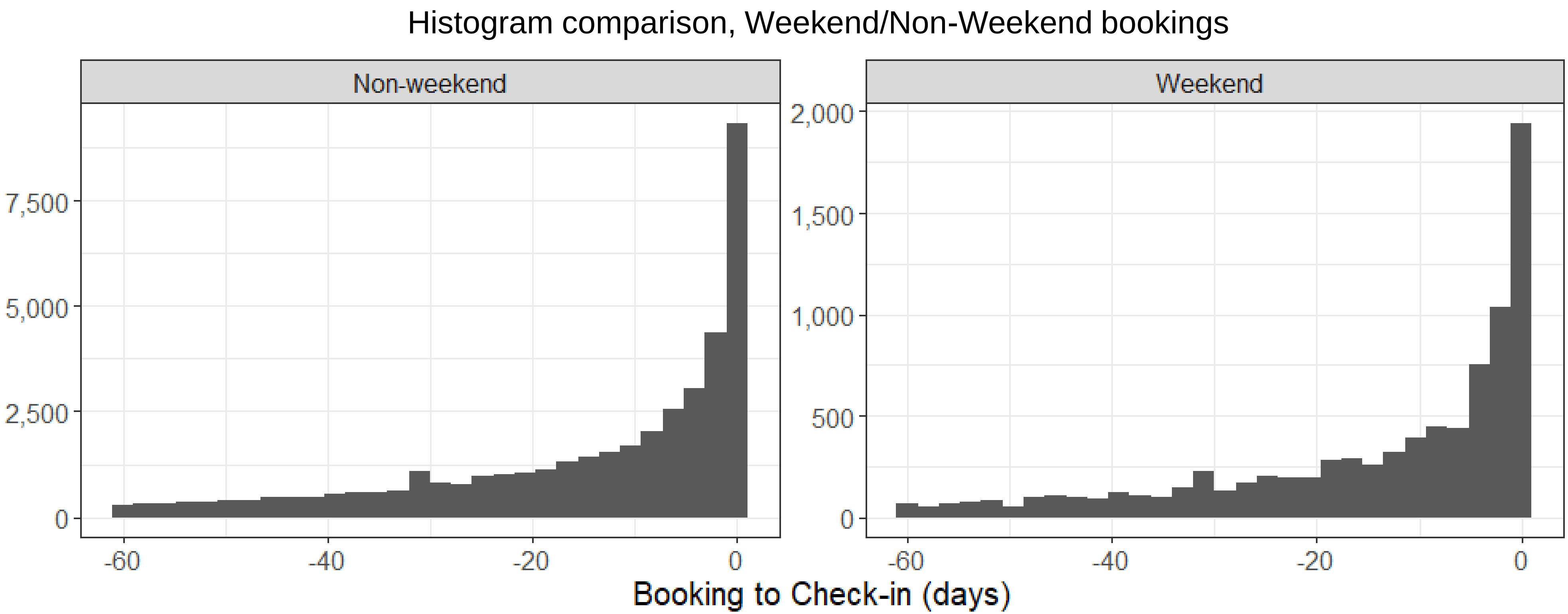
	Area	Last minute bookings[3]	2-weeks+ advance bookings
	Median ADR	88.6	132
% bookings	City A	51	41
% bookings	City B	13	9
% bookings	City C	6	20
% bookings	City D	16	22
% bookings	City E	13	7
% properties	Hotel	69	73
% properties	Hostel	5	6
% properties	Resort	9	7
% properties	Serviced Apt	8	5
% properties	B&B	7	5
% stay days	1	67	40
% stay days	2	23	34
% stay days	3	10	26

[1] In a linear model controlling for city, property type, star rating, hotel chain, booking date, days stayed and day of the week booking. See annex for model output.

[2] All row-wise group differences statistically significant (tested with 2-proportion Z-test)

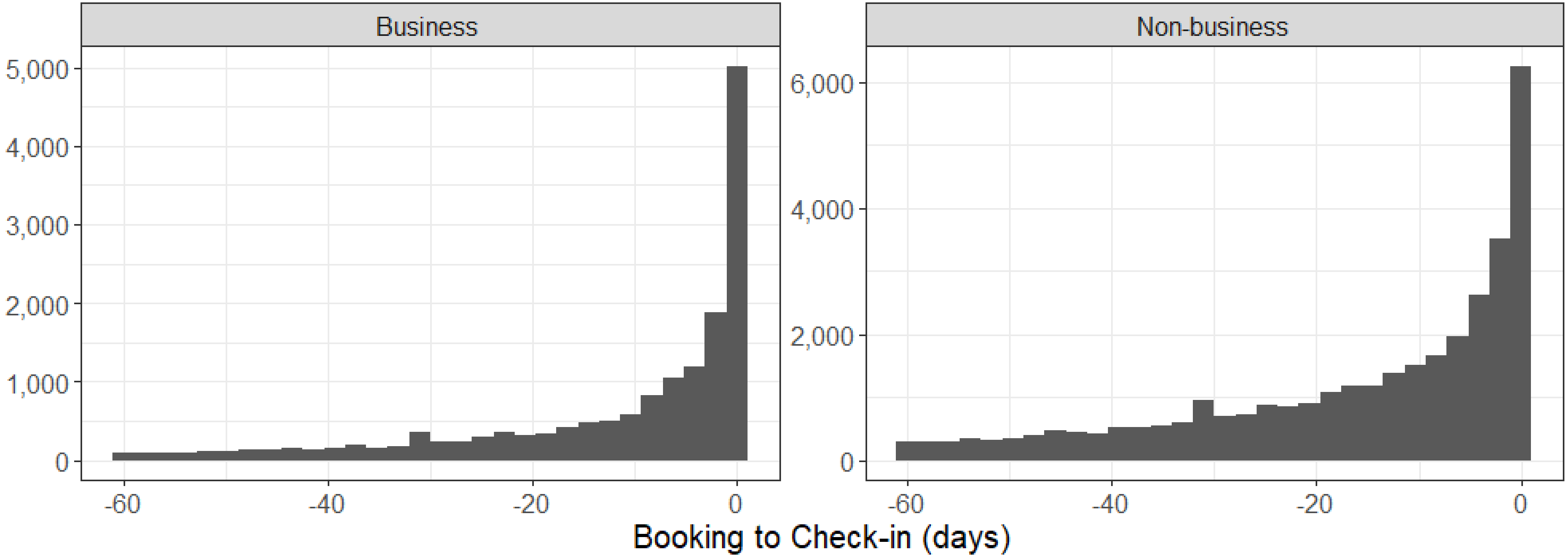
[3] Calculated as bookings made 1 day before check-in or day-of check-in

Weekend trips (Saturday check-in and Sat/Sun check-out) are not more spontaneous than other bookings



Business travelers[1] are more likely to make last-minute bookings

Histogram comparison, Business/Non-Business bookings



- Business bookings are made on average **4.3 days closer** to check-in time
- Business bookings are lower-value: accounting for city differences, **23 dollars less** per booking[2]

[1] Business trips are defined here as trips with check-in day between Monday and Friday, which last a single day.

[2] See annex for model output

Creating a measure of the lateness of booking, using the days between booking and check-in

Those who book late

Definition

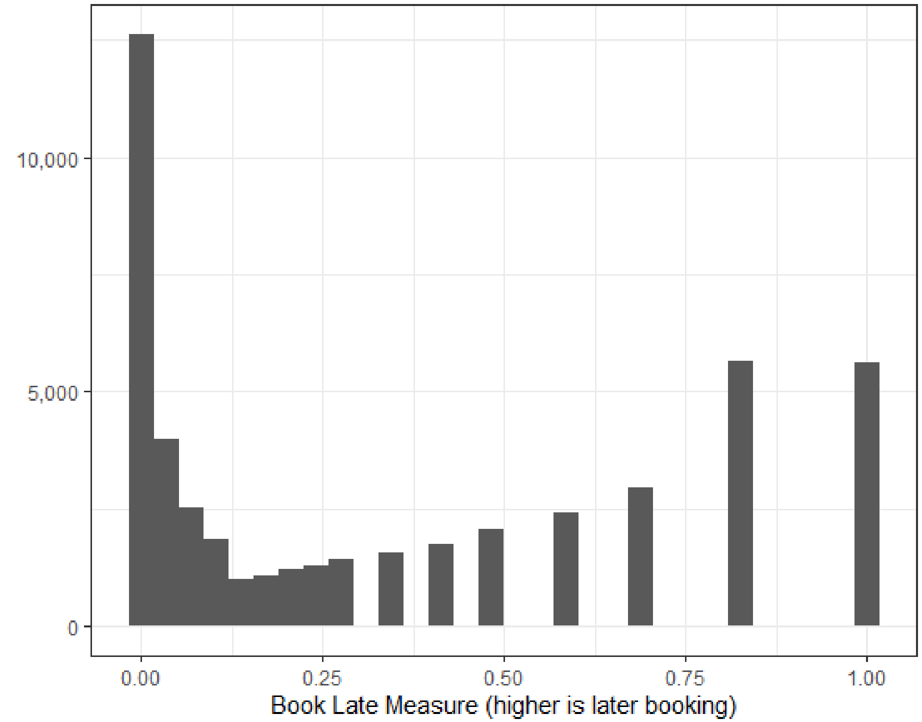
When bookings are done with little time to spare from start of stay (penalizing bookings made farther out less and less)

Measurement

$$\text{Late booking} = \lambda^k$$

Where k is the booking-to-check-in days and λ is a constant (here $\lambda = 1.2$ is chosen)

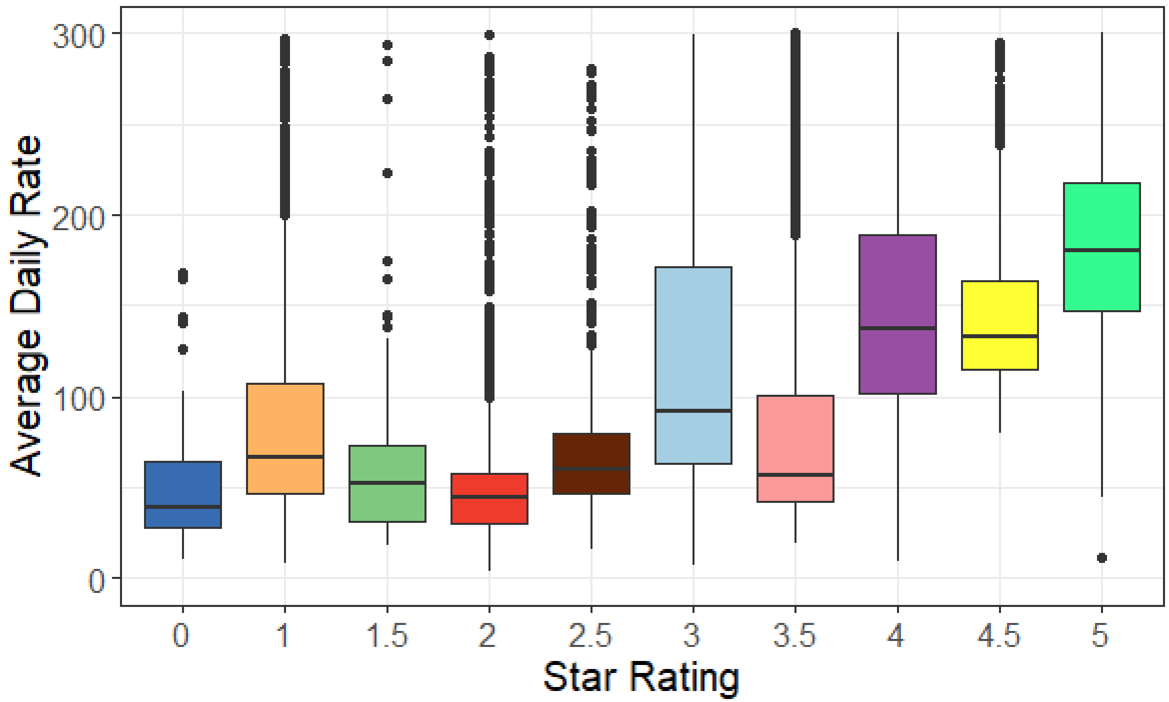
Histogram of late booking measure



Some low-starred bookings are not 'worth it' –the same price yields similar high-and low-star bookings

- When controlling for city, the average **2-starred booking** is statistically indistinguishable from the average **0-star booking** in ADR

- We could consider bookings at good cost-to-star ratios as done by 'deal-seekers'.



Note: boxplot outliers were excluded above ADR = 300 (total 10% of the sample)

Creating a measure of the deal-seeking bookings, using the residuals on a regression of ADR on city and star rating

Deal seekers

Definition

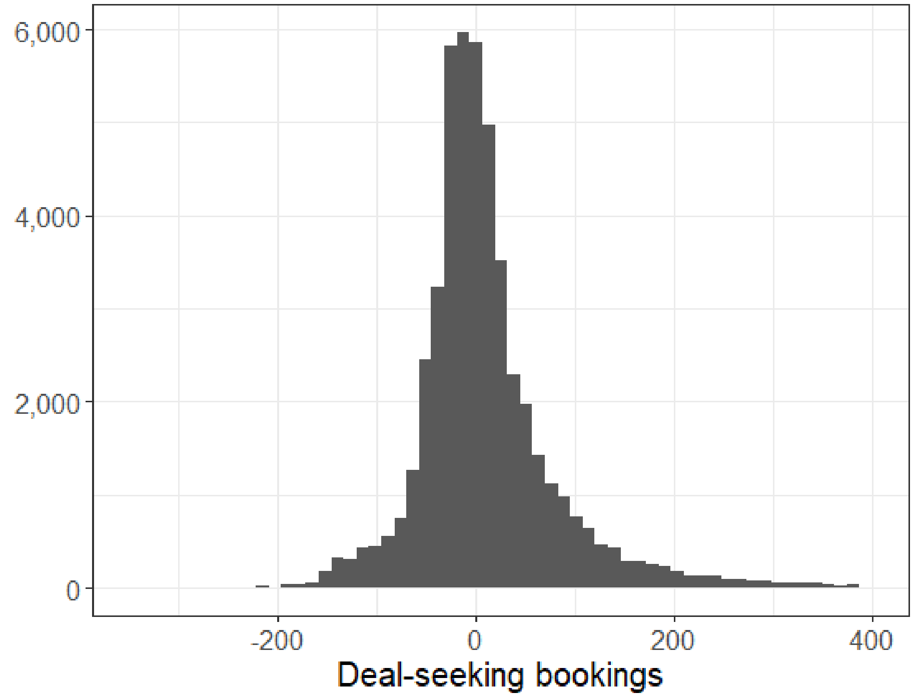
When bookings perform well in price compared to their star rating, controlling for the city

Measurement

$$ADR_{actual} - ADR_{predicted}$$

Where ADR_{actual} is the booking's ADR and $ADR_{predicted}$ is the fitted value of log ADR regressed on city and star rating

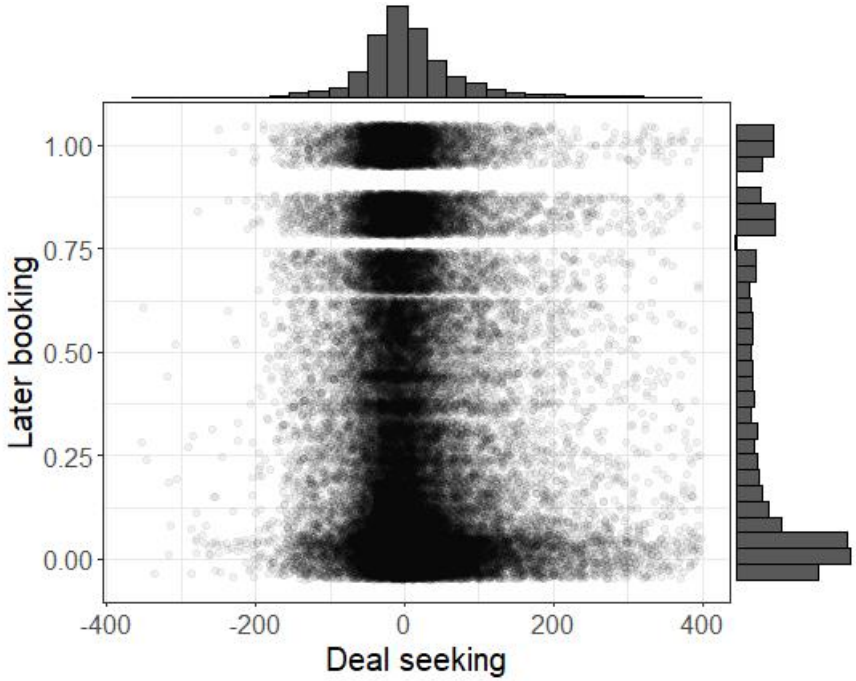
Histogram of deal-seeking measure



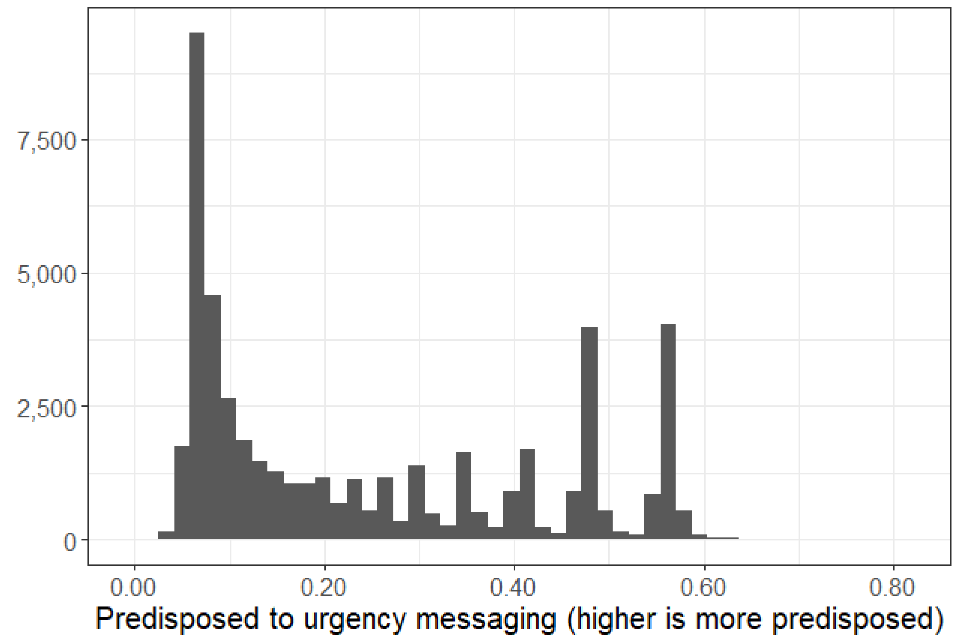
Note: Outliers were excluded above 400 for the histogram

Scaling and averaging 'late bookings' and 'deal-seeking' yields an index of predisposition to urgency messaging

Two measures...

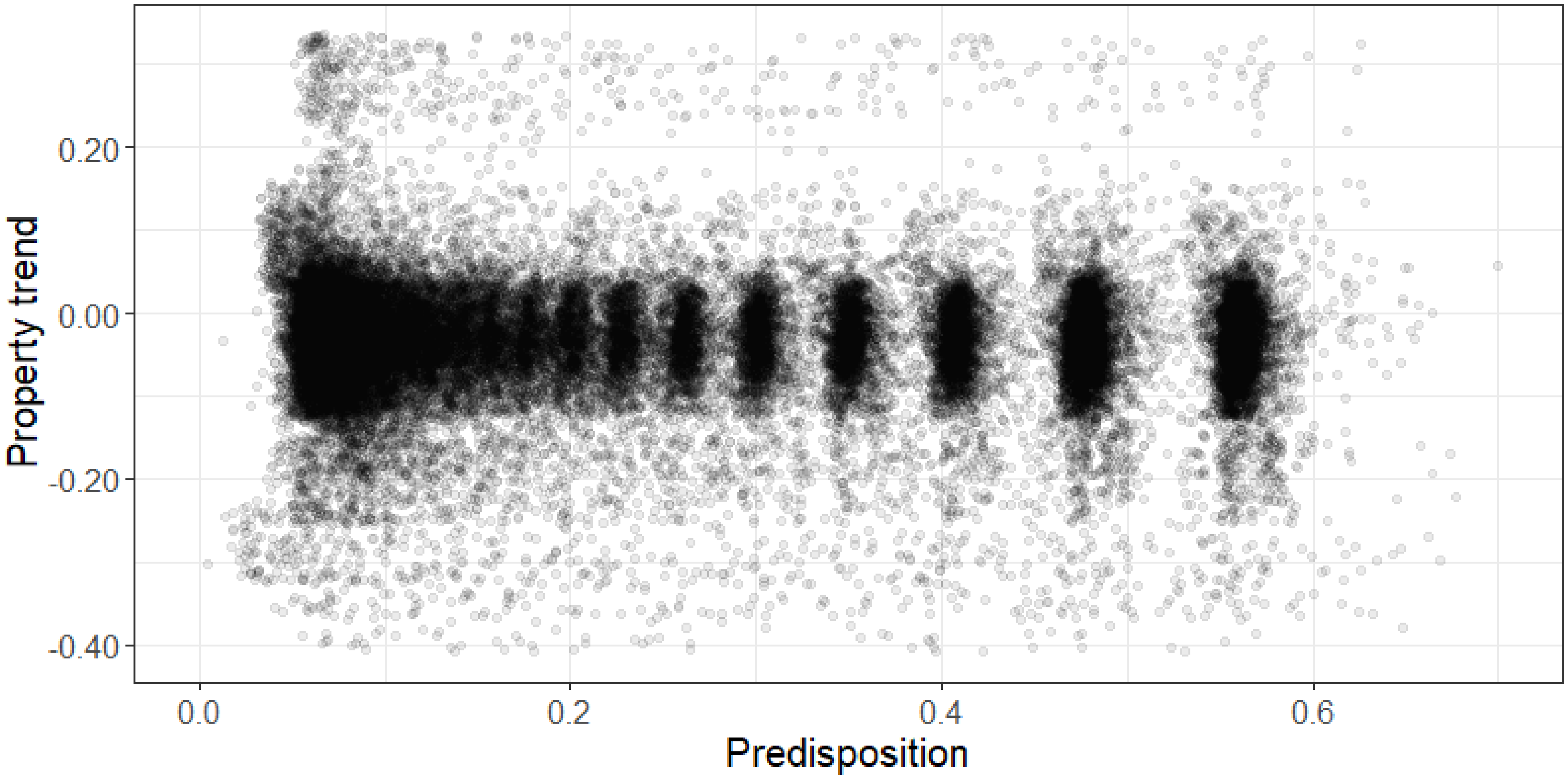


... combined into a single index



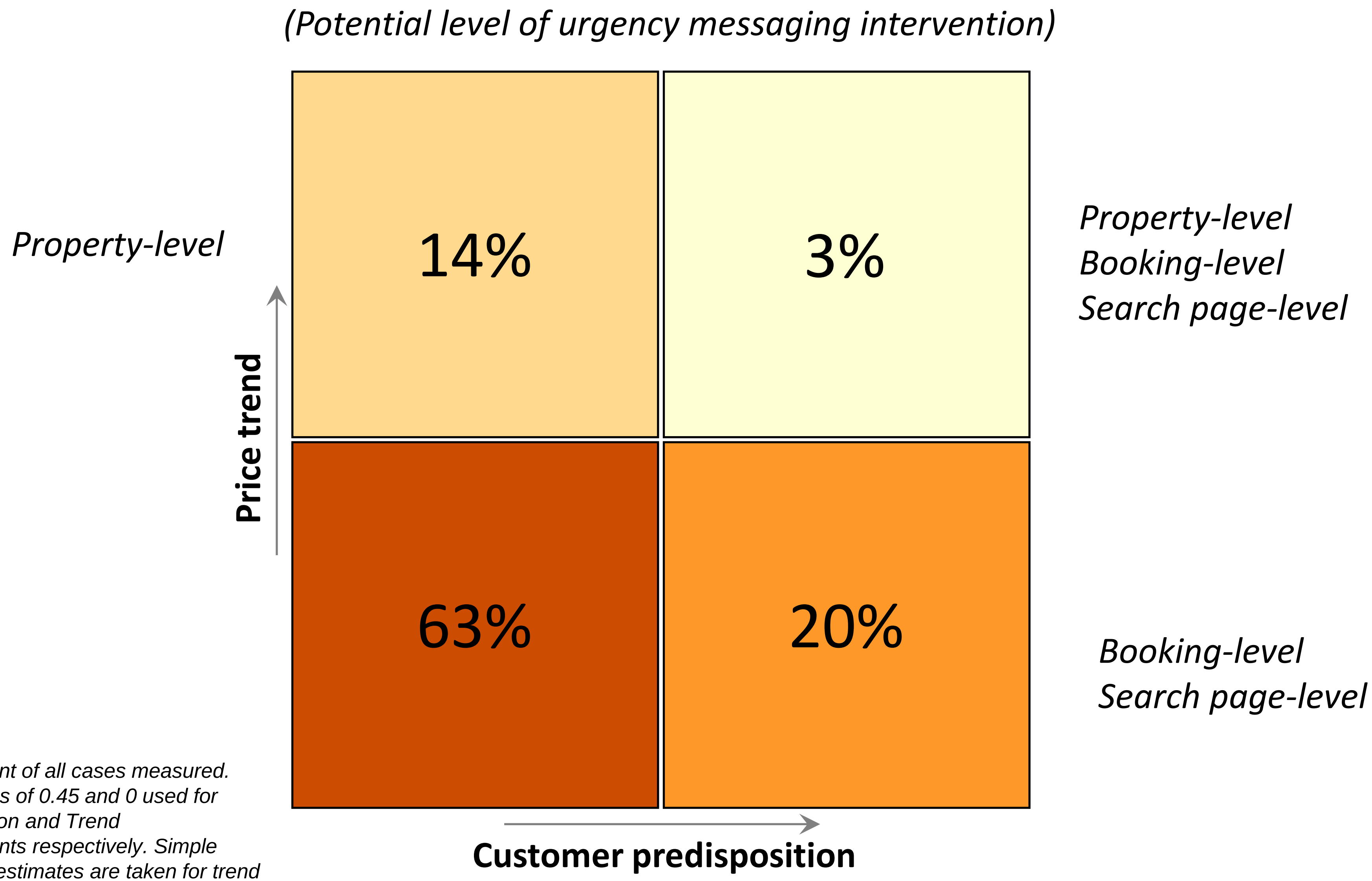
Note: Variables scaled using min-max scaling and combined by simple average

Predisposition to urgency messaging and property trend framework together



Note: total 1,345 bookings excluded (properties with less than 10 bookings + a few x-axis outliers)

Predisposition to urgency messaging and property trend framework together



Note: Percent of all cases measured. Cutoff values of 0.45 and 0 used for Predisposition and Trend measurements respectively. Simple slope point estimates are taken for trend

Table of contents

Findings and recommendations

Main findings

- There is a slightly **negative relationship** between booking-to-checkin time and ADR at property level
- **ADR decrease steepens** as booking-to-checkin days near 0
- **23% of all bookings** are made the day before or day-of check-in
- **1 in 4 bookings** were made by ‘deal-seeking’ customers

Recommendations

1. Understand what drives booking decisions
   - Explore more complete dataset
   - Update framework assumptions
2. Test urgency messaging
   - Target high-value sales first
   - Type
   - Level
3. Calculate net benefit of urgency messaging
   - ‘Are gains from this messaging hurting us elsewhere?’

Next steps

Data required

Use of data

- Room price over time at different b-to-c levels
- Room differentiators
  - Room square meters
  - Room bed size
  - Room is suite or not
- Time from search to booking
- Method of payment
- Time of booking (hour)
- Customer previous bookings

- Room-level time series by date and b-to-c

Room at B-to-C_x	10/1/2016	10/2/2016	10/3/2016	10/4/2016
Room_a_b2c_1	[ADR]	[ADR]	[ADR]	[ADR]
Room_a_b2c_2	[ADR]	[ADR]	[ADR]	[ADR]
Room_a_b2c_3	[ADR]	[ADR]	[ADR]	[ADR]
Room_a_b2c_4	[ADR]	[ADR]	[ADR]	[ADR]
Room_a_b2c_5	[ADR]	[ADR]	[ADR]	[ADR]

- Customer clustering analysis, profiling of customers

Framework testing

- Specific booking/customer level data (above)
- Level at which booking stopped

- Group covariate balance tests
- Construct ‘booking success’ variable

***A/B testing**: effect of urgency message on booking*

ANNEX

Annex regression output 1

```
Call:
lm(formula = log(ADR_USD) ~ book_2_checkin, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.1652 -0.5685  0.0223  0.5270  3.2537 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     4.6144567  0.0045687 1010.01   <2e-16 ***
book_2_checkin -0.0067555  0.0002137  -31.61   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7423 on 49059 degrees of freedom
Multiple R-squared:  0.01996,	Adjusted R-squared:  0.01994 
F-statistic: 999.2 on 1 and 49059 DF,  p-value: < 2.2e-16
```

Annex regression output 2

```
Call:
lm(formula = log(ADR_USD) ~ city + book_2_checkin + I(book_2_checkin^2) +
    I(book_2_checkin^3) + city * book_2_checkin, data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-2.88259 -0.43681  0.00734  0.43796  3.12139

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)           4.332e+00  6.914e-03 626.523  < 2e-16 ***
cityB                -3.352e-02  1.364e-02  -2.457   0.0140 *
cityC                 5.241e-01  1.398e-02  37.478  < 2e-16 ***
cityD                 8.290e-01  1.088e-02  76.174  < 2e-16 ***
cityE                 2.418e-01  1.345e-02  17.977  < 2e-16 ***
book_2_checkin       -1.790e-02  1.309e-03 -13.675  < 2e-16 ***
I(book_2_checkin^2)  -5.759e-04  6.331e-05  -9.097  < 2e-16 ***
I(book_2_checkin^3)  -5.649e-06  7.943e-07  -7.112 1.16e-12 ***
cityB:book_2_checkin -5.020e-03  6.467e-04  -7.763 8.44e-15 ***
cityC:book_2_checkin -6.720e-03  5.578e-04 -12.047  < 2e-16 ***
cityD:book_2_checkin  2.623e-03  5.146e-04   5.098 3.44e-07 ***
cityE:book_2_checkin  1.473e-03  7.337e-04   2.008   0.0446 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.657 on 49049 degrees of freedom
Multiple R-squared:  0.2323,	Adjusted R-squared:  0.2322
F-statistic:  1350 on 11 and 49049 DF,  p-value: < 2.2e-16
```

Annex regression output 3

```
Call:
lm(formula = log(ADR_USD) ~ city + book_2_checkin + I(book_2_checkin^2) +
    I(book_2_checkin^3), data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.8744 -0.4381  0.0099  0.4401  3.2029 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)          4.324e+00  6.394e-03 676.255  < 2e-16 ***
cityB                3.500e-02  1.037e-02   3.373 0.000743 ***
cityC                6.560e-01  9.252e-03  70.904  < 2e-16 ***
cityD                7.880e-01  7.905e-03  99.687  < 2e-16 ***
cityE                2.271e-01  1.048e-02  21.667  < 2e-16 ***
book_2_checkin      -1.768e-02  1.299e-03 -13.609  < 2e-16 ***
I(book_2_checkin^2) -5.417e-04  6.329e-05  -8.559  < 2e-16 ***
I(book_2_checkin^3) -5.484e-06  7.954e-07  -6.894 5.47e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6589 on 49053 degrees of freedom
Multiple R-squared:  0.2278,	Adjusted R-squared:  0.2277 
F-statistic:  2067 on 7 and 49053 DF,  p-value: < 2.2e-16
```

Annex regression output 4

```
Call:
lm(formula = book_2_checkin ~ log(ADR_USD) + city + type + star_rating +
    chain_hotel + booking_date + dayweek_booked + stay_days,
    data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-49.890  -7.315   2.469   9.885  31.272

Coefficients:
                                   Estimate Std. Error  t value Pr(>|t|)
(Intercept)                      -3.687e+03  3.518e+01 -104.792  < 2e-16 ***
log(ADR_USD)                     -3.683e+00  1.373e-01  -26.821  < 2e-16 ***
cityB                             1.960e-01  2.658e-01    0.737 0.460884
cityC                            -2.129e+00  2.625e-01   -8.111 5.15e-16 ***
cityD                             1.580e+00  2.229e-01    7.089 1.37e-12 ***
cityE                             1.761e+00  2.403e-01    7.328 2.38e-13 ***
typeBungalow                      1.407e+00  1.458e+00    0.966 0.334286
typeCapsule Hotel                -3.024e+00  1.023e+00   -2.957 0.003106 **
typeGuest House / Bed & Breakfast -1.378e+00  9.230e-01   -1.493 0.135489
typeHoliday Park / Caravan Park  -7.878e+00  7.865e+00   -1.002 0.316537
typeHome                         -2.425e-01  3.726e+00   -0.065 0.948107
typeHostel                       -3.619e+00  9.341e-01   -3.874 0.000107 ***
typeHotel                        -1.434e+00  9.131e-01   -1.571 0.116245
typeLove Hotel                   -2.083e+00  7.863e+00   -0.265 0.791094
typeMotel                        -1.189e+00  2.964e+00   -0.401 0.688247
typePrivate Villa                 8.408e-01  3.322e+00    0.253 0.800217
typeResort                       -2.604e+00  9.513e-01   -2.737 0.006200 **
typeResort Villa                 -1.716e+00  1.146e+00   -1.497 0.134418
typeRyokan                       -3.988e+00  1.604e+00   -2.486 0.012925 *
typeServiced Apartment            3.281e-01  9.470e-01    0.346 0.728975
typeVilla                         1.589e+01  1.356e+01    1.172 0.241177
star_rating1                      4.164e-01  1.171e+00    0.356 0.722139
star_rating1.5                    2.989e-01  1.306e+00    0.229 0.819041
star_rating2                     -1.304e+00  1.165e+00   -1.119 0.262936
star_rating2.5                   -2.060e+00  1.178e+00   -1.749 0.080275 .
star_rating3                      1.056e-01  1.145e+00    0.092 0.926527
star_rating3.5                   -1.545e+00  1.154e+00   -1.339 0.180520
star_rating4                      3.141e+00  1.154e+00    2.722 0.006493 **
star_rating4.5                    1.904e+00  1.229e+00    1.549 0.121453
star_rating5                      4.132e+00  1.182e+00    3.497 0.000471 ***
chain_hotelnon-chain             -5.371e-01  1.580e-01   -3.400 0.000675 ***
booking_date                      2.500e-06  2.373e-08  105.376  < 2e-16 ***
dayweek_bookedMon                -1.105e+00  2.244e-01   -4.925 8.48e-07 ***
dayweek_bookedSat                -4.273e-01  2.373e-01   -1.800 0.071790 .
dayweek_bookedSun                -2.058e+00  2.366e-01   -8.699  < 2e-16 ***
dayweek_bookedThu                -7.632e-01  2.254e-01   -3.386 0.000710 ***
dayweek_bookedTue                -1.027e+00  2.239e-01   -4.589 4.47e-06 ***
dayweek_bookedWed                -9.964e-01  2.248e-01   -4.432 9.35e-06 ***
stay_days                        -2.657e+00  8.100e-02  -32.798  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.53 on 49022 degrees of freedom
Multiple R-squared:  0.2563,  Adjusted R-squared:  0.2557
F-statistic: 444.6 on 38 and 49022 DF,  p-value: < 2.2e-16
```

Annex regression output 5

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: ADR_USD ~ book_2_checkin + (book_2_checkin | hotel_id) + city +
    star_rating + type + month_booked + dayweek_checkin + dayweek_booked +
    stay_days
   Data: df_prop

     AIC      BIC   logLik deviance df.resid
 66779.7  67192.0 -33342.8  66685.7    47674

Scaled residuals:
     Min       1Q   Median       3Q      Max
-11.9080  -0.3770  -0.0759   0.2199  20.0635

Random effects:
 Groups   Name           Variance Std.Dev. Corr
 hotel_id (Intercept)    0.692358 0.83208
          book_2_checkin 0.006485 0.08053  -0.22
 Residual                0.223778 0.47305
Number of obs: 47721, groups:  hotel_id, 455

Fixed effects:
                                   Estimate Std. Error t value
(Intercept)                       -0.579213   0.643744  -0.900
book_2_checkin                    -0.043277   0.005437  -7.959
cityB                              0.085038   0.130471   0.652
cityC                              1.367986   0.132543  10.321
cityD                              1.291616   0.161772   7.984
cityE                              0.139216   0.133164   1.045
star_rating1                      -0.613187   0.505551  -1.213
star_rating1.5                    -0.543357   0.558565  -0.973
star_rating2                      -0.011748   0.493456  -0.024
star_rating2.5                     0.102849   0.512102   0.201
star_rating3                       0.171387   0.487819   0.351
star_rating3.5                     0.366594   0.511093   0.717
star_rating4                       0.751140   0.493473   1.522
star_rating4.5                     1.481361   0.610059   2.428
star_rating5                       2.907733   0.515195   5.644
typeBungalow                       0.040764   0.597924   0.068
typeCapsule Hotel                 -0.809310   0.543991  -1.488
typeGuest House / Bed & Breakfast -0.199067   0.437433  -0.455
typeHome                          -0.032480   0.944857  -0.034
typeHostel                        -0.411249   0.443529  -0.927
typeHotel                         -0.321680   0.432231  -0.744
typeMotel                         -0.355723   1.048289  -0.339
typeResort                        -0.092185   0.450079  -0.205
typeResort Villa                   0.192144   0.498921   0.385
typeRyokan                        -0.363056   0.645876  -0.562
typeServiced Apartment            -0.304157   0.456849  -0.666
month_booked9                      0.063229   0.016926   3.736
month_booked10                     0.076544   0.016837   4.546
month_booked11                     0.092249   0.016926   5.450
month_booked12                     0.134637   0.017606   7.647
dayweek_checkinMon                -0.206756   0.008421 -24.554
dayweek_checkinSat                 0.040200   0.007222   5.566
dayweek_checkinSun                -0.196842   0.008120 -24.241
dayweek_checkinThu                -0.068536   0.007857  -8.722
dayweek_checkinTue                -0.193809   0.008399 -23.076
dayweek_checkinWed                -0.139251   0.008212 -16.958
dayweek_bookedMon                  0.029400   0.008103   3.628
dayweek_bookedSat                 -0.001501   0.008546  -0.176
dayweek_bookedSun                  0.012542   0.008555   1.466
dayweek_bookedThu                  0.013624   0.008076   1.687
dayweek_bookedTue                  0.017625   0.008083   2.180
dayweek_bookedWed                  0.020532   0.008123   2.528
stay_days                          0.014187   0.003056   4.643
```

Annex regression output 6

```
Call:
lm(formula = book_2_checkin ~ business, data = df_business)

Residuals:
    Min      1Q  Median      3Q     Max 
-48.350  -8.092   5.908  11.650  15.908 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)          -11.6503     0.1235  -94.37   <2e-16 ***
businessNon-business  -4.2575     0.1501  -28.36   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.55 on 49059 degrees of freedom
Multiple R-squared:  0.01613,	Adjusted R-squared:  0.01611 
F-statistic: 804.5 on 1 and 49059 DF,  p-value: < 2.2e-16
```

Annex regression output 7

```
Call:
lm(formula = ADR_USD ~ business, data = df_business)

Residuals:
    Min      1Q  Median      3Q     Max 
-151.77  -82.91  -35.36   43.26 3000.47 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)           130.736      1.070  122.18   <2e-16 ***
businessNon-business   25.658      1.301   19.72   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 134.8 on 49059 degrees of freedom
Multiple R-squared:  0.007866,	Adjusted R-squared:  0.007846 
F-statistic:   389 on 1 and 49059 DF,  p-value: < 2.2e-16
```

Annex regression output 8

```
Call:
lm(formula = ADR_USD ~ business + city, data = df_business)

Residuals:
    Min      1Q  Median      3Q     Max
-212.22  -59.73  -24.66   34.36 2925.25

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)            85.522      1.131  75.619   <2e-16 ***
businessNon-business   23.140      1.194  19.373   <2e-16 ***
cityB                  16.235      1.943   8.357   <2e-16 ***
cityC                 122.949      1.711  71.878   <2e-16 ***
cityD                 118.003      1.477  79.907   <2e-16 ***
cityE                  39.072      1.965  19.888   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 123.4 on 49055 degrees of freedom
Multiple R-squared:  0.1689,	Adjusted R-squared:  0.1688
F-statistic:  1994 on 5 and 49055 DF,  p-value: < 2.2e-16
```

Annex regression output 9

```
Call:
lm(formula = log(ADR_USD) ~ city + relevel(star_rating, ref = "2"),
    data = df)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.87391 -0.30978 -0.00985  0.28890  3.11019 

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)                         3.634725   0.009908 366.841   <2e-16 ***
cityB                               0.073464   0.007549   9.732   <2e-16 ***
cityC                               1.056841   0.006983 151.345   <2e-16 ***
cityD                               0.917540   0.006202 147.952   <2e-16 ***
cityE                               0.219714   0.007502  29.286   <2e-16 ***
relevel(star_rating, ref = "2")0    0.029547   0.036731   0.804    0.421    
relevel(star_rating, ref = "2")1   -0.247551   0.013132 -18.851   <2e-16 ***
relevel(star_rating, ref = "2")1.5 -0.347405   0.023921 -14.523   <2e-16 ***
relevel(star_rating, ref = "2")2.5  0.401168   0.014355  27.946   <2e-16 ***
relevel(star_rating, ref = "2")3    0.565068   0.010498  53.826   <2e-16 ***
relevel(star_rating, ref = "2")3.5  0.479180   0.011390  42.071   <2e-16 ***
relevel(star_rating, ref = "2")4    1.072017   0.010486 102.232   <2e-16 ***
relevel(star_rating, ref = "2")4.5  1.372928   0.017097  80.304   <2e-16 ***
relevel(star_rating, ref = "2")5    1.658667   0.012042 137.740   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4636 on 49047 degrees of freedom
Multiple R-squared:  0.6178,	Adjusted R-squared:  0.6177 
F-statistic:  6098 on 13 and 49047 DF,  p-value: < 2.2e-16
```