
Predicting Your Hotel Experience
kaggle
Booking.com
yelp
Google Maps

How does it work?

It starts with a booking

The only way to leave a review is to first make a booking. That's how we know our reviews come from real guests who have stayed at the property.

Followed by a trip

When guests stay at the property they check out how quiet the room is, how friendly the staff are and more.

And finally, a review

After their trip, guests tell us about their stay. We check for naughty words and verify the authenticity of all guest reviews before adding them to our site.

DATA. EXPLAINED.

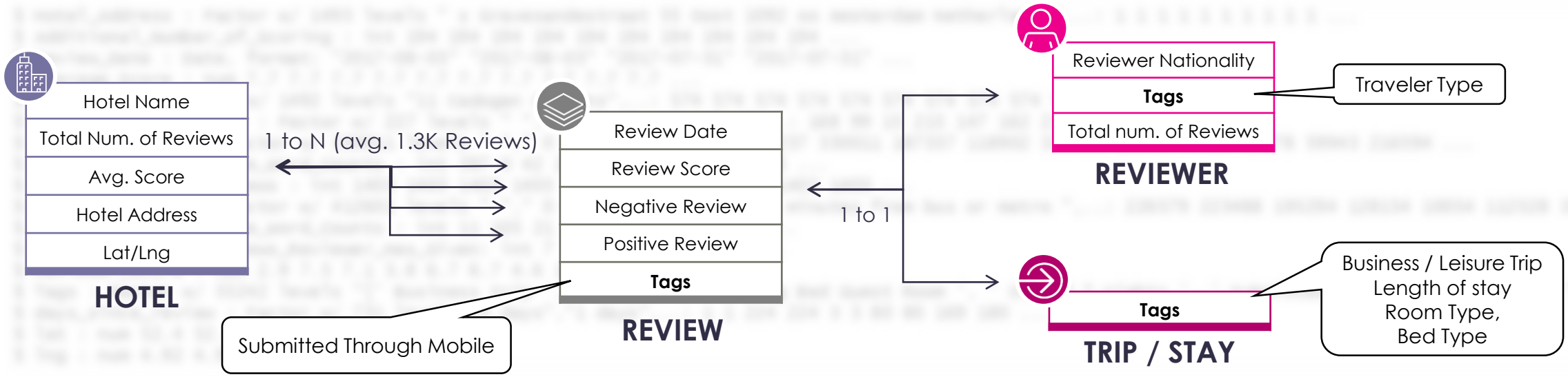
Booking.com

kaggle

```
$ Hotel_Address : Factor w/ 1493 levels " s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands",..: 1 1 1 1 1 1 1 1 1 1 ...
$ Additional_Number_of_Scoring : int 194 194 194 194 194 194 194 194 194 194 ...
$ Review_Date : Date, format: "2017-08-03" "2017-08-03" "2017-07-31" "2017-07-31" ...
$ Average_Score : num 7.7 7.7 7.7 7.7 7.7 7.7 7.7 7.7 7.7 7.7 ...
$ Hotel_Name : Factor w/ 1492 levels "11 Cadogan Gardens",..: 574 574 574 574 574 574 574 574 574 574 ...
$ Reviewer_Nationality : Factor w/ 227 levels " "," Abkhazia Georgia ",..: 168 99 15 215 147 162 215 215 23 152 ...
$ Negative_Review : Factor w/ 330011 levels " "," 0 00 Comments ",..: 80237 330011 187337 118902 329850 16033 44426 13278 59943 216594 ...
$ Review_Total_Negative_Word_Counts : int 397 0 42 210 140 17 33 11 34 15 ...
$ Total_Number_of_Reviews : int 1403 1403 1403 1403 1403 1403 1403 1403 1403 1403 ...
$ Positive_Review : Factor w/ 412601 levels " "," 0 noises Good sleep 10 minutes from bus or metro ",..: 226579 223488 195294 128154 10054 112528 335320
$ Review_Total_Positive_Word_Counts : int 11 105 21 26 8 20 18 19 0 50 ...
$ Total_Number_of_Reviews_Reviewer_Has_Given: int 7 7 9 1 3 1 6 1 3 1 ...
$ Reviewer_Score : num 2.9 7.5 7.1 3.8 6.7 6.7 4.6 10 6.5 7.9 ...
$ Tags : Factor w/ 55242 levels "[' Business trip ', ' Couple ', ' 1 King Bed Guest Room ', ' Stayed 2 nights ', ' Submitted from a mobile device ']",..:
$ days_since_review : Factor w/ 731 levels "0 days","1 days",..: 1 1 224 224 3 3 80 80 169 180 ...
$ lat : num 52.4 52.4 52.4 52.4 52.4 ...
$ lng : num 4.92 4.92 4.92 4.92 4.92 ...
```

DATA. EXPLAINED.

Booking.com



Around ~1.4K hotels in Europe

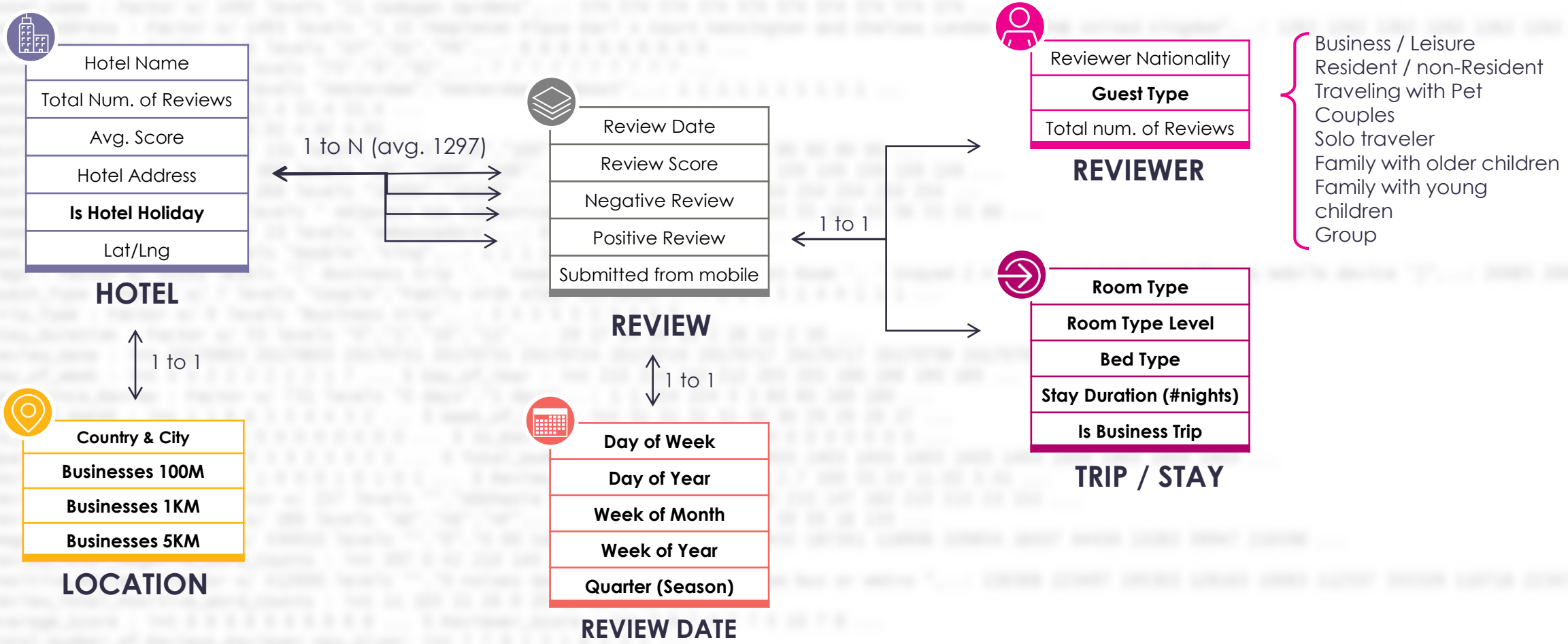
Over ½ Million Reviews

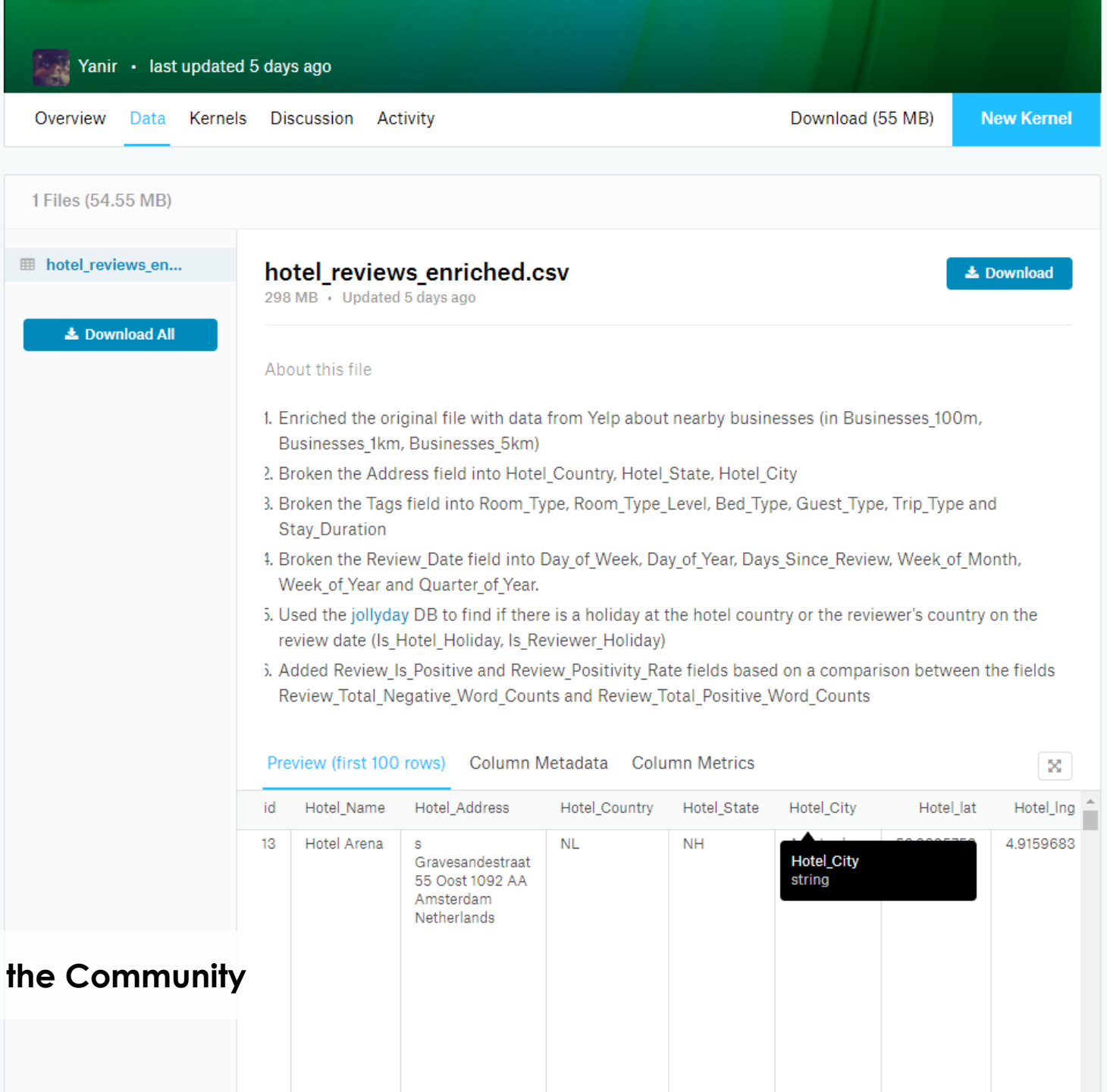
DATA. ENRICHED.

Booking.com

yelp

Google Maps

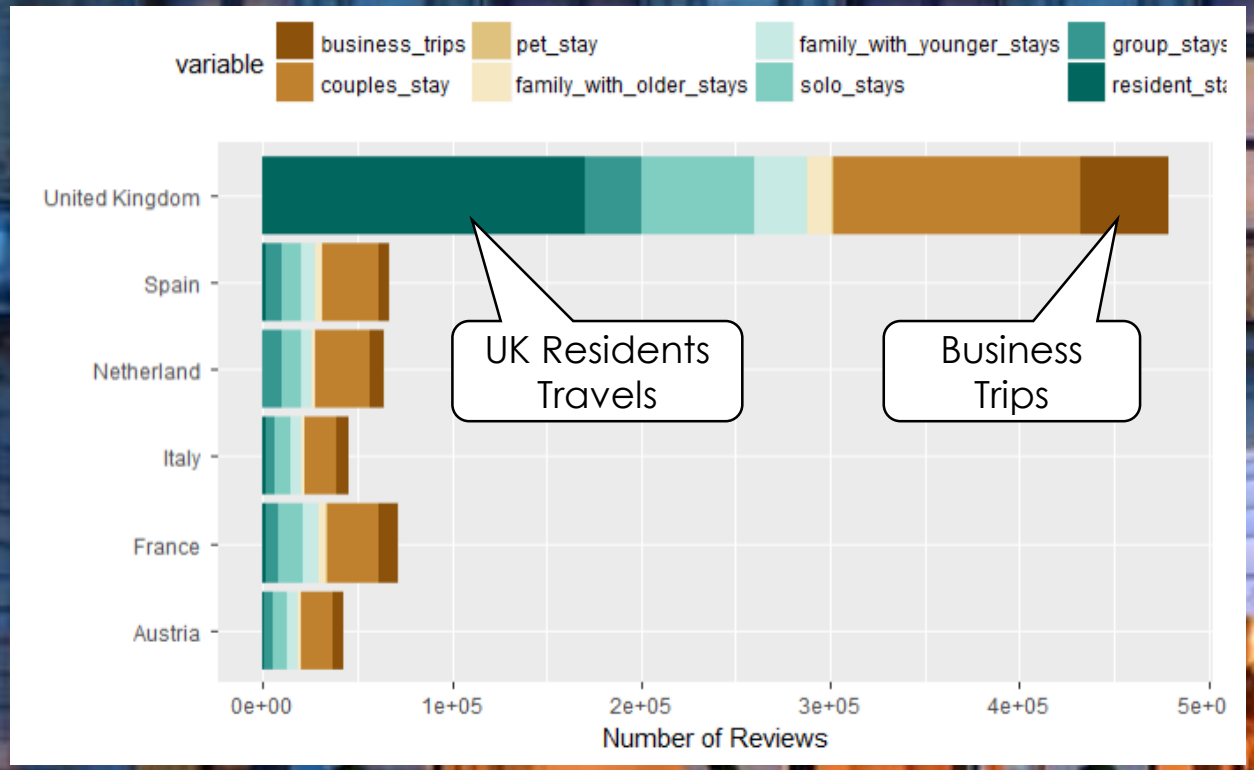


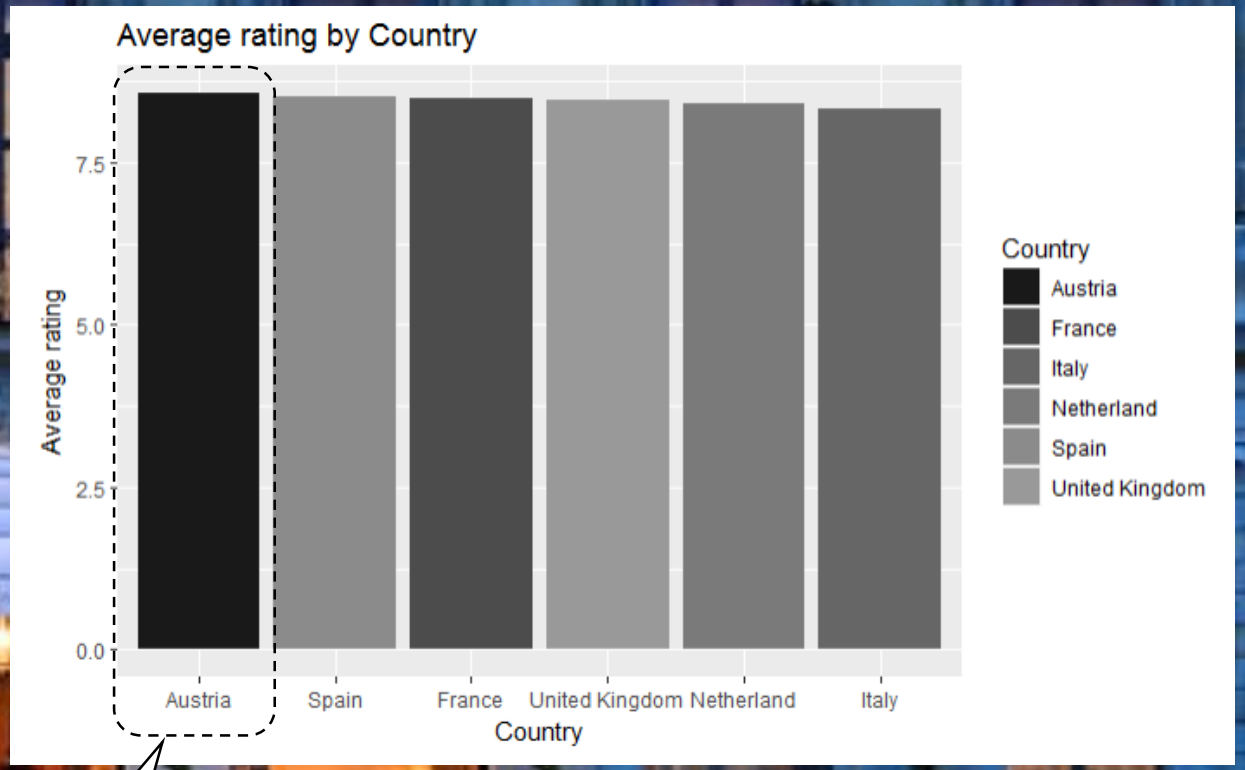


kaggle

Contributing Back to the Community

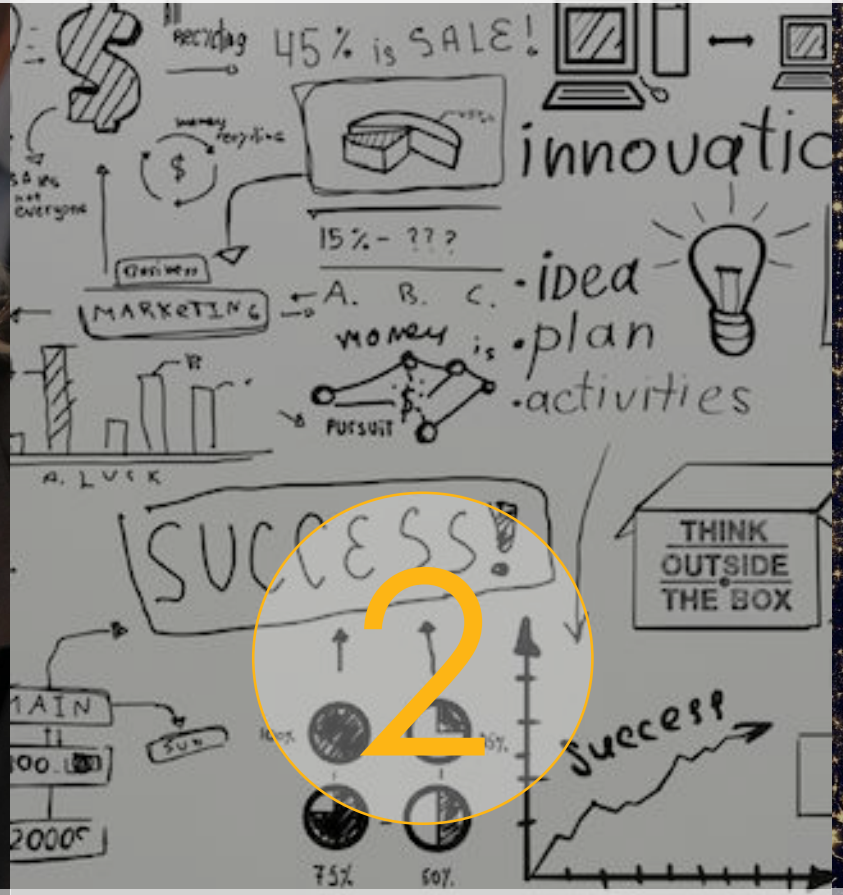
DATA. UNDERSTAND. VISUALIZE.

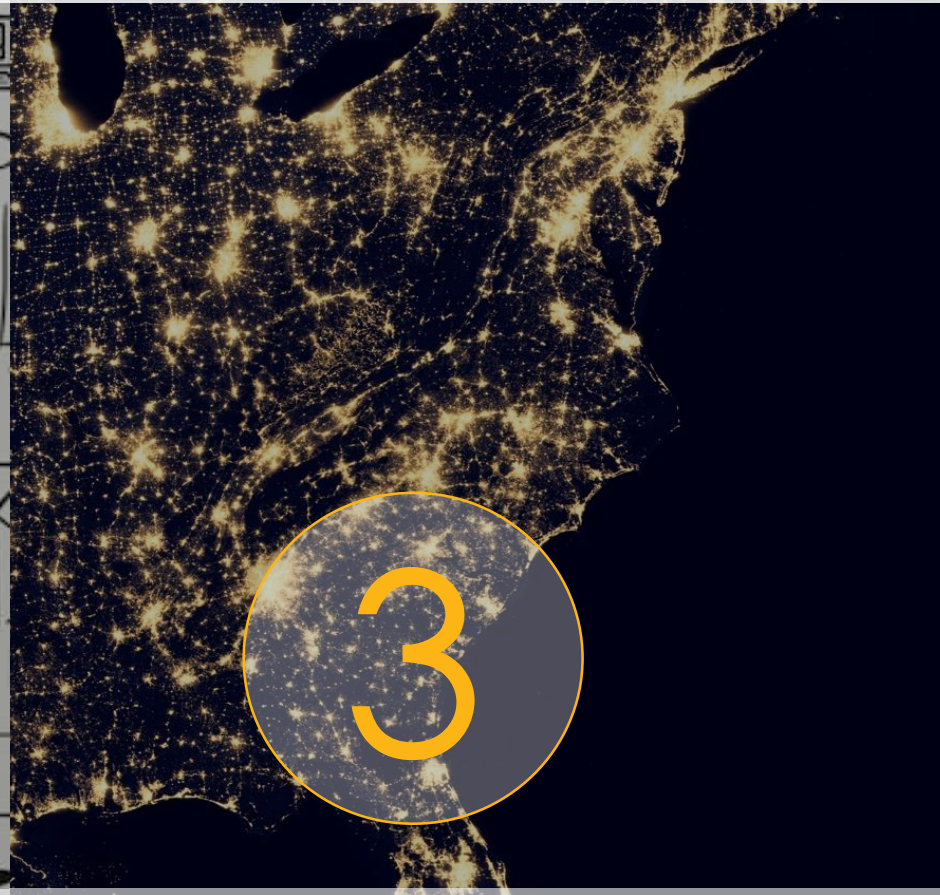




Austria Hotels ranked highest

3 Business Questions





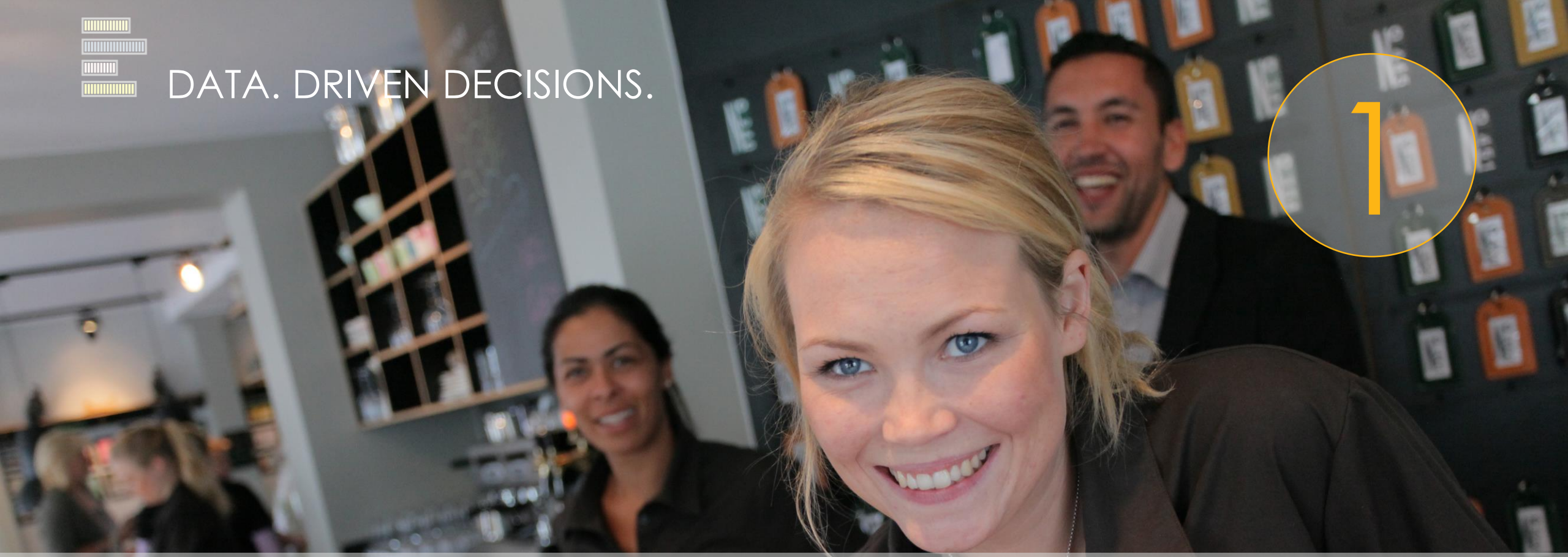
Traveler

Booking.com

Hotels

DATA. DRIVEN DECISIONS.

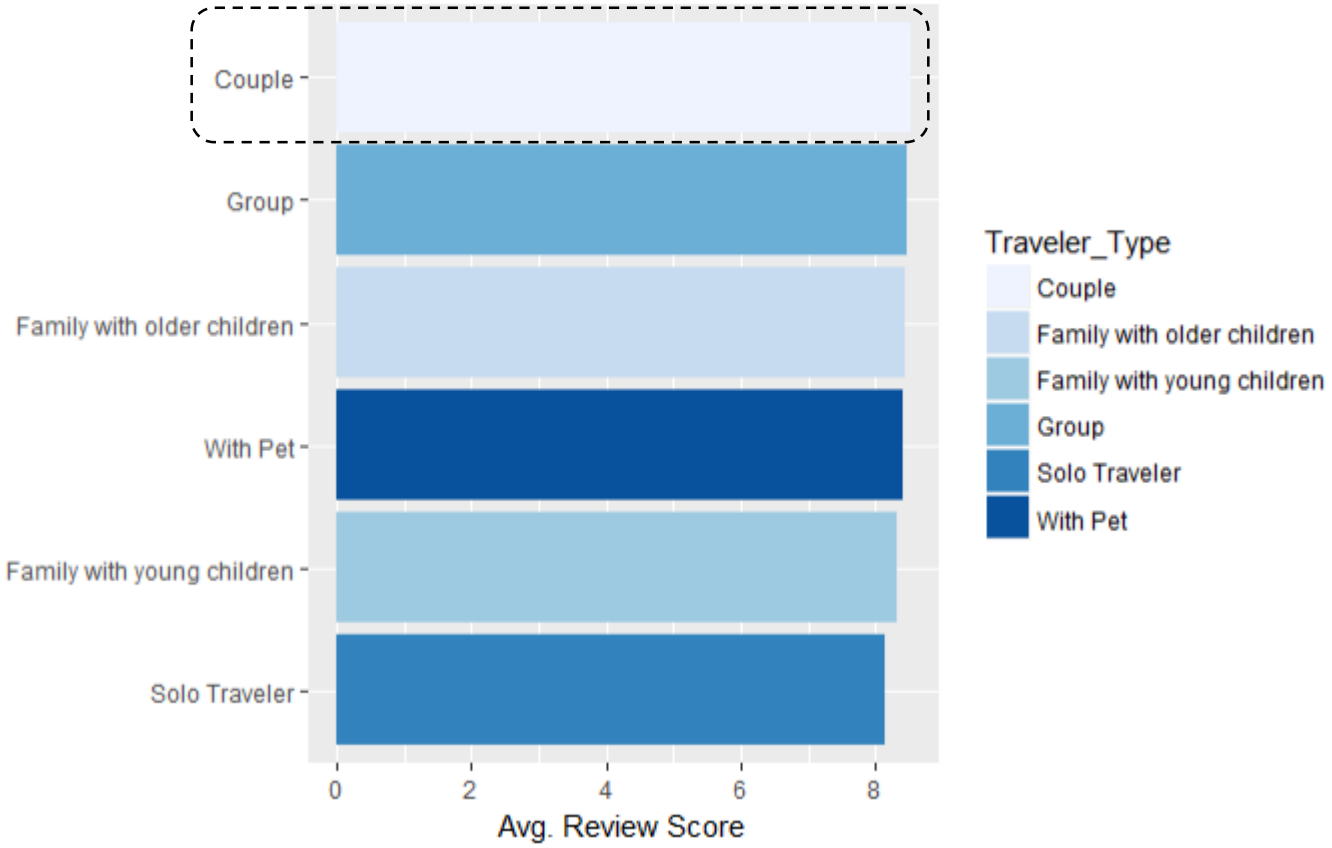
1

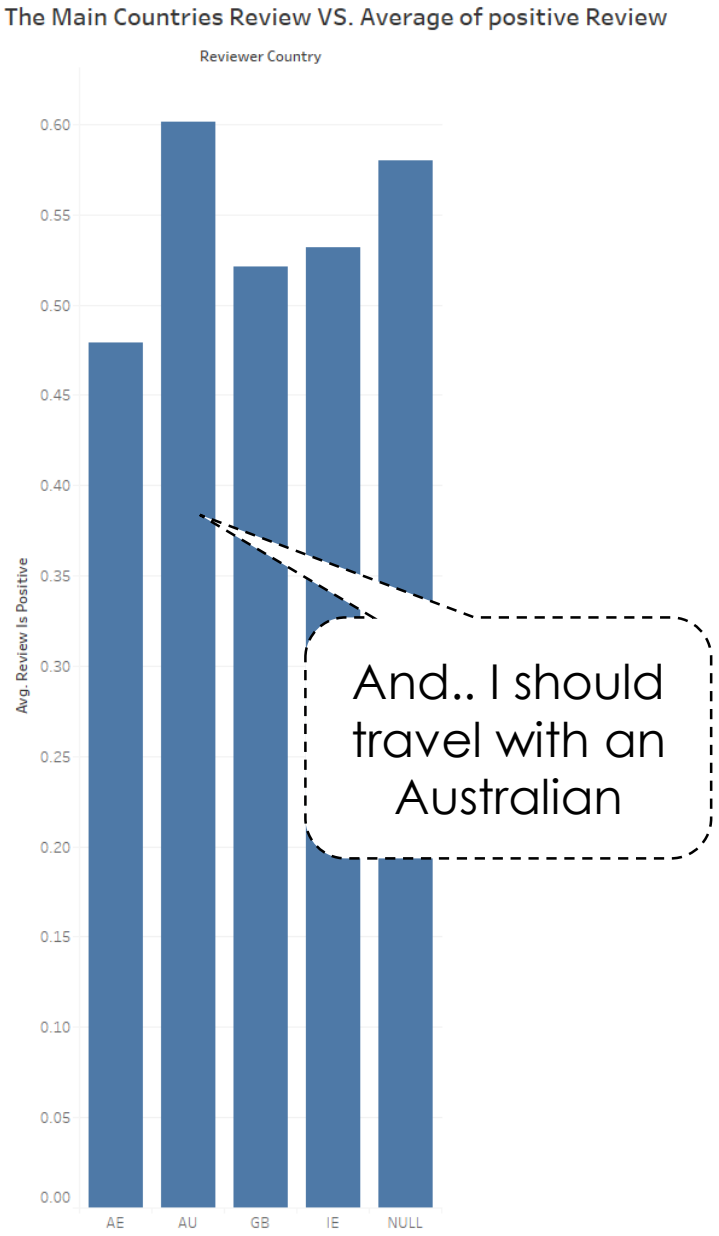


How to maximize my travel Experience?

DATA. DRIVEN DECISIONS.

People traveling as couple tend to enjoy their trip best





Booking.com

Data Driven Design & Growth

DATA. DRIVEN DECISIONS.

Pam Miles
Booking.com
Product Manager, Growth

COMPETITION:

KAYAK
Expedia
Hotels.com
tripadvisor
travelocity

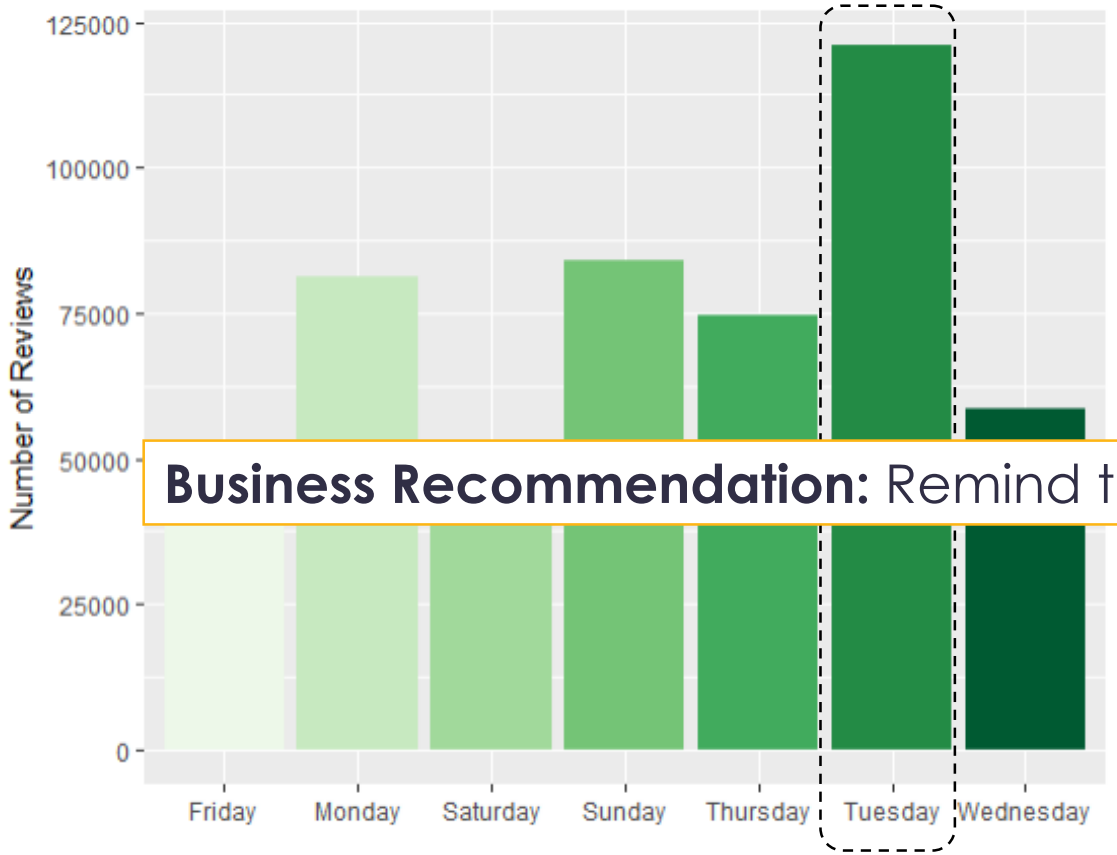
BUSINESS PROBLEM:

Retention and acquisition of unique website visits

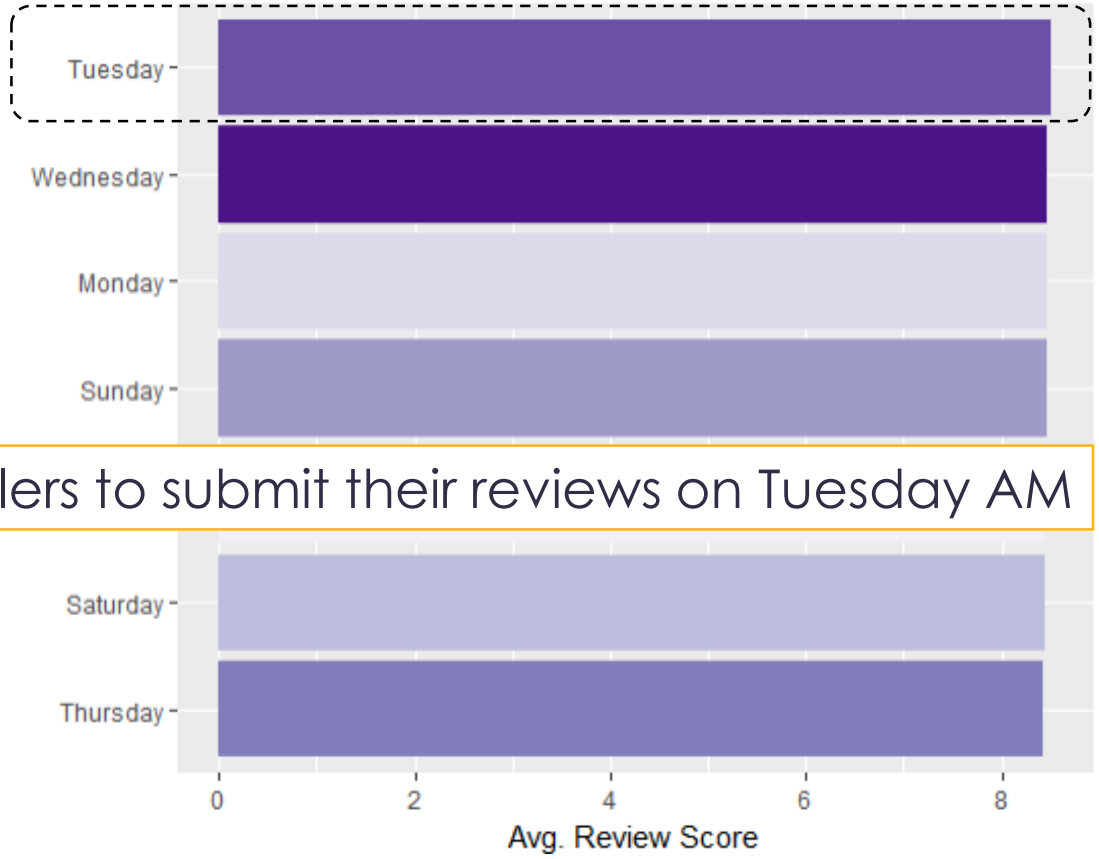
(More Reviews ⟷ Better Engagement)

DATA. DRIVEN DECISIONS.

Tuesday is the Most Popular day to submit a review

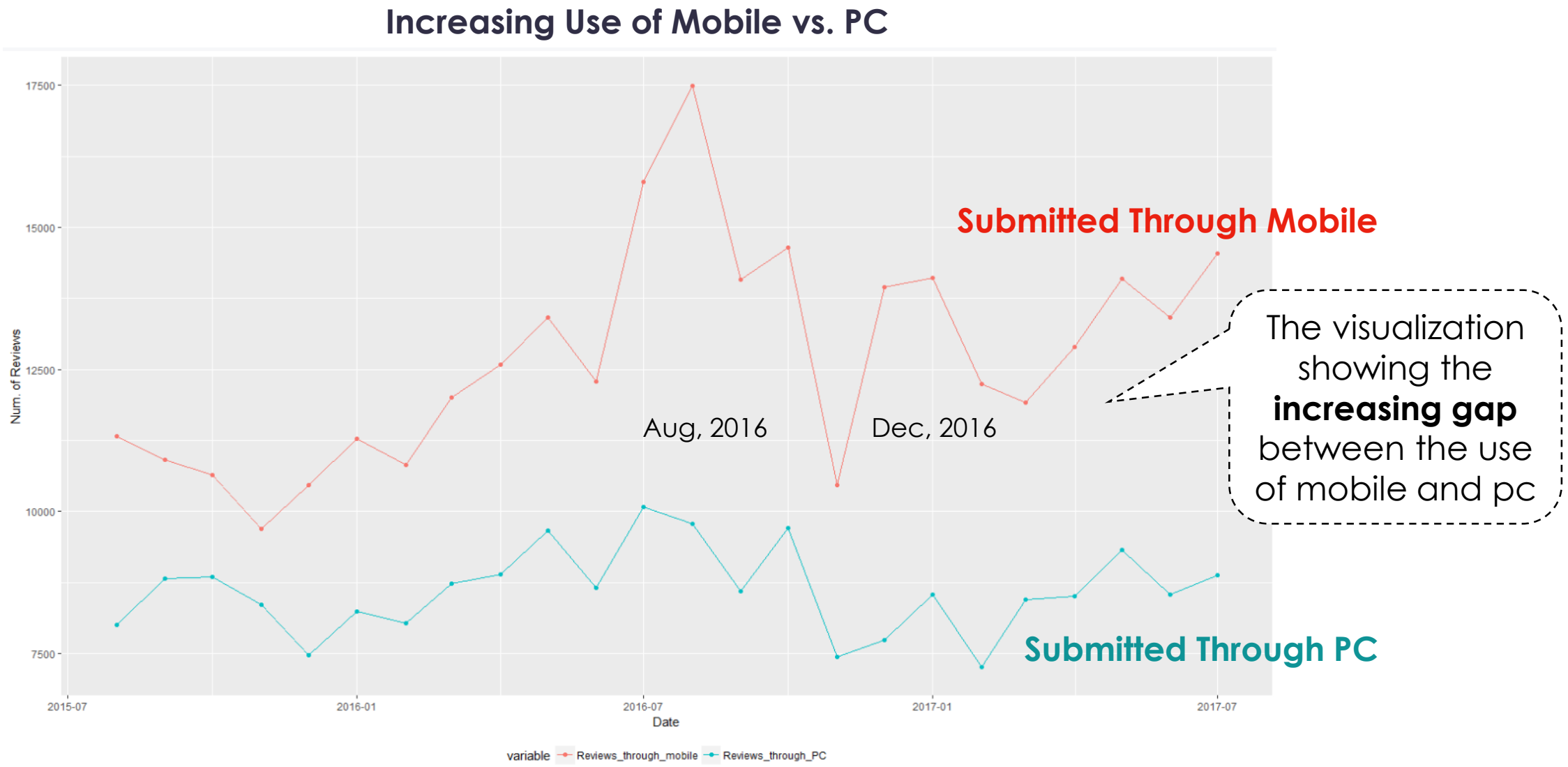


...Its also the day the review will be **most positive**



Business Recommendation: Remind travelers to submit their reviews on Tuesday AM

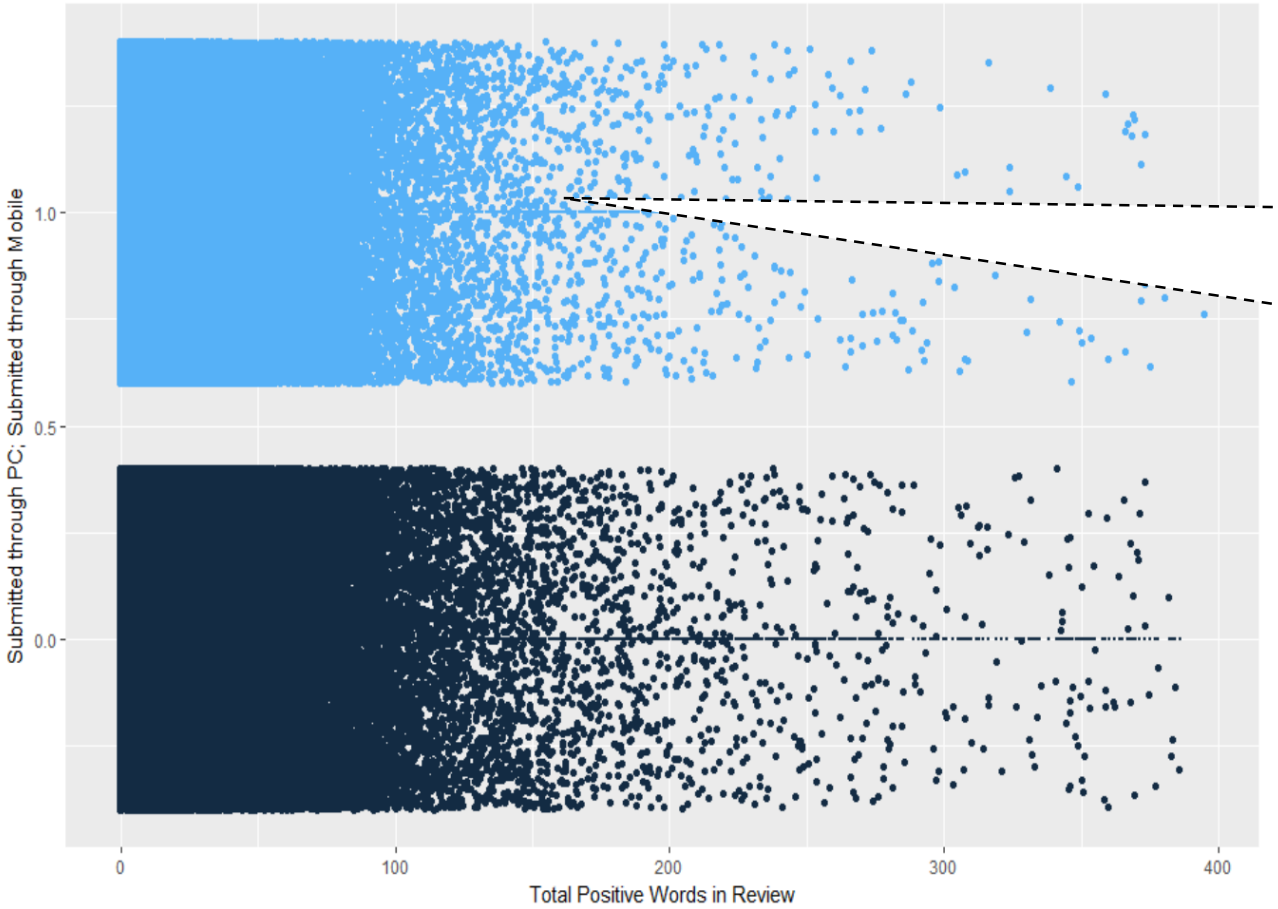
DATA. DRIVEN DECISIONS.



The visualization showing the **increasing gap** between the use of mobile and pc

DATA. PREDICTION

Slight Less Use of words for Positive reviews **Through Mobile**



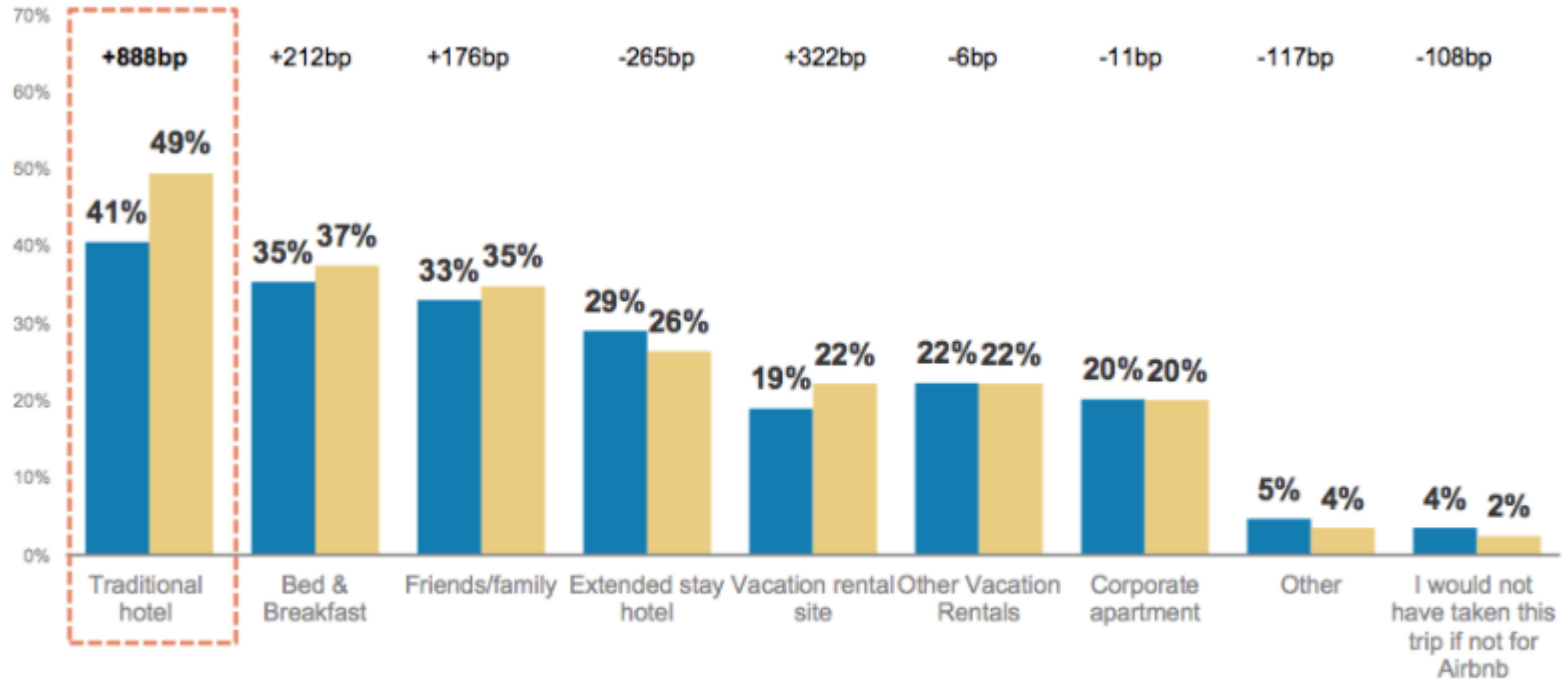
While in the Negative reviews we see almost similar number of words through different channels, here we see **slight decrease in number of words** if review submitted through mobile

Looking At The Bigger Picture

DATA. DRIVEN DECISIONS.

Airbnb Is Becoming an Even Bigger Threat to Hotels

For now, Airbnb users also appear to be using Airbnb for longer stays as opposed to one-night stays.
In its survey, Morgan Stanley found 6% of Airbnb stays are one night and 22% are for six nights or more.
26% of hotel users, by comparison, stay for just one night



AlphaWise, Morgan Stanley Research

Airbnb users are substituting away from hotels more than any other accommodation type — and more quickly.

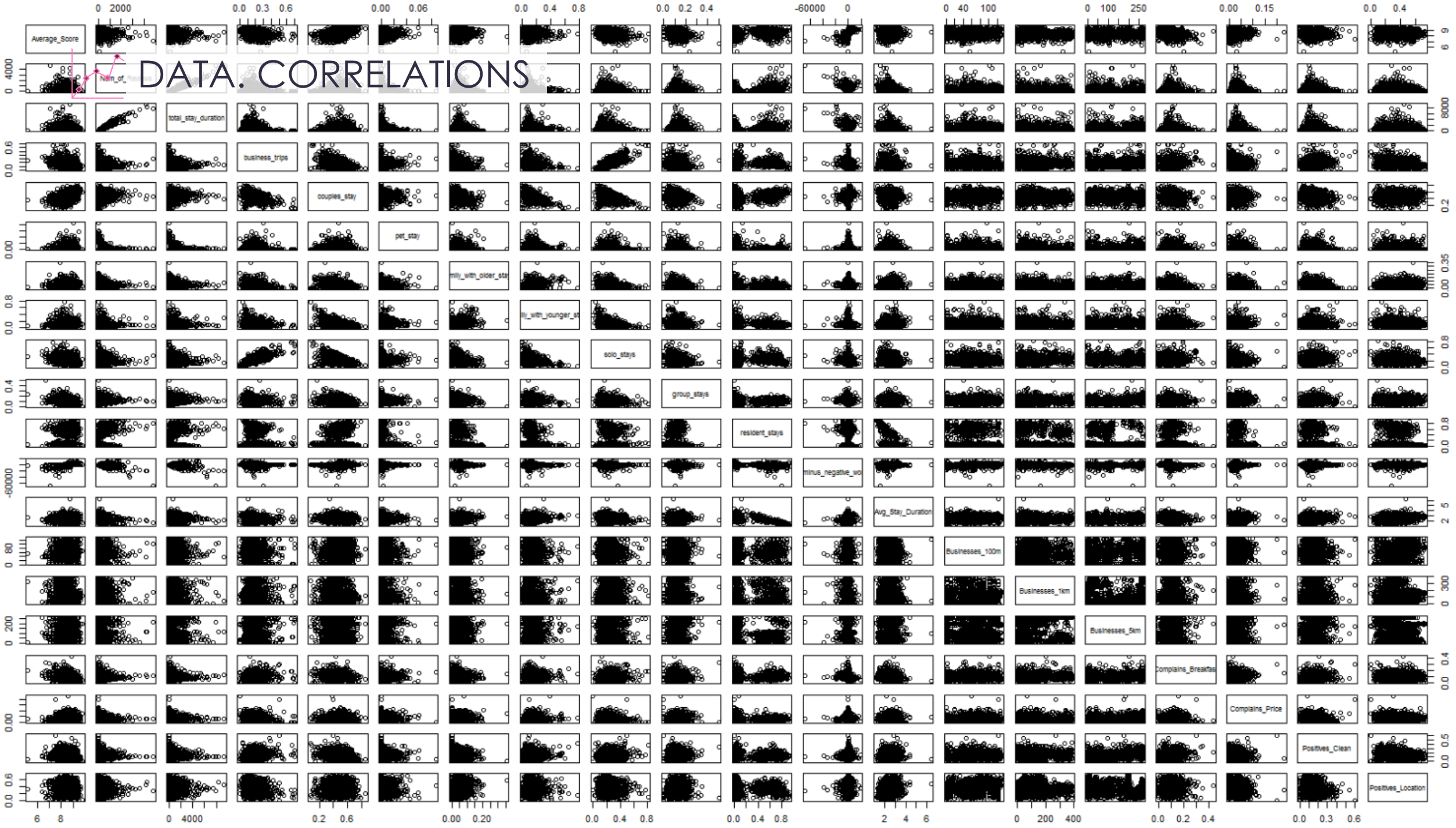
DATA. UNDERSTAND. VISUALIZE.

What matter for Business travelers?

Business travels mention **'breakfast' less**
However emphasize on '**friendly**', '**clean**', '**comfortable' & 'bed'**

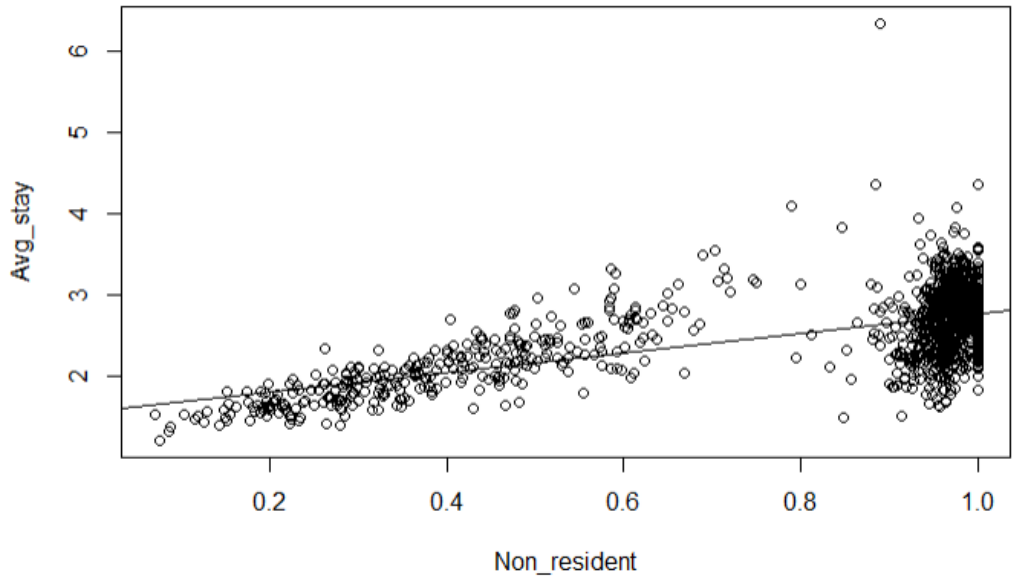
Although its mostly the same
Business travels mention **'bar' less**
However emphasize on '**Expensive**' & '**time**'

DATA. CORRELATIONS



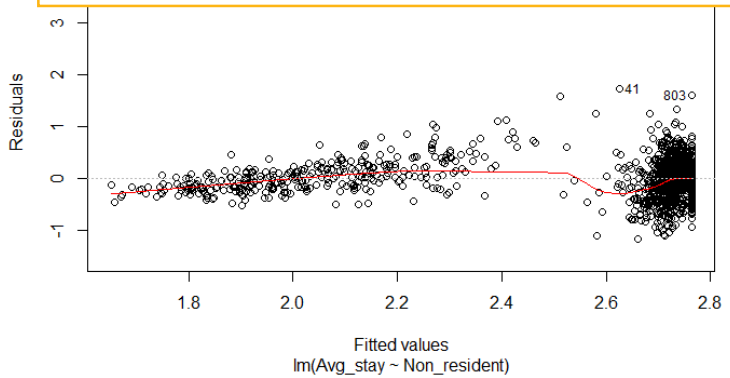
DATA. PREDICTION

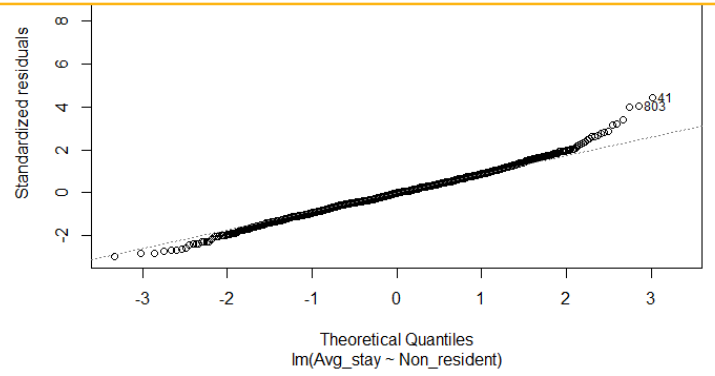
Average Length of Stay by number of Non-Resident Travelers

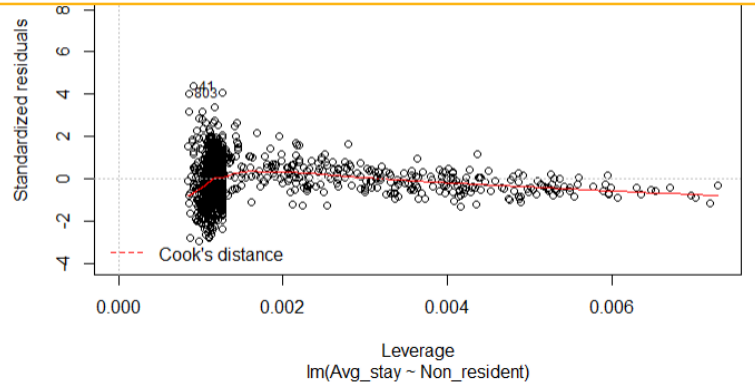



```
Call:
lm(formula = Avg_stay ~ Non_resident)

Residuals:
    Min      1Q  Median      3Q     Max
-1.1577 -0.2315 -0.0030  0.2250  3.7079

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.56768    0.03639   43.07   <2e-16 ***
Non_resident  1.19595    0.04261   28.07   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3934 on 1177 degrees of freedom
Multiple R-squared:  0.4009,    Adjusted R-squared:  0.4004
F-statistic: 787.7 on 1 and 1177 DF,  p-value: < 2.2e-16
```

Business Recommendation: Target non-residents with promotional offers for longer stay

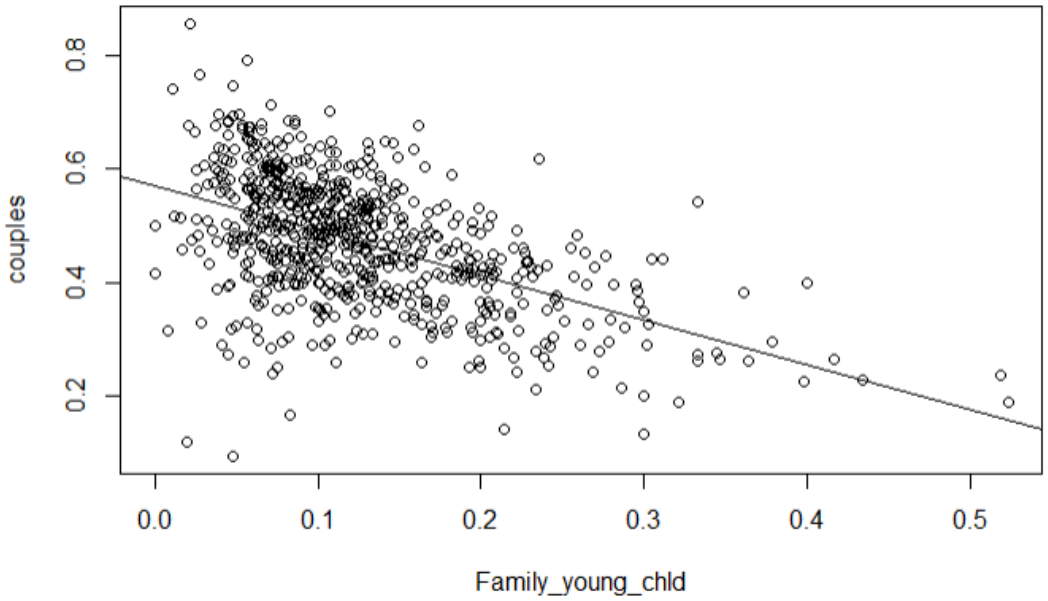






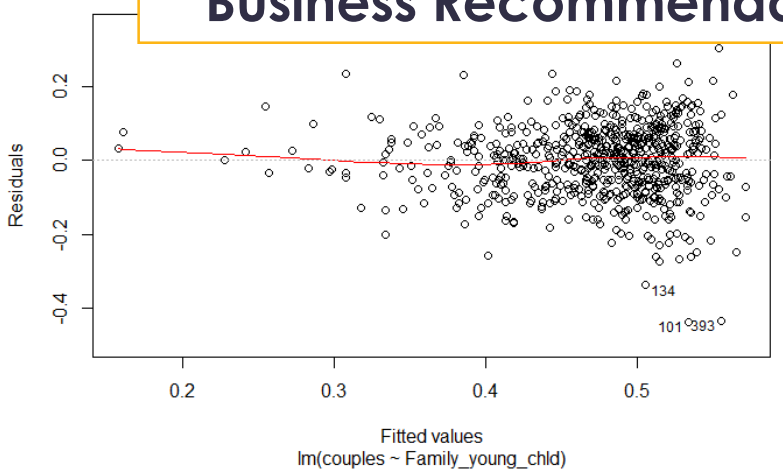
DATA. PREDICTION

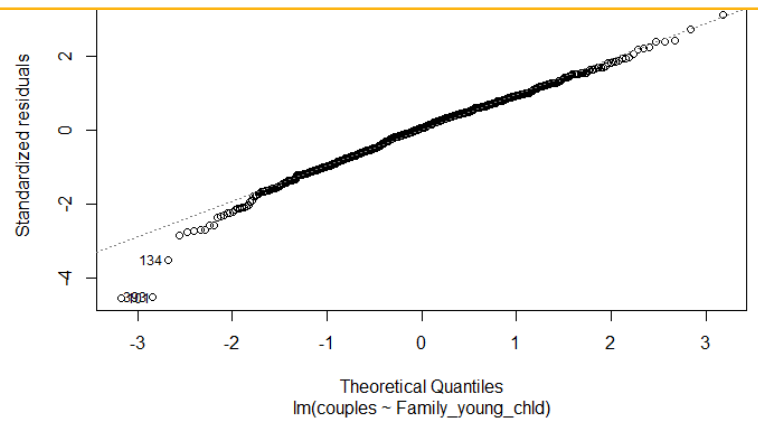
The More Families with Young Children <-> Less Couples
(At Hotels where average stay is higher than mean)

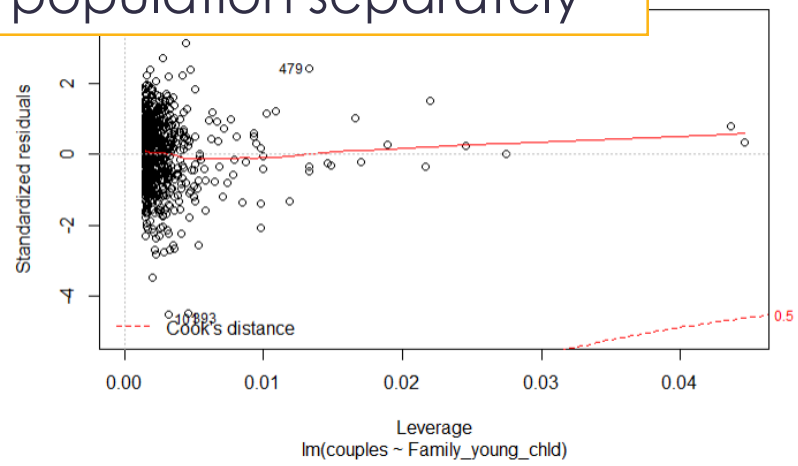



```
Call:
lm(formula = couples ~ Family_young_chld)

Residuals:
     Min       1Q   Median       3Q      Max
-0.43809 -0.06211  0.00704  0.06383  0.30354

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.570984   0.007381   77.36   <2e-16 ***
Family_young_chld -0.790683   0.050610  -15.62   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09685 on 671 degrees of freedom
  (140 observations deleted due to missingness)
Multiple R-squared:  0.2667,    Adjusted R-squared:  0.2656
F-statistic: 244.1 on 1 and 671 DF,  p-value: < 2.2e-16
```

Business Recommendation: Segment and target each population separately







The Team

Yanir Caliser

Head of Intelligence
& Partnerships

Noa Barbiro

Head of Product
& Strategy

Kobi Shamama

Head of Analytics
& Machine Learning

What's Next:

Social Media Sentiment Analysis

Recommendation | Closing-the-loop


To Be Continued...