

# Optimized Hotel Booking

Analyzing Hotel Reviews over Booking.com

kaggle™

Booking.com

yelp\*

Google Maps

 SO, WHERE WERE WE?

**Booking.com**

Around ~1.4K hotels in Europe

Over ½ Million Reviews

10

**“Fantastic experience”**

• Business trip

• Solo traveller

• Superior Double or Twin Room

• Stayed 1 night

+ This was one of the best hotels I have ever stayed at. The Location was superb and the decor outstanding. The staff were fabulous.

SO, WHERE WERE WE?

Booking.com



	Hotel Name
	Total Num. of Reviews
	Avg. Score
	Hotel Address
	<b>Is Hotel Holiday</b>
	Lat/Lng


**HOTEL**

	Country & City
	<b>Businesses 100M</b>
	<b>Businesses 1KM</b>
	<b>Businesses 5KM</b>

**LOCATION**

1 to 1



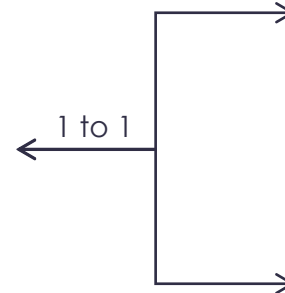
	Review Date
	Review Score
	Negative Review
	Positive Review
	Submitted from mobile


**REVIEW**

	Day of Week
	Day of Year
	Week of Month
	Week of Year
	Quarter (Season)


**REVIEW DATE**

1 to 1



	Reviewer Nationality
	<b>Guest Type</b>
	Total num. of Reviews

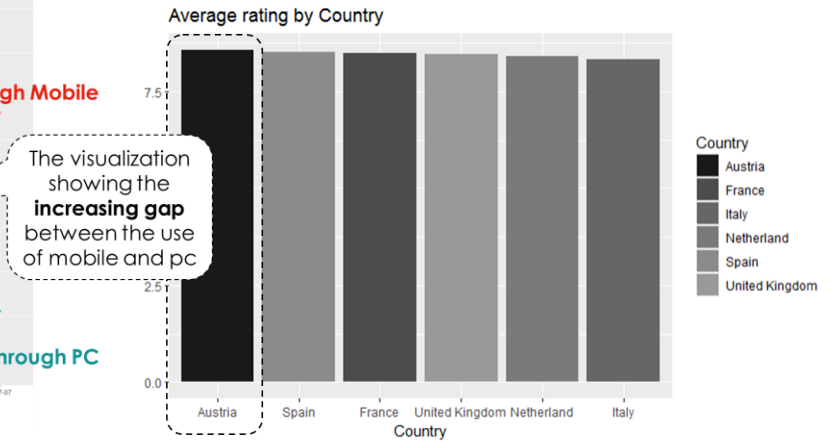
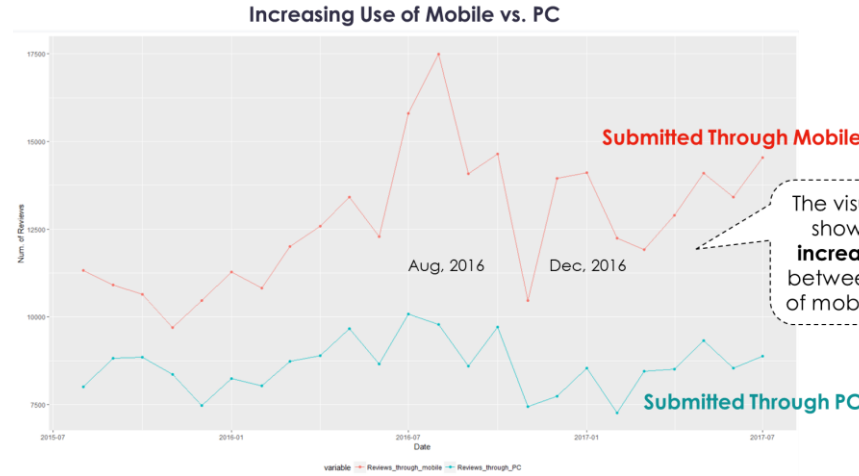
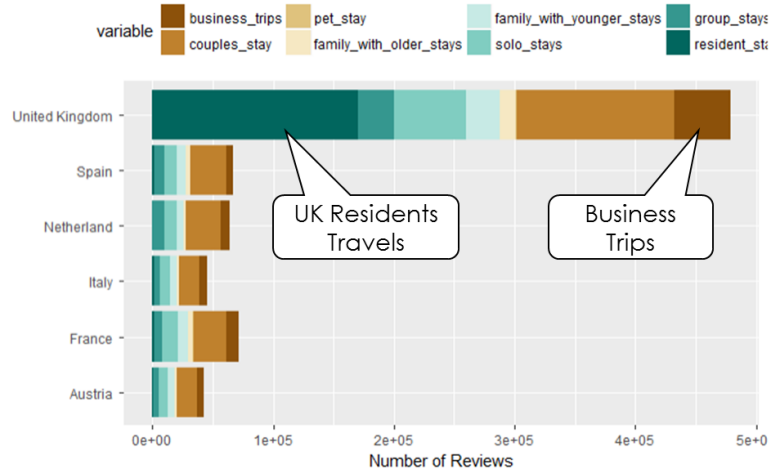
**REVIEWER**

	Room Type
	Room Type Level
	Bed Type
	Stay Duration (#nights)
	Is Business Trip

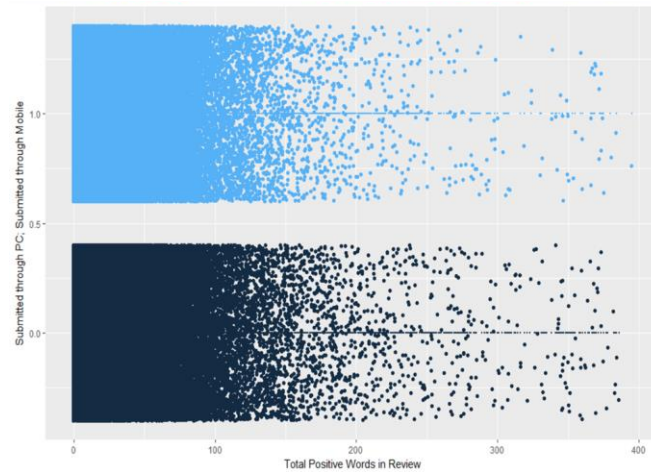
**TRIP / STAY**

Business / Leisure  
Resident / non-Resident  
Traveling with Pet  
Couples  
Solo traveler  
Family with older children  
Family with young children  
Group

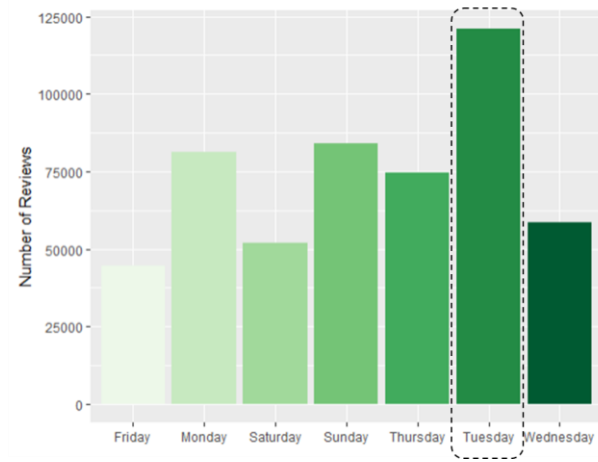
# EDA IN ACTION



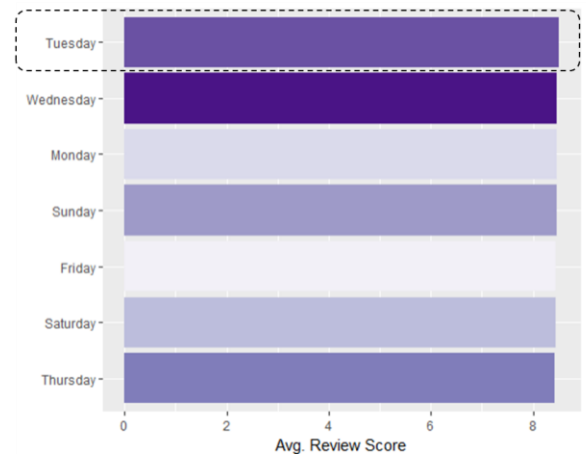
Slight **Less Use** of words for **Positive reviews** Through Mobile



Tuesday is the Most Popular day to submit a review



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





A high-angle, wide shot of a large crowd of people, likely at an airport or a large public event. The people are organized into several parallel lines or queues, separated by black stanchions and retractable belts. The crowd is diverse in age and appearance, with many individuals carrying luggage or bags. The floor is a light-colored, polished surface. The overall scene depicts a high-density, organized gathering of people.

# CLUSTERING





## CLUSTERING. **BUSINESS GOAL**

Increase Customer Engagement

### COMPETITION:



# TRUST

Relevance & Correctness

# CLUSTERING. BUSINESS PROBLEM

Automatically assign location tag to the hotel

The image displays two screenshots of hotel listings from a travel platform, with various annotations highlighting specific features and user feedback.

**Left Screenshot: DoubleTree by Hilton London – Docklands Riverside**

- Header:** "DoubleTree by Hilton London – Docklands Riverside" with a "Reserve" button and a "Couple friendly" icon.
- Address:** "265 Rotherhithe Street, Southwark, London, SE16 5HW, United Kingdom".
- Image:** A photograph of a modern bar area with a curved stainless steel counter and brick walls.
- Annotations:**
  - A yellow box labeled "Great location - show map" points to the address.
  - A yellow box labeled "Couples like what this property has to offer" points to the "Couple friendly" icon.
  - A white box labeled "beautiful views b comfortable room" is overlaid on the image.
  - A white box labeled "Staff" is overlaid on the image.

**Right Screenshot: Dorsett City London**

- Header:** "Dorsett City London" with a "Reserve" button and a "Couple friendly" icon.
- Address:** "9 Aldgate High Street, City of London, London, EC3N 1AH, United Kingdom".
- Image:** A photograph of a modern dining area with red chairs and wooden tables.
- Annotations:**
  - A yellow box labeled "Excellent location - show map" points to the address.
  - A white box labeled "Superb 9.1" with "1,051 reviews" and a blue box labeled "8.8" are overlaid on the image.
  - A white box labeled "Great bed, superb bathroom, love the opaque window in the shower." is overlaid on the image.
  - A yellow box labeled "Score from 402 business travellers" points to the "8.8" score.
  - A white box labeled "Excellent location!" with a blue box labeled "9.2" is overlaid on the image.

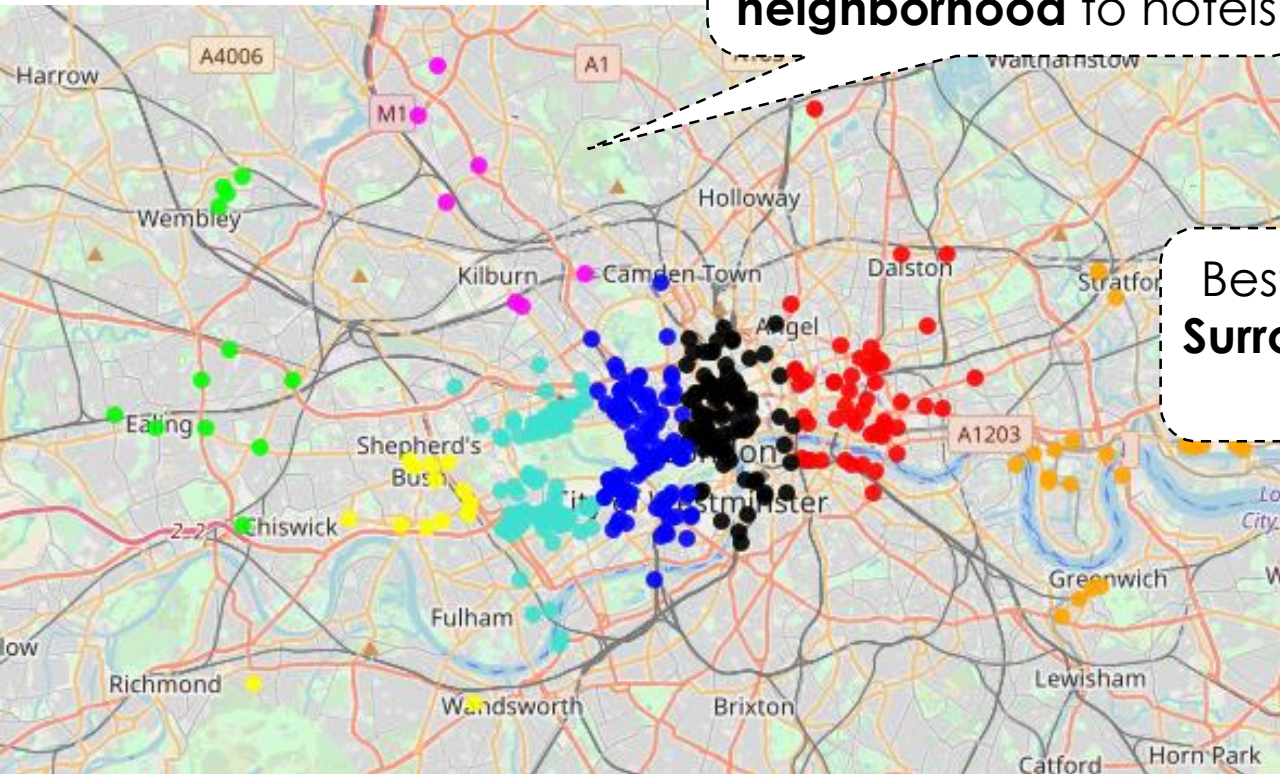




# CLUSTERING. METHODS

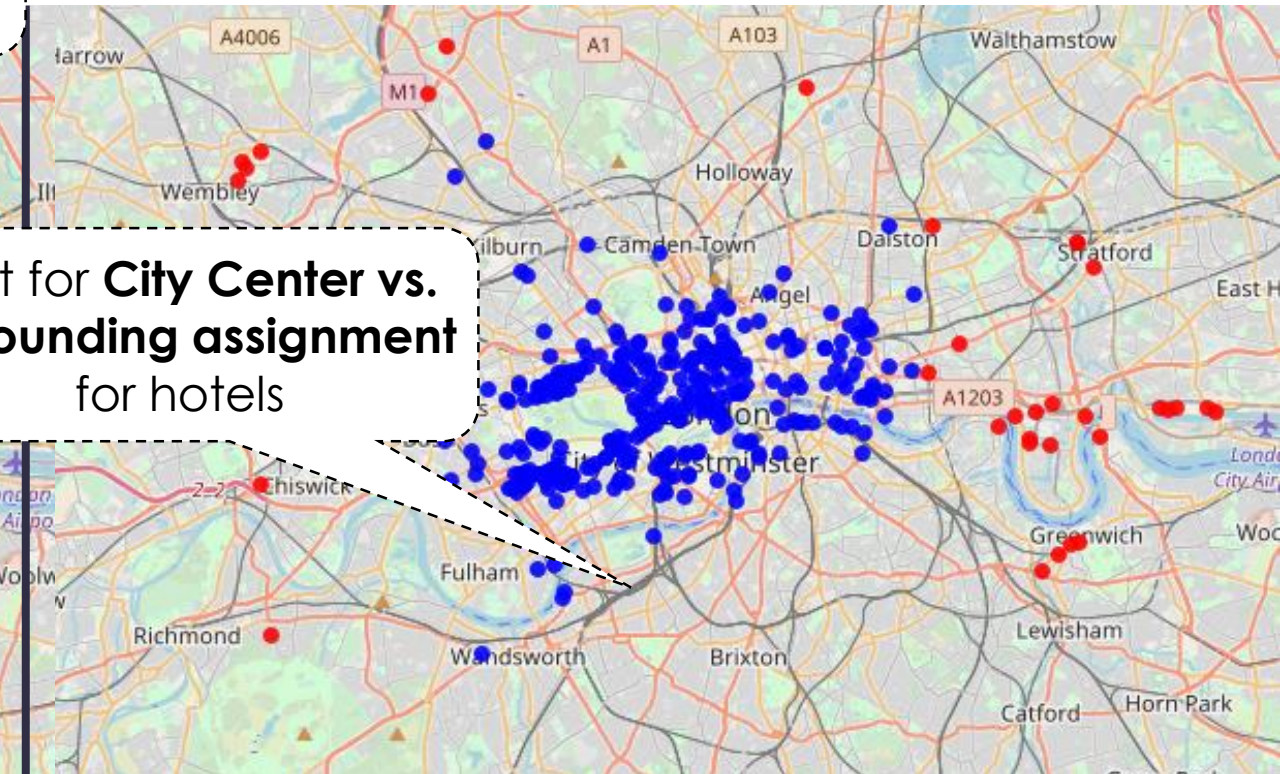
## KMEANS

Best for **automated assignment of neighborhood to hotels**



## DBSCAN

Best for **City Center vs. Surrounding assignment for hotels**



# WHAT ELSE?

## Machine learning is driving growth at Airbnb



Stephanie Pandolph

Jun. 16, 2017, 10:43 AM 4,571



FACEBOOK



LINKEDIN



TWITTER



EMAIL



PRINT

This story was delivered to BI Intelligence "E-Commerce Briefing" subscribers. To learn more and subscribe, please [click here](#).

Machine learning (ML) has had a "profound" effect

### Where Global Airbnb Users Are Switching From

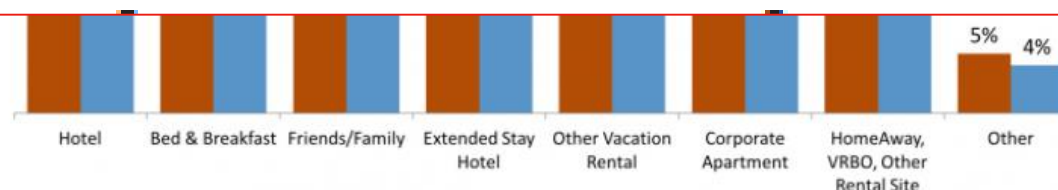
Airbnb leisure users, the past 12 months Airbnb leisure users, the next 12 months

47%

43%

- The company uses a machine-learned search ranking model to personalize results for guests. The model factors in guests' tendencies

Airbnb might look at whether customers favor specific types of décor in places they book. The company feeds more than 100 characteristics into the model, which then uses the data to identify patterns and personalize search rankings.



Source: Morgan Stanley

BI

BI INTELLIGENCE

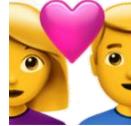


# CLUSTERING. **BUSINESS CHALLENGE**

Personalize hotel search results / recommendations



Yanir

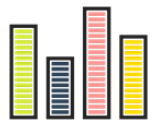


Kobi



Noa





# CLUSTERING. METHODS

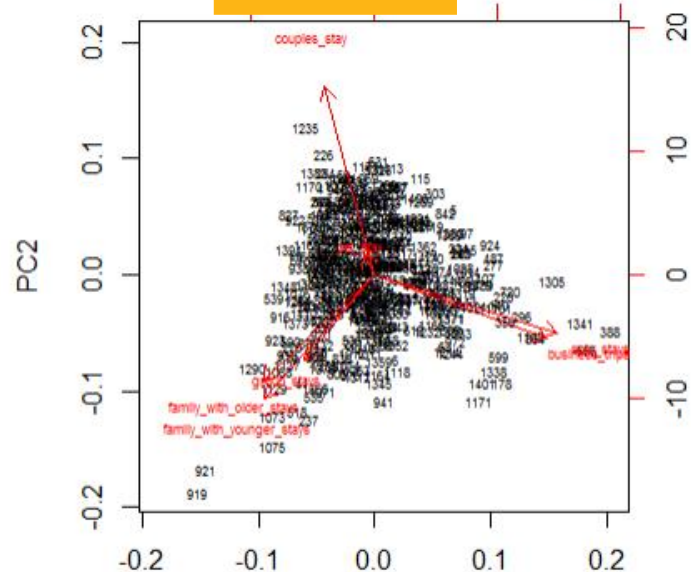
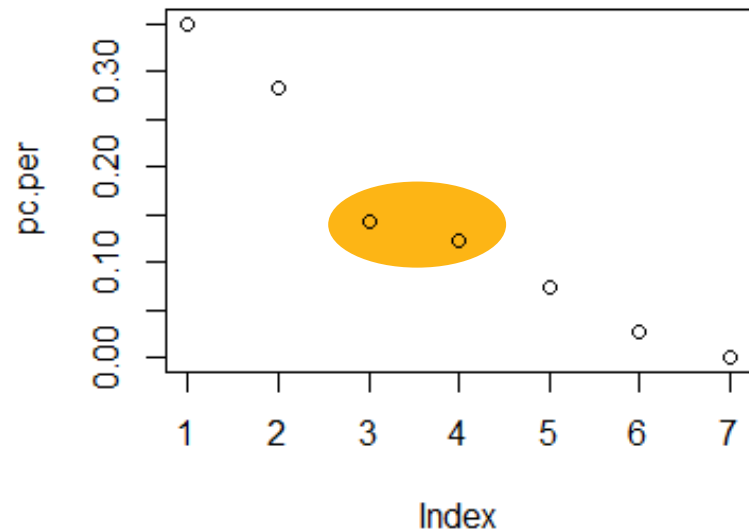
## Dimensionality Reduction - PCA

standard deviations (1, ..., p=7):

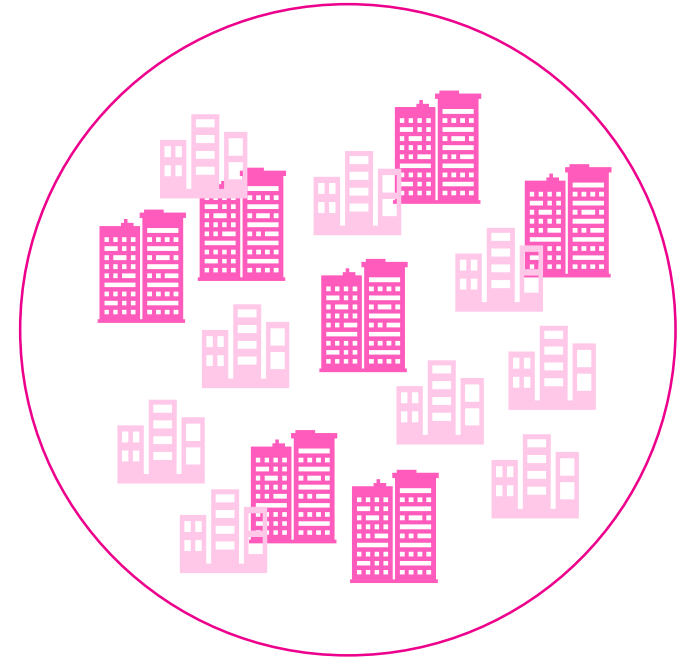
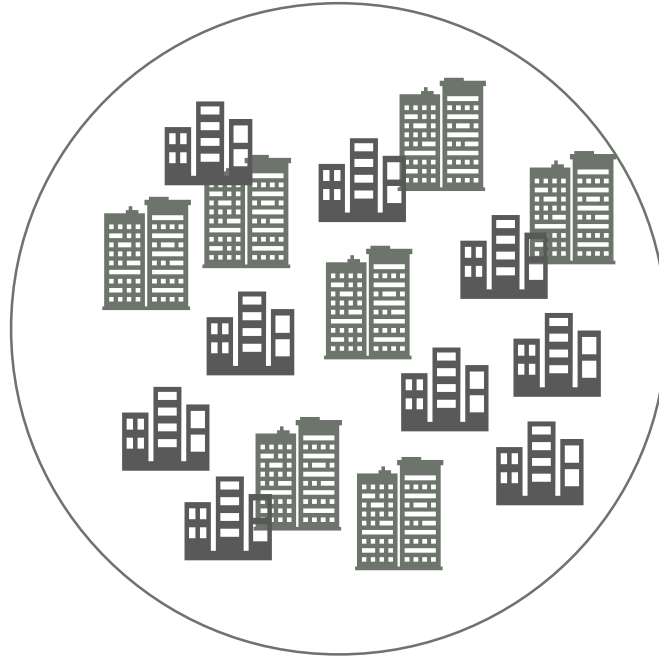
[1] 1.56640718 1.40685881 0.99908959 0.92565552 0.72493386 0.43140421 0.02144696

Rotation (n x k) = (7 x 7):

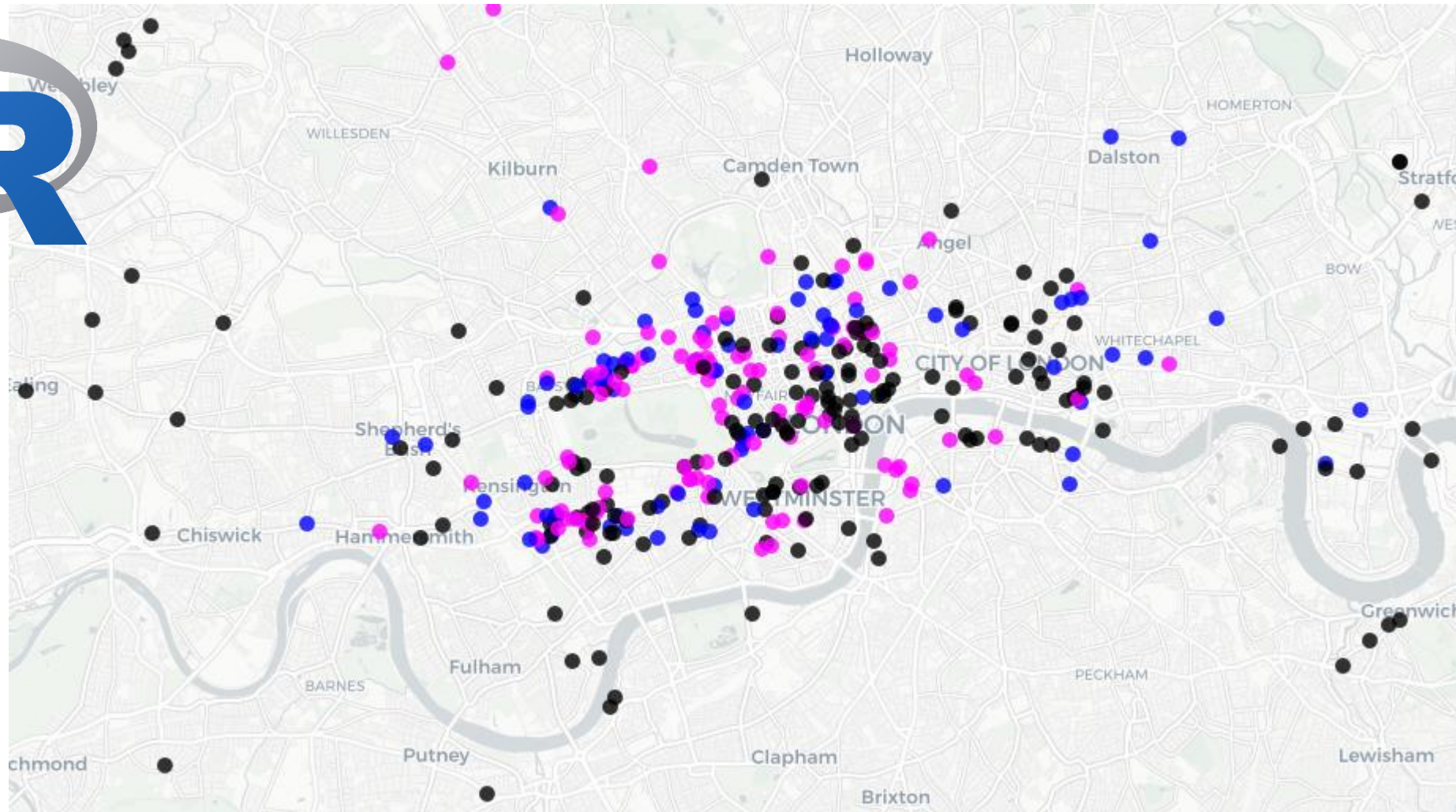
	PC1	PC2	PC3	PC4	PC5	PC6	PC7
🕶️ business_trips	0.55977527	-0.22907968	-0.06405424	0.08461212	0.05777924	0.78712880	0.001347057
💕 couples_stay	-0.16159414	0.67840677	0.05334131	0.03222923	0.08345737	0.30601295	0.639654996
👨👩👦 pet_stay	-0.03539911	0.07769089	-0.98933010	0.06462701	0.08737650	-0.04608224	-0.001449973
👨👩👦 family_with_older_stays	-0.36454270	-0.38116160	0.02728986	-0.27986336	0.77503813	0.12345083	0.163804791
👨👩👦 family_with_younger_stays	-0.35503287	-0.44382457	-0.10640842	-0.31828719	-0.61029915	0.19299941	0.393227788
🕶️ solo_stays	0.59072207	-0.21226768	-0.01386827	-0.08819196	0.09234030	-0.48132550	0.598179531
group_stays	-0.22670158	-0.30647822	0.04492096	0.89454869	0.01964406	-0.02230787	0.227065444



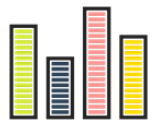
# CLUSTERING. **METHOD**



# CLUSTERING. ON MAP







## CLUSTERING. **BUSINESS OBJECTIVES MET**

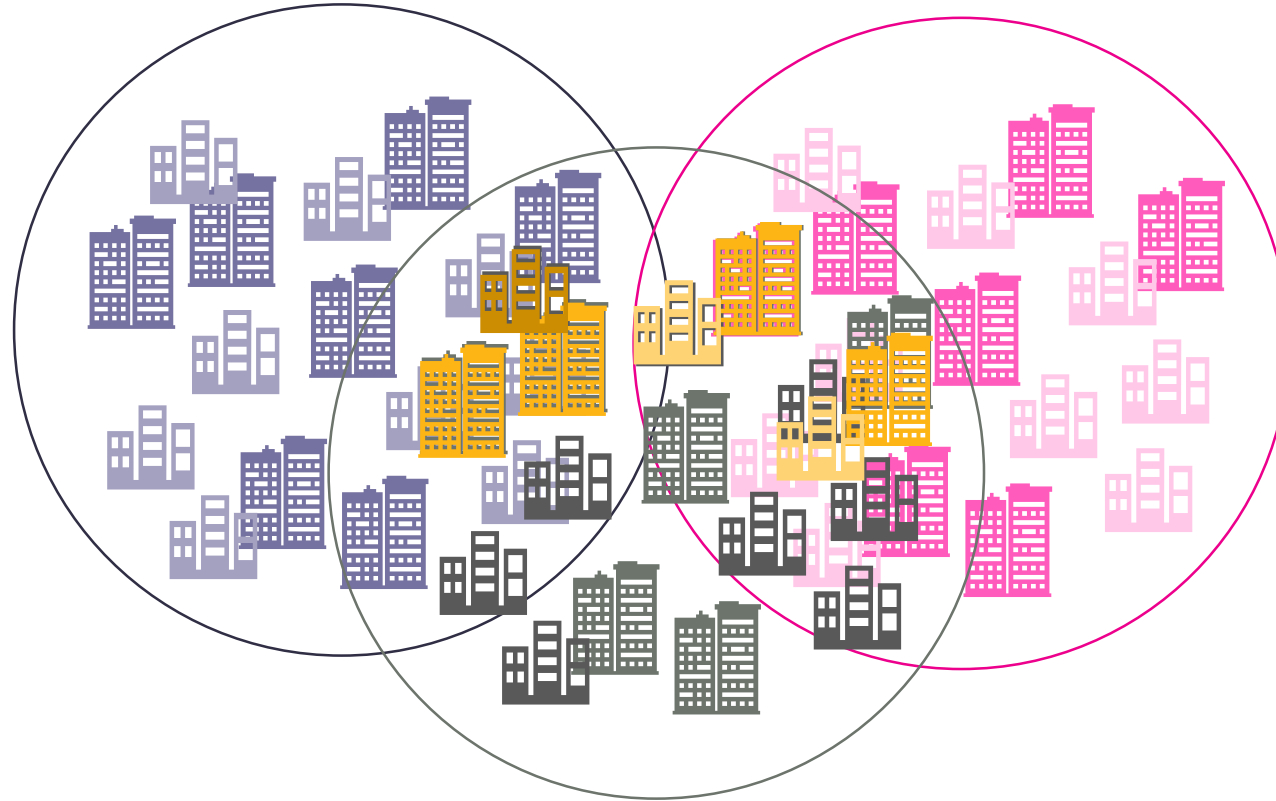
Which business problems / predictions we could fulfill?

1. Machine Learning model (clustering) to **personalize search results**
2. **Prediction model for Segmentation** correlated to geo-location
3. **Recommendation engine for hotels by proximity** (when one is unavailable)



## CLUSTERING. **BUSINESS SOLUTION – REVIEW MODEL**

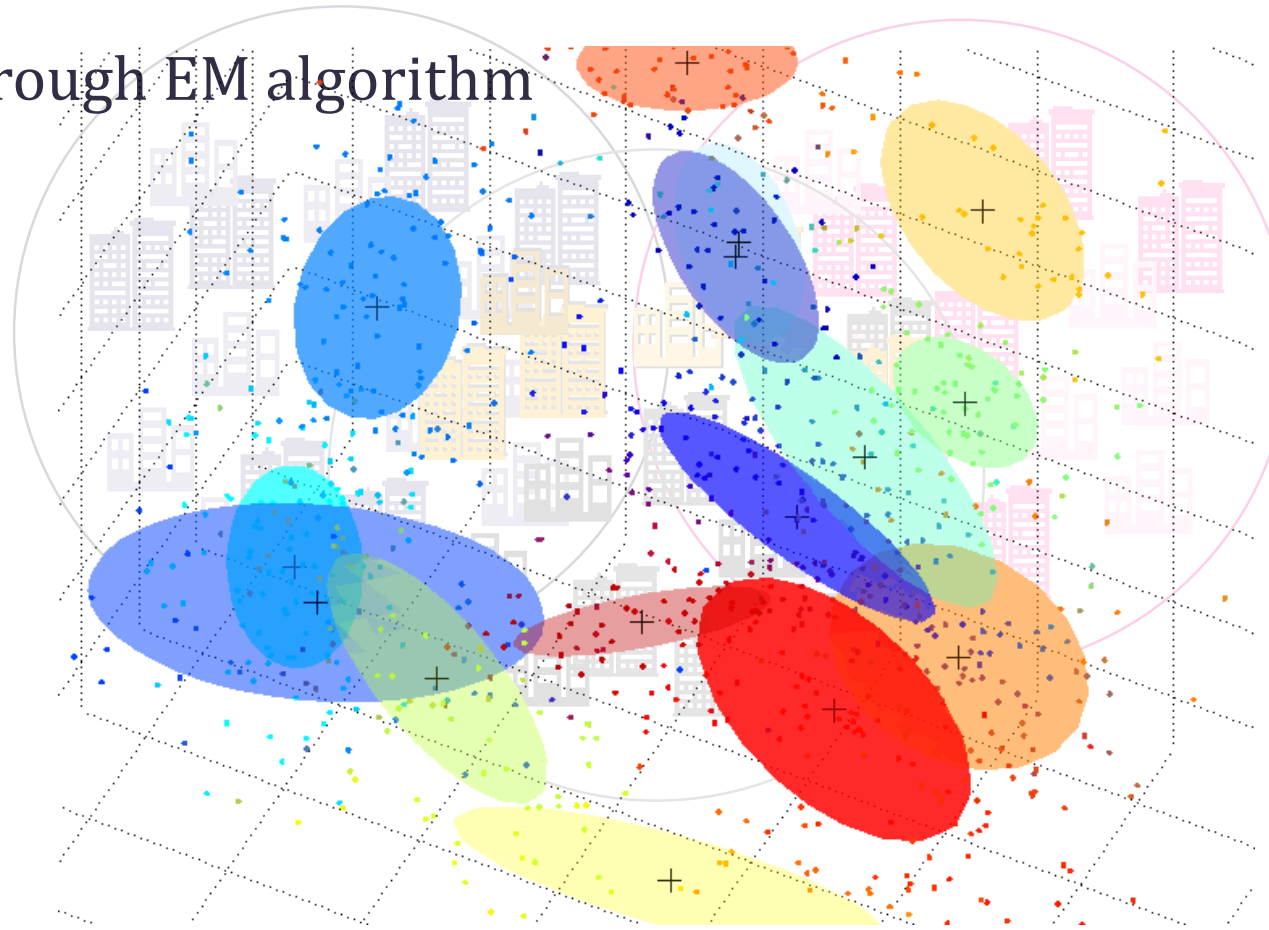
What are we missing with the previous algorithm?





## CLUSTERING. BUSINESS SOLUTION – REVIEW MODEL

Suggested approach: Maximization of the search results using Gaussian Mixture Model (GMM) through EM algorithm







To Be Continued...