

Hospitality Intelligence

Extracting Insights
from European Hotel

Reviews
BA820 Team 09

Eric Bai
Olimpia Borgohain
Chen He
Sricharan Mahavadi
Min Xu



INDEX

Structure

01. Introduction

02. Data Description

03. Market Basket
Analysis

04. Recommendation
System

05. Clustering

06. Sentiment Analysis

07. Conclusion

08. Challenges &
Next Steps



Introduction

- Motivation
- Problem Statement
- Business Values

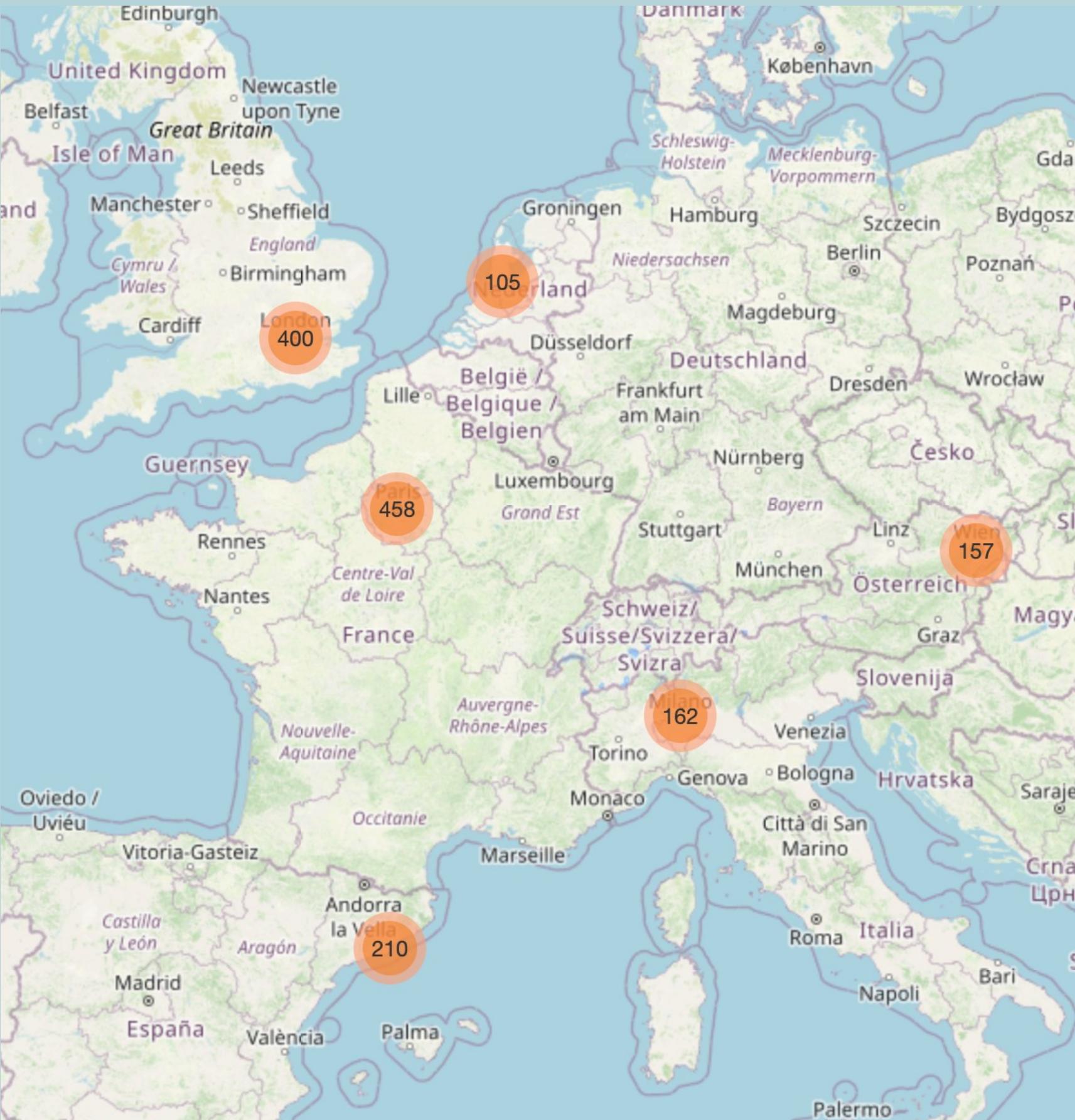
Data Description

- ✓ Hotel Geographical Spread
(Google Maps API to impute missing data)
 - 1492 Hotels in total
 - Majority in France and UK
 - Rest dispersed across Spain, Italy, Austria and Netherlands

- ✓ Review Time Distribution
 - A total of 515k Review
 - 3 years review ranging from Y2015-Y2017
 - Seasonal trends

- ✓ Review Score Distribution
(Filtered out outliers by capping ratings at 10)
 - Median around 8.5
 - Most scores between 7.5 and 9.5
 - Outliers between 3 and 5

- ✓ Reviewers Nationality Distribution
 - UK reviewers at 66.3%
 - Followed by the USA at 9.6%, Australia at 5.9%
 - Other countries ranging from 2.1% to 4.0%



PART 1

Market Basket Analysis

- Treated each dataset row as an individual transaction,
- Applied the Apriori rules, with hotel names as antecedents to highlight commonly associated tags in reviews.
- Developed a function allowing customers to input a hotel name to see the most frequently mentioned tags.

antecedents	consequents	support	confidence	lift
{One Aldwych}	{ Aldwych Double Twin Room }	0.007	0.522	73.529
{Apex Temple Court Hotel}	{ City King Room }	0.074	0.713	9.643
{Apex Temple Court Hotel}	{ Couple }	0.072	0.693	1.517



PART2

Recommendation System

Text/Reviews

breakfast
location
staff
bed
hotel
room
Tag : ✓ station

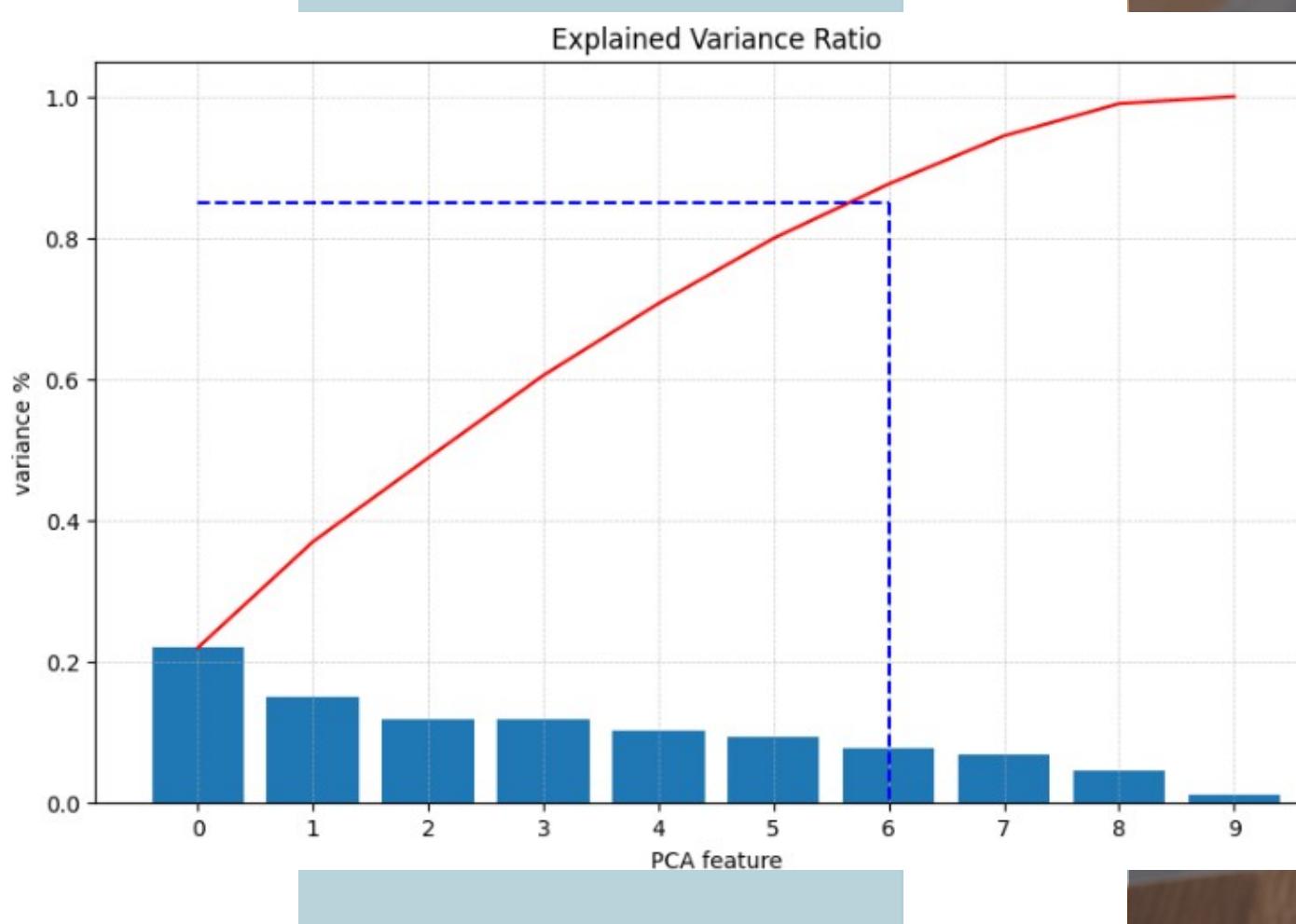
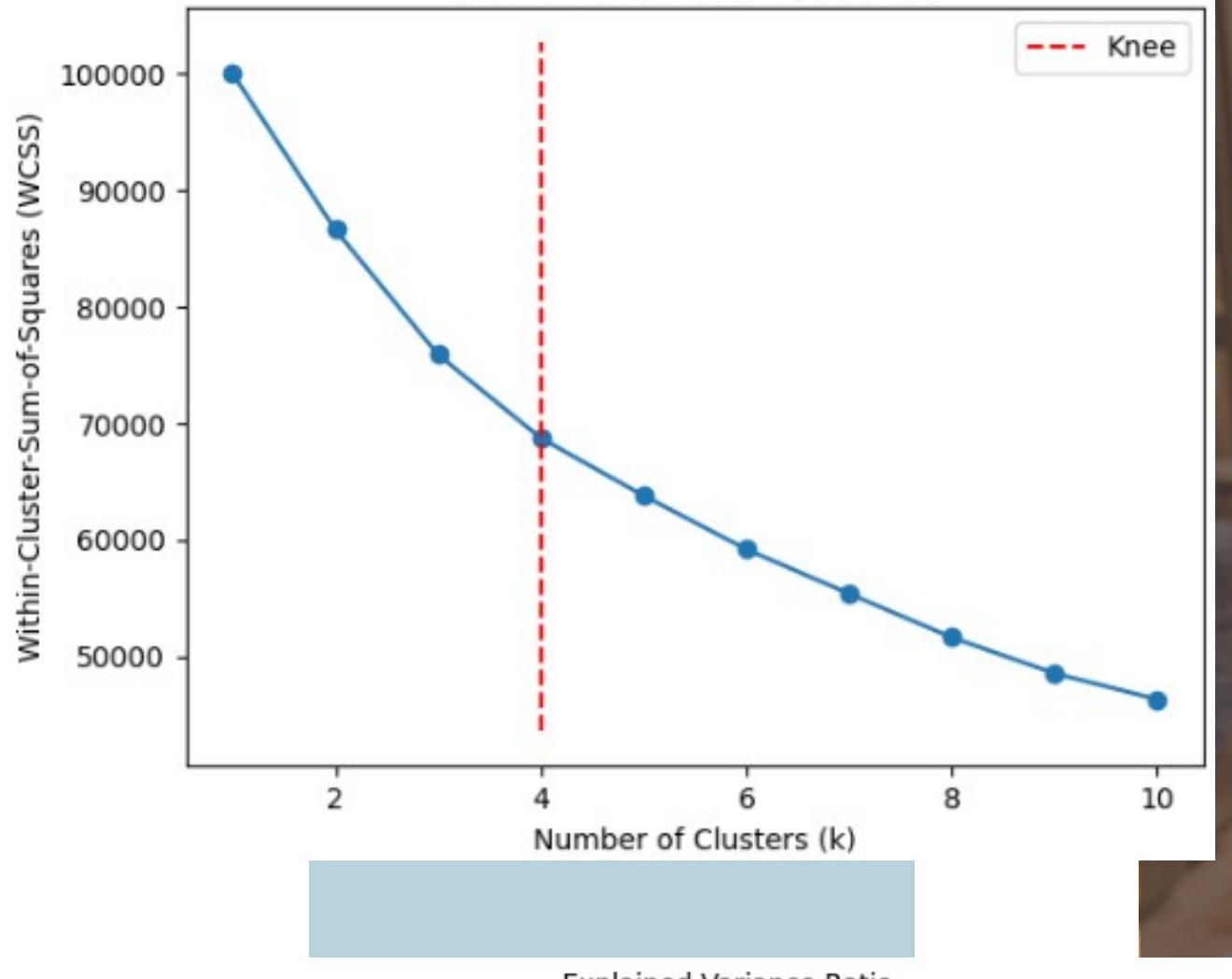
✓ 0.0s

Recommend Hotels

Recommended hotel: the grosvenor
Recommended hotel: nh collection amsterdam barbizon palace

Tags

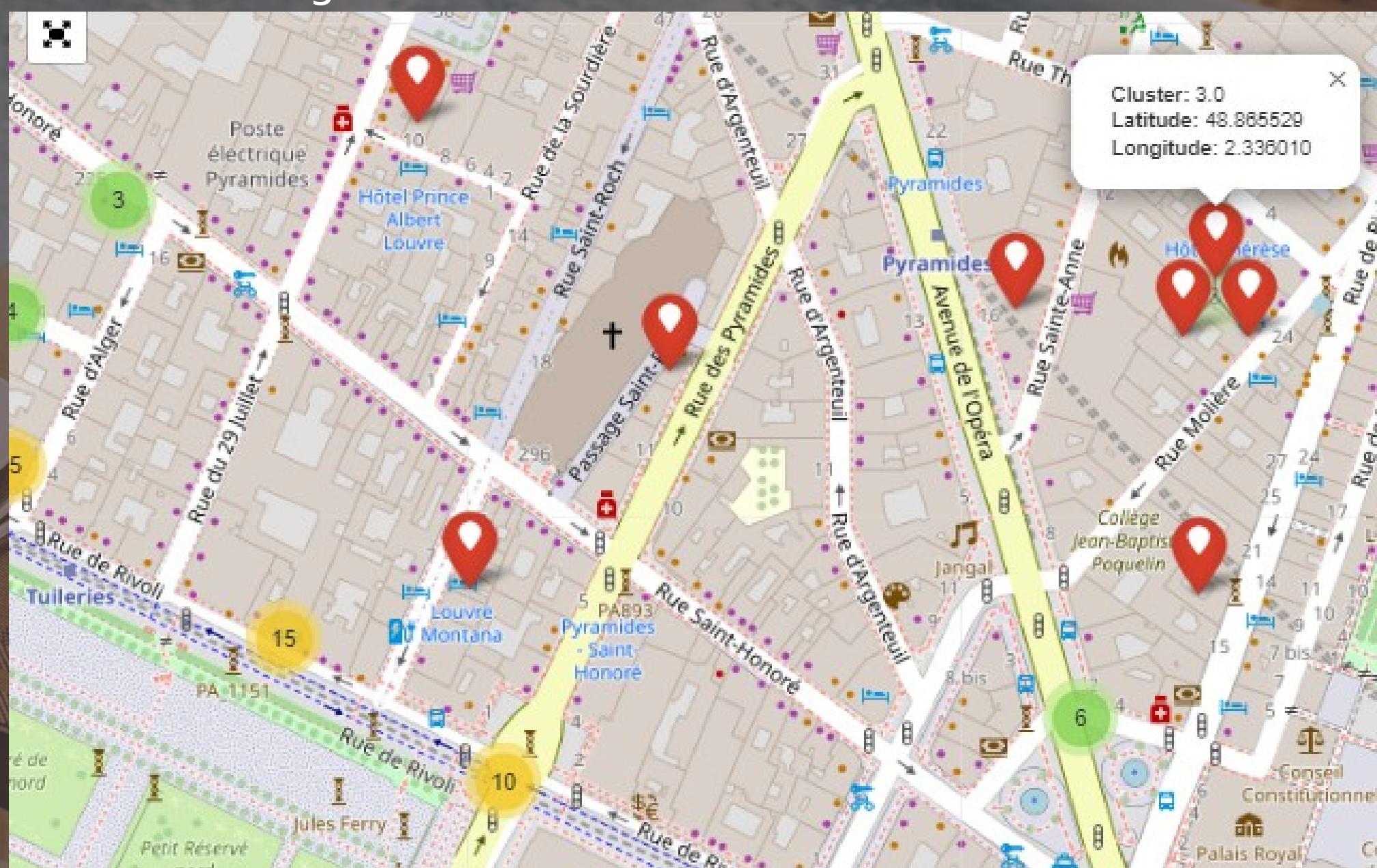
Elbow Method for Optimal k



PART 3

Clustering

- K-means Clustering
- K-means++ Clustering
- Hierarchical Clustering
- DBSCAN Clustering



PART 4

Sentiment Analysis

STEP 1

Preprocess the Reviews

- Convert all text to lowercase
- Remove punctuation, numbers, other non-letter characters
- Tokenize using NLTK's word tokenizer
- Remove stop words
- Lemmatizing

STEP 2

Text Representation

- Use Gensim library to load the "word2vec-google-news-300" model
- Define a Function for Text to Word2Vec Conversion
- Calculate the mean of vectors, serving as the numerical representation
- Convert the clean, tokenized text of each review into a numerical form

STEP 3

Obtain Sentiment Embeddings

- Defined Lists of Sentiment Words
- Calculate the centroid of sentiment
- Obtain sentiment embeddings for positive and negative Words



PART 4

Sentiment Analysis (cont.)

STEP 4

Calculate Similarity Scores

- Calculate Cosine Similarity
- Values closer to 1 indicate greater similarity

STEP 5

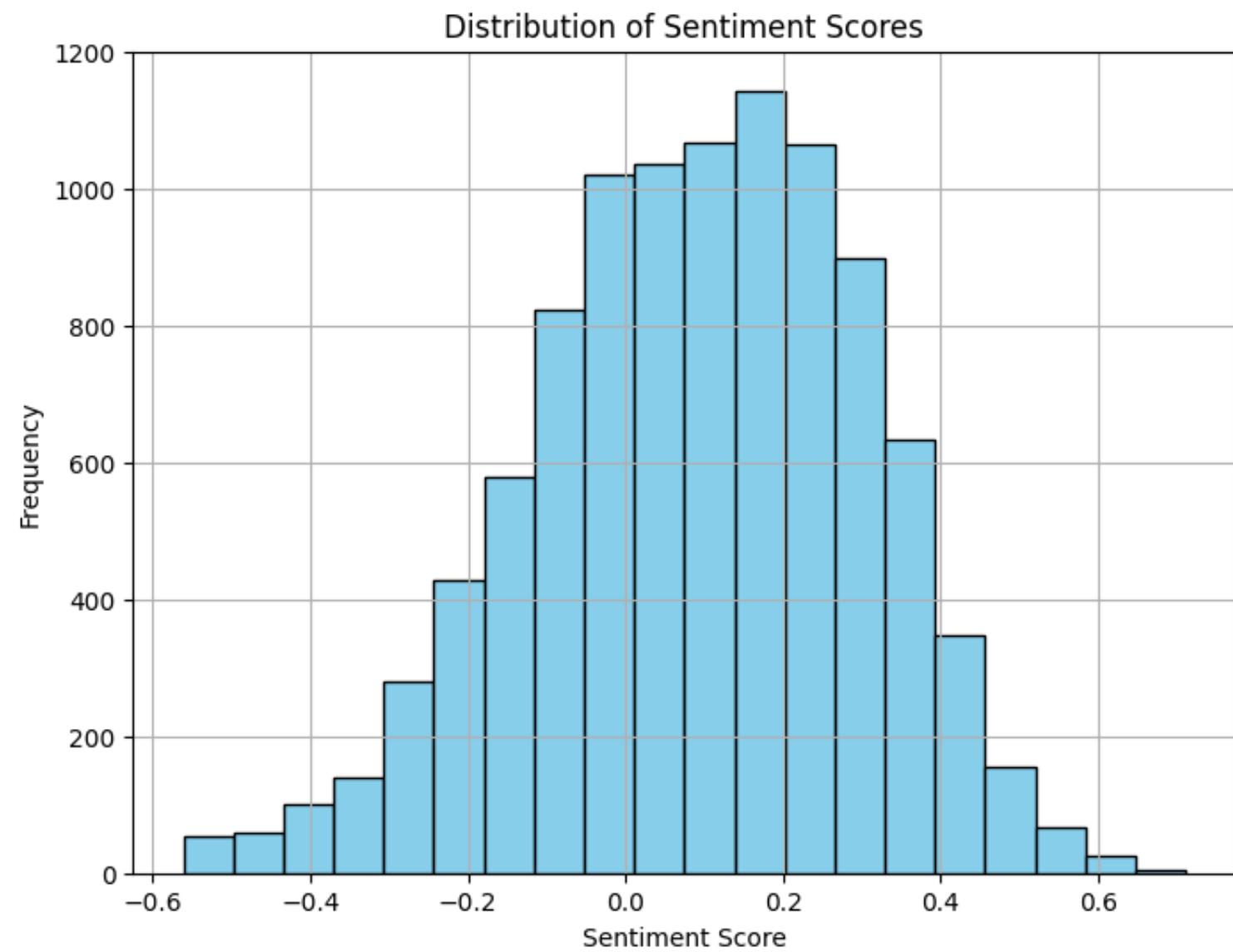
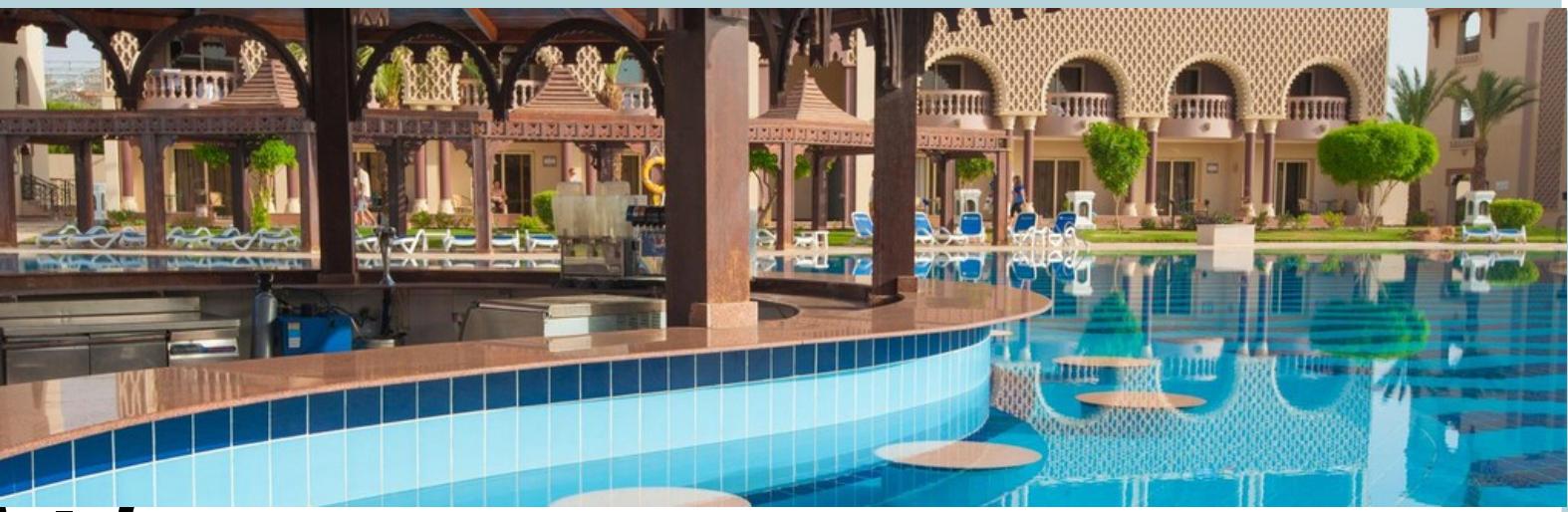
Compute Sentiment Scores

- Calculation of Sentiment Scores by measuring how much closer a review is to positive sentiment relative to negative sentiment

STEP 6

Obtain Sentiment Embeddings

- Overall Sentiment: Mean 0.089 - a slightly positive overall sentiment
- Sentiment Range: Range from 0.561 (most negative) to 0.711 (most positive)



Challenges

Lack of
Computing
Power

Tuning Apriori
Rules (RS)

Word2Vec
(SA) Taking
Too much
Time



Conclusions

Hotel Review
Dataset

Sentiment Analysis

Clustering

Recommendation
System

- Generally positive sentiment for hotels.
- Reviews mostly mildly positive.
- Wide score range: negative to positive.
- Diverse guest experiences.

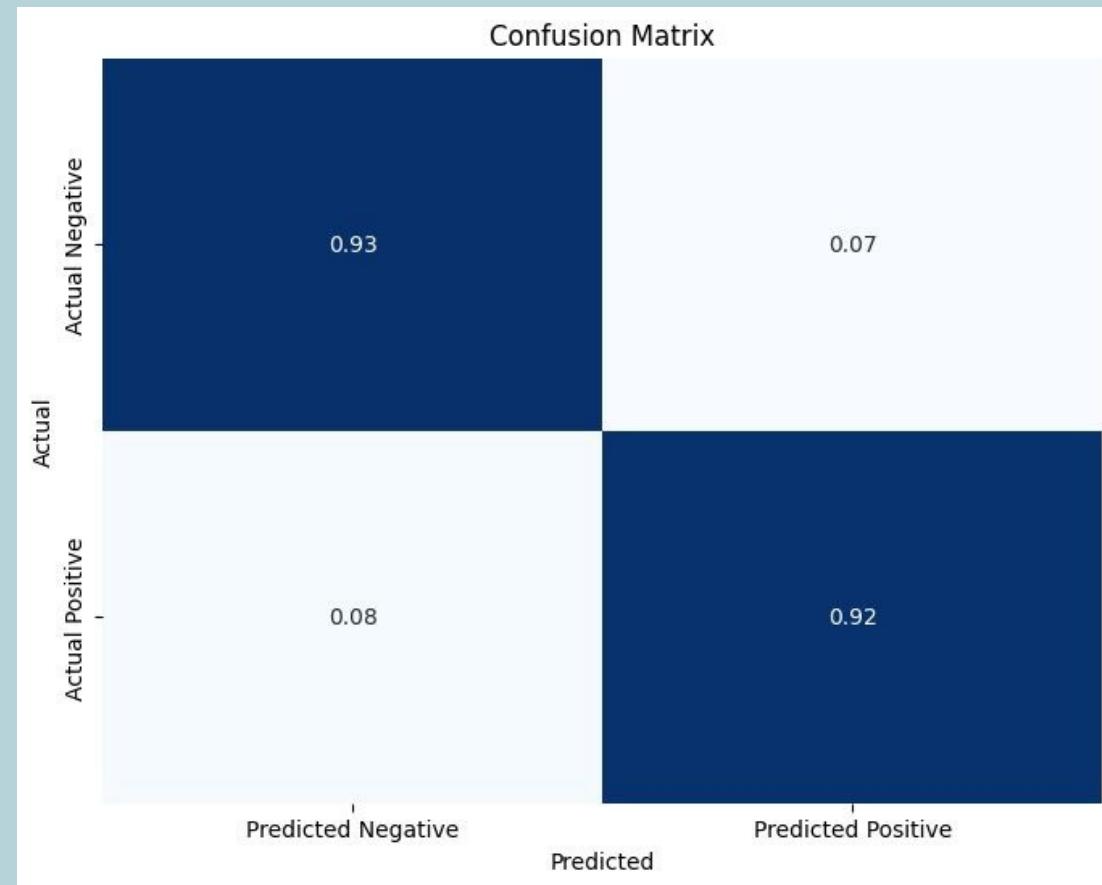
- Applied PCA for dataset's high dimensionality reduction.
- Found only 6 features needed to capture 85% variance.

- Accomplished a Market Basket Analysis Based RS
- Accomplished a Sentiment Analysis Based RS



Thank You

Appendix



City: ▼

Rating:

Recommended Hotel:

Name: Le Narcisse Blanc Spa
Address: 19 Boulevard De La Tour Maubourg 7th arr 75007 Paris France
Average Score: 9.5
Reason for Recommendation: This hotel has a high average score and positive sentiment based on customer reviews.

```
↳ /usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will no
    and should_run_async(code)
[nltk_data]  Downloading package punkt to /root/nltk_data...
[nltk_data]  Package punkt is already up-to-date!
Enter your review: The Hotel is awesome
The review is positive.
```