CUSTOMER REVIEW
ANALYSIS

For Hôtel Trianon Rive Gauche

Object :
Analysis of customer reviews to gain insights into key issues impacting hotel ratings

Gabriel CHAIX
Malo LEROY
Tiberio ZOLZETTICH

10/12/2024

Introduction

In the hospitality industry, customer satisfaction is critical for maintaining a positive reputation and attracting new guests. Online reviews provide valuable insights into guests' experiences, offering hotels an opportunity to identify strengths and address areas of improvement. This report presents an analysis of customer reviews for Hôtel Trianon Rive Gauche, focusing on uncovering key issues that impact overall guest satisfaction and hotel ratings.

By analyzing customer feedback, we aim to gain a deeper understanding of the recurring problems highlighted by guests and provide actionable recommendations that can enhance the hotel's services. The insights gained from this analysis will not only help improve the guest experience but also contribute to a higher overall rating on platforms like Booking.com and TripAdvisor, ultimately boosting the hotel's competitive position in the market.

Data Overview

The dataset used for this analysis consists of customer reviews collected from Booking and Tripadvisor for Hôtel Trianon Rive Gauche. The data was scraped programmatically using one python scraper for each website.

The Booking dataset contains a total of six columns that capture key aspects of each review. To ensure customer privacy and maintain unbiased analysis, personally identifiable information such as names has not been included in the dataset.

The six columns of the dataset are as follows:

- **Content**: The title of the review, which is often correlated with the score provided by the customer (e.g., a score of 10 typically has a title like "Excellent"). However, the content of this field can be customized by the user.
- **Score**: The grade allocated to the hotel by the customer.
- **Country**: The country of residence of the customer, providing a geographical context for the review.
- **Positive**: A section that highlights positive aspects of the customer’s experience.
- **Negative**: A section detailing negative feedback, which is the focus of this analysis as it provides valuable insights into areas where the hotel can improve.
- **Date**: The date when the review was published.

The Tripadvisor dataset contains only two columns that are enough to complete the content of Booking:

- **Review**: The whole review posted by each customer.
- **Score:** Same as previous.

We then preprocessed and cleaned these two datasets apart from each other. Since the objective of this analysis is to identify issues that negatively impact the hotel’s ratings, we chose to focus only on the negative reviews. To do so, we used the "Negative" column of Booking dataset and for the reviews

of Tripadvisor, we used the TextBlob library. This library enables to compute several NLP functions like the sentiment analysis. That's why we put all the reviews one by one into the sentiment analysis to keep only the negative and neutral reviews. We went from 583 reviews and an average score of 7.8 to 216 remaining reviews having an average score of 6.4.

As the hotel attracts an international clientele, the reviews are in multiple languages. For the purpose of the analysis, all necessary reviews have been translated into English to create a uniform data structure for better analysis and insight generation. For the rest of the preprocessing, we created a 'cleaned_review' column following the above process:

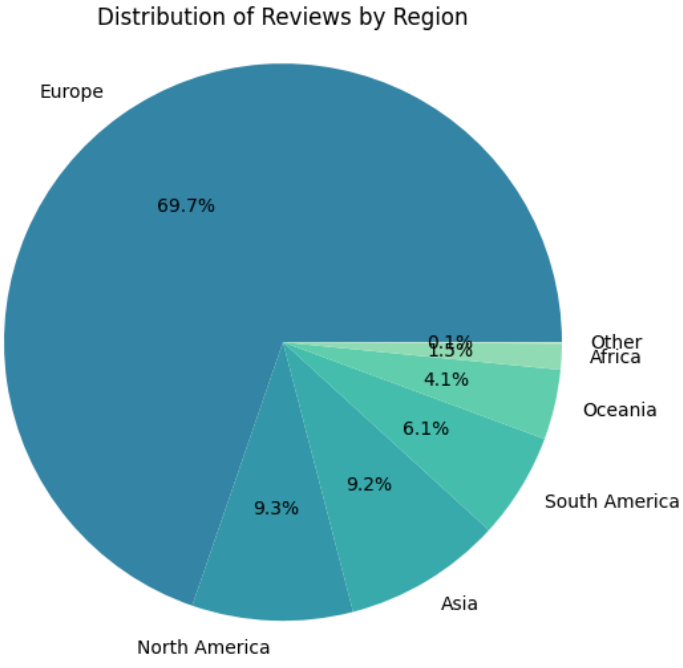
- Lowering text
- Keeping only alphabetic characters: deleting punctuation and digits
- Tokenizing: convert into lists of words using word_tokenize from nltk library
- Lemmatizing: using WordNetLemmatizer from nltk
- Deleting stop words

We then could concatenate both dataframes to obtain a clean dataset composed of two columns : the score and the tokenized and preprocessed reviews.

Analysis and Results

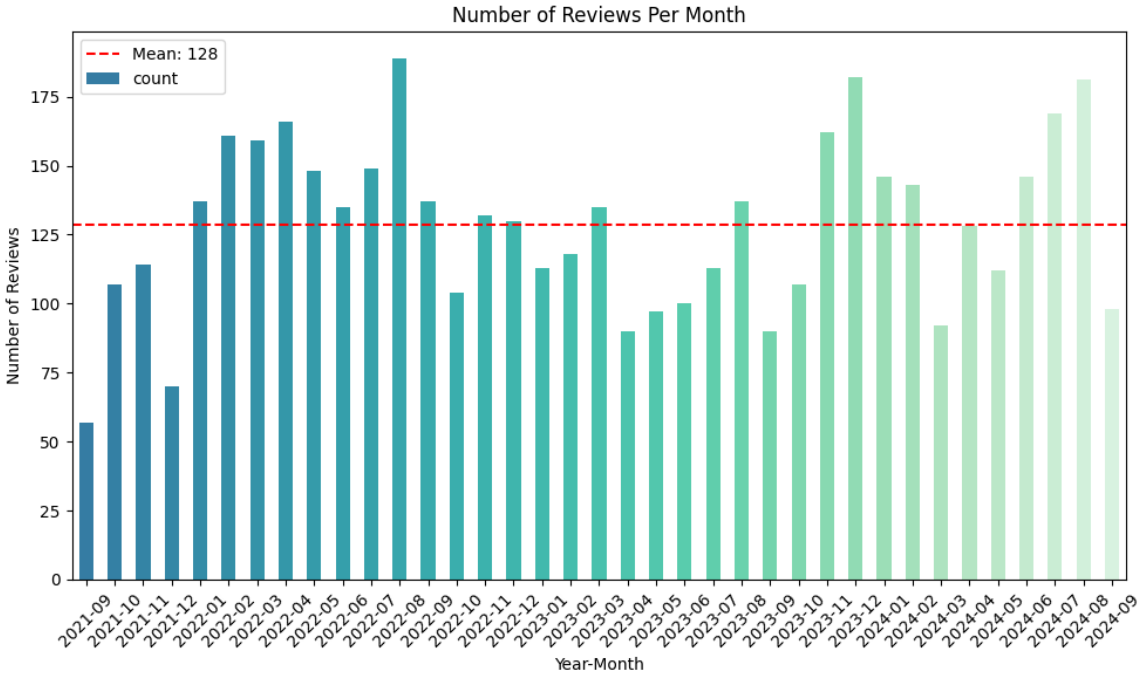
Although geographic timed data is not essential to understanding the hotel's performance, it provides valuable context about its clientele. The following pie chart explores the geographic distribution of the hotel's guests to uncover potential regional trends and preferences.

With 118 different countries represented, Europe accounts for the majority with more than two-thirds of the reviews, while other regions like North America, Asia, South America, and Oceania each contribute smaller portions. This highlights the hotel's predominantly European customers based in the world.



The histogram titled "Number of Reviews Per Month" shows the distribution of reviews received each month. The chart spans from September 2021 to September 2024, with the red dashed line indicating the average number of reviews per month, which is 128.

Fluctuations in the number of reviews are apparent, with some months, such as August 2022, exceeding 175 reviews, while others fall below the average. This variability could be attributed to seasonal trends or other factors affecting guest attendance and review submissions.



Rooms are too small

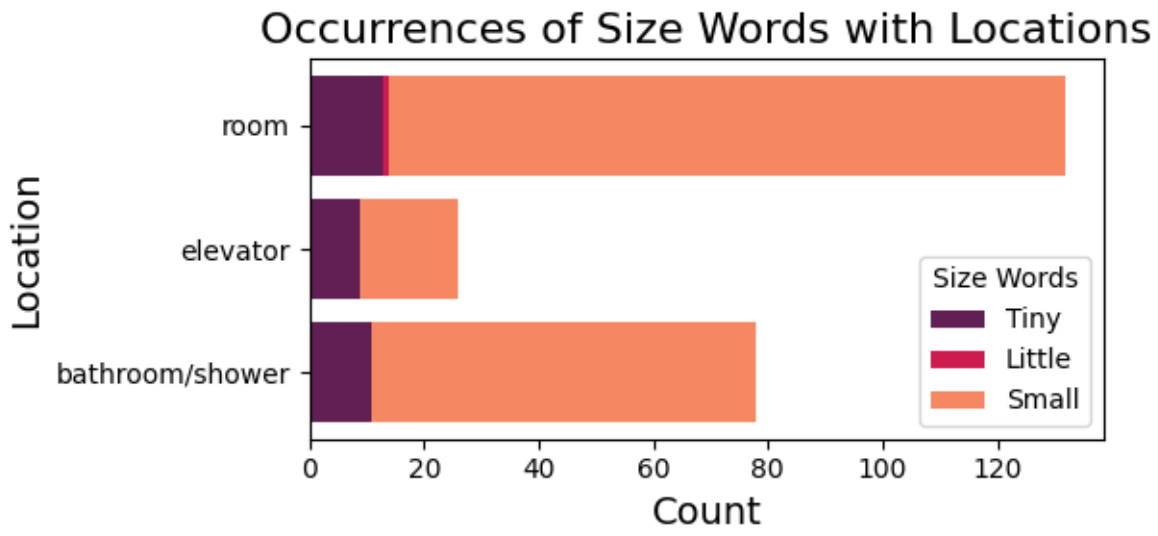
It is now time to begin the analysis of the negative reviews. Naturally, we first wanted to check what are the most common words and bigrams in the dataset. This is the best way to understand a little more about common thinking. On the left (respectively on the right) you can appreciate the most used single words (respectively bigrams) in the dataset.

The lexical field clearly stands about one thematic, narrowness of the rooms. Indeed, we observe a majority of “small room”. As the bigram identification is made on the rid-of-stopwords dataset, we find quite weird spellings like “room small” that stands for “the room is small”. We can hold the same conclusions for the bathroom and the elevator.

	Bigram	Frequency
0	room small	132
1	small room	120
2	air conditioning	69
3	small bathroom	50
4	bathroom small	48
5	star hotel	48
6	room little	40
7	breakfast room	39
8	room bit	36
9	reception staff	35
10	room bathroom	34
11	little small	31
12	hotel located	31
13	eiffel tower	29
14	bit small	27
15	th floor	27

	word	frequency
0	room	183.941105
1	small	145.763992
2	bathroom	107.658870
3	hotel	68.531373
4	breakfast	64.188070
5	little	56.935286
6	shower	49.456733
7	staff	43.201647
8	elevator	42.437045
9	old	41.147902
10	bed	39.272822
11	bit	39.126833
12	could	37.994893
13	everything	36.474176
14	good	31.481887
15	water	30.628224

The following barplot represents the usage of the words “small”, “little” and “tiny” in pair with another words. Here, the room is clearly concerned but the size of the bathroom is not neglectable.



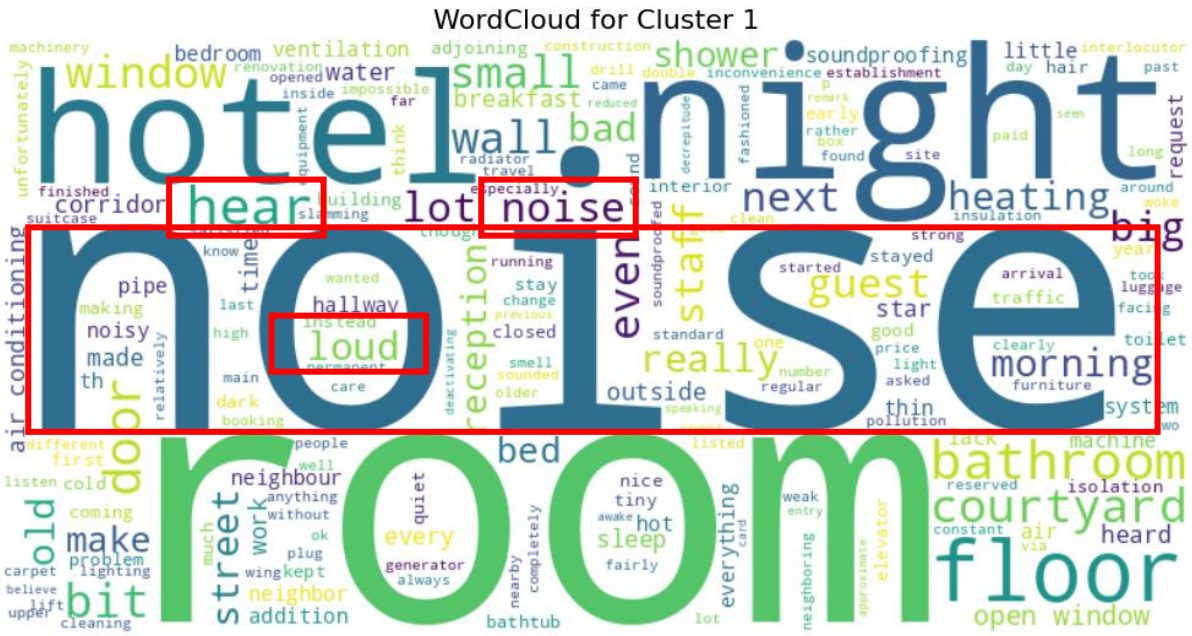
➔ While it is impossible to physically expand the rooms or "push the walls," there are two viable strategies to mitigate this perception. The first one consists in leveraging interior design techniques to create a perception of more spaciousness. For instance, carefully selecting light, reflective paint colors can visually deepen the space. Incorporating strategic lighting and mirrors can also enhance this effect. We invite you to dig this possibility as it could have a real impact.

The second one consists in ensuring that the photos used on booking platforms are honest and accurately depict the size and appearance of the rooms. While this might initially deter some potential customers, it will attract guests with realistic expectations. These guests are more likely to leave positive reviews, leading to a long-term improvement in the hotel's reputation.

By implementing these solutions, the hotel can address concerns about room size indirectly, enhancing customer satisfaction and fostering a stronger, more trustworthy image.

The noise

We identified 2 main clusters of reviews using unsupervised Machine Learning techniques (k-means). While one concerns the smallness of rooms, the other concerns the noise :



As you can see, the lexical field of the noise is present : "noise", "noisy", "hear", "loud". Let's deep dive in some reviews that leverage the problem. With an average score of 6.7 by people complaining of the noise, we can find the reviews giving a 4 like the following :

*"**Very noisy, bad sound insulation, we hear the voices of our room neighbors**. Stwaux in progress on our floor, caulking everywhere unwilling, noise. Prus of shower bath product in the distributor In flexibility to make the room at 12:10 p.m. instead of 12:00 p.m."*

*"**Zero soundproofing we hear the neighbors speak and too many noises coming from the outside.** Tiny shower. Wait too long to be taken care of upon arrival and departure"*

➔ Addressing noise-related complaints is essential for improving guest satisfaction. While some sources of noise (construction work) are temporary and unavoidable, there are several measures that can be implemented to mitigate these issues.

- Double-glazed windows can significantly reduce external noise, particularly in urban environments with heavy traffic or street activity.
- Acoustic insulation can be added between walls and ceilings to reduce noise transmission between rooms.
- Install thicker doors with rubber seals to minimize hallway noise.
- Use carpets or sound-absorbing panels in hallways and communal areas to dampen echoes and footsteps. (be careful also to have clean carpets as certain customers complain about the dirt in the hallway carpets !)

Contribution to sustainable Development Goals

This analysis aligns with several United Nations Sustainable Development Goals (SDGs), demonstrating its broader societal and environmental relevance:

1. **Ensure healthy lives and promote well-being for all at all ages**

By identifying and addressing recurring issues such as cleanliness, service quality, and sleep conditions, this analysis directly contributes to improving the well-being of guests. Enhancing these aspects ensures a healthier and more comfortable experience for customers, positively impacting their overall mental and physical health during their stay.

2. **Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation**

The innovative approach used in this project, including natural language processing (NLP) and sentiment analysis, supports the development of smarter, data-driven hospitality infrastructures. Such advancements enable hotels to become more adaptive to customer needs while fostering sustainable operational practices that ensure long-term resilience in the industry.

3. **Ensure sustainable consumption and production patterns**

This analysis promotes sustainability by helping the hotel optimize its resource allocation and minimize waste. For instance, addressing guest concerns proactively reduces redundancies in operations, supports efficient use of resources, and aligns the hotel's practices with sustainable consumption and production goals.

4. **Promote sustained, inclusive, and sustainable economic growth, full and productive employment, and decent work for all**

By improving customer satisfaction, the hotel can achieve higher occupancy rates, which fosters economic growth and stability. Additionally, improved service quality often requires enhanced staff training and engagement, creating opportunities for employment and ensuring decent working conditions within the hospitality sector.

In summary, the methodologies and insights from this analysis not only enhance the guest experience but also contribute to achieving significant sustainable development goals. This reinforces the importance of data-driven strategies in creating a more inclusive, sustainable, and prosperous future for the hospitality industry and beyond.