

BA820 - Group 9 - Project Proposal

Hospitality Intelligence: Extracting Insights from European Hotel Reviews

By Sricharan Mahavadi, Chen He, Eric Bai, Min Xu, Olimpia Borgohain

Project Motivation :

Hotels must exceed diverse guest expectations to thrive today. By unlocking insights from customer segments and hotel reviews at scale, hotels can understand what customers truly value, guiding improvements for more personalized, cozy stays.

Problem Statement :

Can machine learning unlock insights from hotel reviews to help travelers find their perfect cozy stay in Europe and help hoteliers enhance guest service and their rating?

- **Recommendation system:** The results will display hotels recommended for families based on positive reviews of kids amenities.
- **Customer segmentation:** The results will cluster customers into groups like business travelers, couples, and families based on review preferences.
- **Sentiment analysis:** The results will identify top positively reviewed amenities and commonly criticized pain points.

Dataset and Source :

- The dataset is from Kaggle, with the original data web scraped from **Booking.com**. It's a comprehensive collection of ~**515K** records from **1450+** luxury hotels across Europe, featuring customer reviews, ratings and geographical data. The source can be found [here](#);
- The dataset contains 17 features, including text-based customer reviews, numerical data such as average score, latitude, longitude, as well as categorical data such as hotel name, reviewer nationality, and review tags.

Proposed methodologies :

- K-means clustering to group customers based on review patterns [\[1\]](#).
- Hierarchical clustering to identify hotels with similar reviews [\[2\]](#).
- Sentiment analysis using NLP techniques like tokenization, lemmatization with NLTK to classify review sentiment and identify positive/negative hotel features [\[3\]](#).

Process flow chart : A more detailed approach to our project can be found here - [Process Flow](#)

Timelines : The proposed timelines for the project can be found here - [Timelines](#)

Business Relevance and Implications :

By extracting insights from reviews, this project helps hotels **boost guest satisfaction** and **revenue** through **targeted improvements**, **personalized recommendations**, **sentiment analysis** and data-driven competitive advantages, tapping into the Trillion Dollar\$ Hospitality Industry.