Reviews analysis on European Luxury Hotels

### Data resource:

**Size:**

515,000

**Fields**:

* **Hotel\_Address:** Address of hotel.
* **Review\_Date**: Date when reviewer posted the corresponding review.
* **Average\_Score**: Average Score of the hotel, calculated based on the latest comment in the last year.
* **Hotel\_Name**: Name of Hotel
* **Reviewer\_Nationality**: Nationality of Reviewer
* **Negative\_Review:** Negative Review the reviewer gave to the hotel. If the reviewer does not give the negative review, then it should be: 'No Negative'
* **ReviewTotalNegativeWordCounts**: Total number of words in the negative review.
* **Positive\_Review**: Positive Review the reviewer gave to the hotel. If the reviewer does not give the negative review, then it should be: 'No Positive'
* **ReviewTotalPositiveWordCounts**: Total number of words in the positive review.
* **Reviewer\_Score**: Score the reviewer has given to the hotel, based on his/her experience
* **TotalNumberofReviewsReviewerHasGiven:** Number of Reviews the reviewers has given in the past.
* **TotalNumberof\_Reviews**: Total number of valid reviews the hotel has.
* **Tags**: Tags reviewer gave the hotel.
* **dayssincereview**: Duration between the review date and scrape date.
* **AdditionalNumberof\_Scoring**: There are also some guests who just made a scoring on the service rather than a review. This number indicates how many valid scores without review in there.
* **lat**: Latitude of the hotel
* **lng**: longtitude of the hotel

### Questions to Explore

* Which words are most frequently for each city?
* How are reviewers’ feelings based on the reviews?
* How do factors related with reviewer and their review text influence scores that they gave?
* How to classify different customers based on their reviews and other behaviours, so booking can apply different strategies to them.
* How to classify hotels based on customers’ feelings (reviews)? How can booking improve their recommendation system based on this.

### Methods

* Words frequency analysis and plot a keyword cloud
* Perform a sentiment analysis on the reviews
* Fit a regression model to predict reviewer scores
* Clustering viewers based on recency, frequency, monetary;

Clustering hotels based on reviews

* Informative visualization

### Requirement Satisfaction

* The analysis of the data is reasonably complex, involving multiple steps (geospatial joins/operations, data shaping etc).
* Use techniques of clustering.
* Perform a machine learning analysis with scikit-learn as part of the analysis.
* The webpage includes a significant interactive component (cross-filtering, interactive widgets, etc)