Hotel Reviews Analysis

for Enjoy Hotel Group

Lillian Leong
Data Science Inc.

July 2022

Bio

- Lillian Leong
- Chartered Accountant (ACCA UK)
- Regional Finance Manager, partnering business leaders across various industries in Asia Pacific to achieve their financial and business goals

Presentation Outline

- Problem Statement
- Overview of Dataset
- EDA and Key Insights
- Modelling & Evaluation of Models
- Conclusion and Future Works
- Jupyter Notebook
- Q&A

Business Problem

- Business has suffered due to backlash from their hotel guests
- Conflicting reviews left on the partner booking websites www.booking.com.
- Requirements:
  - provide some insights to these review ratings in Europe hotels
  - key areas of concerns in the guests' reviews
  - any particular segments of the unhappy guests.
  - Whether the client is likely to recommend the hotel
- Stakeholders: Client, customer, hotel staffs

Data Science Questions

We attempt to answer these questions through this project:

1. How the hotels across Europe were doing as a whole in terms of rating score?
2. Which countries are the guests from and is there any differences in rating scores across different nationalities?
3. Any other characteristics with guests giving bad reviews?
4. Is there any common topics/ characteristics in the bad reviews
5. Build a classifier to predict the sentiment of the reviewer text to assess if they would recommend the hotel.

Dataset

Overview of Dataset

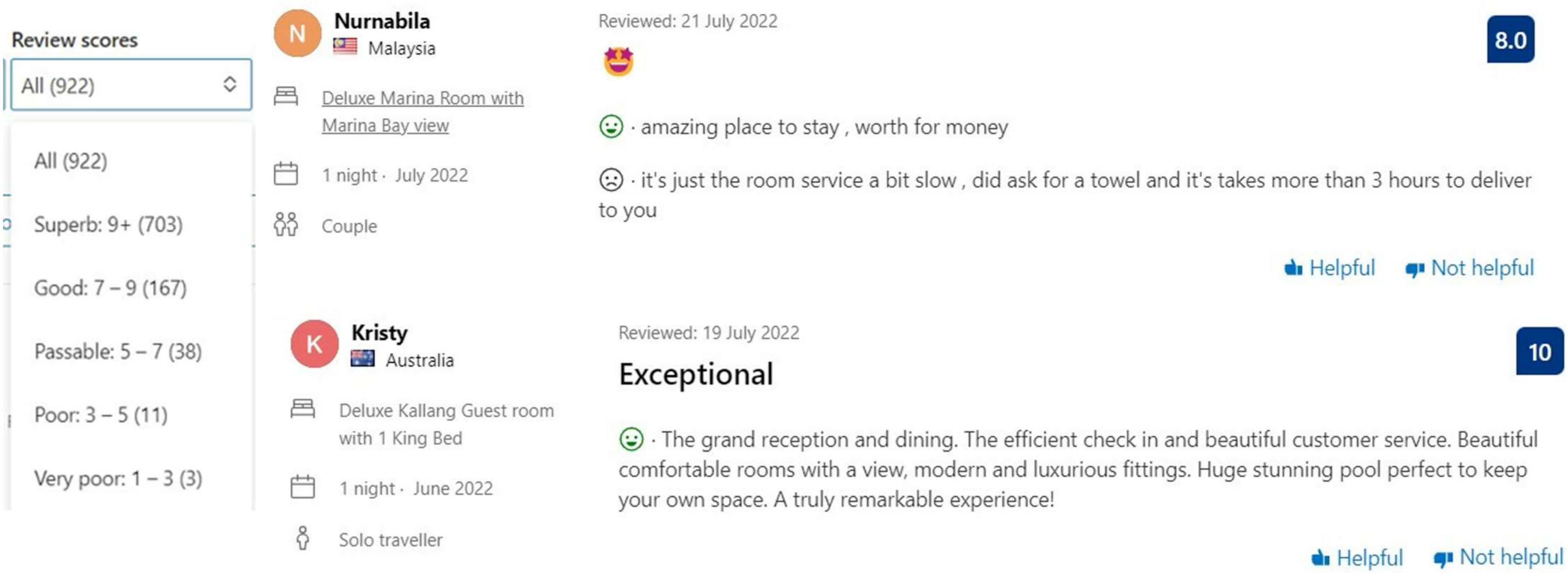
This dataset contains 515,738 customer reviews and scoring of 1493 Luxury Hotels across Europe. The csv file contains 15 fields.

1. Hotel_Address
2. Review_Date: Date when reviewer posted the corresponding review.
3. Average_Score: Average Score of the hotel, calculated based on the latest comment in the last year.
4. Hotel_Name: Reviewer_Nationality: Nationality of Reviewer
5. 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'
6. Review_Total_Negative_Word_Counts: Total number of words in the negative review.
7. 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'
8. Review_Total_Positive_Word_Counts: Total number of words in the positive review.

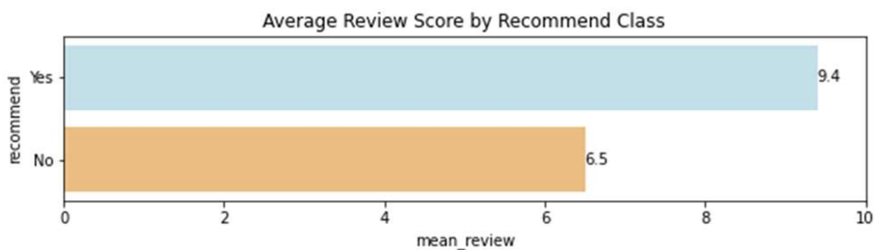
Overview of Dataset

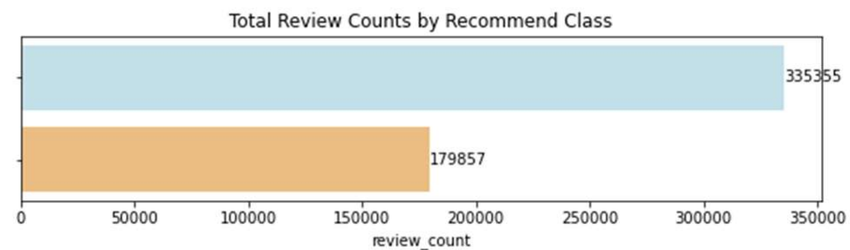
9. Total_Number_of_Reviews_Reviewer_Has_Given: Number of Reviews the reviewers has given in the past.
10. Total_Number_of_Reviews: Total number of valid reviews the hotel has.
11. Tags: Tags reviewer gave the hotel.
12. days_since_review: Duration between the review date and scrape date.
13. Additional_Number_of_Scoring: 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.
14. lat: Latitude of the hotel
15. lng: longtitude of the hotel

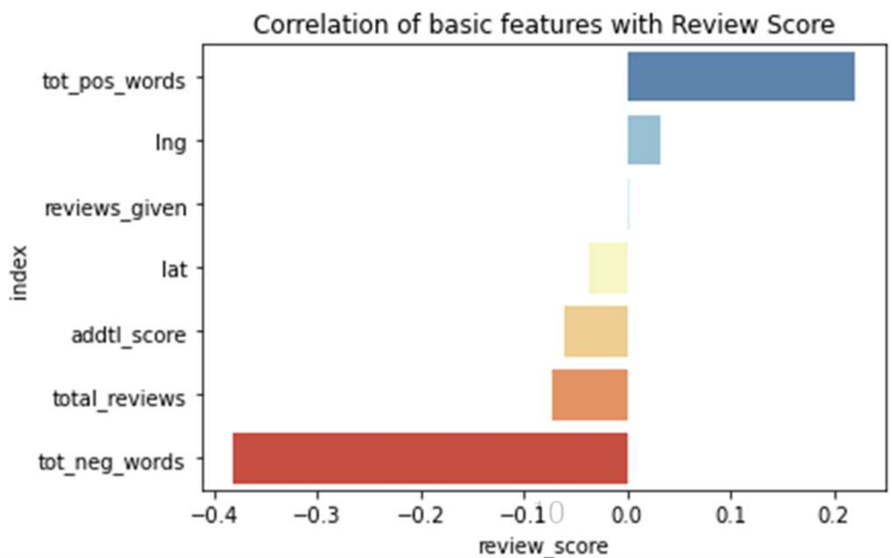
Sample Review on Website



Target Column – Recommend (Review Score > 8/10)

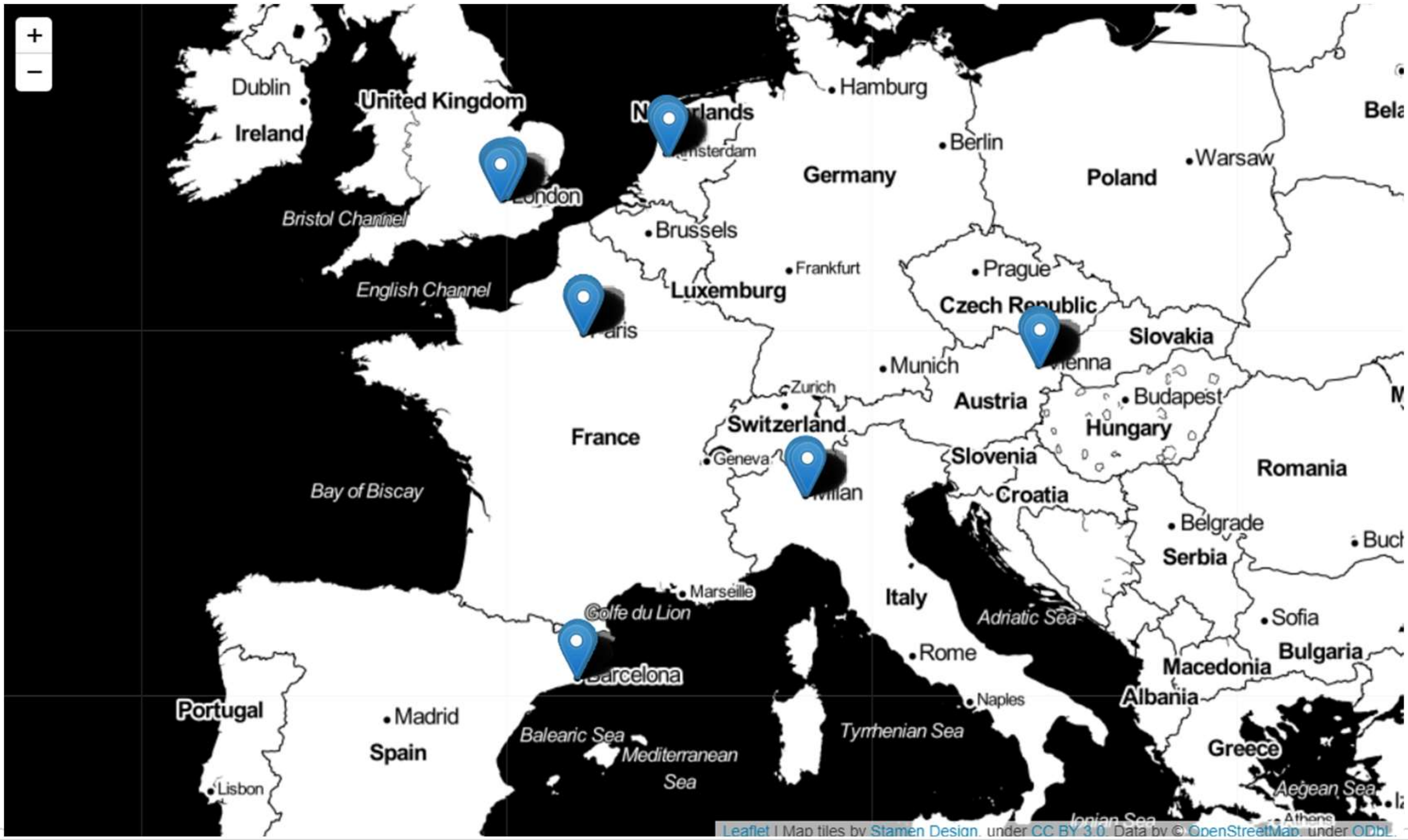




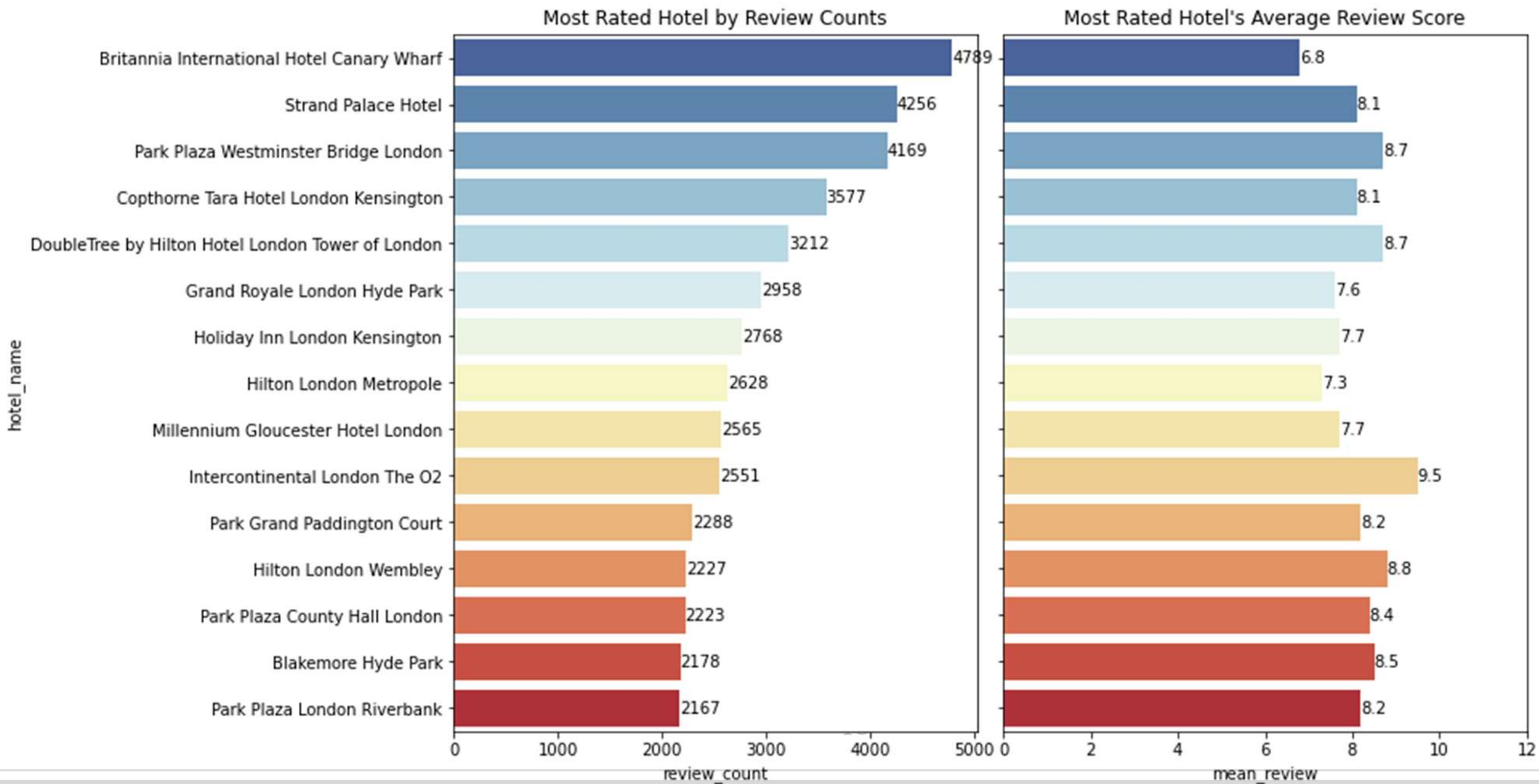


EDA & Business Insights

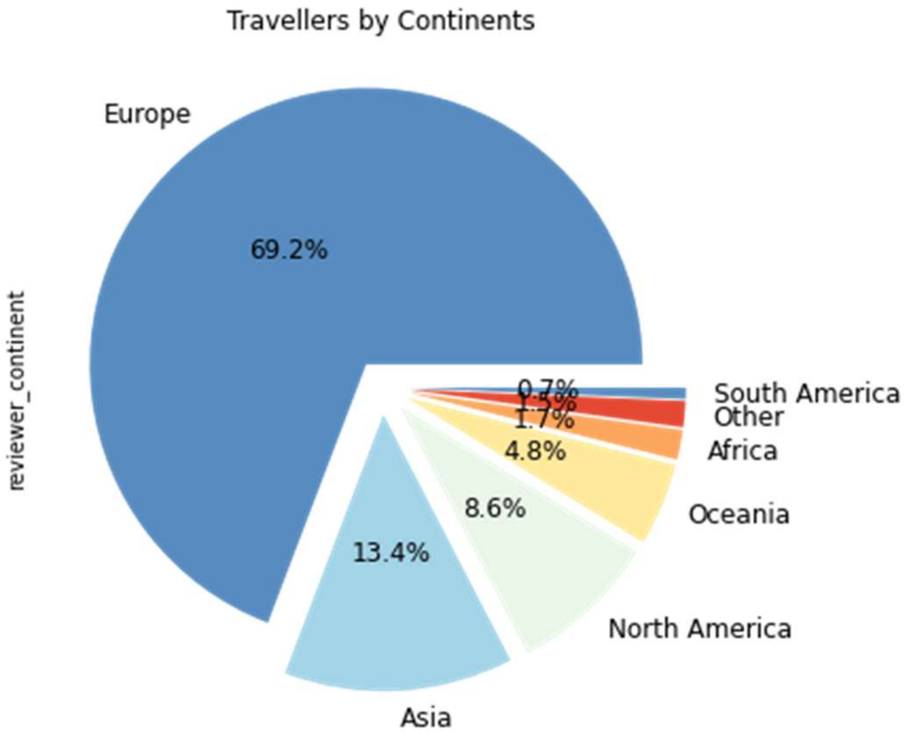
Hotel Locations

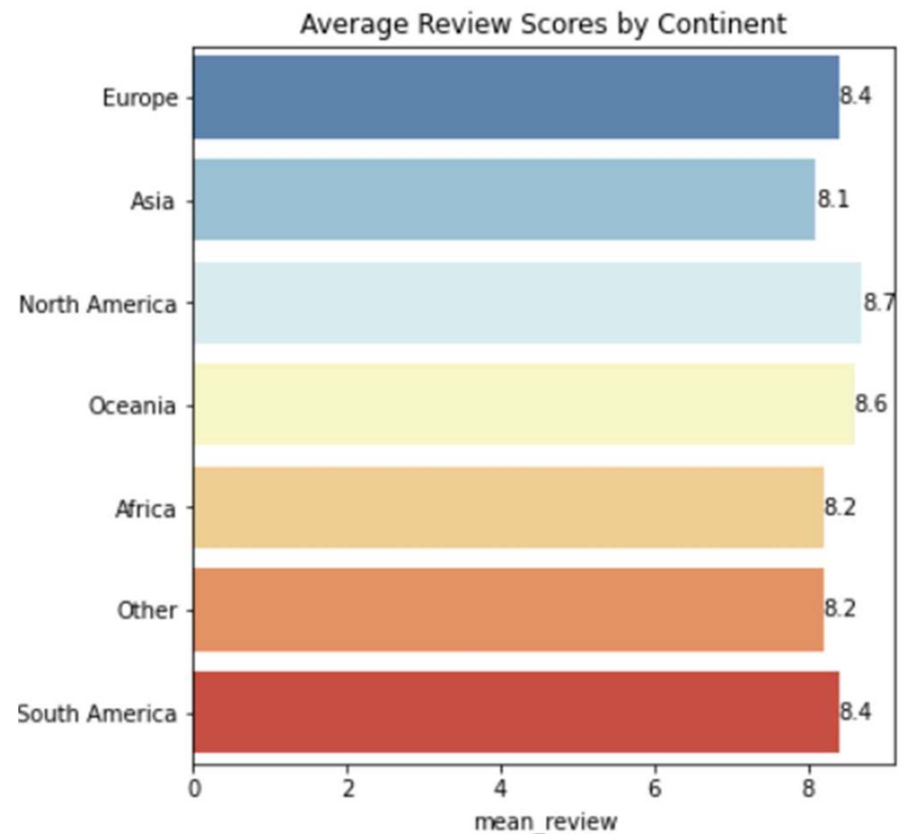


Most Rated Hotels with Review Scores

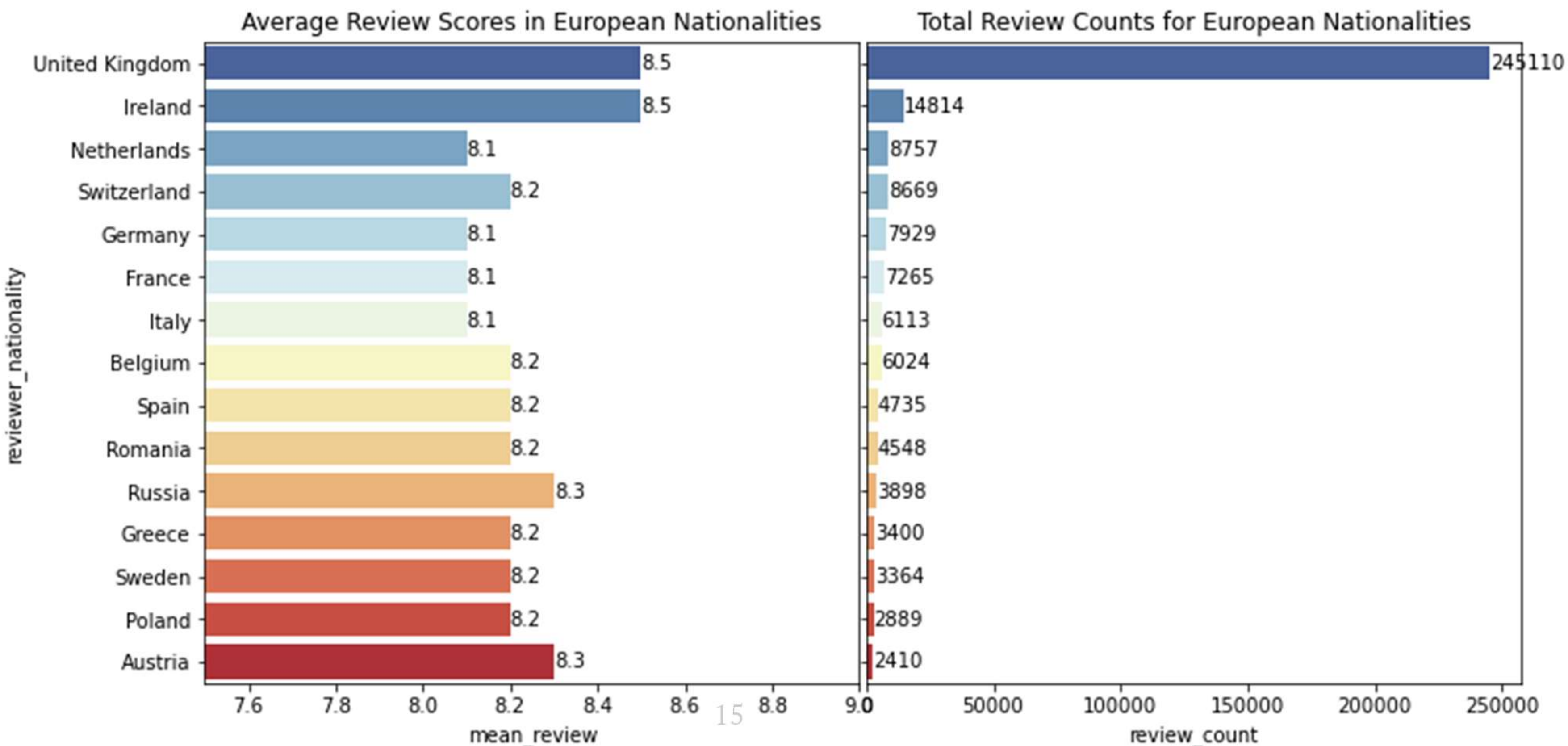


Guest's Geographical Profiles

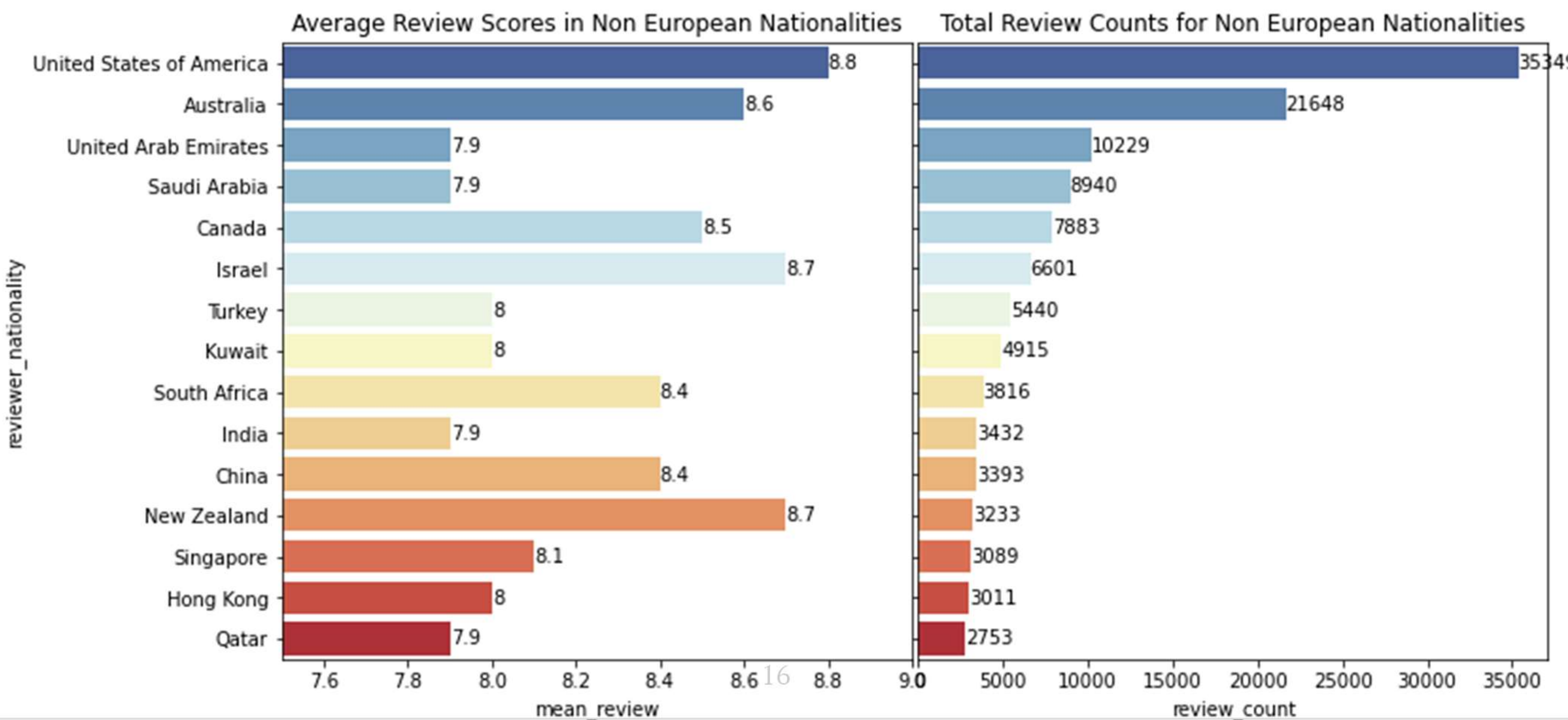




Traveller's Nationality (Europe) Rating Scores

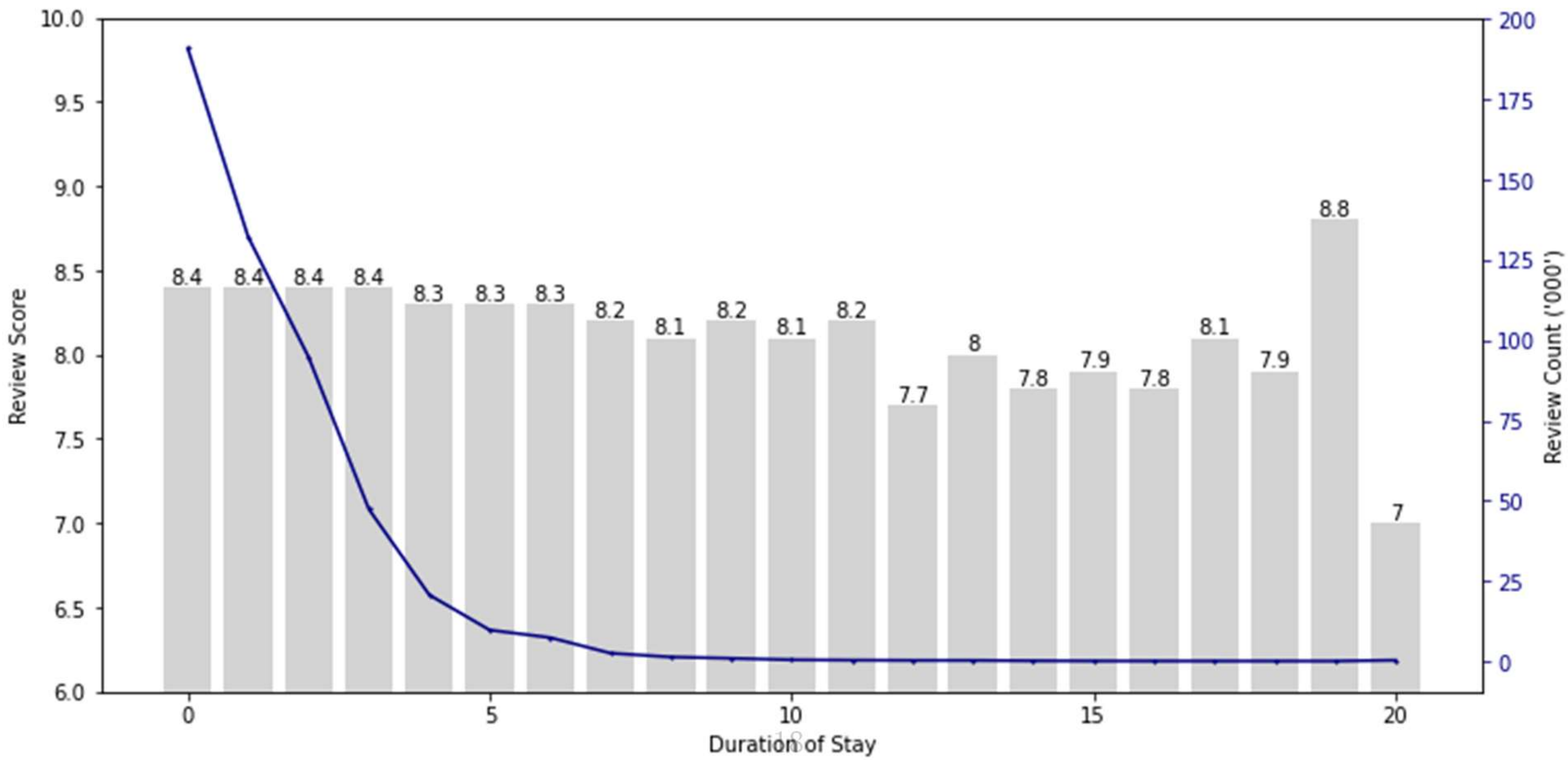


Traveller's Nationality (Non Europe) Rating Scores

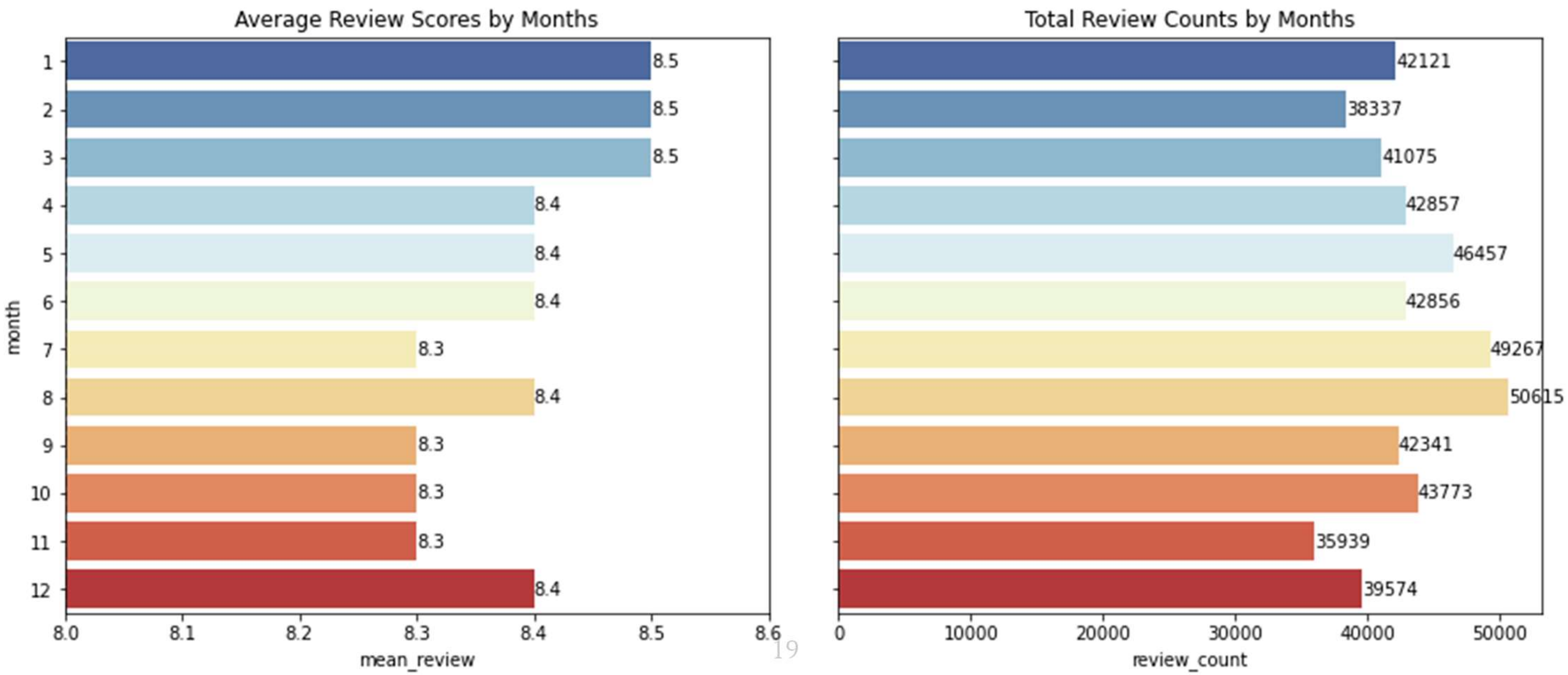


Rating by Traveller's Type

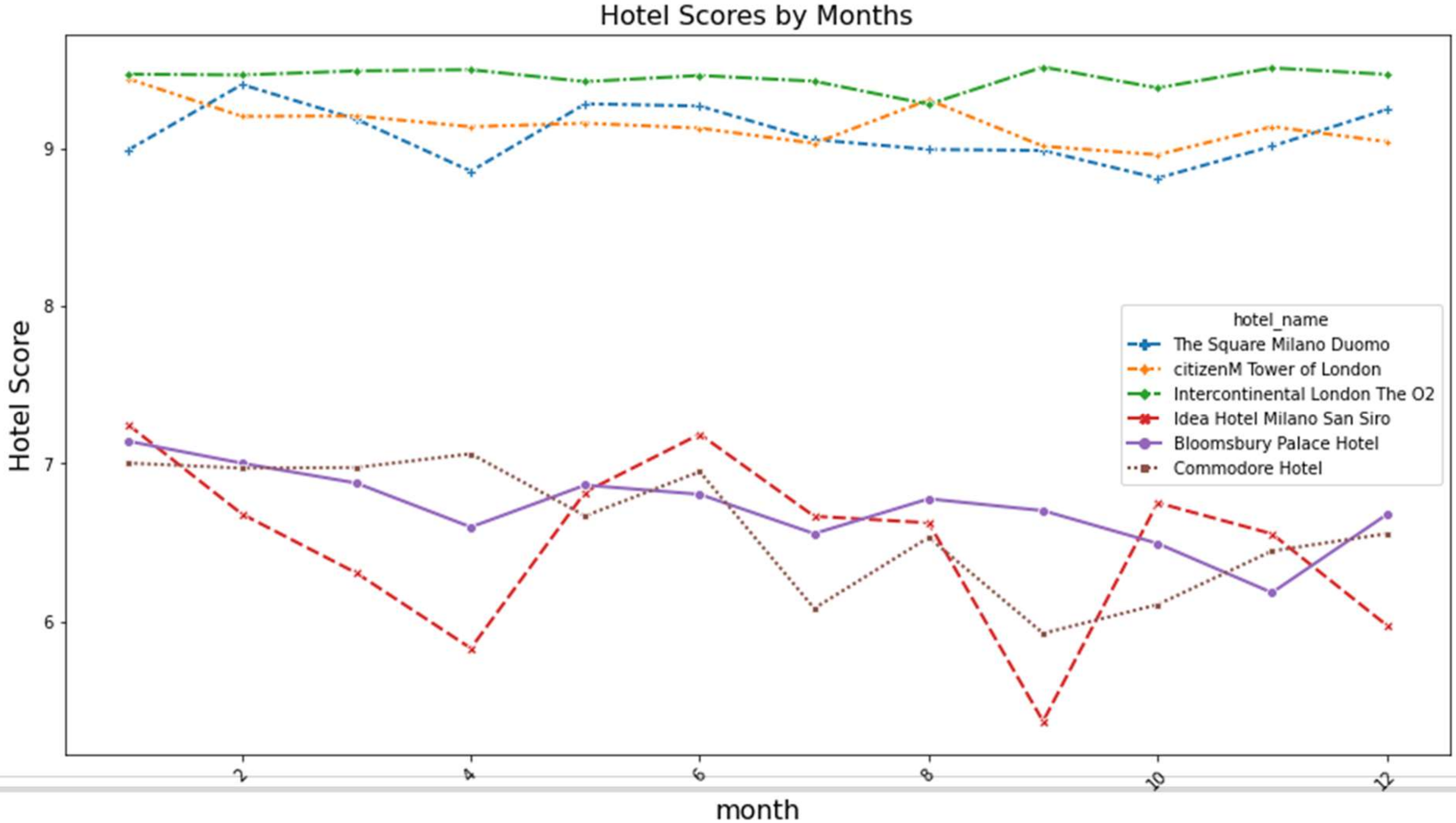
Duration of Stay vs Review Scores & Counts



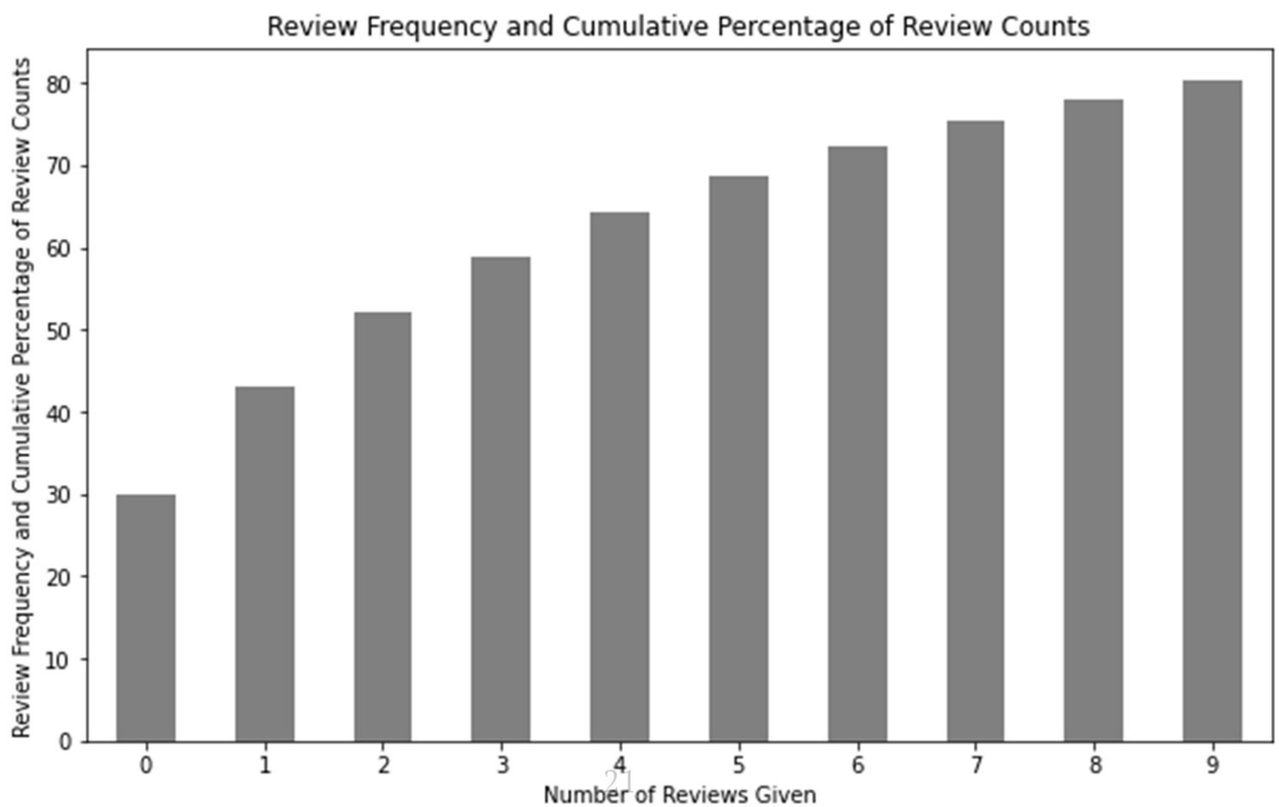
Rating Scores by Months



Rating Scores by Months – Selected Hotels



Frequent Reviewers – New & Repeat Customers?



Key Insights

- Most rated hotels does not translate to best rated hotels.
- 70% of the travellers were from Europe but the best reviews came from outside Europe (NA, Oceania)
- Within the travellers in Europe (mostly from UK) , yet those from Netherlands, Italy, Germany, France gave lower score than their counterparts.
- For non Europe travellers, the lower rating travellers came from Middle East (Saudi Arabia, Qatar, UAE) and India.
- There is more 4x as many leisure vs business travellers who would recommend.
- There is also a higher proportion of solo who had gave bad reviews compared to the segments.
- The review scores drops a little every 3 days and beyond day10 of the hotel stay.
- The lowest scoring review months (where review counts were highest) were Sep & October, possibly after the summer break vacations.
- The hotel with higher review scores were more consistent through the months as compared to the hotels with lower review scores.
- About 30% of the guests was giving the reviews for the first time which might suggest new customer funneling through the booking website

Review Text Analysis

Review Text Analysis

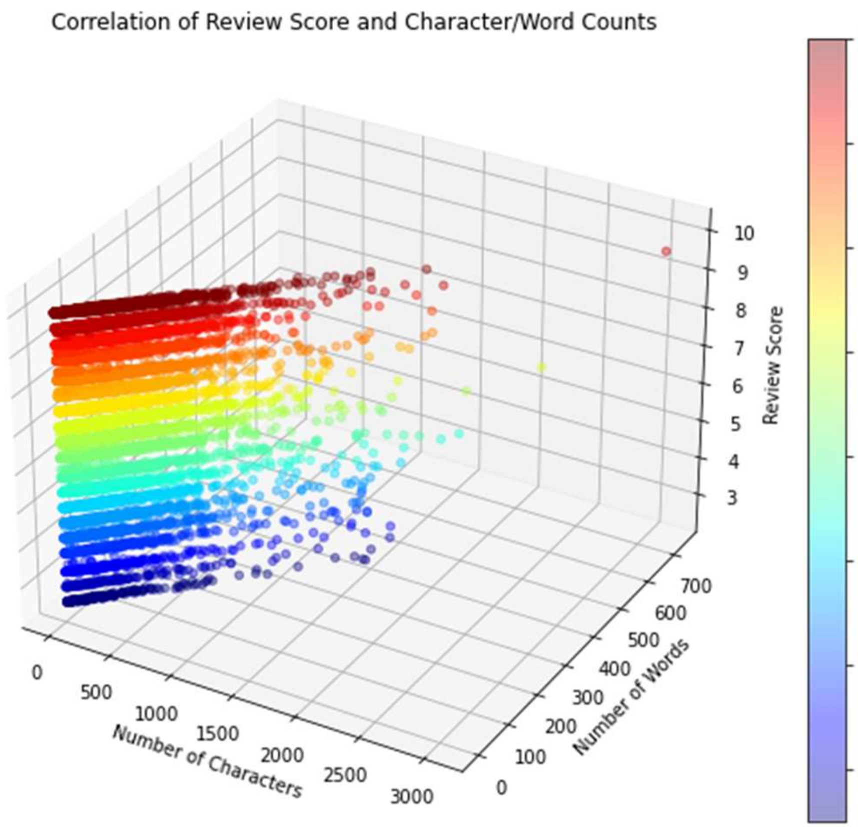
Text Processing Tasks (10% sample of the dataset)

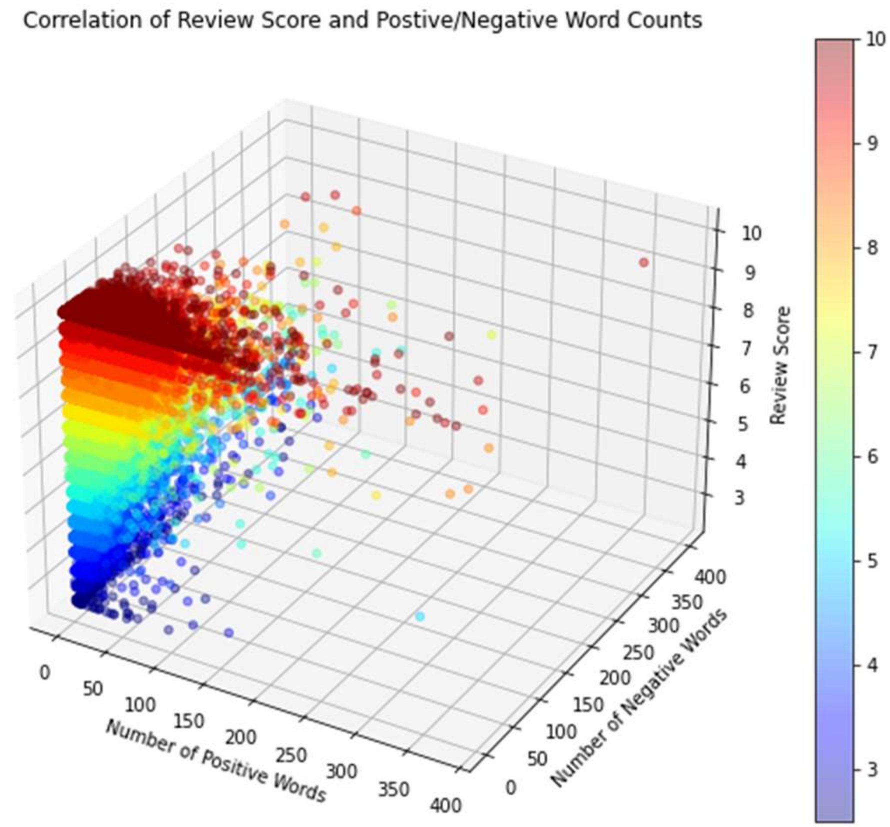
1. Transform the text into lower case
2. Tokenize the text into words and remove the punctuation
3. Remove useless words that contains numbers.
4. Remove useless words like “the”, “a” which are in the stopwords
5. Tag Part of Speech (POS) whether the word is a noun or verb
6. Lemmazation of the text by transforming the word into their root form.

Feature Engineering:

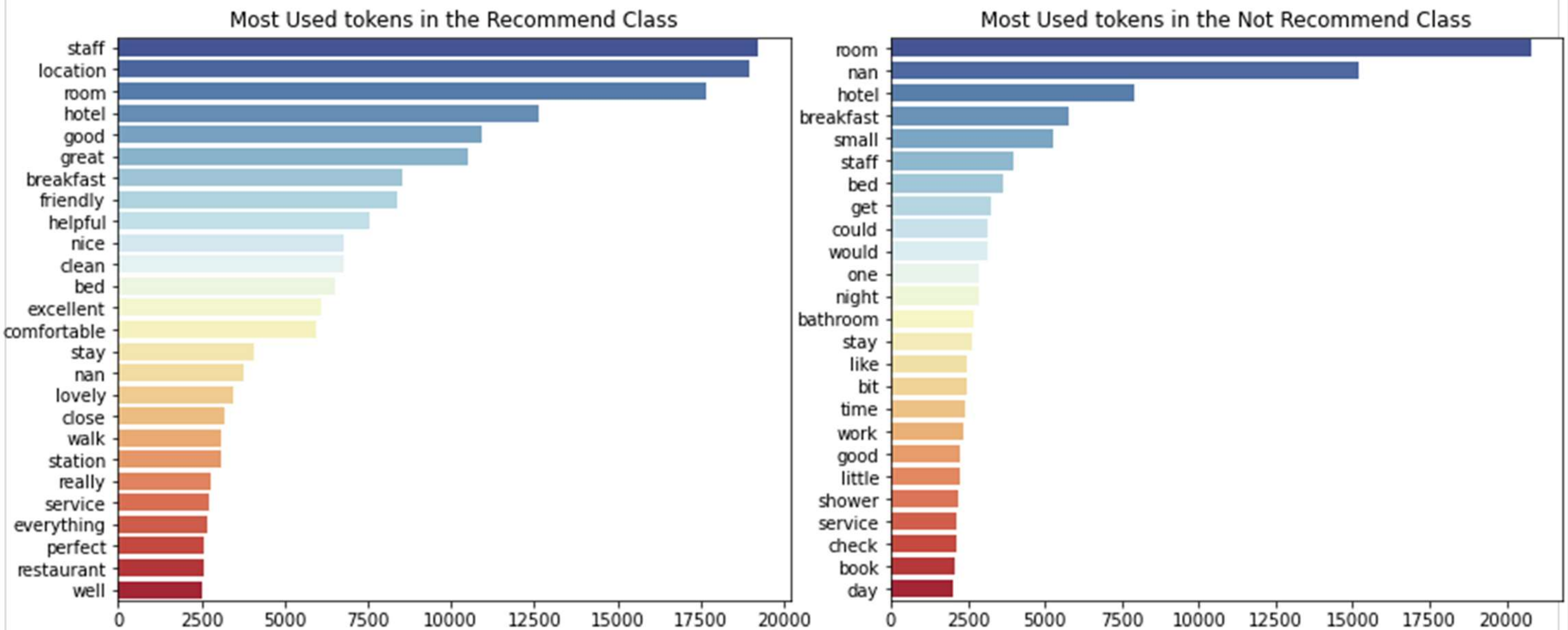
- Character, Word Counts,
- Tokens, Bigram, Trigram Analysis
- Vader Sentiment Analysis + TFIDF Vectorizer

Positive & Negative Words Count vs Review Score



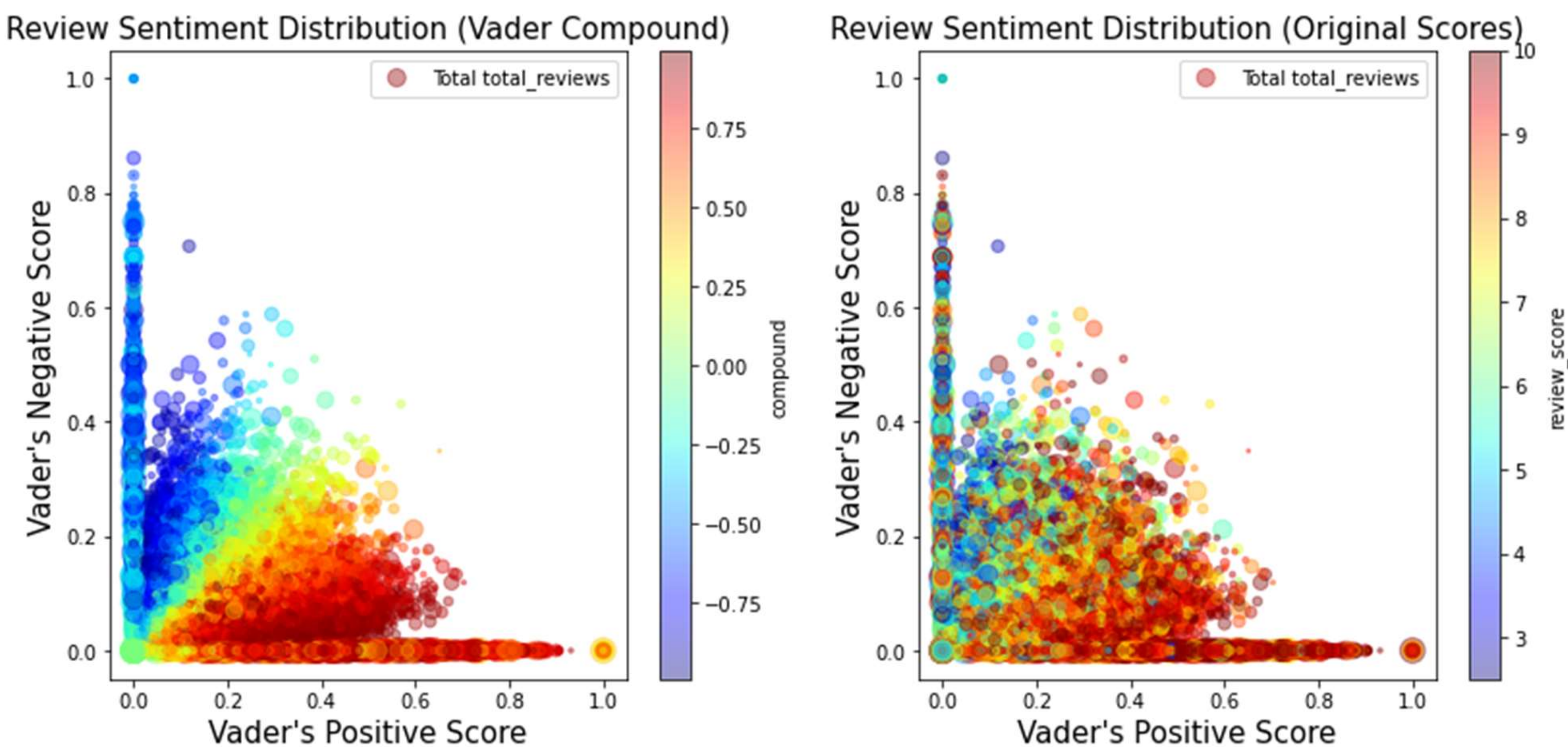


Most Frequent Words in Both Classes



WordClouds –Tokens, Bigram, Trigram

Sentiment Analysis with Vader



TFIDF

	air	also	amaze	area	around	arrive	ask	away	back	bad	bar	bathroom	beautiful	bed comfortable
0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.490878	0.000000	0.0	0.000000	0.000000	0.000000
1	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
2	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
3	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.389873	0.000000	0.0	0.000000	0.000000	0.224892
4	0.0	0.224382	0.0	0.0	0.0	0.0	0.0	0.0	0.472137	0.000000	0.0	0.000000	0.174409	0.136172
5	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.595889	0.0	0.000000	0.000000	0.000000
6	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
7	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.378866	0.000000	0.000000

Summary of Text Classification Analysis

- Most reviews averaged 250 words through top reviews can go beyond 450 words.
- Positive reviews tend to be longer than negative reviews.
- From the unigram analysis, it was observed that unigram removes context from the original review especially when they are used along with words like “not, no” which reverse the original intent of the review.
- From bigram, trigram analysis, we discovered several areas which are often mentioned in the reviews:
  - Room size being small
  - Air conditioning
  - Room amentities like mini bar, wifi, bed, room service, bed
  - Hotel location, walking distance
  - Helpful, friendly staff.
- While Vader Sentiment Analysis is very useful, it might not fully converge to the eventual review score.

Recommendations on guest's concerns

- Size of hotel room ➔ 3D room layout for better visualization.
- Hotel locations being far from their destinations ➔ shuttle bus service
- Address quality of service staff, invest in staff training or even look into staff retention strategies to avoid high turnover rate
- Improve the negative aspect in their breakfast offerings.
- Last but not the least, if the client has budget to uplift or renovate, we would propose that they look into the air-conditioning system, bed condition.

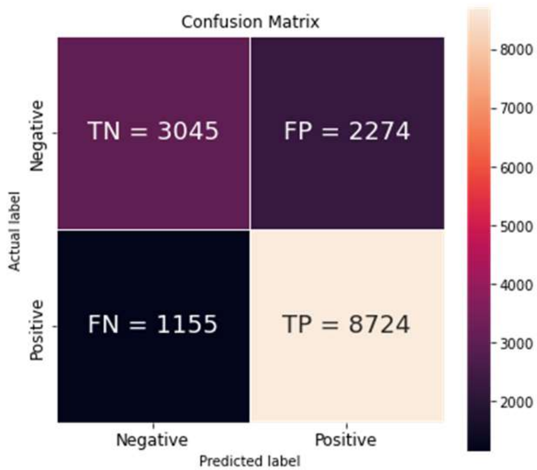
Modelling

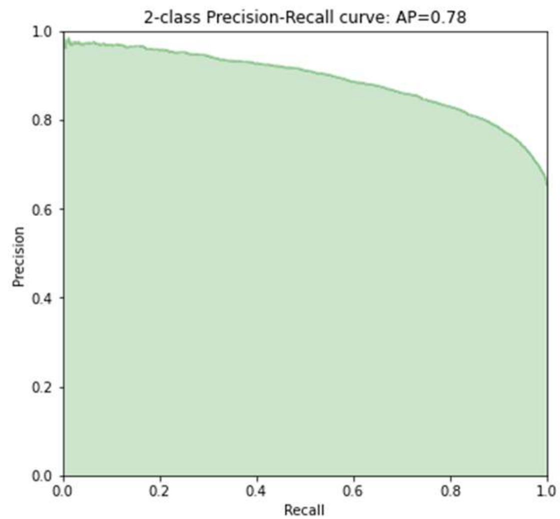
Modelling

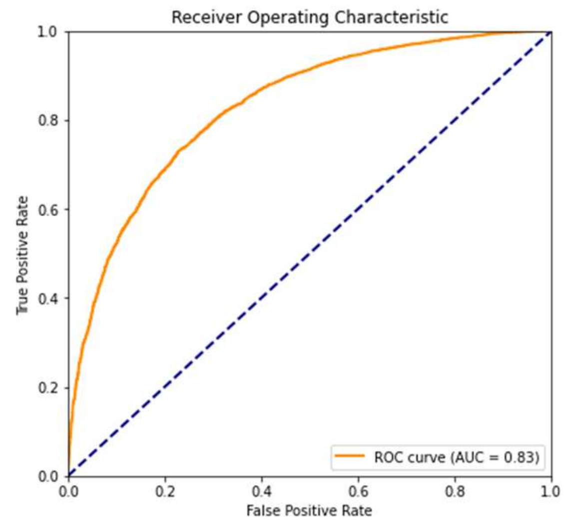
1. TFIDF & Vader Scores as features ➔ Train Size 25% with SMOTE
2. Classification Models
   - Logistic Regression
   - Naïve Bayes Classifier
   - Random Forest Classifier
   - XGBoost Classifier
   - LightGBM Classifier
3. Evaluation of Models
   - Performance Metrics – F1 score > 85%
   - Precision & Recall Tradeoff
   - Confusion Matrices at Different Thresholds
4. Future & Conclusion

Base Model – Logistic Regression

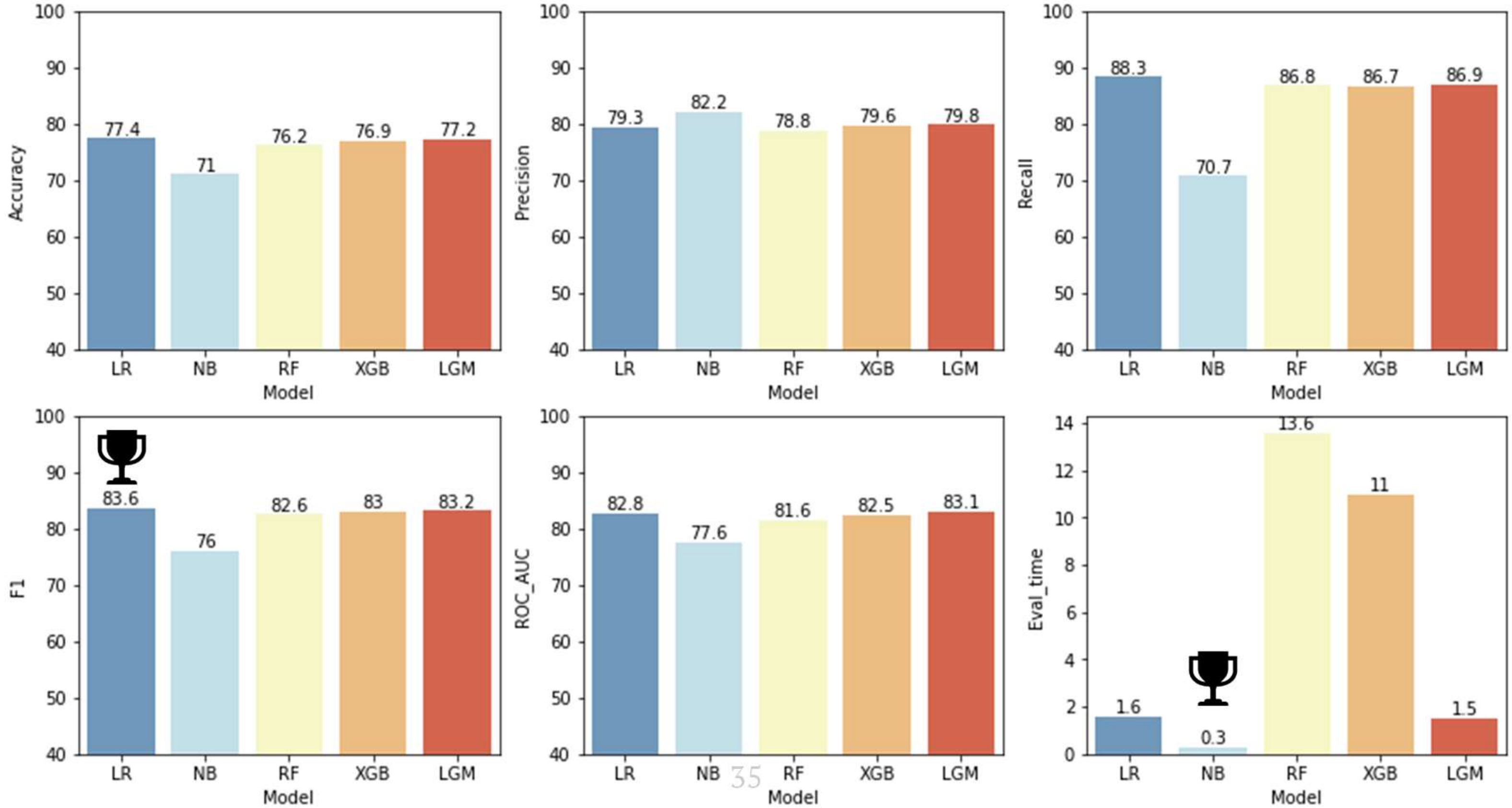
Accuracy	Precision	Recall	F1	ROC_AUC	Eval_time
77.4	79.3	88.3	83.6	82.8	1.6



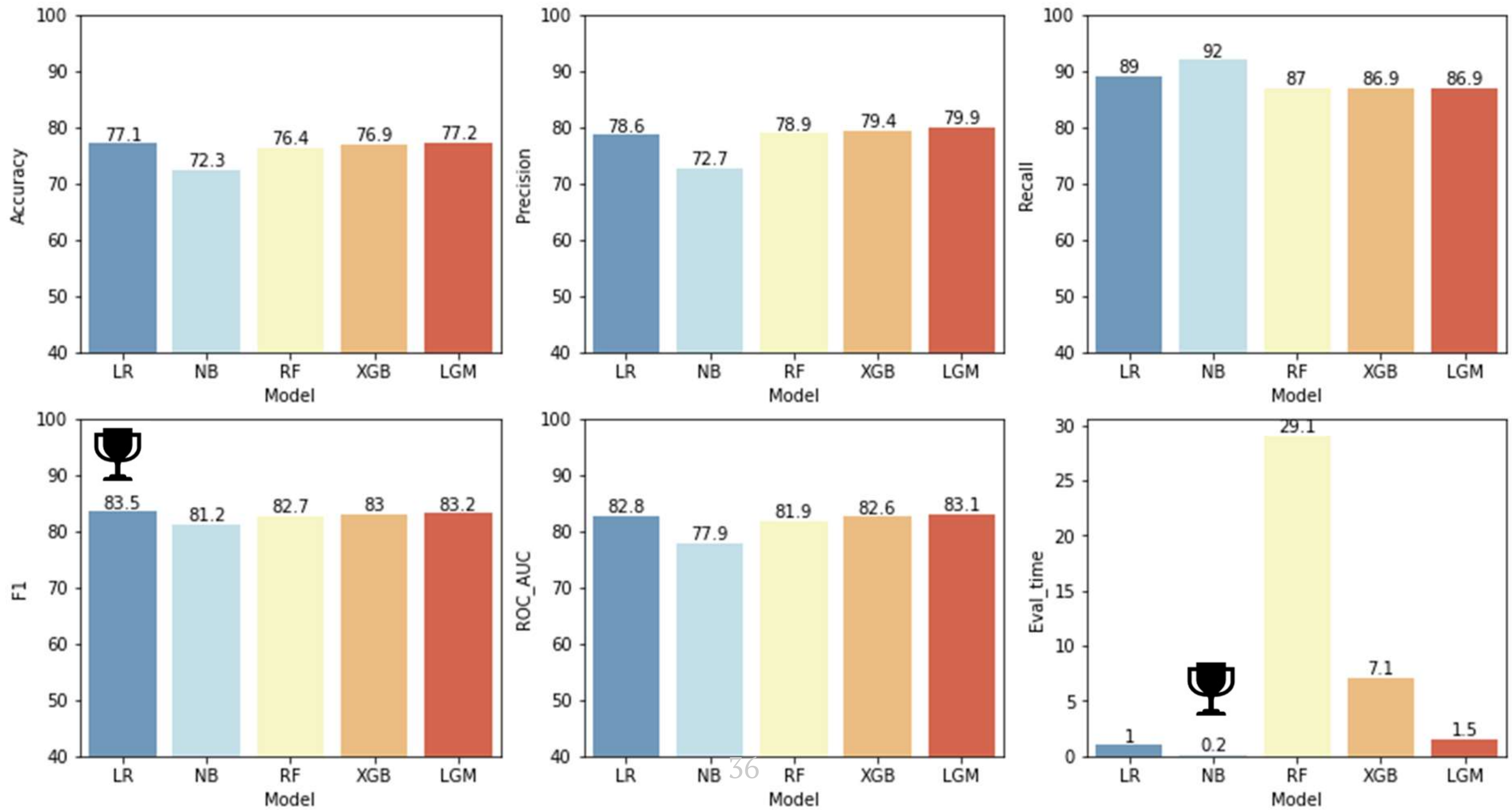




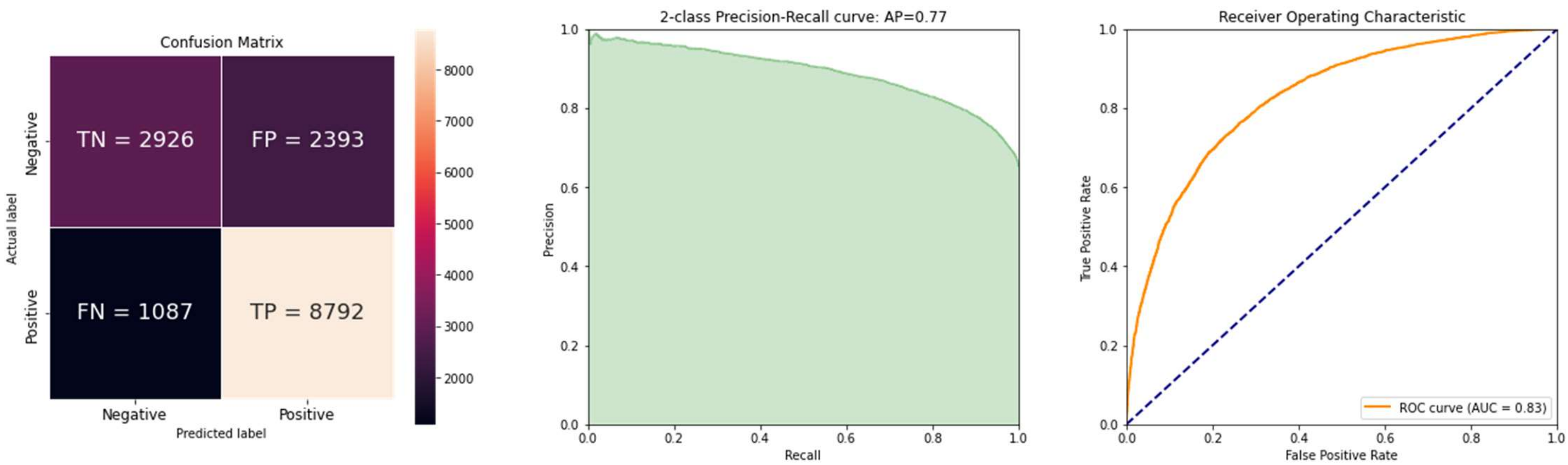
Comparison of Base Model Results



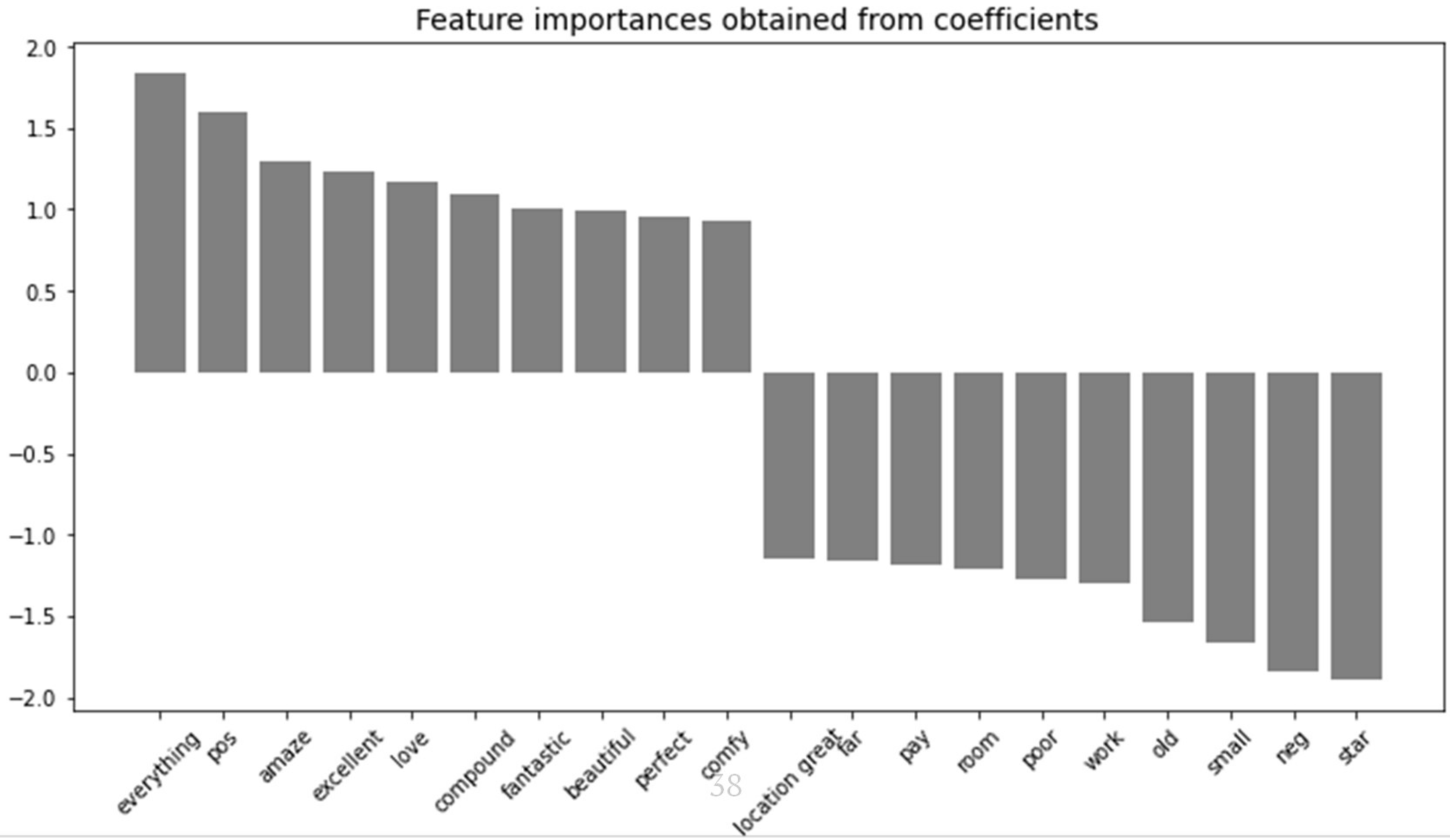
Comparison of Tuned Model Results



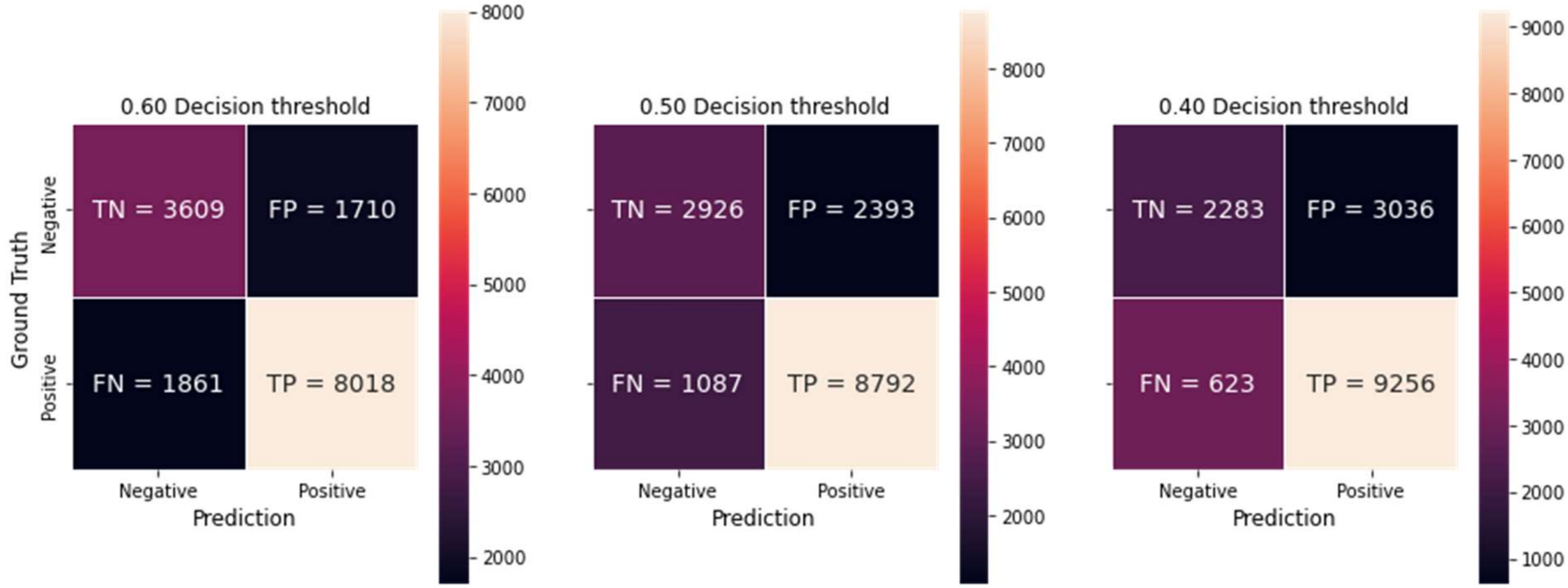
Final Model Selection – Confusion Matrices & PR Curves



Feature Importances – Logistic Regression



Final Model Selection – Confusion Matrices & Different Thresholds



Summary of Models

- Models used: Logistic Regression and Naïve Bayes before applying Ensemble Models (Random Forest, XGBoost, LightGBM).
- Logistic Regression topped the F1 score and clocked a relatively short evaluation time.
- The feature importance in Logistic Regression surrounds Vader's sentiment scores (pos, neu, neg, compound) as well as certain keywords like location, room, staff.
- For every 0.10 upshift in the decision threshold, we see a drop of 500 false positive but at the expense of an increase of false negatives.

Future Works

1. Consider the precision recall tradeoff and finalise the decision threshold
2. Apply to a larger training set to assess the consistency of the scores.
3. To perform further multivariate analysis to find the connection between the text classification features and the existing features to uncover more business insights.
4. Relook at the text pre processing to clean the data further.
5. Relook at hyper parameter tuning to see if F1 can be improved further.
6. Apply the model to a larger sample size to see if the score differs.

References

Detecting bad customer reviews with NLP (Jonathan Oheix, 2018)

https://towardsdatascience.com/detecting-bad-customer-reviews-with-nlp-d8b36134dc7e

Sentiment Analysis using NLP on Hotel Review Dataset (Rishan Sanjay, 2021)

https://medium.com/analytics-vidhya/sentiment-analysis-using-nlp-on-hotel-review-dataset-fa049e23de29

Text analysis basics in Python (Sophia Yang, 2020)

https://towardsdatascience.com/text-analysis-basics-in-python-443282942ec5

How to solve 90% of NLP problems: a step-by-step guide (Emmanuel Ameisen, 2018)

https://medium.com/insight-data/how-to-solve-90-of-nlp-problems-a-step-by-step-guide-fda605278e4e

Jupyter Notebook