Hotel Recommendation System: A mixed Approach

Nittin Yadav, Dishant Yadav, Meenal Gurbaxani, Raman Yadav, Harsh, Daevaang Khairwal

Introduction

While travelling, we book hotels from websites like Oyo, Trivago etc., and they often suggest hotels according to their relevance, paid promotions, or according to the reviews of other customers, and there is nothing personalised. Can you imagine yourself in a hotel with a bathroom made of glass? We don't. So we thought of personalising the recommendation according to the user. Everyone has different tastes in the colours of walls, types of beds, ceilings or even showers, basically the whole vibe of the room, hotel and its ambience. Our recommendation system intends to filter you out of the hotels according to the basic parameters such as amenities in the hotel, your budget, reviews or ratings, nearby amenities and go to places, then show you some pictures of the room and hotel itself, now you have to choose hotels by imagining you in that picture(just kidding), now it will rank the already filtered hotels according to the images you selected.
In real-life scenarios, At the Industrial level, companies generally use only Amenities and reviews of other people to recommend top hotels to a user. But in our Approach, we are using a mixture of two approaches. We are mixing a ranking system based on amenities and reviews of a hotel with a similar image Detection model.
Firstly we intend to provide users with some parameters like Price, location, some amenities etc, and then we filter our hotels from our dataset based on these parameters. Then, we would rank these hotels based on user reviews and amenities as a parameter to our ML model. Then, some of the top hotels would be showed to the user, and the user would need to select multiple photos(any number) based on their interest. Then, we would feed these hotel data into our similar image detection model and again filter out some photos (say 20) and sort them based on their rank we already calculated in the previous step. Then, we would again show the user these top images, but this time we have removed the shortlisted photos from our dataset. Then again, the user will select some photos from these top photos of their interest, and again, these selected images will be fed to our similar image detection algorithm. This process will end if the user wants at any stage when he/she has finalised his hotel or the photos in the dataset left are less than the displaying capacity of our app.

Problem statement:

When travelling, we've noticed that, for the most part, when we book rooms for ourselves, they aren't personalised according to our suggestions or preferences (for instance, the room's colour wouldn't be calming, or the space provided isn't suitable for our needs), which leads to disappointing outcomes or experiences. Numerous hotels offer comparable amenities while considering the customer's needs, and they try to show their finest services that are essentially the same. However, specific work is needed to tailor the room to the customer.

Literature Review:

With the aim of providing recommendations to customers, Collaborative Filtering has been investigated across a variety of industries, including social media (Jiang et al., 2020), restaurants (Liu et al., 2013), and travel (Zheng, Burke, & Mobasher, 2012). The evaluation approaches used include Pearson Correlation Coefficient, Spearman Rank Correlation Coefficient, Cosine Similarity, and Mean-Square Difference in order to identify similar items or users (Pappas & Popescu-Belis, 2015). In the hospitality sector, collaborative filtering is a well-liked method for creating recommendation systems, notably for hotel recommendations.

[1] Within this paper, the context-aware hotel recommendation (CAPH) approach is considered to recommend hotels based on their features and the type of traveller who will be using them. The system's accuracy is obtained and compared to the traditional user and item-based collaborative filtering models.

[2]This paper takes traditional user-based collaborative filtering and decomposes it into three context-sensitive components, and proposes a hybrid contextual approach. It can be seen that choosing an appropriate relaxation of the contextual constraints for each component of an algorithm outperforms the strict application of the context.

[3]This paper proposes a hybrid recommendation model based on deep emotion analysis and multi-source view fusion. This model creates individualised recommendations using user-post interaction evaluations, implicit feedback, and auxiliary data. The DMHR algorithm proposed within this paper performs significantly better at score prediction and suggestion.

[4]This paper proposed a hotel recommendation system that combines collaborative filtering and the RankBoost algorithm. The system's ability to accurately recommend hotels to users was tested using reviews of hotels. The article shows the usefulness of merging various recommendation algorithms and the possibility of personalised recommendation systems in the hotel sector.

[5]This paper introduced VADER (Valence Aware Dictionary for sEntiment Reasoning), a rule-based sentiment analysis model that analyses social media material using a

sentiment lexicon with over 9,000 lexical elements. Researchers and practitioners of natural language processing (NLP) find VADER a popular tool for sentiment analysis across various fields and domains. Multiple models like NB, ME, SVM-C and SVM-R were used

[6]This research used VADER to analyse sentiment in restaurant reviews. According to the study, VADER had an accuracy rate of over 85% when categorising reviews as good, negative, or neutral. The study demonstrates how VADER can be used to analyse the sentiment of consumer feedback

[7]This paper used TensorFlow 2.0 to create a model for image similarity detection. The model, built on a convolutional neural network, detected similar photos with high accuracy. In particular, applications where picture similarity detection is crucial, are highlighted in the research as examples of where deep learning techniques might be used for computer vision tasks.

[8]This research suggests a subject model and sentiment analysis-based visual recommendation system for peer-to-peer accommodation. To offer comparable accommodations, the study blends textual information from internet evaluations with visual elements taken from photographs. The study highlights the value of combining several recommendation techniques and the potential of visual characteristics in the lodging sector.

Overall, the use of image similarity in hotel recommendation systems shows promising results in improving recommendation accuracy and user satisfaction. Research on collaborative filtering-based hotel recommendation systems shows that this method is beneficial for providing individualised recommendations in the hospitality sector. Research in this area will likely benefit from deep learning techniques, hybrid recommendation systems, and matrix factorisation, as they are all intriguing directions for this field's future study. However, there are still challenges to be addressed, such as the high computational cost of image processing and the need for large amounts of image data.

The Proposed Method :

Methodology

For our Dataset Collection, we intend to use web scraping from the websites like agoda.com, tripadvisor.com, Agoda.com, Expedia.com etc. and by using Selenium and beautiful soup(BS4) library.

Our first step after preprocessing the data is to create a ranking system which will then be filtered based on the needs and requirements of amenities by the customer.

For said ranking system, we began by creating scores for each of the hotels mentioned depending upon the basis of reviews and ratings provided. For proper analysis of the text given in the reviews, we used VADER (Valence Aware Dictionary and sEntiment Reasoner)

VADER is a rule-based sentiment analysis model that uses a sentiment lexicon to calculate the polarity of the text and was introduced by researchers Hutto and Gilbert in 2014. It analyses text, including social media writing, using a sentiment vocabulary with different sentiment intensities. 9,000 lexical qualities have now been added to VADER's vocabulary using automated and hand annotation processes. Several tests have shown that the model's performance is on par with or even better than other sentiment analysis algorithms. VADER has been used for political analysis, maintaining brand reputation, and assessing customer feedback since it is computationally efficient.

Based on the text's polarity, the algorithm gives each review a sentiment score; a positive score denotes a positive feeling, and a negative score indicates a negative sentiment. Once we obtained the sentiment scores provided by the algorithm for each review, we then used the corresponding ratings as weights and multiplied them to each other. We then added the results and generated a single aggregate score for each hotel in our dataset. These scores were stored within a column alongside our dataset and then were used to rank the hotels. Hotels with greater scores are ranked higher; thus, we obtain our ranked dataset.

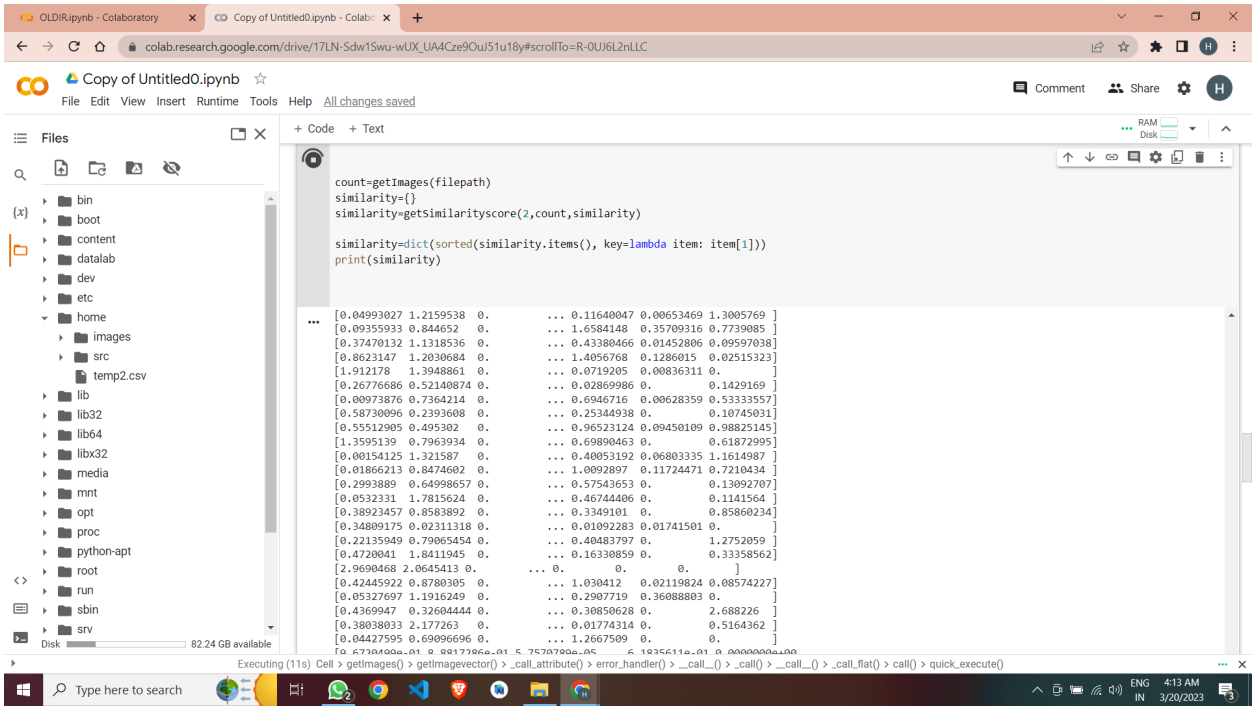
	details	Location	Name	reviews	Score	price	img_link	room_size	amenities	city	...	Private beach access	Carbon monoxide detector	Shared bathroom	Radio
406	{'Location': 'No 1380 1 Near Morjim Football G...	No 1380 1 Near Morjim Football Ground, Tembwad...	Tropical Wave Hostel	[{'Review_Rating': '10.0', 'Review_Heading': '...	9.912	2,498	https://q-xx.bstatic.com/xdata/images/hotel/84...	14m²	['Garden view', 'Balcony/terrace', 'Shower', '...	Goa	...	0	0	0	0
415	{'Location': 'Arambol Beach, Goa, India', 'ove...	Arambol Beach, Goa, India	Goym Resort	[{'Review_Rating': '10.0', 'Review_Heading': '...	9.814	3,820	https://pix8.agoda.net/hotelImages/35712803/58...	10m²	['Garden view', 'Balcony/terrace', 'Shower', '...	Goa	...	0	0	0	0
465	{'Location': "Tito's lane 2, Saunta Vaddo, Bag...	Tito's lane 2, Saunta Vaddo, Baga Near Snow Pa...	Baga Bae	[{'Review_Rating': '10.0', 'Review_Heading': '...	9.783	1,800	https://q-xx.bstatic.com/xdata/images/hotel/84...	0	['Shower and bathtub', 'Free Wi-Fi', 'Toiletri...	Goa	...	0	0	0	0
183	{'Location': 'Cola Goa Beach Resort, Agonda, G...	Cola Goa Beach Resort, Agonda, Goa, India, 403702	Cola Goa Beach Resort	[{'Review_Rating': '10.0', 'Review_Heading': '...	9.694	20,900	https://q-xx.bstatic.com/xdata/images/hotel/84...	30m²	['Sea view', 'Balcony/terrace', 'Shower and ba...	Goa	...	1	0	0	0
410	{'Location': 'Survey No, 155/1, Arpora Baga, O...	Survey No, 155/1, Arpora Baga, Opp - Radisson ...	Caravan Baga Aqua Resort	[{'Review_Rating': '10.0', 'Review_Heading': '...	9.694	3,000	https://q-xx.bstatic.com/xdata/images/hotel/84...	30m²	['Pool view', 'Balcony/terrace', 'Shower', 'To...	Goa	...	0	0	0	0

5 rows × 106 columns

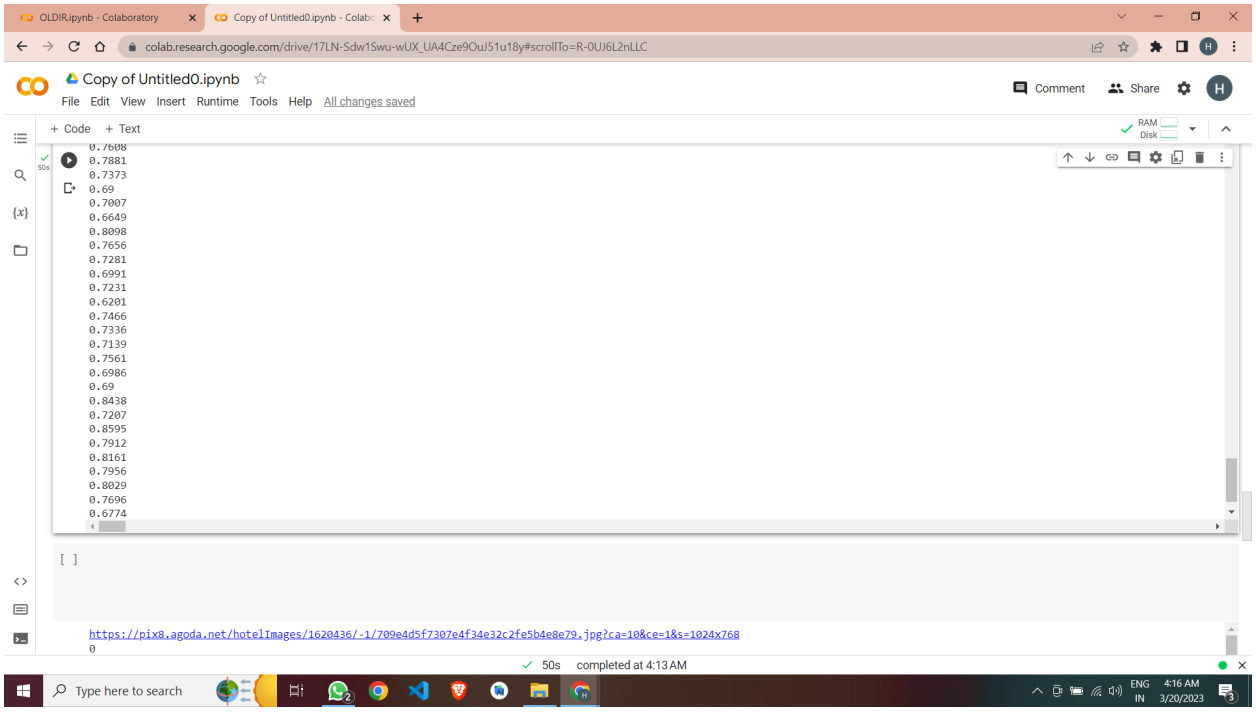
This dataset will further be reduced when the customer selects specific criteria (price, location, amenities etc.). Only the hotels fulfilling those criteria will be passed on for the image similarity step.

[Ranked Hotel Dataset]

For Similar image detection, We are using the Tensorflow library's "Tensorflow Hub" to generate 'image feature vectors' of the images of hotels. Then we are planning to use the "Annoy" library and image feature vectors to calculate the similarity scores and filter out hotels based on similarity scores.



This output is the vector representation of image which will further will be used as similarity detection.



This is the similarity of the image as output with others in a serial manner

Testing

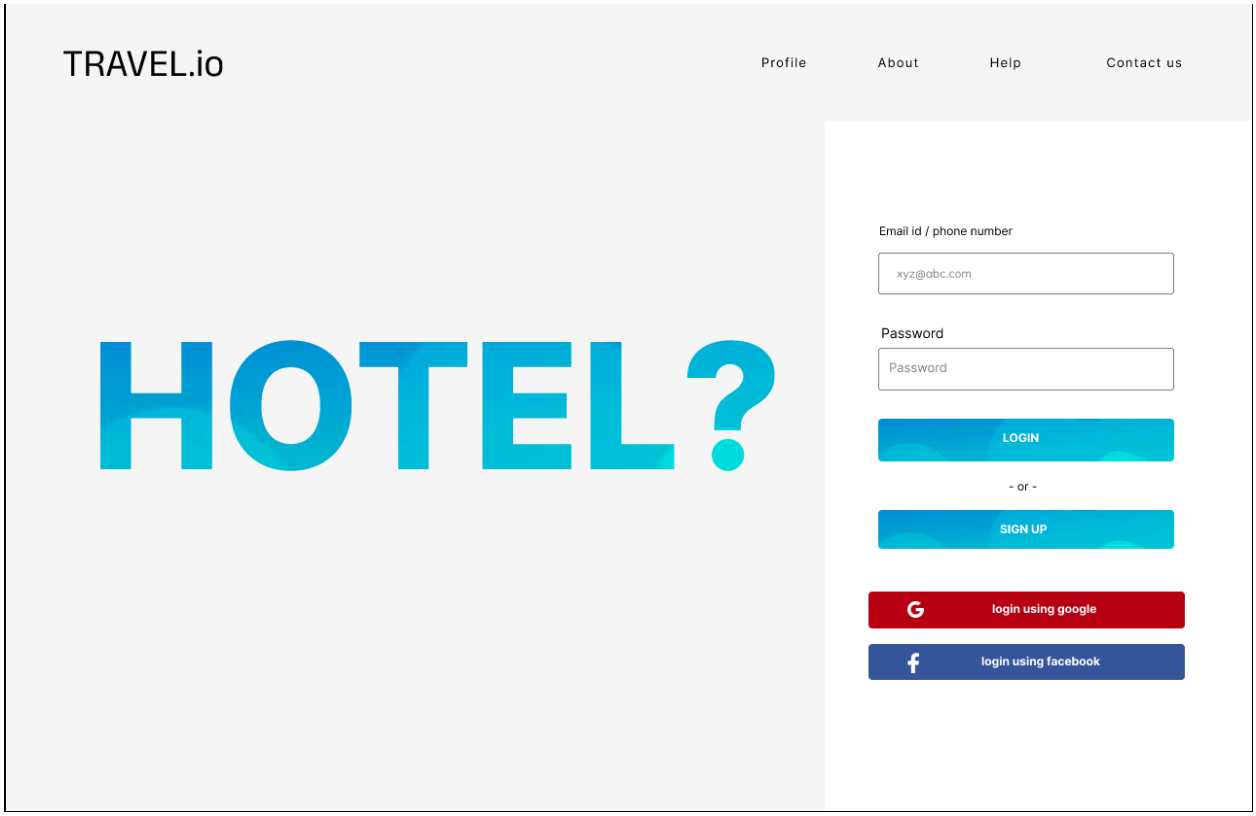
At first, we intend to evaluate it on a small scale by providing the proposed model to our friends and relatives to find out if there are any loopholes or missing features or possible lagging and to make sure they can filter out hotels of their interest or not. Then, we intend to fill those gaps where we were lagging before planning to test or use it on a large scale.

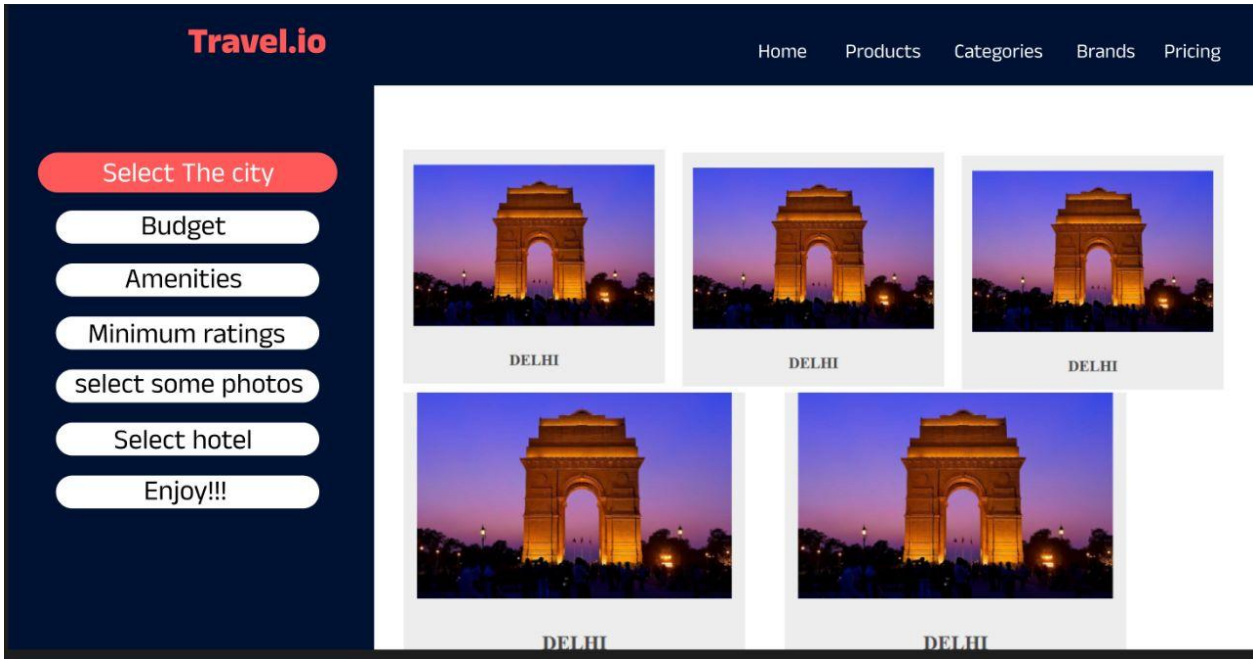
Baseline results(system/prototype)

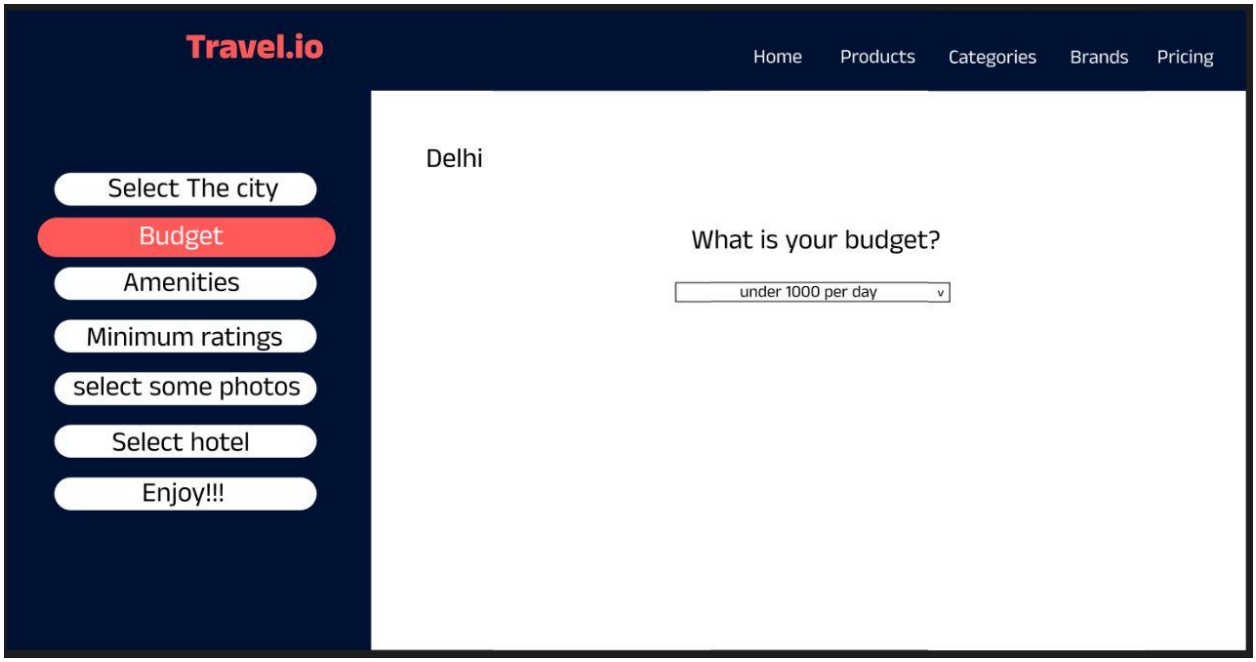
We have web scrapped the data of hotels from websites like Oyo, Trivago etc using Selenium and beautiful soup(BS4) library and extracted the Raw Data and then in pre-processing we cleaned it, removed unwanted amenities, repeated amenities with different names for better training and results and sorted that unformatted data in the form of CSV files. For now, we have submitted Raw Data and processed Data on hotels from Goa which include all the details of the hotels we will need in future like Names, address, images, amenities, price etc. We also explored some more papers, articles related to our work(already mentioned in literature reviews).

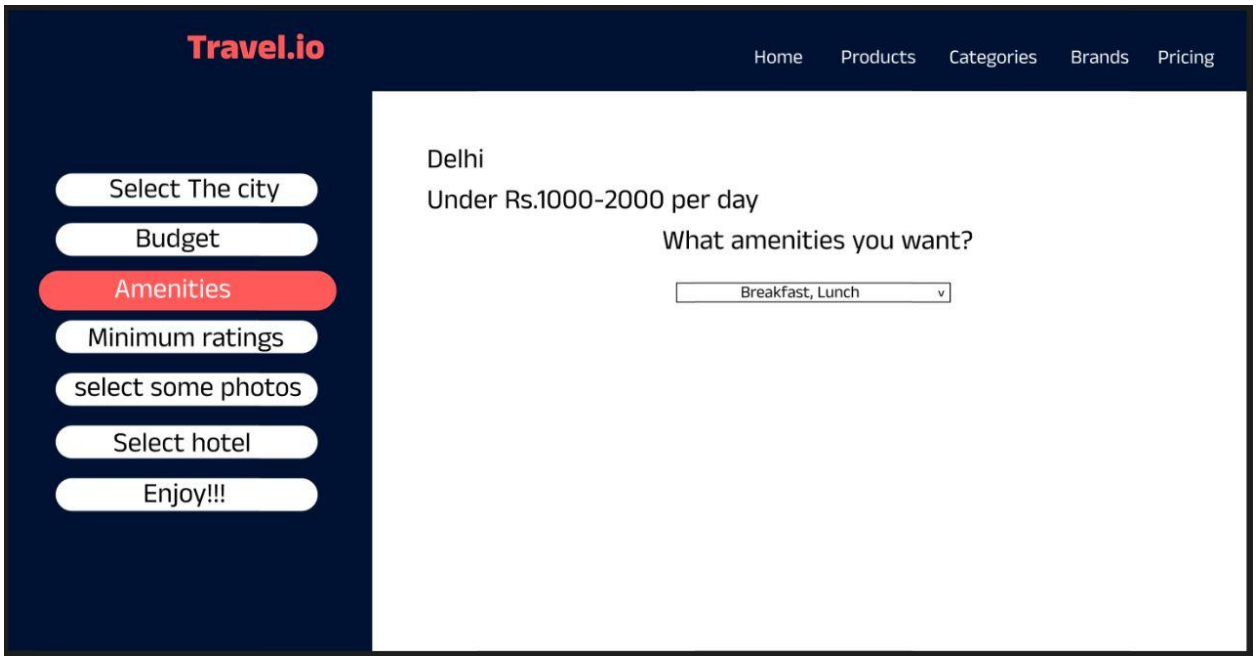
We have designed the protoype on Figma for the website.
You can explore more on this at this link.

We have improved the system prototype as it is not soothing for the eyes (After collecting feedback from our guinea pigs)


TRAVEL.io
Profile
About
Help
Contact us
HOTEL?
Email id / phone number
xyz@abc.com
Password
Password
LOGIN
- or -
SIGN UP
login using google
login using facebook


Travel.io
Home
Products
Categories
Brands
Pricing
Select The city
Budget
Amenities
Minimum ratings
select some photos
Select hotel
Enjoy!!!
DELHI
DELHI
DELHI
DELHI
DELHI


Travel.io
Home
Products
Categories
Brands
Pricing
Select The city
Budget
Amenities
Minimum ratings
select some photos
Select hotel
Enjoy!!!
Delhi
What is your budget?
under 1000 per day


Travel.io
Home
Products
Categories
Brands
Pricing
Select The city
Budget
Amenities
Minimum ratings
select some photos
Select hotel
Enjoy!!!
Delhi
Under Rs.1000-2000 per day
What amenities you want?
Breakfast, Lunch

Travel.io

Home Products Categories Brands Pricing

Select The city
Budget
Amenities
Ratings
select some photos
Select hotel
Enjoy!!!

Delhi

Under Rs.1000-2000 per day

What is your minimum rating?

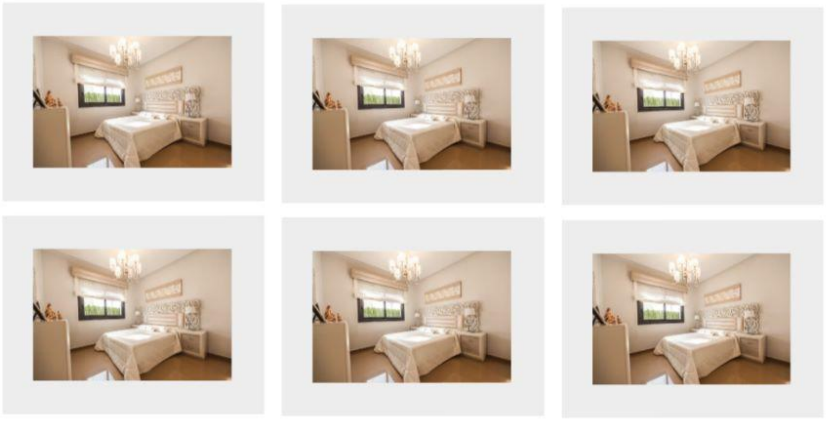
5 Rating

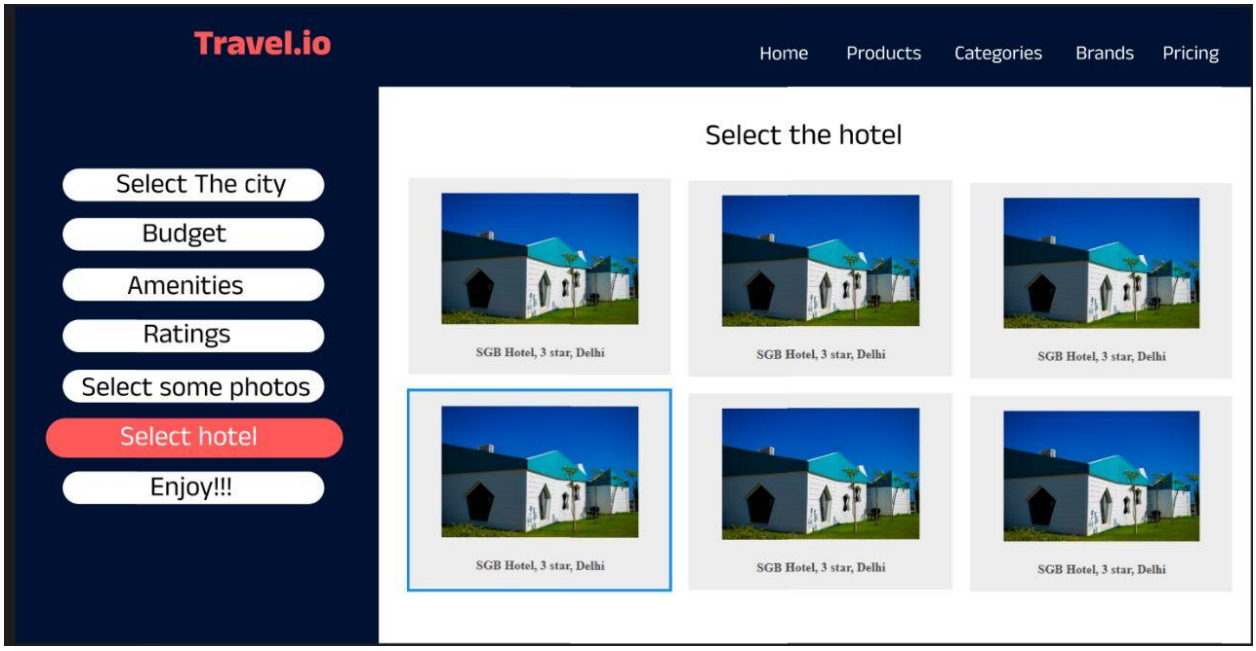
Travel.io

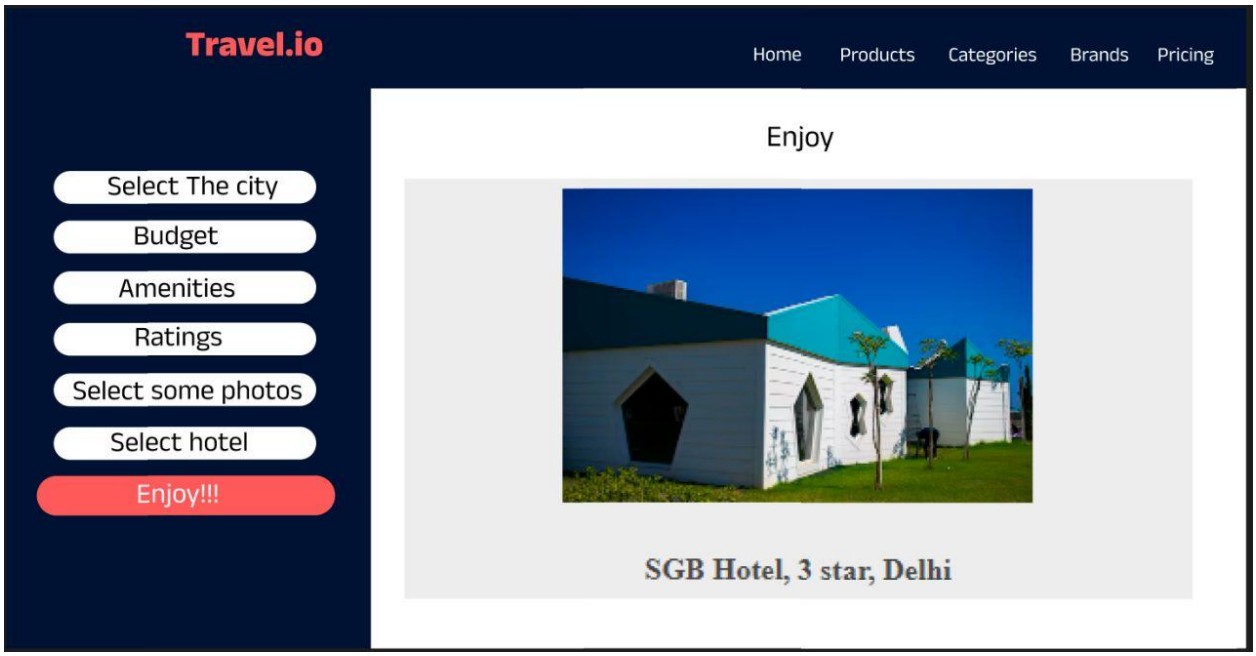
Home Products Categories Brands Pricing

Select The city
Budget
Amenities
Ratings
Select some photos
Select hotel
Enjoy!!!

Select some photos




Travel.io
Home
Products
Categories
Brands
Pricing
Select The city
Budget
Amenities
Ratings
Select some photos
Select hotel
Enjoy!!!
Select the hotel
SGB Hotel, 3 star, Delhi
SGB Hotel, 3 star, Delhi
SGB Hotel, 3 star, Delhi
SGB Hotel, 3 star, Delhi
SGB Hotel, 3 star, Delhi
SGB Hotel, 3 star, Delhi

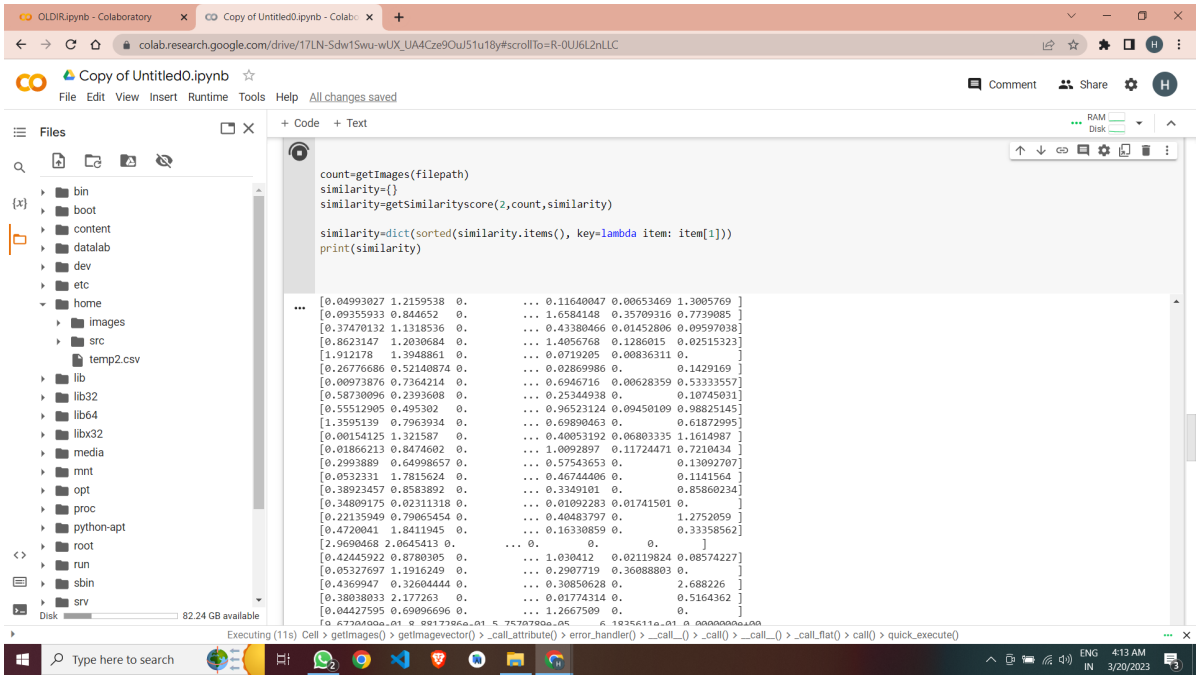

Travel.io
Home
Products
Categories
Brands
Pricing
Select The city
Budget
Amenities
Ratings
Select some photos
Select hotel
Enjoy!!!
Enjoy
SGB Hotel, 3 star, Delhi

Hotel Ranking on the Basis of Scores Obtained:

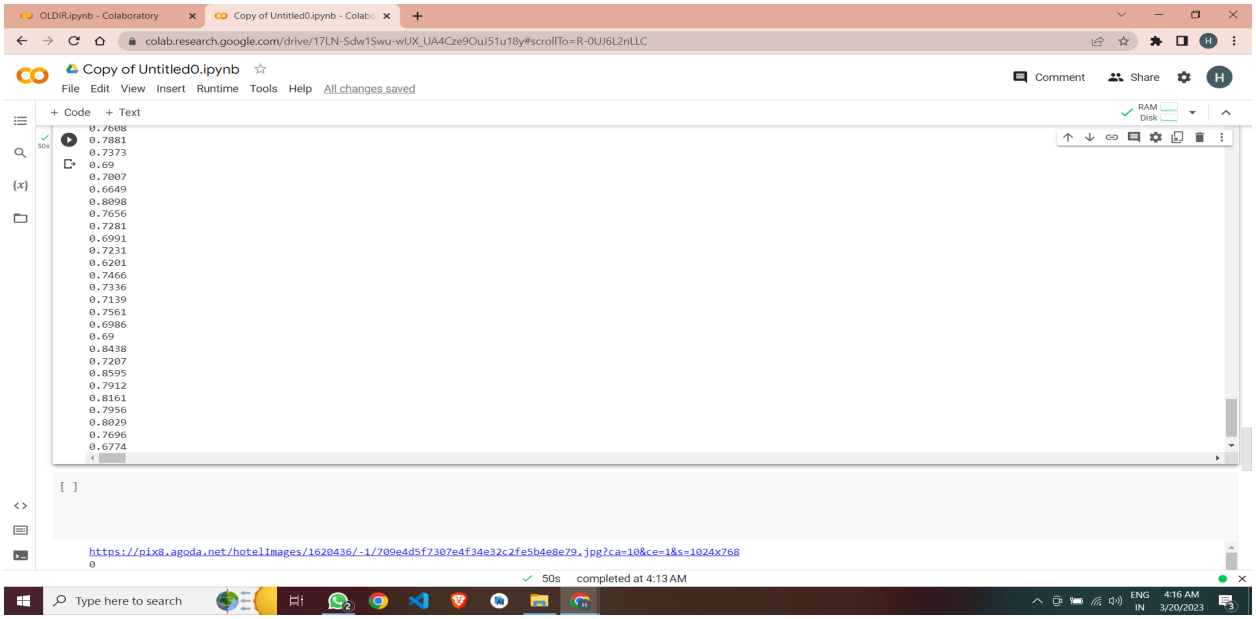
	details	Location	Name	reviews	Score	price	img_link	room_size	amenities	city	...	Private beach access	Carbon monoxide detector	Shared bathroom	Radio	Carpeting	Books/DVDs for ch
406	{'Location': 'No 1380 1 Near Morjim Football G...	No 1380 1 Near Morjim Football Ground, Tembwad...	Tropical Wave Hostel	[{'Review_Rating': '10.0', 'Review_Heading': '...	9.912	2,498	https://q-xx.bstatic.com/xdata/images/hotel/84...	14m²	['Garden view', 'Balcony/terrace', 'Shower', '...	Goa	...	0	0	0	0	1	
415	{'Location': 'Arambol Beach, Goa, India', 'ove...	Arambol Beach, Goa, India	Goym Resort	[{'Review_Rating': '10.0', 'Review_Heading': '...	9.814	3,820	https://pix8.agoda.net/hotelImages/35712803/58...	10m²	['Garden view', 'Balcony/terrace', 'Shower', '...	Goa	...	0	0	0	0	1	
465	{'Location': "Tito's lane 2, Saunta Vaddo, Bag...	Tito's lane 2, Saunta Vaddo, Baga Near Snow Pa...	Baga Bae	[{'Review_Rating': '10.0', 'Review_Heading': '...	9.783	1,800	https://q-xx.bstatic.com/xdata/images/hotel/84...	0	['Shower and bathtub', 'Free Wi-Fi', 'Toiletri...	Goa	...	0	0	0	0	0	
183	{'Location': 'Cola Goa Beach Resort, Agonda, G...	Cola Goa Beach Resort, Agonda, Goa, India, 403702	Cola Goa Beach Resort	[{'Review_Rating': '10.0', 'Review_Heading': '...	9.694	20,900	https://q-xx.bstatic.com/xdata/images/hotel/84...	30m²	['Sea view', 'Balcony/terrace', 'Shower and ba...	Goa	...	1	0	0	0	0	
410	{'Location': 'Survey No, 155/1, Arpora Baga, O...	Survey No, 155/1, Arpora Baga, Opp - Radisson ...	Caravan Baga Aqua Resort	[{'Review_Rating': '10.0', 'Review_Heading': '...	9.694	3,000	https://q-xx.bstatic.com/xdata/images/hotel/84...	30m²	['Pool view', 'Balcony/terrace', 'Shower', 'To...	Goa	...	0	0	0	0	0	

5 rows × 106 columns

Similarity detection:



This is the similarity of the image as output with others in a serial manner:


OLDIR.ipynb - Colaboratory
Copy of Untitled0.ipynb - Colab
colab.research.google.com/drive/17LN-Sdw1Swu-wUX_UA4Cze9OuJ51u18y#scrollTo=R-0UJ6L2nLLC
Copy of Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
0.7608
0.7881
0.7373
0.69
0.7007
0.6649
0.8098
0.7656
0.7281
0.6991
0.7231
0.6201
0.7466
0.7336
0.7139
0.7561
0.6986
0.69
0.8438
0.7207
0.8595
0.7912
0.8161
0.7956
0.8029
0.7696
0.6774
[ ]
https://pix8.agoda.net/hotelImages/1620436/-1/709e4d5f7307e4f34e32c2fe5b4e8e79.jpg?ca=10&ce=1&s=1024x768
0
50s completed at 4:13 AM
Type here to search
ENG
IN
4:16 AM
3/20/2023

Future Challenges

- This project will also be further personalised by additional ranking on the basis of location wherein we might rank districts on the basis of customer requirements(art, party etc) and include it within the ranking algorithm
- While VADER sentiment analysis excels at categorizing a wide range of syntactical patterns, it has a limitation in that it solely focuses on individual words without taking into consideration their surrounding context. As a rule-based classifier, VADER can only recognize specific scenarios that are pre-programmed into its algorithm. This means that it can account for the contextual effects of words like "but," but it may overlook conjunctions such as "although" or "however."

Individual Contribution:

- Nittin Yadav(202093) - Web Scraping Data Collection
- Meenal Gurbaxani(2019434) - Ranking System, Literature Review
- Dishant Yadav(2020057) - Data Preprocessing
- Raman Yadav(2020111) - Frontend UI
- Harsh(2020061) - Similar Image Detection
- Daevaang Khairwal(2020369) - Frontend UI

References:

[1] *Hotel recommendation system based on review and context information: A ...* (n.d.). Retrieved March 27, 2023, from https://core.ac.uk/download/pdf/301369618.pdf

[2] *Differential context relaxation for context-aware travel ... - researchgate*. (n.d.). Retrieved March 27, 2023, from https://www.researchgate.net/publication/230773488_Differential_Context_Relaxation_for_Context-Aware_Travel_Recommendation

[3] Jiang, L., Liu, L., Yao, J., & Shi, L. (2020, October 7). *A hybrid recommendation model in social media based on deep emotion analysis and multi-source view Fusion - Journal of Cloud Computing*. SpringerOpen. Retrieved March 27, 2023, from https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-020-00199-2

[4] *A hotel recommendation system based on collaborative filtering and ...* (n.d.). Retrieved March 27, 2023, from https://www.researchgate.net/publication/238516069_A_Hotel_Recommendation_System_Based_on_Collaborative_Filtering_and_Rankboost_Algorithm

[5] Hutto, C., & Gilbert, E. (n.d.). *Vader: A parsimonious rule-based model for sentiment analysis of social media text*. Proceedings of the International AAAI Conference on Web and Social Media. Retrieved March 27, 2023, from https://doi.org/10.1609/icwsm.v8i1.14550

[6] Beri, A. (2020, May 27). *Sentimental analysis using vader*. Medium. Retrieved March 27, 2023, from https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664

[7] Isbilen, E. (2019, December 29). *Image similarity detection in action with tensorflow 2.0*. Medium. Retrieved March 27, 2023, from https://towardsdatascience.com/image-similarity-detection-in-action-with-tensorflow-2-0-b8d9a78b2509

[8] Li, D., Yin, H., Wang, C., Song, S., Li, K., & Li, C. (2022). *Visual recommendation for peer-to-peer accommodation with online reviews based on sentiment analysis and topic models*. Journal of visualization. Retrieved March 27, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9127497/

[9] Journal, I. R. J. E. T. (2017, June 20). *Hotel recommendation based on User Preference Analysis*. 2015 31st IEEE International Conference on Data Engineering Workshops. Retrieved February 21, 2023, from https://www.academia.edu/33567241/Hotel_Recommendation_based_on_User_Preference_Analysis

[10] *Computing similarity on images using machine learning and cosine similarity*. Computing Similarity on Images using Machine Learning and Cosine Similarity. (2020, July 10). Retrieved March 13, 2023, from https://soumilshah1995.blogspot.com/2020/07/computing-similarity-on-images-using.html

[11] Author links open overlay panelJianshan Sun a, a, b, c, websites, A. S. media, Xia, R., Wang, G., Liu, H., Liu, J. G., Chen, L., Abbasi, A., Aciar, S., Adomavicius, G., Aspden, P., Baccianella, S., Balahur, A., Banker, C., Bell, R., Bennett, J., ... Li, Y. (2014, October 27). *Mining affective text to improve social media item recommendation*. Information Processing & Management. Retrieved March 13, 2023, from https://www.sciencedirect.com/science/article/abs/pii/S0306457314000879

[12] Popescu-belis, A. (2015, August 3). *Combining content with user preferences for non-fiction Multimedia recommendation: A study on ted lectures*. Academia.edu. Retrieved March 13, 2023, from https://www.academia.edu/14599858/Combining_content_with_user_preferences_for_non_fiction_multimedia_recommendation_a_study_on_TED_lectures

[13] *A hotel recommendation system based on collaborative filtering and ...* (n.d.). Retrieved March 13, 2023, from https://ieeexplore.ieee.org/document/5474286