HOTEL RECOMMENDATION SYSTEM

MINI PROJECT DOCUMENTATION

Team Members:
Sneha S (2019103583)
Shreya Ananth (2019103580)
Harini E R (2019103019)

for the course

CS 6029 – SOCIAL NETWORK ANALYSIS

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, COLLEGE OF ENGINEERING, GUINDY, ANNA UNIVERSITY, CHENNAI 600 025.

Contents

ABSTRACT……. 3

INTRODUCTION…… 4

DATASET. ……6

CODE EXECUTION SCREENSHOTS……..……..6

PERFORMANCE MEASURES…… ……..10

CONCLUSION…… 11

TITLE - HOTEL RECOMMENDATION SYSTEM

Abstract:

We all plan trips and the first thing to do when planning a trip is to book a good place to stay. Booking a hotel online can be an overwhelming task with thousands of hotels to choose from, for every destination. A hotel recommendation system aims at suggesting properties/hotels to a user such that they would prefer the recommended property over others. This helps in recommending hotels based upon user's needs instead of showing generalised results to all users. It is an application that understands the purpose of your next trip and recommends the best hotels based on the reviews and ratings of people who have stayed there for the same type of trip. It aims to find the best hotels and saves you time by showing you reviews and ratings of people who have stayed there.For example, suppose you want to go on a business trip, so the hotel recommendation system should show you the hotels that other customers have rated best for business travel. It is therefore also our approach to build a recommendation system based on customer reviews and ratings.

The aim of the project is to learn more about how the multiple attributes that describe a hotel can be used in recommendation systems. A large number of travel industries are benefiting from the recommendation systems in improving customer satisfaction and experience. We make use of content-based filtering to recommend hotels. The advantage of using a content-based recommendation system is that it doesn't have a cold-start problem.

Introduction:

The cold start problem is a well known and well researched problem for recommendation systems , where the system is not able to recommend items to users. due to three different situations i.e. for new users, for new products and for new websites.

Content-based filtering is the method that solves this problem. Our system first uses the metadata of new products when creating recommendations, while visitor action is secondary for a certain period of time. And our systems recommend a product to a user based upon the category and description of the product.

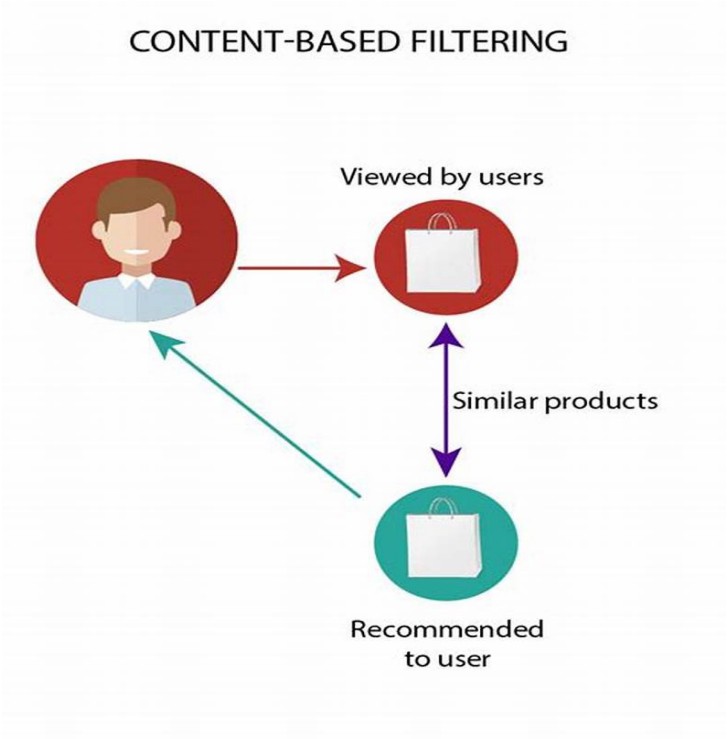
Content-based recommendation systems may be used in a variety of domains ranging from recommending web pages, news articles, restaurants, television programs, and hotels. The advantage of content-based filtering is that it doesn't have a cold-start problem. If you just start out a new website, or any new products can be recommended right away.

Let's assume we are starting a new online travel agency (OTA), and we have signed up thousands of hotels that are willing to sell on our platform, and we start seeing traffic coming from our website users, but we don't have any users history, therefore, we are going to build a content-based recommendation systems to analyse hotel descriptions to identify hotels that are of particular interest to the user.

We would like to recommend hotels based on the hotels that a user has already booked or viewed using the cosine similarity. We would recommend hotels with the largest similarity to the ones previously booked or viewed or showed interest by the user. Our recommender system is highly dependent on defining an appropriate similarity measure. Eventually, we select a subset of hotels to display to the user or to determine an order in which to display the hotels.

The main idea of content-based methods is to try to build a model, based on the available "features", that explain the observed user-item interactions.

User profiles are constructed using historical interactions or by explicitly asking users about their interests. There are other systems, not considered purely content-based, which utilize user personal and social data.

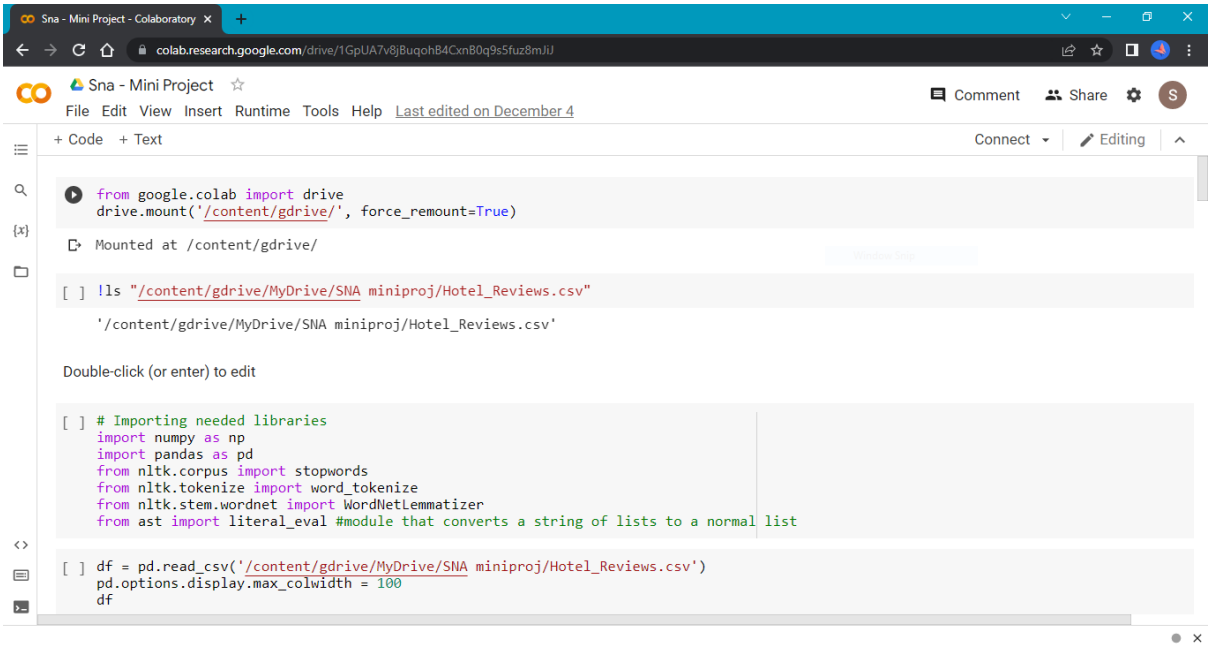


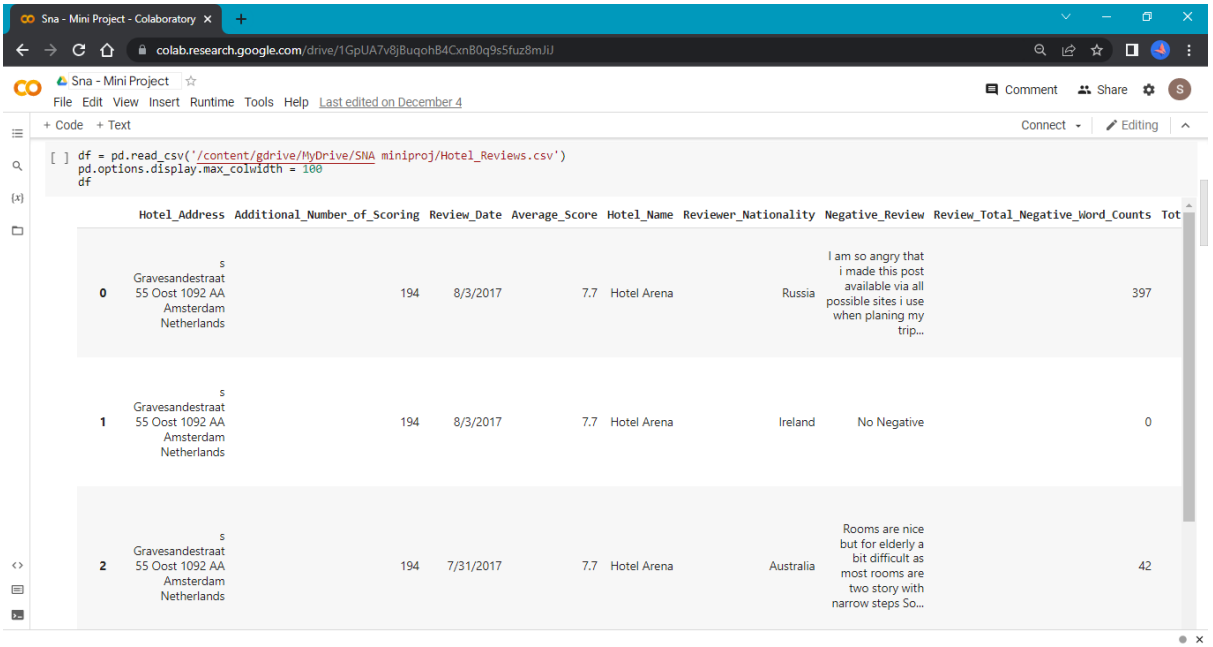
Information about Dataset:

The dataset used in this project was accessed from kaggle name '**515K Hotel Reviews Data in Europe'.** This dataset contains 5,15,000+ reviews that were collected by authenticated hotel reviewers that specify the various features of the hotel. Taking this data into consideration as values for a 'content based recommendation system', we compare the user desired specification to get the top 10 similar hotels that are in the locality along with the average_score of the recommended hotels.

Dataset link - [515K Hotel Reviews Data in Europe]

Screenshots of our complete work:





	Hotel_Address	Additional_Number_of_Scoring	Review_Date	Average_Score	Hotel_Name	Reviewer_Nationality	Negative_Review	Review_Total_Negative_Word_Counts	Tot
0	s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands	194	8/3/2017	7.7	Hotel Arena	Russia	I am so angry that i made this post available via all possible sites i use when planing my trip...	397	
1	s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands	194	8/3/2017	7.7	Hotel Arena	Ireland	No Negative	0	
2	s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands	194	7/31/2017	7.7	Hotel Arena	Australia	Rooms are nice but for elderly a bit difficult as most rooms are two story with narrow steps So...	42	

Sna - Mini Project - Colaboratory

colab.research.google.com/drive/1GpUA7v8jBuqohB4CxnB0q9s5fuz8mJiJ

Sna - Mini Project

File Edit View Insert Runtime Tools Help Last edited on December 4

Comment Share

+ Code + Text

Connect Editing

```
[ ] # Replacing 'united kingdom' with 'UK' for easy use
    df.Hotel_Address = df.Hotel_Address.str.replace('United Kingdom','UK')
    # Splitting the hotel address and picking out the last string which would be the countries
    df['countries'] = df.Hotel_Address.apply(lambda x: x.split(' ')[-1])
    df.countries.unique() # All the hotels are located in six(6) countries
```

```
array(['Netherlands', 'UK', 'France', 'Spain', 'Italy', 'Austria'],
      dtype=object)
```

```
[ ] # Dropping unneeded columns
    df.drop(['Additional_Number_of_Scoring',
           'Review_Date','Reviewer_Nationality',
           'Negative_Review', 'Review_Total_Negative_Word_Counts',
           'Total_Number_of_Reviews', 'Positive_Review',
           'Review_Total_Positive_Word_Counts',
           'Total_Number_of_Reviews_Reviewer_Has_Given', 'Reviewer_Score',
           'days_since_review', 'lat', 'lng'],1,inplace=True)
```

```
<ipython-input-10-18b2681f6ec5>:2: FutureWarning: In a future version of pandas all arguments of DataFrame.drop except for the argume
  df.drop(['Additional_Number_of_Scoring',
```

```
[ ] df
```

Sna - Mini Project - Colaboratory

colab.research.google.com/drive/1GpUA7v8jBuqohB4CxnB0q9s5fuz8mJiJ

Sna - Mini Project

File Edit View Insert Runtime Tools Help Last edited on December 4

Comment Share

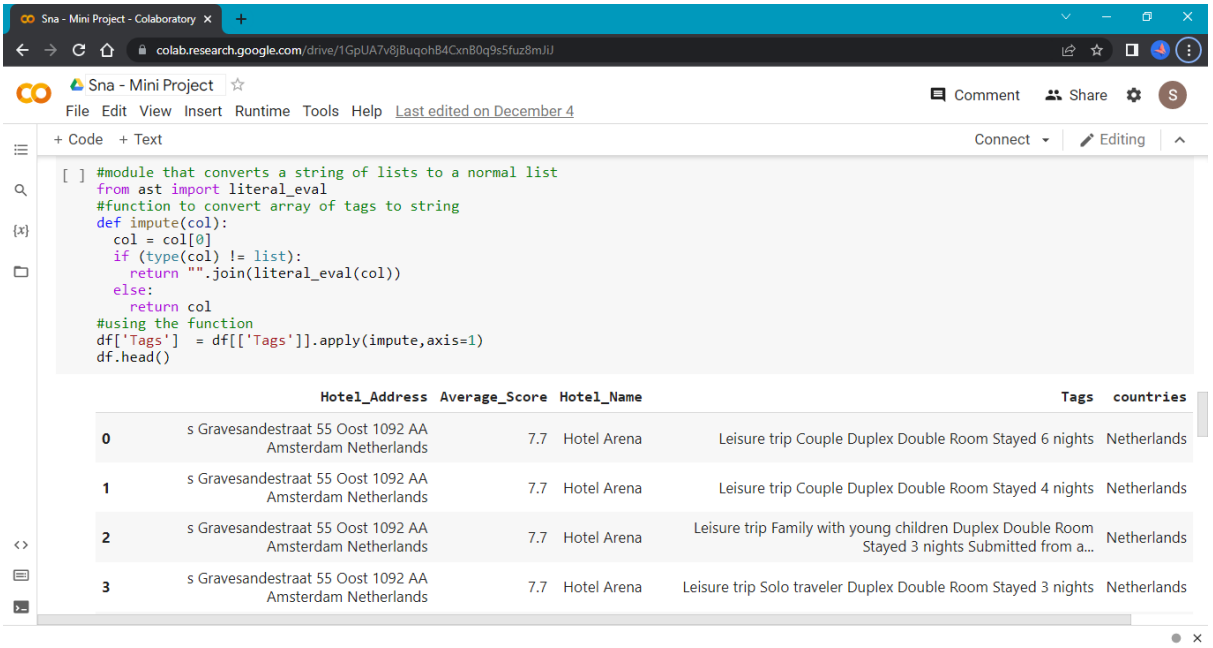
+ Code + Text

Connect Editing

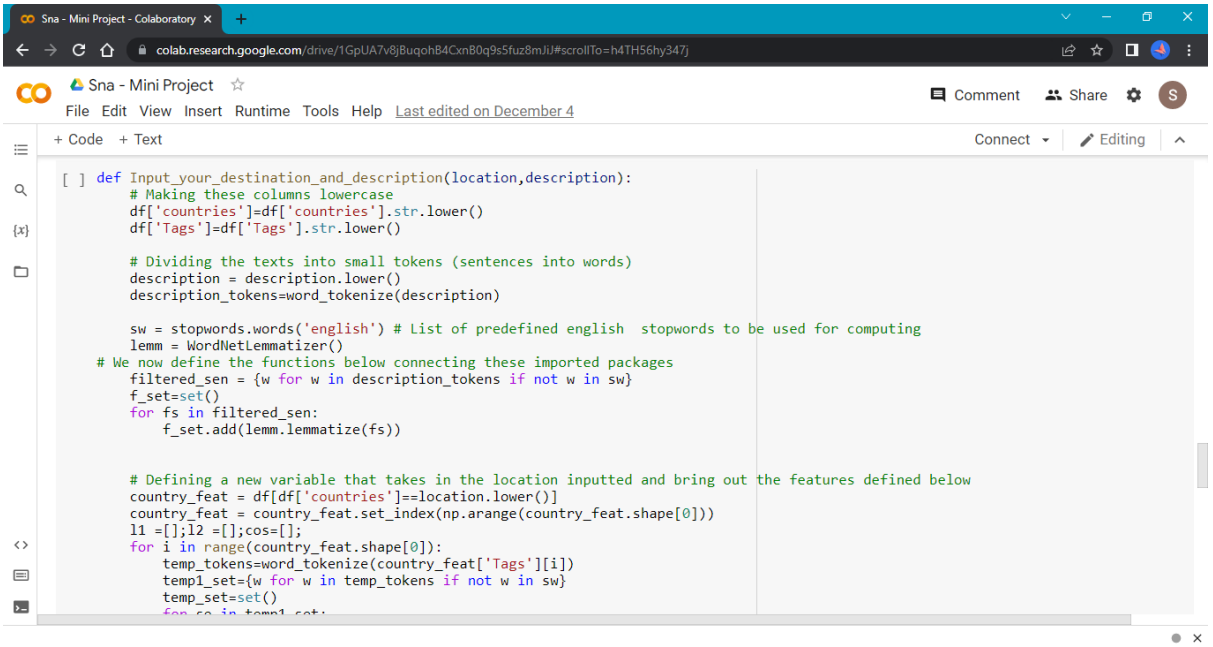
```
df
```

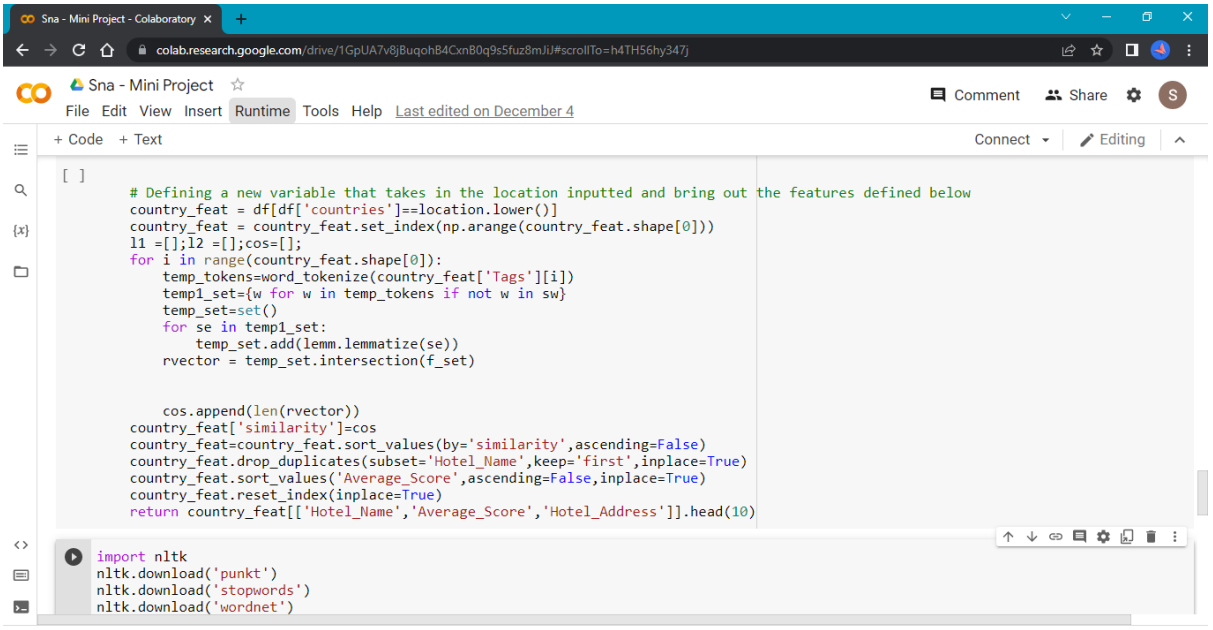
	Hotel_Address	Average_Score	Hotel_Name	Tags	countries
0	s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands	7.7	Hotel Arena	[' Leisure trip ', ' Couple ', ' Duplex Double Room ', ' Stayed 6 nights ']	Netherlands
1	s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands	7.7	Hotel Arena	[' Leisure trip ', ' Couple ', ' Duplex Double Room ', ' Stayed 4 nights ']	Netherlands
2	s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands	7.7	Hotel Arena	[' Leisure trip ', ' Family with young children ', ' Duplex Double Room ', ' Stayed 3 nights ', ...	Netherlands
3	s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands	7.7	Hotel Arena	[' Leisure trip ', ' Solo traveler ', ' Duplex Double Room ', ' Stayed 3 nights ']	Netherlands
4	s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands	7.7	Hotel Arena	[' Leisure trip ', ' Couple ', ' Suite ', ' Stayed 2 nights ', ' Submitted from a mobile device ']	Netherlands
...	...	...	...	...	...
515733	Wurzbachgasse 21 15 Rudolfsheim F nfhaus 1150 Vienna Austria	8.1	Atlantis Hotel Vienna	[' Leisure trip ', ' Family with older children ', ' 2 rooms ', ' Stayed 5 nights ']	Austria
515734	Wurzbachgasse 21 15 Rudolfsheim F nfhaus 1150 Vienna Austria	8.1	Atlantis Hotel Vienna	[' Leisure trip ', ' Family with young children ', ' Standard Triple Room ', ' Stayed 2 nights ']	Austria

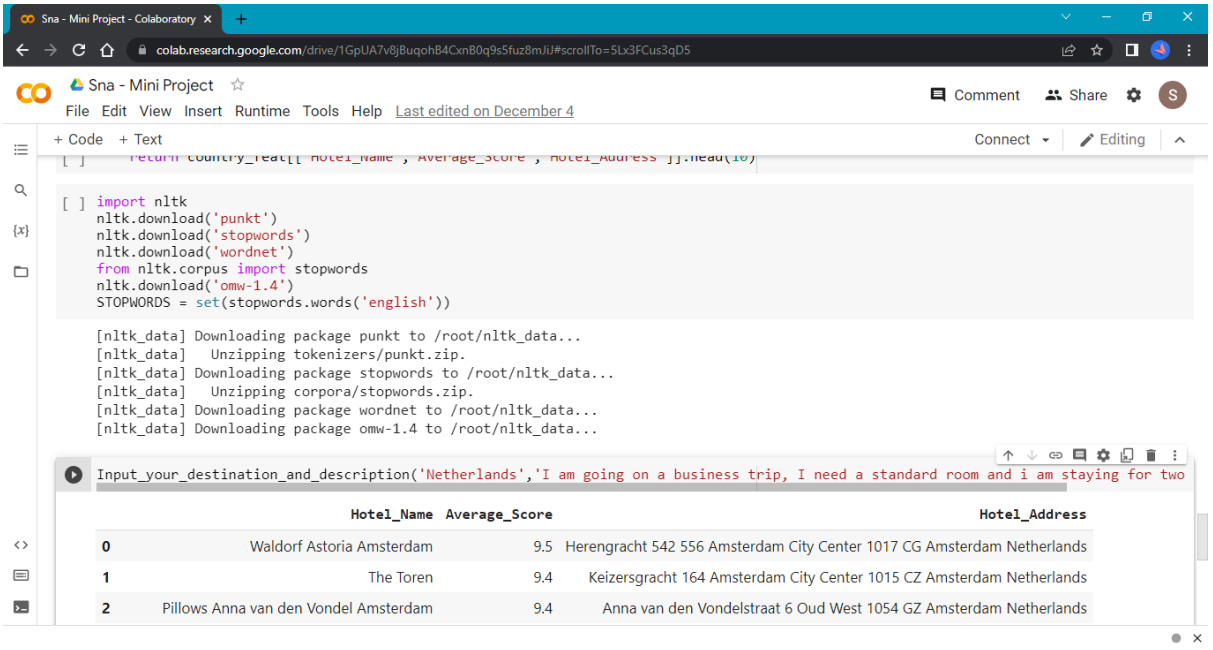
ENG IN 03:01 PM 14-Dec-22



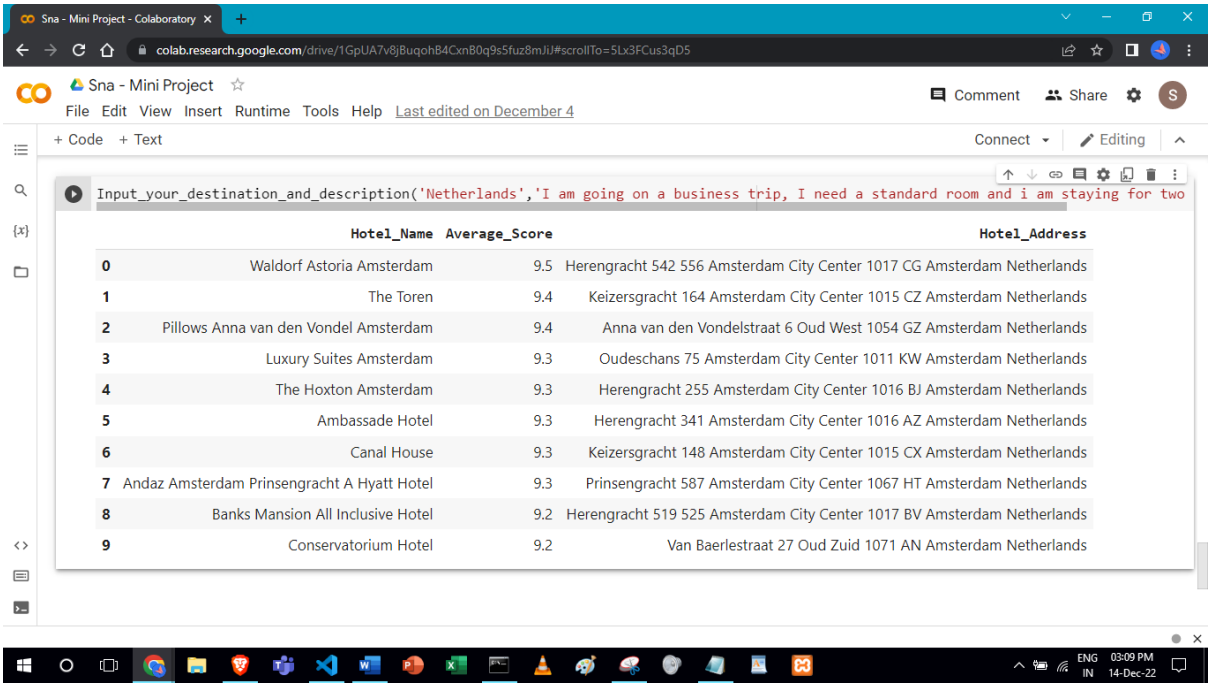
	Hotel_Address	Average_Score	Hotel_Name	Tags	countries
0	s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands	7.7	Hotel Arena	Leisure trip Couple Duplex Double Room Stayed 6 nights	Netherlands
1	s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands	7.7	Hotel Arena	Leisure trip Couple Duplex Double Room Stayed 4 nights	Netherlands
2	s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands	7.7	Hotel Arena	Leisure trip Family with young children Duplex Double Room Stayed 3 nights Submitted from a...	Netherlands
3	s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands	7.7	Hotel Arena	Leisure trip Solo traveler Duplex Double Room Stayed 3 nights	Netherlands





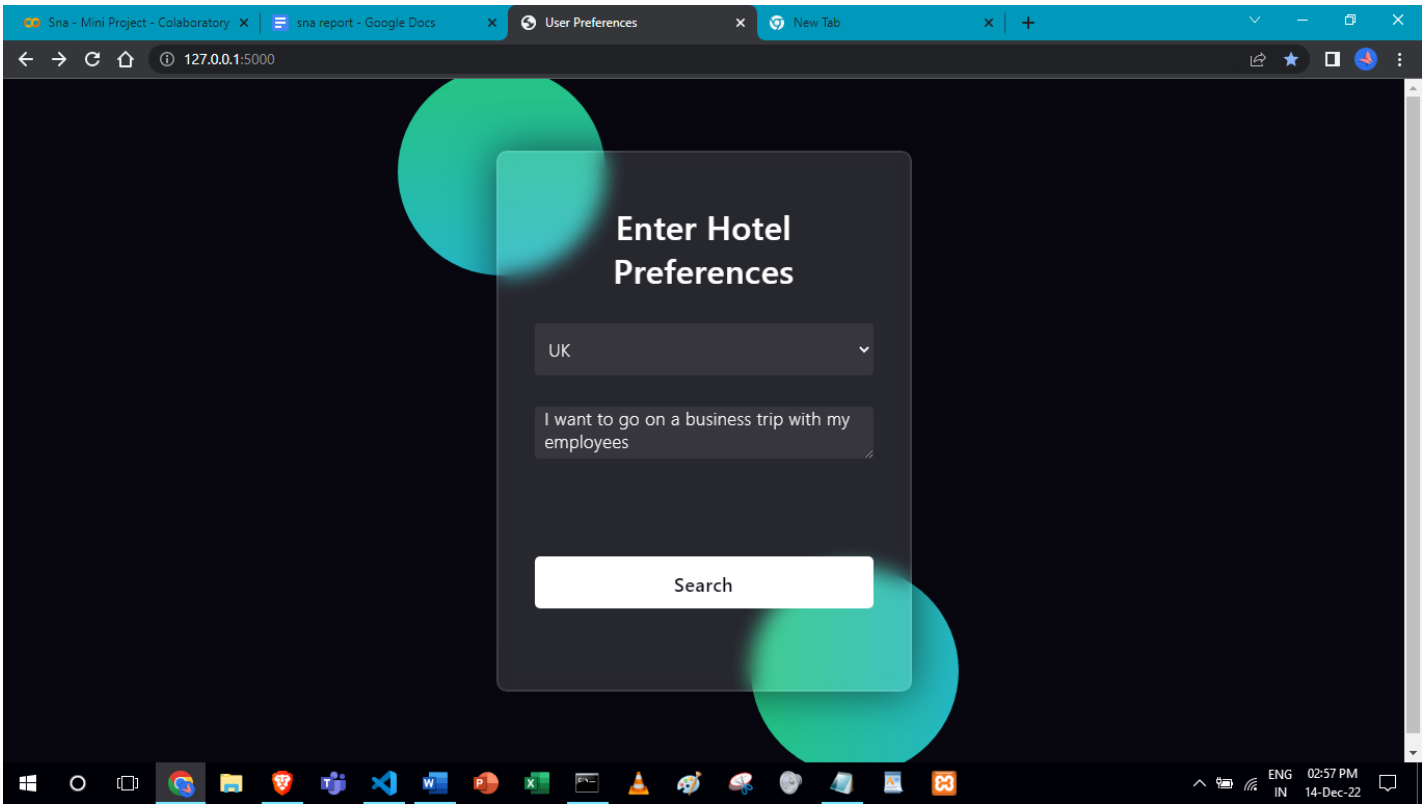


	Hotel_Name	Average_Score	Hotel_Address
0	Waldorf Astoria Amsterdam	9.5	Herengracht 542 556 Amsterdam City Center 1017 CG Amsterdam Netherlands
1	The Toren	9.4	Keizersgracht 164 Amsterdam City Center 1015 CZ Amsterdam Netherlands
2	Pillows Anna van den Vondel Amsterdam	9.4	Anna van den Vondelstraat 6 Oud West 1054 GZ Amsterdam Netherlands

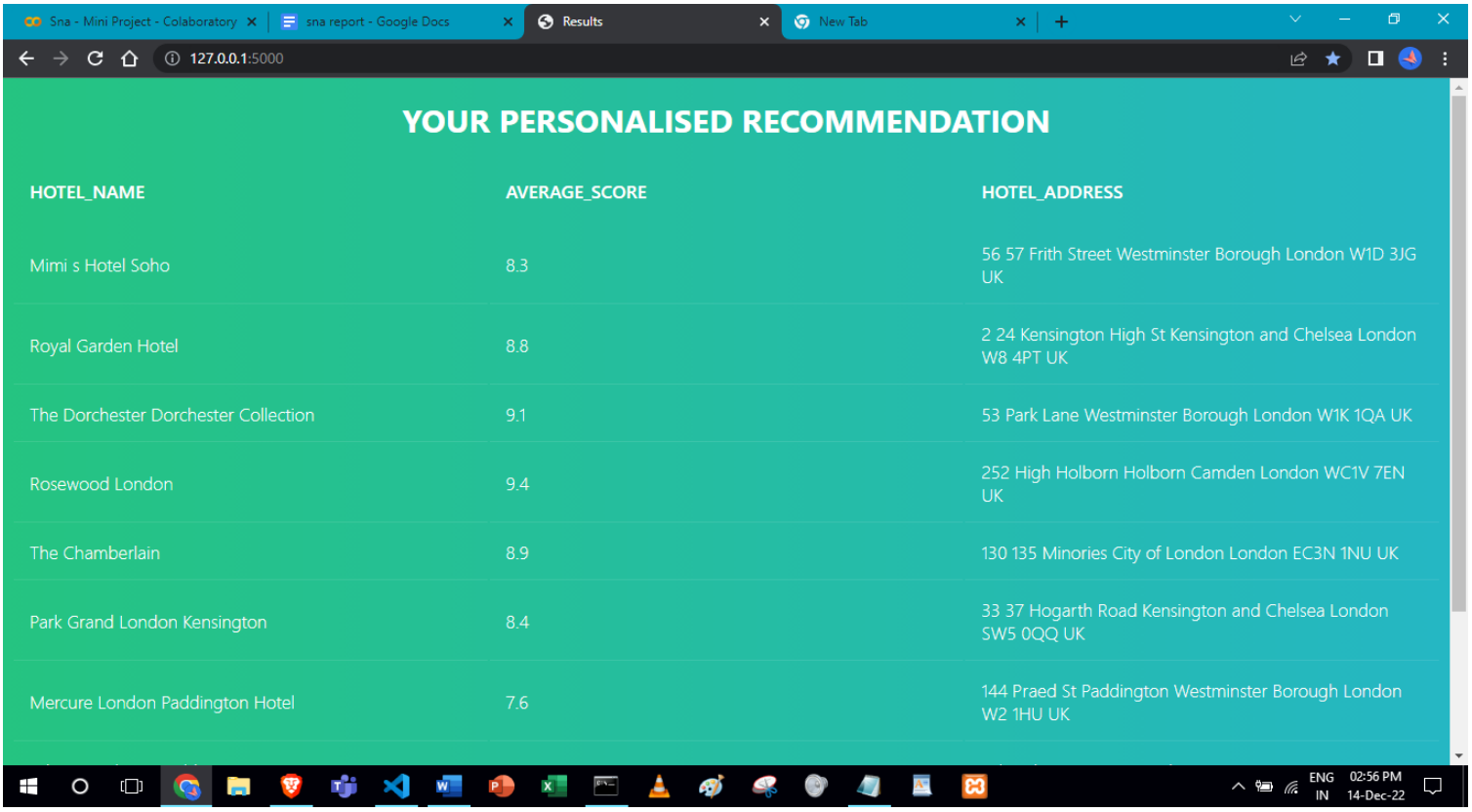


The above code is connected to a Flask Application that gets the user's desired location and the description and displays the top 10 hotels that match the hotel description and are in the same city as the user

The User Input Form:



The Top 10 recommended hotels:



Performance measures:

Recommendation System accuracy is popularly evaluated through two main measures: Root Mean Squared Error (RMSE) and Mean Absolute Error(MAE).

$$\mathrm{MAE} = \frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} |r_{ui} - \hat{r}_{ui}|$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} (r_{ui} - \hat{r}_{ui})^2}.$$

Evaluating the performance of **unsupervised models** is a complex task. While for supervised learning the **labelled** data (ground truth) can be directly used as a target measure, it is much harder to quantify the performance of a certain method when such labelled data is only scarcely available or even not available at all.

With the help of domain experts, we can verify the correctness of our system. Based on our knowledge of these restaurants, we found that the recommendation system gave accurate suggestions.

Conclusion:

Therefore, content-based approaches are highly efficient in hotel recommendation systems. As we can see, the top ten most similar restaurants are returned by the recommendation after considering factors like location and the user specified description.