Mahatma Education Society's
Pillai College of Engineering, Panvel
Department of Computer Engineering's

A
Presentation on

Hotel Recommendation System

Under the Guidance of
Prof. Asmita Deshmukh

Group Members :-
Aniket Surve (711)
Prajwal Patil (704)
Shreyas Varadkar (755)
Yashraj Sinha (749)

Problem Statement

- A hotel recommendation system aims to predict which hotel a user is most likely to choose from among all hotels.
- Recommendation System is an information filtering technique, which provide users with information, which he/she may be interested in.

PROCESS FLOW

- Importing the necessary libraries and reading in the dataset.
- Useless data deleted
- Relevant data is made such that it is easily accessible
- Words are assigned values
- Hotels identified based on user input using text vectorization
- Model identifies the best hotel among the hotels identified

DATASET

- This dataset contains 515,738 hotel's.
- The dataset contain countries of Europe such as United Kingdom, Netherland, Spain, Austria, Italy.
- As we are going to build a recommendation system according to the user ratings so here will be using Natural Language Processing.

Featured Datasets

(8743 rows X 17 columns) Features: Hotel_Name, Hotel_Address, Average_Score, Countries, Tags etc

```
hotel.head()
```

	Hotel_Address	Additional_Number_of_Scoring	Review_Date	Average_Score	Hotel_Name	Reviewer_Nationality	Negative_Review	Review_Total_Negative_Word_Counts	Tota
0	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	194	8/3/2017	7.7	Hotel Arena	Russia	I am so angry that i made this post available...	397.0	
1	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	194	8/3/2017	7.7	Hotel Arena	Ireland	No Negative	0.0	
2	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	194	7/31/2017	7.7	Hotel Arena	Australia	Rooms are nice but for elderly a bit difficul...	42.0	
3	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	194	7/31/2017	7.7	Hotel Arena	United Kingdom	My room was dirty and I was afraid to walk ba...	210.0	
4	s Gravesandestraat 55 Oost 1092 AA	194	7/24/2017	7.7	Hotel Arena	New Zealand	You When I booked with your company on line	140.0	

Data Exploration

Removing Null Values.

Use Column function of pandas.

```
[12] hotel.head()
```

	Hotel_Name	Average_Score	Hotel_Address	Total_Number_of_Reviews	Tags	lat	lng
0	Hotel Arena	7.7	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	1403.0	[' Leisure trip ', ' Couple ', ' Duplex Double...	52.360576	4.915968
1	Hotel Arena	7.7	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	1403.0	[' Leisure trip ', ' Couple ', ' Duplex Double...	52.360576	4.915968
2	Hotel Arena	7.7	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	1403.0	[' Leisure trip ', ' Family with young childre...	52.360576	4.915968
3	Hotel Arena	7.7	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	1403.0	[' Leisure trip ', ' Solo traveler ', ' Duplex...	52.360576	4.915968
4	Hotel Arena	7.7	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	1403.0	[' Leisure trip ', ' Couple ', ' Suite ', ' St...	52.360576	4.915968

```
[12]
```

```
[13] hotel.isnull().sum()
```

```
Hotel_Name                  0
Average_Score               0
Hotel_Address               0
Total_Number_of_Reviews     1
Tags                        1
lat                         1
lng                         1
dtype: int64
```

Data Cleaning

Remove the columns that are not needed.

Drop null values and empty values.

Use the describe function to identify unique and top hotels.

```
[26] hotel.head()
```

	Hotel_Name	Average_Score	Hotel_Address	Total_Number_of_Reviews	Tags	lat	lng	countries
0	Hotel Arena	7.7	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	1403.0	[' leisure trip ', ' couple ', ' duplex double...	52.360576	4.915968	netherlands
1	Hotel Arena	7.7	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	1403.0	[' leisure trip ', ' couple ', ' duplex double...	52.360576	4.915968	netherlands
2	Hotel Arena	7.7	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	1403.0	[' leisure trip ', ' family with young childre...	52.360576	4.915968	netherlands
3	Hotel Arena	7.7	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	1403.0	[' leisure trip ', ' solo traveler ', ' duplex...	52.360576	4.915968	netherlands
4	Hotel Arena	7.7	s Gravesandestraat 55 Oost 1092 AA Amsterdam ...	1403.0	[' leisure trip ', ' couple ', ' suite ', ' st...	52.360576	4.915968	netherlands

```
[27] hotel['Tags'][0]
```

```
'[' leisure trip ', ' couple ', ' duplex double room ', ' stayed 6 nights ']'
```

Apply Cosine Similarity to identify similar document in text analytics.

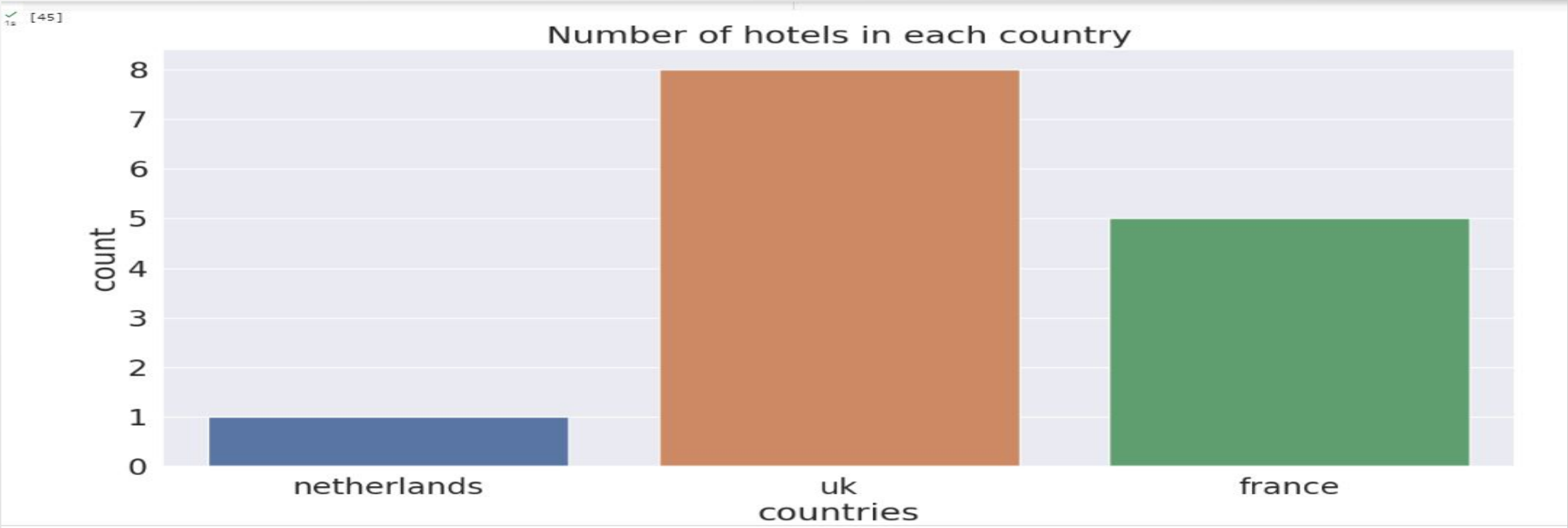
```
def recommend_hotel(location, description):
    description = description.lower()
    word_tokenize(description)
    stop_words = stopwords.words('english')
    lemm = WordNetLemmatizer()
    filtered  = {word for word in description if not word in stop_words}
    filtered_set = set()
    for fs in filtered:
        filtered_set.add(lemm.lemmatize(fs))

    country = hotel[hotel['countries']==location.lower()]
    country = country.set_index(np.arange(country.shape[0]))
    list1 = []; list2 = []; cos = [];
    for i in range(country.shape[0]):
        temp_token = word_tokenize(country["Tags"][i])
        temp_set = [word for word in temp_token if not word in stop_words]
        temp2_set = set()
        for s in temp_set:
            temp2_set.add(lemm.lemmatize(s))
        vector = temp2_set.intersection(filtered_set)
        cos.append(len(vector))
    country['similarity']=cos
    country = country.sort_values(by='similarity', ascending=False)
    country.drop_duplicates(subset='Hotel_Name', keep='first', inplace=True)
    country.sort_values('Average_Score', ascending=False, inplace=True)
    country.reset_index(inplace=True)
    return country[["Hotel_Name", "Average_Score", "Hotel_Address","Total_Number_of_Reviews","lat","lng"]].head()
```

Data Visualization

To identify count of hotels in each country.

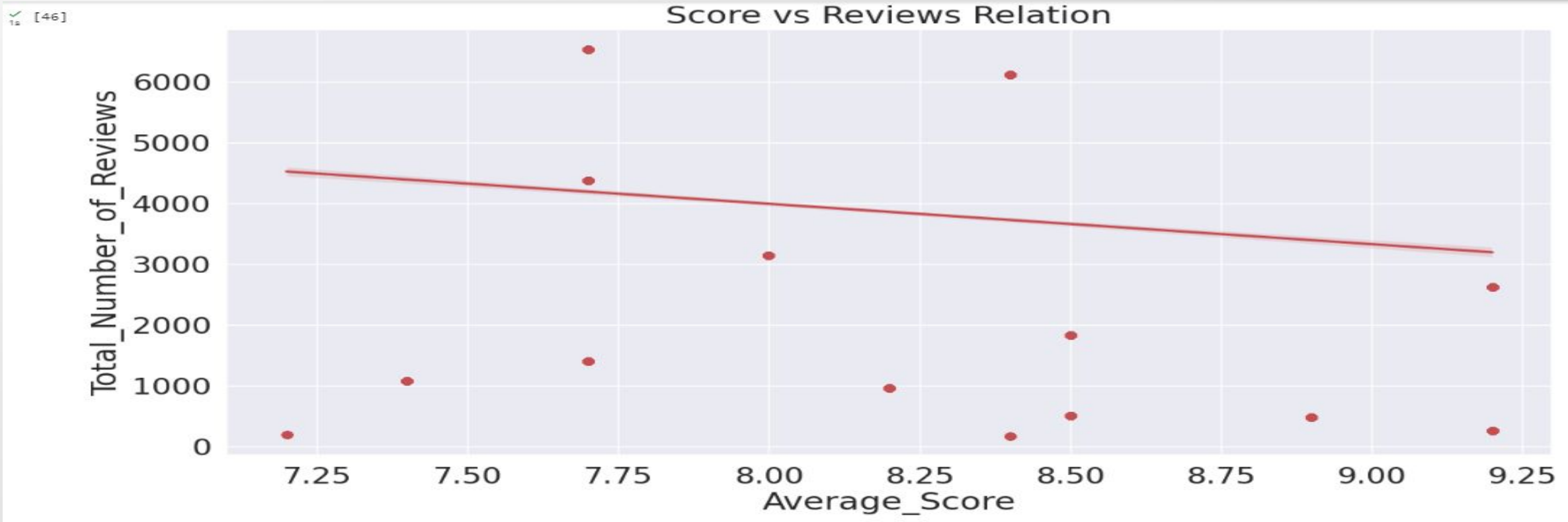
This bar graph shows count of hotels in each countries .



Data Visualization

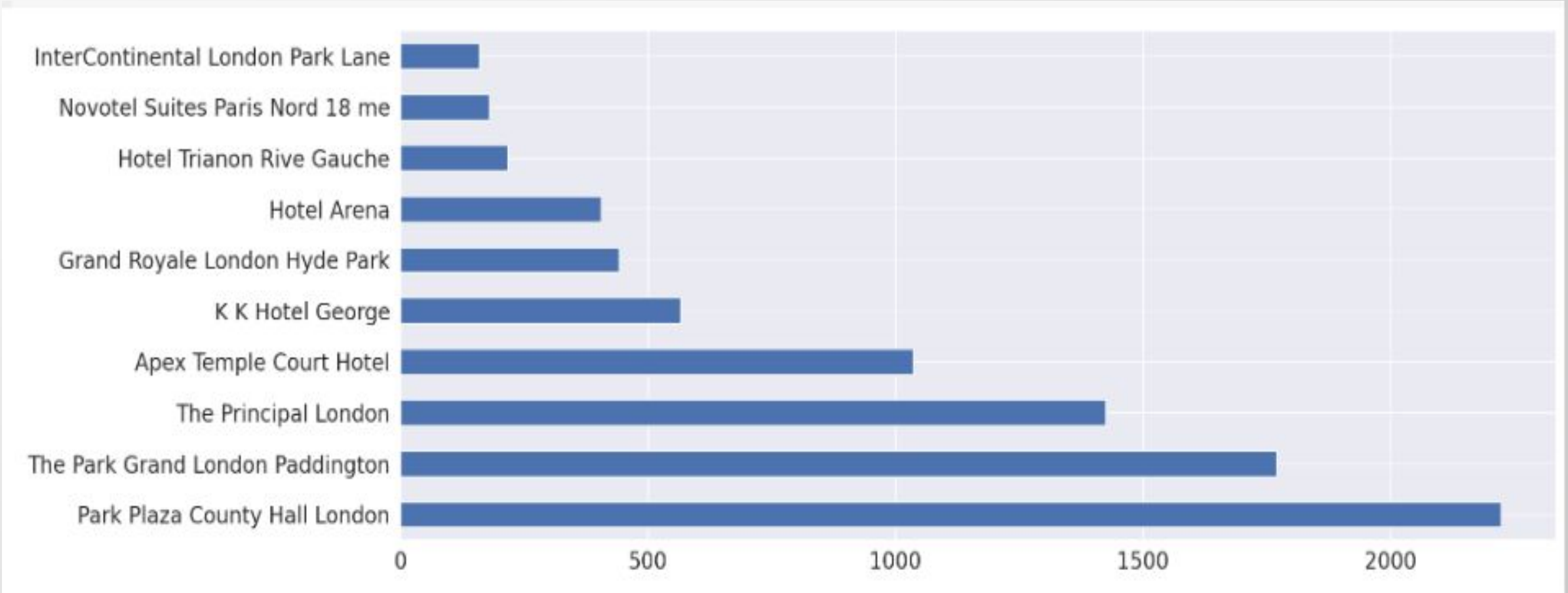
This regression line show the relationship between Average score and

Total Number of Reviews that help users to identify best hotel according to purpose of his visit



Data Visualization

This graph shows the hotels available of specific type



EXISTING SYSTEM

- In the existing system there was no option for the purpose of visit.
- So we have developed a model where the user can state the purpose of visit and based on that the hotels are predicted

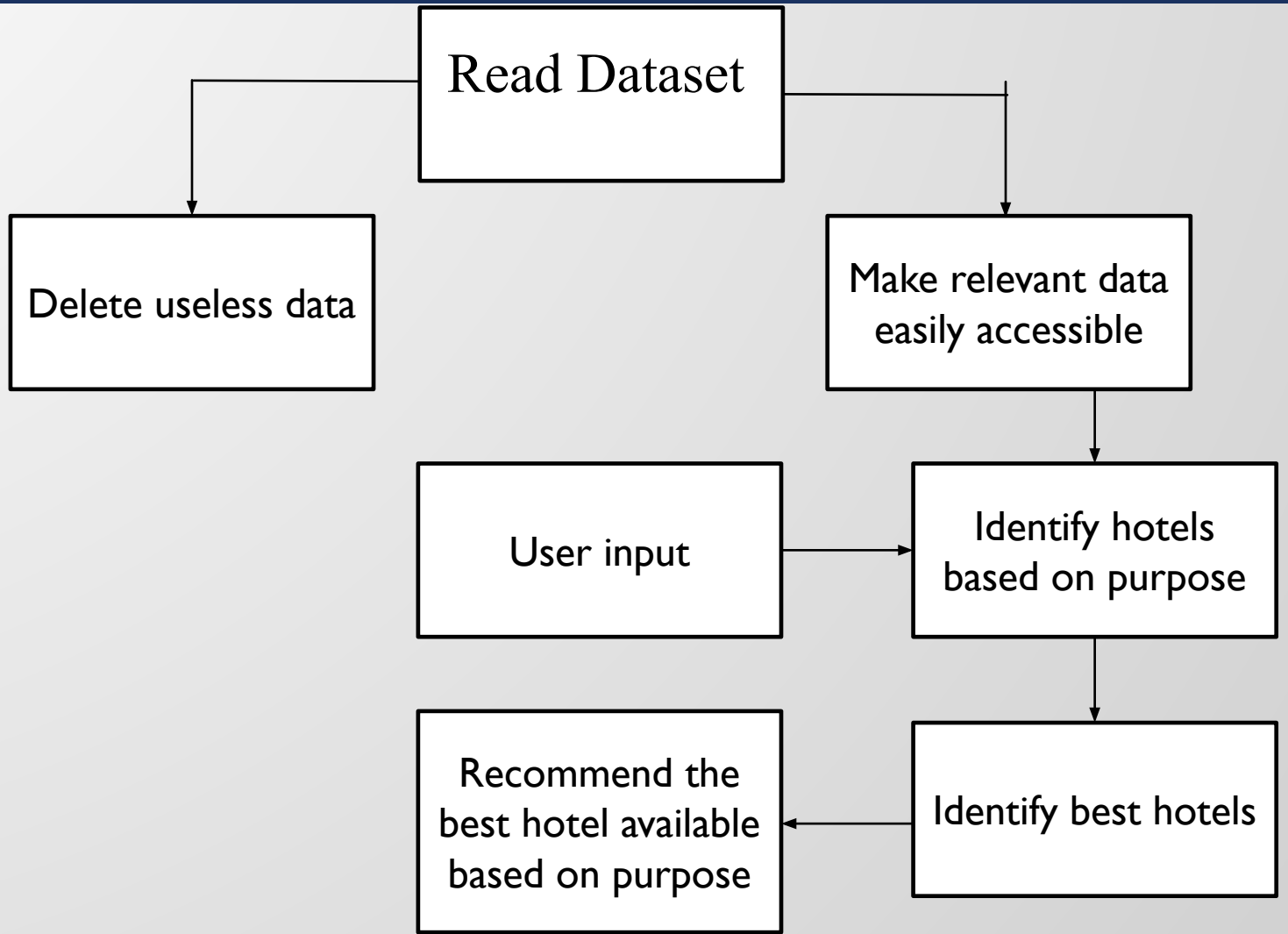
PROPOSED SYSTEM

- Creating a system to recommend a hotel based on client's location and type of trip
- It recommends the best rated hotels based on the following criteria

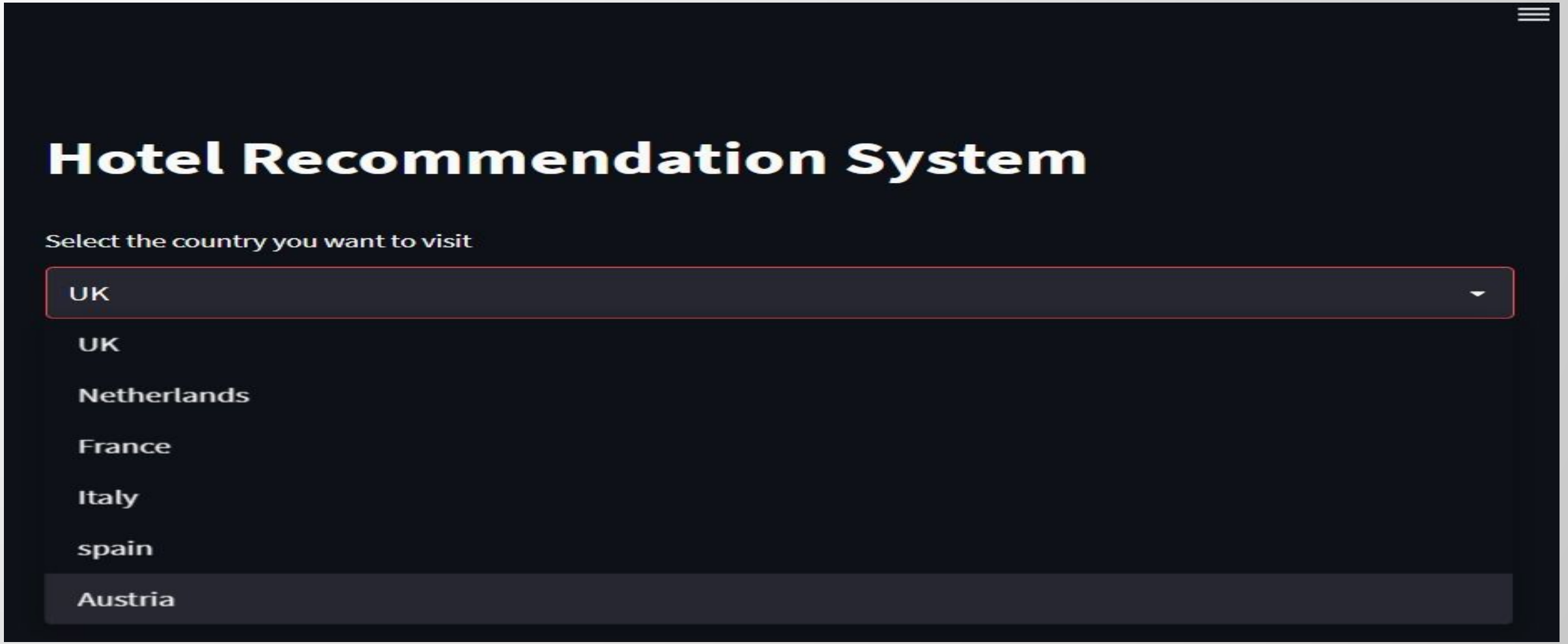
1) Purpose of Visit.

2) Countries.

FLOW DIAGRAM



OUTPUT


Hotel Recommendation System
Select the country you want to visit
UK
UK
Netherlands
France
Italy
spain
Austria

CONCLUSION

- We have successfully implemented a hotel recommendation system.

- Increases efficiency
- User friendly
- Reduces human efforts

HARDWARE / SOFTWARE

- SOFTWARE
  - OS : Windows 10
  - Language used : PYTHON

  Anaconda, Jupyter Notebook, PyCharm

  Frontend : Python, Streamlit

- HARDWARE
  - Minimum 4 GB RAM

THANKYOU