

```
import numpy as np
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn import svm
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from sklearn.metrics.pairwise import linear_kernel, cosine_similarity
from nltk.stem.snowball import SnowballStemmer
from nltk.stem.wordnet import WordNetLemmatizer
from nltk.corpus import wordnet
from nltk.stem import WordNetLemmatizer
```

```
hotel_details=pd.read_csv('/content/Hotel_details.csv',delimiter=',')
hotel_rooms=pd.read_csv('/content/Hotel_Room_attributes.csv',delimiter=',')
hotel_cost=pd.read_csv('/content/hotels_RoomPrice.csv',delimiter=',')
```

```
hotel_details.head()
```

|   | id    | hotelid | hotelname                    | address                | city    | country | zipcode | propertytype  | starrating | latitude  | longitude | Source       |
|---|-------|---------|------------------------------|------------------------|---------|---------|---------|---------------|------------|-----------|-----------|--------------|
| 0 | 46406 | 1771651 | Mediterranean Bungalow Galeb | Vukovarska 7           | Omis    | Croatia | 21310.0 | Holiday parks | 4          | 43.440124 | 16.682505 | 2 https://ww |
| 1 | 46407 | 177167  | Hotel Polonia                | Plac Teatralny 5       | Torun   | Poland  | NaN     | Hotels        | 3          | 53.012329 | 18.603800 | 5 ht         |
| 2 | 46408 | 1771675 | Rifugio Sass Bece            | Belvedere del Pordoi,1 | Canazei | Italy   | 38032.0 | Hotels        | 3          | 46.477920 | 11.813350 | 2 http://www |
| 3 | 46409 | 177168  | Madalena Hotel               | Mykonos                | Mykonos | Greece  | 84600.0 | Hotels        | 3          | 37.452316 | 25.329849 | 5 https://   |
| 4 | 46410 | 1771718 | Pension Morenfeld            | Mair im Korn Strasse 2 | Lagundo | Italy   | 39022.0 | Hotels        | 3          | 46.682780 | 11.131736 | 2 http://v   |

```
hotel_rooms.head()
```

|   | id       | hotelcode | roomamenities                                     | roomtype      | ratedescription                                   |
|---|----------|-----------|---|---------------|---|
| 0 | 50677497 | 634876    | Air conditioning: ;Alarm clock: ;Carpeting: ;C... | Double Room   | Room size: 15 m²/161 ft², Shower, 1 king bed      |
| 1 | 50672149 | 8328096   | Air conditioning: ;Closet: ;Fireplace: ;Free W... | Vacation Home | Shower, Kitchenette, 2 bedrooms, 1 double bed ... |
| 2 | 50643430 | 8323442   | Air conditioning: ;Closet: ;Dishwasher: ;Firep... | Vacation Home | Shower, Kitchenette, 2 bedrooms, 1 double bed ... |
| 3 | 50650047 | 7075      | Air conditioning: ;Clothes                        | Standard      | Room size: 20 m²/215 ft².                         |

Data Cleaning and transformations

```
del hotel_details['id']
del hotel_rooms['id']
del hotel_details['zipcode']
```

```
hotel_details=hotel_details.dropna()
hotel_rooms=hotel_rooms.dropna()
```

```
hotel_details.drop_duplicates(subset='hotelid',keep=False,inplace=True)
```

```
hotel=pd.merge(hotel_rooms,hotel_details,left_on='hotelcode',right_on='hotelid',how='inner')
hotel.columns
```

```
Index(['hotelcode', 'roomamenities', 'roomtype', 'ratedescription', 'hotelid',
      'hotelname', 'address', 'city', 'country', 'propertytype', 'starrating',
      'latitude', 'longitude', 'Source', 'url', 'curr'],
      dtype='object')
```

```
del hotel['hotelid']
del hotel['url']
```

```
del hotel['curr']
del hotel['Source']

hotel.columns

Index(['hotelcode', 'roomamenities', 'roomtype', 'ratedescription',
      'hotelname', 'address', 'city', 'country', 'propertytype', 'starrating',
      'latitude', 'longitude'],
      dtype='object')
```

Recommender system based only on City and ratings about the hotel

```
def citybased(city):
    hotel['city']=hotel['city'].str.lower()
    citybase=hotel[hotel['city']==city.lower()]
    citybase=citybase.sort_values(by='starrating',ascending=False)
    citybase.drop_duplicates(subset='hotelcode',keep='first',inplace=True)
    if(citybase.empty==0):
        hname=citybase[['hotelname','starrating','address','roomamenities','ratedescription']]
        return hname.head()
    else:
        print('No Hotels Available')
```

```
print('Top 5 hotels')
citybased('London')
```

Top 5 hotels

|       | hotelname                          | starrating | address               | roomamenities   | ratedescription   |
|-------|------------------------------------|------------|-----------------------|---|---|
| 48985 | The Colonnade London Hotel         | 4          | 2 Warrington Crescent | Additional toilet;<br>;Air conditioning;<br>;Alarm c... | Room size: 28 m²/301 ft²,<br>Balcony/terrace,<br>Non... |
| 58663 | South Point Suites - London Bridge | 4          | Bermondsey Street     | Air conditioning;<br>;Alarm clock;<br>;Blackout curt... | Room size: 24 m²/258 ft², Street view, Non-smo...       |

```
room_no=[
    ('king',2),
    ('queen',2),
    ('triple',3),
    ('master',3),
    ('family',4),
    ('murphy',2),
    ('quad',4),
    ('double-double',4),
    ('mini',2),
    ('studio',1),
    ('junior',2),
    ('apartment',4),
    ('double',2),
    ('twin',2),
    ('double-twin',4),
    ('single',1),
    ('diabled',1),
    ('accessible',1),
    ('suite',2),
    ('one',2)
]

def calc():
    guests_no=[]
    for i in range(hotel.shape[0]):
        temp=hotel['roomtype'][i].lower().split()
        flag=0
        for j in range(len(temp)):
            for k in range(len(room_no)):
                if temp[j]==room_no[k][0]:
                    guests_no.append(room_no[k][1])
                    flag=1
                    break
            if flag==1:
                break
        if flag==0:
            guests_no.append(2)
    hotel['guests_no']=guests_no
```

```
calc()

def pop_citybased(city,number):
    hotel['city']=hotel['city'].str.lower()
    popbased=hotel[hotel['city']==city.lower()]
    popbased=popbased[popbased['guests_no']==number].sort_values(by='starrating',ascending=False)
    popbased.drop_duplicates(subset='hotelcode',keep='first',inplace=True)
    if popbased.empty==True:
        print('Sorry No Hotels Available\n tune your constraints')
    else:
        return popbased[['hotelname','roomtype','guests_no','starrating','address','roomamenities','ratedescription']].head(10)

pop_citybased('London',4)
```

|        | hotelname                                      | roomtype                          | guests_no | starrating | address                   | roomamenities  | ratedescription |
|--------|--|-----------------------------------|-----------|------------|---------------------------|--|-----------------|
| 1440   | Holiday Inn London - Heathrow T5               | Family Room                       | 4         | 4          | Old Bath Road             | Air conditioning;<br>;Alarm clock;<br>;Blackout curt...    |                 |
| 53496  | Mercure London Hyde Park Hotel                 | Family Room                       | 4         | 4          | 8-14 Talbot Square        | Air conditioning;<br>;Desk; ;Free Wi-Fi in all ro...       | si              |
| 60734  | Hallmark Hotel London Chigwell Prince Regent   | Executive Family Room             | 4         | 4          | Manor Road                | Air conditioning;<br>;Carpeting;<br>;Closet;<br>;Clothe... | t               |
| 107246 | Radisson Blu Edwardian Kenilworth - Bloomsbury | Family Room 2 adults + 2 children | 4         | 4          | 97 Great Russell Street   | Air conditioning;<br>;Alarm clock;<br>;Bathrobes;<br>;C... | ε               |
| 55913  | Acorn Lodge Gatwick                            | Family Stay & Park 8 Days         | 4         | 4          | 79 Massetts Road, Horley  | Air conditioning;<br>;Carpeting;<br>;Closet;<br>;Coffee... | ε               |
| 64171  | Gatwick Cambridge Hotel                        | Family Room                       | 4         | 4          | 19 Bonehurst Road, Horley | Air conditioning;<br>;Alarm clock;<br>;Carpeting; ;C...    | ε               |

```
hotel['roomamenities']=hotel['roomamenities'].str.replace(':', ';', '')

def requirementbased(city,number,features):
    hotel['city']=hotel['city'].str.lower()
    hotel['roomamenities']=hotel['roomamenities'].str.lower()
    features=features.lower()
    features_tokens=word_tokenize(features)
    sw = stopwords.words('english')
    lemm = WordNetLemmatizer()
    f1_set = {w for w in features_tokens if not w in sw}
    f_set=set()
    for se in f1_set:
        f_set.add(lemm.lemmatize(se))
    reqbased=hotel[hotel['city']==city.lower()]
    reqbased=reqbased[reqbased['guests_no']==number]
    reqbased=reqbased.set_index(np.arange(reqbased.shape[0]))
    l1 =[];l2 =[];cos=[];
    #print(reqbased['roomamenities'])
    for i in range(reqbased.shape[0]):
        temp_tokens=word_tokenize(reqbased['roomamenities'][i])
        temp1_set={w for w in temp_tokens if not w in sw}
        temp_set=set()
        for se in temp1_set:
            temp_set.add(lemm.lemmatize(se))
        rvector = temp_set.intersection(f_set)
        #print(rvector)
        cos.append(len(rvector))
    reqbased['similarity']=cos
    reqbased=reqbased.sort_values(by='similarity',ascending=False)
    reqbased.drop_duplicates(subset='hotelcode',keep='first',inplace=True)
    return reqbased[['hotelname','roomtype','guests_no','starrating','address','roomamenities','ratedescription','similarity']].head(10)
```

Requirement And special needs based Recommender

```
requirementbased('London',1,'I need a extra toilet and room should be completely air conditioned.I should have a bathrobe.')
```

|     | hotelname                          | roomtype                  | guests_no | starrating | address                  | roomamenities  |
|-----|------------------------------------|---------------------------|-----------|------------|--------------------------|--|
| 310 | The Colonnade London Hotel         | Studio                    | 1         | 4          | 2 Warrington Crescent    | additional conditioning,alarm  |
| 819 | Amba Hotel Charing Cross           | Studio King Suite         | 1         | 4          | The Strand               | air conditioning,clock,bathrobes   |
| 793 | Holiday Inn Express Park Royal     | 2 Single Beds Non-smoking | 1         | 3          | Victoria RoadNorth Acton | air conditioning,clock,carpeting   |
| 816 | Simply Rooms & Suites Hotel        | Deluxe Studio Suite       | 1         | 4          | 21 Avonmore Road         | additional conditioning,clean  |
| 912 | The Cleveland Hotel                | Single Room - Basic       | 1         | 4          | 39-40 Cleveland Square   | conditioning,carpeting,clock   |
| 408 | South Point Suites - London Bridge | Deluxe Studio Suite       | 1         | 4          | Bermondsey Street        | additional bathroom,alarm,air conditioning,carpeting,clock,extra toilet,extra tv |