RESTAURANT RECOMMENDER SYSTEM

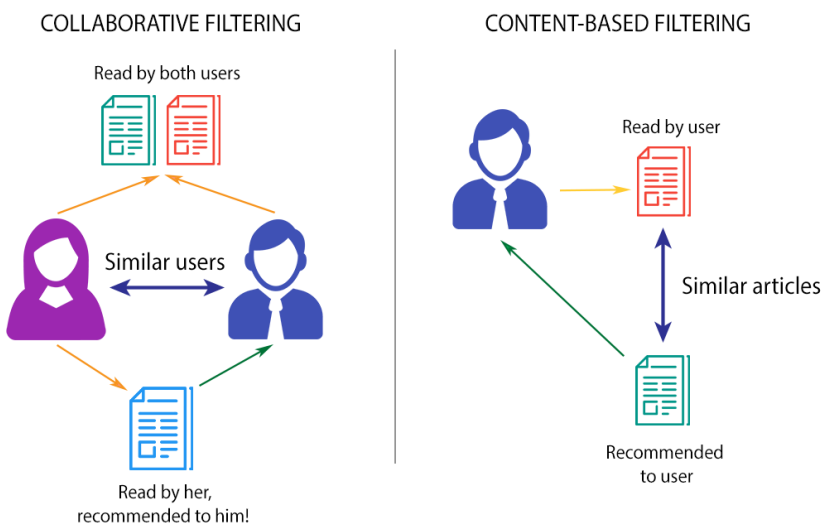
Purpose of the recommender system

The main purpose of this part of is to recommend 2-3 restaurants which are very similar to the restaurant which we choose as a parameter.

RESTAURANT RECOMMENDER SYSTEM

What is the recommendation system we are using and Why?

Our dataset is fit for content based



RESTAURANT RECOMMENDER SYSTEM

Preparing the data for the recommendation system

Out of the 18 features, we only keep 6, as others do not play any role in analyzing the data for the development of a recommendation system.

Trimming the data

We will drop the features which are not required for a recommendation system.

In [35]:
```
drop_cols = ['url','address','phone','book_table','location','reviews_list','listed_in(type)','menu_item','listed_in(ci
dataframe.drop(drop_cols, axis=1, inplace = True)
```

In [36]:
```
dataframe.rename({'approx_cost(for two people)':'cost_for_two'},axis = 1, inplace = True)
dataframe.head()
```

Out[36]:

	name	online_order	rate	votes	dish_liked	cuisines	cost_for_two
0	Jalsa	Yes	4.1/5	775	Pasta, Lunch Buffet, Masala Papad, Paneer Laja...	North Indian, Mughlai, Chinese	800
1	Spice Elephant	Yes	4.1/5	787	Momos, Lunch Buffet, Chocolate Nirvana, Thai G...	Chinese, North Indian, Thai	800
2	San Churro Cafe	Yes	3.8/5	918	Churros, Cannelloni, Minestrone Soup, Hot Choc...	Cafe, Mexican, Italian	800
3	Addhuri Udupi Bhojana	No	3.7/5	88	Masala Dosa	South Indian, North Indian	300
4	Grand Village	No	3.8/5	166	Panipuri, Gol Gappe	North Indian, Rajasthani	600

RESTAURANT RECOMMENDER SYSTEM

Preparing the data for the recommendation system

Identifying and formating the unstructured data

cost_for_two
1,600
1,600
1,600
1,600
1,600

cost_for_two
1600.0
1600.0
1600.0
1600.0
1600.0

rate
4.1/5
4.1/5
3.8/5
3.7/5
3.8/5

rate
4.1
4.1
3.8
3.7
3.8

RESTAURANT RECOMMENDER SYSTEM

Preparing the data for the recommendation system

What about the missing data?

#Quick Question: What should we do with the restaurants for whom the either the rate, vote or the cost is missing?

Let's think in terms of business…

Missing values of the cost column can be replaced by the mean. However, can we do the same with rate and vote? Think, if we replace the missing rate (which could have been 1 or 2) is replaced by average (which is 3.6) and recommend you that restaurant(may be that is your first date), what will you do???

Definitely throw tomatoes at us!!

The unrated restaurants are not taken into account.

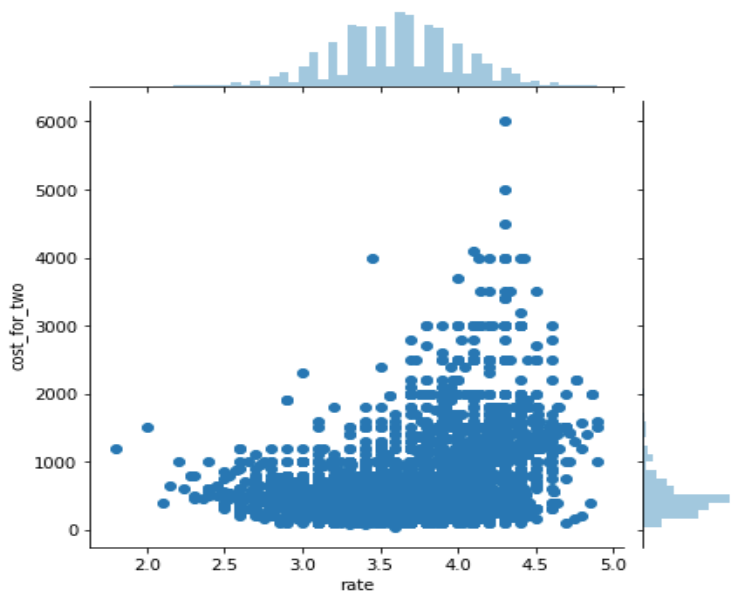
RESTAURANT RECOMMENDER SYSTEM

Information from different features

Does Costlier Means Better?

The jointplot below tells us that the most highly rated restaurants(above 4.5 ratings) are not the costliest ones.

```
In [204]: sns.jointplot(x='rate', y='cost_for_two', data=dataframe)
Out[204]: <seaborn.axisgrid.JointGrid at 0x1b0bdc2048>
```



RESTAURANT RECOMMENDER SYSTEM

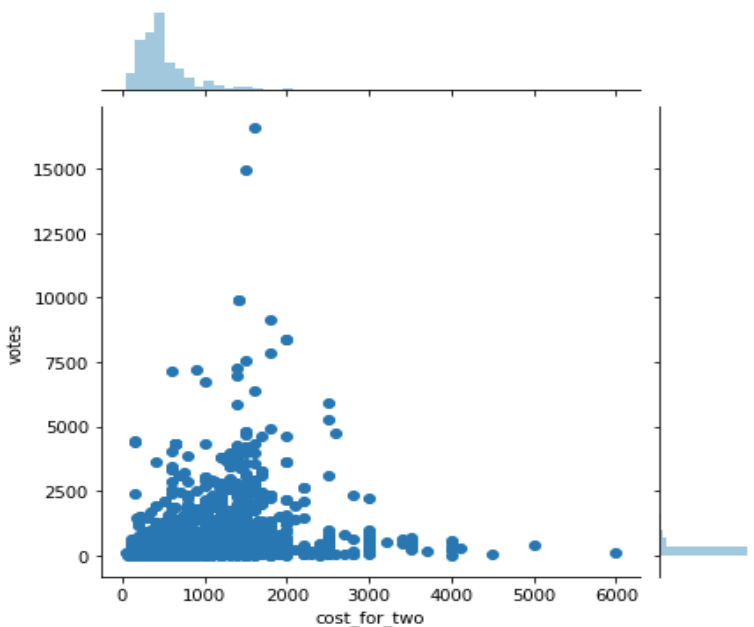
Information from different features

Does cheaper means more popular?

Here is a jointplot comparision which shows the city population is really cost conscious

```
In [205]: sns.jointplot(x='cost_for_two', y='votes', data=dataframe)
```

```
Out[205]: <seaborn.axisgrid.JointGrid at 0x1b0b7824a8>
```



RESTAURANT RECOMMENDER SYSTEM

Developing the recommendation system

STEP-1: TOKENIZE THE CUISINES WITH TF-IDF

#Quick Question: Why TF-IDF, Why not Count-Vectorizer?

RESTAURANT RECOMMENDER SYSTEM

Developing the recommendation system

STEP 2: K-MEANS

The cluster size is 5

#How do you select the cluster size?

RESTAURANT RECOMMENDER SYSTEM

Developing the recommendation system

STEP 2: K-MEANS

The cluster size is 5

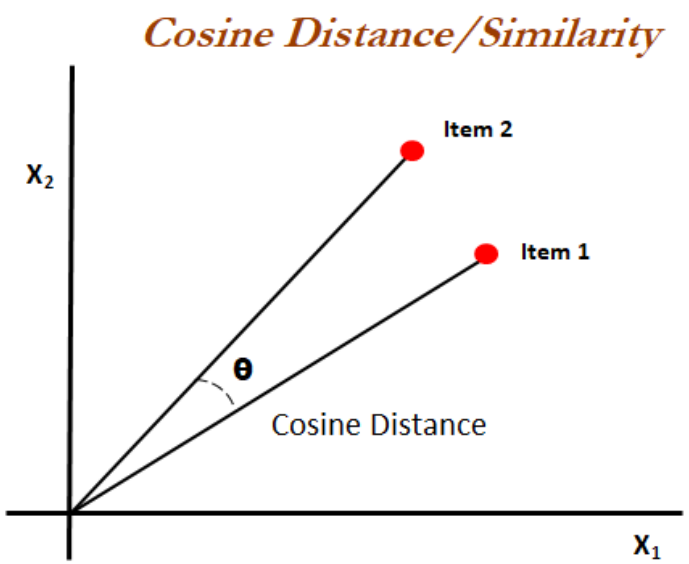
#How do you select the cluster size?

RESTAURANT RECOMMENDER SYSTEM

Developing the recommendation system

STEP 3: Cosine Similarity

What? Why?



RESTAURANT RECOMMENDER SYSTEM

Developing the recommendation system

STEP 3:Results

id							
0	Byg Brewski Brewing Company	1	4.900000	16588.500000	['Continental', ' North Indian', ' Italian', '...	1600.000000	26541600
1	SantÃÂÂÂÂÂÂÂÂÂÂÂÂÂÂÂÂÂÂÂÂÂÂÂÂ...	0	4.900000	246.000000	['Healthy Food', ' Salad', ' Mediterranean']	1000.000000	246000
2	Asia Kitchen By Mainland China	1	4.900000	2223.727273	['Asian', ' Chinese', ' Thai', ' Momos']	1500.000000	3335590
3	Punjab Grill	1	4.866667	1286.666667	['North Indian']	2000.000000	2573333
4	Belgian Waffle Factory	1	4.850000	890.785714	['Desserts']	400.000000	356314
5	Flechazo	0	4.833333	4301.000000	['Asian', ' Mediterranean', ' North Indian', '...	1400.000000	6021400
6	The Pizza Bakery	1	4.800000	1763.333333	['Italian', ' Pizza', ' Beverages']	1200.000000	2116000
7	O.G. Variar & Sons	0	4.800000	1158.500000	['Bakery', ' Desserts']	200.000000	231700
8	AB's - Absolute Barbecues	0	4.790909	4069.250000	['European', ' Mediterranean', ' North Indian'...	1563.636364	6362827
9	Biergarten	0	4.766667	2639.111111	['Continental', ' European', ' BBQ', ' Chinese...	2200.000000	5806044

```
In [231]: rest_recommendations('O.G. Variar & Sons').head(4)

Out[231]: id
          3240      Lassi Darbar
          3239    Karachi Bakery
          3238     Paratha Plaza
          3237           Calvin's
          Name: name, dtype: object
```

RESTAURANT RECOMMENDER SYSTEM

Developing the recommendation system

STEP 3:Results

```
3240      Lassi Darbar
3239    Karachi Bakery
3238     Paratha Plaza
3237           Calvin's
```

Why this order?

5059	Karachi Bakery	1	3.624325	22.384615	['Bakery', ' Desserts']	423.076923	9470	
5050	Calvin's	0	3.628748	240.181818	['Desserts', ' Italian', ' Pizza']	763.636364	183411	

RESTAURANT RECOMMENDER SYSTEM

Other Applications:

A Business Idea: Sponsored Recommendation

Increase visibility and sales on Amazon with advertising