

Restaurant recommendation based on users preferences

Fernando Olvera Pérez | fernandolverap@gmail.com

June 30, 2018

Contents

1	Introduction	3
2	Problem	3
3	Data and treatment	3
4	Similarity	9
5	Assign a score to a new restaurant	9

1 Introduction

This is an approach to use some techniques of data mining in order to recommend a item to user, The first section is about description of available variables and how to treat in order to apply a specific technique.

The second part is about method used and why i chose it.

The third section its about results and conclusions

2 Problem

So I have to find a way to recommend a restaurant to a user in base of user's preferences, To do that I have a some database with information of restaurants and information of users. I will use this information to find the best restaurant to specific user

3 Data and treatment

First I explain data i have,Basically are three types:

- Restaurant features.
- User preferences.
- Score assign to restaurants by users.

In recommendation system there are differing approaches there are:

- User - user- this mean to recommend something on base how much similar is one user to another user , if some user buy something and it likes another user similar to him would like the same things.
- Item-item this is like previous if some user buy something and it likes another thing similar to that item probably likes too.
- User - Item. this is most simple if we have features of items and preferences of users, we can find a metric that told you how similar are.

Our data have similar variables between users and restaurants let's see which variables are similar.

Restaurants	Users
latitud	latiud
longitud	longitud
smoking area	smokoer
dress code	dress preference
alcohol	drink level
Rambiance	ambiance
Rpayment	Upayment
Rcuisine	Ucuisine
price	Budget

Figure 1: There are more restaurants without smoking area, and with can reduce categories to has smoking are or not

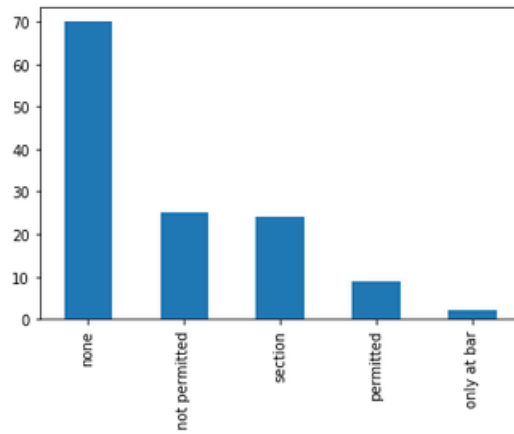


Figure 2: There are more users that are not smoker so this will gives how to replace missing values with more frequent value

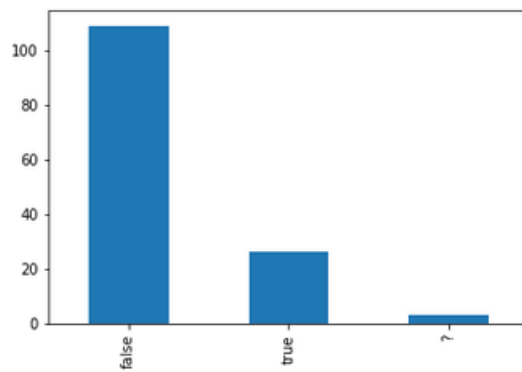


Figure 3: There are more restaurants with formal dress code

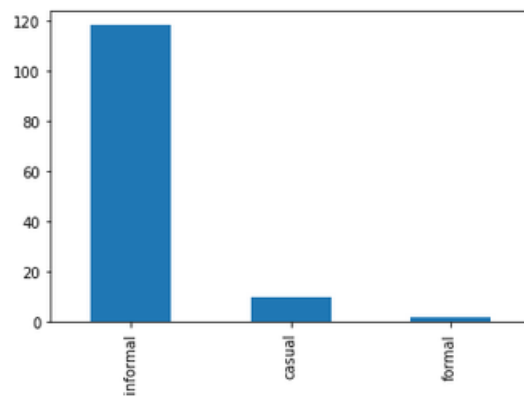


Figure 4: Users with no preference we set to all variables

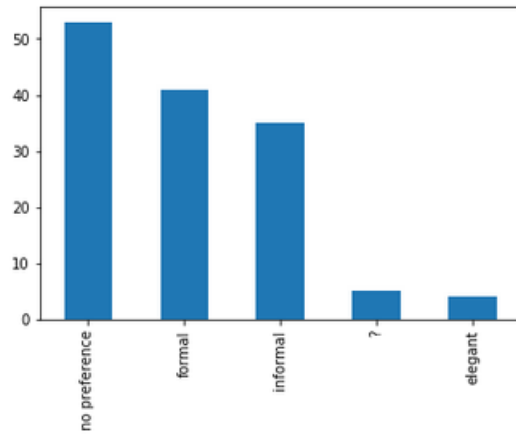


Figure 5: There are more restaurants without alcohol

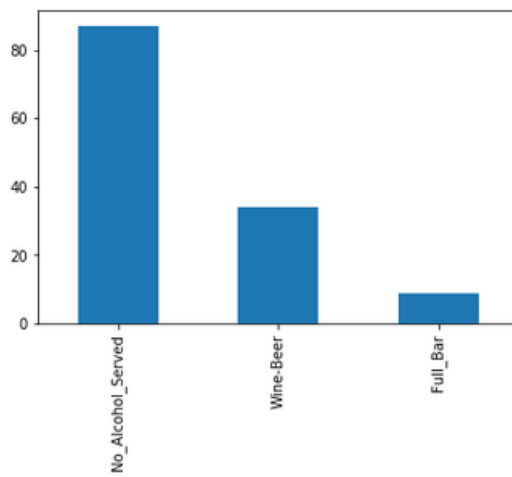


Figure 6: Users preferences are similar but with can reduce to two categories drink or not drink

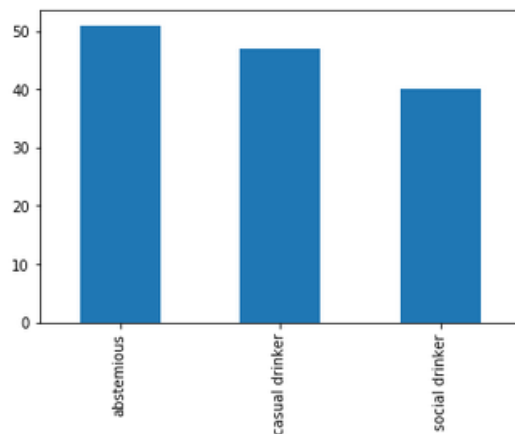


Figure 7: are more family restaurant

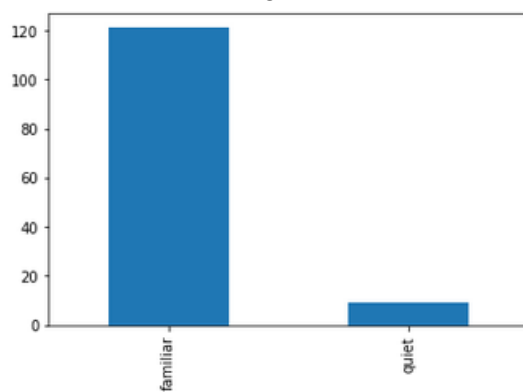


Figure 8: there is one level more than restaurant and is friends. I will do friends similar to family

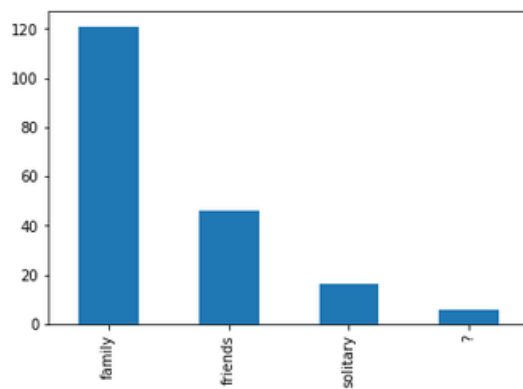


Figure 9: all the restaurants accept cash

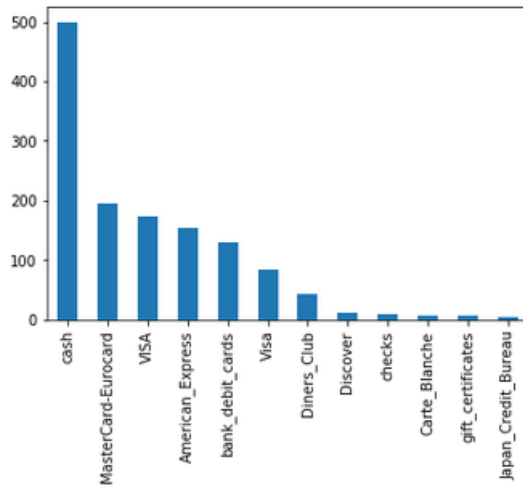


Figure 10: Most of the users prefer cash than other, I will delete pay method that a user has not even used

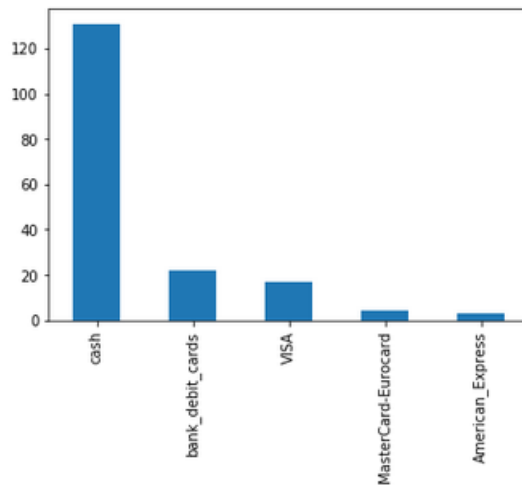


Figure 11: We have more deferents preference of users than restaurants

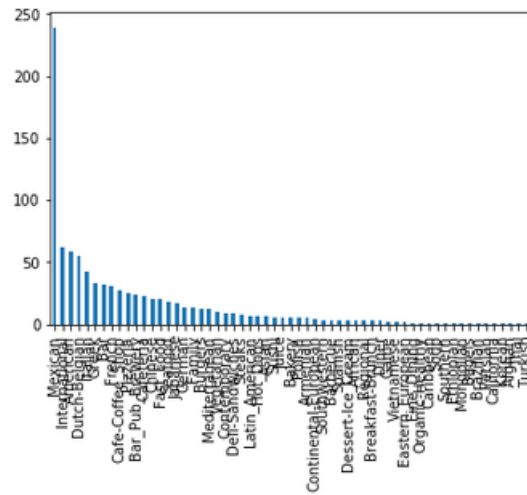
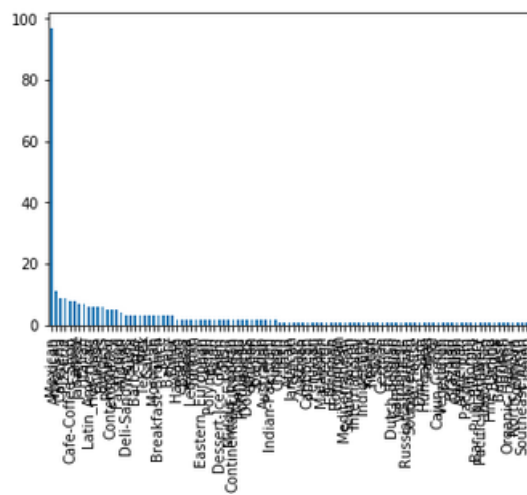


Figure 12: We will delete food that any restaurant have



All of these variables are categorical, except latitude and longitude so let's see how many levels and how they are distributed:

With these variables we can homologue to have an equal number of variables between restaurants and user.

4 Similarity

All the variables that I will use are categorical so we need a similarity metric this different of distance metric because this works with continuous variables, suppose that we have values like red, green and yellow and we assign values 1,2,3 if we will use a distance metric it says us that green is more similar to red than yellow that does not have any mean.

So we are going to use Jaccard metric, It is defined by :

$$J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}$$

In codes that I attached is implementation of all explain here so and results of metrics.

5 Assign a score to a new restaurant

To assign a score a new one approach would be to use a similarity between restaurants, suppose that we have a restaurant and it has a score and come a new restaurant with some features and it's similar to one restaurant that already has a score or difference scores we can assign the score more frequent to this new restaurant.