**To made clusters of business data :-**

**we did preprocesssing of data**:- To fill null values we use mode for that ..and we converted that data into numeric form like by replacing YES by 1 and No by 0...and for columns like Ambience ,smoking,Alcohol we converted that into unicode form.

that's how we got our whole data for applying clustering

Now we have two types of data in our business\_data

**1>categorical data**- cuisines data(Italian Pizza Mexican Tex.Mex Bakery Cafe Deli Seafood American Waffles Belgian Thai Bar Diner Indian Filipino Desserts Korean Chinese Japanese Vietnamese French Mediterranean Greek Malaysian Persian Irish Kosher Lebanese Canadian German British Singaporean Caribbean Moroccan African Egyptian Arabian Afghan Hawaiian Ethiopian Spanish Brazilian Scottish Nepalese Bangladeshi Kebab Taiwanese Polish Indonesian Cambodian Ukranian Salvadoran Argentine Teppanyaki Mongolian Cantonese Dominican Shanghaniese Czech Uzbek Slovakian Iberian)

(it will contain 0 for absence of that data and 1 for presence of these cuisnes in a restaurant)

**2>numeric data**- (review\_count stars BusinessAcceptsCreditCards BusinessAcceptsBitcoin BikeParking Ambience ByAppointmentOnly WiFi RestaurantsDelivery GoodForKids HappyHour smoking Alcohol DogsAllowed HasTV OutdoorSeating WheelchairAccessible RestaurantsTakeOut RestaurantsReservations RestaurantsTableService DriveThru parking pricelevel average\_hours)

In Categorical data we applied KMODES clustering algorithm...so we got 8 clusters from this.

In numeric data we applied CLARA clustering algorithm here we got 15 clusters..

Now we made new clusters by merging both clusters...by combining the centroids from both clustering algorithm and assigning cluster labels so here we got total 45 clusters of 49k+ business data.

For user data we considered YelpUseAttr.csv file which has 9878 users and we selected only those columns from this which are considered for clustering purpose

Now we will assign clusters to user ids from business clusters based on the distance of the user attributes and business cluster centroids attributes…

After assigning clusters we will find the similarity of that user from those restaurants that are present in the cluster which is allocated to that user And based on the similarity we will recommend top 50 restaurant to that user.