

EDA_summary

November 16, 2019

1 Exploratory Data Analysis (EDA)

Our EDA is developed in two parts, business data and review data, since we give recommendations by separately analyzing these two datasets.

1.1 EDA for Business Dataset

We extract business data of all the mexican restaurants which contain 4628 businesses, and transfer the attributes into dataframe. The percentage of missing of each variable is shown below. For the further analysis, we filter out the variables with more than 90% of missing.

For the rest variables, we show their bar plots.

From the above plots, we can see that most businesses are rated as 3-4 stars. Most mexican restaurants offer parking but do not offer romantic/intimate/classy ambience.

Most mexican restaurants offer hipster/divey/touristy/trendy/upscale ambience, but more than half of them offer casual ambience.

We can see that most restaurants are good for lunch and dinner, but not good for dessert/latenight/brunch/breakfast.

Most of the mexican restaurants are good for kids and are available to take out and caters. But most of them do not allow reservation.

Most mexican restaurants are not expensive and offer bike parking and TV, but most of them do not have wifi.

Most of the mexican restaurants have casual ambience, are good for groups and wheelchair accessible, accept credit card, and offer parking.

1.2 EDA for Review Dataset

We extract review data of all the mexican restaurants which contain 403941 reviews, and then tokenize the texts. We give some plots to show the information of our review data.

The distribution of stars in mexican restaurants is shown below, we can see that most ratings are five stars.

This plot shows review counts of 5 mexican businesses which have the most reviews. They all have more than 2000 reviews.

This plot shows review counts of 9 cities which have the most reviews. They all have more than 2000 reviews.

By the wordcloud, we can see that the most frequent words is chips, salsa, carne asada, etc.

Some examples which shows the relationship between stars and words in text are shown below.