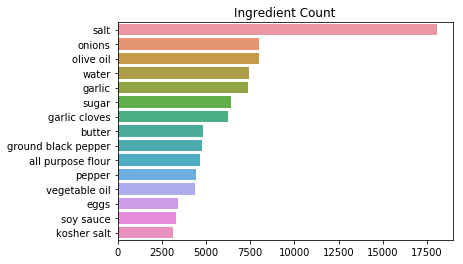


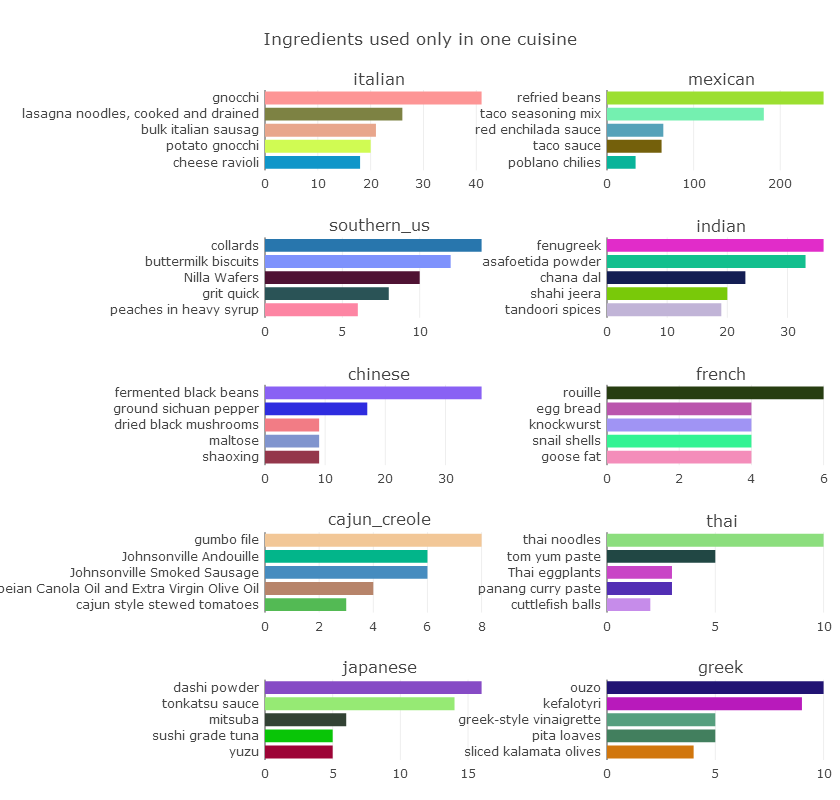
1. Salt, water, onions, garlic are such common ingredients that we expect them to have poor predictive power in recognizing the type of cuisine.



2. Explore which ingredients occur only in one cuisine (ingredients specific to the cuisine).

 From the visualizations above we can make several observations:

* the Mexican cuisine is the only one where we find that the unique ingredients are also found in many recipes part of the cuisine in question. The refried beans are found more than 200 times in Mexican recipes (250 out of 6438 recipes ~ 4% of all Mexican recipes). This finding means that the 'refried beans' and 'taco' (or derivatives of those) ingredients are expected to be strong predictors of Mexican cuisine. On the next step of the analysis we will try to support this observation by using Tf-Idf representation of our training sample in an attempt to find the most important ingredients in a cuisine.
* the above-mentioned observation is also valid for the:
  + the Italian gnocchi;
  + Brazilian cuisine with its traditional cachaca which is the most popular spirit among distilled alcoholic beverage in Brazil
  + Indian fenugreek - in our sample it is found only in the Indian cuisine



* 3. Partition the data into training set (each cuisine with 90% data) and separate test set (each cuisine with 10% data), train the model with Logistic Regression again, and verify the accuracy with the test set.