INTRODUCTION

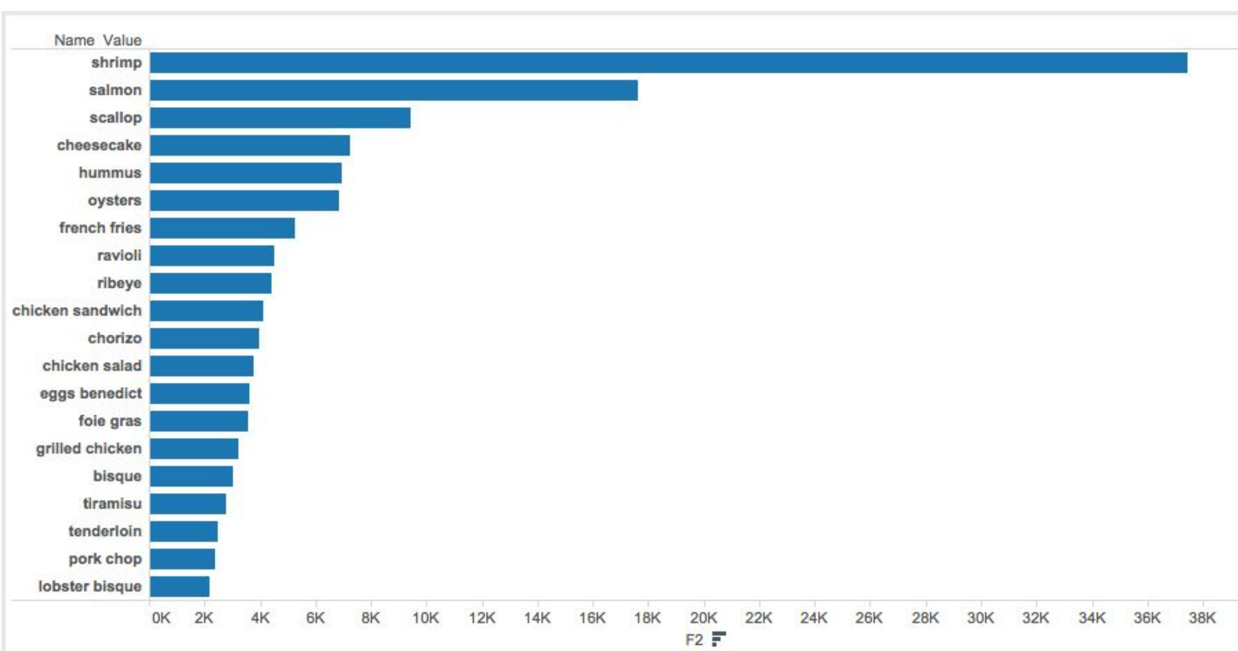
All code mentioned in this report could be found in my github repositiry, and it licensed under MIT, so feel free to use it. For both tasks I used Mediterranean cuisine, because i live in Spain, and familiar with topic.

TASK 4

As entry data, I used list of the dishes from the task 3 (it can be found in the repository linked above). First I wrote a node.js script dishScoreCalculator.js which looked at mongodb with a Yelp data set which I create while working on task 1, and generate next JSON for each dish:

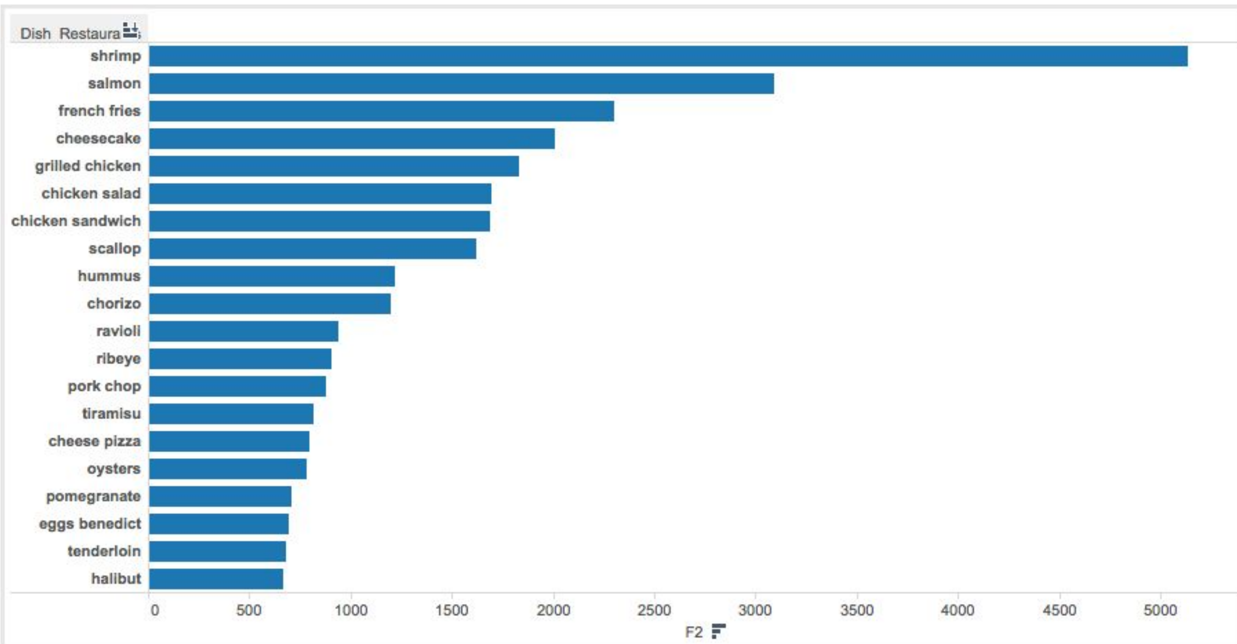
```
{
  name : dish name
  reviews :  number of reviews that mention this dish
  score : average sentiment score of this dish review
  restaurants : list of all restaurants with their sentiment scores
}
```

Using this data I moved to Tableu and made first visualisation of dish popularity based on number of reviews.



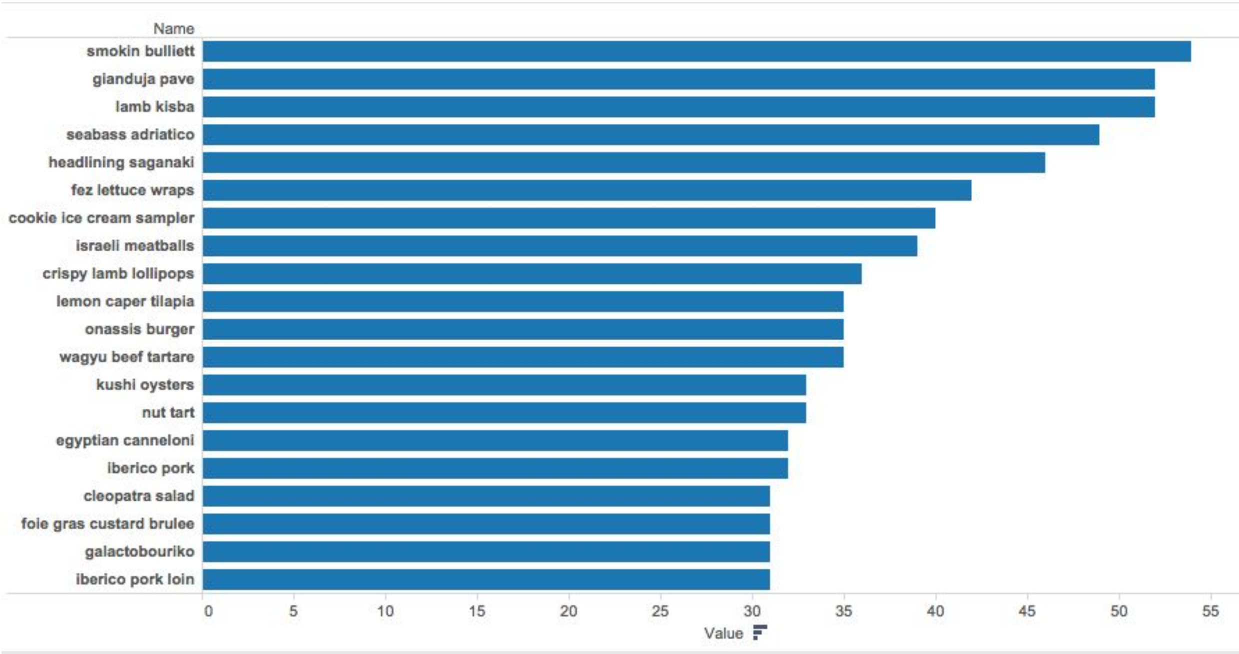
This chart simply shows that people talk a lot about shrimps in their reviews of restaurants with Mediterranean cuisine, but it is not really representative, as we don't know if this is a just a few places famous because of their shrimps, or is this something common to the whole cuisine.

Next attempt was made by incorporating the count of restaurants into the final score of the dish. It was simply made by counting in reviews of how many restaurants this dish is mentioned. And in my opinion this score is more representative than previous one, as dish could be a true representation of a cuisine only if it represented in wide range of the restaurants specialised in this cuisine.



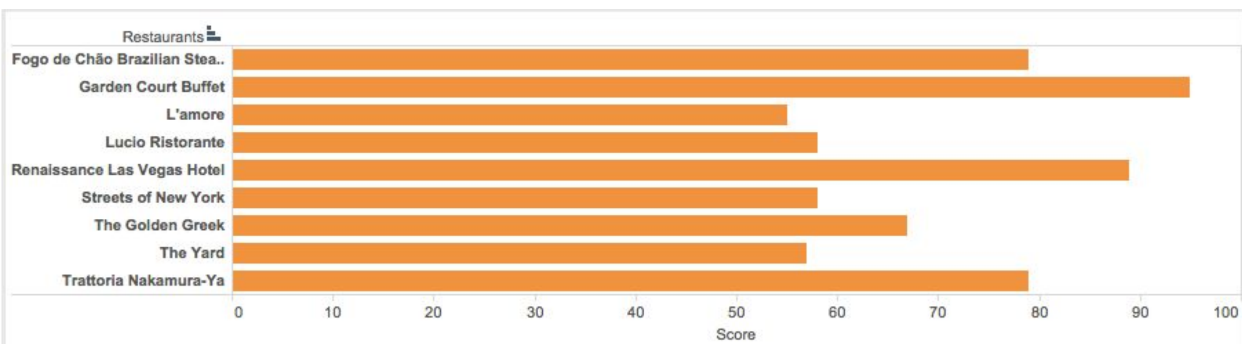
As we can see, not much had changed in the top part of the chart. Sea food is still going strong there. So in the next and last iteration, I calculated score based on the sentiment analysis of the review. I used AFINN-based sentiment analysis package from npm to calculate the sentiment score of the review. The final score of a dish is an average sentiment score across all reviews where this dish were mentioned. With high score corresponding to more positive feedback about this dish in the review.

As you can see below, now we can see completely different set of top dishes, which should better represent the best and the most tasteful and technical dishes of a cuisine, and skip noisy reviews like: "shrimps were not fresh"



TASK 5

All data for this task were prepared in the task 4. As were described above I did prepared JSON file which have list of the dishes with list of the restaurants where you can find this dish and their scores. Restaurant score based on the sentiment score of reviews for chosen dish.



Visualisation above shows top restaurants for tasting “Kebabs”

ONLINE RECOMMENDER ADAPTATION

Code used to generate data for these tasks could be easily adapted for the online recommender. If briefly, a simple single page web application where user first select list of a cuisines, which next asks him to select a specific dish from a cuisine. This last action triggers a call to a server with a name of the dishes he just selected. On a server we do calculations based on the code from the task 5, and return to a user a page with this list. Extra points if prepare this list based on the geographical proximity restaurants to the user in question.