

Viraj Bhalala
Task 4-5
07/07/2019
Data Mining Capstone

Task 4 and 5

4.1 Overview

In Task 4, our goal is to leverage dish names to further help people making dining decisions. People often face challenge when they have to make decision on trying new dish from cuisine they are not so familiar with as well as trying restaurants that they have never visited before. In task 4, we will mine popular dishes in cuisine that are liked by people using Yelp reviews. We will also create visualization plots that can help people making decision.

4.2 Data Preprocessing

We first parsed in yelp_academic_dataset_business.json file that contains information about businesses. In our case, we are only interested in restaurant business. Additional in order to keep scope of this task narrow, we only selected Indian restaurant types. This is because I am very familiar with Indian cuisine can easily validate if results are accurate from data mining. Next, using this filtered business ids of Indian restaurant, we obtained all of their reviews from parsing yelp_academic_dataset_review.json file. Lastly, we filtered reviews of all selected business ids that contains tokens of Indian cuisine that we used in task 3 and that are relevant with having 1s in second column.

4.3 Popular Dishes

We created first plot using the above described preprocess data, we grouped reviews by dishes and aggregated total counts as well as breakdown counts by their stars. This plot helps us to find most popular dishes.

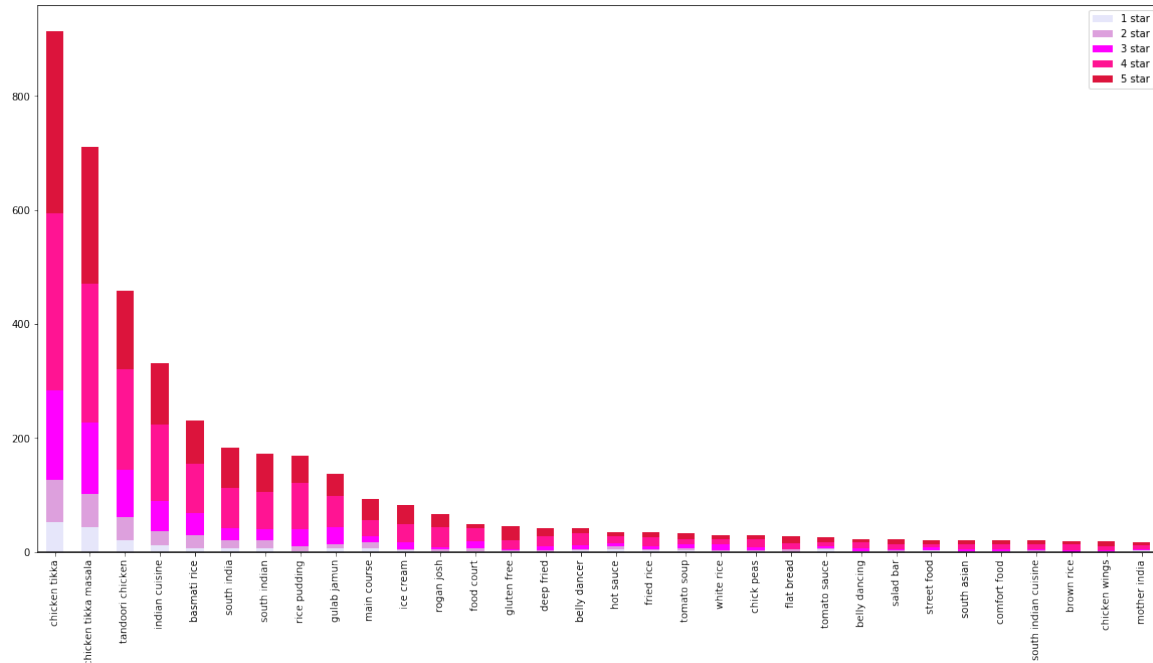


Figure 1: Popular Indian dishes by total counts

Secondly, we created another plot that shows popular dishes by percentage of reviews stars by total review. This metric helps us to see which dishes have good ratings out of all reviews. For example, looking at 4- and 5-stars proportions.

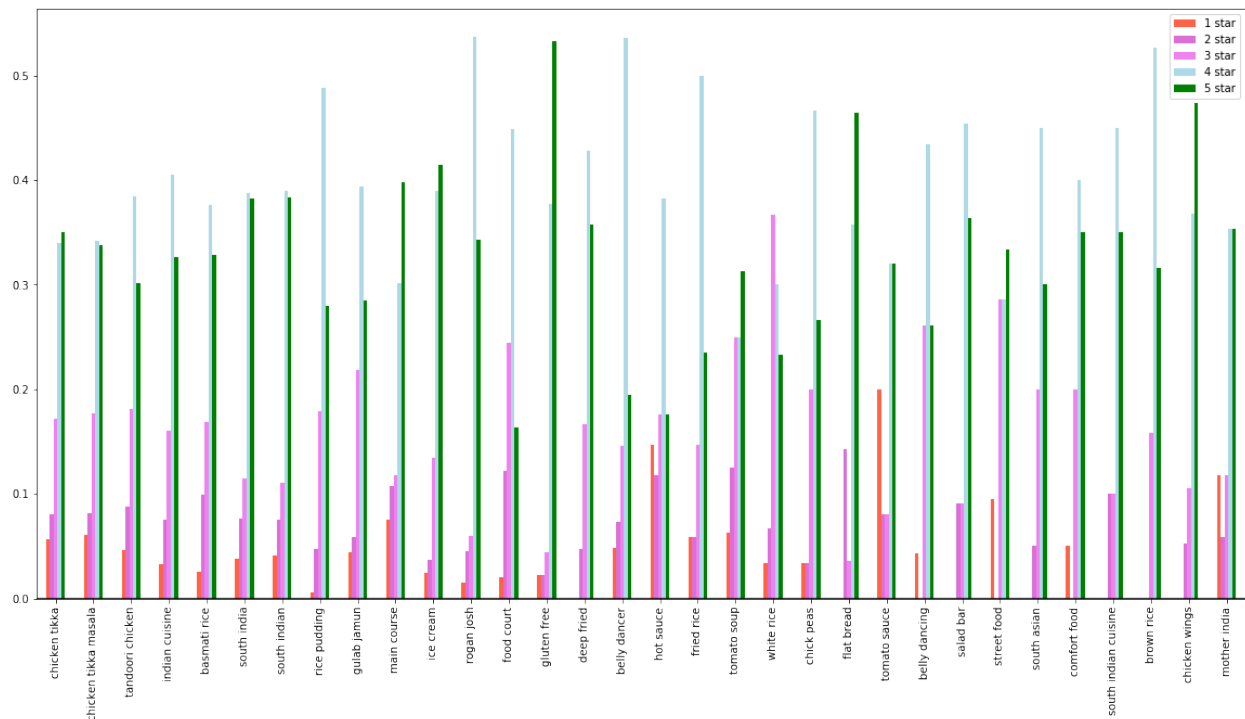


Figure 2: Distribution of rating stars for each popular Indian dish.

4.4 Conclusions

By analyzing both of the above plots, we think that “Chicken Tikka” dish is top recommendation or popular that users should try. “Chicken Tikka” has the greatest number of reviews out of all Indian dishes in Figure 1. We think that user should not rely much on Figure 2 because it can be misleading. For example, Figure 2 can show top ratings to really high proportion if there are really less ratings and if most of those ratings have high starts. If we use figure 1 and figure 2 together then it can strongly help us to make decision. For example, we can see that “Chicken Tikka” dish is most popular dish and it also has decent proportion of 4 and 5 star ratings.

5.1 Overview

Similar to task 4, we will mine popular restaurants for a selected dish. Based on discussion on section 4.4, we have selected “Chicken Tikka” dish to be most popular. In this section, our goal is to rank the most popular restaurants that serves our selected dish

5.2 Popular Restaurants

In the first attempt, we have used all reviews has “Chicken Tikka” in it and have used those reviews to group by restaurants and used total counts. We have also broke down counts by restaurant rating starts and created bar plot of it

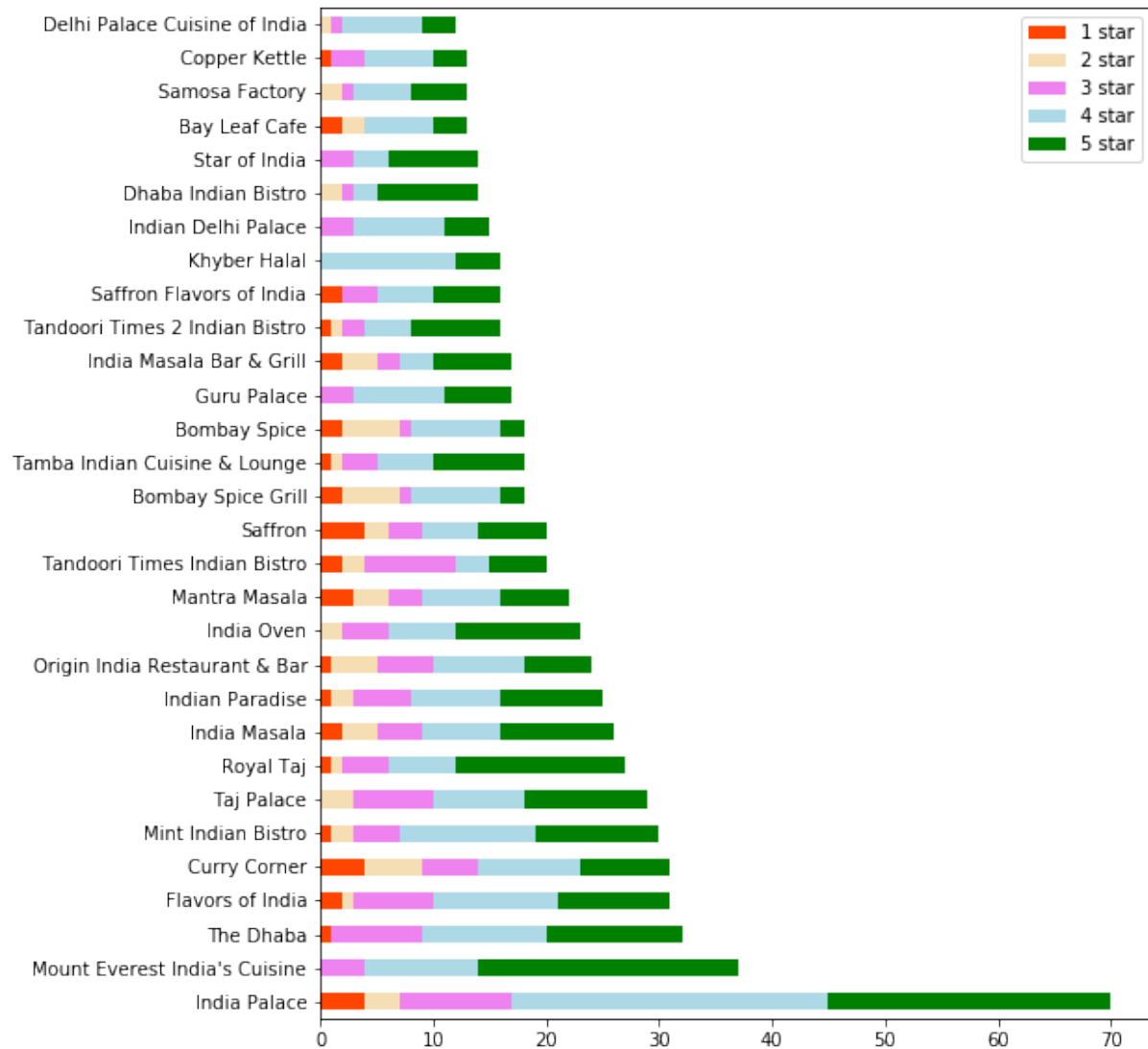


Figure 3: Popular restaurant by total review counts

In the next attempt, we have used results from previous attempt to create another plot. In this plot, we have took total count of reviews for top restaurant that has our selected cuisine and averaged by total ratings of each restaurant. This plot can help us to analyze which restaurant has most proportion of good and bad reviews.

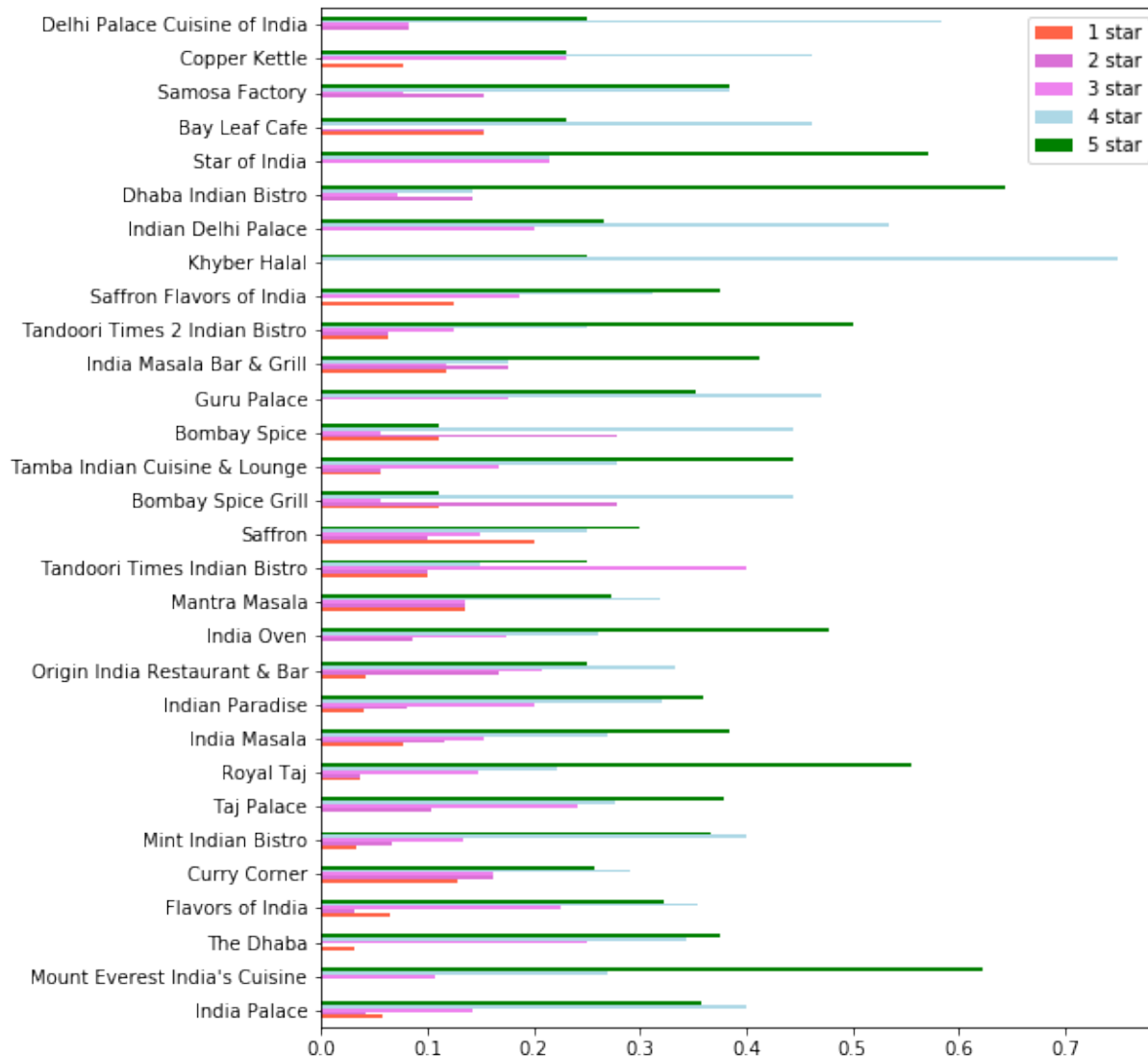


Figure 4: Distribution of rating stars for each popular restaurant that has “Chicken Tikka” in their reviews.

5.3 Conclusion

Using Figure 3 and 4, we came to conclusion that “Mount Everest India’s Cuisine” seems to be the best restaurant to have “Chicken Tikka” dish at. We used both plots to make this decision similarly how we did in task 4. “Mount Everest India’s Cuisine” is the second most popular restaurant by total reviews in Figure 3, and it has one of the top proportions of 5 stars in Figure 2. From this analysis we can effectively, recommend anyone to try “Mount Everest India’s Cuisine” to try “Chikken Tikka” disk.