

Xtern Application 2020
FoodieX Dataset Findings

Finding 1: The food types. There is a good variety of food delivered by the listed restaurant. Further analysis is required to confirm whether food type depends on restaurant location / proximity to one another.

Finding 2: Top rated restaurants. The ratings ranged from 2.5 up to 4.8, with some restaurants not having a rating.

Top Rated Restaurant IDs and Ratings:

ID_4728 with a 4.8 rating
ID_7412 with a 4.8 rating
ID_1166 with a 4.7 rating
ID_2201 with a 4.7 rating
ID_7924 with a 4.7 rating
ID_2051 with a 4.7 rating
ID_1064 with a 4.7 rating
ID_1166 with a 4.7 rating
ID_383 with a 4.7 rating
ID_6278 with a 4.7 rating
ID_6537 with a 4.7 rating
ID_1160 with a 4.7 rating
ID_7433 with a 4.6 rating
ID_7739 with a 4.6 rating
ID_1666 with a 4.6 rating
ID_1350 with a 4.6 rating

Finding 3: Top voted restaurants. The votes go from 0 to 9054.

Top Voted Restaurant IDs and Vote Totals:

ID_1064 with 9054 customer votes
ID_1666 with 4903 customer votes
ID_2885 with 4691 customer votes
ID_2601 with 4606 customer votes
ID_6511 with 4438 customer votes
ID_4202 with 4335 customer votes
ID_4202 with 4335 customer votes
ID_2051 with 3975 customer votes
ID_13 with 3860 customer votes
ID_8087 with 3782 customer votes
ID_4606 with 3535 customer votes
ID_1947 with 3394 customer votes
ID_2041 with 3248 customer votes
ID_7753 with 3124 customer votes
ID_847 with 3021 customer votes
ID_6915 with 2858 customer votes

Finding 4: The geolocations of the restaurant geolocations seems to be random, but they can be clustered into regions based on their proximity and density.



Future Suggestions:

1. Gather data and determine the possible relationship between food frequency and their geolocations.
2. Determine relationship between rating, customer votes, and reviews.
3. Use excel or prewritten libraries (or whatever is most convenient) to cluster restaurants into regions.