

Executive Summary

When we travel, how would we choose hotels? Location and convenience come to the first place. The hotels in Times Square are usually the first option when tourists visit New York City and would like to experience the New Yorker life at Midtown, such as watching Broadway show, shopping, drinking at a bar, enjoying a meal at a steakhouse, and etc. There is nowhere else in the world like New York City presents all kinds of best gourmets from everywhere of the world. If we search the vocabulary “restaurants” in Yelp, there are Italian, American, Mexican, Portuguese, Asian fusion... When we choose a hotel, if we take the foods surrounding the hotel into considerations, what kind of hotel we should choose? Also, gourmets play significant role during travelling. The food quality, restaurant environment, and price are the essential elements for us to judge if a restaurant is “good” or “bad”. If the rating of the hotel is high, are the ratings of the neighborhood restaurants also high? Does hotel affect the development of restaurants?

We would answer the above inquires by analyzing datasets provided by NYC OpenData and Yelp. There are 332 restaurants and 41 hotels chosen located in Times Square. The first graph (“Histogram of Rating on Different District”) gives us a general view of the restaurant data we collected. According to Yelp rating system, there are 5 levels—1 is the lowest and 5 is the highest. Most of the restaurants are rated from level 2 to 4; there are only a few are rated less than 2 or above 4.5. The shape is left skewed, so we can say the majority of the restaurants provide high quality foods. The 10036 zip code area owe nearly half of the restaurants, and 10018 zip code area have the second largest restaurant number. One reason is 10036 area itself occupies half of the Times Square area; the other one would be those area are usually popular tourists attractions, such as Madame Tussauds and theaters.

Here comes another question. What kind of food dominates Times Square food industry? If we take a look at the graph “Mosaic of Category vs. Zip Code”, it demonstrates how the food categories scatters in Times Square area. North American restaurants have the largest portion in the food industry in Times Square, and Delis have the second largest, while European and Asian become the third and the fourth largest category. Majority of the restaurants are located within 10019 and 10036 zip code area, while there are more North American style restaurants in 10036 than 10019 zip code area. Inside 10036 area, Asian restaurant and Deli have smaller portion than 10019 zip code area.

Alluvial diagram, or called Parallel coordinate plot, presents us the overall information we would be interested in and how the data we collected after we combine food categories, ratings, price levels, and locations together. If we are not sure what to eat tonight, we could look at the Alluvial diagram and make a decision. Most of the restaurants in Times Square are North American, and the majority of them are rated at level 4 but with low cost (2 or 3 dollar-sign according to Yelp), located in zip code area 10036. So, we might go to a North American restaurant with ratings at level 4 and price level at 2 and located at 10036 areas after we finish watching a show in the Theater District. If we would like to pick a high-quality steakhouse, we could just walk into any steakhouse we come across in Times Square, but before we do that, we

might want to yelp it first; although 90% of them are rated very high, only few are served with low quality. Likewise, if we want something quick and not too expensive, we can stop by a Deli, because as the top 3 largest categories in Times Square, Delis offer low-cost meals, and spread over all the area, including all zip codes.

After we watch the Broadway show, fantastic North American dinner, it is time to take a rest. Are there any high rating hotels near the fancy restaurant we just have our dinner? Let's take a look at the histogram of "Price distribution at each hotel price level and neighborhood restaurants". As for each of hotel, we find out its surrounded restaurants within a circle area with 100 meters as radius; and then we make a comparison to see how the restaurants spread around each hotel through grouping by hotel price levels. Generally speaking, the shape of how expensive restaurants distributed are nearly identical. No matter how the price level of the hotels is, restaurants with price level 2 are nearly half of the neighborhood restaurants; restaurants with price level 1 are the second largest for each hotel cluster. At each price level, there are only few restaurants are 4 dollar-sign. Overall, the price of hotels is not too expensive in Times Square.

Let's talk about the ratings of hotels and restaurants ("Rating distribution at each hotel rating level and neighborhood restaurants"). Likewise, the restaurant rating distribution has almost the same shape at each hotel rating level. Obviously, most of hotels and restaurants are rated as level 3, and more than half of the restaurants at level 3 hotels are also rated as level 3.

To conclude, North American and European restaurants prefer to locate in 10036 zip code area. Most restaurants and hotels provide high quality service with 3 to 4 stars, but of course, there are some low rating hotels and restaurants rated only 1 or 2 stars, and we might want to avoid those businesses when choosing a place to eat or accommodate. Speaking of the correlation between hotels and their neighborhood restaurants, there is no strong evidence showing the relationship exists. The ratings and price levels of restaurants and hotels are evenly spreading. What we have done so far is analyzing hotels as a group, such as in the group of price level at 1, 2, 3, which might average out some special and individual cases.