1) Problem Background

The international hotel chain recently opened a steakhouse in one of its locations. Because this new establishment was very successful, the owners would be interested to open another steakhouse in one of its hotels, just in another city. In order to maximize the chances of success, they would like to find a location which would be similar to the current one. Our task is to analyze the list of candidate hotels for the new location and make preliminary recommendations.

2) Data description, collection and its application for the problem solution.

Because clientele of the steakhouse consists of the guests who are staying hotel, as well the residents of the city (and possibly tourists residing in other hotels), the similarity of the new location to the current one should somehow address both of these factors:

2.1) We assume, that guests of the hotel might have similar attitude and level of satisfaction when evaluating both hotel and eating establishments. Thus, to evaluate similarity between this part of customer base, we will use publicly available data presented in www.bookings.com. More precisely, for each hotel we will use average user scores based on how the guests evaluated the following characteristics of the hotel:

> 1) staff; 2) facilities; 3) cleanliness; 4) comfort; 5) value for money; 6) location.

These scores will be stored in 6-tuple of numerical values (in range between 0-10), which will serve as our reviews profile of the hotel.

2.2) To address the similarity between other part of clientele base, we will analyze neighborhood of around the hotels, based on the data provided by Foursquare API. Our assumption is that types of locations residing around the hotel forms or at least reveals a distinct profile, which is connected to foot traffic of possible clients. To form such neighborhood profile, we will extract the categories of the venues located around the hotel. We'll group all the categories provided by Foursquare API into six types, which, in our opinion, might be relevant for our task:

1) Various shops and stores; 2) Tourists attractions and other walking destinations, such as parks, museums, monuments etc.; 3) Restaurants (other than steakhouses); 4) Steakhouses; 5) Other food and beverage serving establishments, such as pubs, cafes, pizza places, burger joints etc.; 6) Other locations.

The extracted number of each type of location will when be divided by the overall number of found locations and stored into another 6-tuple, which will serve us as profile of the neighborhood for a particular hotel.

2.3) The collected review and neighborhood profiles of "candidate" hotels will be compared to the respective profiles of the current location. To do so, we will apply Euclidean distance between vectors, which will serve as a measure of dissimilarity between the hotels. To account for both review and neighborhood profiles, we will use combined measures, based on the sums of normalized sums and ranks of the distances. The hotels with the lowest values of these joint measures should be most similar to the current location.

In addition, we will perform data clustering based on the distances between reviews and neighborhood profiles. This additional analysis will help to better determine the list of best locations, which are similar to each other, so that the hotel owners would have a list of alternative solutions for their development plans.