Hong Kong Baptist University 2021/2022

JOUR7280/COMM7780

Big Data Analytics for Media and Communication

**Final Report**

**Instructor: Dr. Xiaoyi Fu**

**Group Members:**

MENG Guosheng 21431280

REN Tongyang 21407748

SHI Minting 21422443

**Abstract**

City homogenization has seriously affected the development of the tourism industry. For people who desire to explore local experiences, facing almost the same scenery will greatly reduce their travel intentions. Food, often regarded as an important reason why people visit a city, is also showing a convergent trend. However, individuals are supposed to enjoy more authentic local food that is different from their daily meals when discovering a new place. Besides, factors such as food genres, food quality, price, rating score, environment all affect people’s decision when picking up a canteen. *Zomato*, as one of the leading online food recommendation platforms in the globe, provides customers with rich restaurant information and objective user evaluation. By scraping and analyzing the data of several global representative restaurants on *Zomato*, this report strives to offer tourists pertinent suggestions when facing a large number of restaurants with no obvious distinctions.

*Keywords:* city homogenization, food, *Zomato,* tourists

**Introduction**

With the rapid advancement of worldwide urbanization and modernization, the phenomena of city homogenization has become more serious. People are more likely to have the feeling that each city appears to have already met previously. The characteristics of each city become less distinctive, resulting in an increasingly blurred impression generated by people. The urban landscape, the functions of public places, and even people's consumption and cultural habits have become standardized, showing a higher trend of convergence.

City urbanization could accelerate urban development, and provide standard integrated services for residents and tourists. However, it has created obstacles in the way of helping bolster the tourism industry. It is not hard to imagine that people will gradually lose the desire to travel when all cities in the world have the same architectural style, food, and even cultural customs. For instance, food, as a key factor in encouraging individuals to travel, could be regarded as one of the indicators of whether a city is worth exploring repeatedly. Thanks to the growth of globalization, people can enjoy tasty food in a city, even if the city is not the origin of the cuisine. However, it is precisely because of globalization and city homogenization, menus in all parts of the world have become similar. In order to meet the tastes of different immigrants, the types of dishes are increasing, while the food with real local flavors gradually disappears. For tourists who come to the city for the first time, whether to provide local foods is a factor in choosing a restaurant. In an industry study conducted by National Restaurant Association (2011), 69% of participants stated that they would like to prefer restaurants that offer locally produced foods. In addition to the genre of the food, other factors driving customers into a canteen like food quality, price, and “soft conditions” also play significant roles. Auty (1992) argued that in the market where food types and quality are almost similar, “soft conditions” including environment and atmosphere are determinant.

It was built on these ever-changing factors that on-line platforms offering food delivery, dining and restaurant discovery services have been emerging. *Zomato* is one of the major market players of these on-line platforms. The platform was launched in 2008 and provides comprehensive restaurant information, menus, as well as user reviews. The Company considers judgment as its core value that ‘it is not the skills that show who we [customers] are, but the choices.’(Panigrahi et al., 2020).

In this regard, by analyzing the data including cuisine genre, rating score, number of shops of several representative cities in the world on *Zomato*, this report aims to conclude the characteristics of the food industry of each city, such as high cost performance, rich local food offering, so as to provide guide to tourists when they are considering their destination among the burry distinction between the cities.

In fact, there has been extensive related works discussing global food trends and customers’ eating habits patterns, providing reference for our research and analysis. In an international report issued by YouGov (Matthew, 2019), by analyzing more than 25,000 participants in 24 countries, it was concluded that Italian cuisine is the most popular food among the world. In the way of asking participants to evaluate 34 national cuisines from “like a lot” to “I have never eaten this cuisine”, Italian cuisine received an average popularity score of 84%, followed by Chinese and Japanese cuisine, with an average score of 78% and 71% respectively. In addition to the global cuisine survey, there are also scholars exploring people’s dining behavior in terms of online restaurant reviews. For instance, in the study of Vu et al. (2017), it was concluded that certain groups, such as Italians, Indians, and French, have a strong predilection for their own national food. Moreover, it was also inferred that cultural background influences the preference of restaurant types (e.g., pubs are more popular among westerners).

**2. Methodology**

To have a thorough understanding of the people’s consumption behavior and insights in terms of eating in major cities of the world, the information in dining-out, nightlife and collection pages of Zomato website for different cities were scraped with Python (mostly with *selenium* library) and related data were extracted, visualized and studied on thereafter.

Dining-out pages in Zomato mainly include the canteens available in a specific country when a user is browsing through the website. The upper limit for the number of canteens on the page is 500. Nightlife refers to the places that run business mostly at night, a majority of which are clubs and inns in the city. Lastly, the collection indicates the thematic activities in the city, which could vary as people have different lifestyles, traditions and religions in different cities of the world.

To have a better comparison, the sample cities are located in North America, Europe, Asia and the Middle East. To ensure the cities to be studied have a different cultural contexts, for each page, the data for three cities from three continents were scraped.

Specifically, for the dining-out part, Sealltte from North America, Istanbul located in the center of the Eurasia Continent and Singapore, the Southeast Asian state-country were sampled. For the nightlife part, the three cities to be studied include Seallte, Rome from Italy and Mumbai, the *Apple City* in the eyes of Indian people. Initially, Istanbul, Seattle, and Singapore were chosen as the three major cities to compare the number of bars in their nightlife to respond to the ‘collection’ of three major cities. But unexpectedly, Singapore didn’t show any nightlife on the website of *Zomato*, maybe with some restrictions due to the current pandemic situation. For Istanbul, in the beginning, we did not take into account the influence of religious factors on the city, but after we scrape Istanbul's webpage, we found that there are only nine-night bars, which is not enough to support our comparison of nightlife bars in the three cities, so we finally gave up choosing Istanbul as the nightlife city. And lastly, for the collection part, since Singapore does not have many thematic activities on *Zomato*, Seattle, Istanbul and Greater Lisbon from Portugal were taken for comparison (Barcelona or other cities in Spain would have been a better unit of analysis if it were available on *Zomato*).

The scraped data for the three groups were cleansed, sorted and visualized mostly in Python. For better illustration, tableau was used to locate the canteens in different cities (the map generated thereby was included in the PPT). Based on the results, analysis on the relationship of food genre and rating, the overall number of clubs and thematic activities were carried out, which aims to reflect the lifestyle difference among people from different parts of the world by inductive reasoning. The results could pave the way for further study in the future as the sample size for this research was not big enough to cover every major city in the world.

**3. Storytelling and Outputs**

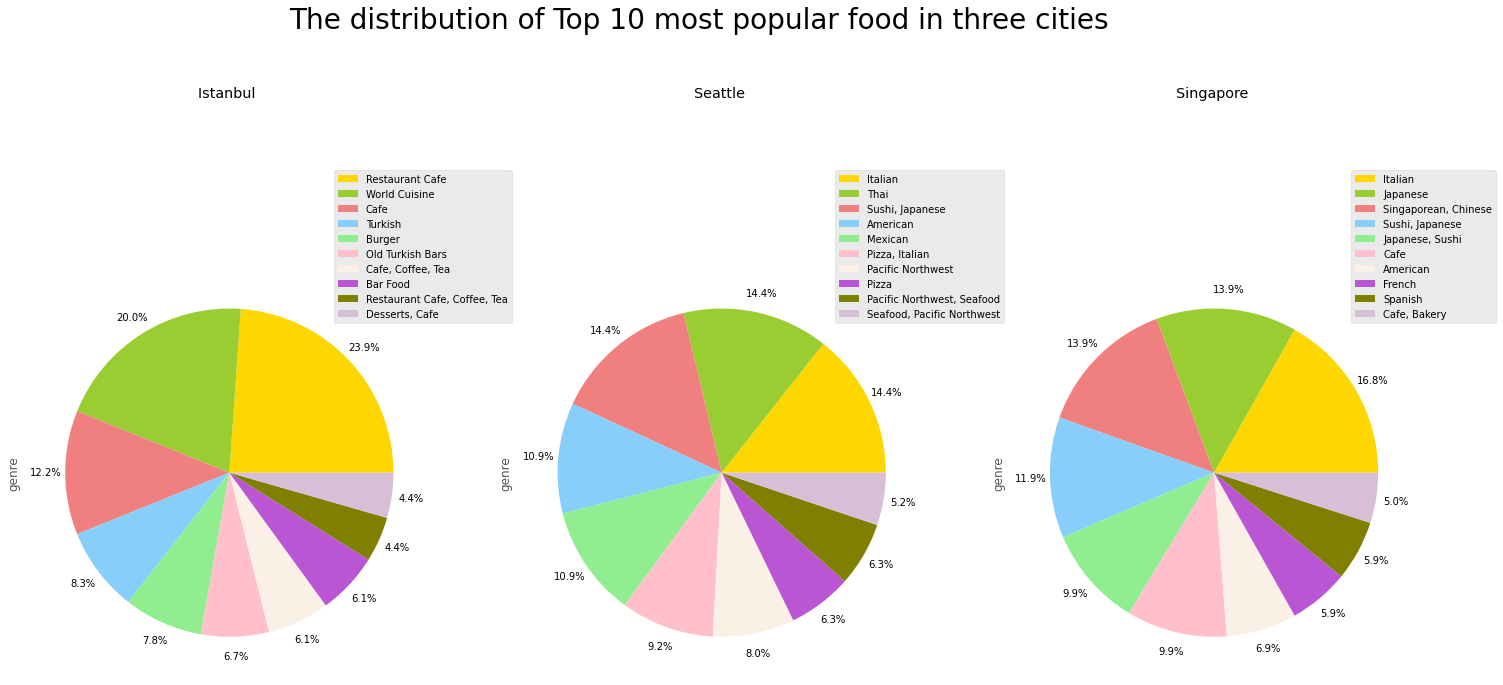
***Dining out***

As a newcomer to a city, the first thing you want to do may be to taste the yummy local food. A question on what is the most popular cuisine in the city may pop up into your mind. The question could well serve as the first research question on the scraped data in dining out pages. Calculating the average mean of the canteen rating by genre, we attribute the first ten genres with the highest ratings as the top 10 highest-rating foods in the city.

From the bar chart of *Top 10 most highest rating food in three cities*, it is concluded that, in Istanbul, Cafe, Desserts, Beverages, Turkish Pizza, Kebab, Steak, Izgara, Doner, Desserts, Turkish, Grill, Beverages, and Patisserie, Cafe, Turkish are most welcomed food, with a high rating score of 4.9. Giblets, Turkish, Balknas, Old Turkish Bars, and Cafe, Desserts, Coffee, Tea are the second in popularity, with an average rating score of 4.8. In Seattle, the type of restaurant with the highest rating score 4.9 is Sandwich, Cafe, followed by Russian, Bakery, Coffee and Tea, and Cafe, French, Bakery, with a score of 4.8. The third welcomed are the restaurants of Greek, Mediterranean, Pacific, Northwest, and Bakery, French, Desserts, rating as 4.7. Restaurant genres rated as 4.6 come the forth, including Sushi, Japanese, Latin American, Caribbean, Sandwich, Cuban, Cafe, Bakery, and Burger, Diner, American. In Singapore, dishes of Singaporean, European, Seafood, Japanese, Continental , and South India, India are the most popular, with an average score of 4.3. American, Continental, French, Seafood, and Mediterranean, Asia, Japanese, Sushi, Seafood, French ranked second and third, with an average rating score of 4.2 and 4.1 respectively. The cuisine with a score of 4.0 comes third, including European, Spanish, American, Cafe, Indonesian, Singapore, Chinese, Vietnamese, Seafood, Japanese, India, South India, North India, Asian, and Asian, Continental, Seafood, Bakery. India, North Indian and Hainanese, Chinese ranked fourth, with a rating score of 3.9.

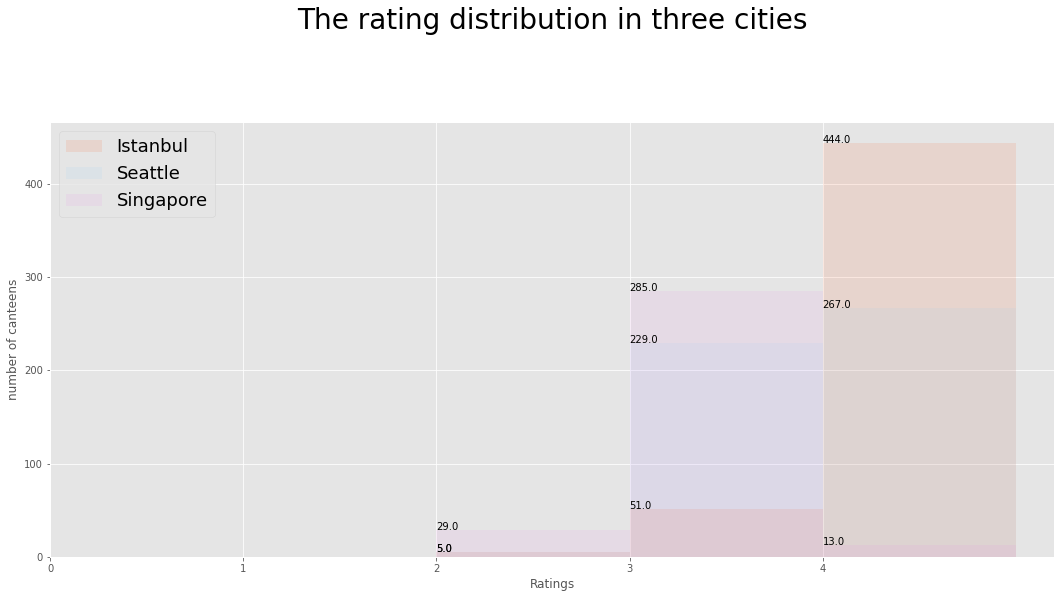


From the pie chart of *The distribution of Top 10 most popular food* *in three cities*, it has been found that in Istanbul, Restaurant Cafe has the largest number, accounting for 23.9%. World Cuisine comes second, accounting for 20.0%. Cafes with a distribution proportion of 12.2%, ranks third. In Seattle, genres of Italian, Thai, Sushi and Japanese are the most widely distributed, reaching 14.4%. American and Mexican food are the second, with 10.9% of total genres. The third is Pizza, Italian, accounting for 9.2%. In Singapore, the number of canteens with Italian food is the largest, accounting for 16.8% of the total number of restaurants. Japanese, Singapore and Chinese restaurants are the second most widely distributed. The third is Sushi, Japanese, accounting for 11.9% of the total.



Knowing what the most popular cuisines are in each city, the second RQ is to explore the rating and price distribution for Istabul, Seattle and Singapore.

From the histograms of the *rating distribution in three cities*, there are 29 restaurants rated between 2 and 3 in Singapore, 5 in Seattle and 5 in Istanbul. Besides, Singapore has 285 restaurants rating between 3 and 4, Seattle has 229, and Istanbul has 51. For the canteens rating above 4.0, Istanbul is the winner, with 444 (in total of 500) such restaurants. Seattle came in second with 267. In contrast, Singapore has only 13 restaurants with scores above 4.0.

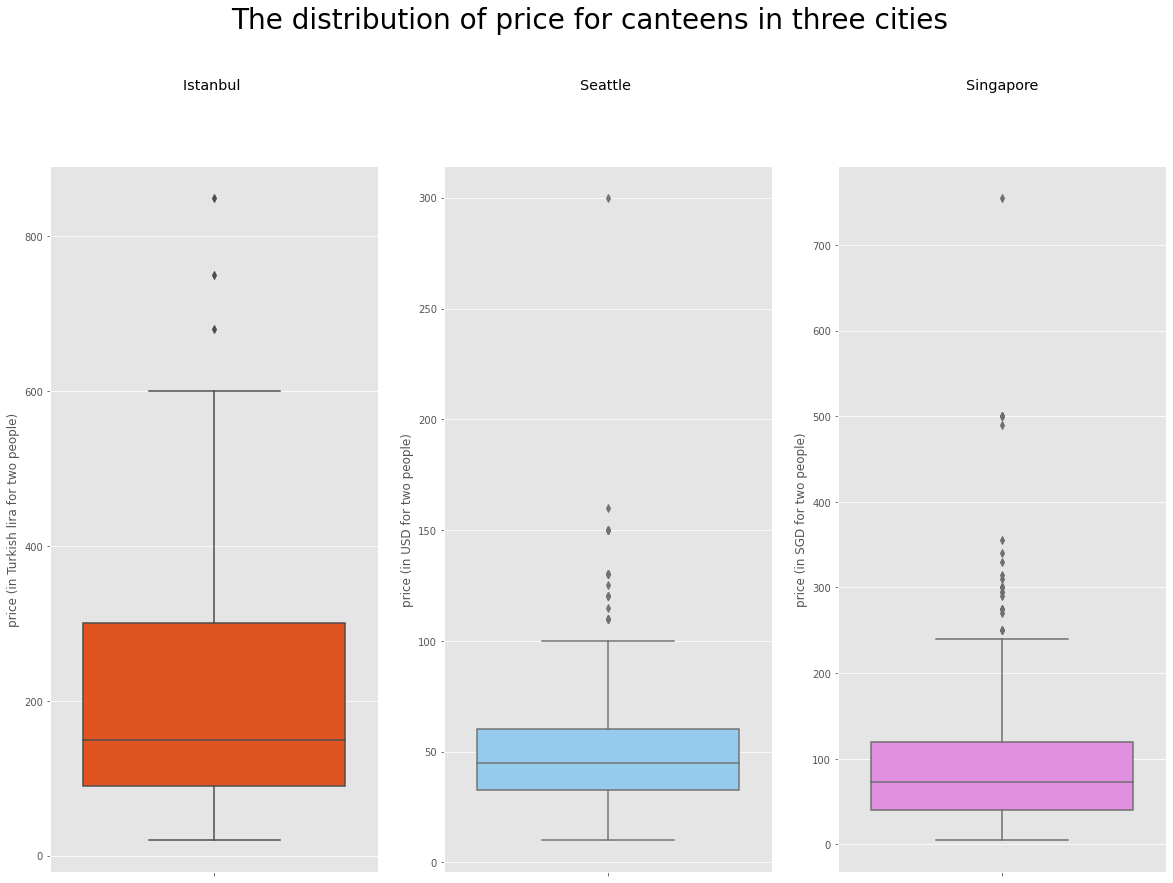


As for the price for eating in these cateens, from the histograms of *distribution of price for canteens in three cities*, it can be seen that, in Istanbul, restaurants with prices of 10-100 Turkish lira account for the most, reaching about 34%. More than half of the restaurants are priced between 100 and 340. The number of restaurants with prices between 340-600 accounts for about 13%. Less than 1% of restaurants set the price at more than 600 Turkish lilar. In Seattle, 50% of restaurants cost between $35 and $70. Restaurants priced between $15 and $35 accounted for about 31%, ranking second. Restaurants with consumption between 70-155 ranked third, accounting for about 20% of the total number of restaurants. Restaurants pricing at $155-180 and $275-300 together account for less than 1%. In Singapore, about 52% of restaurants set food prices at 10-80 SGD. Restaurants that set the price between 80-150 and 150-300 account for about 34% and 12% respectively. Restaurants that consume between 300-380 and 460-520 together account for about 2%. Restaurants with an average consumption of more than 700 only account for less than 1%.

Given the fact that 100 Turkish lilar equals 6.78 USD, and 100 SGD equals 72.78 USD, the result implies the cost for a meal in Turkey should be much lower than in USA and Singapore.

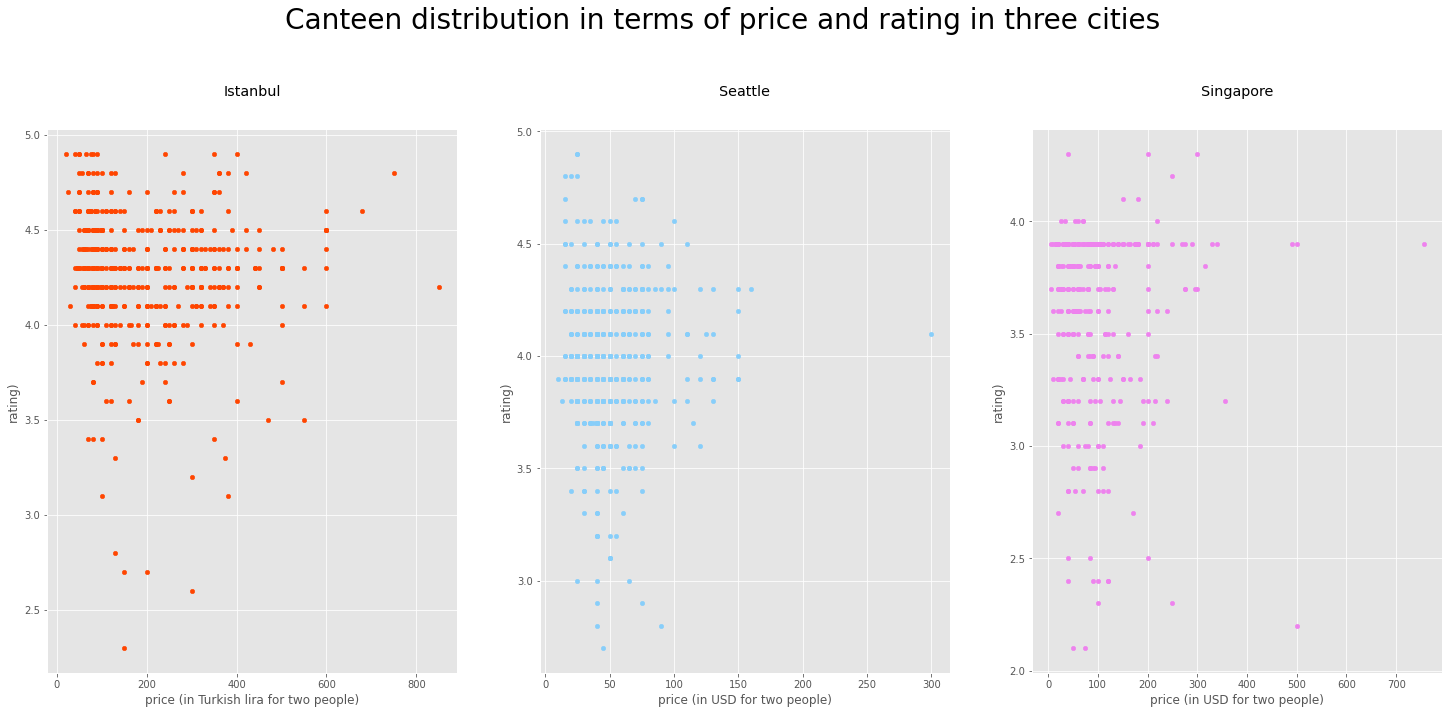


Specifically, from the boxplot of *The distribution of price for canteens in three cities*, it can be clearly seen that compared to Seattle and Singapore, the median price of restaurants in Istanbul is the lowest, which is around 160 Turkish lira (about 11.2 USD). Singapore has the highest median price with 70 SGD (about 51 USD), while Seattle has a moderate median price at 45 USD. Besides, there are more outliers for Singaporian restaurants, whose prices are higher than the maximum of the plot, and reaching 760 SGD at most (almost 555 USD).

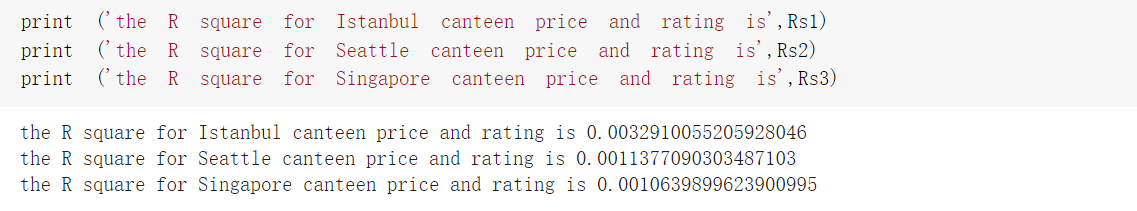


The last RQ for this part is to explore the relationship between price and rating as you may be wondering whether a canteen that costs you hundreds of bucks necessarily means that it serves the best quality food. Based on the data of the three cities, there comes a conclusion that price for food has little to do with its quality.

The *Canteen Distribution in terms of price and rating in three cities* demonstrates that canteens are evenly distributed in terms of price and rating in these three cities, that is, in each price range, the possibility of the canteen to be rated as good, mediocre and bad is perhaps just the same. It can be proved from the scatter plot below as none of the three cities generate a clear regression line for the relationship between the price for canteen and its rating.



As per calculation in Python, the R square in regression analysis for Istanbul is .03, Seattle .001, and Singapore, .001.Suffice to say, the price for each meal in the canteens account for less than 1% of its rating (which indicates the canteens’ overall quality) in all of the three cities.



Taking all the results into account, it can be seen that, compared to Seallte and Singapore, Istanbul is an ideal place for food as it is endowed with the most high-rated canteen and the cost is relatively low. However, an expensive canteen does not necessarily mean that it serves the best quality food.

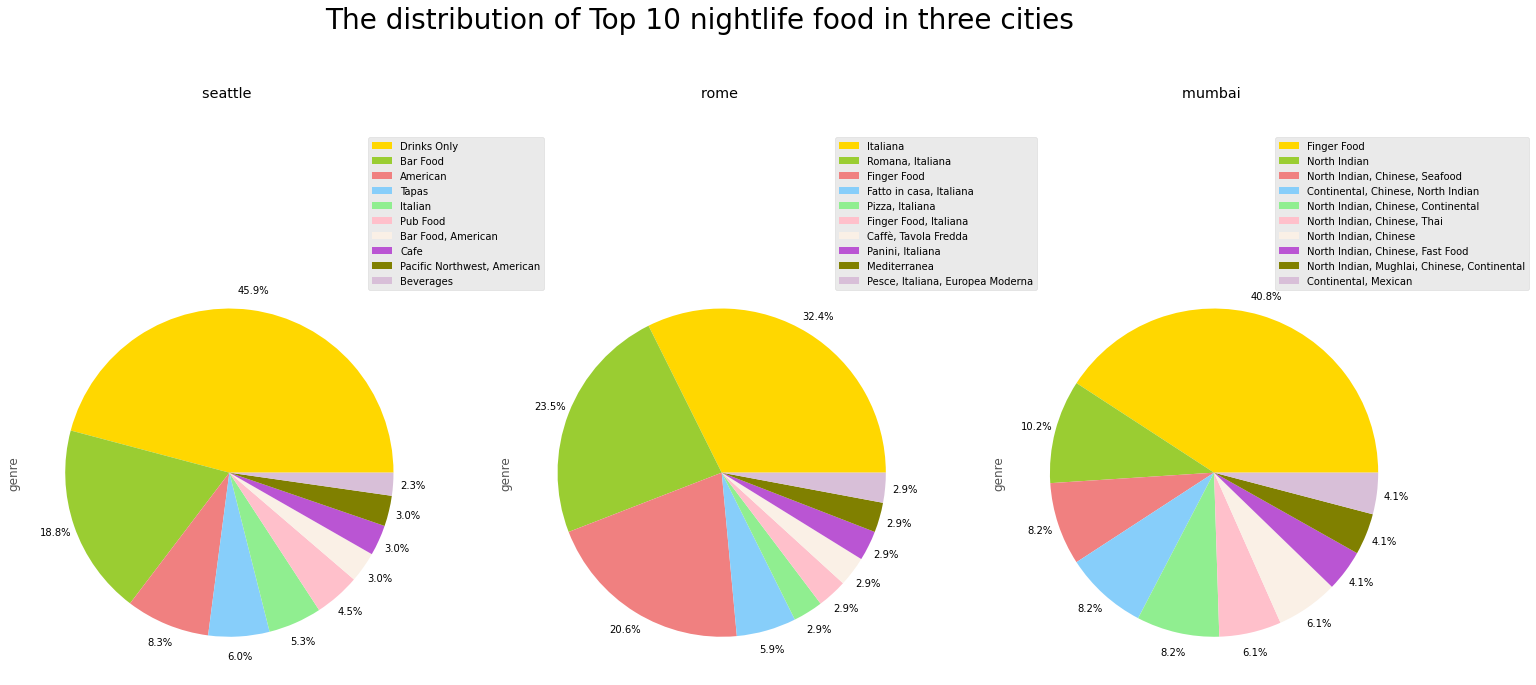
***Nightlife***

In addition to knowing the local cuisine, experiencing the local nightlife is also an essential part of the traveler's experience of native conditions and customs.

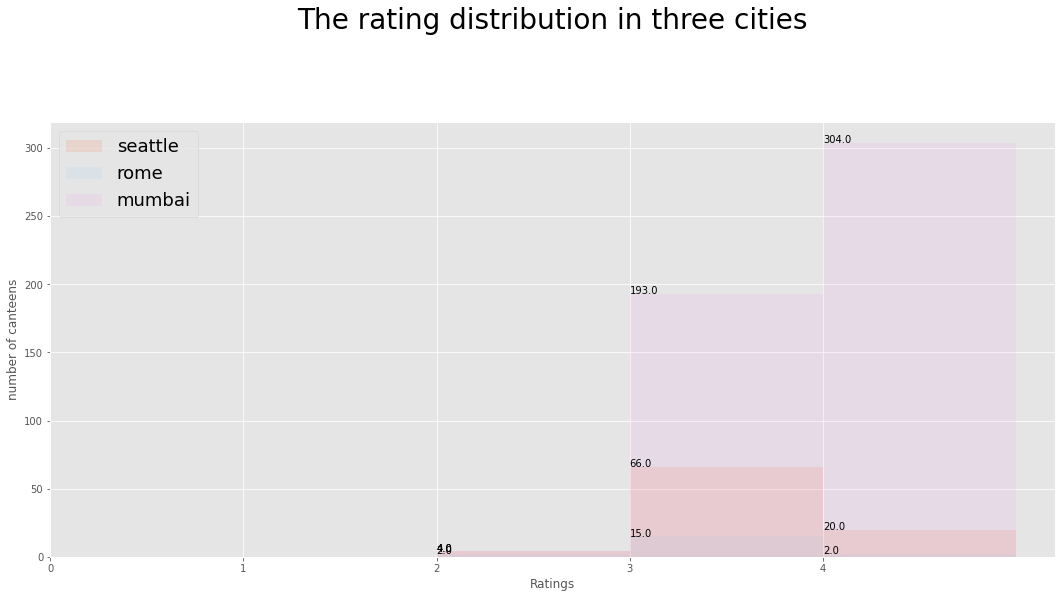
The plot shows the ten most popular foods for nightlife in Seattle, Rome, and Mumbai. As we can see, most of the evening favorites have local specialties. Bar Food, American received a favorable 4.10-grade point average, and Pacific Northwest, American obtained an average rating of 3.95. And as the capital of Italy, Rome, nine out of ten nightlife cuisine types of Italian meals are loved by customers. Europea, Pesce, Italiana, Piemontese, and Pugliese are all those types of nightlife dishes with an average rating score of 3.90. For the city of Mumbai, the top four nighttime foods are dominated by Middle Eastern and Asian cuisines. For example, North Indian food, Middle Eastern food, and Thai food received an average rating of 4.90. Of course, as the protagonist of the night beverages is also an indispensable part of Mumbai nightlife, the average rating score reached 4.90.



From the pie chart, the distribution of night food genres among Seattle, Rome, and Mumbai can be seen more clearly. In Seattle, 45.90% of customers order drinks only during midnight. A few people around 2.30% like to enjoy beverages in the evening. 18.80% of people in Seattle would select bar food in the midnight restaurant. At night restaurants, Italians especially love the cuisine of their homeland, the figure was taken 32.40%. Only 20.6% of people chose finger food that doesn't belong to Italian cuisine. You can see how much people in Rome like to order local cuisine. The pie chart for Mumbai also shows this trend, the top ten genres of nightlife food are mainly aggregated with North Indian, Chinese, and Finger food. 40.80% of people order finger food in the midnight inn. 10.20% of people choose North Indian cuisine as their midnight snack.



If you are night owls, you may wonder which city is the most suitable place for hunting tasty night restaurants. From this histogram, you can directly see that there are three hundred and four midnight restaurants in Mumbai that received rating scores above 4.00, and twenty restaurants in Seattle obtained a rating of 4.00 or higher. Unfortunately, the capital of Italy – Rome doesn't have any nightlife inns that received a score of more than 4.00. It may be due to the local people having high standard food labeling or any other reason like there are not enough night restaurants in Rome etc. Travelers can, according to this histogram, find favored night restaurants to decide which cities they want to go to in the future.

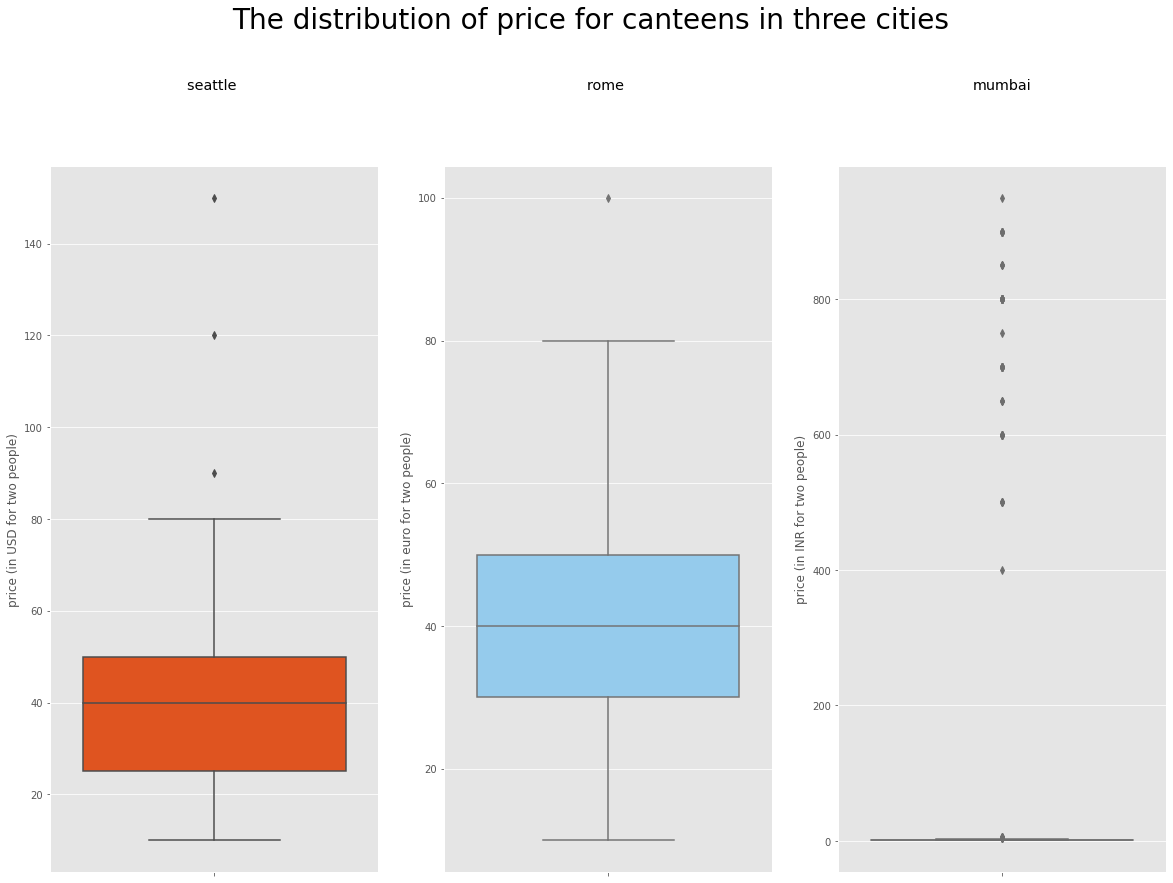


Price is also a necessary point for people to hunt night restaurants. The three histograms show the distribution of prices for restaurants in three cities. Most of the prices in Seattle night restaurants are in the range of 30.00 US dollars to 50.00 US dollars for two people. In Rome, around 24% of night restaurants range in price from 30.00 euros to 50.00 euros for two people. The prices of restaurants in Mumbai are much cheaper than those in Seattle and Rome. Over 80% of night restaurants range in price from 00.00 INR to 100.00 INR for two people. You can carefully consider this factor for which cities you would like to go.

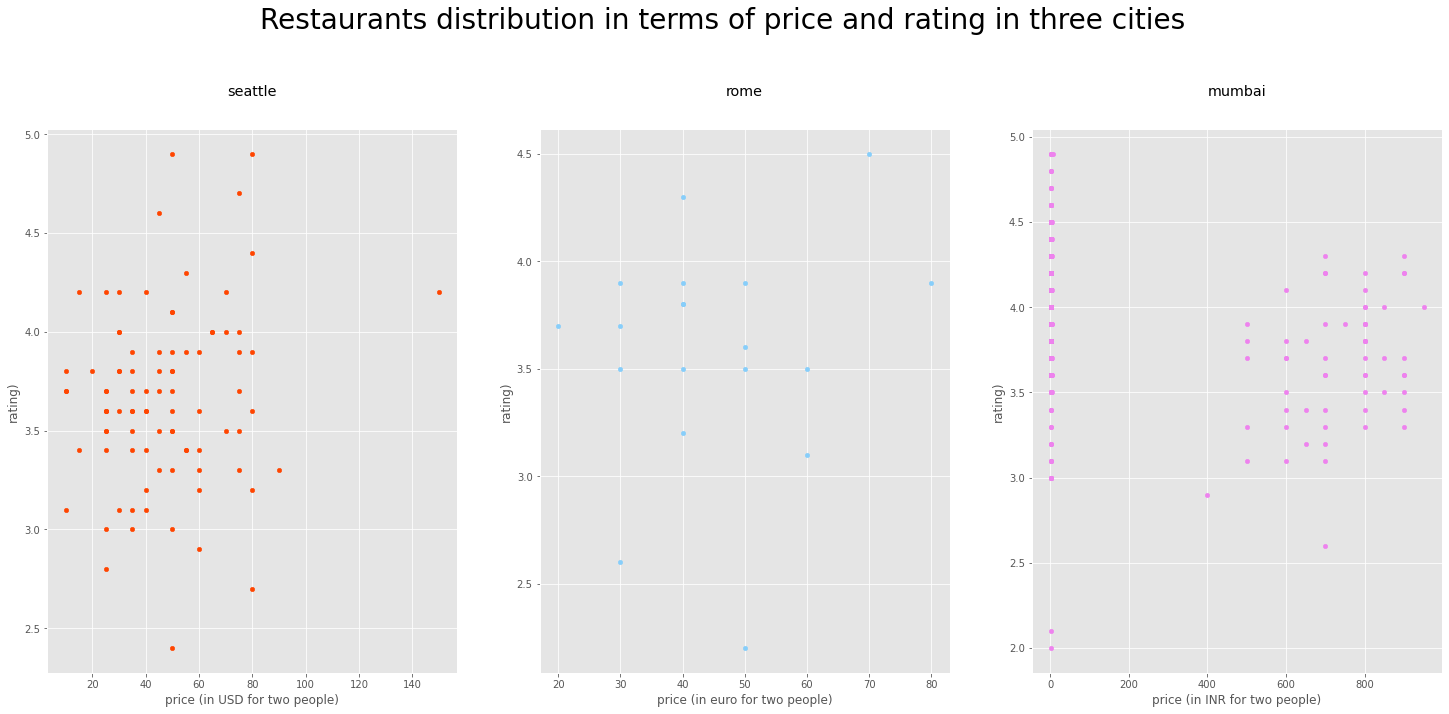
Here are the currency exchange rates, which will help you get a better idea of the average nightly restaurant price in Seattle and Rome: 7.83 euro equals 8.46 USD. So that the price in Rome for night restaurants is higher than night restaurants in Seattle.



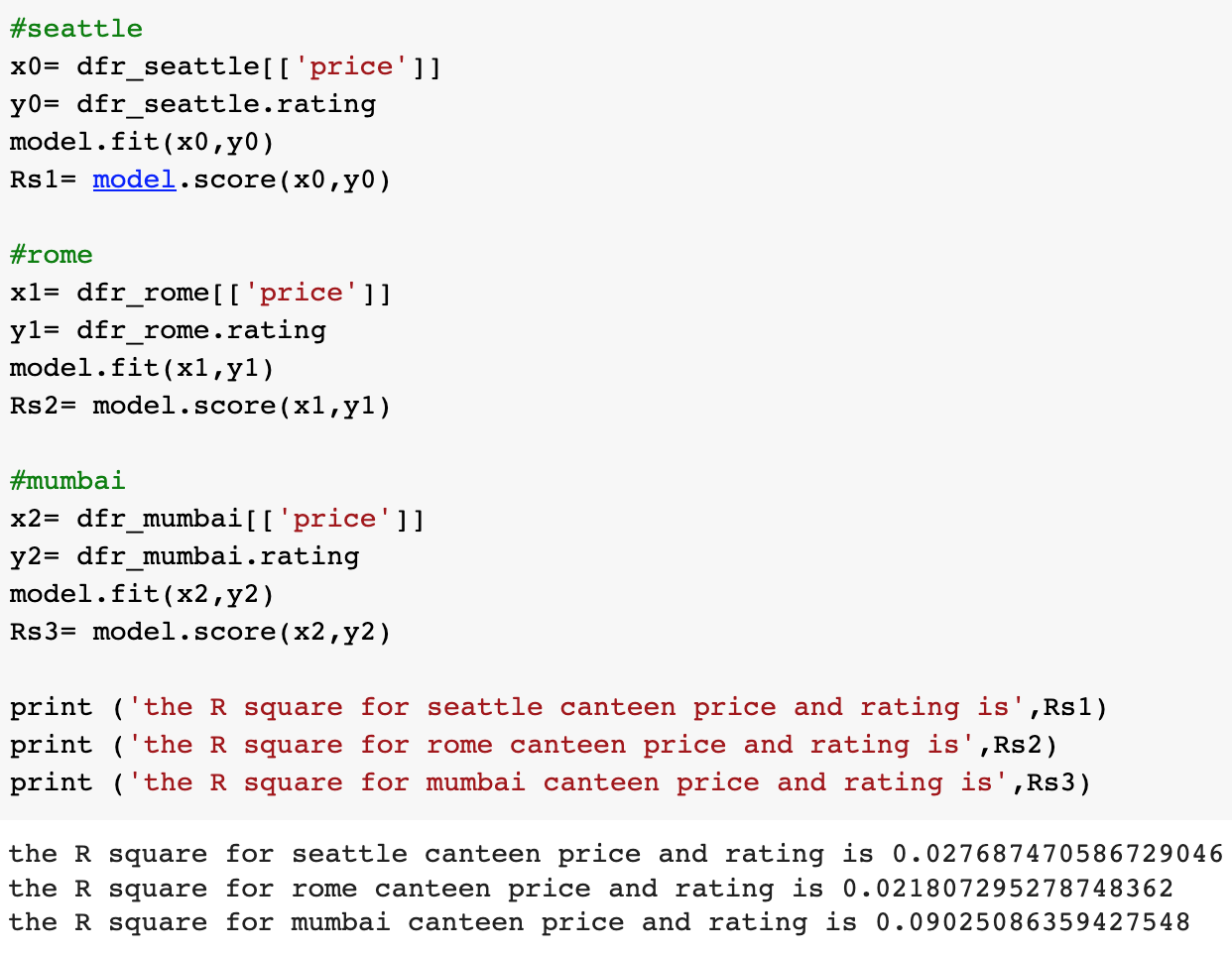
The boxplot shows more clearly the distribution of prices in three cities of night restaurants. The median price for Seattle and Rome are all equal to 40.00, but due to the different currencies, it can be seen that the median price in Rome for night restaurants is higher than for night restaurants in Seattle. Compared to the three cities, night restaurants in Mumbai are cheaper than those two(Seattle and Rome).



Lastly, using a dot plot to compare the relationship between prices and ratings of restaurants in the three cities. Assuming the rating score above 4.00 is a high grade. Seattle and Mumbai have more and better night restaurants, as shown by the density of the dot divisions, but Mumbai's restaurants are more affordable in terms of price. Rome's night restaurants have higher prices and lower ratings.



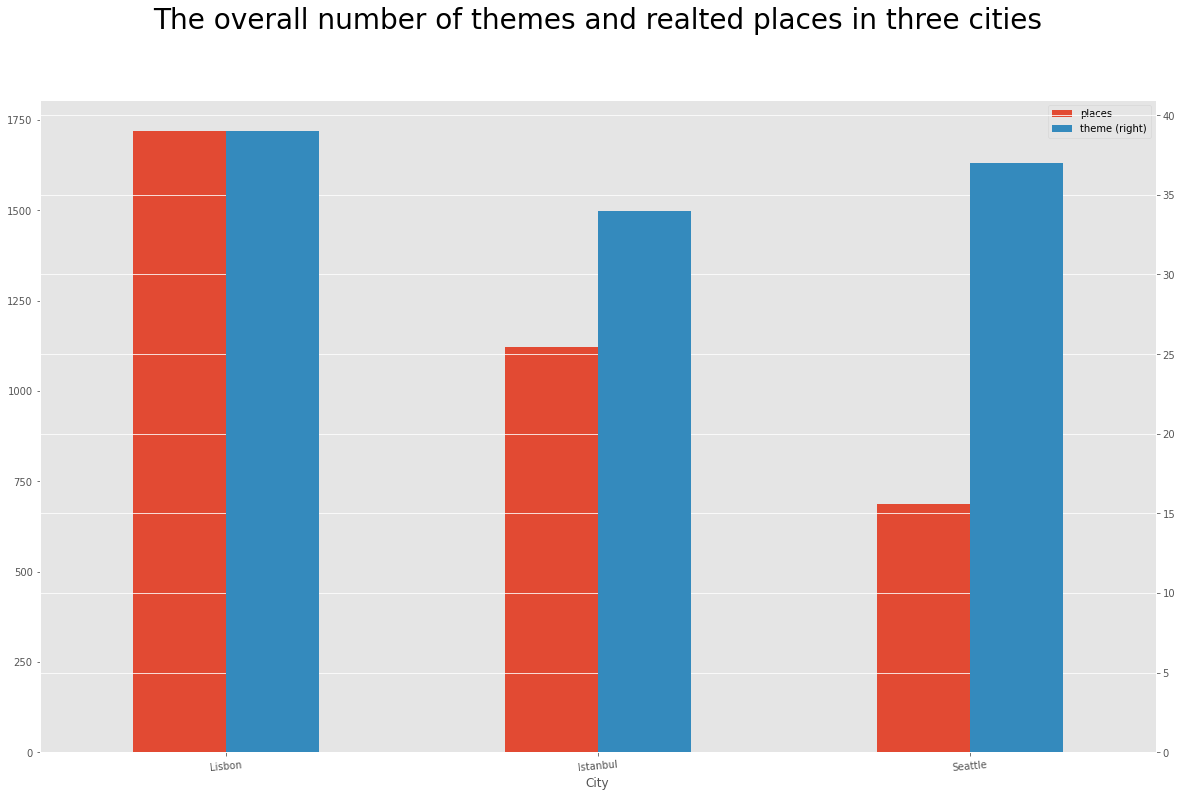
As per calculation in Python, the R square in regression analysis for Seattle is 0.027, Rome is 0.021, and Mumbai is 0.090. Obviously, night restaurants' prices are related to their rating. Mumbai has the best night restaurant value among the three cities.



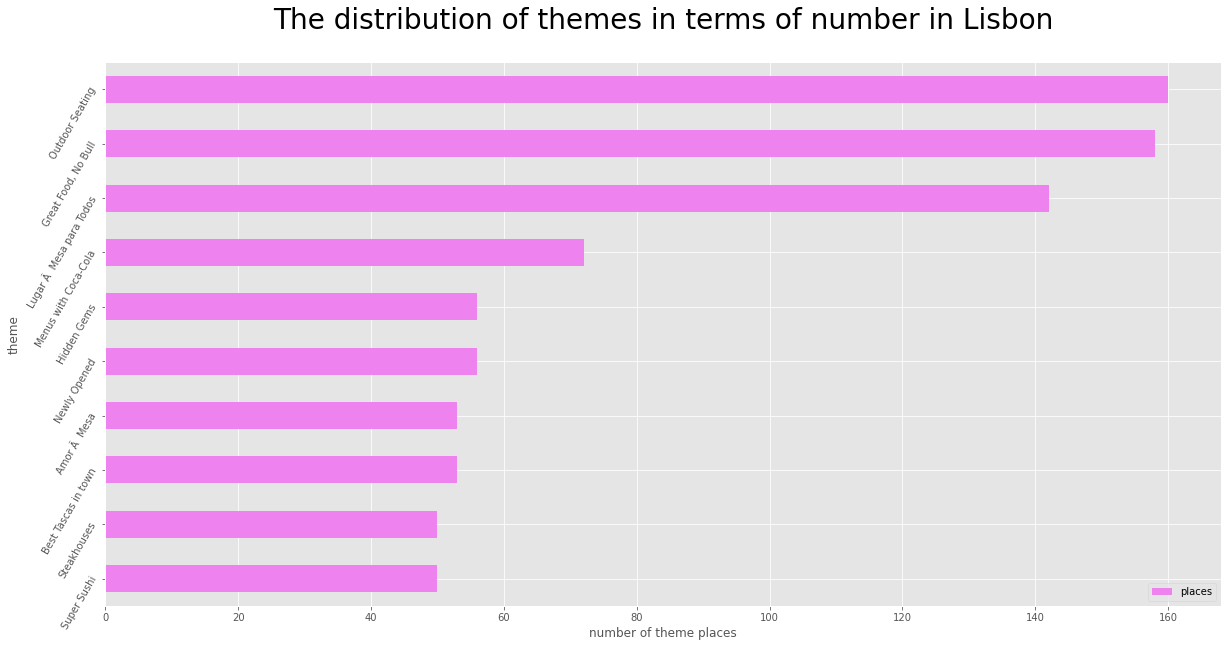
***Collection***

The units of analysis in this part are Seattle, Istanbul and Lisbon. The thematic activities vary in these cities. Therefore, the visualization results could provide a map on the lifestyle of the people in North America, Europe, and the Middle East.

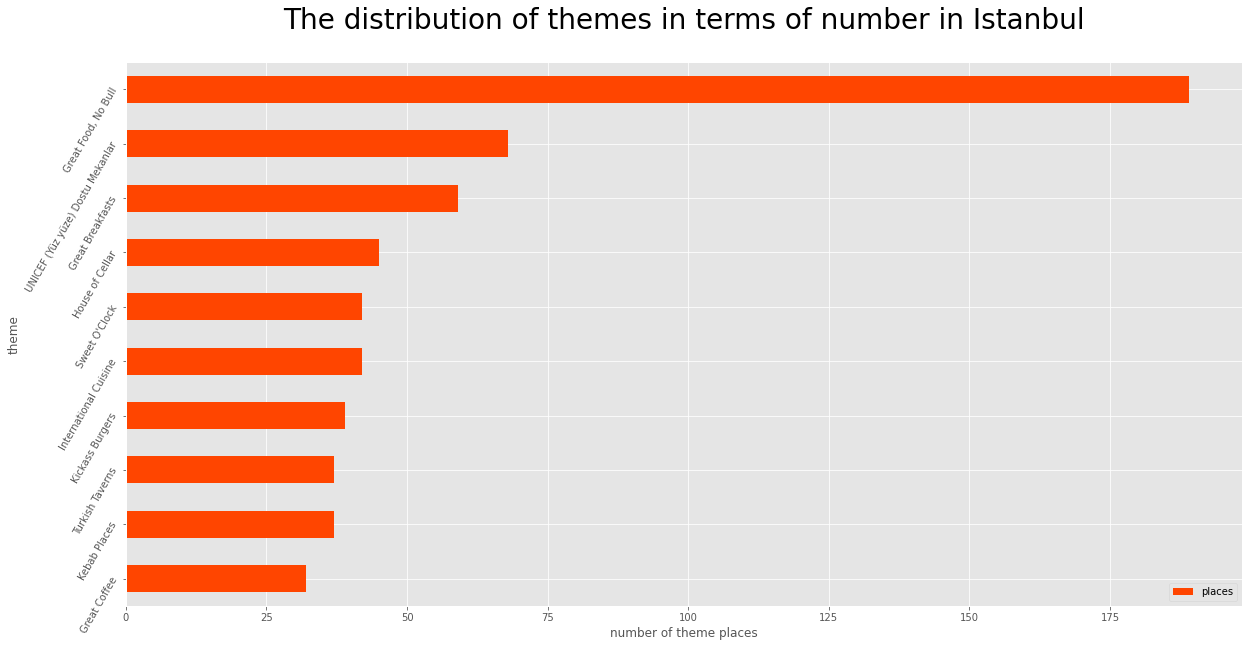
Not surprisingly, Lisbon (short for Greater Lisbon) is the city with the most places for fun. The capital in Portugal enjoys 39 thematic activities and 1719 related places on *Zomato*. Isbanbul, a capital city of a Muslim country comes the second as it has 34 themetic activities and 1121 places in the collection page. And surprisingly, Seattle, being an American city, is the most ‘boring’ place to be as there are just 37 thematic activities including 687 places to chill on the city’s website.



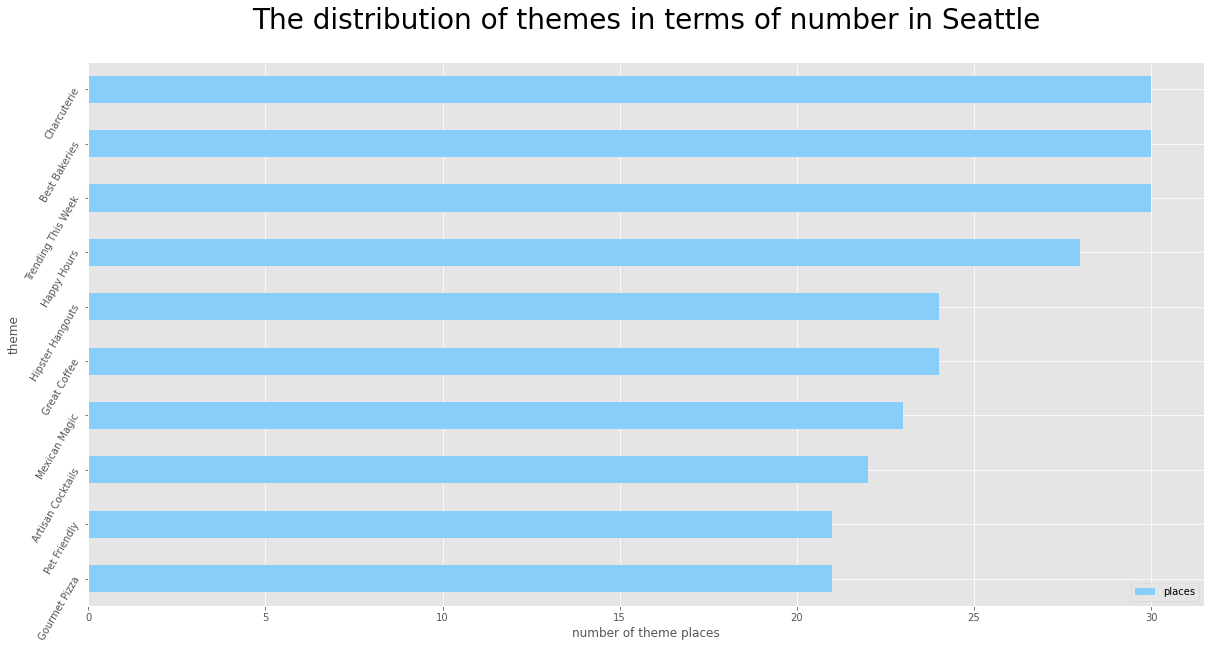
In terms of consumption behavior, the Portuguese (or perhaps it could be generalized to most of the Southern Europeans) favor Outdoor Seating the most as there are as many as 160 related places in Lisbon.The Great Food comes after as there are 158 related places. Lugar Ã Mesa para Todos (*a place at the table for all* in English) ranks third as there are 142 places which include not only some canteens but desserts and cafeterias. What comes fourth is the Menus with Coca-Cola, where you could find everything that could be savored with a Coca-Cola, such as pizza, Japanese Sushi (*what?*) and burgers.



The Turks like Great Food the most as there are 189 related places, and then the UNICEF (Yüz yüze) Dostu Mekanlar (*UNICEF (Face to Face) Friendly Places* in English) with 68 places, including coffee shops, desserts and even bar food. Unfortunately, the bar food shops only supply alcohol-free drinking such as Soda and Cola in Istanbul. Great Breakfasts comes the third most popular thematic activities in Istanbul as there are 59 related places.



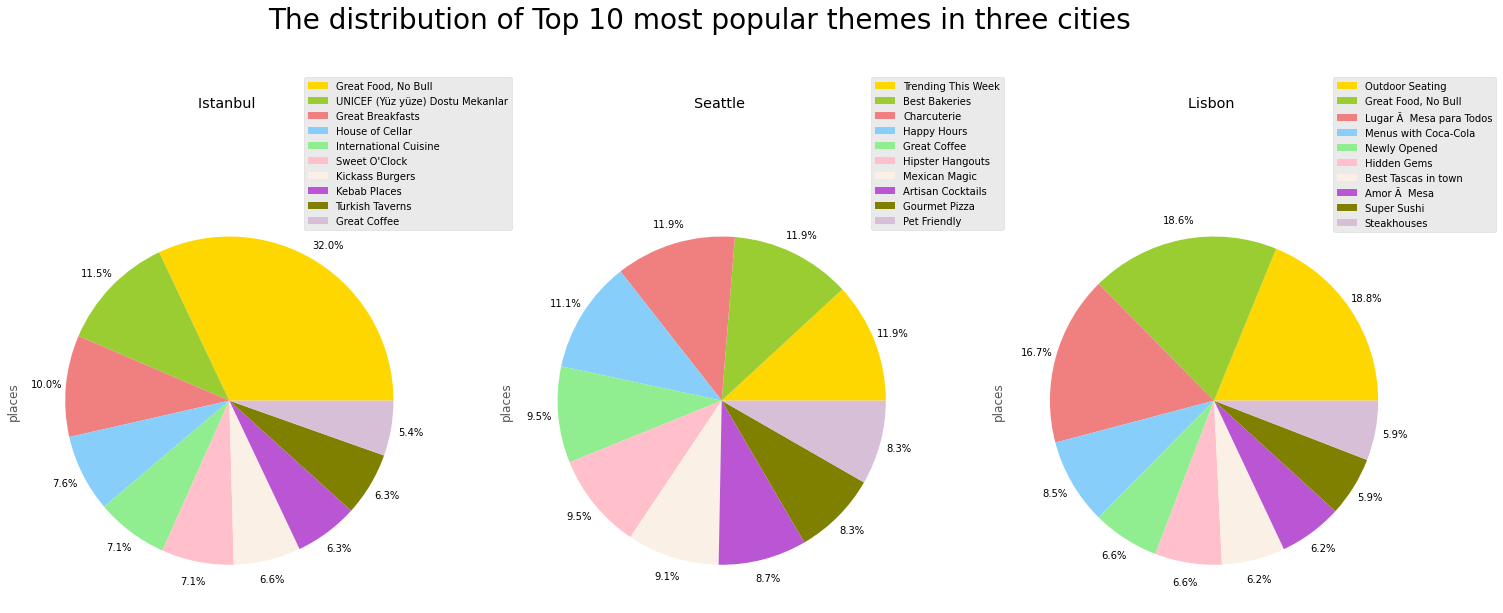
Life in Seattle is rather boring, as there are 30 places for Best Bakeries, 30 for Charcuterie (a French term referring to a branch of cooking devoted to prepared meat products, such as bacon, ham and sausage), and 28 for Happy Hours. A closer look reveals that the Happy Hours includes some seafood, Italian food and some taverns.



From the pie chart below, it can be seen that people in Lisbon like outdooring the most, as outdoor seating accounts for 18.8% of the overall top 10 thematic activities, and the Great Food, 18.6%. Turkish muslims like Great Food(32.0%), Face to Face Friendly Places(11.5%) and Great Breakfast (11.5%) the most. They are all about eating.

While lastly, for the people in America, they do not show a favor to a particular theme, as Bakeries, Charcuterie, Happy Hour and Great Coffee all share a similar percentage at around 11%.

To sum up, according to the data from *Zomato* collection pages, the Southern Europeans are believed to have a cozy lifestyle, and the people in the Middle East just love THE FOOD. While for Americans, work might weigh much more in their daily life as coffee is the thing on their top 10 most popular thematic activity list. So if you are a traveler looking for a funny experience, Lisbon and Istanbul might be your good options.



**4. Limitation**

There are several limitations of this study that could draw your attention.

Firstly, due to the jet lag, when scraping the data from different cities, the results reflected the information about the canteens which were available at different time of one day in a specific city, which could not be best for comparison.

Secondly, since the sample size of this study was small, the findings of this study may not be generalized and apply to other cities. For example, It is not certain that the finding of people in Portugal leading a cozy lifestyle could be applied to their counterparts in Iberian Peninsula or Penisola appenninica. Further studies are encouraged to fill such gaps.

Lastly, as *Zomato* does not have a high penetration rate in some markets, the sample of this study may not be representative. As per the financial report of *Zomato*, India contributes most of its revenue, UAE the second and the rest of the world are categorized as a reportable segment of ROW. As such, *Zomato* may not be popular in countries besides India and UAE, and the data scraped from cities other than Mumbai may not truly reflect the consumption behavior, insights and lifestyle of all the people in all of the sampled cities.

***Labor distribution matrix among the team members***

| Job | MENG Guosheng 21431280 | REN Tongyang  21407738 | SHI Minting  21422443 |
| --- | --- | --- | --- |
| Data Scraping | ☑️ | ☑️ | ☑️ |
| Data Visualizaion | ☑️ | ☑️ | ☑️ |
| Report | ☑️ | ☑️ | ☑️ |
| Powerpoint | ☑️ | ☑️ | ☑️ |

***Reference:***

Auty, S. (1992). Consumer choice and segmentation in the restaurant industry. The Service

Industries Journal, 12, 324–339.

Matthew, S. (2019, Mar 12). *Italian cuisine is world’s most popular.* YouGov.

[https://yougov.co.uk/topics/consumer/articles-reports/2019/03/12/italian-cuisine-worl](https://yougov.co.uk/topics/consumer/articles-reports/2019/03/12/italian-cuisine-worlds-most-popular)

[ds-most-popular](https://yougov.co.uk/topics/consumer/articles-reports/2019/03/12/italian-cuisine-worlds-most-popular)

National Restaurant Association. (2011b). Restaurants by the numbers. Retrieved from

<http://www.restaurant.org/pdfs/research/2011forecast_pfb.pdf>

Panigrahi, C. M. A. (2020). A case study on Zomato–The online Foodking of India. *Journal*

*of Management Research and Analysis*, 7(1), 25-33.

Vu, H. Q., Li, G., Law, R., & Zhang, Y. (2017). Exploring tourist dining preferences based on

restaurant reviews. Journal of Travel Research, 58(1), 149–167.

<https://doi.org/10.1177/0047287517744672>