

Descriptive analysis of the presence of the restaurant industry across major cities in the world

Achyut Sankhe

April 20, 2020

1 Introduction:

1.1 Background

The most densely populated cities in the world attract businesses, as the latter see an opportunity to maximize profits. This applies to the restaurant industry as well. Many fast food chains, restaurant chains and restaurant owners look to expand their business across various major cities. However, they have to adjust to the eating habits and general eating culture of the city they want to expand to. For example: Some cities might have eating habits that encourage them to eat at home more often than eating out while some cities it's the opposite. To understand this concept more, a preliminary analysis of existing venues that represent the restaurant industry within such largely populated cities needs to be conducted so as to understand which major cities across the world have more restaurants i.e. are more conducive to restaurants and where do restaurants/ fast food chains need to focus more energy in terms of advertising and marketing to make their business more profitable as well as the potential causes that would validate such action.

1.2 Problem

To conduct this analysis we will go through lots of location and trending venues data as well as population statistics. The aim of this project is to set up data frames and charts that will provide preliminary understanding regarding trending restaurant industry venues across various cities and how the population of these cities is linked to these venues.

1.3 Interest

The main stakeholders in this project are owners and workers of the restaurant industry. This understanding will give them some basic understanding of the current state of the industry in major cities across all continents and what obstacles they might face setting up a successful business in those respective cities.

2 Data requirements and source:

I have chosen one of the most densely populated cities from each continent. The choice of cities is as follows: (a) Asia: *Tokyo* (b) Africa: *Lagos* (c) Europe: *Istanbul* (d) Australia: *Sydney* (e) North America: *New York* (f) South America: *Sao Paulo*

The location data for each of these cities i.e. latitude and longitude will be required for this analysis. Population data for each of these cities was also utilised and was adjusted to check against the restaurant density within a certain radius in each city. I also extracted and created a dataset containing all the venues present in the respective cities and particularly restaurants and eateries which was evaluated to get a preliminary understanding of the dependency of a particular city on restaurants

2.1 Data source:

For this descriptive analysis, I have used the services of the Foursquare API to get the venue data for each of the aforementioned city. The data has been constrained to top 100 trending venues within a 10 km radius of the city coordinates. The coordinates for each city have been derived from the geopy library in python. Furthermore, the population statistic used for analysis in this project will be the population density per square kilometre. Numbers pertaining to each city are based on census data taken from various websites mentioned in the References section of this document.

3 Methodology:

3.1 Data cleaning:

I initially derived latitude and longitude data for the corresponding city via geopy. Then, using the Foursquare API service, I created a data frame of the top 100 trending venues in a 10 km radius. The data frame contained the name of each venue and the venue category that corresponds to each venue. I was interested in venues corresponding to the restaurant industry. Firstly I duplicated this data frame. The reason behind this is I needed to carry out two separate analyses: (1) An analysis of the restaurant industry with other venue categories representing other industries. (2) An analysis of competition from within the restaurant industry as well.

3.2 Exploratory data analysis:

The initial part of the data analysis was focused on the venue category statistics from each city. The analysis was divided into parts using two data frames that were identical to each other.

- (1) For the first part of the initial analysis I grouped all the venue categories that I deemed fit to be included in the restaurant industry. This included categories such as Bar and Café. I then replaced all of these categories in the data frame to a single category called 'Restaurant Industry'. I also dropped the latitude longitude data as it was not required for this part of the analysis. Using this modified data frame I applied data visualization techniques to make

statistical inferences on how much competition the restaurant industry faced within each city

- (2) For the second part of the initial analysis I dropped all the venue categories that did not belong to the restaurant industry and again dropped the latitude and longitude data. This part of the analysis focused on which trending venue categories corresponding to the restaurant industry were currently popular within the respective cities.

This process of creating and cleaning the data frames and applying data visualization techniques to understand the impact of the restaurant industry on the corresponding city was repeated for all six aforementioned cities. Furthermore, these data frames were then combined to understand the overall distribution of Restaurant venues across all continents.

The latter part of the analysis focused on incorporating population data into the created data frames. For this I created a new data frame using relevant variables such as 'Number of Restaurants ', 'Overall Venue Categories' from the initial part of the analysis and also incorporated new variables based on population statistics such as 'Potential population in a 10 km radius' whose variables were extrapolated from the population per square kilometre statistic so as to account for the 10 km radius set up in our Foursquare API call. Using this data frame I created a bubble plot to understand the how the three aforementioned variables are interacting with each other in each city. This bubble plot can potentially give preliminary insight into the interaction of respective city population with the restaurant industry.

4 Results:

Beginning with the initial part of the analysis the results focus on the restaurant venues and overall venue categories for each city. Venue categories such as 'Bars' and 'Bed & Breakfast' were included as part of the industry for the sake of this analysis

4.1 Tokyo, Japan

In Tokyo there were overall 24 venue categories within a 10 km radius including the restaurant industry. 59 venues represented the restaurant industry category. So 59 % of the venues belonged to the restaurant industry with competition from 23 other venue categories. The main competition was from the hotel industry and the second highest competitor was the Art museum venue category (Figure 1).

Within the 59 venues representing the restaurant industry the highest representation was from 'Coffee Shops' at 5 and the second highest was tied between 'Chinese Restaurant' and 'Sake bar' at 4 each. Furthermore there is healthy competition from venue categories such as

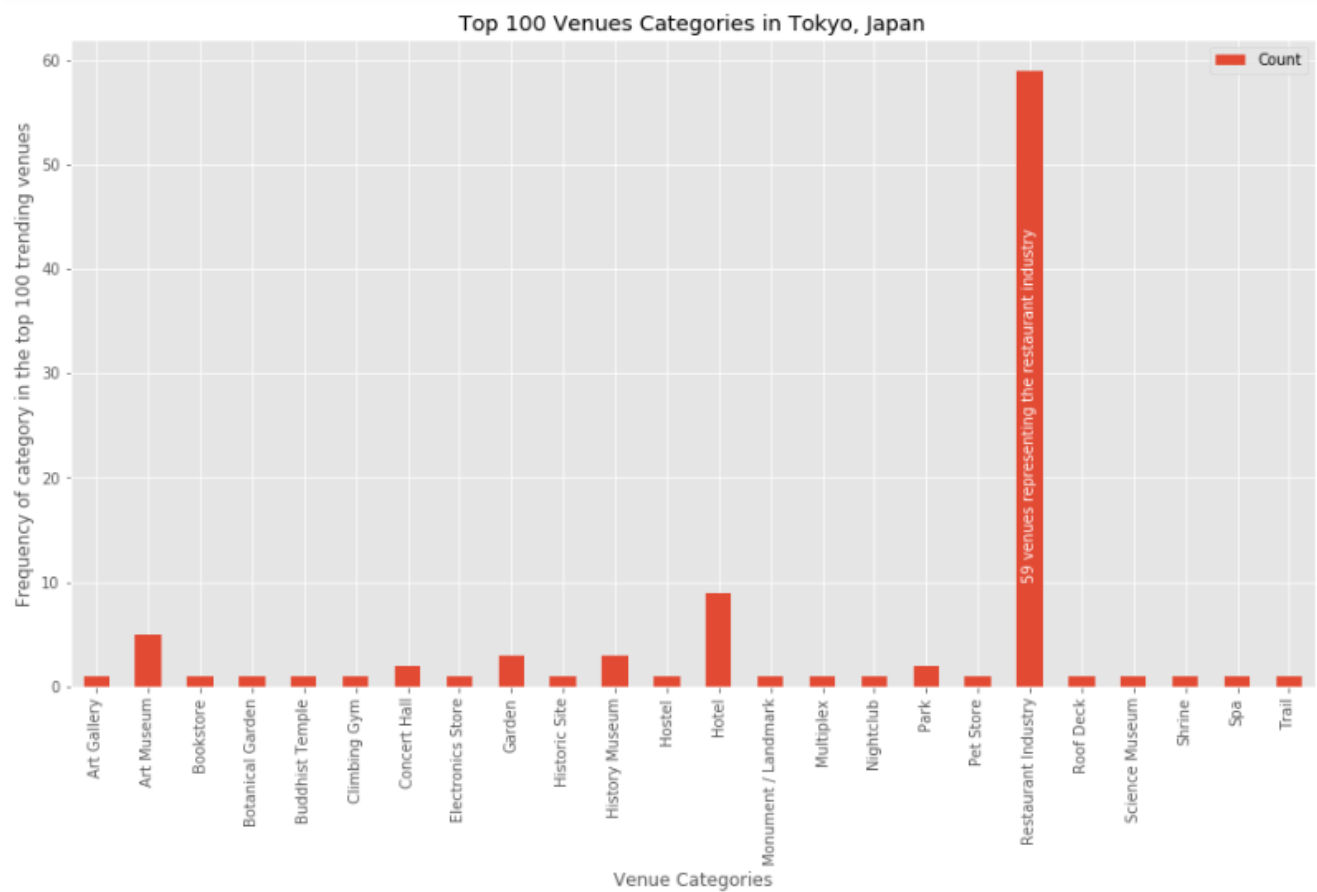


Figure 1: Venue Categories in Tokyo, Japan

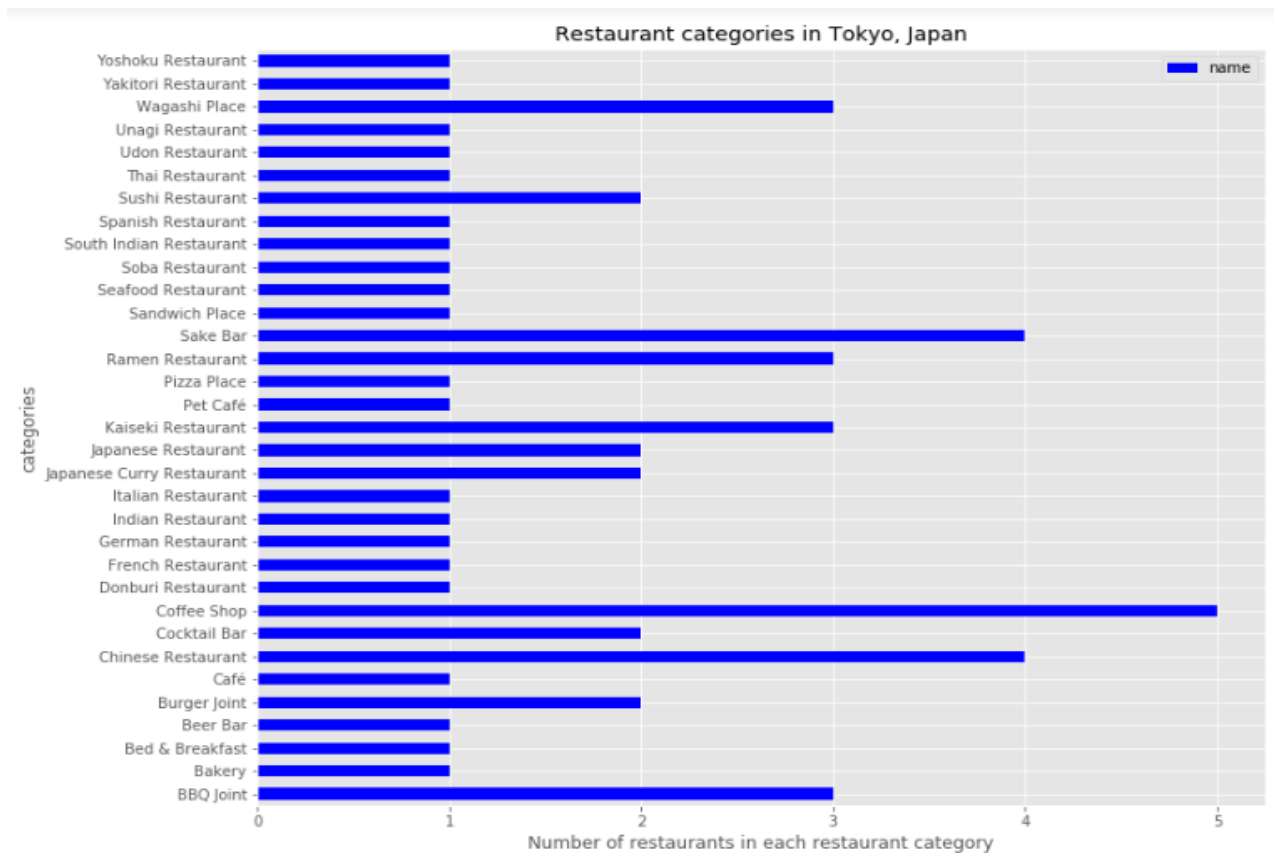


Figure 2: Restaurant Categories within the restaurant industry in Tokyo, Japan

4.2 Lagos, Nigeria

In Lagos there were overall 29 venue categories within a 10 km radius including the restaurant industry. 47 venues represented the restaurant industry category. So 47 % of the venues belonged to the restaurant industry with competition from 28 other venue categories. The main competition was from the hotel industry followed closely by the 'Lounge' and 'Nightclub' venue categories (Figure 3).

Within the 47 venues representing the restaurant industry the highest representation was from 'African Restaurant' category at 8 and the second highest was tied between 'Pizza Place' and 'Ice Cream Shop' categories at 4 each. Furthermore there is healthy competition from venue categories such as 'Café', 'Bar', 'Fast Food Restaurant' etc. representing the general food categories and exotic cuisine (Figure 4).

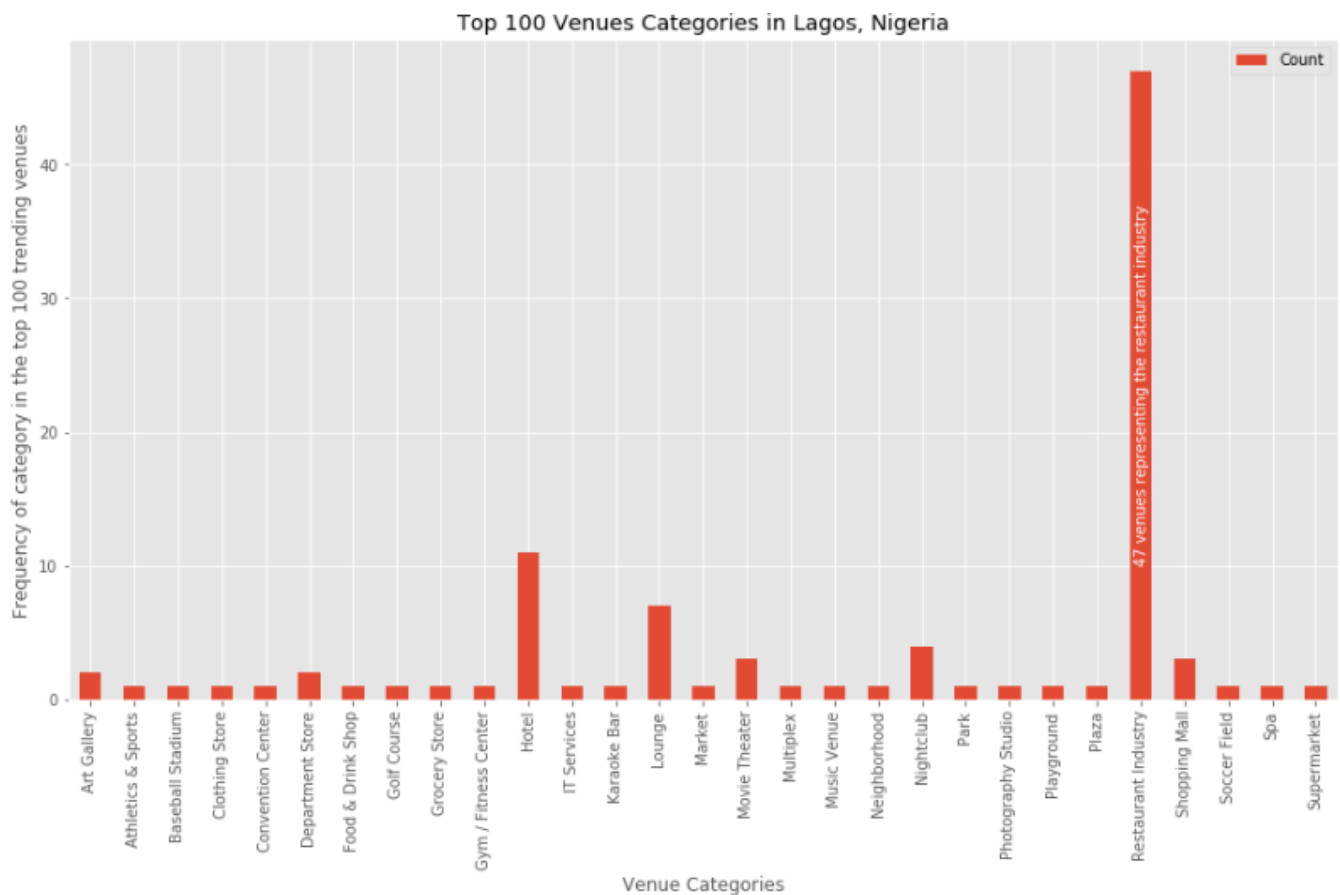


Figure 3: Venue Categories in Lagos, Nigeria

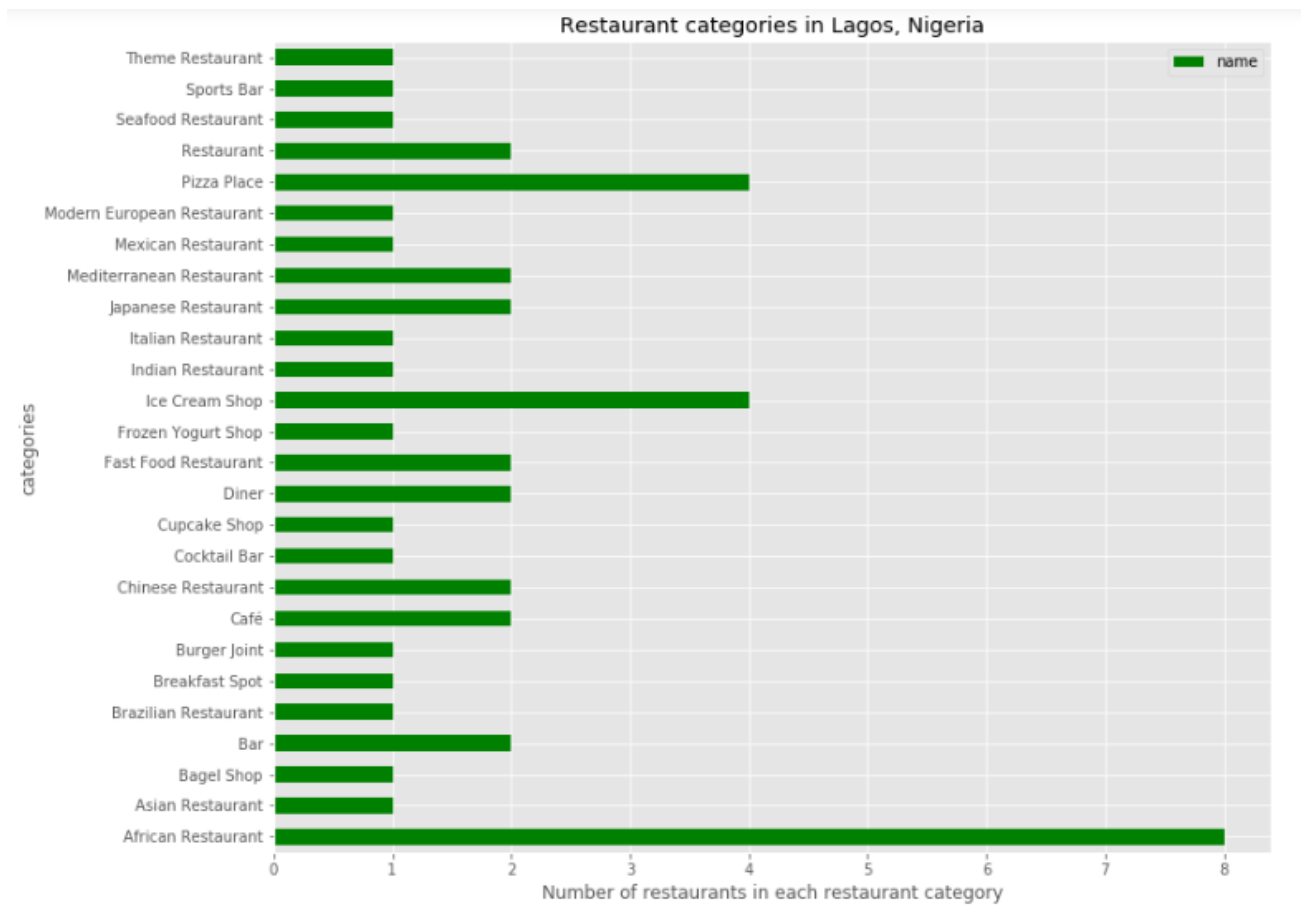


Figure 4: Restaurant Categories within the restaurant industry in Lagos, Nigeria

4.3 Istanbul, Turkey

In Istanbul there were overall 49 venue categories within a 10 km radius including the restaurant industry. 17 venues represented the restaurant industry category. So 17 % of the venues belonged to the restaurant industry with competition from 48 other venue categories. The main competition was from the 'Waterfront' venue category at 13 followed by the 'Hotel', 'Park' and 'Concert Hall' venue categories (Figure 5).

Within the 17 venues representing the restaurant industry the highest representation was from 'Coffee Shop' category at 3 and the second highest was tied between 'Steakhouse', 'Restaurant', 'Seafood Restaurant' and 'Café' categories at 2 each. Furthermore there is competition from venue categories representing local cuisine: such as 'Doner Restaurant' and 'Cajun / Creole Restaurant' at 1 each (Figure 6).

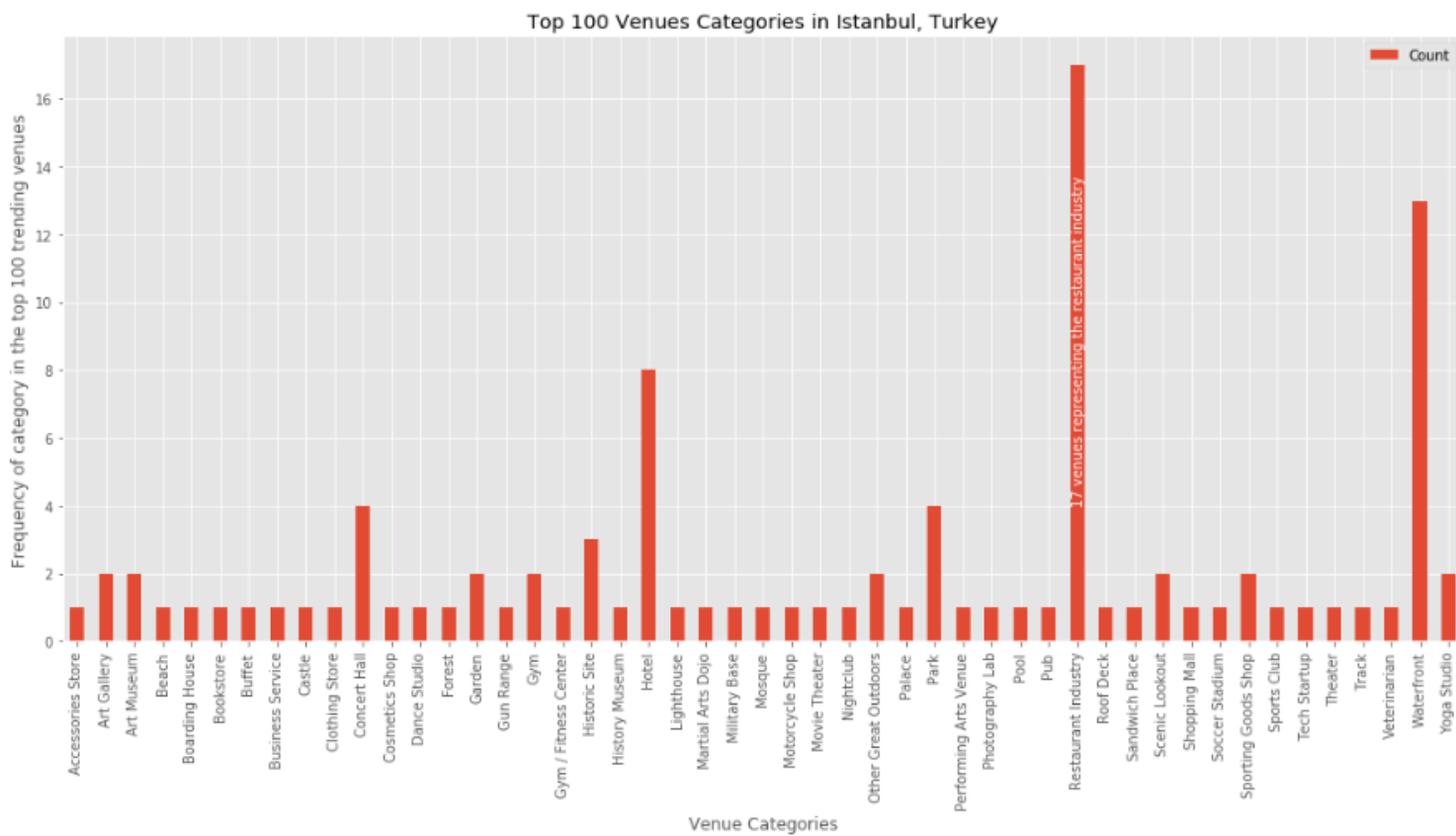


Figure 5: Venue Categories in Istanbul, Turkey

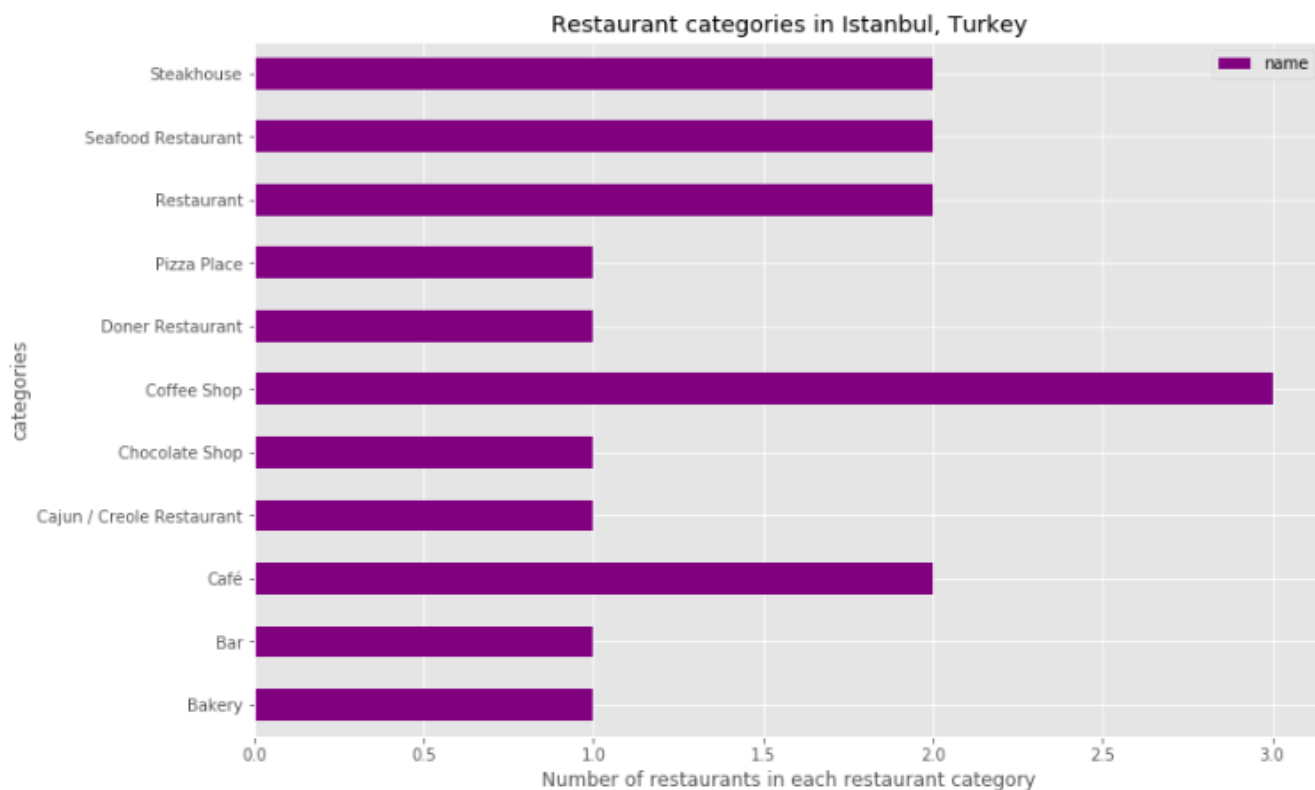


Figure 6: Restaurant Categories within the restaurant industry in Istanbul, Turkey

4.4 Sydney, Australia

In Sydney there were overall 32 venue categories within a 10 km radius including the restaurant industry. 42 venues represented the restaurant industry category. So 42 % of the venues belonged to the restaurant industry with competition from 31 other venue categories. The main competition was from the 'Park' venue category followed by the 'Scenic Lookout', 'Theatre' and 'Pool' venue categories (Figure 7).

Within the 42 venues representing the restaurant industry the highest representation was from 'Café' and 'Coffee Shop' category at 10 and 5 respectively followed by 'Bakery' and 'Cocktail Bar' at 4. Furthermore there is competition from venue categories such as 'Sandwich Place', 'Pub', 'Pizza Place', 'Thai Restaurant' and 'Bar' at 2 each (Figure 8).

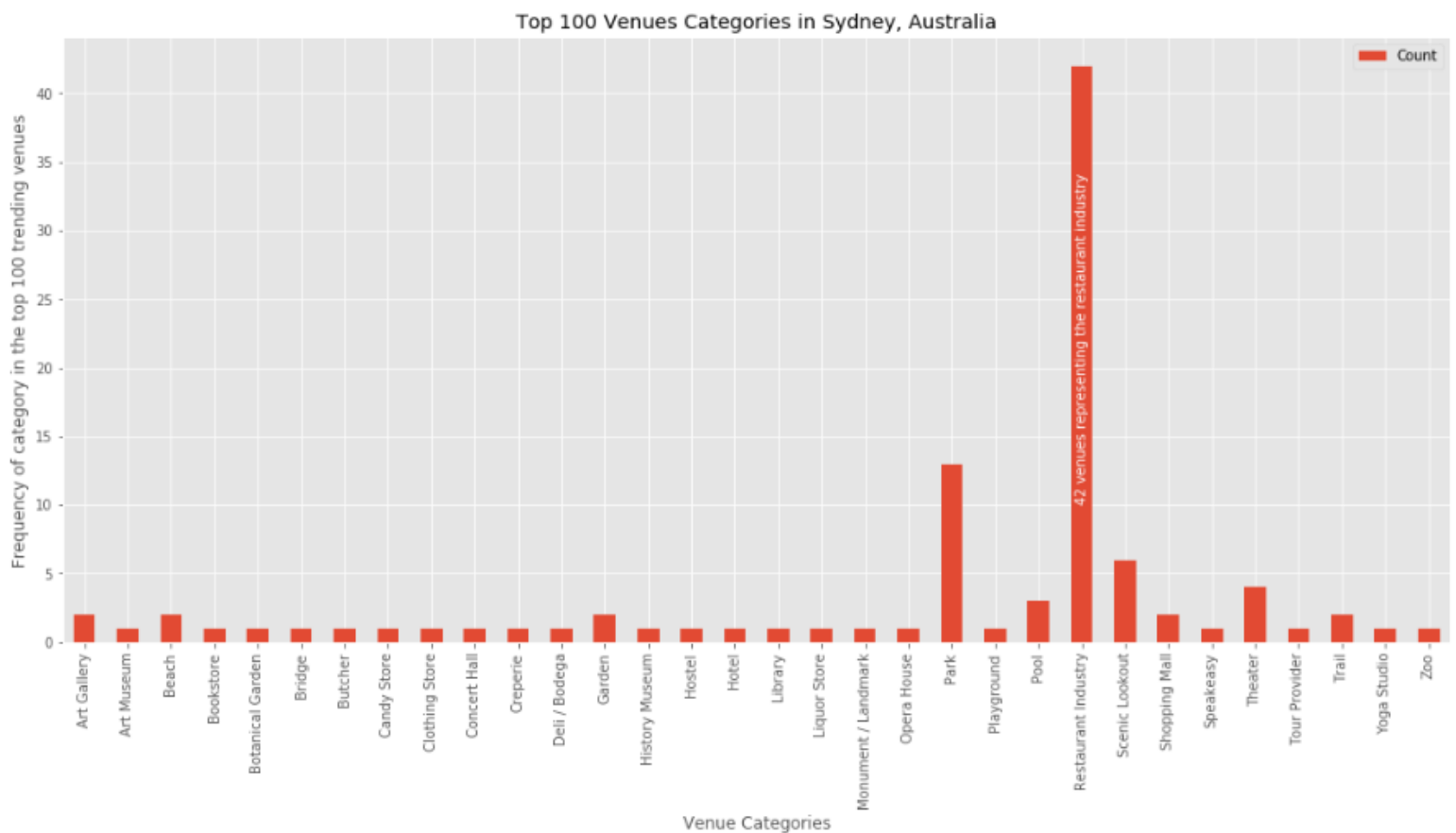


Figure 5: Venue Categories in Sydney, Australia



Figure 8: Restaurant Categories within the restaurant industry in Sydney, Australia

4.5 New York, United States of America

In New York there were overall 38 venue categories within a 10 km radius including the restaurant industry. 30 venues represented the restaurant industry category. So 30 % of the venues belonged to the restaurant industry with competition from 37 other venue categories. The main competition was from the 'Park' venue category followed by the 'Bookstore', 'Movie Theatre' and 'Scenic Lookout', venue categories (Figure 9).

Within the 30 venues representing the restaurant industry the highest representation was from 'Gourmet Shop' category at 5 followed by 'Bakery', 'Italian Restaurant' and 'Ice Cream Shop' at 4, 3 and 2 respectively. The rest of the competition is from a variety of venue categories representing general categories such as 'Pizza Place', such as 'Sandwich Place', 'Pub', 'Pizza Place', 'Beer Bar' as well as categories representing local delicacies such as 'Hot Dog joint' all at 1 each (Figure 10).

Top 100 Venues Categories in New York, United States of America

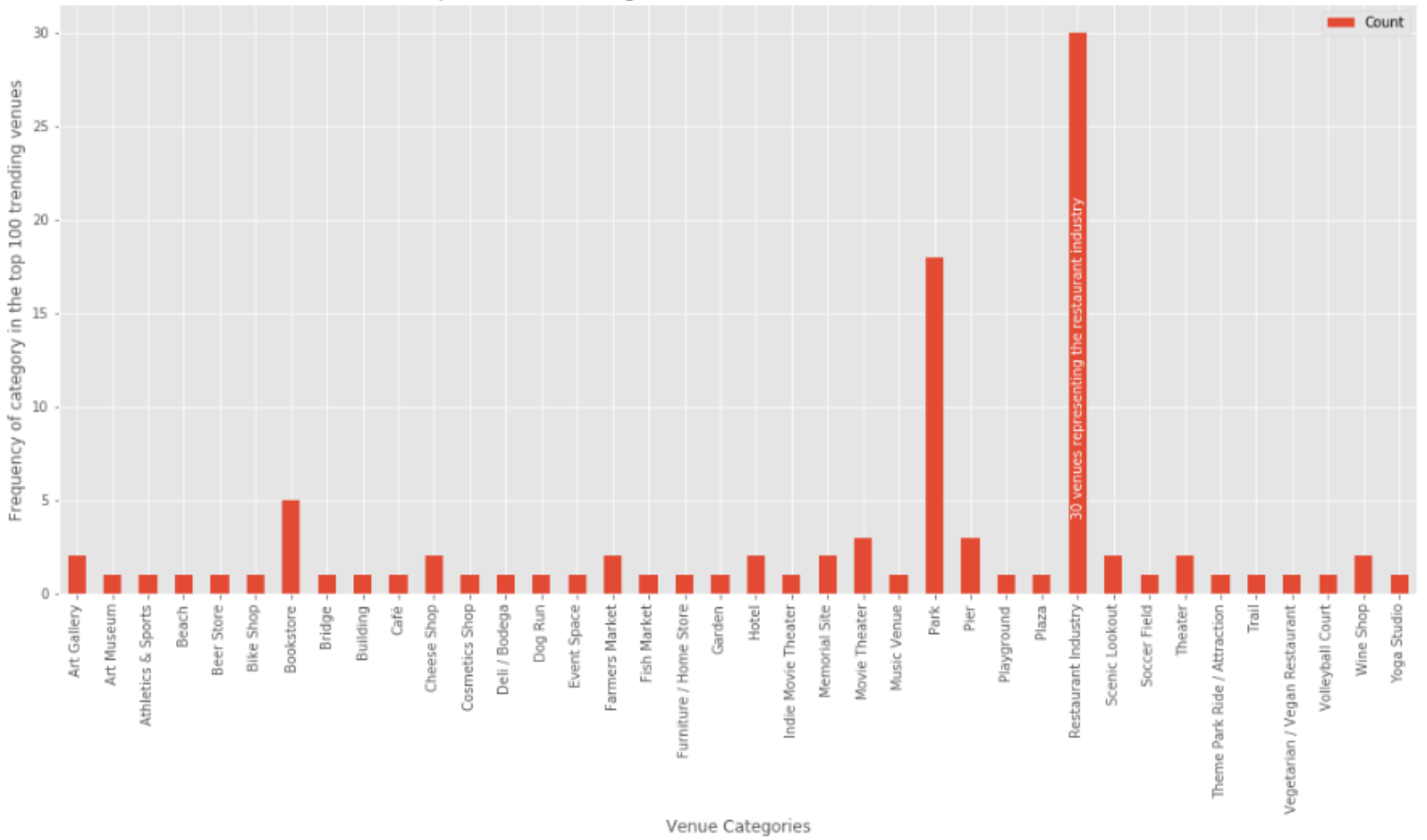


Figure 9: Venue Categories in New York, United States of America

Restaurant categories in New York, United States of America

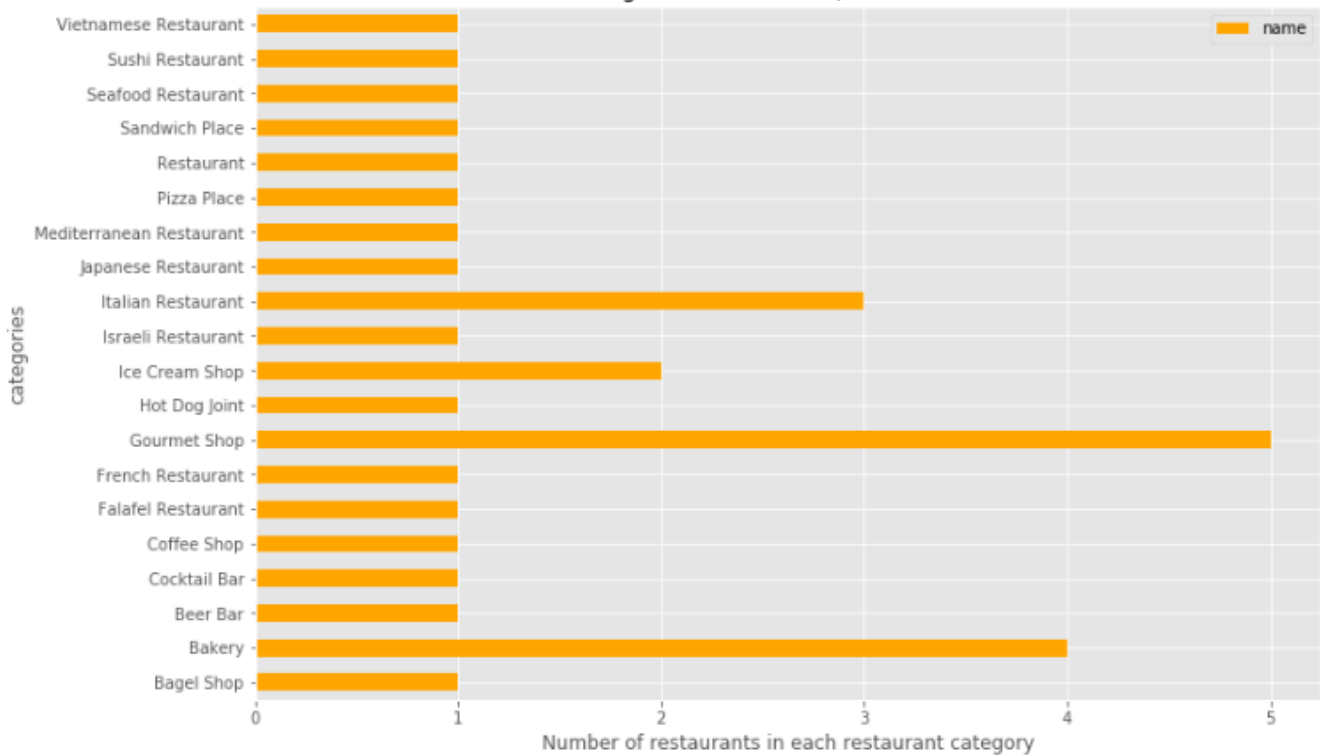


Figure 10: Restaurant Categories within the restaurant industry in New York, United States of America

4.6 Sao Paulo, Brazil

In Sao Paulo there were overall 36 venue categories within a 10 km radius including the restaurant industry. 37 venues represented the restaurant industry category. So 37 % of the venues belonged to the restaurant industry with competition from 35 other venue categories. The main competition was from the 'Art Museum' and 'Fruit & Vegetable Store' venue categories followed by the 'Park', 'Pet Store' and 'Theatre' venue categories (Figure 11).

Within the 37 venues representing the restaurant industry the highest representation was from 'Ice Cream Shop' category at 5 followed by 'Italian Restaurant' at 4 and 'Bakery', 'Steakhouse', 'Pizza Place' at 3. The rest of the competition is from a variety of venue categories representing general categories such as 'Burger joint', 'Bar' and 'Restaurant' as well as categories representing local delicacies such as 'Brazilian Restaurant' all at 2 each (Figure 12).

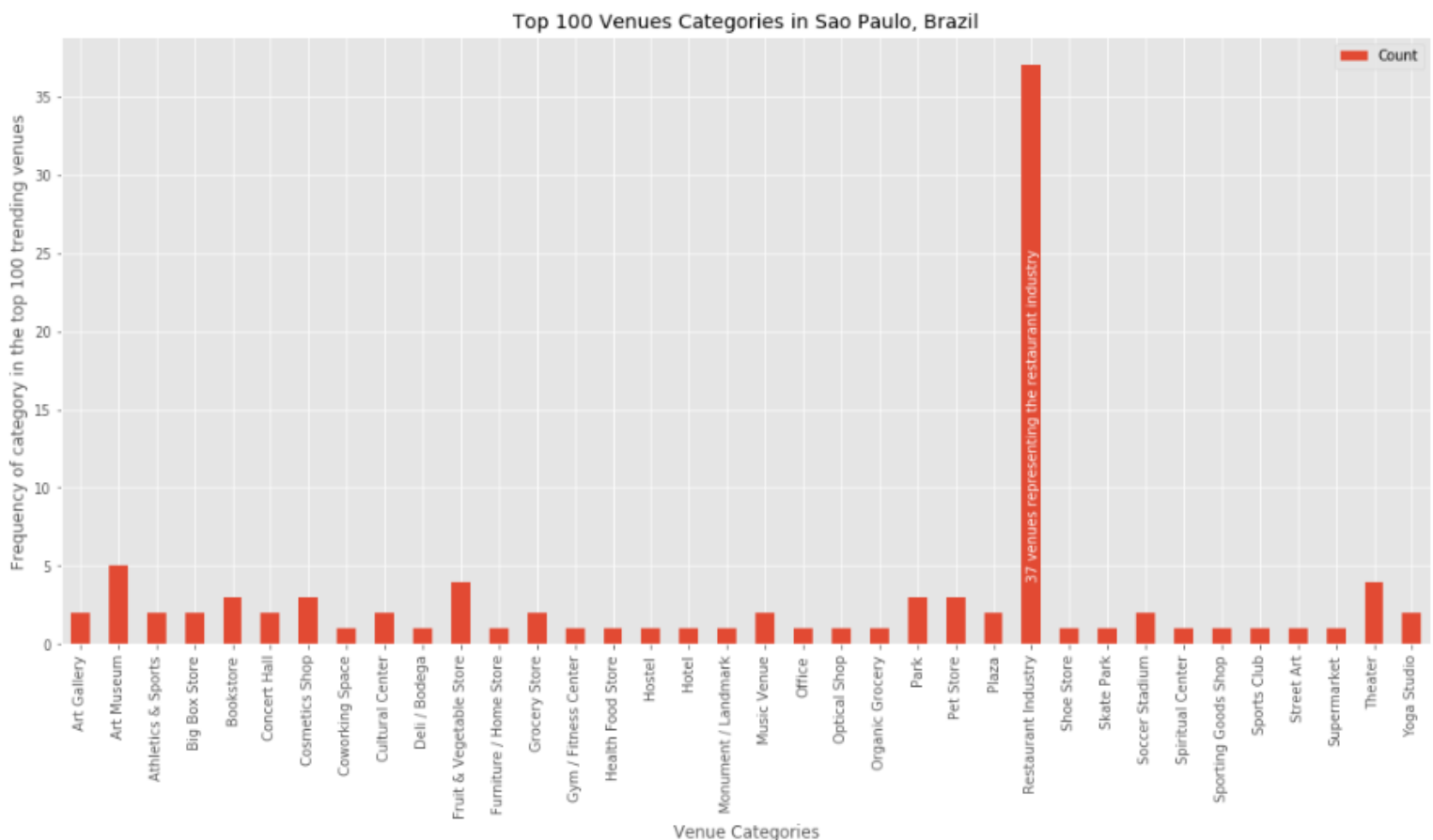


Figure 11: Venue Categories in Sao Paulo, Brazil

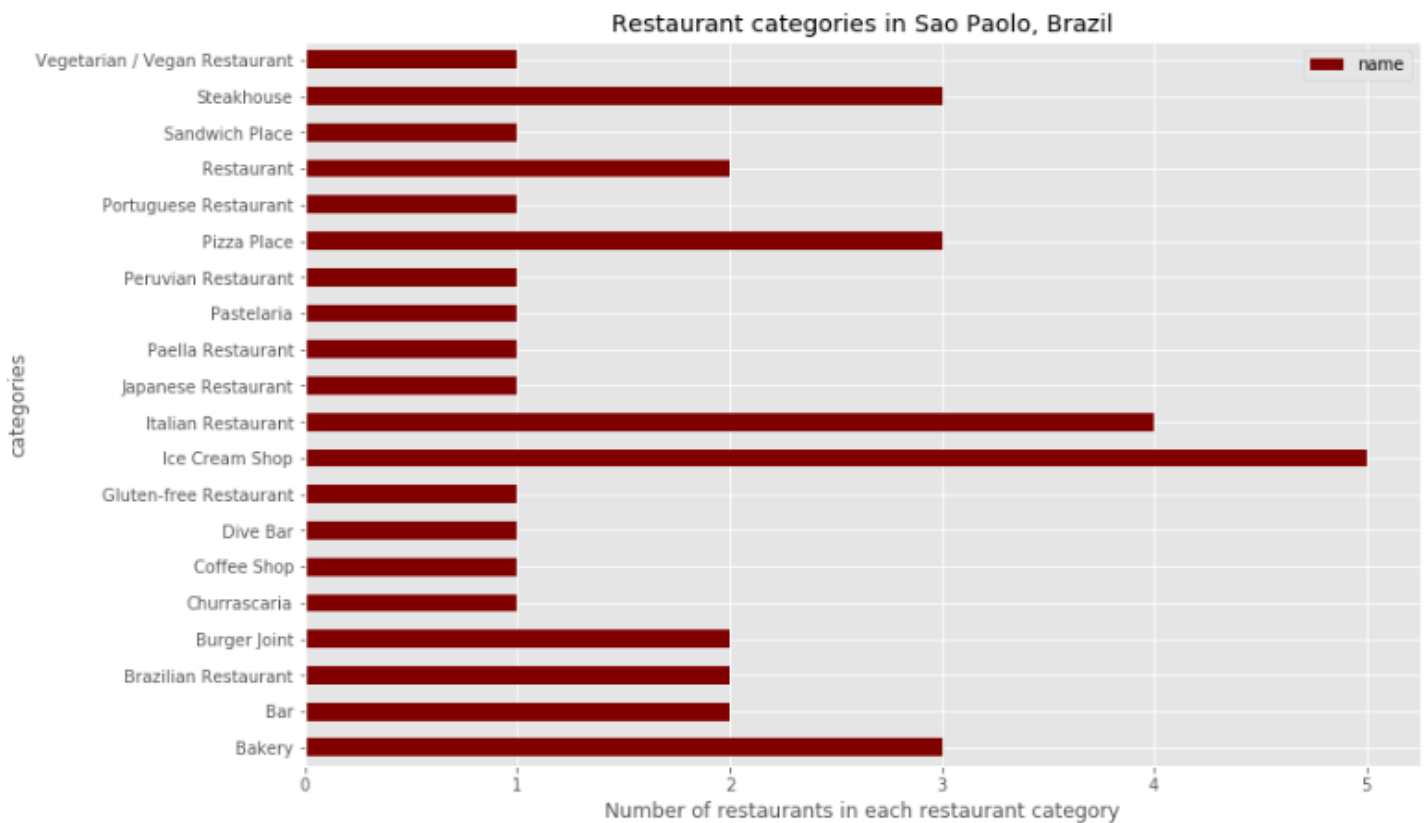


Figure 12: Restaurant Categories within the restaurant industry in Sao Paulo, Brazil

4.7 Restaurant venue frequency

If we append the data frames with all the restaurant venues present and across all the aforementioned cities the top most 5 venues with a highest numbers across all the six cities are Café , Coffee Shop, Ice Cream Shop, Bakery and Pizza Place (Figure 13)

```
frequency = venues.categories.value_counts()
frequency.head()
```

Café	15
Coffee Shop	15
Ice Cream Shop	14
Bakery	13
Pizza Place	12

Name: categories, dtype: int64

Figure 13: Top 5 restaurant venue categories across the six cities

If we go on to merge the data frames for all six cities based on venue categories based on the restaurant venue categories, we the restaurant venue common to all six cities. In my analysis the only common restaurant venue category across all six cities was 'Pizza Place'

4.8 Incorporating population data

The population interaction with the restaurant industry for each city is depicted in the following bubble plot (Figure 14):

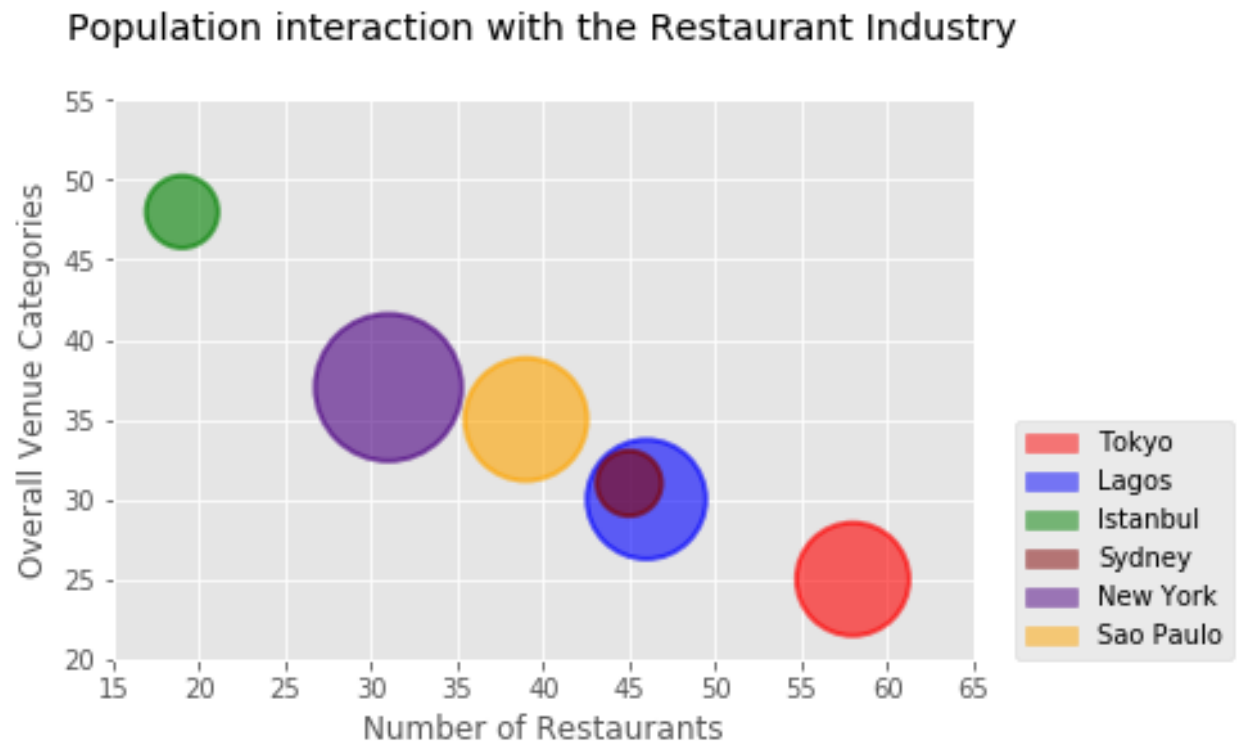


Figure 14: Correlation of venue categories, Restaurant industry and city population for all six cities

Figure 14 indicates that the highest population is in New York city followed by Sao Paulo, Lagos, Tokyo, Istanbul and finally Sydney respectively. The population in this case is per 1000 people. Therefore the bubble size does not directly indicate the population but is a relative indication within a 10 km radius. Istanbul has the most variety in terms of the overall venue categories present including the restaurant industry. The Istanbul population is not the highest but is larger than Sydney. The number of restaurant industry venues is the lowest for Istanbul as well. New York has the second highest variety in terms of overall venue categories and is highest in terms of population. It also has the second least amount of restaurant categories. Sao Paulo has the third most variety in overall venue categories and has the third highest population. It also has third least amount of restaurant categories. Sydney and Lagos fall in the same ball park of overall

venue categories and restaurant categories. However, there is a massive difference in population. Lagos has the third highest population whereas Sydney has the least in the six cities being compared. Tokyo has the highest amount of restaurant venues and the lowest overall venue categories along with the fourth highest population.

5 Discussion

5.1 Correlation

Three factors have been compared in the final part of the analysis including the number of overall venues categories in each of the six cities, number of restaurant venues and the total population. If we consider the first two factors from Figure 14, it is clear that there is an inverse relationship between the number of overall venue categories and restaurant venues. It is not surprising since I have kept a constant area (of 10 km radius) while deriving numbers for both these variables. However we understand the dominance of the restaurant industry in each city. Highest dominance is within Tokyo that has the highest diversity in restaurant venues (59% venues represent restaurant) and with the least competition from other venue categories potential indicating minimal competition from other industries (Figure 1). This dominance can potentially indicate the affinity of Tokyo citizens for the restaurant industry. However, for the stakeholders it is also important to remember more competition from within the restaurant industry that could hinder the business growth in such an area. In terms of cuisine the Tokyo data represents 12 venue categories that represent local cuisine which again shows that there is high competition within the restaurant industry in Tokyo along with preference for local cuisine.

The second highest dominance of the restaurant industry is in Lagos and Sydney with 47 % and 42% of the venues belonging to the restaurant industry respectively (Figure 3). The competition from other venue categories i.e. other potential industries is higher than Tokyo in both these cases. However the population of both cities is very different with the population of Lagos being almost 3.5 times larger than that of Sydney. This is important information for the stakeholders as it could highlight that the Lagos population is more prone to spend on eating and other outdoor activities compared to the Sydney population because both cities have similar amounts of venues within the restaurant industry and overall. Therefore this potentially highlights that Sydney will be a tough spot for a potential restaurant business owner to be successful in as the citizens there are not much prone to eat out compared to citizens in Lagos. Within Lagos itself the restaurant industry is dominated by local cuisine (Figure 4) which is important to understand as it further highlights the eating habits of the citizens there. On the other hand the Sydney restaurants are dominated by Cafés. The local cuisine is not much dominant as there is stiff competition from general restaurants such as Pizza places and burger joints along with exotic restaurants such as Thai and Brazilian restaurants. Therefore although it can be tough for a restaurant owner to succeed in Sydney, the chances increase if their restaurant adds to the diversity of the industry within the city.

In the case of New York and Sao Paulo the restaurant industry further reduces its dominance on the city. Sao Paulo has less competition from other venue categories i.e. potential other industries compared to New York but again faces stiff competition from within the restaurant industries. It also has plenty of south American cuisine themed restaurants with categories such as 'Brazilian Restaurant', 'Peruvian Restaurant', 'Churrascaria', 'Paella Restaurant' etc present in the city (Figure 12). Therefore, the Sao Paulo citizens could have an affinity for South American cuisine which can be an important insight for a stakeholder. Contrarily, in New York the restaurant industry has heavy competition with overall second highest amount of venue categories i.e. potential competition from other industries. However New York also has the highest population of all the cities being considered for this analysis (Figure 14). Therefore even though the competition from other industries is a lot higher compared to a city like Sydney, it will still be easier for stakeholders to get more people to invest i.e. potential customers in New York. Within the restaurant venues, the venues are diverse from general categories like 'Pizza place', 'Coffee Shop' etc, to more exotic categories like 'Israeli Restaurant', 'Japanese Restaurant' etc. to local delicacies like 'Hot Dog joint', suggesting a very diverse palate for the New York citizens.

Lastly, Istanbul has the highest overall venue categories at 49. Therefore the highest competition from non-restaurant industries is here. It is also important to understand that Istanbul has the second lowest population amongst the six cities only more than Sydney. It also has the lowest amount of restaurant venues (17) with strong competition from venue categories such as Waterfronts (Figure 5). This suggests that the eating habits of people in the city do not encourage them from going out to eat. Therefore it can potentially be very difficult for a restaurant owner to set up a successful business in that area. Within the restaurant industry, the most dominating venues are general category restaurants such as 'Coffee Shop', 'Pizza Place' etc with few venues representing local cuisine such as 'Doner Restaurant' and 'Cajun/ Creole Restaurant' (Figure 6). Even though competition from within the industry is lower as compared to Tokyo, Lagos and other cities, the Istanbul citizens potentially have more affinity towards aforementioned general cuisines compared to local cuisines which gives further insight into the eating habits of people.

5.2 Limitations and future work

This preliminary analysis only provides initial insight into the presence of the restaurant industry across major cities in the world and how the respective populations interact with it. There are many more factors that need to be taken into account that will require more detailed analysis of the data. For example a major factor overlooked in this scenario is the economics of these cities and how citizen income affects the eating habits of the population. This can provide more insight in the subject matter.

Another limitation is the lack of understanding of the population distribution within the city. To make this analysis more robust and meaningful, a better understanding of the population density in respective areas of the city will be very beneficial. This will allow us to constrain the eating habits of the population in a more specific way. Indeed, the selected cities do not reflect

the eating habits of the entire continent. Therefore this preliminary process can be applied to furthermore cities so as to gather more data frames and reflect broader eating habits that potentially impact the restaurant industry.

6 Conclusion

In this study I conducted preliminary analysis of the interaction of population with the restaurant industry in major cities in the world across all continents. Using Foursquare API and population statistics I created and edited data frames that depicted the number of restaurant venues and overall venues in each city as well as the potential population within a 10 km radius. For the relevant stake holders this study suggests the following eating habits in the respective cities:

- (a) High competition within the restaurant industry in Tokyo along with preference for local cuisine.
- (b) Similar (second highest) dominance of the restaurant industry is in Lagos and Sydney. However, the vast population difference indicates that Lagos population is more prone to spend on eating and outer outdoor activities compared to the Sydney population.
- (c) Sao Paulo citizens could have an affinity for South American cuisine as indicated by the presence of a wide variety South American themed restaurants.
- (d) New York has the highest population of all the cities. Therefore even though the competition from other industries is higher compared to a city like Sydney, it will still be easier for stakeholders to get potential customers in New York. New York citizens have a very diverse palate as well.
- (e) Eating habits of people in Istanbul do not encourage them from going out to eat. Therefore it can potentially be very difficult for a restaurant owner to set up a successful business in that area.
- (f) The most common eating habit and therefore the maximum potential for success across all continents is restaurant venue 'Pizza Place'.
- (g) The most frequent restaurant venue across all six cities is the 'Ice Cream Shop'

References

Tokyo population density:

<https://www.metro.tokyo.lg.jp/ENGLISH/ABOUT/HISTORY/history03.htm>

Lagos population density:

<https://worldpopulationreview.com/world-cities/lagos-population/>

Istanbul population density:

<https://worldpopulationreview.com/world-cities/istanbul-population/>

Sydney population density:

https://en.wikipedia.org/wiki/Demographics_of_Sydney

New York population density:

https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population_density

6. Sao Paulo population density:

<https://worldpopulationreview.com/world-cities/sao-paulo-population/>