CLUSTERING OF BUSIEST INTERNATIONAL AIRPORTS IN THE WORLD

By Parikshit Verma

Email: parikshit.iitvaranasi@gmail.com Linkedin GitHub

Contents

List of figures

1. Introduction & Background

It has always been the dream of mankind from time immemorial to fly in the sky like the birds do. Umpteen number of unsuccessful attempts had been made by humanity in this quest upto the very beginning of 20th century. Finally in 1903 two American brothers made history when they flew over North Carolina sky first time in the history of mankind.

*Figure 1: First flight of the Wright Flyer, December 17, 1903. (*Pic courtesy Wikipedia)

Up until this date fascination of flying in the sky has not died but only has increased. The fact that is clearly established by the World Air Transport Statistics, 2018 by IATA. According to report, since 2004 the revenue growth of the airlines has outpaced world GDP continuously. Since the people are flying too frequently airports are not just the places to catch flight anymore. Airports have become more of a micro city offering all types of facilities & services that one can imagine and need. These facilities and services not only to promote businesses but also are necessary as many a times one has to spend a considerable amount of time on airport due to procedural requirements, non-availability of direct flight to intended destination etc.

Foursquare is an online platform that provides plethora of services including details about places of interest within a specified distance of a geographic location. As of the date of writing this article there were 941 different categories of places of interest covered by Foursquare. This huge information when combined with the airports data may reveal useful information for the stakeholders.

1.1. Target audience

In view of above it is quite useful from a commuter's point of view to have an idea that which airports in the world are likely to have the facilities he or she requires while at the airport. This may help people to choose from various flights options available to them. On the other side it may help business owners and multinational corporations to decide whether they should choose one airport or the other for their next store.

1.2. Problem statement

To classify busiest international airports of the worlds on the basis of certain set of establishments/ facilities/ services/ places of interest using foursquare APIs and any other relevant data.

2. Data Description

Data required for this project comes from different sources. Since our aim is to group airports on the basis of different categories of places of interest like different types of restaurants, services, places to stay etc. this shall constitute major part of our data and shall be obtained using foursquare API. Other data like list of busiest airports and their location shall be obtained by scraping Wikipedia pages and other relevant reports if necessary.

2.1. FourSquare API

Since one of the busiest international airports are quite huge hence to search for relevant places of interest a radius of 2km has been chosen. Four square API shall be used to get the number of places of interest of a particular category for each airport. But before that we required a list of relevant target categories. FourSquare categories page was scraped to accomplish this task giving us total 941 categories of places available in FourSquare data bases. Out of these 941 categories 442 categories are deemed to be relevant for our business problem and have been grouped into following 31 groups :-

1 Eateries Food BBQ Joint Bistro Breakfast Spot Diner Food Court Food Stand Food Truck Friterie Halal Restaurant Molecular Gastronomy Restaurant Restaurant Salad Place Seafood Restaurant Theme Restaurant Vegetarian / Vegan Restaurant Food & Drink Shop Airport Food Court
2 Asian_restaurant Afghan Restaurant Asian Restaurant Korean Restaurant Bossam/Jokbal Restaurant Bunsik Restaurant Gukbap Restaurant Janguh Restaurant Samgyetang Restaurant Mongolian Restaurant Tibetan Restaurant Comfort Food Restaurant Gluten-free Restaurant
3 African_restaurant African Restaurant Ethiopian Restaurant Moroccan Restaurant
4 NAmerican_restaurant(*North American*) New American Restaurant American Restaurant Cajun / Creole Restaurant Caribbean Restaurant Cuban Restaurant Fried Chicken Joint Hawaiian Restaurant Poke Place Mac & Cheese Joint Mexican Restaurant Botanero Burrito Place Taco Place Tex-Mex Restaurant Yucatecan Restaurant Poutine Place Southern / Soul Food Restaurant Steakhouse Wings Joint
5 SAmerican_restaurant(*South American*) Latin American Restaurant Arepa Restaurant Empanada Restaurant Salvadoran Restaurant South American Restaurant Argentinian Restaurant Brazilian Restaurant Acai House Baiano Restaurant Central Brazilian Restaurant Churrascaria Empada House Goiano Restaurant Mineiro Restaurant Northeastern Brazilian Restaurant Northern Brazilian Restaurant Pastelaria Southeastern Brazilian Restaurant Southern Brazilian Restaurant Tapiocaria Colombian Restaurant Peruvian Restaurant Venezuelan Restaurant
6 Australian_restaurant Australian Restaurant
7 European_restaurant Austrian Restaurant Bagel Shop Belgian Restaurant Czech Restaurant Dutch Restaurant English Restaurant Fondue Restaurant German Restaurant Bavarian Restaurant

Bratwurst Joint Currywurst Joint Franconian Restaurant German Pop-Up Restaurant Palatine Restaurant Rhenisch Restaurant Schnitzel Restaurant Silesian Restaurant Swabian Restaurant Hungarian Restaurant Modern European Restaurant Polish Restaurant Portuguese Restaurant Scandinavian Restaurant Scottish Restaurant Slovak Restaurant Spanish Restaurant Paella Restaurant Tapas Restaurant Swiss Restaurant

8 IndianSub_restaurant(*Indian sub-continent*) Himalayan Restaurant Bangladeshi Restaurant Indian Restaurant Andhra Restaurant Awadhi Restaurant Bengali Restaurant Chettinad Restaurant Dhaba Dosa Place Goan Restaurant Gujarati Restaurant Hyderabadi Restaurant Indian Chinese Restaurant Irani Cafe Jain Restaurant Karnataka Restaurant Kerala Restaurant Maharashtrian Restaurant Mughlai Restaurant Multicuisine Indian Restaurant North Indian Restaurant Northeast Indian Restaurant Parsi Restaurant Punjabi Restaurant Rajasthani Restaurant South Indian Restaurant Udupi Restaurant Pakistani Restaurant Sri Lankan Restaurant

9 Chinese restaurant Chinese Restaurant Anhui Restaurant Beijing Restaurant Cantonese Restaurant Cha Chaan Teng Chinese Aristocrat Restaurant Chinese Breakfast Place Dim Sum Restaurant Dongbei Restaurant Fujian Restaurant Guizhou Restaurant Hainan Restaurant Hakka Restaurant Henan Restaurant Hong Kong Restaurant Huaiyang Restaurant Hubei Restaurant Hunan Restaurant Imperial Restaurant Jiangsu Restaurant Jiangxi Restaurant Macanese Restaurant Manchu Restaurant Peking Duck Restaurant Shaanxi Restaurant Shandong Restaurant Shanghai Restaurant Shanxi Restaurant Szechuan Restaurant Taiwanese Restaurant Tianjin Restaurant Xinjiang Restaurant Yunnan Restaurant Zhejiang Restaurant

10 MiddleEastern_restaurant Falafel Restaurant Jewish Restaurant Kosher Restaurant Kebab Restaurant Middle Eastern Restaurant Egyptian Restaurant Iraqi Restaurant Israeli Restaurant Kurdish Restaurant Lebanese Restaurant Persian Restaurant Ash and Haleem Place Dizi Place Gilaki Restaurant Jegaraki Tabbakhi Shawarma Place Syrian Restaurant Yemeni Restaurant

11 Japanese_restaurant Japanese Restaurant Donburi Restaurant Japanese Curry Restaurant Kaiseki Restaurant Kushikatsu Restaurant Monjayaki Restaurant Nabe Restaurant Okonomiyaki Restaurant Ramen Restaurant Shabu-Shabu Restaurant Soba Restaurant Sukiyaki Restaurant Sushi Restaurant Takoyaki Place Tempura Restaurant Tonkatsu Restaurant Udon Restaurant Unagi Restaurant Wagashi Place Yakitori Restaurant Yoshoku Restaurant Noodle House

12 SouthEastAsia_restaurant Burmese Restaurant Cambodian Restaurant Filipino Restaurant Hotpot Restaurant Indonesian Restaurant Indonesian Restaurant Balinese Restaurant Betawinese Restaurant Indonesian Meatball Place Javanese Restaurant Manadonese Restaurant Padangnese Restaurant Sundanese Restaurant Malay Restaurant Mamak Restaurant Satay Restaurant Thai Restaurant Som Tum Restaurant Vietnamese Restaurant

13 French_restaurant Creperie French Restaurant Alsatian Restaurant Auvergne Restaurant Basque Restaurant Brasserie Breton Restaurant Burgundian Restaurant Catalan Restaurant Ch'ti Restaurant Corsican Restaurant Estaminet Labour Canteen Lyonese Bouchon Norman Restaurant Provençal Restaurant Savoyard Restaurant Southwestern French Restaurant

14 Italian_restaurant Italian Restaurant Abruzzo Restaurant Agriturismo Aosta Restaurant Basilicata Restaurant Calabria Restaurant Campanian Restaurant Emilia Restaurant Friuli Restaurant Ligurian Restaurant Lombard Restaurant Malga Marche Restaurant Molise Restaurant Piadineria Piedmontese Restaurant Puglia Restaurant Romagna Restaurant Roman Restaurant Sardinian Restaurant Sicilian Restaurant South Tyrolean Restaurant Trattoria/Osteria Trentino Restaurant Tuscan Restaurant Umbrian Restaurant Veneto Restaurant Pizza Place

15 Turkish_restaurant Turkish Restaurant Borek Place Cigkofte Place Doner Restaurant Gozleme Place Kofte Place Kokoreç Restaurant Kumpir Restaurant Kumru Restaurant Manti Place Meyhane Pide Place Pilavcı Söğüş Place Tantuni Restaurant Turkish Home Cooking Restaurant Çöp Şiş Place

16 Greek/Mediterranean_restaurant Greek Restaurant Bougatsa Shop Cretan Restaurant Fish Taverna Grilled Meat Restaurant Kafenio Magirio Meze Restaurant Modern Greek Restaurant Ouzeri Patsa Restaurant Souvlaki Shop Taverna Tsipouro Restaurant Mediterranean Restaurant

17 Russia/East Eaurope_restaurant Eastern European Restaurant Belarusian Restaurant Bosnian Restaurant Bulgarian Restaurant Romanian Restaurant Tatar Restaurant Russian Restaurant Blini House Pelmeni House Ukrainian Restaurant Varenyky restaurant West-Ukrainian Restaurant Caucasian Restaurant
18 Desserts/Bakery Bakery Dessert Shop Cupcake Shop Frozen Yogurt Shop Ice Cream Shop Pastry Shop Pie Shop Donut Shop Indian Sweet Shop Candy Store Chocolate Shop
19 Cafe Bubble Tea Shop Cafeteria Coffee Shop Tea Room Turkish Coffeehouse
20 Fast food/Snack Burger Joint Deli / Bodega Dumpling Restaurant Fast Food Restaurant Fish & Chips Shop Hot Dog Joint Sandwich Place Snack Place Soup Place
21 Pub/Bar/Brewries Gastropub Apple Wine Pub Irish Pub Bar Beer Bar Cocktail Bar Pub Whisky Bar Wine Bar Champagne Bar Brewery
22 Beverages/juices/shakes Juice Bar Beer Store Liquor Store Wine Shop
23 Waiting/Stay Lounge Airport Lounge Hotel Bed & Breakfast Boarding House Hostel Inn Motel Vacation Rental Travel Lounge
24 General_Shopping Shop & Service Baby Store Convenience Store Department Store Cheese Shop Gourmet Shop Grocery Store Leather Goods Store Music Store Toy / Game Store
25 Gifts/ Souvenirs Antique Shop Flower Shop Gift Shop Souvenir Shop Thrift / Vintage Store
26 Electrnics_stores Camera Store Electronics Store Mobile Phone Shop Mobility Store
27 Lifestyle Clothing Store Accessories Store Boutique Kids Store Lingerie Store Men's Store Shoe Store Women's Store Costume Shop Cosmetics Shop Fabric Shop Health & Beauty Service Jewelry Store Watch Shop Salon / Barbershop
28 Books Bookstore Comic Shop
29 Tour & Travel Tour Provider Travel Agency Bus Station Bus Line Bus Stop Metro Station Rental Car Location Taxi Stand Tourist Information Centre Train Station Train Transportation Service
30 Duty Free shops Duty-free Shop
31 Services ATM Currency Exchange Baggage Locker Bike Rental / Bike Share

Only getting the names of categories was only half the work done as for API calls we require code of category also arranged in a similar manner as the groups. For arranging names in group as mentioned, excel was used. Similarly, a corresponding sheet was populated using ‘Index’ and ‘Match’ function of excel to find the corresponding code for each member of the newly formed group. Both these excel sheets were loaded in two different data sets for further processing.

To get the number of places of interest corresponding to each group, every element of the group has been queried using ‘Explore’ function of FourSquare API for each of the airport. And the number so obtained is stored in separate excel sheets. Building this data for number of places of interest for each of the group member required 62255 API calls. However before making API calls we needed to have list of airports of our interest.

2.2. List of busiest international airports in the world

To find the list of busiest airports in the world two Wikipedia pages were scraped. First pages ‘List of cities by international visitors’ and second one ‘ List of international airports by country’.

List of cities by international visitor was the basis to decide which airports shall be part of our project and provided us with the data ‘Country’ and ‘City’.

List of international airports by country was then used to find the ‘Name’ and ‘IATA Code’ of airports of our interest. However, it was observed that as per wikipedia list, one airport may have multiple city names and hence same had to be resolved before comparing the same to list of cities by international visitors.

Finally the both the lists were merged to form one pandas dataframe and rearranged, resulting in a data frame containing ‘Country’, ‘City’, ‘Airport’ and ‘IATA Code’ for 155 airports around the world.

To the list of airports so obtained, latitude and longitude for each airport was added using Geopy library. For some of the airports geopy didn’t seem to work and same had to be done manually.

Note:- All the data available is for the year 2018.

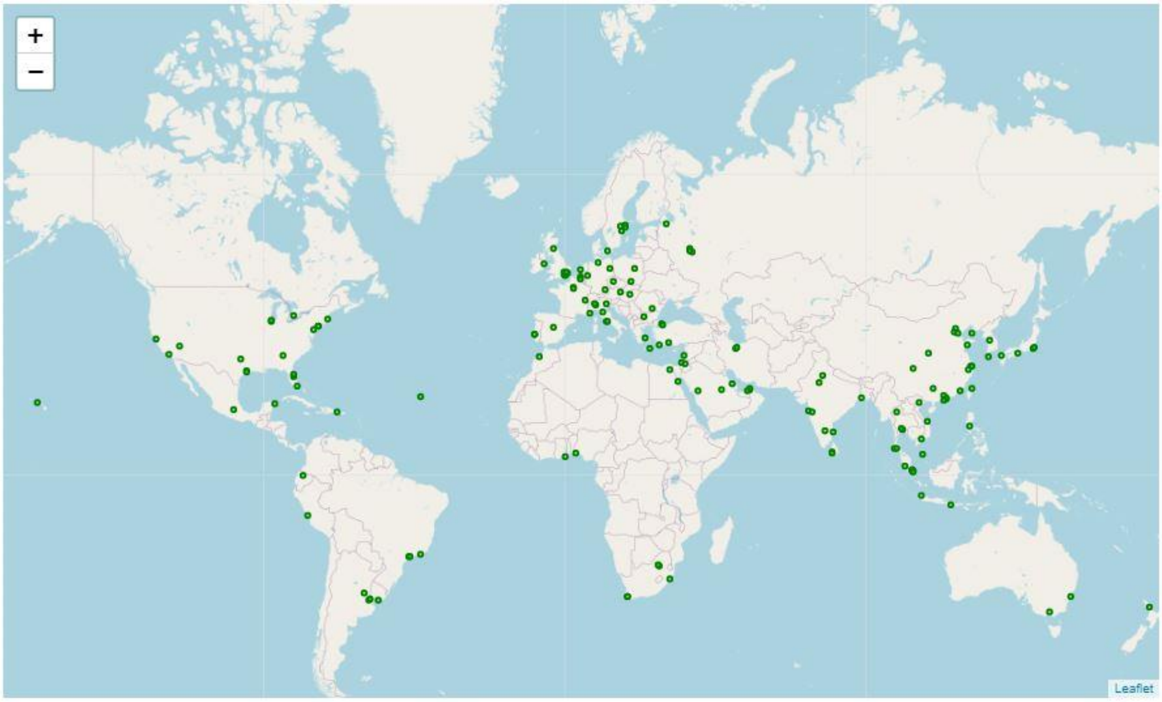


Figure 2: Busiest airports in world

2.3. Birds eye view of data

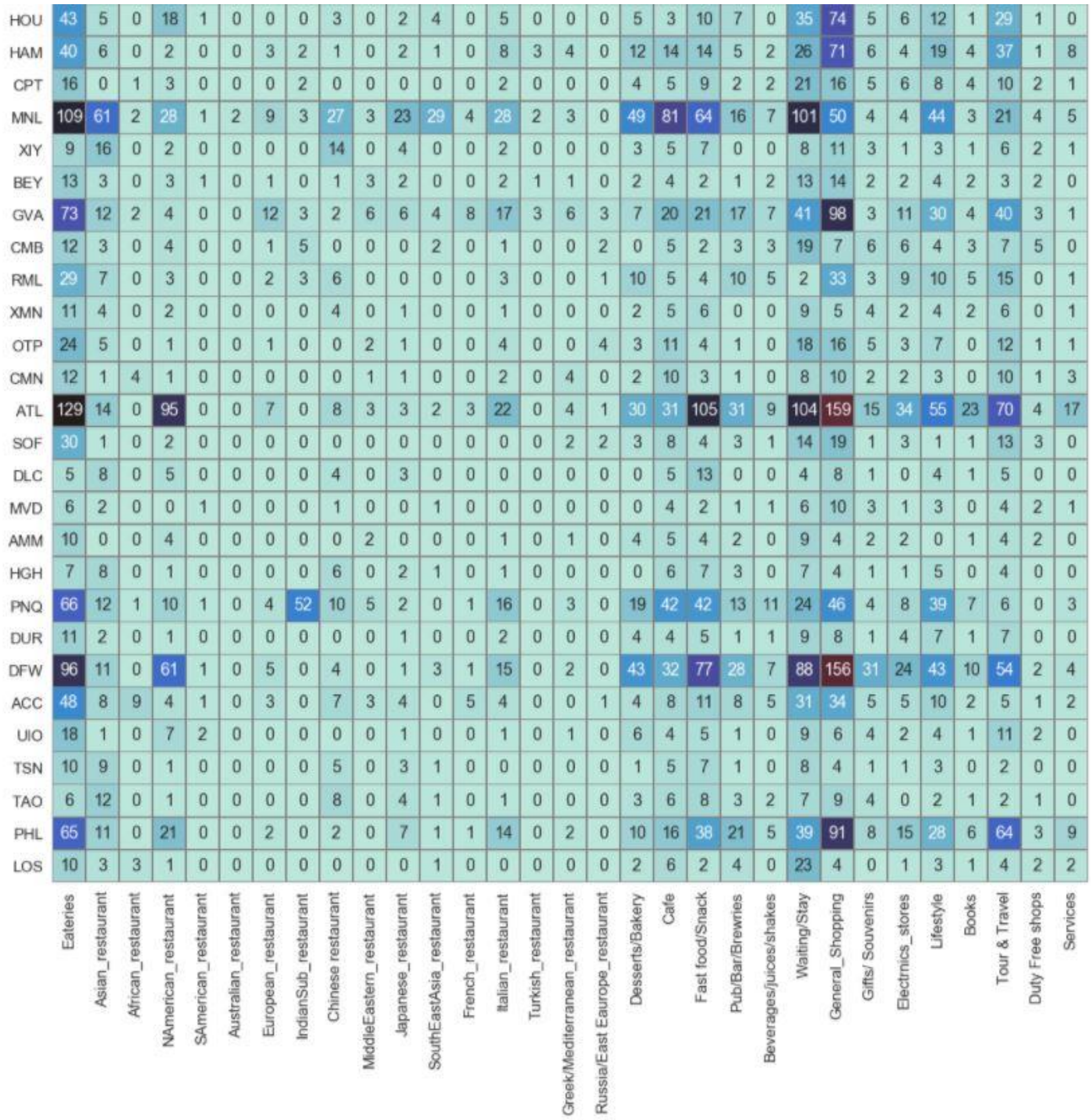
For our final data set ‘IATA Code’ column from section 2.2 Data was taken and corresponding to each airports all 31 categories of section 2.1 were added containing count of each category group within 2 km radius of the location (latitude and longitude) obtained using Geopy library. Additionally a column ‘distribution’ is added to the end of this dataframe to get an idea of distribution of facilities in terms of average distance from central point.

Our final data base comprised of ‘IATA Code’, 31 groups of categories and finally a distribution column for each of the 155 airports. And same was populated with integer values representing number of a places of that particular category located within 2km from the centre (latitude longitude location form Geopy) of airport.

3. Methodology

3.1. Data wrangling

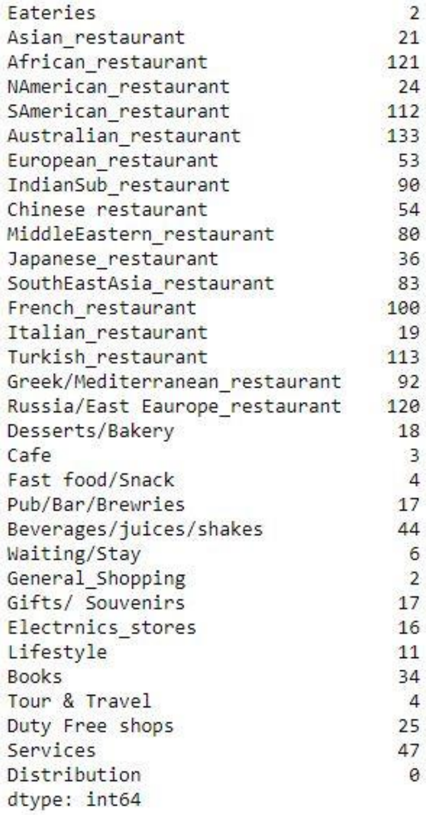
Since over data set is self-curated in a way that there are no null values hence we don’t have to deal with null value manipulation however there may be certain outliers in a sense that many airports may have very less facilities or many facilities may be limited to very small number of airports. These may be termed as outlier because at this point of time we are not sure whether to use spherical clustering algorithms like kMeans or density-based algorithms like DBSCAN etc. Additionally, all the values are count of places hence all the values are integers. Hence, we can get a better idea of dataset by heatmap.



	Eateries	Asian_restaurant	African_restaurant	NAmerican_restaurant	SAmerican_restaurant	Australian_restaurant	European_restaurant	IndianSub_restaurant	Chinese restaurant	MiddleEastern_restaurant	Japanese_restaurant	SouthEastAsia_restaurant	French_restaurant	Italian_restaurant	Turkish_restaurant	Greek/Mediterranean_restaurant	Russia/East Europe_restaurant	Desserts/Bakery	Cafe	Fast food/Snack	Pub/Bar/Brewries	Beverages/juices/shakes	Waiting/Stay	General_Shopping	Gifts/ Souvenirs	Electrnics_stores	Lifestyle	Books	Tour & Travel	Duty Free shops	Services
HOU	43	5	0	18	1	0	0	0	3	0	2	4	0	5	0	0	0	5	3	10	7	0	35	74	5	6	12	1	29	1	0
HAM	40	6	0	2	0	0	3	2	1	0	2	1	0	8	3	4	0	12	14	14	5	2	26	71	6	4	19	4	37	1	8
CPT	16	0	1	3	0	0	0	2	0	0	0	0	0	2	0	0	0	4	5	9	2	2	21	16	5	6	8	4	10	2	1
MNL	109	61	2	28	1	2	9	3	27	3	23	29	4	28	2	3	0	49	81	64	16	7	101	50	4	4	44	3	21	4	5
XIY	9	16	0	2	0	0	0	0	14	0	4	0	0	2	0	0	0	3	5	7	0	0	8	11	3	1	3	1	6	2	1
BEY	13	3	0	3	1	0	1	0	1	3	2	0	0	2	1	1	0	2	4	2	1	2	13	14	2	2	4	2	3	2	0
GVA	73	12	2	4	0	0	12	3	2	6	6	4	8	17	3	6	3	7	20	21	17	7	41	98	3	11	30	4	40	3	1
CMB	12	3	0	4	0	0	1	5	0	0	0	2	0	1	0	0	2	0	5	2	3	3	19	7	6	6	4	3	7	5	0
RML	29	7	0	3	0	0	2	3	6	0	0	0	0	3	0	0	1	10	5	4	10	5	2	33	3	9	10	5	15	0	1
XMN	11	4	0	2	0	0	0	0	4	0	1	0	0	1	0	0	0	2	5	6	0	0	9	5	4	2	4	2	6	0	1
OTP	24	5	0	1	0	0	1	0	0	2	1	0	0	4	0	0	4	3	11	4	1	0	18	16	5	3	7	0	12	1	1
CMN	12	1	4	1	0	0	0	0	0	1	1	0	0	2	0	4	0	2	10	3	1	0	8	10	2	2	3	0	10	1	3
ATL	129	14	0	95	0	0	7	0	8	3	3	2	3	22	0	4	1	30	31	105	31	9	104	159	15	34	55	23	70	4	17
SOF	30	1	0	2	0	0	0	0	0	0	0	0	0	0	0	2	2	3	8	4	3	1	14	19	1	3	1	1	13	3	0
DLC	5	8	0	5	0	0	0	0	4	0	3	0	0	0	0	0	0	0	5	13	0	0	4	8	1	0	4	1	5	0	0
MVD	6	2	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	4	2	1	1	6	10	3	1	3	0	4	2	1
AMM	10	0	0	4	0	0	0	0	0	2	0	0	0	1	0	1	0	4	5	4	2	0	9	4	2	2	0	1	4	2	0
HGH	7	8	0	1	0	0	0	0	6	0	2	1	0	1	0	0	0	0	6	7	3	0	7	4	1	1	5	0	4	0	0
PNQ	66	12	1	10	1	0	4	52	10	5	2	0	1	16	0	3	0	19	42	42	13	11	24	46	4	8	39	7	6	0	3
DUR	11	2	0	1	0	0	0	0	0	0	1	0	0	2	0	0	0	4	4	5	1	1	9	8	1	4	7	1	7	0	0
DFW	96	11	0	61	1	0	5	0	4	0	1	3	1	15	0	2	0	43	32	77	28	7	88	156	31	24	43	10	54	2	4
ACC	48	8	9	4	1	0	3	0	7	3	4	0	5	4	0	0	1	4	8	11	8	5	31	34	5	5	10	2	5	1	2
UIO	18	1	0	7	2	0	0	0	0	0	1	0	0	1	0	1	0	6	4	5	1	0	9	6	4	2	4	1	11	2	0
TSN	10	9	0	1	0	0	0	0	5	0	3	1	0	0	0	0	0	1	5	7	1	0	8	4	1	1	3	0	2	0	0
TAO	6	12	0	1	0	0	0	0	8	0	4	1	0	1	0	0	0	3	6	8	3	2	7	9	4	0	2	1	2	1	0
PHL	65	11	0	21	0	0	2	0	2	0	7	1	1	14	0	2	0	10	16	38	21	5	39	91	8	15	28	6	64	3	9
LOS	10	3	3	1	0	0	0	0	0	0	0	1	0	0	0	0	0	2	6	2	4	0	23	4	0	1	3	1	4	2	2

Figure 3: Part of Heat Map

Looking at the heatmap it is clear that our apprehensions were correct as there are many groups as well as many airports which are having very few non zero values. To quantify the same for groups first as in case of group of categories we can combine two groups to generate new group. This would enable us to categorise without loss of information.



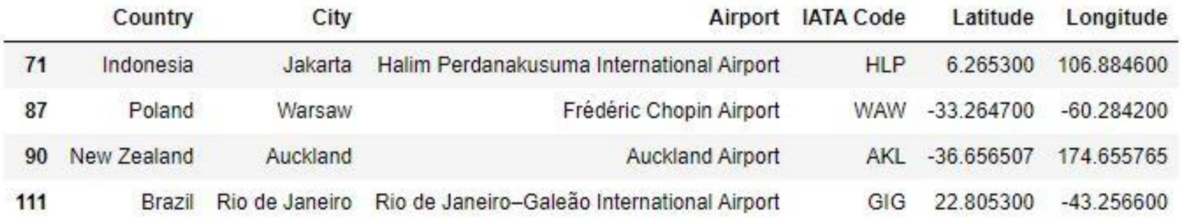
```
Eateries                            2
Asian_restaurant                   21
African_restaurant                121
NAmerican_restaurant               24
SAmerican_restaurant              112
Australian_restaurant             133
European_restaurant                53
IndianSub_restaurant               90
Chinese restaurant                 54
MiddleEastern_restaurant           80
Japanese_restaurant                36
SouthEastAsia_restaurant           83
French_restaurant                 100
Italian_restaurant                 19
Turkish_restaurant                113
Greek/Mediterranean_restaurant     92
Russia/East Eaurope_restaurant    120
Desserts/Bakery                    18
Cafe                                3
Fast food/Snack                     4
Pub/Bar/Brewries                   17
Beverages/juices/shakes            44
Waiting/Stay                        6
General_Shopping                    2
Gifts/ Souvenirs                   17
Electrnics_stores                  16
Lifestyle                          11
Books                              34
Tour & Travel                       4
Duty Free shops                    25
Services                           47
Distribution                        0
dtype: int64
```

Based on above values following categories were merged to arrive at final categories:-

- African & Australian restaurant merged into Eateries
- North american and South american restaurants merged into new category American restaurants
- French_restaurant, Turkish_restaurant, Greek/Mediterranean_restaurant & Russia/East Eaurope_restaurant merged into European restaurants
- Chinese and South east asian restaurant merged into new category Chinese/ SEAsian_restaurant
- Indian and Middle eastern restaurant merged to Asian restaurant

This reduces total number of group of places from 31 to 23.

Same can be true for each airport also, hence it is obvious to count and remove airports that have excess zero values. If more than 50% of groups for a particular airport have '0' values then those shall be dropped. We Find only following 4 airports out of 155 fall in such criteria hence removed from further processing.



	Country	City	Airport	IATA Code	Latitude	Longitude
71	Indonesia	Jakarta	Halim Perdanakusuma International Airport	HLP	6.265300	106.884600
87	Poland	Warsaw	Frédéric Chopin Airport	WAW	-33.264700	-60.284200
90	New Zealand	Auckland	Auckland Airport	AKL	-36.656507	174.655765
111	Brazil	Rio de Janeiro	Rio de Janeiro–Galeão International Airport	GIG	22.805300	-43.256600

However, it is quite interesting to note that all the four airports are prominent airports and it is highly unlikely that these airports really lack so many facilities/ services but it may be the case of non-availability of data in FourSquare API.

Before moving on to the next step lets see how the data stacks up for all the airports.

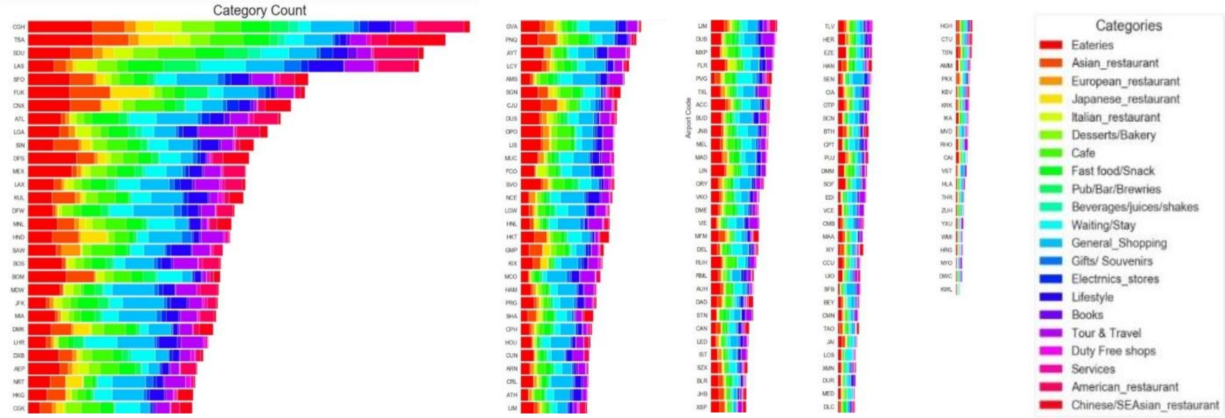
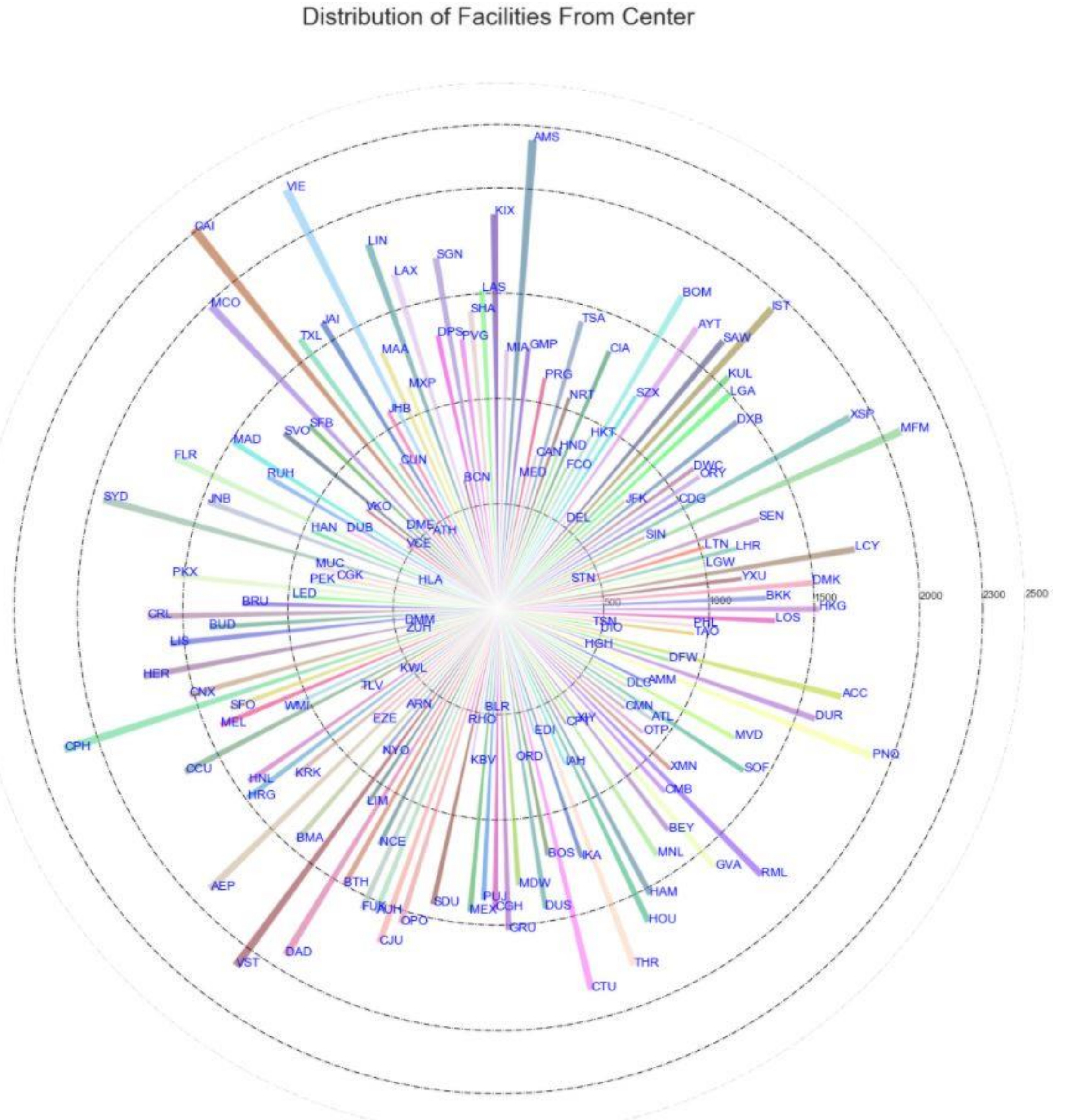


Figure 4: Airports sorted by total number of facilities

Now let us see how is the distribution of facilities from the centre of each airport. This tells us how far apart are the facilities from central point on an average.



3.2. Normalisation

By looking at out data we can see that some of the groups have much higher values than other for each of the airports. For example, ‘Eateries’, ‘services’ or ‘general_shopping’ have very high values when compared to ‘duty free shops’, ‘services’ etc. Hence there is a need to normalise our data for each airport. For this we use normaliser from scikit learn.

3.3. Model and parameter selection

Now we feed our normalised data to one of the clustering algorithms. As we have already processed data to remove skewness and outliers any of the two most popular clustering algorithms kMeans or DBSCAN can be applied. However, as in case of any unsupervised learning problem, it is not sure whether there are really any well-defined clusters to be segmented for our data or not. And in our case by common knowledge It can be said that most of the airports have more or less similar facilities hence clearly defined clusters are highly unlikely.

kMeans

For using kMeans we have to decide optimum number of clusters. To obtain optimum number of clusters we used have to run kMeans for various number of clusters and then plot ‘inertia’ to look for elbow in the graph. However, several times we cannot observe any clear elbow hence we have to decide optimum number of clusters based on trial and error and our intuition. This approach is crude hence for better results we must other clustering algorithm

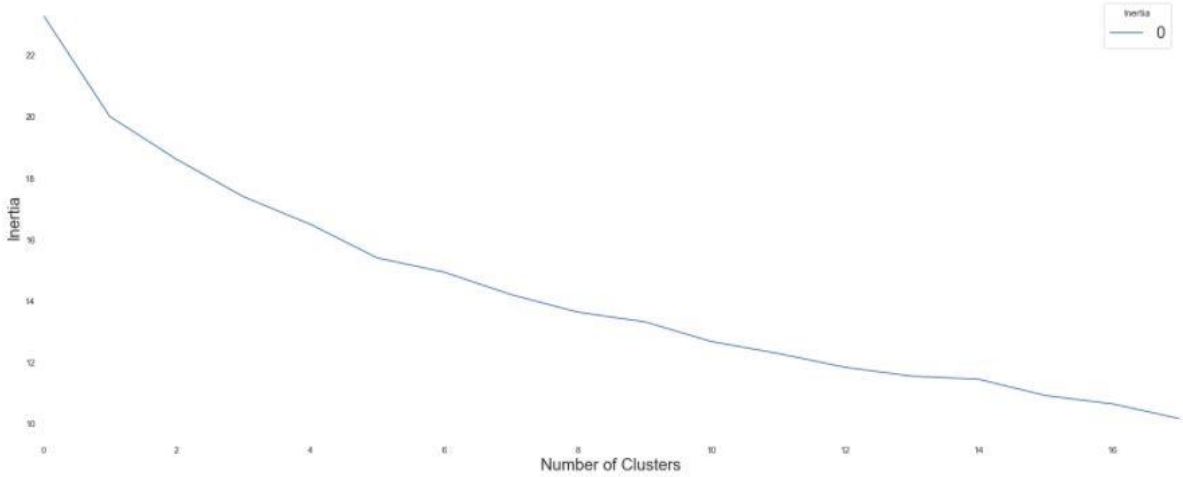


Figure 5: Inertia vs cluster plot

DBSCAN

For applying DBSCAN we need to decide upon two parameters *‘epsilon’* and *‘min_samples’*. *Epsilon* is nothing but the maximum distance within which data points can be considered belonging to one cluster. For determining epsilon a histogram, of distance of one point from rest all points, is plotted. And whatever distance has highest frequency, that value would be close to epsilon value to be used in analysis.

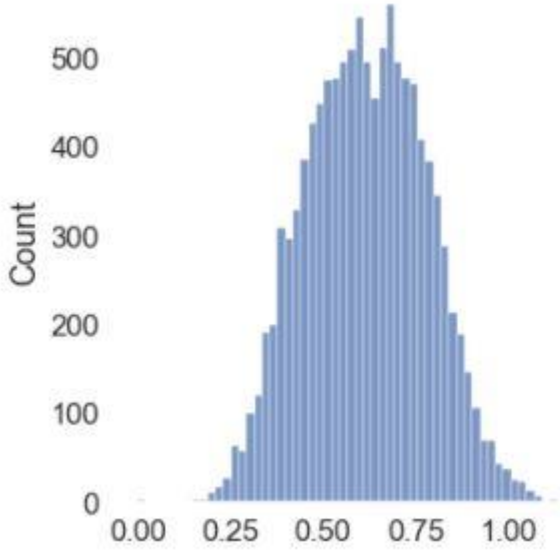


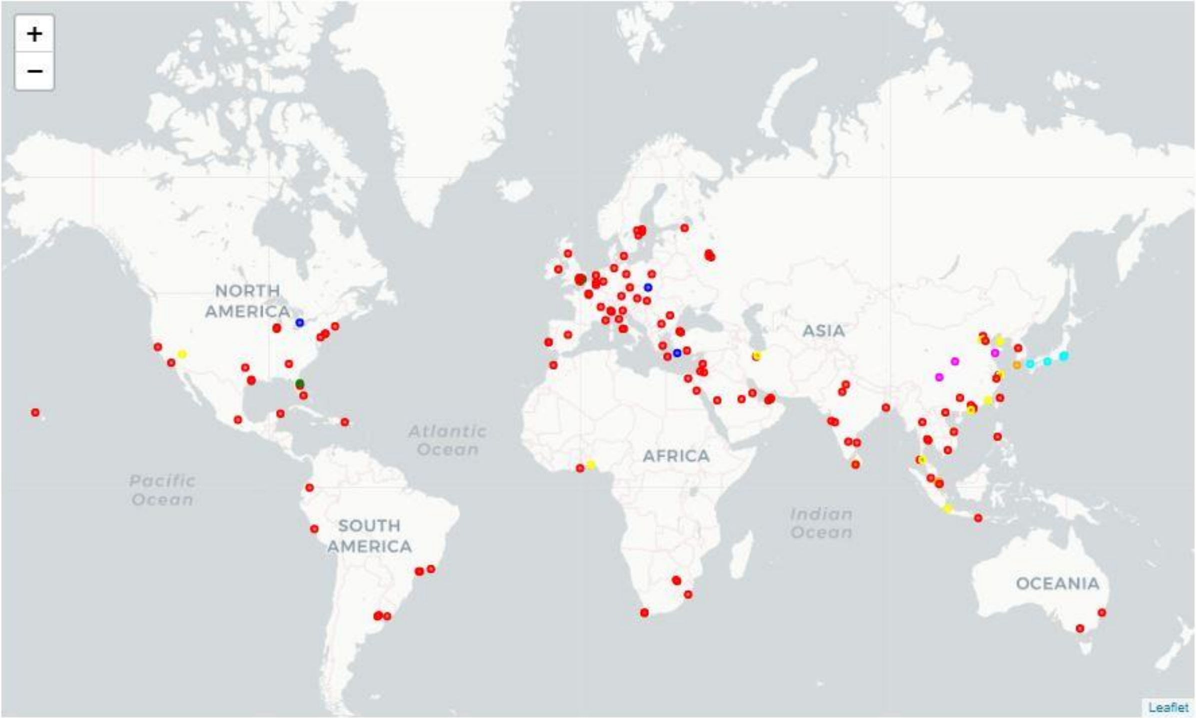
Figure 6: Frequency distribution of distance between two sample points

Hence by trial and error it is inferred that *'epsilon'* value of 0.36 is best fit and to maximise number of clusters *'min_samples'* is considered as 2.

4. Results

After running DBSCAN on final data it results in total 6 different clusters. Additionally, 11 out of 155 airports are such that they could not be assigned to any of the clusters. Seeing at the clusters obtained by our analysis it is quite clear that there is not much dissimilarity between most of the international airports in the world depending on the categories under consideration.

However it is also interesting to note that there are many peculiar features for each of the clusters that we have obtained and also the airports that could not be clustered. Which would be further elaborated in Discussion section.



5. Discussion

In this section we look clusters more closely and try to find patterns, peculiar features that differentiate one cluster from the other.

For Custer 1 first thing we notice that it is quite a big cluster comprising almost 80% of all the airports we considered. This is expected as we have considered all of the busiest international airports around the world and many of them have most of the facilities with similar frequency and same is represented by the results we got.

Contrary to Cluster1, Cluster2 comprises of only 3 airports and it is clear that these three airports lack, particularly "Asian Restaurant, Japanese restaurant, Beverages/juice/shakes, services & Chinese/ SE Asian restaurant". So, in case if someone requires these facilities it is better to avoid the airports of Cluster2. But this presents an opportunity for business owners too to open outlets in this field however that requires further study into specific area of interest

Like Cluster2, Cluster3 is also small cluster with 2 airports. Cluster3 has quite high numbers of "gifts/souvenirs shops and pub/Bar/Breweries" in comparison to total number of facilities across all categories on these airports.

Cluster4 comprises of 3 airports and they have comparable number of "Eateries & Asian restaurant". We can observe that in most of the clusters number of eateries are much more than number of Asian restaurants.

Cluster5 comprises of 4 airports and cluster6 has 3 airports. Both clusters are similar in the way that both represent airports belonging to same country. Cluster 5 airports are all in Japan whereas cluster 6 airports are all in China. It goes without saying that Cluster 5 airports have abundance of “Asian and Japanese restaurants “. However, one very interesting this is that these cluster 5 airports lack in “American and Chinese restaurants”. Considering the fact that Japan and America have very close ties and Japan is located very near to China it is very surprising to see such small number of “American and Chinese” restaurants. However this also represents opportunities for American and Chinese restaurant owners to head to Japan for their new outlets and till that time people looking for good Peking duck or fried chicken are out of luck travelling or in transit in Japan.

Cluster 6 airports on the other hand can certainly do with some European restaurants to start with as they have none. Also noteworthy is the lack of places to have alcoholic as well as non-alcoholic beverages.

6. Conclusion

Based on results and discussion later on it can be concluded that almost 80% of the world’s busiest international airports are more or less similar in terms of facilities offered by them. Certain airports in Japan (Cluster5) and China (Cluster6) do have potential for business owners of a particular category of restaurants. However as far as travellers are considered they have more information to make choices for their future connecting flights. However, as we know that for every individual the requirements may be different and this leads to the future scope section of this project.

7. Road Ahead

As mentioned in conclusion section as needs of different user is different hence this model can be deployed as web-based application allowing using to select the categories of places of interest relevant to them resulting in re-clustering of airports. This may be much more useful and can be developed into a full product from a project.