Clustering World Top 50 Universities Based on Venues

Donghuai Li

June 19, 2020

1.Introduction

1.1 Business problem

Imagine that you are a postdoc who will choose one university to live and work in, or a parent with children who will go to a college far away. You may be curious about the new environment around different universities and wonder if you or your children will adapt to life there easily. Will there also have restaurants that suit your taste, parks, cafes, bars, convenience stores just like home? You can help you do this by studying the cluster results of the surrounding environment of different universities. And now I will explain the need and reasons to do this research.

1.2 Background

For the world's top universities, different rankings have their own basis. For example, there are some most prestigious global rankings：

QS World University Rankings

Times Higher Education World University Rankings

Each of the score universities from their publication, peer review from scholars and academics and the ratio of faculty/students. Often, we also see some not so serious rankings of universities on the aspect of campus life's such as canteens, dormitories and entertainments. But most of these rankings are for universities in a certain region. As for me, I haven't seen a ranking list based on the surrounding environment of the university. That's the reason I want to explore the topic: clustering universities based on surrounding venues.

1.3 Personal Interest

As a traveller, every time I visit a new city, I will go to a local university to see it. I think a university in a city will not only reflect the spirit of the city, but also express the history of the city and even the country. University students and faculty members also usually arrange their daily activities within a certain distance near the university. It can be said that the distribution of different types of places near the university is affected by the combination with the spirit of the university and the preference of people and vice versa.

2. Data acquisition and preparation

2.1 Data sources

Different university rankings belong to business information. Fortunately, they have similar choices for the best universities in the world. So, we get the **Times Higher Education World University Rankings** which published on the wiki as our universities list. The graph below is the screen capture of the Top 50 universities from Times Higher Education World University Rankings' Wikipedia webpage.

World rankings [edit]

Institution	2010–11[42]	2011–12[43]	2012–13[44]	2013–14[45]	2014–15[46]	2015–16[47]	2016–17[48]	2017-18[49]	2018–19[50]	20
University of Oxford	6	4	2	2	3	2	1	1	1	1
California Institute of Technology	2	1	1	1	1	1	2	3	5	2
University of Cambridge	6	6	7	7	5	4	4	2	2	3
Stanford University	4	2	3	4	4	3	3	3	3	4
Massachusetts Institute of Technology	3	7	5	5	6	5	5	5	4	5
Princeton University	5	5	6	6	7	7	7	7	7	6
Harvard University	1	2	4	2	2	6	6	6	6	7
Yale University	10	11	11	11	9	12	12	12	8	8
University of Chicago	13	9	10	9	11	10	10	9	10	9
Imperial College London	9	8	8	10	9	8	8	8	9	10

Following is the full list of all 50 universities we used in analysis:

1	University of Oxford	26	University of Washington
2	California Institute of Technology	27	Carnegie Mellon University
3	University of Cambridge	28	London School of Economics and Political Science

4	Stanford University	29	New York University (NYU)
5	Massachusetts Institute of Technology	30	University of Edinburgh
6	Princeton University	31	University of California, San Diego
7	Harvard University	32	Ludwig Maximilian University of Munich
8	Yale University	33	University of Melbourne
9	University of Chicago	34	University of British Columbia
10	Imperial College London	35	University of Hong Kong
11	University of Pennsylvania	36	King's College London
12	Johns Hopkins University	37	University of Tokyo
13	University of California, Berkeley	38	École Polytechnique Fédérale de Lausanne
14	Swiss Federal Institute of Technology in Zurich	39	Georgia Institute of Technology
15	University College London	40	University of Texas at Austin
16	Columbia University	41	Karolinska Institute
17	University of California, Los Angeles	42	McGill University
18	University of Toronto	43	Technical University of Munich
19	Cornell University	44	Heidelberg University
20	Duke University	45	Katholieke Universiteit Leuven
21	University of Michigan	46	Université Paris Sciences et Lettres
22	Northwestern University	47	Hong Kong University of Science and Technology
23	Tsinghua University	48	University of Illinois at Urbana–Champaign
24	Peking University	49	Nanyang Technological University
25	National University of Singapore	50	Australian National University

2.2 Data transform

First, we use wiki package in python scratch the whole table and select universities' name.

Second, we find the relative location of each university via google map API.

	University	Latitude	Longitude
0	University of Oxford	51.758708	-1.255668
1	California Institute of Technology	34.137102	-118.125275
2	University of Cambridge	52.199852	0.119739
3	Stanford University	37.431314	-122.169365
4	Massachusetts Institute of Technology	42.358396	-71.095678

Then we plot universities on a world map according to their latitude and longitude.

Third, we analyse the university's surrounding venues via FourSquare api. After we fetch the surrounding venues, we count the total number of venues of each university. We can see roughly every university has 100 venues.

Universities	
Australian National University	76
California Institute of Technology	100
Carnegie Mellon University	100
Columbia University	100
Cornell University	100

Then we find the unique categories for venues and get dummies variables from them. After that, we get the top 10 most dominant venues as identification of universities. The result like the graph below.

```
———Australian National University———
                 venue   freq
0          Coffee Shop   0.105
1                 Park   0.105
2                 Café   0.092
3     History Museum    0.039
4                Hotel   0.026
```

We now finished the data preparation part and ahead to modelling.

3. Modelling

Because we wanted to explore the internal grouping of universities, we chose K means algorithm. After we set 6 clusters in total, we get the result quickly.

	Universities	Latitude	Longitude	Cluster Labels
0	University of Oxford	51.758708	-1.255668	2
1	California Institute of Technology	34.137102	-118.125275	0
2	University of Cambridge	52.199852	0.119739	2
3	Stanford University	37.431314	-122.169365	3
4	Massachusetts Institute of Technology	42.358396	-71.095678	3
5	Princeton University	40.338675	-74.658365	0

4. Analysis of Results

4.1 plot the clustered universities
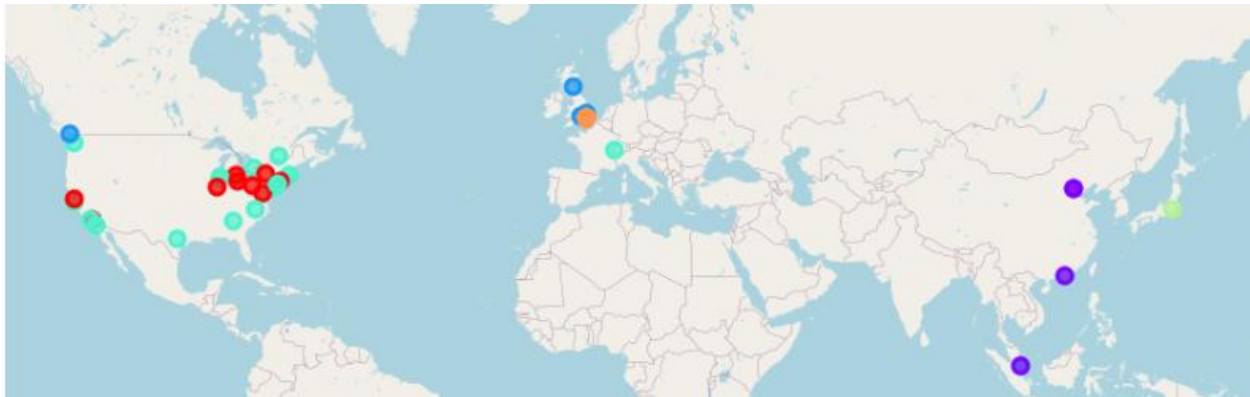
4.2 Analysis each cluster

Cluster 1 --- Pizza

We can see 'Pizza Place' 'Coffee Shop' have high rank in venues in cluster 1. The universities under cluster 1 mainly distributed on east part of United States.

	Universities	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	California Institute of Technology	Pizza Place	Garden	Burger Joint	Mexican Restaurant	American Restaurant	Coffee Shop	Gym / Fitness Center	Café	Mediterranean Restaurant	Bookstore
5	Princeton University	Pizza Place	Clothing Store	Hotel	Rental Car Location	Park	Coffee Shop	Ice Cream Shop	Trail	Mexican Restaurant	Lingerie Store
7	Yale University	Fast Food Restaurant	American Restaurant	Mexican Restaurant	Grocery Store	Pizza Place	Deli / Bodega	Gym / Fitness Center	Beach	Coffee Shop	Pharmacy
11	Johns Hopkins University	Pizza Place	Sushi Restaurant	Coffee Shop	Mexican Restaurant	Sandwich Place	Hotel	American Restaurant	Bakery	Ice Cream Shop	Asian Restaurant
12	University of California, Berkeley	Pizza Place	Coffee Shop	Café	Theater	Park	Scenic Lookout	Japanese Restaurant	Salad Place	Ice Cream Shop	Indian Restaurant
17	Cornell University	American Restaurant	Sandwich Place	Pizza Place	Coffee Shop	Park	Thai Restaurant	Café	Bagel Shop	Ice Cream Shop	Diner
19	University of Michigan	Sandwich Place	Korean Restaurant	Park	Coffee Shop	Indian Restaurant	Gym / Fitness Center	Sushi Restaurant	Chinese Restaurant	Pizza Place	Nature Preserve
25	Carnegie Mellon University	Coffee Shop	Pizza Place	New American Restaurant	Ice Cream Shop	Mexican Restaurant	Bar	Park	American Restaurant	Sandwich Place	Burger Joint
40	Heidelberg University	Pizza Place	Bar	Convenience Store	Coffee Shop	Ice Cream Shop	Gas Station	Asian Restaurant	Theater	Bank	Office
42	University of Illinois at Urbana–Champaign	Coffee Shop	Bar	Mexican Restaurant	Park	Ice Cream Shop	Pizza Place	Chinese Restaurant	Bakery	Hotel	Sandwich Place

Cluster 2 --- Chinese Restaurant

'Chinese Restaurant' counts the main part of venues in cluster 2. Universities under cluster 2 are deeply influenced by Chinese's culture.

	Universities	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
21	Tsinghua University	Chinese Restaurant	Café	Coffee Shop	Fast Food Restaurant	Sandwich Place	Pizza Place	Bar	Bakery	Korean Restaurant	Clothing Store
22	Peking University	Chinese Restaurant	Fast Food Restaurant	Pizza Place	Sandwich Place	Historic Site	Café	Coffee Shop	Korean Restaurant	Bar	Hotel
41	Hong Kong University of Science and Technology	Fast Food Restaurant	Chinese Restaurant	Cantonese Restaurant	Coffee Shop	Shopping Mall	Cha Chaan Teng	Clothing Store	Park	Beach	Hong Kong Restaurant
43	Nanyang Technological University	Fast Food Restaurant	Coffee Shop	Food Court	Japanese Restaurant	Chinese Restaurant	Asian Restaurant	Shopping Mall	Sandwich Place	Supermarket	Dessert Shop

Cluster 3 --- Pub/Cafe/Park/Coffee Shop

From the names, we can find that these universities are located in the United Kingdom or the original Commonwealth region.

And the 'Pub' and 'Cafe' definitely stand for its characteristic of cluster 3. Coffee shop and park also show a lot in the top 3 most common venues.

	Universities	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	University of Oxford	Pub	Coffee Shop	Café	Bakery	Movie Theater	Thai Restaurant	Restaurant	Ice Cream Shop	Sandwich Place	Chinese Restaurant
2	University of Cambridge	Pub	Café	Coffee Shop	Gastropub	Burger Joint	Park	Indian Restaurant	Grocery Store	Bakery	Deli / Bodega
28	University of Edinburgh	Café	Pub	Park	Coffee Shop	Hotel	Cocktail Bar	Beer Bar	Restaurant	Art Gallery	Sandwich Place
30	University of Melbourne	Café	Coffee Shop	Wine Bar	Bar	Deli / Bodega	Vegetarian / Vegan Restaurant	Cocktail Bar	Japanese Restaurant	Chinese Restaurant	Bookstore
31	University of British Columbia	Coffee Shop	Park	Beach	Garden	Café	Scenic Lookout	Restaurant	Frozen Yogurt Shop	Botanical Garden	Deli / Bodega
44	Australian National University	Park	Coffee Shop	Café	History Museum	Pub	Restaurant	Hotel	Art Gallery	Wine Bar	Monument / Landmark

Cluster 4 --- Delicious food around world

We can see there no domain venue category can select from cluster 4. Restaurants of various different tastes illustrate the cultural fusion among these universities. No matter where you come from, you can find your own taste in these universities.

	Universities	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Stanford University	Park	Gym / Fitness Center	Coffee Shop	Ice Cream Shop	Café	Art Museum	French Restaurant	Steakhouse	New American Restaurant	Furniture / Home Store
4	Massachusetts Institute of Technology	Bakery	Coffee Shop	Seafood Restaurant	Salad Place	Pizza Place	New American Restaurant	Spa	Clothing Store	Mediterranean Restaurant	Italian Restaurant
6	Harvard University	Bakery	Pizza Place	Café	Park	Grocery Store	Ice Cream Shop	Pub	Vegetarian / Vegan Restaurant	Bookstore	Seafood Restaurant
8	University of Chicago	Park	Pizza Place	Grocery Store	Coffee Shop	Lounge	Bank	Art Gallery	History Museum	Bookstore	Café
10	University of Pennsylvania	Coffee Shop	Park	American Restaurant	Trail	Grocery Store	Art Museum	Pizza Place	Hotel	Ice Cream Shop	Outdoor Sculpture
14	Columbia University	Park	Coffee Shop	Southern / Soul Food Restaurant	Italian Restaurant	Pizza Place	Wine Shop	Ice Cream Shop	Spa	Farmers Market	Cocktail Bar
15	University of California, Los Angeles	Coffee Shop	Italian Restaurant	Gym	Grocery Store	Hotel	Mediterranean Restaurant	Sandwich Place	Ice Cream Shop	Sushi Restaurant	Garden
16	University of Toronto	Coffee Shop	Bakery	Park	Sandwich Place	Café	Dessert Shop	Bookstore	Dance Studio	Caribbean Restaurant	French Restaurant
18	Duke University	Sandwich Place	Hotel	Mexican Restaurant	Gastropub	Café	Coffee Shop	Burger Joint	BBQ Joint	Bakery	Mediterranean Restaurant
20	Northwestern University	Coffee Shop	Beach	American Restaurant	Pizza Place	Bakery	Gym	Sushi Restaurant	Mexican Restaurant	Park	Brewery
23	National University of Singapore	Indian Restaurant	Chinese Restaurant	Italian Restaurant	Coffee Shop	Ice Cream Shop	Asian Restaurant	French Restaurant	Japanese Restaurant	Gym / Fitness Center	Dessert Shop
24	University of Washington	Park	Italian Restaurant	Seafood Restaurant	Pizza Place	Bubble Tea Shop	Brewery	Grocery Store	Ice Cream Shop	Indian Restaurant	Thai Restaurant
27	New York University (NYU)	Italian Restaurant	Bakery	New American Restaurant	Pizza Place	Mediterranean Restaurant	Yoga Studio	Garden	Bookstore	Salad Place	French Restaurant
29	University of California, San Diego	Coffee Shop	Cosmetics Shop	New American Restaurant	Grocery Store	Burger Joint	Furniture / Home Store	Juice Bar	Lounge	Resort	Beach

Cluster 5 --- Tokyo-One must be like a team

Maybe the convenience stores imply that Tokyo has the world's leading convenience stores. I believe that students studying at the University of Tokyo can eat hot Kanto cooking at any time.

	Universities	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
34	University of Tokyo	Convenience Store	Drugstore	Ramen Restaurant	Coffee Shop	Discount Store	Supermarket	Restaurant	Japanese Restaurant	Italian Restaurant	Shopping Mall

Cluster 6 Art, Movie, Coffee and ...... Hotel!

The universities in London are classified as a separate group, probably because they are too close together. Another explanation is the venues distribution shows that London is rich in tourism resources (numerous hotels) and art resource. We can conclude that number one common venue is the hotel and followed by theatre, coffee shop and etc.

	Universities	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
9	Imperial College London	Hotel	Garden	Café	Ice Cream Shop	Plaza	Park	Pub	Coffee Shop	Monument / Landmark	Restaurant
13	University College London	Hotel	Coffee Shop	Grocery Store	French Restaurant	Ice Cream Shop	Theater	Bookstore	Pedestrian Plaza	Plaza	Movie Theater
26	London School of Economics and Political Science	Hotel	Theater	Coffee Shop	Plaza	Grocery Store	Art Gallery	Ice Cream Shop	History Museum	Art Museum	Liquor Store
33	King's College London	Hotel	Coffee Shop	Theater	Grocery Store	Cocktail Bar	History Museum	Plaza	Bakery	Pedestrian Plaza	Performing Arts Venue

5 Conclusion

From the above results, we can easily obtain the results of different universities according to the environment. It can also be applied to a wider range of problems. For example, we can explore the surrounding venues of all universities in the United States, or universities with similar venues distribution around Asia. For special needs, for example, if you want to select more Chinese restaurants around the university, it can also be easily done by modify the code. It can be said that we have solved the problems raised at the beginning: Find similar universities based on surrounding venues.