

Comparison of G20 Largest Cities

Similarities and differences of venues in G20 largest cities





















Introduction

- ▶ The top 20 economies of the world are those of the G20 which consists of 19 countries plus the European Union.
- ▶ The purpose of this study is to compare the largest cities of these 19 countries and see how similar or different they are to each other.
- ▶ The cities will be clustered into four distinct clusters based on most common venues.

Data & Methodology

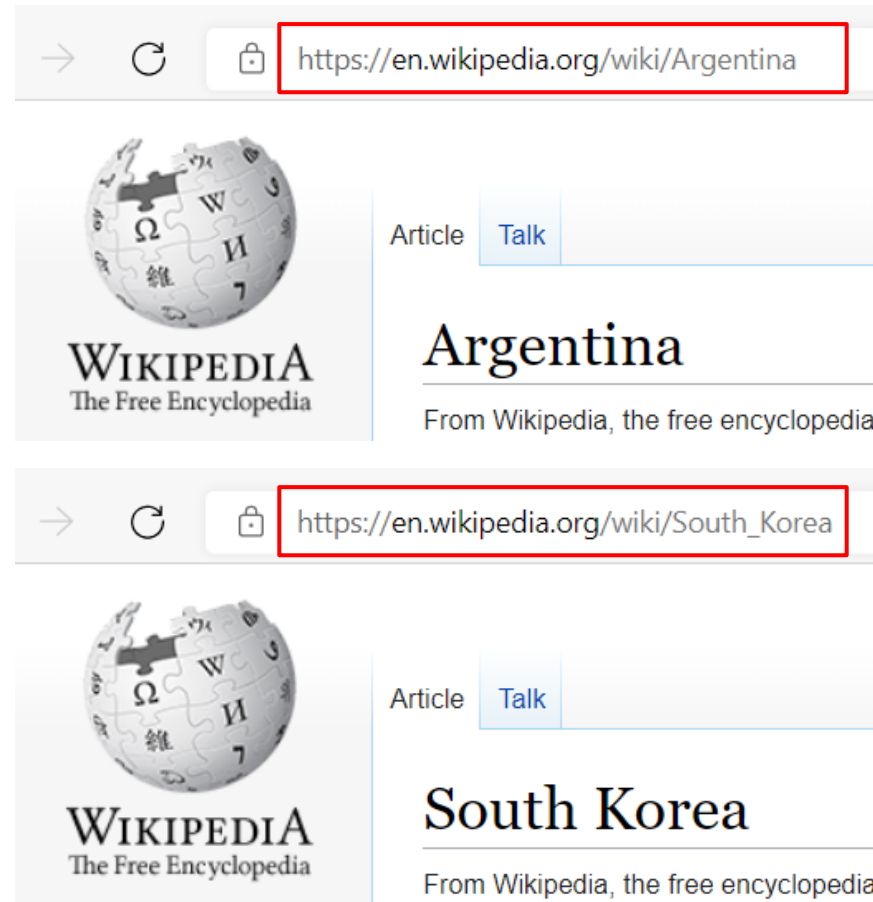
- Data is obtained by web scraping Wikipedia for information on the largest cities in G20 countries.
- The Wikipedia page for G20 is <https://en.wikipedia.org/wiki/G20>
- The list of countries is available inside a table on that page.

Member country data [edit]

Member	Trade bil. USD (2018) <small><i>citation needed</i></small>	Nom. GDP mil. USD (2020) ^{[46][47]}	PPP GDP mil. USD (2019) ^[48]	Nom. GDP per capita USD (2019) ^{[49][50]}	PPP GDP per capita USD (2019) ^[51]	HDI (2019)	Population (2018) ^[52]	Area km ²
 Argentina	127	388,279	920,209	10,667	20,537	0.845	44,570,000	2,780,400
 Australia	481.1	1,359,330	1,369,392	56,698	52,373	0.944	25,182,000	7,692,024
 Brazil	650.0	1,434,080	3,596,000	8,955	17,016	0.765	210,869,000	8,515,767
 Canada	910	1,643,410	1,896,725	46,733	49,651	0.929	37,078,000	9,984,670
 China	4,629	14,722,840	27,331,166	10,276	18,110	0.761	1,396,982,000	9,596,960
 France	1,227.4	2,598,910	3,054,599	42,931	45,775	0.901	65,098,000	640,679
 Germany	2,834	3,803,010	4,467,238	48,670	52,559	0.947	82,786,000	357,114
 India	830.7	3,050,000	11,468,022	2,171	7,874	0.645	1,334,221,000	3,287,263
 Indonesia	368.9	1,059,640	3,743,159	4,120	13,230	0.718	265,316,000	1,904,569
 Italy	1,047.4	1,884,940	2,442,144	34,349	39,637	0.892	60,756,000	301,336
 Japan	1,486.6	5,048,690	5,749,550	39,306	44,227	0.919	126,431,000	377,930
 South Korea	1,140.4	1,630,870	2,229,779	32,046	41,351	0.916	51,665,000	100,210
 Mexico	915.2	1,076,160	2,658,041	9,807	20,602	0.779	124,738,000	1,964,375
 Russia	687.5	1,473,580	4,135,000	11,601	28,184	0.824	146,850,200	17,098,242
 Saudi Arabia	369.1	701,467	1,924,253	23,187	55,944	0.854	33,203,000	2,149,690
 South Africa	187.8	302,114	875,100	6,560	15,239	0.709	57,420,000	1,221,037
 Turkey	391	719,537	2,274,072	9,346	27,956	0.820	71,867,000	783,562
 United Kingdom	1,157.1	2,710,970	3,128,185	42,261	45,705	0.932	66,466,000	242,495
 United States	4,278	20,932,750	21,344,667	62,606	62,606	0.926	328,116,000	9,833,517
 European Union		15,167,820	22,761,233	33,715	41,091	0.900	512,600,000	4,233,262

Data & Methodology

- ▶ The Wikipedia pages for countries follow a standard format for their URLs that is easy to follow.
- ▶ The Wikipedia pages URL is “https://en.wikipedia.org/wiki/” followed by the name of the country.
- ▶ If the country’s name is made of more than one word, the words are separated by an underscore (_).



Data & Methodology

- ▶ The largest city of each country is listed on the Wikipedia page of the country.
- ▶ The city's name can be found in a table that can be found using BeautifulSoup.
- ▶ The coordinates for each city are obtained using a geocoder.
- ▶ All data collected through web scraping is stored in a data frame to be studied.



Data & Methodology

- ▶ Information on venues for each city is collected using the Foursquare API.
- ▶ One-hot encoding is used to create a separate column for each venue and city.
- ▶ The data is grouped by city using the mean function to make cities comparable to each other.
- ▶ The top ten venues for each city shall be considered for the comparison.
- ▶ The cities are clustered into four clusters using K-means clustering.

Results

- ▶ The search returns 1712 venues in 292 unique categories.
- ▶ The cities are not evenly distributed among the four clusters.
- ▶ The largest cities in the G20 countries are very similar in categories of most common venues.
- ▶ Most cities have restaurants, coffee shops, and hotels in their top ten most common venues.

Results

► Cluster 0 represents 11% of G20 countries

Largest City	Cluster labels	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
Shanghai	0	Coffee Shop	Fast Food Restaurant	Hotel	Chinese Restaurant	Café	Lounge	Indian Restaurant	Asian Restaurant	Gym	French Restaurant
Jakarta	0	Indonesian Restaurant	Fast Food Restaurant	Asian Restaurant	Café	Hotel	Coffee Shop	Padangnese Restaurant	Bakery	Food Truck	Noodle House

Results

► Cluster 1 represents 21% of G20 countries

Largest City	Cluster labels	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
Buenos Aires	1	Café	Pizza Place	Argentinian Restaurant	Bakery	Ice Cream Shop	Burger Joint	Coffee Shop	Gym	Italian Restaurant	Indie Theater
Sydney	1	Café	Australian Restaurant	Scenic Lookout	Hotel	Japanese Restaurant	Italian Restaurant	Ice Cream Shop	Cocktail Bar	Theater	Thai Restaurant
Toronto	1	Café	Coffee Shop	Japanese Restaurant	Restaurant	Sushi Restaurant	Clothing Store	Gym	Furniture / Home Store	Plaza	Middle Eastern Restaurant
Johannesburg	1	Café	Fast Food Restaurant	Portuguese Restaurant	Breakfast Spot	Art Gallery	Historic Site	Coffee Shop	Hotel	Scenic Lookout	Public Art

Results

► Cluster 2 represents 63% of G20 countries

Largest City	Cluster labels	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
São Paulo	2	Japanese Restaurant	Cultural Center	Café	Sake Bar	Grocery Store	Theater	Bookstore	Dessert Shop	Bakery	Snack Place
Paris	2	French Restaurant	Ice Cream Shop	Plaza	Bookstore	Restaurant	Art Museum	Bakery	Tea Room	Lebanese Restaurant	Bar
Berlin	2	History Museum	Drugstore	Hotel	Coffee Shop	Bookstore	Monument / Landmark	Cocktail Bar	Art Museum	Theater	Art Gallery
Rome	2	Historic Site	Italian Restaurant	Plaza	Sandwich Place	Ice Cream Shop	Monument / Landmark	Wine Bar	Temple	Garden	Church
Tokyo	2	Hotel	Café	Japanese Restaurant	Chinese Restaurant	Chocolate Shop	Italian Restaurant	French Restaurant	Nabe Restaurant	Coffee Shop	Historic Site
Seoul	2	Hotel	Korean Restaurant	Coffee Shop	Café	Chinese Restaurant	Japanese Restaurant	Sushi Restaurant	Plaza	Historic Site	Bakery
Mexico City	2	Mexican Restaurant	Ice Cream Shop	Art Museum	Museum	Arts & Crafts Store	Hotel	Restaurant	Jewelry Store	Clothing Store	Boutique
Moscow	2	Boutique	Hotel	Coffee Shop	Plaza	Italian Restaurant	Cosmetics Shop	History Museum	Art Gallery	Beer Bar	Caucasian Restaurant
Riyadh	2	Jewelry Store	Asian Restaurant	Hotel	Middle Eastern Restaurant	Park	Historic Site	Shopping Mall	Electronics Store	Toy / Game Store	Market
Istanbul	2	Hotel	Turkish Restaurant	Mosque	Café	Historic Site	Restaurant	Jewelry Store	Kebab Restaurant	Bookstore	Seafood Restaurant
London	2	Hotel	Ice Cream Shop	Garden	Bakery	Gelato Shop	Steakhouse	Lounge	Coffee Shop	Plaza	Cocktail Bar
New York City	2	Coffee Shop	Wine Shop	Spa	Gym / Fitness Center	Memorial Site	Café	French Restaurant	Park	Gym	Burger Joint

Results

► Cluster 3 represents 5% of G20 countries

Largest City	Cluster labels	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
Mumbai	3	Bar	Indian Restaurant	Coffee Shop	Flea Market	Multicuisine Indian Restaurant	Mexican Restaurant	Pizza Place	Italian Restaurant	Food & Drink Shop	Food Court

Conclusion

- ▶ Since Cluster 2 has the biggest representation of G20 countries, following the lead of the countries in that cluster in developing a city would be the safest bet for the customer to achieve their goals.