COURSERA CAPSTONE FINAL ASSIGNMENT

Oytun Altasli Widmer

24 July 2020

1. INTRODUCTION

1.1 Background

In the last 20 years, the coffee shops have spread in most metropolitan areas globally, spawning multiple coffee chains as well as a new “Café Culture”. This has gone hand in hand with changes in consumer preferences both in terms of consumption of coffee (over tea or other drink) and how it is consumed (take-away vs. sit-down). In business centres, cafés have emerged as places where business meetings take place. After the advent of wireless internet, coffee shops have become increasingly flexible office space as well as study hall for students. While it remains to be seen if pandemic will have a long-lasting effect in our relationship with cafes, it is nevertheless interesting to explore how this social phenomenon compare globally, across different cities and cultures.

1.1 Idea

As prevalent as the spread of coffee houses may be, there are significant differences in terms of intensity in city centres which may be driven by culture, climate and focus such as being a financial or tourism centre. As a first point of interest, I would like to collect sufficient data to see if the intensity of coffee houses in city centres can be used as a predictor of economic activity or a cultural group. This can in turn be used for a future project to predict up-and-coming regional centres if historical location data can be gathered. A secondary interest is the chain vs. independent breakdown as well as prevalence of specific chain brands.

1.2 Audience and Stakeholders

I think this data would be useful for entrepreneurs who are considering opening cafes, restaurants or other competing venues as it would allow them 1) to identify a new developing areas, 2) avoid overserved centres, 3) to understand competition and the character of the area. By the same token, this data would also be useful for chain coffee shop managers who are looking to optimise their locations. Finally, coffee-loving travels should find it interesting.

2. DATA

2.1 Data description and acquisition

I utilised FourSquare to extract the current list of coffee houses for major city centres across the globe. To do this, I defined a 500m radius for city centre, and generated a list of coffee houses using search for 'coffee' and 'café' on FourSquare.

The search for 'coffee' vs. 'café' yields significantly different results for many cities which I will discuss in more detail later in the report. A combined search of 'coffee and café' was not feasible as FourSquare limits results at 50 venues, and a combined search yielded more than 50 venues for almost all cities even with a smaller search radius. At the end, I settled to use 'coffee' for my search as it seemed to yield more meaningful results. In addition, FourSquare offers this as default search in its web search engine.

As cities differ significantly in terms of their populations and density, I decided to adjust my numbers for city density. For this, I used London Datastore webpage at this link. Clearly, there are other factors such as number of tourists, but I did not include it in my analysis in the scope of this analysis.

2.2 Data cleansing

Before preparing a final selection of cities and preparing my main dataset, I ran individual test-runs on some of the cities that I am familiar with such as London, New York, Istanbul and Paris. The results were surprisingly distinct which made me dig deeper into the type of data. The first observations were:

1) New York City data included corner carts which resulted in having a very big count of coffee houses,

2) A big metropolis like Paris returned surprisingly small number of coffee shops. A search for 'cafes' rather than 'coffee' gave a better representation. The same search gave a low quality data for London whereas 'coffee' search was more appropriate,

3) The cities in India gave very few results considering their populations.

On final analysis, I decided to use 'coffee' as my search item as I decided this yielded a better quality of data.

As a second step, I prepared a short-list of cities from different continents taking care to include a wide-range of cultures. In this step, I continued to quality-check data where the numbers did not seem reasonable, and finalised my list accordingly. One challenge was downloading the right coordinates for some of the cities in Africa, which I have been able to only partially resolve. I downloaded coordinated from geolocator, and I assumed these were good representations for city centres. My quality check on individual cities seem to support that view.

My main dataset included a list of 20 cities, and the total number of coffee houses they had within a 500m radius of their centre. The final list from FourSquare was relatively clean, and required limited cleansing.

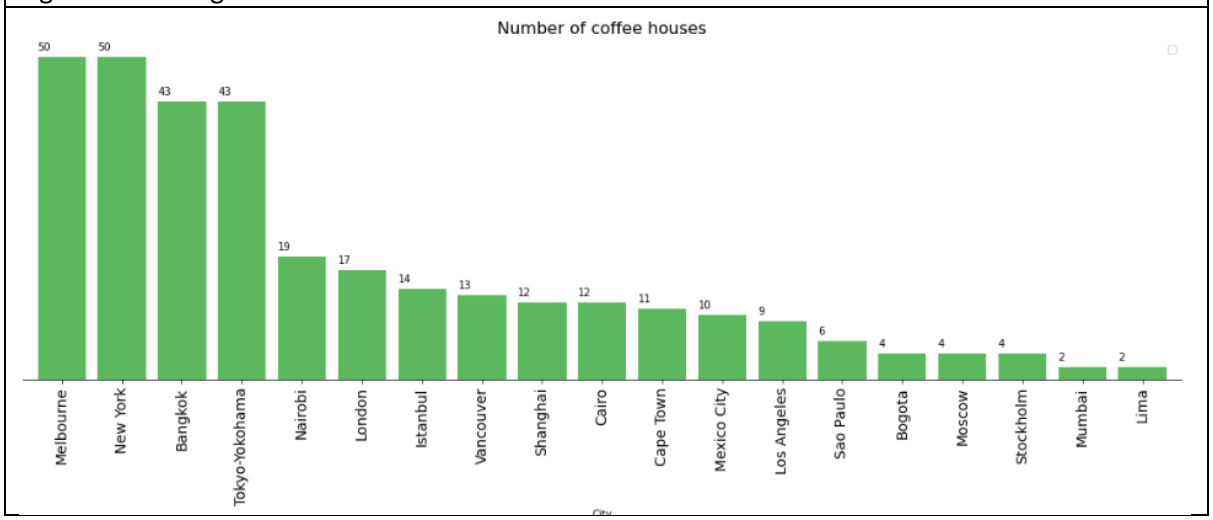
I accessed data about global city population density by read_excel. After deleting irrelevant data, I had to wrangle the city names as they included commas, brackets and such.

3. METHODOLOGY

3.1 Determining the number of coffee houses in city centres

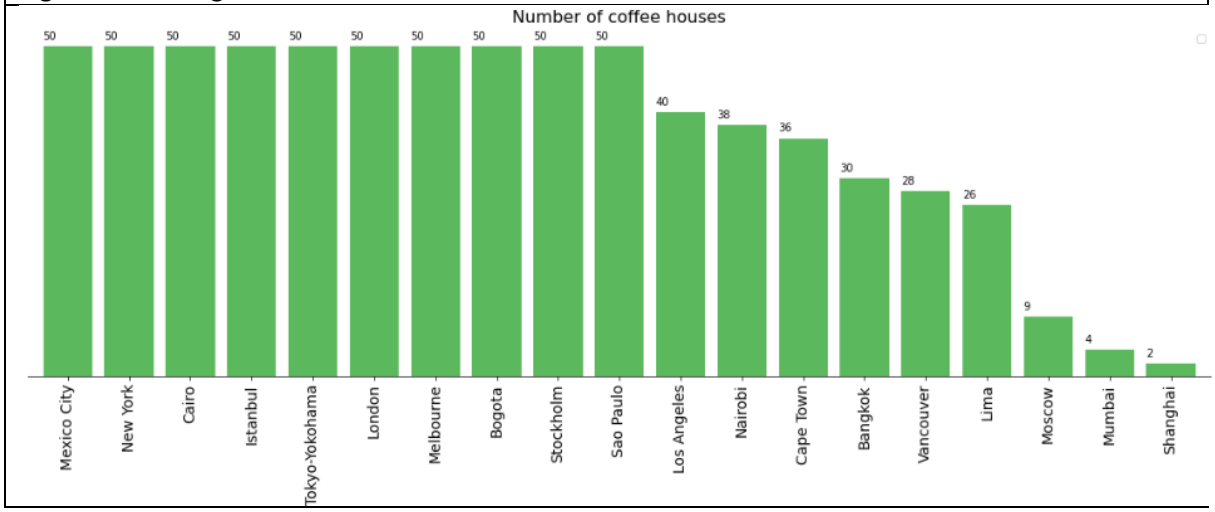
After grouping coffee houses by name, I sorted and created bar charts to see how cities compared. It was not very surprising to see New York City at the top, while Melbourne's top ranking was somewhat interesting. It was also noteworthy that Nairobi came ahead of London while it was almost suspicious that big metropoles like Sao Paolo could be so low in the ranking. In fact, Paris had less than 10 coffee houses according to this search.

Figure 1: Ranking under 'coffee' search



I made a second search with 'café' as the keyword which yielded more results in more categories. In fact, most cities hit the 50 results per search limit of FourSquare – even under a smaller radius. A closer look at breakdown of data revealed that many of the venues were in fact would not qualify as coffee houses. Therefore, I decided to progress with my first search 'coffee'.

Figure 2: Ranking under 'café' search



3.2 Comparison of top two cities: New York vs. Melbourne

A more detailed look into individual data reveals additional insight. In New York City, It is common to see carts that sell various food items, including coffee, at the corner of blocks. These carts offer only

take-away, and would not normally qualify as a coffee houses. While it may make sense for FourSquare to list them under 'coffee' search, it does not satisfy the coffee house definition, and is therefore artificially boosting New York City numbers. Melbourne, on the other hand, seems to indeed offer a very wide selection of genuine coffee houses.

Figure 3.1: Top 15 venues in NYC

	City	Latitude	Longitude	Cafe_name
0	New York	40.712728	-74.006015	Blue Spoon Coffee Co.
1	New York	40.712728	-74.006015	Proof Coffee Roaster
2	New York	40.712728	-74.006015	For Five Coffee Roasters
3	New York	40.712728	-74.006015	Coffee Cart
4	New York	40.712728	-74.006015	Hudson/Chambers Coffee Cart
5	New York	40.712728	-74.006015	John & William St. Coffee Cart
6	New York	40.712728	-74.006015	Mary's Coffee Shop
7	New York	40.712728	-74.006015	WFM Coffee Bar
8	New York	40.712728	-74.006015	Nick's Coffee Truck
9	New York	40.712728	-74.006015	Periscope Coffee On John Street
10	New York	40.712728	-74.006015	Blue Bottle Coffee
11	New York	40.712728	-74.006015	Birch Coffee
12	New York	40.712728	-74.006015	Blue Spoon Coffee Co.
13	New York	40.712728	-74.006015	Dante Coffee Shop
14	New York	40.712728	-74.006015	Think Coffee

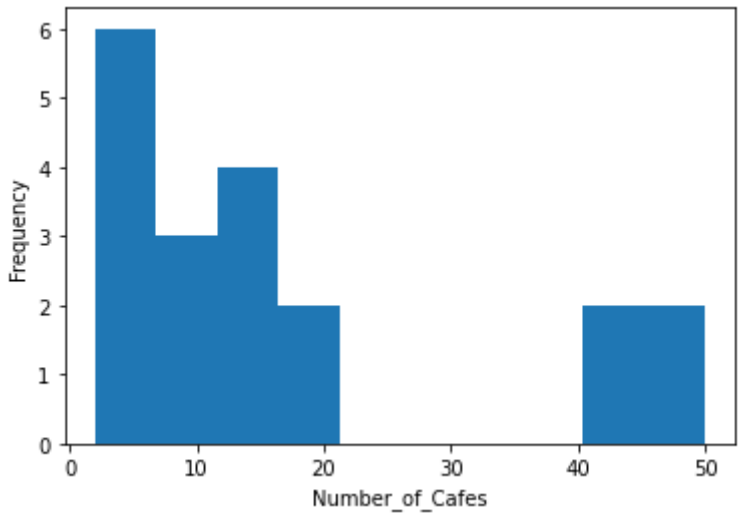
Figure 3.2: Top 15 venues in Melbourne

	City	Latitude	Longitude	Cafe_name
142	Melbourne	-37.814218	144.963161	Federal Coffee Palace
143	Melbourne	-37.814218	144.963161	No 1 Coffee
144	Melbourne	-37.814218	144.963161	Hash Specialty Coffee
145	Melbourne	-37.814218	144.963161	Coffee & Katsu Sando Bar (サンドワルー)
146	Melbourne	-37.814218	144.963161	Hudsons Coffee
147	Melbourne	-37.814218	144.963161	Waffee Waffles+Coffee
148	Melbourne	-37.814218	144.963161	Black Coffee Shop
149	Melbourne	-37.814218	144.963161	The Fix Coffee & Bake
150	Melbourne	-37.814218	144.963161	Quists Coffee
151	Melbourne	-37.814218	144.963161	Hudsons Coffee
152	Melbourne	-37.814218	144.963161	Map Coffee
153	Melbourne	-37.814218	144.963161	Coffee HQ
154	Melbourne	-37.814218	144.963161	Romano's Coffee and Bagels
155	Melbourne	-37.814218	144.963161	Milano Coffee House
156	Melbourne	-37.814218	144.963161	Cartel Coffee Roasters

3.3 Histogram and scatter plots

I utilised histogram to understand how the number of cafes clustered. Accounting for the artificially high New York numbers, it seems a lot of global cities have less than 10 cafes in their centres. I would like to highlight again that the accuracy of data for this analysis is not perfect, and actual number is likely to be slightly higher than this – perhaps in the 10-20 range.

Figure 4: Histogram of number of cafes in city centres

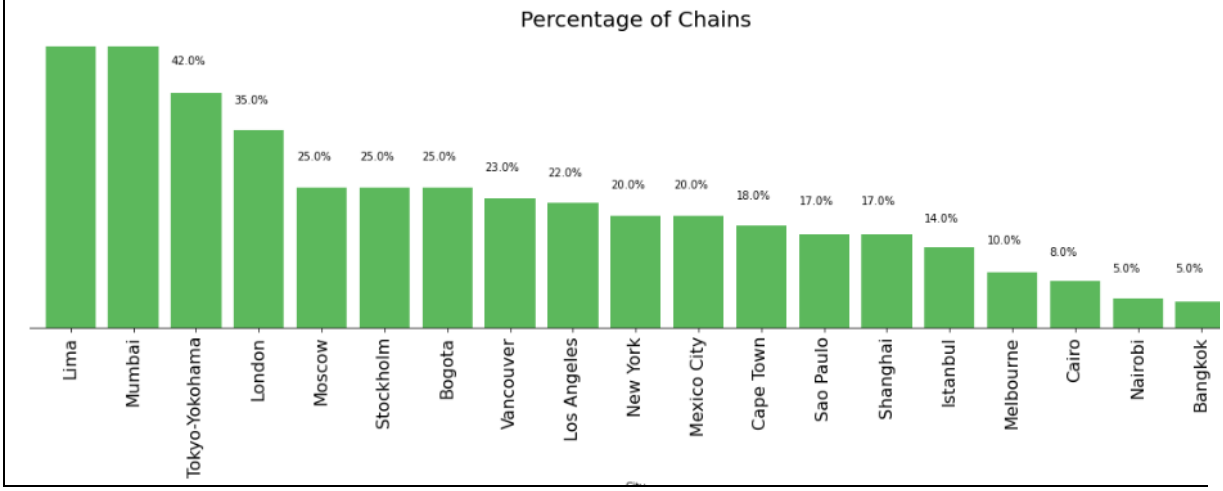


3.4 Calculation of chain presence in city centres

By counting unique coffee house names, I was able to calculate the percentage of coffee houses that belonged to a chain. This analysis puts Mumbai and Lima at the top of the list. Given that Mumbai and Lima each have only two coffee houses each, this is rather misleading. The next in the list are

Tokyo and London which has strong chain presence with 42% and 35% penetration. In the case of London, Costa Café seems to have a very dominant presence in the city centre.

Figure 5: Percentage of coffee chains in city centres

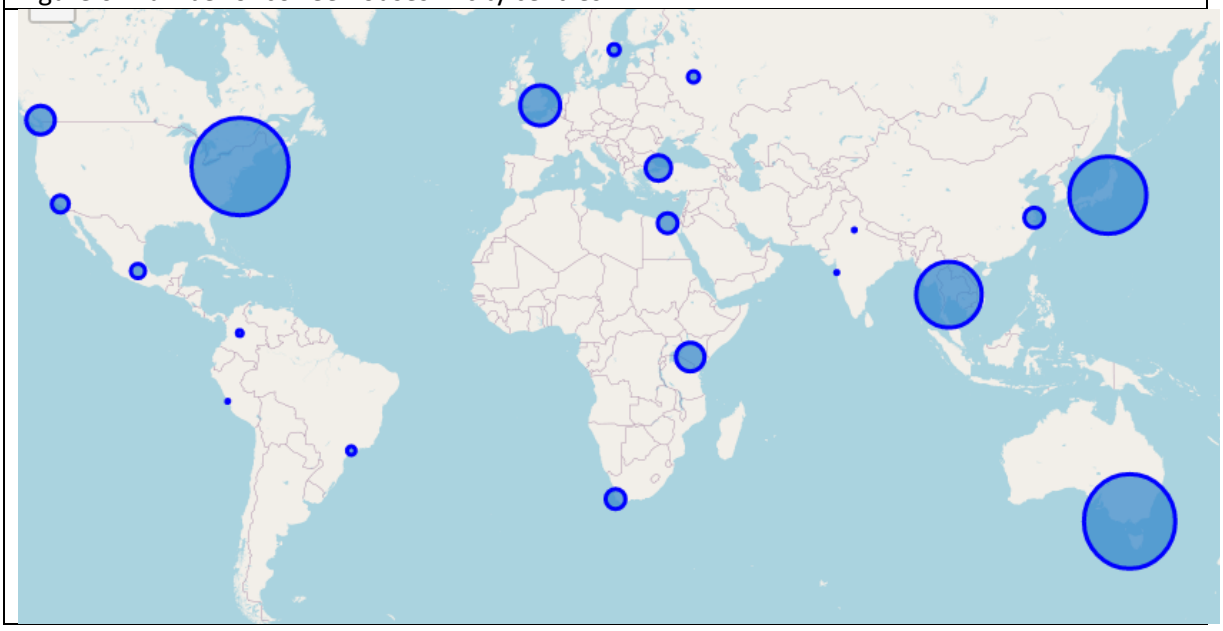


3.5 Bubble charts

For a visual overview of my dataset, I used Folium Maps charts with bubble size representing the number of coffee houses in city centres. This was a useful tool to see geographical and potential cultural differences.

It is striking to see how few coffee houses the Indian cities have according to FourSquare data, especially considering their sizes. I found it equally surprising to see small bubbles in South America. However, a search for 'café' immediately increases venue numbers in this geography which is not the case for indian cities in our list.

Figure 6: Number of coffee houses in city centres

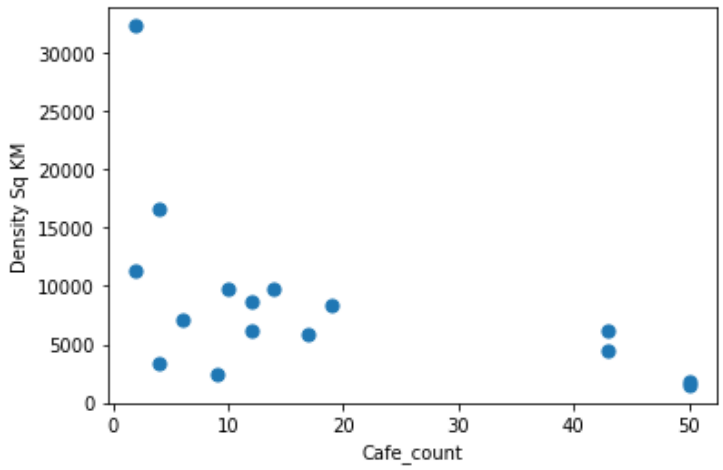


3.6 Accounting for city density

Finally, I adjusted my dataset to account for city density which is population over urban area. To achieve this, I calculated 'café density' which is essentially number of coffee houses over per capital per area.

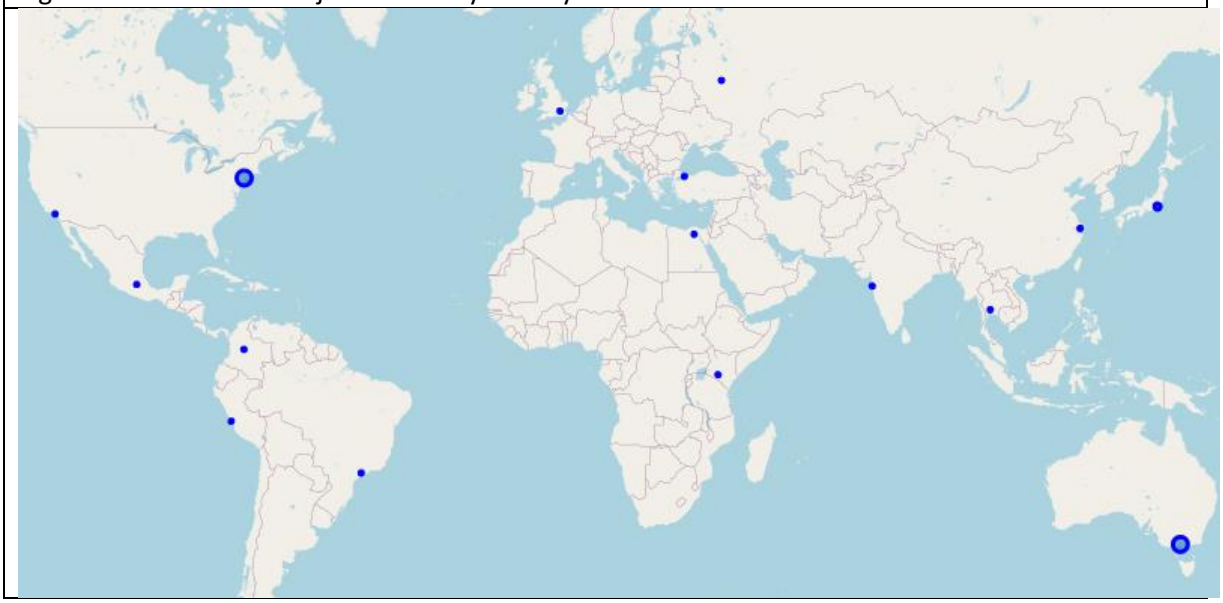
A scatter plot of city density vs. number of cafes suggests that it is generally positively correlated, if I exclude outliers. The outliers in this instance are Mumbai and Bogota which are both dense and have a limited number of coffee houses according to our FourSquare search. I also note that New York density data is underestimated (at lower right hand of the chart with 50 coffee houses) as it includes a bigger metropolitan area.

Figure 7: City density vs. number of coffee houses in the city centre



The resulting bubble chart confirms the finding of the scatter plot, namely the density adjusted coffee shop count results in a more uniform distribution.

Figure 8: Bubble chart adjusted for city density



4. RESULTS

This project provides confirmation of some our initial expectations while yielding other surprising outcomes:

- As expected, main financial centres such as New York and London have a big number of coffee houses. In the case of New York, the number is overestimated because of coffee carts included in FourSquare data. I preferred to keep the carts in the data as I would be unable to identify similar venues in other cities.

- There were some surprising locations such as Melbourne and Ulaanbaatar (not shown on the charts here) that have very large number of coffee shops. According to FourSquare data, Melbourne has a very vibrant independent coffee scene.

- In terms of chain penetration, Tokyo and London top the ranks with 42% and 35% chain market share, respectively. In London, Costa Coffee is the only chain in the city centre which is located at Trafalgar Square. In contrast, Tokyo has a more variety of coffee chains in its centre. Moscow, Stockholm and Bogota also have relatively high coffee chain presence in their centres although Bogota figures looked skewed due to only four coffee shops.

- Indian cities has almost suspiciously limited number of coffee houses which requires additional investigation. We think this may be to do with city centre coordinates provided by Geolocator.

5. DISCUSSION

Based on this project, I noted that coffee market as defined by coffee shops is a quite fragmented market. Although most city centres are penetrated by coffee houses, most are independent names, and there are pockets of areas where chains can still potentially expand and consolidate.

However, there are some cities such as London and Tokyo where chain penetration at the centres have arguably reached its limits.

FourSquare data proved to be a practical source for the analysis of chains, and the project can be extended on that front.

On the other hand, the data was insufficient to compare cities on the basis of their coffee cultures which we tried to identify by the number of coffee shops. The differences in classification, be it driven by city-specific venues (like the carts in New York City) or other concerns, requires further cleansing before I can build confidence in my conclusions.

6. CONCLUSION AND FUTURE DIRECTION

In the West, we are living in a coffee culture where cafes have become an integral part of our daily lives: we meet, work, study and eat at coffee shops. In the beginning of this project, my goal was to find differences across cultures and continents as well as explore the dominance of coffee chains globally.

As I became more familiar with FourSquare data, I realised that there are significant variations in categorisations of venues by country. Ultimately, the type and quality of FourSquare data makes it hard to draw solid conclusions regarding consumer preferences in different cities.

The project provides a basis for additional work on coffee chains. While the FourSquare data has its short comings, it is still a good source for further analysis on the penetration of the coffee chains. The project can be expanded to include more cities in the regions. It is also possible to branch out the project to analyse brand dominance in different cities.