

DECODING OF CHARACTERISTICS OF URBAN CONSUMPTION BASED ON “BRAND CLUSTER”

Maxim Tang^{*}, Octavia Wei²

¹*Dataway Horizon, Shanghai Bigdata Department, GM*

²*Dataway Horizon, Shanghai Bigdata Department, Senior Consultant*

Abstracts

Brand Cluster is proposed based on the background of evolved consumption modes and concepts as well as brand preferences of different categories of consumers. With the support of inter-urban, inter-category and inter-brand big data, after deep learning and profound analysis of consumption relations of different brands, *Brand Cluster* was born to reflect characteristics of diverse consumers.

We try to understand the inner features of 18 clusters of brands and how these clusters look like in different cities, which underlies the practice of city siting of brand owners.

Brand Cluster is believed to reveal the relationships between “allies” of brands in a whole new angel of view and in the large. In addition, the make-up of brand clusters in different cities indicate whether a new city is appropriate for brand owners to expand into.

*Keywords:*Consumption relations, Brand cluster, City siting

* Corresponding author: tanghao@idataway.com

I. Background

In Shanghai, you may probably hear conversations that go as follows:

“Half a year of hard work! I finally saved some money. I want to pamper myself with Burberry.”

“Wonderful! What about shopping in Lujiazui tomorrow?”

“By the way, my brother is going college soon. I plan to buy him Converse shoes.”

“No problem. Let’s go to Huaihai Road after Lujiazui.”

“Besides, I need some underwear from sixty8ight.”

“Next time, we can go to ShinMay Union Square in Yaohan.”

“ ... ”

Traditionally, it’s thought that purchases on different brands usually take place in business districts that exhibit different consumption characteristics. With the development of modernization and urbanization, a city can be deemed as a collection of business districts. Brands of different levels and targeted at different individuals are not evenly distributed in cities at different developmental phases. That is to say, the brands that a city possess is a good manifestation of its consumption development level. Also, pattern of brand consumption in a city is an indicator of its economic diversity, attractiveness, openness and inclusiveness.

Recently, with economic development and the increase of disposable incomes, words such as “consumption upgrade”, “individuation” and “bespoke” suggest a great change in consumer attitudes and an entry into a customer-oriented era. In this context, consumption structure, consumer demands and attitudes are evolving rapidly.

Diversity and individuation of consumer needs has posed a challenge to sellers as well. The conventional seller’s market gradually gives way to a diversified and personalized market that have to cater to various consumers with mixed needs.

Shopping malls and brand owners have devoted themselves to studying consumer behaviors and exploring potential needs. Restructuring product models, tries at personalized marketing, crossover and combination of online and offline services are not rare. Even though, we still need to further deepen our understanding of consumers, their attitudes and behaviors.

To better explore the relationship between needs for different products and categories in daily life, we delve into the correlations between brands and thereby propose the concept of “Brand Cluster”.

II. Research Methods

1 Data Description

This research mainly adopted interbank transaction data. These data keep a good track of consumers' transactions and have the following characteristics:

- (1) Covers a large population. Nowadays, almost everyone in China has at least one debit or credit card supported by the Union Pay (The total number of cards in China is in billions). As the biggest settlement organization, the Union Pay has a marvelous coverage of population, cards, and transactions.
- (2) Covers a great many consumption categories. Card transactions cover almost all consumption categories and a huge number of stores, and thus can keep a comprehensive, accurate and timely track of consumer behaviors. Nearly 40 overall sections, 200 more specific categories (named as MCC), and tens of thousands of stores in the data, altogether provide a solid basis for feature analysis, correlation analysis and more advance deep learning models.
- (3) Captures any time trend if present. Transactions are recorded at any time, and hence any macro trend of consumption environment or micro trend of consumer groups can be well captured.

Below are some examples of the transaction data fields.

Table. Examples of Bank Transaction Data

Transaction id	Card class	Card number	City of Issue	City of Store
Store ID	Store Name	Transaction Category(mcc)	Amount	Time

2 Sampling

Due to the high data volumes, we adopt the stratified probability sampling method, which considers card class and spending power, to obtain a subset that covers one million cards. All the subsequent analyses are based on the transactions of these selected cards.

Compared to average studies, our data not only has a huge sample size, but also is highly representative of the whole population, and therefore can provide a more credible and reliable basis for research.

3 Modeling

3.1 Objective

We divide more than 300 brands into groups according to their correlations, including multiple similarity measures, via clustering methods. For brands in the same group, if they are in the same industry, a competitive relationship is implied, otherwise, a correlative relationship can be inferred. The results are a good manifestation of the consistencies in brand styles and consuming propensity within the same group.

3.2 Basic Idea

First, a bunch of widely-known and representative brands are selected to form a brand pool. A grid is established for each card on all the brands in the pool. Then from the grid, a correlation network is created for all the one-to-one combinations of brands. Afterwards, *taboo search* is applied, together with qualitative methods, to divide brands into clusters. Each cluster is fully summarized for its characteristics and probed into for cross-cluster relationships.

Network clustering algorithm is used, with qualitative methods for validation. Such methods are first applied to brand networks, and then extended to city evolvement networks.

We do not consider common clustering methods, such as K-Means or Hierarchical Clustering (which requires input like some sales-related figures or brand value labels), because they only consider the problem from a shallow perspective and ignore the underlying correlations between brands. But with network clustering, the similarity matrix is obtained from cross-brand transactions and can be used to help in facilitating a more neutral analysis of any local correlation network to achieve the objective.

3.3 Work Flow

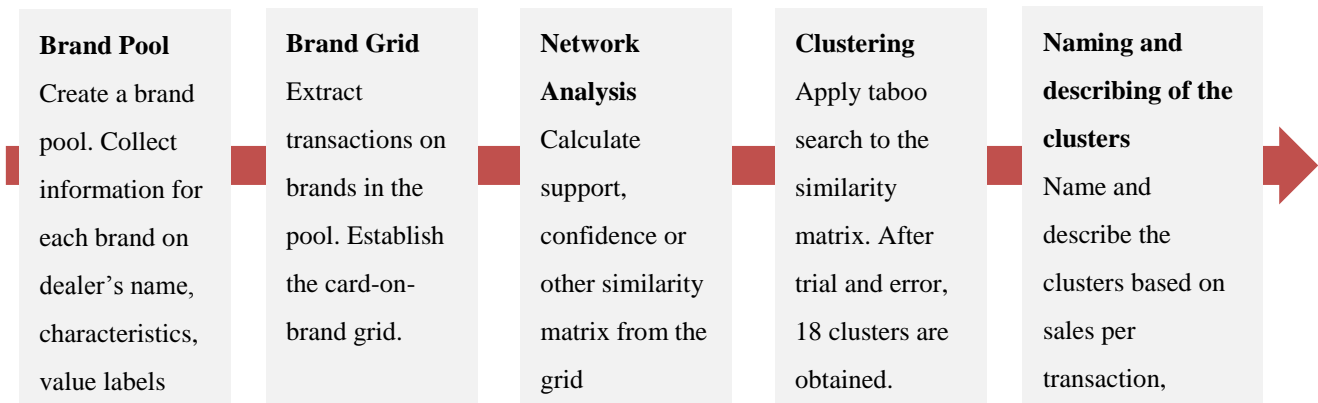


Chart. Chart Flow of Brand Cluster

3.3.1 Brand Pool

Since there are too many brands in the market, and some of them are too minor to be analyzable, or the research itself is unimportant and even meaningless, in order to prevent such brands from disturbing our analysis, to make the research more efficient and generalizable, we first need to establish a pool of representative brands that share a solid customer base and potential for analytically meaningful outcomes.

Considering past research experience (on consumer goods, commercial properties and so on), and brand performance in the market, after preliminary analysis and filtration, we obtained a pool of representative brands. For each brand in the pool, more information is collected on such as, category, Chinese name, English name, characteristics (features and population coverage), value labels and so on. All the information provides a basis for the clustering of brands in the future.

3.3.2 Brand Grid

In order to find brand correlations, first of all, we extract all the transactions on the brands in the pool from the big sample data. To achieve this, we create a list for searching where each brand may involve several keywords. This process is extremely time-consuming and labor-intensive. After the comprehensive search, a grid with rows representing cards, columns brands, is attained. This grid exhibits a general picture of consumer behaviors on the brands.

Card01-Card09 in the following table correspond to different cards respectively. In the columns, 1 indicates the card owner buys the brand, 0 the otherwise.

Table. Example of Brand Grid

Brand	Aigle	Alexander McQueen	BAPE	Bottega Veneta	Burberry
card01	0	0	1	0	1
card02	0	0	1	0	0
card03	0	0	1	0	0
card04	0	0	0	0	0
card05	0	0	1	0	0
card06	0	1	0	0	0
card07	0	1	0	0	0
card08	0	1	0	1	1
card09	1	0	0	0	0

3.3.3 Network Analysis

Calculate the similarity matrix from the grid. This $n \times n$ symmetric similarity matrix measures the closeness between the brands. It's obvious that similarities between some brands are extremely weak, while strong between some other brands. Such accurate results cannot be obtained from traditional field search. Such improvements again demonstrate the advantage of big data analysis on brand.

Table. An example of similarity matrix between brands

(figures omitted and color scale used instead to measure the similarities, the darker the stronger)

	361	6ixty8ight	AAPE	ADIDA, Oakley	ALISA	APM Monaco	ASOBIO
361							
6ixty8ight							
AAPE							
ADIDAS, Oakley							
ALISA							
APM Monaco							
ASOBIO							

3.3.4 Brand Cluster

Taking goodness of fit and business thinking into consideration, we meticulously scrutinize each possible combination of brands, and finally group all the brands into 18 clusters.

To sum up, Brand Cluster makes use of transaction data from 1 million cards, which covers more than 300 cities, 281 specific categories, and manages to decipher the consumption correlation among over 300 national brands. Brands within the same cluster are highly correlated and shares some latent characteristics, while brands from different clusters do vary in some consumption characteristics.

The best part of Brand Cluster lies in its global view. No assumptions presumed, Brand Cluster tackles the issue of consumer behaviors in a more general, global and comprehensive perspective. The alliance of brands as inferred can help enterprises to find potential collaborators and can help achieve a better union of inter-industry brands which shares some underlying characteristics that appeal to similar consumers.

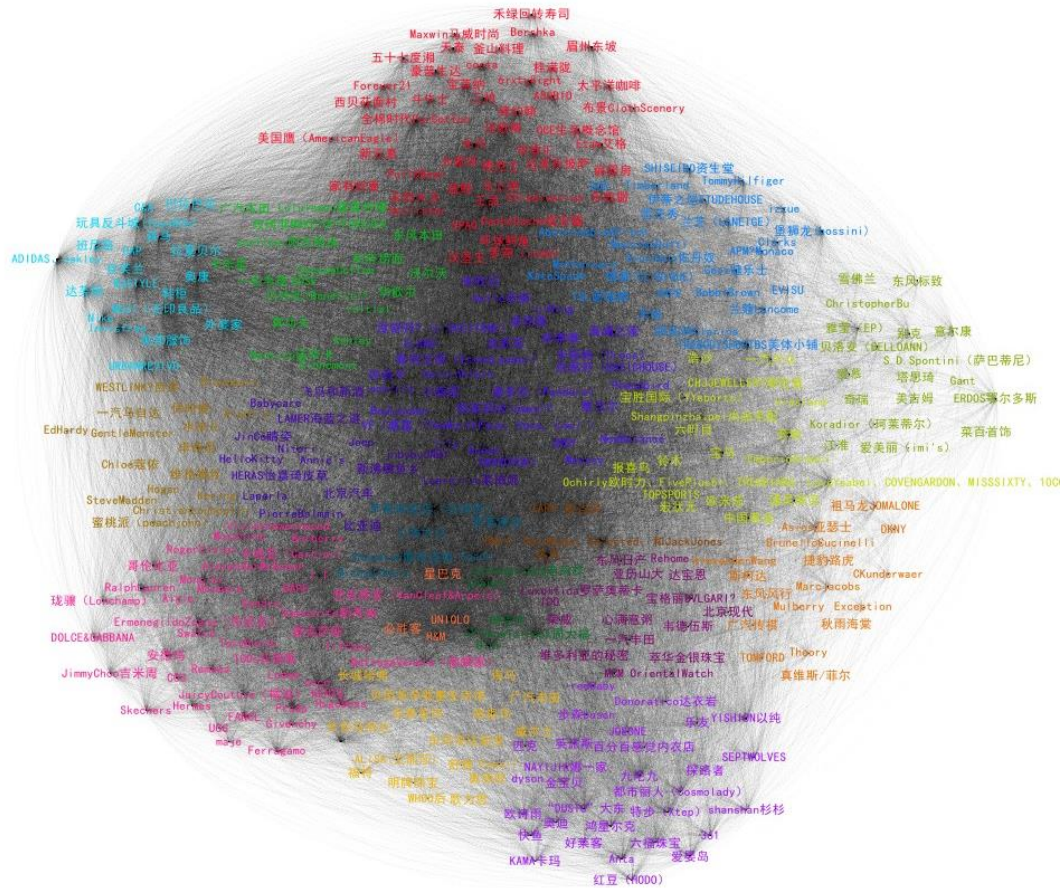


Figure. Overview of Brand Networking

III. Research Results

1 Naming and Describing of the Clusters

We need to summarize the characteristics of each cluster, entitle a proper name and form a consumer profiling. This step is a description of some summary statistics, an interpretation of the results, and a validation of network analysis. The naming and describing requires a good awareness of each brand's DNA, as well as a good understanding of consumer perceptions.

We find that brands within the same cluster do share some intra-cluster characteristics, which facilitates naming, and that brands from different clusters are significantly different, which suggests the partition is reasonable and solid. Brand Cluster can easily accommodate new brands or new clusters when necessary.

By studying the average amount spent by each card within the same cluster, structure of categories and standardized sales, we summarize the overall characteristics of each cluster and conceive an appropriate name.

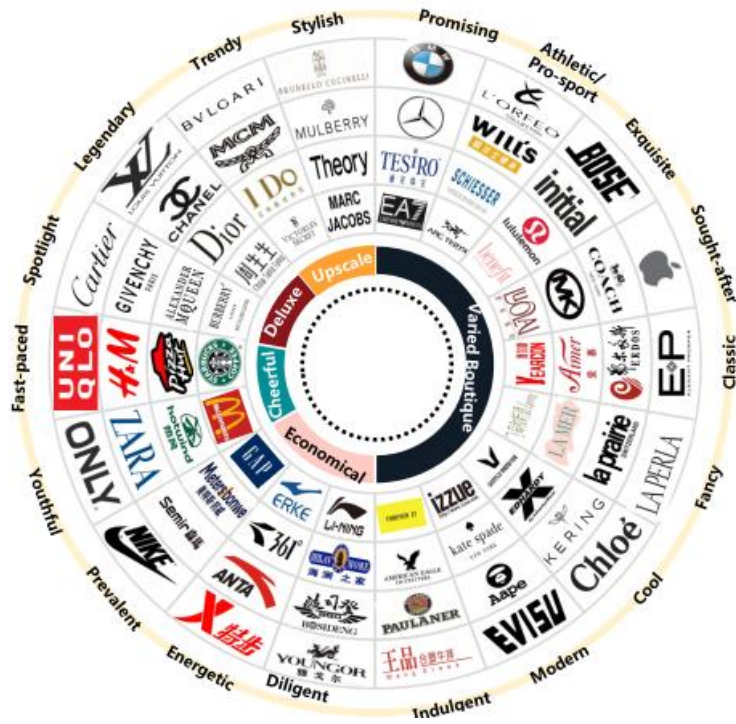
1.1 Names of Clusters

- First, divide all the brands into five main classes based on their levels and target market: Deluxe, Upscale, Varied, Economical, Cheerful
- Second, divide brands within the same main class into smaller groups based on their geographic, socioeconomic, psychographic and behavioral targets.

Below are the details of our naming systems.

Table. Naming of Brand Cluster (In a descending order of the average amount)

Cluster No.	Name	Main Class
1	Spotlighted	Deluxe
14	Legendary	Deluxe
16	Trendy	Upscale
6	Stylish	Upscale
18	Promising	Varied
2	Athletic	Varied
4	Exquisite	Varied
11	Sought-after	Varied
8	Classic	Varied
17	Fancy	Varied
7	Cool	Varied
5	Modern	Varied
9	Indulgent	Varied
13	Diligent	Economical
3	Energetic	Economical
10	Prevalent	Economical
12	Youthful	Cheerful
15	Fast-paced	Cheerful



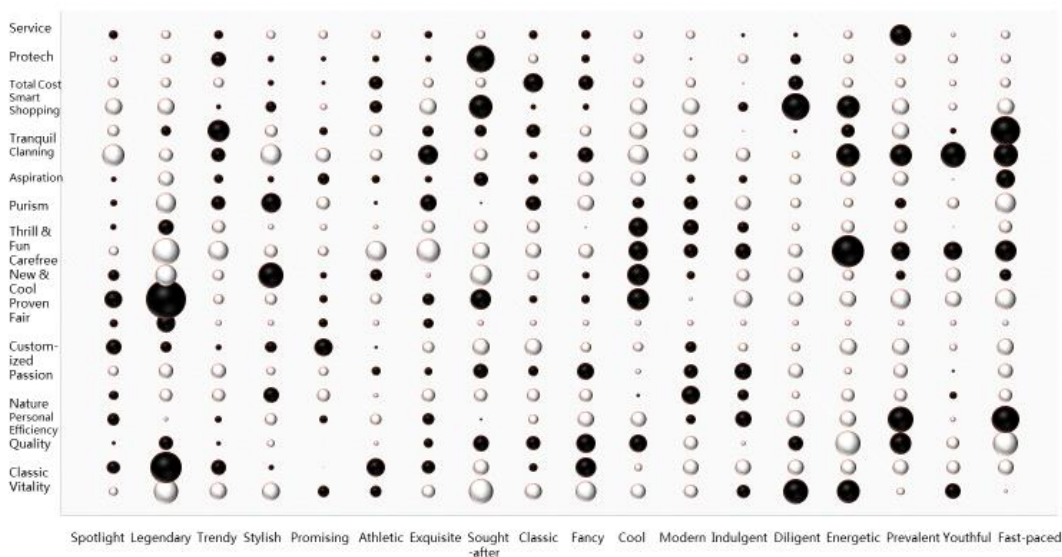
The following diagram exhibits 4 representative brands for each cluster.

Figure. Overview of Brand Cluster

1.2 Labels of Values

The values for each cluster are shown in the following bubble plot. Black bubbles indicate above-average performance, while white bubbles indicate below-average performance. It can be inferred that clusters do vary in features. Each cluster has several outstanding values and exhibit quite distinct performance. For instance, the *Legendary* cluster has prominent performance on “proven” and “classic” values, while the *Sought-after* cluster tend to be “pro-tech” and “smart shopping”.

Figure. Bubble Plot for Value Labels of Brand Cluster



The name and characteristics of each cluster is a good manifestation of values of the brands that constitute the cluster, is also an indication of the underlying consumer propensity and market positioning strategies. For instance, luxuries like Louis Vuitton, Dior, or Chanel and the *Legendary* cluster where they belong have certain features in common. Consumers who buy such luxuries prefer the brands that deliver sense of dignity and elegance which undoubtedly will highlight their identity and status.

1.3 Examples of Brand Clusters

Name : *Deluxe-Spotlighted*

Description: Top scale as the focus of attention

Examples of Brands : Samsonite, Bottega Veneta

Implication :

Luxuries of new fashion are proud of themselves. They pursue general visual enjoyment but meanwhile have their own style. They display a unique ingenuity in design and provide supreme experience in use. Wealthy youngsters/the middle-aged are their target consumers. This group are inclined to be different from the mass and up-to-date on the purchases of products.

In addition, people with a habit of overdraft are likely to be customers of brands in *Spotlighted* cluster. They can be middle or senior managers of young age, who have unique pursuit in life quality.

Sample Portrait of Consumers :

Tom, 31 years old, came back from America 6 years ago and now undertakes the responsibility of BD director in an internet company.

He pays attention to his outfits and appearance because of frequent participation in social occasions. Reading luxuries websites and fashion magazines, shopping Burberry and Givenchy at Wangfujing are his hobbies during leisure time. Also, he will buy his girlfriend a Cartier necklace as an important festival gift to show his affection.

Luxuries in new fashion equips him with confidence and personal charisma.

Name : *Deluxe-Legendary*

Description: Dignity from the start and for ever

Examples of Brands : Chow Sang Sang, Lao Feng Xiang, Dior, Chanel

Implication :

It is domestically widely recognized that classic luxuries are famous for its long history, abundant cultural background and impressing brand story. They prefer style of seriousness rather than exaggeration and therefore are popular among high-flyers.

all over the country with a relatively higher percentage in third-tier and fourth-tier cities.

Sample Portrait of Consumers :

Sara, 31 years old, is currently living in Shandong, with a pursuit of life quality and a personality of straightforwardness.

She is fond of Dior Eau de Parfum which reflects her willingness to be her true self. Her mother chose Lao Feng Xiang as her wedding ring which delivers her blessing that the new couple should live together happily for good and all.

Classic luxuries have been widely accepted in common life as they are representative, low-key and of inheritance. They have occupied people's mind and will still take a seat in the future.

Name : Economical-Energetic

Description : Like a new start of life with energy and passion

Examples of Brands : Lukfook Jewelry, ERKE

Implication :

With the arrival of consumer era, purchase of named brands has become a new trend. Brands in the *Energetic* cluster usually display a sense of vitality and comfortability, and a simple but unique design, and has therefore been the choice of the majority.

These brands are frequently seen in daily life and speak best to the hearts of customers who think of staying comfortable as important as improving the quality of life. The brands are also competitively priced and easily accessible to the youngsters and the working class. Such customers usually exhibit a great deal of energy and vigor.

Sample Portrait of Consumers :

Mubai, a campus student living in Shenyang, is very into basketball. He is a student athlete, and good at multiple sports. He is tall and good-looking, and therefore undoubtedly the dream guy for many girls. Mubai has a fondness for sports brands. He prefers to dress comfortably and casually, and especially loves playing basketball with a pair of Anta sneakers. He plays basketball like a duck to water.

2 Performance of Brand Clusters in Cities

2.1 Urban Coverage of Brand Clusters

Brand distribution in clusters of every city has its own characteristics. As shown in the below data (see Table. *Examples of urban coverages of brand cluster*), maturely developed brand clusters, such as *Energetic*, *Legendary* and *Fast-paced*, account for high proportion in each city. In contrast, the percentages of some other brand clusters such as *Exquisite* and *Spotlighted*, vary in different cities, which reflects the phase of city evolution.

In view of the urban coverage of different brand clusters, first-tier and second-tier cities have a wider range of clusters, while third-tier and forth-tier cities have relatively fewer clusters. This implies more diverse consumption patterns of first-tier and second-tier cities than the other cities. Besides, attention to traditional mass luxury goods has been transferred from first-tier and second-tier cities to cities of lower tiers, while personalized and trendy luxuries have been more popular in first-tier and second-tier cities.

Table. Example of urban coverage of brand clusters
(figures omitted; the greener the shades, the larger the coverage)

Cities/Clusters	Spotlighted	Athletic	Energetic	Exquisite	Stylish	Prevalent	Legendary	Fast-paced
Beijing	Light Green	Orange	Light Green	Light Green	Yellow	Light Green	Dark Green	Dark Green
Tianjin	Light Green	Red	Light Green	Orange	Red	Light Green	Dark Green	Dark Green
Shijiazhuang	Red	Red	Light Green	Orange	Red	Light Green	Dark Green	Dark Green
Qinhuangdao	Red	Red	Yellow	Red	Red	Orange	Light Green	Dark Green
Baoding	Red	Red	Light Green	Red	Red	Light Green	Dark Green	Dark Green
Zhangjiakou	Red	Orange	Light Green	Red	Red	Orange	Light Green	Dark Green
Langfang	Red	Red	Light Green	Red	Red	Yellow	Light Green	Dark Green
Linfen	Red	Orange	Light Green	Red	Red	Light Green	Dark Green	Yellow
Yuncheng	Red	Red	Light Green	Red	Red	Orange	Light Green	Dark Green
Hulun Buir	Red	Red	Light Green	Red	Red	Light Green	Dark Green	Yellow
Changzhou	Red	Red	Light Green	Red	Red	Light Green	Dark Green	Dark Green
Hangzhou	Yellow	Yellow	Light Green	Light Green	Orange	Light Green	Dark Green	Dark Green
Ningbo	Orange	Orange	Light Green	Red	Red	Light Green	Dark Green	Dark Green

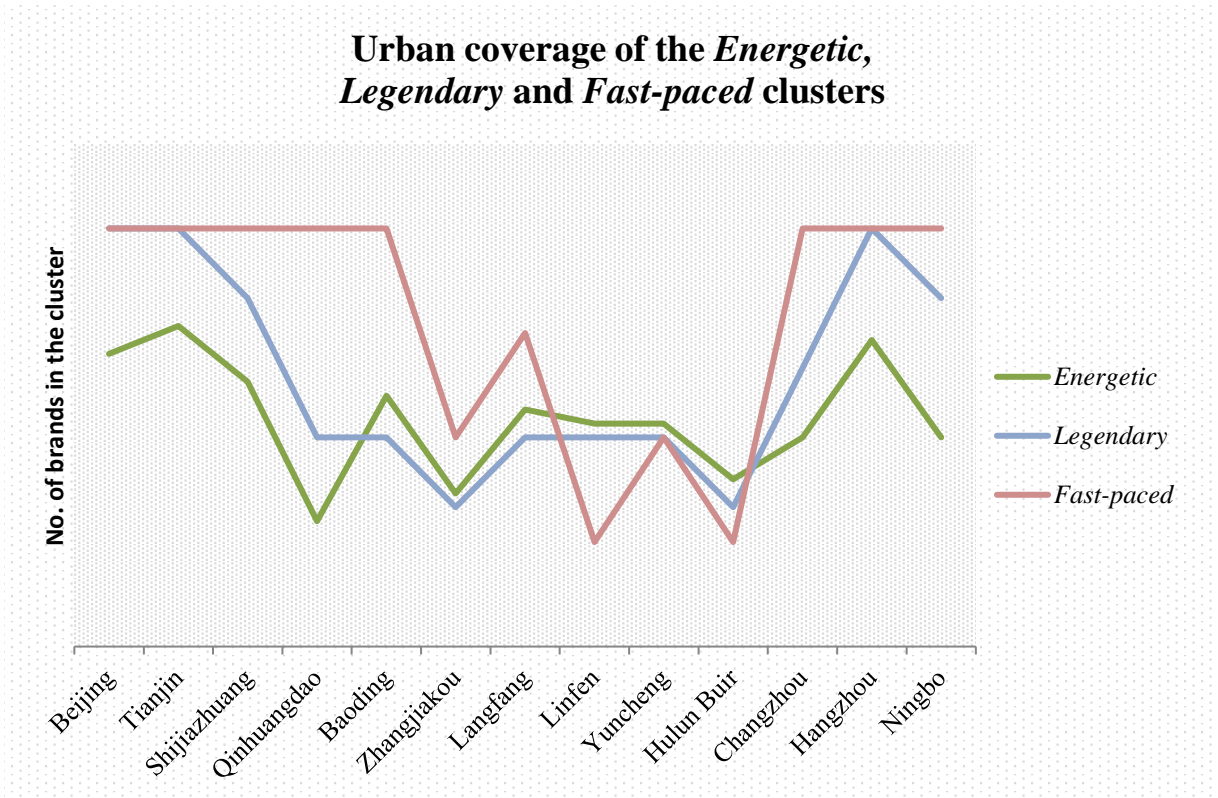


Chart. Example of urban coverage of the *Energetic*, *Legendary* and *Fast-paced* clusters

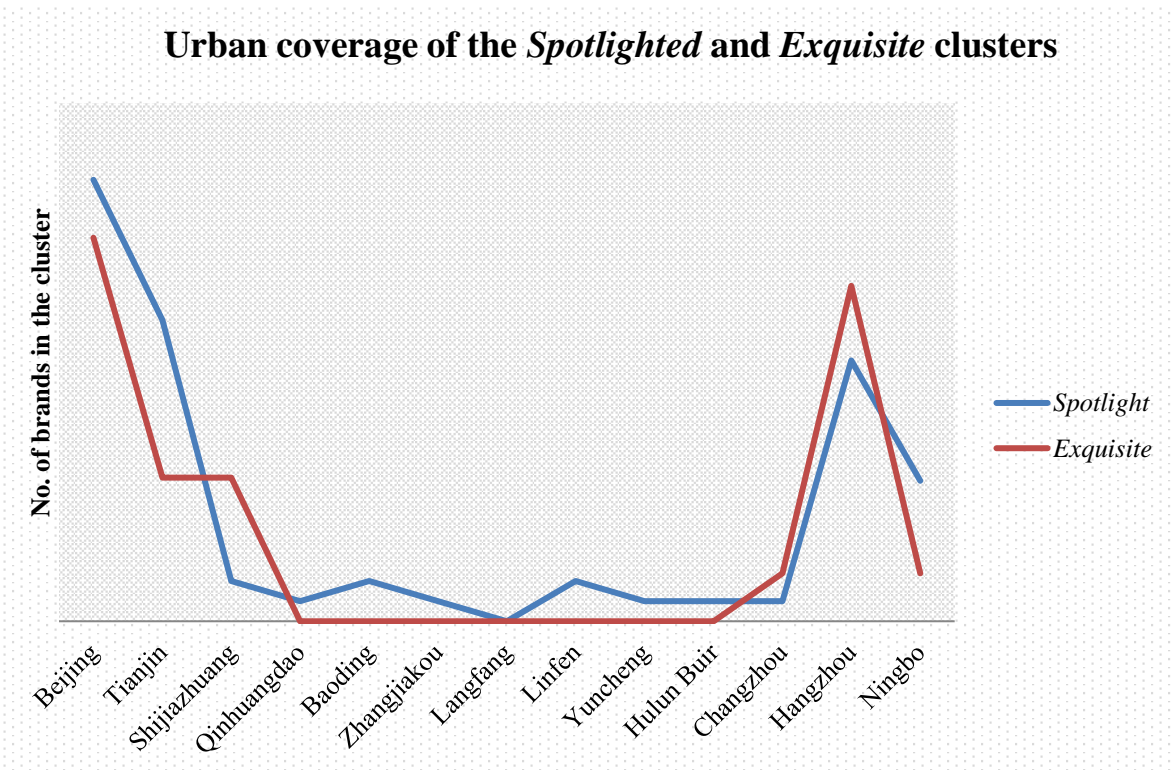


Chart. Example of urban coverage of the *Spotlighted* and *Exquisite* clusters

2.2 Comparison between Scales of Brand Clusters in Different Cities

Given each cluster involves different categories, we designed a standardized algorithm to calculate the standardized scores of the total brand sales in each cluster, in order to remove the influence of disparity of per customer transaction in categories. In this way, it is possible to objectively compare the performance of each cluster in different cities. Besides, use of logarithm processing makes the percentage of each cluster in different cities becomes comparable.

In order to optimize the comparison between cities, we introduced the *Line of Hu Huanyong*. This is a line drawn to tell the difference of China's population density and was created by geographer Hu Huanyong (1901-1998) in 1935. The line was originally proposed for demographic geography. East to the line accounts for over 95% of domestic population. Afterwards, the line was also taken as the boundary to discriminate between cities of different levels of urbanization. East to the line presents a performance better than the average level of the whole country, while west to the line stands for lower level than the national performance.

Furthermore, visualized methods are adopted to show the performance of brand clusters among different cities. Figure 2.2-1 displays what the *Youthful* cluster looks like in each city. The results tell us a city in Sichuan province to the left of the Line of Hu Huanyong line plays a most prominent part. This also manifests the unevenly development of different clusters in different cities. In general, to the east of the line, the *Youthful* cluster seems more robust.



Figure 2.2-1 Performance of the *Youthful* cluster in each city

Figure 2.2-2 shows the performance of the *Spotlighted* cluster all over the country. Some cities around the Bohai Bay, the Yangtze River Delta and part of Southwest China display higher coverage and dense distribution. Same as the *Youthful* cluster, cities on the right of the line have better performance, while the left side looks looser and smaller.



Figure 2.2-1 Performance of the *Spotlighted* cluster in each city

3 Similarity between Cities

Urban Similarity constructs an evaluation system of 3 dimensions based on a certain benchmarked city. The dimensions include urban infrastructure and living facilities, urban vitality and consumption environment. Upon the comprehensive assessment of the similarity between alternative cities and benchmark cities, we evaluate their similarity in urban development and business environment.

(1) Infrastructure and living facilities mainly consist of 6 aspects and relevant points of interest (POIs) have been listed as below:

- Fundamental living services: convenience stores; supermarkets; public restrooms; stores for appliances, building materials; pharmacy, hospital and etc.;
- Public transportation services: subway, bus stations and etc.;
- Education and training services: middle schools, primary schools, kindergartens;
- Fashion services: pet grooming, animal clinics, cars decoration, bars, nightclubs and etc.;
- Financial business services: banks, insurance companies, security companies, accounting firms, law firms and etc.;
- Exhibition and media Services: conference centers, convention and exhibition centers, exhibition halls

N.B. Coverage density of each POI is used as a measure.

(2) Urban vitality mainly includes 3 basic economic indicators as well as a series of indicators on service facilities. After expert discussion and confirmation, growth rates of GDP and of fixed assets investment, and increase in the number of visitors received become the three main indicators. The supporting indicators regarding service facilities are listed as follows:

- R&D innovation services: universities, incubators and shared work space;
- Other exhibition and media Services: newspaper offices, publishing houses and advertising companies;

- Other financial and business services: Hotels;
- Stylistic leisure services: parks, fitness, sports venues, scenic spots, bookstores, cinemas, theaters, art galleries, concert halls;
- Other education and training services: art museums, science and technology museums, libraries, training institutions.

N.B. Coverage density of each POI is used as a measure.

(3) Evaluation of the consuming environment

Contrast the structures of the 18 brand clusters in candidate cities with that in benchmark cities, in order to understand the degree of similarity between the two kinds of cities.

For example: By setting Shanghai as the benchmark city and with the help of the indicator system, we compare multiple cities with Shanghai and obtain the score of each system of “most alike Shanghai” (referring to the similarity degree between the candidate city and Shanghai)

- Benchmark city: Shanghai;
- Candidate cities: Nanjing, Hangzhou, Wuhan, Suzhou, Changzhou, Wuxi, Xi’an, Fuzhou

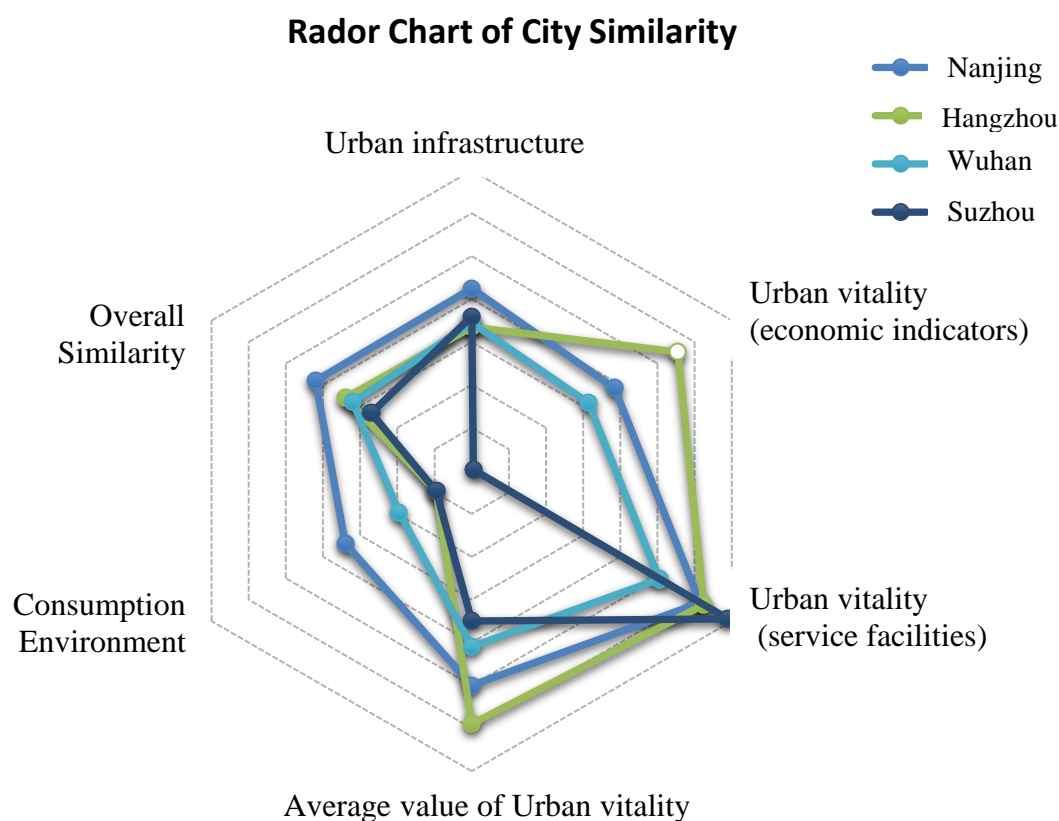


Figure. Similarity between candidate cities and Shanghai

As can be seen from the above figure, four of the eight candidate cities are close to Shanghai, with Suzhou being the most similar, followed by Wuhan, Hangzhou and Nanjing. In different dimensions, they reflect how they are connected to Shanghai.

The other four cities are somewhat similar to Shanghai. However, in some dimensions, the differences are obviously shown as below.



Figure. Illustration on where the 4 other cities look unlike Shanghai

4 Study of City Evolution

4.1 City Evolutionary Mapping

As mentioned, different cities display diversified coverage rates of brand clusters, and cities at different development phases constitute different grades of brands. Let's take Hulun Buir, Shijiazhuang and Beijing for example:

In Hulun Buir, brand clusters such as *Energetic*, *Prevalent* and *Legendary* are of prominent performance. The former two clusters are not so expensive while the latter one is made up of early luxury brands in China. Comparatively, Shijiazhuang has a different situation, the coverage ratios of the *Energetic*, *Prevalent* and *Legendary* clusters are above 60%, while brand clusters like the *Exquisite*, *Stylish*, *Spotlighted* and *Athletic* which represents personalized options and fashion trends look just like at an initial start. In regard to Beijing, all the above clusters look more saturated, among which the coverage of *Legendary* and *Fast-paced* already reach a percentage of 100%.

Accordingly, we assume that whether there exists an evolutionary relationship between cities. **This sort of "evolution" indicates a maturation of urban commercial environment. It represents a raise of citizens' spending power, an upgrade of consumer cognition or even a progress of sense of style. At a macro level, it also signifies the development of urban economy.**

Over 300 cities were put in an ascending order based on the number of brand clusters each city has. Cities with an equal number of clusters are classified into a same row, and thereafter arranged in an ascending order based on the average price of each cluster. If the make-ups of clusters are also identical, the cities will then be arranged in an ascending sequence of the number of brands each cluster has.

In short, cities across different rows are lined according to the number of clusters, while cities within a same row are arrayed due to price level and the number of brands. In the light of such logic, we can organize approximately 300 cities in the shape of a fish tail.

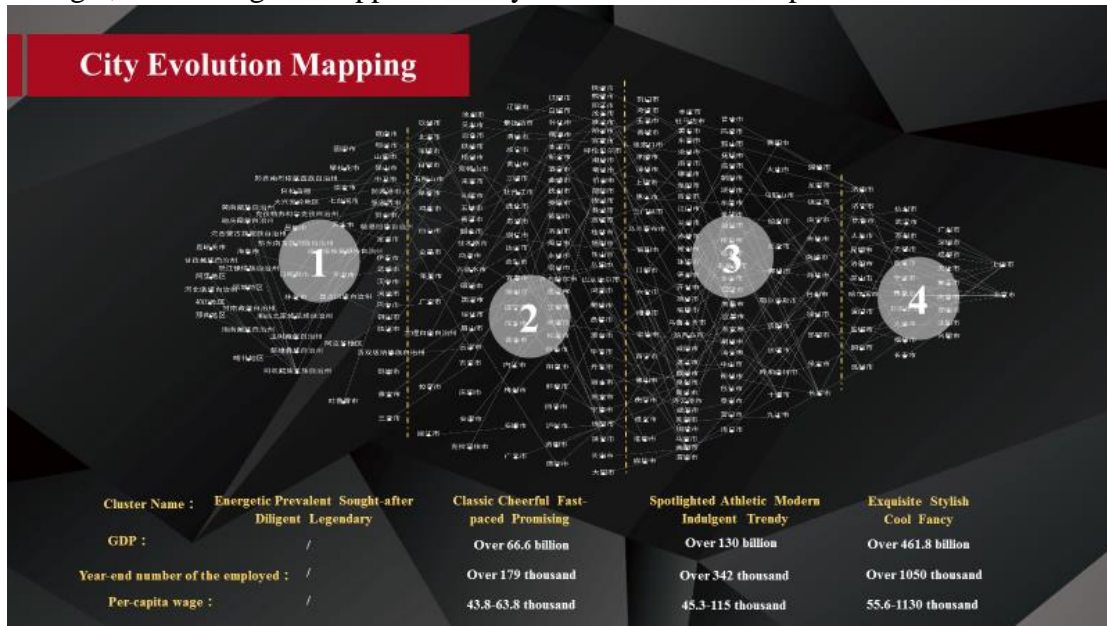


Figure. City Evolution Mapping

Cities are ranked due to maturity of their consumption environment and thereafter divided into 4 stages (see above picture from left to right). At each of the stages exist 4-5 clusters which on average have a coverage of 30%-50%.

We find that the way we split the cities into different developmental stages corresponds to the conclusion that macro-economic statistics of different cities will reach. Therefore, we believe the supposed “city evolution theory” based on brand cluster is valid and verifiable.

4.2 City Evolutionary Paths

So now we have got a City Evolutionary Map, then we need to work out the evolutionary direction of each city, i.e. how cities at lower levels are connected to those at upper levels. We would like to draw paths of evolution between cities based on below two factors:

- (1) **The brand clusters each city incorporates.** For example, City A owns 3 clusters and City B has exactly the same 3 clusters besides one more cluster. In this case, it is highly probable that A should evolve into B.
- (2) **“Gravity Formula”, whose concept was borrowed from physics.** When a city at a lower level seems to have several evolutionary directions as alternatives (for instance, City A is not only likely to evolve into City B but also into City C), “gravity formula” is adopted to help make a better choice.

We find that the closer the economic communications between two cities, the more similar their brand cluster structures. In the meanwhile, the closer the geographic distance between two cities, the more similar their cultural backgrounds, which also contributes to the similarity between the brand clusters that people of the two cities may consume. In this regard, the “gravity formula” borrowed from physics works out in “city evolution”.



Figure. City Evolutionary Paths

City evolutionary Map and Paths are quite useful for city siting of brand owners. This will be explained in detail in the following.

IV. Example of Practice

1 Dilemma of Data Used in Traditional Expansion Decision

When an enterprise plans for expansion, it usually needs to work out target cities and specific strategies used in the cities. This usually consist of 3 key jobs: 1) Locating the most valuable city candidates which are best appropriate for the brand owners to enter into or enlarge their operating areas in; 2) Prioritizing the candidates according to their values; 3) Comparing to benchmark cities where brands have performed well and making strategic decisions on the markets of target cities.

Traditionally, in a thinking logic from macro level to micro, the enterprise is likely to collect information of the 3 things respectively: GDP and population of target cities, current and future investment environment of the industry to which the brand belongs, as well as competitive analysis of the target market and demand analysis of the target consumers.

However, in this way, there arise some dilemmas that the enterprise has to overcome. First, decisions supposed to be based on data may turn into experience-based judgement under influence of personal preference and knowledge base. Second, in addition to lack of standardized and systematic methods to make statistical analyses of the concerned factors, such as developmental levels, spending power and market demands, there is also lack of comparability of data of different time periods. This leads to a system of decision making unadaptable to market dynamics. Third, no sufficient or accurate data can be acquired because of temporal or financial limits.

Therefore, strategic decisions of the enterprise tend to become balances between qualitative experience and quantitative data. This cannot ensure the precision of the decision.

2 City Siting Decision based on Brand Cluster

The *City Siting Methodology* based on brand cluster, created by Dataway, can alleviate the above situation to some extents. Consumer big data was applied in *brand cluster* to analyze variety and vitality of brands, characteristics of pattern of urban consumption as well as city evolutionary paths. These data results are of great help for enterprises to devise strategic goals and optimize resource allocation.

Compared to traditional methods of brand strategy design, brand cluster analyzes urban consumers in a medium-level view, which brings about more efficiency and effectiveness. Also, compared to traditional consumer demand analysis, urban evolutionary analysis based on brand cluster and “adsorption vs spillover” analysis of brand consumption can give a more clear and accurate depiction of latent and unsatisfied demands of consumers.

Specifically speaking, the three main jobs of City Siting Methodology constitute the following key judgement.

First, locating the most valuable target cities requires answers to 3 questions: **whether the cities are in a suitable developmental phase, whether spending power of the cities is suited for the target brand, and whether the external market environment is friendly.**

Second, optimizing the city candidates according to their values, focuses on 2 key issues: *City Evolution* and *Time Line* of target brand cluster consumption. In regard to City Evolution, as mentioned before, we studied the evolutionary directions of cities. For example, Hangzhou and Wuhan are more likely to approach the consumption levels and characteristics of Shanghai, while Xi’an and Chongqing tend to come after Beijing. With regard to Time Line, we pay attention to the consumption changes of the cluster where target brand belongs with time advancing.

Third, benchmarking and strategy devising requires us to **construct an index system to detect the most reasonable benchmark cities.** This demands data of patterns of brand clusters, data of urban infrastructure and living facilities, as well as data that depict urban vigor (which include fundamental economic indicators & data of urban service facilities pointing to upper level of Maslow’s hierarchy of needs).

Now let’s take the first job “locating the most valuable target cities” as an example, and see how the strategic decision on city siting works out in reality. It is worthy of emphasis that the key point here is the **timing for entry.**

Regarding “whether target cities are in a suitable developmental phase”, it resolves the concern if the cities satisfy the development of the brand cluster. The adaptability between cities and brand clusters result in the fact that some brands grow rapidly in certain cities.

As discussed before, Dataway divided 300 or so cities in China into 4 stages with each stage represented by 4-5 typical brand clusters. “Typical” refers to the status of fast development. For example, the *Spotlighted* cluster commences to grow in cities where GDP reaches 130 billion. Instead, brands in the *Stylish* cluster start to spring up in volume at the point of GDP reaching 461.8 billion. Therefore, when we figure out how brand development

is related to city development, brand cluster theory can be adopted to position what phase of cities a brand fits in.

Regarding “whether spending power is suited”, it examines if consumers in target cities have already made a large amount of purchases on the target brand cluster and how strong their consuming intentions are.

Take the *Legendary* cluster consumed in Shanghai for example, the below picture gives out information on at least 3 aspects:

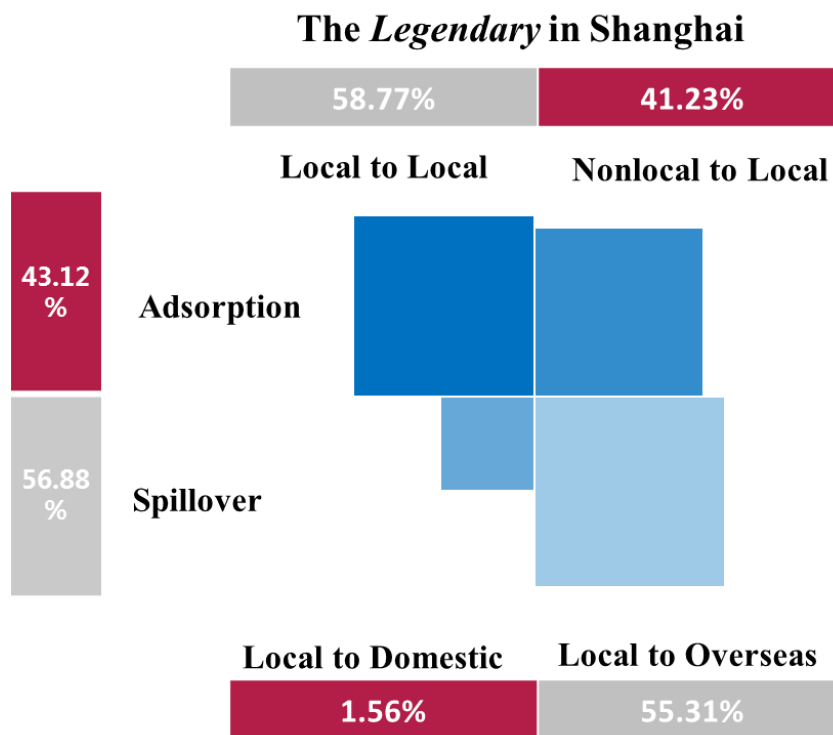


Figure. “Adsorption-Spillover” analysis of the *Legendary* cluster in Shanghai

- (1) High level of spending power of local consumers;
- (2) Plenty of consumption at both local and overseas places from local consumer;
- (3) A large amount of consumption from non-local consumers on brands of the *Legendary*.

The above information identifies Shanghai as a place with apparent adsorption and spillover effects on consumption in the *Legendary*. For brands in this cluster, consumers moving to Shanghai should be made good advantage of regardless of where they come from, and in the meanwhile, Shanghai is worth being taken as an important source to export brand image and culture. When the amount of adsorption is smaller than spillover, we claim this is a favorable case for brand entry.

Last but not least, “whether the external environment is friendly” tells us to what extent the cluster has covered the city, which means we need to know if a city is worth entry, according to the performance of other related brands in the same cluster.

Take Under Amour in the *Spotlighted* cluster as an example, the coverage of the *Spotlighted* in Jining and Jiaozuo is 7% and 2.3% respectively. The figures indicate Jining can provide more market opportunities for Under Amour and its related brands.

On the basis of the above 3 judgement, we can distinguish those with the most appropriate timing among 300 or so cities. As such, if we continue to implement Job 2 and Job 3, we will find a series of satisfying answers to the question of brand expansion and city siting.

V. Prospects

Having reviewed the modelling process, theoretical framework and application, based on our own cognition and feedback from customers, we believe that Brand Cluster can be further reformed as follows:

1 Improvement and Update of the Modelling Process

First, in terms of data cleansing, collecting brand-related information and extracting relevant transactions can be refined and optimized, which can contribute to improve accuracy and reliability of subsequent analyses.

Second, in terms of categories and brands included, more can be added. Moreover, the coverage of brands in each category can be thereafter checked and updated.

Third, in terms of clustering algorithm, that the transaction data should be standardized can lead to a more scientific and revealing correlation matrix.

Fourth, an ensemble of data from different sources can be considered, which could optimize and generalize the results. For instance, when classifying and naming brand clusters, we can combine data from traditional consumer survey or search data from dianping.com, a leading Chinese shopping review site.

2 Inclusion of new brands and its Discriminant Models

It is necessary to find a reasonable way of data collection (for example, using Dadui, a customer survey app based on social interaction devised by Dataway) and an efficient discriminant model for inclusion of new brands, to help explore brand correlations and improve the model's extendibility and applicability.

3 Future Application

In the future, we plan to try different angles of applications of Brand Cluster. It can be used, for instance, to monitor the annual performance of a brand, such as brands moving from one cluster to the other, and measure changes in customer base, brand image and reputation, and to explore possible solutions to such changes. These are all recommended applications.

References

- [1] Wu Jin, Li Dongjin & Du Liting.(2015). Formation and development mechanism of luxury brands, based on multiple case studies of top-scale European luxuries with a history of over 150 years. *Nankai Business Review*, 2015(06).
- [2] Ma Xiangyang, Liu Xiao & Jiao Jie.(2014). Trans-regional brand development new strategies, under the umbrella of regional brand unions. *Soft Science*, 2014(01).
- [3] Wang Caiyu.(2013). Implications, formation mechanism and influence on the “self-brand” correlation of consumers. *Advances in Psychological Science*, 2013(05).
- [4] Lei Liang, Peng Zhen & Li Hong.(2015) Application of big data in regional brand marketing. *Library & Information*, 2015(02).
- [5] Yang Nannan.(2015).The influence of business district where international “fast fashion” brands choose-A case study of Beijing market. *Contemporary Economics*, 2015(03).
- [6] Liu Wei. (2018). A study of discrimination against local brands within CBD in Beijing and Shanghai. *Consumption Daily (Business Version)*, 2018/1/7.
- [7] Zonghe.(2017). Changes in business districts like Nanjing Road and Huaihai Road: Shops closed and upgraded to introduce new fashionable brands. *Shanghai Business*, 2017(03).
- [8] Xu Pengfei, Sun rong, Tie Haoyue, Yang Pei & Han Sige.(2015) A study of brand strategy in interdisciplinary marketing. *Management and Science of SME*, 2015(12).
- [9] Jiang Wenjun.(2006). Management of brand relationship in multiple-brand strategy. *Commercial Times*, 2006(30).
- [10] Huang Jingsong, Zhao Ping.(2005). Application of clustering analysis in brand positioning. *Application of Statistics and Management*, 2005(01).
- [11] Lei Jing, Li Xia.(2014) A research on market segmentation based upon factor analysis and clustering analysis-A case study of an e-commerce ladies’ brand in Jiangsu. *Journal of Nanjing University of Posts and Telecommunications (Social Science)*, 2014(04).
- [12] Zhang Mengxia, Feng Yuefeng. A clustering analysis of attributes of SCE brands with appeals to print ads of international luxuries. *An international academic conference on information technology, service science and engineering management*
- [13] Chen Jimeng, Chen Jiajun, Liu Jie, Huang Yalou, Wang Yuan & Feng Xia.(2015) Large-scale social network clustering algorithm based on structural similarity. *Journal of Electronics & Information Technology*, 2015(02).
- [14] Zeng Cheng, Sun Yaqian, Xu Yuzhu & Zhang Damin.(2015) An overview of optimized complex network clustering methods. *Communications Technology*, 2015(08).
- [15] Lovro Šubelj, Nees Jan van Eck, Ludo Waltman.(.).*Clustering scientific publications based on citation relations: A systematic comparison of different methods*. PLoS ONE 11(4), e0154404 (2016).

