Social Listening Analytics in Chinese Car Market

Group RiceBall

Member:

3035674812 ZHOU Zezhong
3035675359 NI Mengmeng
3035675969 WANG Jingting
3035674680 LIN Hongying
3035675062 ZHUANG Danyun

THE UNIVERSITY OF HONG KONG
Faculty of Business and Economics

1 Project Idea

There is a popular saying that "social media is the world's largest focus group". As a result of the booming digital age, social media platforms contain a lot of valuable and easily-accessed user-generated information. By analyzing this information, we can get to know users' thoughts, sentiment, interests, and even their personalities. Therefore, collecting and analyzing the contents from the social media platform is a great way for a company to know and understand your current and potential customers.

In this project, we act as a company that provides consultancy services for the car manufacturer. We use Social Listening Analytics to help the company to get insights from popular social media platforms and support its decisions on product development, customer management, and marketing strategy in China. This time, our client is Mercedes-Benz and we focus on analyzing two of its models Mercedes C-Class and E-Class in Chinese Market. The average prices of C-Class and E-Class are ￥330 thousand and ￥530 thousand, respectively, representing Mid-End and High-End models in its company.

We will implement our analysis in 7 dimensions: Market Overview and Competitor Analysis, Purchase Purpose Analysis, Topics modeling, Product defect and advantage, Comment sentiment analysis, Customer Service and Retention, and Customer Portrait.

The analysis methods we used included: Correspondence Analysis, Co-Occurrence Analysis, TF-IDF, Clustering, Sentiment analysis, Topic modeling. Given the descriptive features, we found from the data, we can advise the company to make specific improvements based on (potential) customers' feedback.

2 Data Description

The information on the Internet also follows the Pareto Principle, e.g. the 80% of valuable content is in the Top 20% websites. We aim to collect the customers' feedback from several very popular car forums in China:

1. Autohom.com: Established in 2005. The world's most viewed car information website according to the statistics of iUserTracker.
2. XCar.com: Established in 2002, up to 7.5 million averaged daily pageviews.
3. PCAuto.com: Established in 2002, aims to provide the most professional information about cars to users.

From the above 3 websites, we will use Selenium to collect User Generated Content in 3 different aspects:

1. Comments and evaluations on our target car models such as appearance, power, controllability, and service quality.

2. Comments and evaluations on the similar car models of our main competitors.

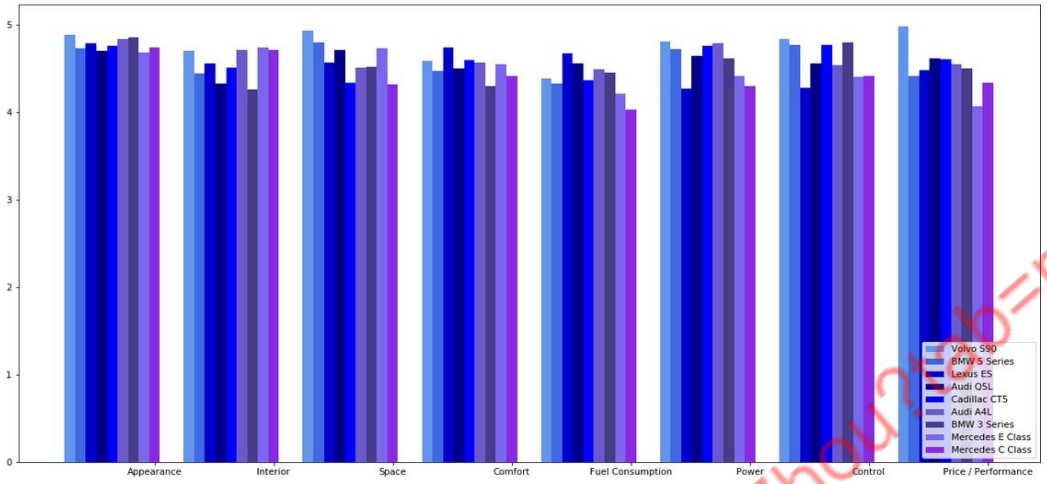
3. The information of users who are interested in our targeted car models, such as gender, location, and other interests.

Besides, we also use Selenium to collect the account information from Weibo.com. A detailed description sheet of each variable in our data is attached in the Appendix.

3 Analyze

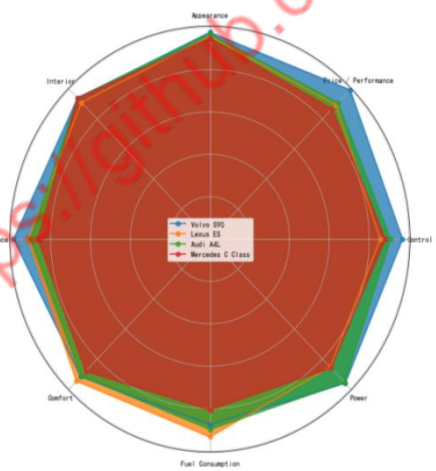
3.1 Market Overview and Competitor Analysis

There are several similar players in the market: Volvo S90, BMW 5 Series, Lexus ES, Audi Q5L, Cadillac CT5, Audi A4L, and BMW 3 Series. Their rating on the website are shown in graph 1:

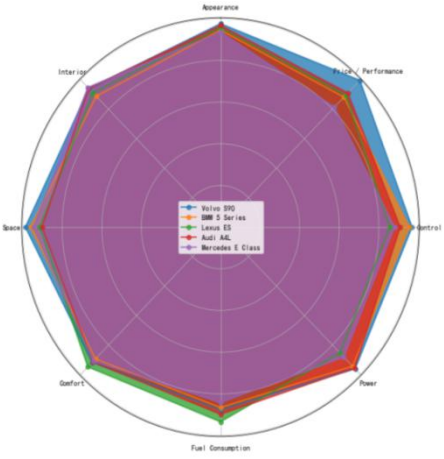


Graph 1

As it's impossible to be the NO.1 in every aspect, we believe Mercedes need to base on its advantage, and further catch up with other cars with similar advantages. Mercedes C and E class both have high rates of the interior. Mercedes E also has a high rate of space and comfort. For Mercedes C, as the interior is its advantage, we will further compare it with those cars that also have a high rate of the interior(Volvo S90, Lexus ES, Audi A4L). For Mercedes E, as interior, space, and comfort are its advantages, we will compare it with Volvo S90, BMW 5 Series, Lexus ES, and Audi A4L.



Graph 2



Graph 3

Finding: As the graph on the left shows, Mercedes C Class needs to catch up with Volvo S90 in 'Price / Performance', 'Control', 'power', and 'space', catch up with Lexus ES in 'Comfort' and 'Fuel Consumption'.

As the graph on the right shows, Mercedes E Class need to catch up with Volvo S90 in 'Price / Performance', 'Space' and 'Control' and catch up with Lexus ES in 'comfort' and 'fuel consumption'. Also Audi A4L is better at 'power'.

3.2 Consumer Purchase Purpose

In the "consumer purpose" analysis, as mentioned before we select 2 sets of automobile brands with different consumer target and price level: Volvo S90, Mercedes-Benz E-Class, BMW 5 Series, Lexus ES, and Audi Q5L as the analysis object sets for Benz E-class, and Cadillac-CT5, Audi-A4L, BMW 3 Series, LEXUS-ES, Benz C-class as the analysis object sets for Benz C-Class. And we want to know what differences consumers have in the using scenarios of these five automobile brands in the same set, which is also a more concerned aspect of automobile manufacturers - Their product positioning depends on whether the positioning in the consumer's mind is consistent, and whether the propaganda strategy needs to be strengthened or adjusted.

To understand the purpose of the customers' purchase from the perspective of use scenario, we extract the purchase purpose tags each comment gives at Autohom.com. In the comments of the word-of-mouth channel, there is a field of "purchase purpose", which is a semi-structured option. Commentators can choose to fill in the application scenarios they like to purchase cars. We choose 8 most common tags from the official waiting options: Road Trip, Long-distance journey, Business trip, Commute, Take or send kids, Shopping, Racing, Off-road riding.

We extract each cars' purchase purpose tags of 300 comments, some of which contain more than one scenario(tags), and give a pivot table statistics on the tags, then organize into a cross-contingency table. The results are shown in the table below:

	Volvo S90	Benz E-class	BMW 5 Series	LEXUS-ES	Audi-Q5L
Road Trip	51.33%	38.00%	39.67%	38.67%	47.33%
Long-distance journey	23.00%	17.00%	22.00%	11.00%	18.33%
Business trip	31.33%	31.33%	28.33%	20.33%	17.67%
Commute	83.67%	78.67%	78.67%	88.67%	73.00%
Take or send kids	35.67%	22.33%	28.00%	33.33%	26.00%
Shopping	29.33%	20.00%	25.67%	26.33%	21.67%
Racing	0.67%	0.67%	1.33%	0.33%	2.00%
Off-road riding	0.33%	0.33%	0.33%	0.33%	5.33%

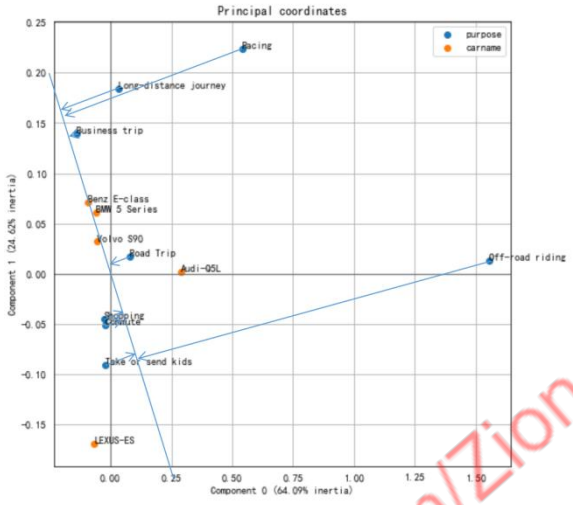
	Cadillac-CT5	Audi-A4L	BMW 3 Series	LEXUS-ES	Benz C-class
Road Trip	59.67%	49.33%	38.67%	46.67%	34.67%
Long-distance journey	13.00%	22.67%	11.00%	13.67%	10.67%
Business trip	11.67%	10.00%	20.33%	14.00%	6.33%
Commute	92.33%	90.33%	88.67%	88.67%	88.33%
Take or send kids	30.67%	29.33%	33.33%	25.00%	19.33%
Shopping	43.00%	30.00%	26.33%	39.00%	23.00%
Racing	7.33%	1.33%	0.33%	1.33%	1.33%
Off-road riding	2.00%	1.00%	0.33%	1.67%	0.33%

Table 1

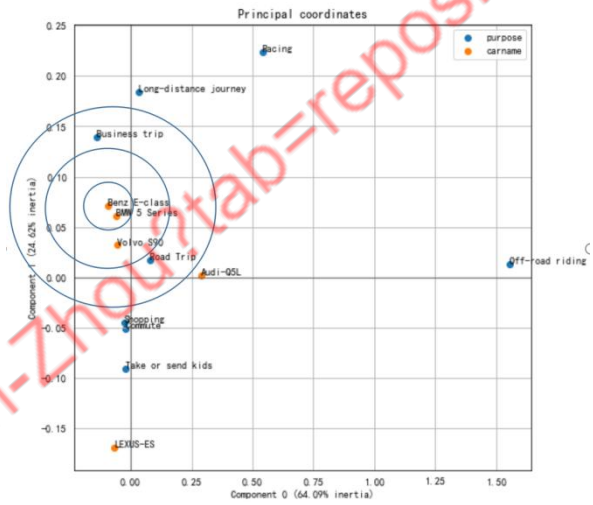
From the picture we can see:

- Commuting is the main purpose for all ten types of cars, especially for Volvo S90 and LEXUS-ES. The second purpose is Road Trip, particularly for Volvo S90, Audi Q5L, Cadillac CT5, and Audi A4L.
- For Benz E-class, commuting, road trip, and the business trip is its customers' main purchase purpose.
- For Benz C-class, commuting, road trip is its customers' main purchase purpose. Compared with Benz E-class, Benz C class is less business and more for daily use.

Only from these pictures, it's hard to get all the features of each car. Therefore, in order to reveal the differences between the purposes of the same car and the correspondence between the cars of different purposes, we use the Correspondence analysis method, which is a powerful data visualization technology and mainly used in market segmentation, product positioning, geological research and computers In engineering and other fields.



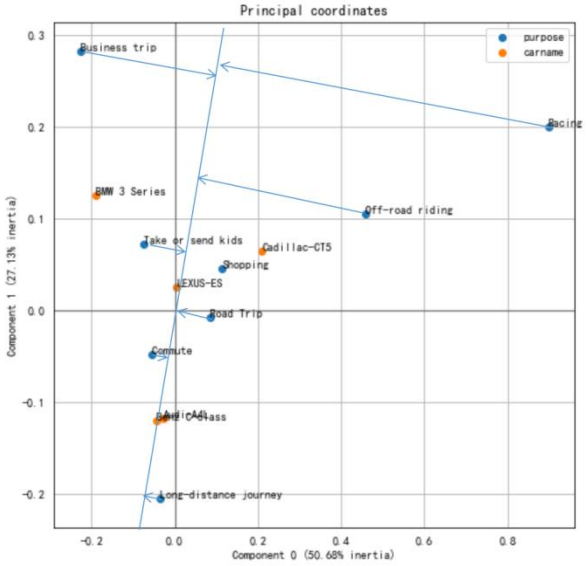
Graph 4



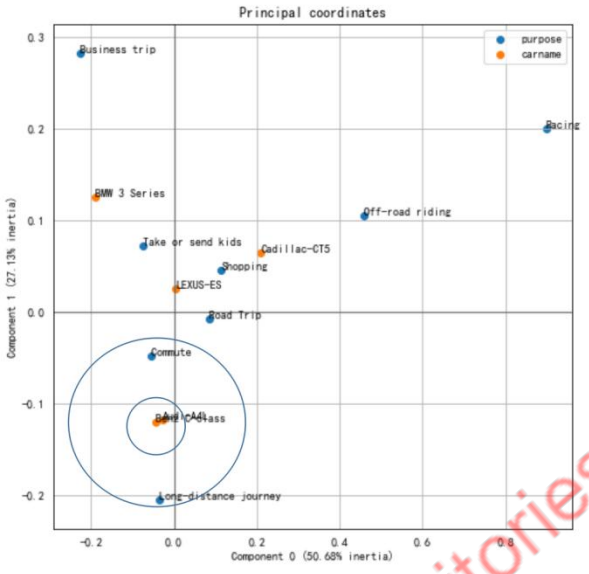
Graph 5

This Correspondence analysis picture above is for Benz E-class car set, capturing 24.62%+64.09%of the real features. From the picture we can capture the features of customer purchase purposes:

- From the perspective of purchase purpose or use scenario, BMW 5-series is the biggest competitor and then the Volvo S90, as they are closer to each other in the graph. Besides Mercedes, Audi, and LEXUS are quite different from the using scenario.
- If a customer wants to purchase a car only for business trips or long-distance journeys, we can obtain his preference for those cars in order: Benz E-class, BMW 5-series, Volvo S90, Audi Q5L, and last LEXUS ES.



Graph 6



Graph 7

This Correspondence analysis picture above is for Benz C-class car set, capturing about 78%(27.13%+50.68%)of the real features.

From the picture we can capture the features of customer purchase purposes:

- From the perspective of purchase purpose or use scenario, Audi A4L is the biggest competitor, as they are very close to each other in the graph.
- BMW 3-series is more for business trips, Benz C-class and Audi A4L is more for commuting and long journeys, while LEXUS ES is more for shopping.

Findings:

- Nearly all the customers purchase cars for commuting
- From the purchase purposes, the Audi A4L shares a similar customer use scenario with Benz C-class, while BMW 5-series shares a similar customer using scenario with Benz E-class.
- Main use scenario for Benz E-class is business trip

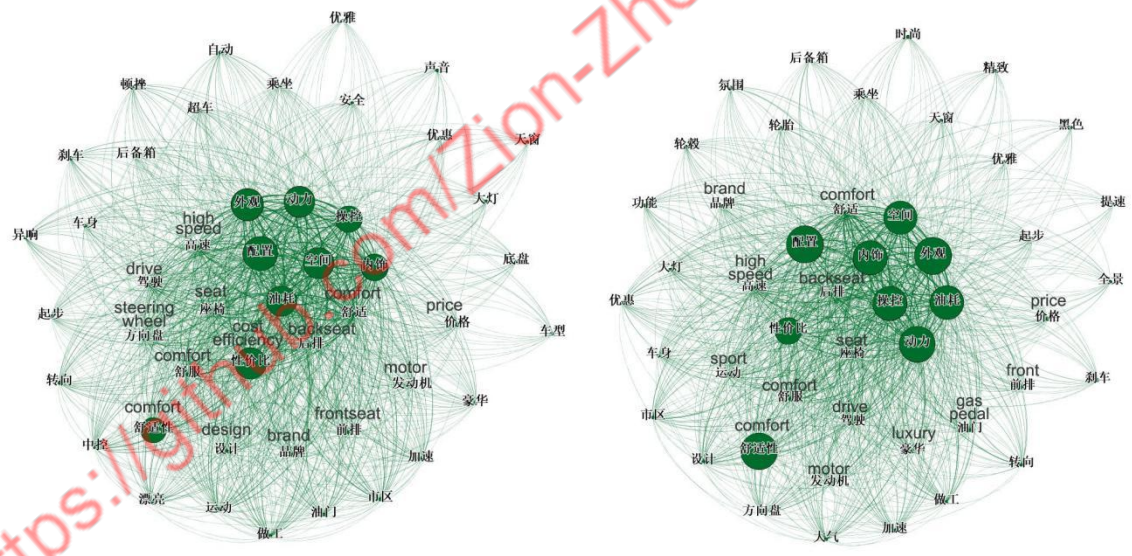
3.3 Topic Modeling

We extracted satisfied and unsatisfied comments from customers' reviews of Mercedes Benz C-class and made a text clustering analysis based on TF-IDF values using the DBSCAN method. For the satisfied comments, we got five clusters, from which we conclude customers are mostly pleased by safety, brand values, and power. Whereas the unsatisfied comments, people think improvements are needed to reduce noise and upgrade center control.

Graph 8 *Graph 9*

We did a similar analysis for Benz-E class. The figures reveal that customers are satisfied with brake and control. And they want to see some improvements in noise reduction and engine. (Figure attached in Appendix Graph 30-31)

Then we merged all text-type fields to do further text mining, and see the strength of association between keywords in the comments. We extract the TOP47 keywords with the highest TF-IDF. Then, according to the lexical succession relationship (chronological order), we do vocabulary co-occurrence analysis, and finally use Gephi to generate the following figure:



Graph 10 *Graph 11*

All dots with a large green background are keywords the system gives customers to separate comments. The more green lines mean there are more connections between the word pairs. And we can see that 'comfort' is away from other system-given keywords, which means 'comfort' is a more independent factor when buyers are considering a car. There are also some words in the center of the graph except those system-given keywords, like 'seat' and 'backseat'. And 'comfort' appears really a lot.

Finding: We recommend Mercedes to focus on improving 'comfort' which customers care about and is a more independent feature. Also when comparing these two graphs, we can see that for class C buyers, they care more about brand and steering wheel. And for class E buyers, they care more about luxury and gas pedals.

3.4 Identify Vehicle Defects

The word cloud in the previous part is an analysis from a bird-eye view, which can seize the focus of massive comments in a short period. However, if we want to know more about the shortcomings that consumers are concerned about Mercedes C-Class and E-Class, that is, to dig out consumers' typical opinions on products' defects, we need to analyze the field of negative comments.

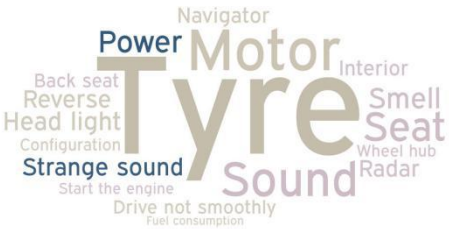
Here, we want to find out the most representative ones from 1643 negative reviews of C-Class and 1097 negative reviews of E-Class to Identify defects of both classes. The method we apply here is Text clustering with Sklearn's TFIDF((Term Frequency-Inverse Document Frequency) vectorizer and KMeans. First, we use the TFIDF-Vectorizer after getting rid of stop words to pull out unique words that can be used in clustering from each negative review. Since KMeans requires that we specify the number of clusters, we try K=1 through K=80 and record testing error to look for the "elbow". Here we choose 10 clusters. Lastly, we cycle through the clusters and print out the top 10 keywords based on their TF-IDF score of each cluster to see if we can spot any trends. By locating the keywords in the whole negative review, we finally extract one representative opinion of each cluster. The result (Take 3 keywords from C Class for instance. The complete result of C and E Class is attached in the Appendix table 7 and 8) is shown below.

Cons(C CLASS)	Keyword
The landing price in many cities is much cheaper than ours, this is too pitted! **The reversing radar is unresponsive, and the reversing vehicle has encountered a tree once.** The steering wheel is heavy.	Radar Reverse
The speed is a bit slow at the start, and the steering wheel is a bit heavy at low speed. Secondly, **fuel consumption is higher.**	Fuel consumption
The function buttons are a bit messy, so you have to look carefully. **The back-seat cushions are a bit short and not very comfortable,** and some mid-range parts squeak during driving. **The space inside the car is not particularly large!**	Space; Back-seat

Table 2

In order to identify the defects more intuitively, we extract the top 20 keywords with the highest TF-IDF value throughout the whole negative comments of both cars to make word clouds. As we can see from the two-word clouds below, for Class C, the defects mentioned the most is space(space, back seat and seat) and sound(sound and strange sound), and for Class E, the defects mentioned the most is tyre, motor(motor and power) and sound(sound and strange sound).

Graph 12



Graph 13

Finding: The typical comments and defects above should be an important consideration while determining product improvement for the car manufacturers, which could greatly enhance the user experience and satisfaction.

3.5 Identify Vehicle Advantages

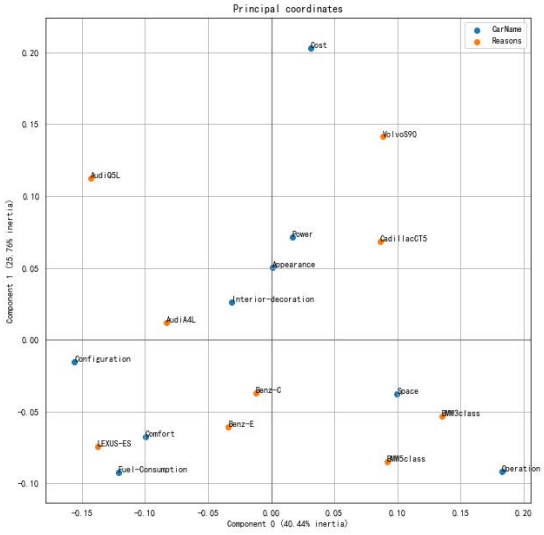
In the vehicle advantages identification analysis, we choose the same 2 sets of automobile brands as the 2 sets in 3.2. And we want to know what perspectives consumers have on the vehicle features advantages of these different brands' vehicles, which is also a more concerned aspect of automobile manufacturers - They can identify their advantages on vehicle features, where their criticized, and competitors' with similar features, to obtain their product positioning, whether the vehicle features should be strengthened or adjusted for catering consumers preference and creating competitive advantages.

In order to understand the customers' perspectives of cars' performance, we extract the performance score each comment gives at PCAuto.com and Xcar.com. In the comments, commentators can rate 10 aspects of the car's performance, scale from 1 to 5: Appearance, Interior-decoration, Space, Configuration, Power, Operation, Fuel-Consumption, Comfort, Price/Cost, Off-road Riding.

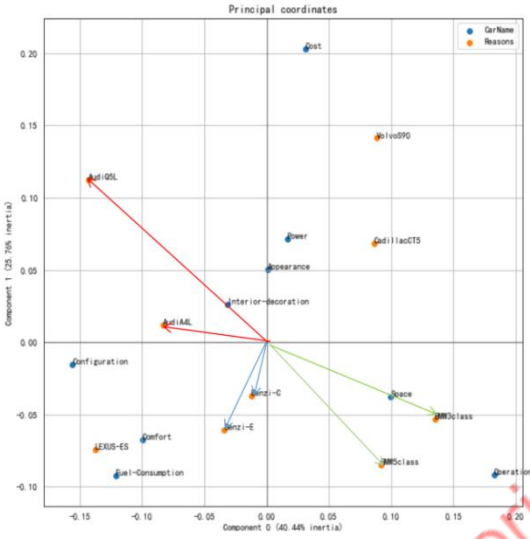
Similar to what we did in 3.2, we still give a pivot table statistics on the scores of five, organize into a cross-contingency table, and then use Correspondence Analysis to recognize the competitors and advantages.

CarName\Reasons	Benz-E	Benz-C	AudiA4L	AudiQ5L	BMW3class	BMW5class	VolvoS90	LEXUS-ES	CadillacCT5
Appearance	74.60%	66.98%	77.05%	73.70%	72.05%	75.94%	74.27%	75.50%	86.99%
Interior-decoration	58.78%	67.08%	50.95%	46.70%	47.95%	45.93%	60.67%	61.50%	53.57%
Space	65.54%	47.53%	56.75%	45.55%	77.75%	75.30%	70.36%	60.90%	49.82%
Configuration	36.27%	35.96%	51.90%	43.90%	34.50%	40.90%	33.90%	51.20%	27.27%
Power	52.72%	48.79%	62.00%	62.50%	59.55%	61.11%	56.35%	44.70%	56.33%
Operation	51.50%	59.39%	48.65%	30.90%	78.20%	73.25%	53.62%	51.00%	60.88%
Fuel-Consumption	49.73%	42.46%	53.25%	45.90%	50.00%	46.33%	32.57%	65.20%	36.76%
Comfort	56.94%	49.37%	43.90%	46.70%	44.45%	47.83%	40.17%	63.50%	38.86%
Cost	23.89%	25.28%	30.00%	35.00%	34.40%	24.43%	48.18%	31.90%	32.08%
Off-road Riding	0.00%	0.00%	0.00%	59.90%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 3



Graph 14



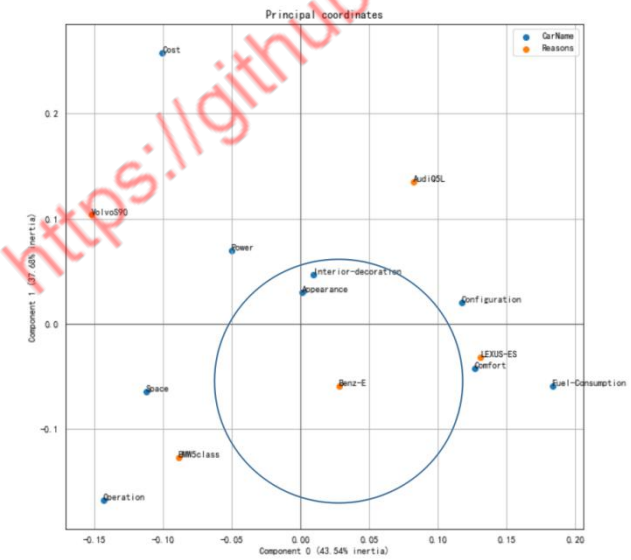
Graph 15

This Correspondence analysis picture above, we can capture the features of vehicles performance:

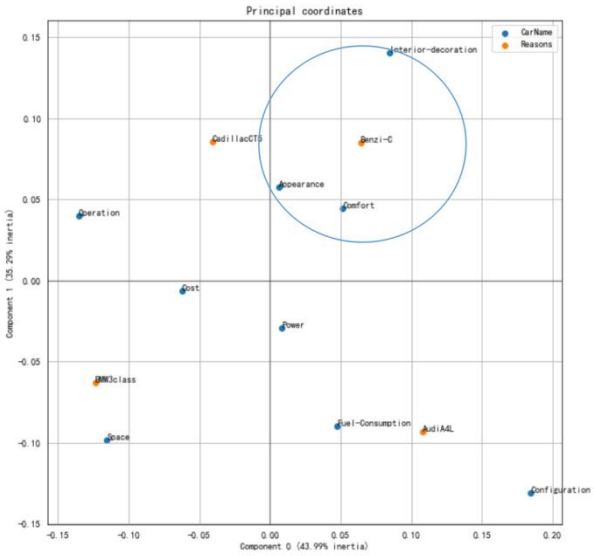
- Different cars under the same brand, tend to have similar performance regardless of slight differences. The angle between rays from origin to points of same brand cars, is quite small, representing the similarity of two cars' performances.
- The competitive edge of Volvo might be the cost performance, LEXUS: Fuel Consumption and Comfort, Cadillac: the Power, BMW: the Operation and Space.

It's not accurate or fair to just compare them at the same time, as they have different target customers with different price levels. Thus, below we separate those into two parts as we did before.

For Benz E-class, we compare it with Volvo S90, BMW 5 Series, Lexus ES, and Audi Q5L:



Graph 16



Graph 17

This Correspondence analysis picture on the left above is for Benz E-class car set, capturing 37.68%+43.54%of the real features.This Correspondence analysis picture on the right above is for Benz C-class car set, capturing 35.29%+43.99%of the real features.

From the picture on the left, we can capture the features of Benz E-class set vehicle performances:

- From the picture on the left, we can obtain that the main competitive advantages of Benz E-class are Appearance and Inter-decoration, and it is poor at cost performance.
- From the perspective of vehicle performances, its potential competitors are LEXUS-ES and BMW 5-series. In the picture, the Benz E-class point is far away from other brands, therefore Benz E is unique enough. However, it's edge is not close enough to the Benz E-class, therefore it could still improve the appearance or comfort from this perspective.
- The competitive edge of each brand: BMW 5-series: Space and Operation, LEXUS-ES: Comfort and Fuel Consumption, Volvo S90: Cost-performance and Power, Audi Q5L: Off-road driving Power, interior-decoration, Benz E-class: Appearance.

From the picture on the right, we can capture the features of Benz C-class set vehicle performances:

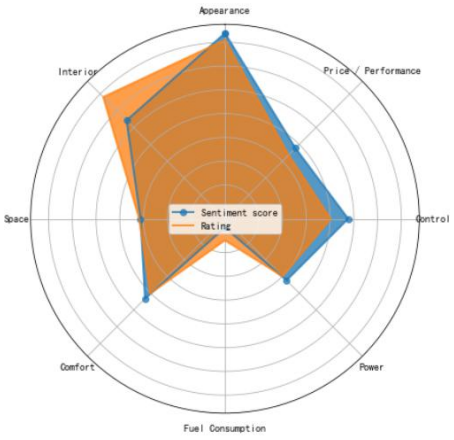
- From the picture on the right, we can obtain that the main competitive advantages of Benz C-class are Appearance, Inter-decoration, and comfort, and it is poor at Space. Compared with Benz E-class, it actually shows different positioning on vehicle performances.
- From the perspective of vehicle performances, Benz C-class's potential competitor is Cadillac. In the picture, the Benz C-class point is not very close to other brands' point, therefore Benz C-class might be unique enough and has its own niche market.
- Competitive edge of each brand: BMW 3-series: Space and Cost, Audi A4L: Fuel Consumption and Configuration, Cadillac: Appearance and Operation, Benz C-class: Appearance, Inter-decoration, and comfort.

Findings:

- Both Benz E-class and Benz C-class have its unique vehicle features or combination of features: Benz E-class is good at Appearance, not so good at Cost performance, Benz C-class is good at Appearance, Inter-decoration, and Comfort, not so good at Space and Configuration.
- Different cars under the same brand, seem to share similar great vehicle features, which might be the brand uniqueness.

3.6 Comment Sentiment Analysis

We further do sentiment analysis on the comments and compare the sentiment score with rating. First we separate comments into different aspects (Appearance, Interior, Space, Comfort, Fuel Consumption, Power, Control and Price / Performance). Then we use 'jiagu' and 'snownlp' packages to do sentiment analysis. After standardizing the sentiment score and rating, we compare them in the Radar chart:



Graph 18

Graph 19

From the left plot, the sentiment score and rating are similar to each other. And from the right plot, the sentiment score of Price/Performance is far higher than the Price/Performance rating and the sentiment score of Power is far lower than the Power rating. We further check these differences. For Power, customers discuss a lot about traffic jams in the city, causing the sentiment score to be low. And for cost efficiency, we found something interesting. There are some comments like: ‘Benz is not cost-effective. But it’s a childhood dream’ and ‘The last thing you should pay attention to when buying this car is cost efficiency ... I believe many E260 owners will not care about this?’. And both ratings are 2(Full score is 5). Even though buyers give low rates for cost efficiency, they still buy it due to the brand.

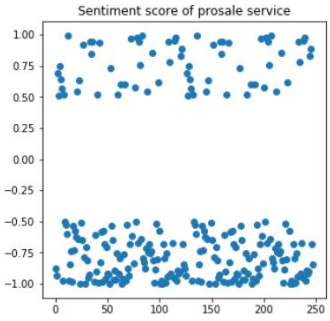
Finding: As a result, Mercedes don’t need to lower their price immediately, but they need to maintain the brand value.

3.7 Customer Service and Retention

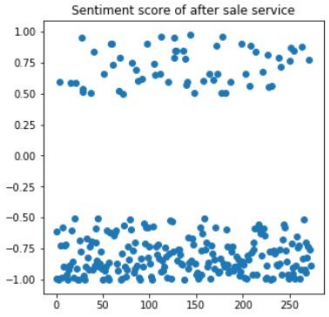
In China, consumers buy and maintain cars mainly through car dealerships(4s stores). The websites we scraped don’t have scores about service in 4s stores, so we use NLP to further analyze customer service.

As most 4s stores sell both Class C and Class E and provide after-sale service together, we combine Class C comments and Class C comments together. Among all customers' comments, we first extract comments about service, like "服务'(service), '售后'(After-sale service), '售前'(Pre-sale service), '维修'(repair),'保养'(maintain). Among all service-related comments, we further separate them into 'Presale service' and 'After-sale service'. After applying sentiment analysis, the positive sentiment ratio in after-sale service is lower than that in pre-sale service. Even though there are selection biases

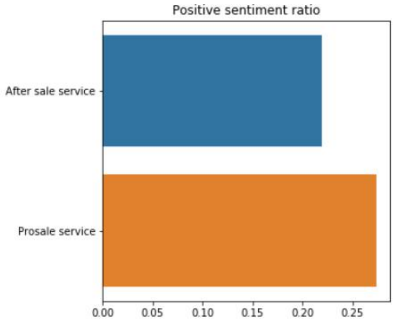
in service comments, comparing these two ratios, we can easily tell that after-sale service is much worse than pro-sale service.



Graph 20



Graph 21



Graph 22

Then we dive into comments related to after-sales service and see what Mercedes's dealerships can improve. We select the top 20 words with the highest TF-IDF value in pro-sale comments and after-sale comments.

From the word cloud on the left, we can tell that during pro-sale service, customers care about service quality and attitude the most (green color). Also, customers interact with salespeople a lot (blue color). From the word on the right, we can see that in after-sale service, customers come with purposes like repairing and maintaining cars, and Insurance (yellow color). Also, customers complained a lot about the high price (purple color).

Graph 23

Graph 24

Finding: Mercedes need to supervise salespeople to provide quality service with a good attitude during pre-sale service. In after-sales service, Mercedes and car dealerships need to lower the price of car repairing, maintenance, and accessories replacement.

3.8 Consumer Portrait and Interest Mining

Understanding customers' interests could help the company to adjust the brand image or promotion style to match the customers' tastes. An appropriate brand image could not only attract the right potential buyers but also retain the old customers who already accept you. One of the most effective ways to know someone's interests is to know what kind of information he or she is following or willing to know.

Graph 25

Sina Weibo has a function called Interest-Tag, which can be set by users themselves. Once you add one Interest-Tag in your account, Basketball, for example, the weibo will automatically provide you with a lot of corresponding information. Therefore, we can use the Interest-Tag to know what kind of information the user is interested in. In this section, we will analyze the interests of 6 car brands' customers(Audi, BMW, Mercedes, Cadillac, Lexus, Volvo).

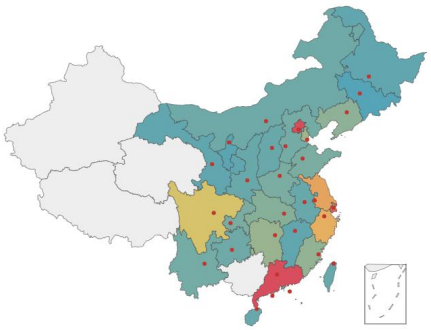
For each brand, we first find the users whose Interest-Tag contains our target and collect their other interests and the account's region. After collecting this information, exclude some noisy tags such as the brand itself and some are frequently mentioned in all brands. we apply the same method, Correspondence Analysis, on these cleaned tags data and get the graph 25.

As there are a lot of tags in the graph, we don't translate them into English but about some interesting findings from the graph. First of all, the points of Audi, BMW, or Mercedes are very close, meaning that people who follow them have very similar interests. For example, they like road trips, photography, and automobile repacking. The phenomenon is reasonable, as people in China always mention these three brands together and simply call them BBA(Benz, BMW, Audi). Most of BBA's cars are in business use, so it's also reasonable that their owners have relatively more costly interests as I mentioned above. Second, people who like Cadillac and Lexus have very different tastes from those who like BBA. For instance, we can see the tag Basketball is very close to the Cadillac and the tag freedom is close to both Cadillac and Lexus. Food, Fashion, and Music are also very close to Lexus. We can conclude that even Cadillac and Lexus are also the High-End brand, their target audiences are much younger than BBA's. We also create a word cloud for our client (Graph 26).

Graph 26

Like what we mentioned before, people like Mercedes also have huge interests in BMW, traveling, fashion. Thus for Mercedes' future advertisement strategy, when talking about its traditional car models, it's better to avoid the style of food, sports, or freedom, because based on our analysis, your target audiences have little interest in these fields. On the contrary, Mercedes can focus on presenting the style of high art or describing the comfortability of driving experience.

Besides, we also want to know the geographical distribution of these brands' online followers. Knowing the geographical distribution of your (potential) customer could help the company to improve its efficiency of promotion and customer retention strategy. We use the data from Weibo to get the geographical distribution of each brand's online followers and draw heat-maps to visualize the distribution Due to the space limit, we only attach the heat-map of Mercedes, the other heat-maps are attached in the Appendix part (Graph 32-37).



Graph 27

As the size of each province is too small, the number of followers and region names are not included in the heat-map. We created HTML files in our data folder for these maps which have an interaction function that shows the province name and the number of followers when your mouse moves to the region. The heat-maps represent the density of online followers from low to high(blue < yellow < red). From the 6 heat-maps(in Appendix), we can find that Guangdong province always has the most online followers. Beijing and Shanghai also have a large followers group, so the company can try to convert some offline promotion events to online platforms in Guangdong, Beijing, and Shanghai. A good characteristic of online promotion is that once having a sufficient audience, promoters can use less money to get larger exposure volume compared with traditional offline advertisements. However, if you don't have enough audience in that place, T-bet in these graphs, for example, online promotion is not a good choice at this moment because you waste money on people who don't care about your brand. Considering the customers' habits and age distribution of each brand is different, we will not make a comparison between the brands.

For our client, Mercedes, we suggest to classify all regions into 3-period types: Online-Mature Period(red and yellow part), Online-Semi-Mature Period(blue part), Online-Immature Period(gray part). Regions in Mature Period have large numbers of online followers now, so Mercedes needs to put more effort into building the online connection with these people. The more connection on the Internet between brand and customers, the less communication cost will have for the brand itself. The brand can implement 7-24 customers' feedback collection and analysis with the help of AI and seek the potential customers by detecting the purchase signal, like someone asking the price or driving experience of a car. If we believe the purchase activity has a golden time that the customer has a very high probability to pay, the online platform can help the brand to grab the chance within the golden time more efficiently. For the regions at Semi-Mature Period, we suggest Mercedes change more services online, developing and strengthening customers' online habits. For example, build the online booking system for repairing services, build the online accessories shop with special discounts. Once people are more and more familiar with solving the problem online, the region is not far from the Mature Period.

For regions at Immature Period, there are few active online followers in Weibo. We suggest the company do more research on the reasons for sparse online followers. Is it because of the culture or society's development level? Then customize a specific promotion strategy for each region.

It should be noted that the conclusions and statements in this section might not be very accurate, as our source is just from Sina Weibo and we don't know the portion of online customers against offline customers. For example, a brand in region A has 10000 customers and 500 of them are online active users, but in region B the brand has 400 customers and all of them are online active users. If we just depend on the data from heat-map, region A might be classified as in Mature Period and region B might be classified as in Semi-Mature period. Online promotions are more appropriate to implement in region B since all of their customers are online. Thus, it's more appropriate to use the proportion of online followers to total number of customers in this region instead of the absolute values, but our method is also effective when we monitor the change of the online users' amount and make the adjustment if we have more information.

4 Limitations

1. Due to time limits and anti-scraping techniques on some websites, the volume of Data is not very high. We can also collect more information from other social media platforms such as Sina Weibo or Baidu. Besides, all data we collect is user-generated content and there might be some bias in our sample. For example, younger customers are more likely to comment on these websites, so we need to be very careful when we compare the scores of two different car models.

2. The data we analyzed are only from online platforms. Online data have some bias as we mentioned before, and after comparing with company internal data, we can draw more precise conclusions.

3. As the main findings above come from analyzing historical data, we cannot be 100% sure the suggestions are effective or not. If possible, we want to do some AB tests to further check our results.

4. The sentiment analysis models we used are based on BiLSTM models and trained by large-scale corpus. However, as there are some proper nouns in the car industry, the sentiment score may not be accurate enough. For example, '车子起步比较肉' is a special expression, meaning 'The car starts to move slowly'. It should be negative, but the sentiment score is positive 0.95017. In the future, we need to further train the LSTM model by car industry related corpus.

5. There are always fake reviews on the website. Car brands and 4s stores may use fake reviews to attract customers and increase sales. It’s hard to tell fake reviews as companies spend large amounts of money to customize fake reviews in order to make it look real.

6. The rating score and reviews may vary according to comment time. For example, people tend to give high ratings right after buying the car. After a few months of driving, some problems appeared, so the rating score decreased.

Reference:

1. Brown, B. (2020, March 31). Market Research: 5 Ways Social Listening Can Help. Retrieved from https://mention.com/en/blog/market-research-social-listening/
2. Hao, J. (2019, January 31). Correspondence Analysis: What is it, and how can I use it to measure my Brand? (Part 1 of 2). Retrieved from https://www.qualtrics.com/eng/correspondence-analysis-what-is-it-and-how-can-i-use-it-to-measure-my-brand-part-1-of-2/
3. Minkner, K. (n.d.). Why You Need Social Listening In Your Market Research ... Retrieved from https://interq-research.com/why-you-need-social-listening-in-your-market-research-quiver/
4. 苏格兰折耳喵 . (2019, June 6). 干货丨 如何利用 Social Listening 从在线垂直社区提炼有价值的信息？：人人都是产品经理. Retrieved from http://www.woshipm.com/data-analysis/2408624.html
5. https://github.com/ownthink/Jiagu
6. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html

Appendix:

Description sheet of each variable in data

Consumer Purchase Purpose	
Variable	Description
rank	The rank of the car within competitor models
car_name	The name of the car model
score	The average score of the car model
Advantage	The biggest advantage of the car
purpose	The purpose of buying the car for each customer
total_post	The number of posts we collect from the websites

Table 4

User Reviews	
car type	The type of the car models
purch time	The time the blogger bought the car
purch place	The place the blogger bought the car
purch price	The price the blogger bought the car
avg rate	The average Fuel consumption rate of the car
publish time	The time the post was published
score	The score of the car in 8 dimensions graded by the blogger
purch purp	The purpose of buying the car for this blogger
simple review	The simple evaluation of this car
review	The detailed evaluation of this car
cum mileage	The accumulated mileage the car has

Table 5

Weibo Interests	
name	The name of the account user
uid	The link of the account with its user-id
gender	The gender of the account user
region	The region of the account user from
tag	The interests tag of the account user

Table 6

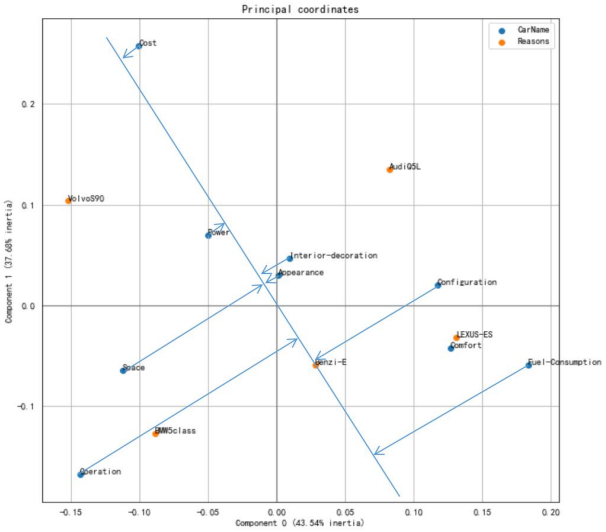
Cons(C CLASS)	Keyword
The landing price in many cities is much cheaper than ours, this is too pitted! **The reversing radar is unresponsive, and the reversing vehicle has encountered a tree once.** The steering wheel is heavy.	Radar Reverse
The speed is a bit slow at the start, and the steering wheel is a bit heavy at low speed. Secondly, **fuel consumption is higher.**	Fuel consumption
The function buttons are a bit messy, so you have to look carefully. **The back-seat cushions are a bit short and not very comfortable,** and some mid-range parts squeak during driving. **The space inside the car is not particularly large!**	Space; Back-seat
There is a strange sound when adjusting the air conditioner. It 's quite obvious. Another thing is that when **the car keeps stepping on the brake, the watch turns old and rises, and the return is falling.**	Sound; Brake
The new car **has a big smell**, the control is very general, and the shock absorption is relatively poor!	Smell
The trunk is a bit small. For me is enough, but from the perspective of home, it is really not enough.	Trunk
Not drive smoothly at low speed, which may be related to insufficient torque. When the steering wheel turns at a low speed, there is abnormal noise and a vibrating feel. **The back door needs to be closed tightly.** Besides, **the start engine sound is slightly louder.**	Not drive smoothly; Start engine Back Door
There is **no electric rearview mirror folding.**	Rear view mirror
It does not drive smoothly at low speed, a frustrating feeling. The trunk is smaller for me! **Fuel consumption is indeed higher for a 1.5T car. Maintenance and repair costs are relatively large.**	Maintenance
I have a classmate who is currently in Mercedes-Benz. According to him, the C200L is not worth it. The ordinary car is covered with a Mercedes-Benz car shell. **It is not worth the price in terms of configuration or other aspects.**	Price

Table 7

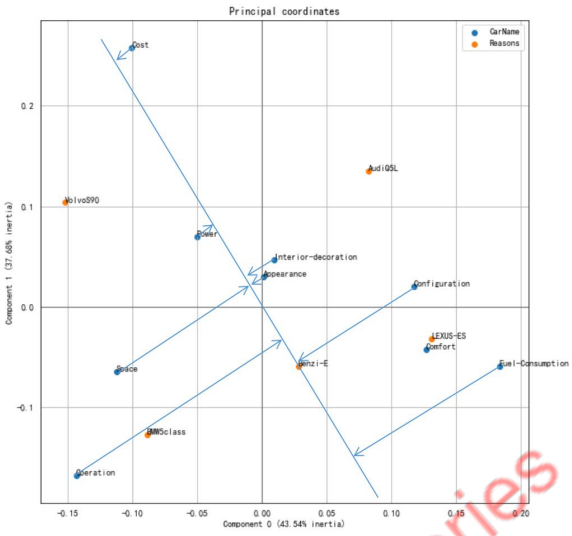
Cons(E CLASS)	Keyword
The most dissatisfied is that s**uch an expensive car does not have keyless entry and one-click start, navigation, reverse images, seat heating, and folding Rear View mirrors are not standard!** There are two or three hundred thousand cars! In addition, Beijing Benz C300GE has AMG wheels, but this is also an E-class coupe, with a chant! It's not a problem for Mercedes-Benz!	Rearview mirror; Reverse

The interior style is relatively old-fashioned, and the precision is not enough. It gives a strong sense of cheapness. The color of the match is younger. More owners will choose the car. **If the seats are raised, the armrest box can not be reached.** The rear seat is not as soft as the front seat. Chassis tuning is hard	Seat; Interior
First of all, **the taste of the new car, um, let's say it is not slight, and it has always been there,** and then when the vehicle is started, the sound of the engine, I don't know that I fucked bought a million tractor! Second, the new car was bought for 11 days, passed a speed bump, 60 yards speed, and the left front wheel punctured! ! ! ! Flat tire! ! Then it didn't take long, most of my own reasons, and there was a small traffic violation, although my own reasons dominated!	Taste
There was a noise when starting the cold car, and I also encountered it. It will be fine after a few days. It is estimated that the running-in of the belt has eliminated the problem of insufficient debugging. SSSS has also added lubricant, and the effect is not obvious. **The radar response of the vehicle in front lags behind, sometimes it does not call when it is close, and it calls only when everyone leaves, and lights up. The navigator upgrade is slow**, the German road construction is not fast, and it is normal to upgrade every 2 years.	Navigator; Radar
In the late 2017, the manufacturer simply equipped all the 19-inch wheels with an 18-inch hot wheel except the E320. Although it is disguised to reduce the risk of late tire bulging and the cost of tire replacement, it reduces the probability of this common disease. But it is also a simple match in disguise. **It is still the quality of assembly after domestic production, the abnormal noise of the central control, the workmanship and the overall sense will make you feel a bit off-price**	Domestic
The most unsatisfactory thing is that **the engine noise is a bit loud when the car is started in a cold car.**	Engine
The tire noise is still more obvious. I don’t know if it ’s the brand of the tire or the cause of the car itself. Another shortcoming is obvious. **Everyone knows that the noise and jitter of the cold engine start are similar. It’s the same as a tractor, haha, it's a bit exaggerated,** but the jitter is gone after the hot car, and the noise is much better.	Noise
Abnormal brake sound. After a period of time, when the speed is slow, I don't know if the brake pad is a little creaky. **Higher prices**; poor fuel economy. In addition, **it must be said that Mercedes-Benz's maintenance is quite expensive.**	Maintenance; Price
The wheels are so ugly, flat tires are said to be easy to puncture, and you must drive carefully; in terms of mute, the others are very good, that is, the tire noise is very large.The radar is also a little bit bad, especially on both sides, the response speed feels very strange, and it only starts to drip when it has passed the narrow place.	Tire; Wheels
I drive not smoothly in the first and second gear. I think it is because there are more gears. **The first gear uses a larger scale ratio to alleviate the lack of power at the start**. The rear middle seat is very uncomfortable, the seat cushion is too hard, and it is also a rear-drive car.	Power; Drive not smoothly

Table 8



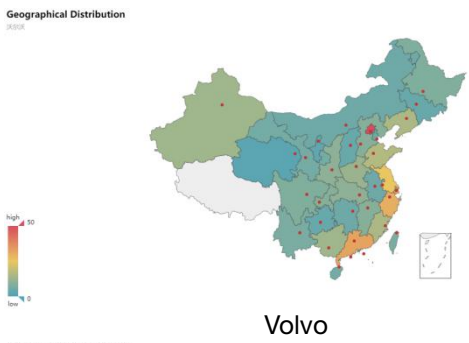
Graph 28



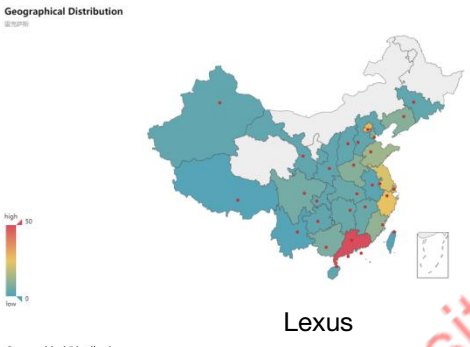
Graph 29

Graph 30

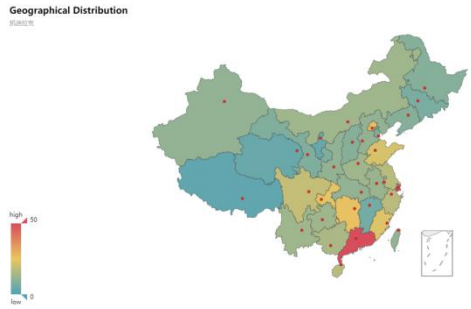
Graph 31



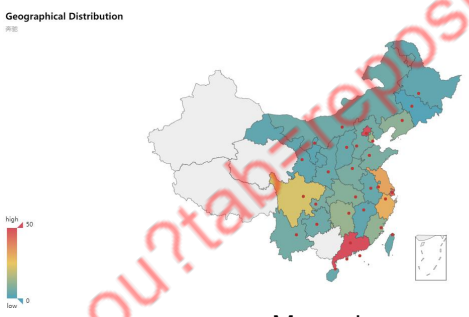
Volvo



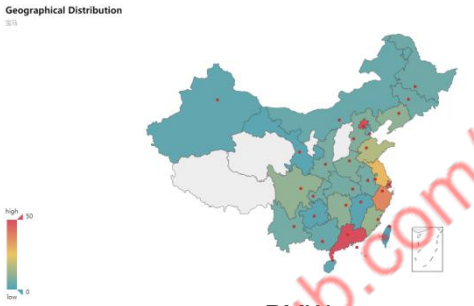
Lexus



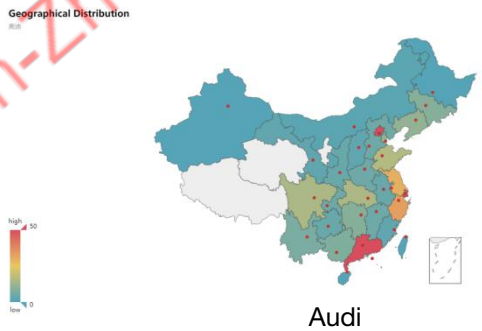
Cadillac



Mercedes



BMW



Audi

Graph 32-37