Journal of Marketing Management
December 2017, Vol. 5, No. 2, pp. 1-14
ISSN: 2333-6080(Print), 2333-6099(Online)

Published by American Research Institute for Policy Development
DOI: 10.15640/jmm.v5n2a1
URL: https://doi.org/10.15640/jmm.v5n2a1

Ethnicity in Car Purchase Decisions

Xin Zhao[1]

Abstract

Ethnicity is found to play an important role in consumer's purchase decision. However, more research is needed to broaden product category in academic research on ethnic consumption (Ogden et. al., 2004). In this paper, we analyze the impact of ethnicity in consumer choice in terms of car purchase. We approachthis question linking consumers' ethnicity to their cars' brand country of origin. In particular, we are interested in Asian American consumers'automotive purchasing decisions - do they drive Asian carbrands such as Honda, or Hyundai? We employ a large geographically diversified data set of registered car-dealerships in the United States and combine spatial cluster analysis, trade area analysis and regression models to identify the determinants of brand origin affinity based on socio-economic attributes across trade-areas. The results show that Asian Americans on average tend to buy Asian cars, but the affinity decreases with income, and the strength of this relationship is weaker than expected. Nonetheless, our proposed method helps finding promising new markets based on location specific attributes.

Keywords: ethnicity, car purchase, country of origin, spatial cluster analysis

1. Introduction

"You are what you eat" is a phrase that dates back to the 19th, possibly even to the 16th century[2], with the notion that your nutrition determines your health. The German poet Goethe suggested pretending to be what one appears to be, as one is defined by one's reputation.[3]From a marketing perspective, this boils down to stereo-typing consumers and identifying consumer groups by the types of products they consume. Cars are considered to make an especially strong statement about their buyers. For example, according to a Forbes study (Forbes, 2008), Porsche drivers are thought to be self-rewarding achievers with a household income for around $390,000. And if you drive a Porsche 911, you are most likely a man (87% of 911 buyers are).The idea is that consumers are drawn to certain types of products, i.e. country-of-origin (Roth and Romeo, 1992; Saeed, 1994; Hsieh et. al., 2004; Lin and Chen, 2006; Wang and Yang, 2008). What is much harder and costly is to identify the characteristics of buyers that tend to buy a certain brand or product, and a substantial share of marketing research expenditures fall into this category. The most common approach is to use surveys, and it has been adopted both by consulting firms as well as academic research. For example, Wu (2011) found out that Chinese Americans prefer Japanese brand to American brand, with European brand in between in term of car purchase decision making. However, surveys for that purpose have a range of undesirable properties. (1) It is hard to draw a large random sample, specifically for big ticket luxury items such as cars. Consequently, many studies use convenience samples instead (Wu, 2011 had a convenience sample of 150 participants), which often have been hard to identify statistical properties, and thus may lead to highly unreliable results.(2) If surveys are conducted only in a specific region, for example due to cost considerations, it may be hard to generalize the results.

[1] Associate Professor in Marketing, School of Business, University of Redlands,1200 E. Colton Ave, Redlands, CA, 92373, USA xin_zhao@redlands.edu

[2]http://www.phrases.org.uk/meanings/you%20are%20what%20you%20eat.html, accessed 8/4/2012

[3]Seid, was ihr scheint! Man ist, wofür man gilt." (Johan Wolfgang von Goethe).

To choose an obvious example, sales of convertible cars are - all else equal - expected to be much higher in warmer climates. (3)Past or current experience influences responses to avoid cognitive dissonance.A current BMW driver may not want to admit to himself that she would have actually preferred a Porsche, which tends to increase errors on survey questions that link buyer attributes with actual purchasing behavior. (4) Visible consumption by others in the same geographic area of a particular type of product or brand may increase demand through the "keeping up with the Jones" effect, which is hard to identify without the knowledge of exposure to the product.

This paper suggests an alternative- and generally much more affordable - method of identifying consumer attributes by assigning consumers to geographic market areas serviced by many retailers that form a spatial cluster. The paper describes in detail the methodology, data and results using car purchases by Asian American consumers as an example. It overcomes many issues listed above: Instead of drawing a random sample, it uses actual purchasing behavior across the target geography. Geographic biases are also eliminated, as the universe of car sales in an entire market, in our case car purchases in the United States in a particular year are used. Instead of asking about preferences, revealed preference of purchases forms the basis of analysis, thus eliminating both aspiration and ex post rationalization biases. Moreover, provided that the required data is available, the costs of executing this type of study are a fraction of the cost of a survey. In addition, results can be tracked over time whenever new data becomes available. These benefits come at some cost: the types of questions that can be answered depend on data availability and quality and are limited to macro-aspects of consumer behavior. For example, questions about product specifics are impossible as there is no data available at this scale. If data at the required or wished for level of detail is not available, it also requires some assumptions which need to be revisited after completion of the empirical exercise. For example, we would have liked to know for each dealership what their sales of economy, midsize, and luxury cars were. Since weobtained this information only at the national level, we had to assume that this sales structure was the same for all dealerships. In our exercise, this is a cause of concern if, for example, Ford dealerships in a large number ofregions mostly sold luxury version of their cars, while in others only economy versions and those two did not cancel each other out. However, market analysts within large corporations should have access to that data at least for their own brands, and thus be able to obtain more precise estimates than we were able to. Small businesses can still gather insights based on the publicly available data (albeit most of it is for purchase) and make up for the lack of detail with the help of assumptions.

There is another caveat: To the best of our knowledge, this paper is proof of concept for a new method of how one canstudy consumer behavior. We apply it to studying general features of markets and characteristics of product groups.Product and brand specific analysis with this method is possible, but has not been carried out yet, so its suitability needs to be investigated further. At this point, we therefore think of it as a complementary type of analysis to confirm or question results obtained from surveys or other marketing research methods.Thebig question is: how do consumers' characteristics correspond with dominant product characteristics? To answer this question, we first need to assign the set of products, brands, or dealers under consideration into one or more categories, such as low -, medium- or high price segment and sports car vs. sedan. A combination of categories and their segments is possible, as we will show below. Next, we need to identify the relevant geographic markets that are served by these products. In particular, we need to create market areas that are served by identifiable points of sales for which data is available. In our case, we identified dealer clusters and the market areas each cluster would likely serve, however, different approaches can be more appropriate, such as neighborhood boundaries or local availability of substitutes (as in the case of highway gas stations). Next, we calculate consumer characteristics for the delineated market areas. Finally, we use statistical analysis, including correlation and regression analysis to identify how consumer characteristics influence sales by product or dealer category.

The main contribution of this paper is therefore methodological - exploiting the geographic heterogeneity of consumer characteristics to explore and test which characteristics influence purchasing decisions. The value of it is driven by three components - overcoming problems with survey analysis, affordability, and the ability to execute this analysis even for small businesses.The paper is organized as follows: section two discusses the research design and the data in terms of the chosen application, namely the car industry. This includes attribute construction of dealers, cluster identification and assignment of trade areas. It also discusses how consumer characteristics were identified. Section three presents the empirical exercise, followed by section four with the results and caveats. Section five concludes.

2. Research Design

In our sample study, we were ultimately interested in testing our hypothesis whether Asian Americans buy Asian cars, and whether this conjecture would hold through across all income categories and car segments, henceforth called tiers. Thus, the consumer characteristics we were interested in testing are ethnicity and income, while the car characteristics we were interested in analyzing were car brands' County-of-Origin (Asian, US, and Other-European) and car segments or tiers (Luxury, Middle Class, and Economy). To test our hypothesis, we first need to identify car dealer clusters and trade areas, which requires the following steps:

(1) Assignment of product characteristics: car-tier and origin; each dealership is then characterized by a tier and an origin. For example, a Lexus dealership would be considered an Asian Luxury car dealership.
(2) Assigning dealerships to geographic clusters through statistical cluster analysis. In our case, this required 16 cluster analyses: overall, by tier, by origin, and by cross-combinations
(3) Construction of trade areas around dealerships. For ease of calculation, we assumed non-overlapping trade areas.
(4) Calculation of consumer characteristics and sales by product characteristics for each trade area, such as total sales of Asian Luxury cars in trade areas around luxury car clusters.
(5) Statistical analysis, namely regressions of sales within trade areas (by product characteristics) on consumer characteristics and controls.
(6) Interpretation of results and robustness checks.

This section describes steps (1) to (4) and the required data sources for these steps. Steps (5) and (6) are described in sections 3 and 4, respectively.

2.1 Tier and Origin Assignment

Our method of assigning origin was straightforward by "sound". If a brand sounded European, we sorted it into the "European / other" category. Many European brands are predominantly engineered, designed and made in Europe: Mercedes, VW, Ferrari, all fit this category well. Saab, however, had been acquired by General Motors in the year 2000. In 2008, the year for which we could obtain consistent data, it was a whole owned subsidiary of General Motors (GM). Due to its European origin, we still called it European.[4] We called a brand American when it was directly marketed by one of the big three car companies, General Motors, Ford, and Chrysler. These include the whole range of brands, such as Cadillac, Lincoln, Hummer, Buick, Oldsmobile, Ford, and Saturn. And we called a brand Asian when its origin was in Asia, again regardless of ownership or production location. Controversial examples here include Mitsubishi, Nissan, Suzuki, which are partly foreign owned or Toyota, which produces a large share of their cars in the United States.Tier Assignment was more involved. American car companies created brands with brand image in mind, for example, Ford created the Lincoln and GM the Cadillac as upper tier brands. The Japanese followed suit with Acura, Infinity and Lexus as the upper tier brands for Honda, Nissan, and Toyota respectively. Some European (sounding) brands also have easy assignments, such as Jaguar, Porsche, and Mercedes. VW, on the other hand, produces cars from economy, such as the Golf, to luxury class, such as the Phaeton, although with questionable success. So which tier should VW be assigned to? Moreover, all brands across all origins offer upgrades and extras, which can- in the extreme - result in a price tag for an economy class car being considered upper middle class, at the least. We therefore needed to have a consistent methodology to assigning tiers. We were able to obtain sales data for North America for all models and brands, such as BMW 3 and 5 series, Mini, and Rolls Royce from the Automotive News Data Center, and combined that with average car prices in that year by model from various web-sources (see appendix). We used this to calculate the weighted average price of cars by brand, using sales figures as weights and then ordered brands by weighted average price. We then chose cut-offs by value sold so that the dollar sales volume across all three groups would be close to being equally split (around 33%).[5]

---

[4] Recall that we consider the contribution of the paper conceptual. To be able to do such assignment correctly, one would have to construct surveys about origin perception or drop potentially controversial brands. Due to the increased internationalization of the car industry with consumers mostly unaware of engineering, design, and production locations, we consider this a minor problem.

[5] Due to uneven cut-off values, the middle segment ended up somewhat larger than the other two.

This resulted in the top 17% ofunit sales to be counted in the luxury segment, the next 40% counted in the middle class segment, and the remaining 43% economy segment. One particular problem is posed by multi-brand dealers, which constituted about 10% of the dealerships in the data we had obtained. We used parallel methodology to determine the tier and origin of a multi-brand dealer. In particular, we used the market-share data to determine the tier in the following way: We calculated the value-market-shares per brand a dealer carries and added them by tier and origin. For example, if a dealer had two Asian and three US brands in his brand portfolio, we added the national dollar value sales of the two Asian brands and the same for the US brands. If the total dollar value of sales of the two Asian brands was higher than the three US brands, the dealer was considered an Asian brand dealer, otherwise it was considered an US brand dealer. We performed the same calculations by tier and assigned the brand the tier with the highest total value of sales. This procedure biases results towards national average, and likelyperforms better for volume-brands (tier 2 and 3), but not for low volume brands, since it overemphasizes the former. In the robustness checks, we dropped multi-brand dealers entirely, without substantive change in the results. The specific cut-offs values by brand are listed in the appendix.

2.2 Assigning Dealerships to Geographic Clusters

Choosing the type of cluster analysis is a critical step since different methods of cluster construction reflect different assumptions of consumer behavior. The first assumption is about the general type of cluster assignment - random or hierarchical. In a random cluster assignment, the system randomly chooses the number of starting points (in our case, car dealers) given and then allocates all other points / dealers to the existing clusters till it passes a threshold value in a heuristic process of minimizing some criteria, such as minimal distance within and maximal distance between clusters. A random cluster assignment assumes that consumers perceive the structure of their shopping environment differently each time they leave the door to go shopping for the particular good they are after. This can only be rationalized by incomplete search and very short memory for the consumers, which are somewhat unrealistic. Moreover, results are not replicable, since each time a new set of starting clusters is chosen. Hierarchical cluster analysis implies that consumers perceive the retail structure as fixed in geographic space and they will likely shop in the cluster that is next to them. While these assumptions are still strong, they appear more realistic and produce replicable results in terms of cluster assignment of dealerships. The three most common methods of hierarchical cluster allocation are nearest neighbor analysis, centroid analysis and average distance analysis. All of these correspond to different search behavior within versus across clusters. Nearest neighbor analysis is probably the most commonly used method. It assumes that a consumer enters the cluster at a specific point, and is then willing to drive no more than x kilometers from her current store to the next store. Once she arrived there, she is again willing to drive up to x km to reach the next one.While this may mimic actual purchasing behavior of one specific customer, it assumes that customers enter the cluster at specific points and continue a particular path. This can lead to odd-shaped clusters resulting in unintuitive trade areas. Alternatively one can assume that consumers drive to the center of a cluster and then drive a radius of x kilometers from there. This would be reflected in the centroid method of cluster assignment. Finally, consumers may want to only visit stores or dealer shops which are on average no more than x miles apart from each other, that is the average distance between stores cannot exceed x miles. Consumers enter a cluster at their nearest access point and then visit stores within this cluster, as long as they are no more than x miles apart from each other (on average). We chose the last method as we think it reflects average consumer behavior the best, as cluster construction from a nearest neighbor model with different starting points can look very much like average distance cluster assignments.

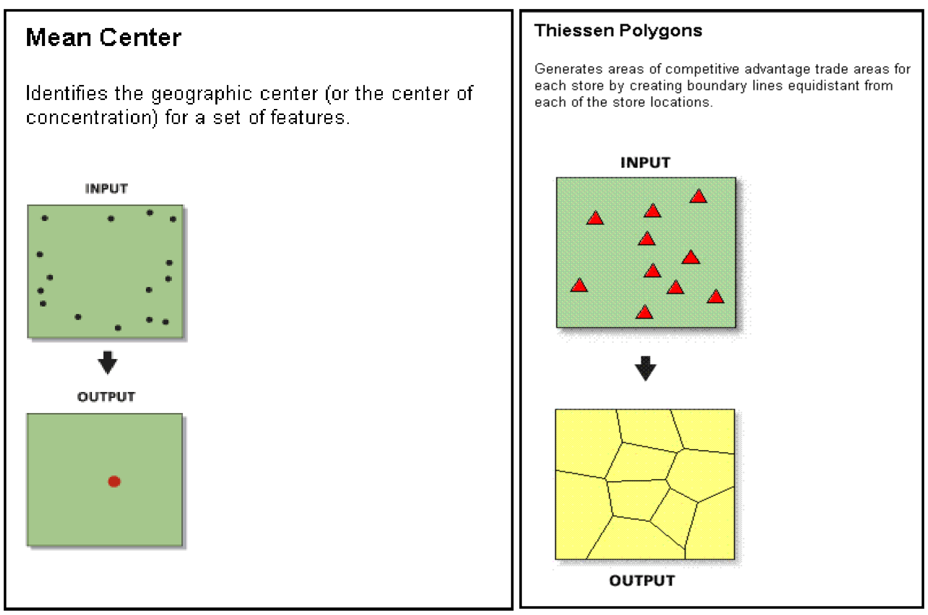
Next, we need to determine which dealers should belong to a cluster. Cobalt (2007) suggests that consumers are willing to drive about 30 kilometers (20 miles) on average, with some consumers willing to drive up to 100 kilometers (60 miles) for car purchases. Distances in that order of magnitude implies that some brands may be represented more than once in a cluster, which we call brand overlap. It may also lead to unreasonable large clusters with very large numbers of dealerships. Cobalt (2007) also points out those consumers on average visit six dealerships, but only shop for one brand per dealership. In order to see which cut-off distance to use, we calculated cluster assignments for all dealers for average distances between 5 and 40 km. Since statistical diagnostics results provided no guidance, we chose an average distance of 15 km between dealers, as this cut-off leads to cluster properties we believed most consistent with the findings of Cobalt (2007). Moreover, it happened to also offer “reasonable” brand overlap and cluster size distribution, as the following statistics indicate.

Number of Clusters:	3,526
Average Count of Dealers in Cluster: Standard Deviation	21.54, 20.41
Min/Max Count of Dealers in Cluster:	1 – 135
Share of Clusters with brand-overlap:	39%
Average number of Brand-overlaps (for clusters with brand-overlap):	6.85

2.3 Construction of Trade Areas

Taking cluster assignments as given, the next step is to decide how consumers will decide on their shopping location. Will they drive to the nearest cluster and shop there or drive to a further away cluster if that cluster has properties that they find more desirable, such as a larger number of dealers? While these variations can be modeled in principle, for a first pass to demonstrate the method, we assumed that consumers will drive to the cluster with the center of the cluster closest to them. This allows constructing trade areas in a simple way. Consumer characteristics in these trade areas can simply be calculated. The first step for this is to find the cluster mean center of dealer locations.[6]Then the area between the cluster mean centers is divided in such a way that lines run along equidistant points between cluster centers. The resulting geometric construct is called a Thiessen polygon. The following graphics illustrates the process:

Figure 1: Mean center and Thiessen polygon construction

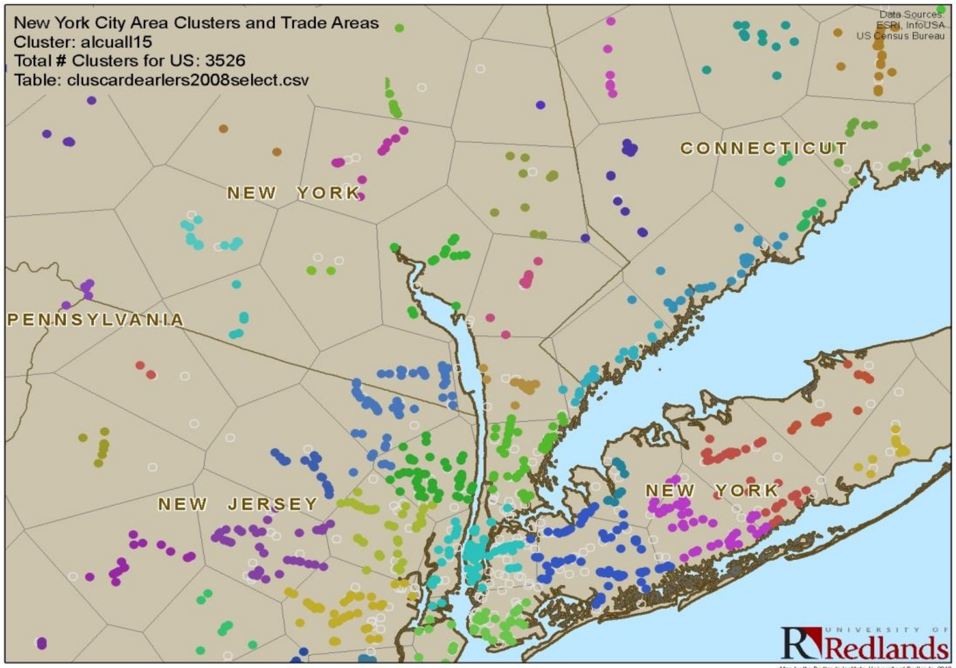


Source: Esri

As illustrated in the figure, the dealership clusters are first condensed in one point (left panel) and then Thiessen polygons of equidistant lines are used to separate the trade areas around the mean centers of each cluster. Those lines extend until they join up with other lines between points, thus trade areas cover the whole geography, regardless of population distribution. A particular nice feature of Thiessen polygons is that no other input was required to calculate the trade areas besides the set of points representing dealerships. However, issues like population distribution or travel times within the polygon are assumed away, read as: are ignored.Two more issues deserve special mention: dealership clusters leaking into neighboring trade areas and large geographic barriers. Let us first consider the issue that clusters may be leaking into neighboring trade areas. This problem is illustrated with a zoom into the New York area:

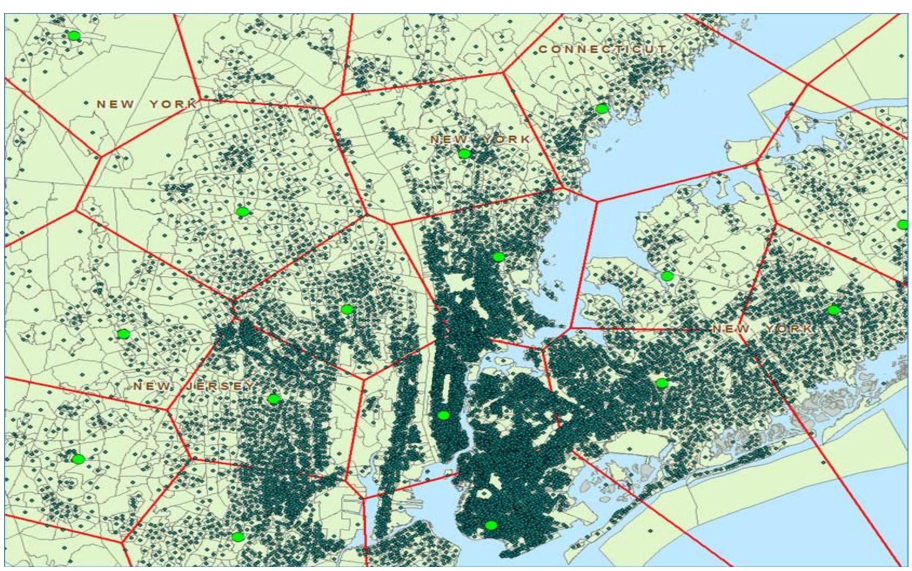
[6]Note that this is different from using the centroid method for constructing clusters.

Figure 2: Dealership clusters around New York



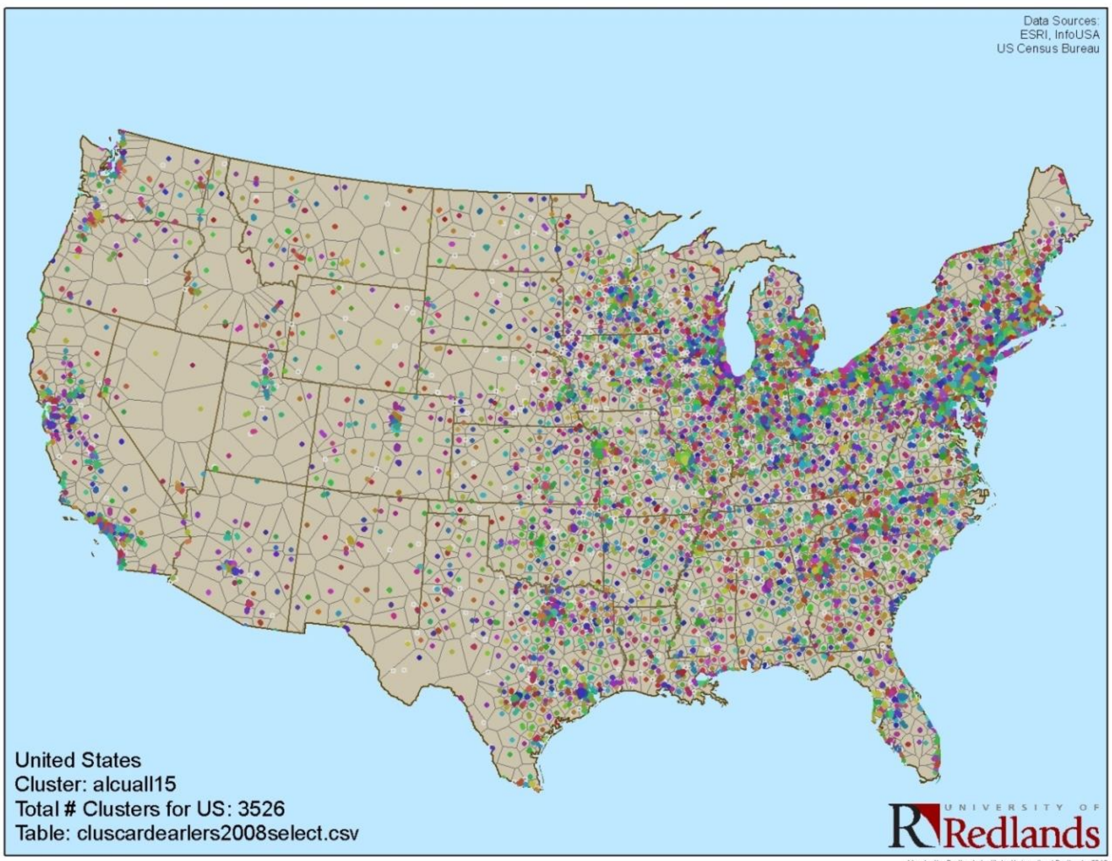
New York City is densely populated, and thus the cluster analysis based on statistical properties resulted in cluster separations that can almost appear arbitrary. In the figure, each point represents a dealership; with dealerships belong to the same cluster being shown in the same color. Trade area borders are indicated by the thin dark grey lines. In many instances, outliers from one cluster, that is dealerships relatively far from the center, may be located in the trade area of another cluster. This resulted from our simplified construction procedure of trade areas which only uses the cluster mean centers as information to construct trade areas.This phenomenon is especially likely in high density populated areas. While it is possible (but nontrivial) to adjust trade areas to avoid this issue, for the purpose of our study we expect only small distortions and therefore left it unattended.Another issue that deserves mentioning is the existence of hard to overcome barriers. In principle, these could be waterways, mountainous areas with no roads or any other kind of geographic hindrance. The following figure, again of New York, slightly extended, illustrates the issue:

Figure 3: Trade areas, mean centers and block-groups around New York



The mean centers are shown as neon green points, census tract centers are dark green points and trade areas are separated by red lines. Clearly, some trade areas extend across the water despite the fact that dealerships on the other side of the water are unlikely candidates for shopping visits. As outlined earlier, these problems can arise with many types of geographic barriers, but with waterways, this problem is likely to be the most severe as population densities tend to be higher along waterways as compared to, for example, mountainous areas. We thus manually adjusted those trade areas to make them more consistent with expected search behavior of consumers.All these adjustments lead to the following representation of cluster assignments and trade areas for the entire US. The following map shows all car dealers, with dealers belonging to the same cluster shown in the same color (if located next to each other). Dealers without a brand attachment are shown in white. As one can see from the map, the size of trade areas varies substantially, with small trade areas in the highly populated areas and large trade areas, often with only a few dealerships, located in the thinly populated areas.

Figure 4: Dealerships and their trade areas, entire United States



2.4 Variable Construction for Consumer Characteristics

The last step in the data construction for the empirical exercise was to calculate consumer characteristics for each trade area. For this, we used data from the American Community Survey (ACS) at what is called a block-group level. The Census bureau divides the entire United States into small units called census blocks for which it keeps track of demographic characteristics. However, it only publishes data at the block-group level, which is an aggregation of census blocks.

According to the Census Bureau, each block-group is constructed to contain between 600 and 3,000 people. There are more than 200,000 block-groups in the United States.[7] Figure 3 displays block-group outlines for the area around New York. One can see that more densely populated areas have a larger number of block-groups. In order to aggregate consumer characteristics, we used the block-group centroids. If the centroid of a block-group fell into a trade area, it was considered part of the trade area in its entirety, regardless of the possibility that some fraction of the block-group may have fallen outside of it and belonged into a different trade area. We did that for two reasons: first, we felt that - given the visual inspection of Figure 3 and many other areas - potential errors are expected to be small and cancel each other out. Moreover, we could have only based assignment of population on area, as we had no access to more precise data, which introduces other types of errors. Finally, we simply had to manage computing time - calculating the area share for each block-group that falls into one or another trade area for the entire United States was too strenuous for the computing resources we had available.Once all block-groups had been assigned to trade areas, we simply summed the counts of subjects with the same characteristics, for example the number of households with a household income larger than a threshold value or the number of married couples. We then divided it by the total number of subjects in each trade area to obtain the share of subjects with the same characteristic in the population of the trade area. Specifically, we chose the income tier thresholds to match our tier categories for luxury, mid-size, and economy cars.[8]

2.5 Data Description

Raw data comes from several sources. The first component is a large geographically diverse data-set from InfoUSA® that contains 12 million private and public US companies as geo-coded points.Each point has an 8 digit NAICS code attached, which allows selecting car dealerships only. Moreover, car dealerships have franchise codes attached, such as S for Saturn and T for Toyota. Multi-brand dealers were identified as those who had several codes attached. For each of the brands listed in Tables 1 and 2 in the appendix there was a separate franchise code available. This required aggregating some brand sales data, such as Lincoln and Mercury to match the brand codes available in the data-set.The data-set also has the number of employees and sales volume available. It is updated annually. In the raw data for 2008, there were 22,290 car dealers with franchise codes available. About 2,200 additional car dealers in the data-set had no franchise codes and were therefore not included in the study. We aggregated dealer characteristics by cluster. For example, we calculated total sales by all dealers in the cluster, but also the share of Asian, US or European cars as well as luxury, mid-size and economy sales in the cluster based on our assignment of brands to tiers and origins.

Consumer characteristics were derived from the American Consumer Survey (ACS) data with the estimates by block-group provided by ESRI. Those estimates combine information from the last available Census, which was 2000 for the data-set used, with the ACS data. ESRI's team of demographers and statisticians adjusts and projects the data to the desired level of geography, in our case block-groups. Our two main variables of interest were ethnicity as well as household income. However, we also calculated the data for a large set of control variables, such as employment, education, type of residence, gender, and so on. We finally matched the trade area data with the dealership data by trade area to perform our empirical exercises, which we describe next.

3. Correlation Analysis

We performed our empirical exercise in three steps: first, we run correlation analysis, next we performed regression analysis in log-levels (not reported here), and finally in shares. We first correlated the total counts of ethnic population with overall car sales as well as Asian car sales in each cluster / trade area. The results are presented in Table 2:

---

[7] http://www.census.gov/geo/landview/lv6help/pop_estimate.html, andhttp://factfinder2.census.gov/help/en/americanfactfinder help. htm#glossary/glossary.htm, both accessed 7/12/2012

[8] However, there are small discrepancies due to the available income classes available in the data from the Census Bureau versus the aggregation cut-offs from the car sales data.

Table 2: Correlation Analysis: car sales in dollars versus ethnic population

	Totalsales(n = 3526)	**Asian sales (n=1216)**
Asian	0.611	0.555
White	0.875	0.751
Black	0.564	0.455
Hispanic	0.617	0.571

Not surprisingly, total car sales are positively correlated with all ethnic groups across all trade areas. The correlation is the strongest for the white population. As for Asian car sales, interestingly, while all positive and significant, the correlations are weaker across the board, and white population still displaying the strongest correlation between ethnic population and car sales. As this could be driven by the simple sheer dominance of white population in absolute numbers, we were interested in how this analysis would fare if we replaced counts with shares. We thus correlated ethnic population shares with share of Asian car sales in each cluster. We distinguished between areas where there were no Asian car sales due to lack of available dealerships and those who had Asian car sales. The results are displayed in Table 3

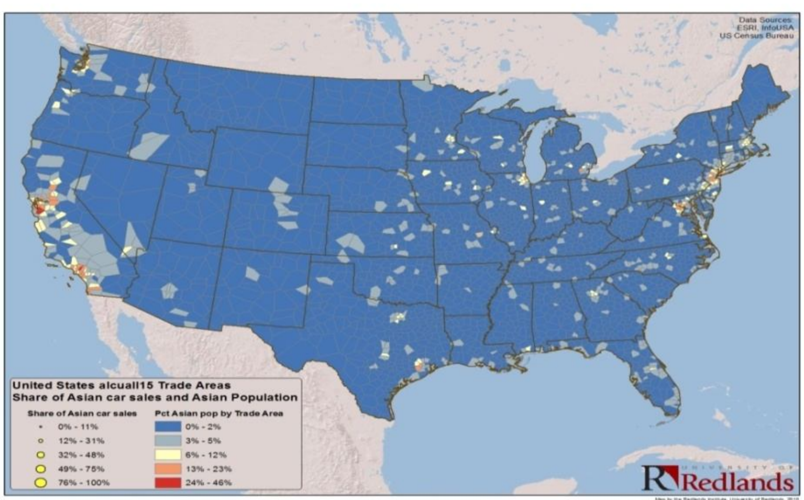
Table 3: Correlation Analysis: share of car sales in dollars versus share of ethnic population

	Share of Asian Car Sales (all clusters n=3526)	**Share of Asian Care Sales (Cluster>0, n=1216)**
Share of Asian	0.357	0.083
Share of White	-0.241	-0.106
Share of Black	0.133	0.028
Share of Hispanic	0.134	0.079

The results indicate that the share of Asian car sales is positively correlated with Asian population share, and this is the strongest correlation across groups. These results are consistent with our hypothesis that Asians tend to buy Asian cars. This is more striking as Asian car brands are mostly aligned in the economy segment, while Asians tend to have higher income than other ethnic groups of the population. The results also indicate that zeros are very important - once one removes all clusters without Asian car sales, the correlations are much lower, but still positive. If confirmed, Asian population shares could be indicative of market potential for Asian cars. Thus, a relatively high Asian population share may signalan opportunity to open an Asian brand car dealership in a particular area. Before we move on to regression analysis in the next section, we wanted to see whether our correlation results would be consistent with visual inspection of our results. We first created a map of the United Sates with Asian population shares by trade areas, which is displayed in Figure 5:

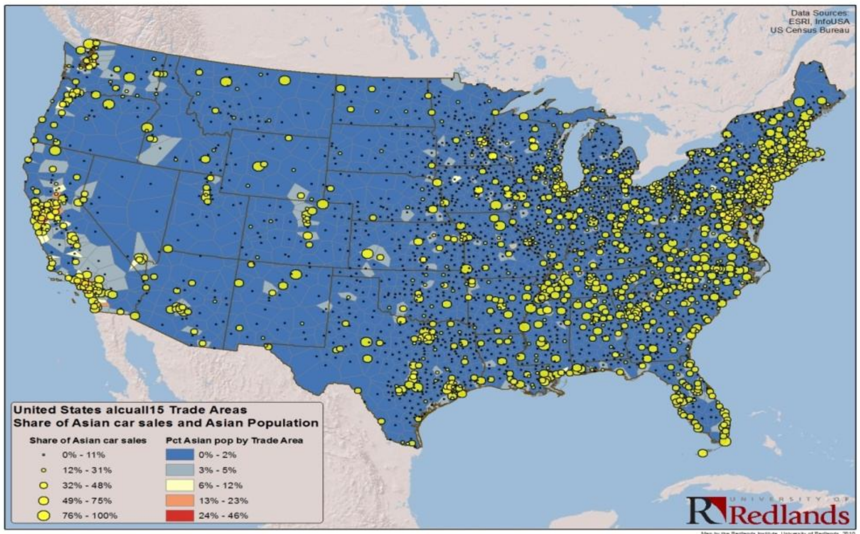
\

Figure 5: Asian population shares by trade areas



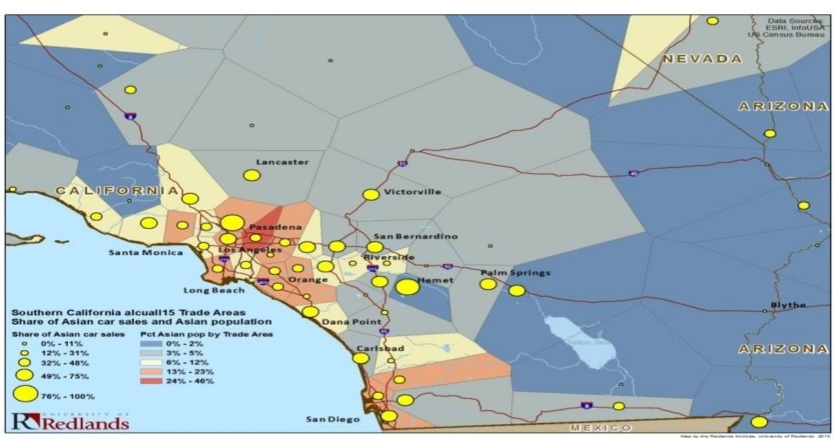
The map shows that Asian population is concentrated on the coasts, the Los Angeles, San Francisco, Seattle, and areas in the New England States dominating. Smaller pockets can be found in Chicago and Detroit and Minneapolis in the Mid-West. Overlaying this map of consumer characteristics with the share of car purchases of Asian brands (Figure 6) creates an interesting visual confirmation of our results - areas with larger shares of Asian population see larger dots, representing larger shares of Asian car purchases, smaller dots can be seen in areas with lower shares of Asian population.

Figure 6: Asian population shares and market shares of Asian brands by trade areas



The University of Georgia Selig Center for Economic Growth (2009) also notes that Asian American buying power has increased substantially in certain sates over the decade from 2000 to 2009. Some states saw dramatic increases in Asian American buying power, such as Wyoming with 187%, Nevada with 154%, and North Dakota with 146%. Combining these numbers with our study suggests that all these states may offer the potential for new Asian car dealerships. The large scale of the United States map does not allow for learning about some more intricate features of the possible connections between dealership sales and Asian consumer locations. We therefore visually inspect this relationship for Southern California in Figure 7, as it allows us to learn about some important features of our data. We find that the larger share of Asian brand car purchases may actually not happen right in the same trade area with the higher Asian population share, but right next to it. This suggests that our results could be biased towards zero, that is, actually importance of Asian population shares could be higher than the results we find in our regression analysis, which we will turn to next.

Figure 7: Asian population shares and market shares of Asian brands by trade areas, Southern California



4. Regression Results and Caveats

Here we present our core regression results. We conducted regressions of total car sales per trade area, on total number of Asians, age, education, employment status, home ownership and a large number of control variables. We repeated these regressions for Asian car sales, as well as in logs. All results confirmed the earlier correlation analysis. We repeated the analysis in logs again with the same results. Thus, we only focus on the more interesting results here where we used the same variables, but now calculated in shares instead of in levels. This means that we regressed the share of Asian car sales (in dollar values) onto the share of Asian American population, share of the population younger than 30 and older than 65, the unemployment rate, the share of home-ownership and the share of lower income earners (less than $ 47,000 household income), and additional controls such as gender, shares of other ethnic groups, and education.The results indicate that Asian population shares are positively related to shares of Asian car sales. While we report the results for all clusters, one has to realize that more than 2/3rds of those clusters have no Asian car sales. Thus, the significance of the results is likely largely driven by the fact that there are zeros or very small numbers for Asian cars and Asian population shares on both sides of the equation. This is still an important result, as it indicates that car dealers and companies do take ethnicity of the constituencies in their trade areas into account. However, in order to see the non-zero effects, one needs to look at the regression with positive sales activity. The first item to note is that there is no effect on Asian car sales through Asian population in our sample when all controls are included and we suspect education to be one of the more important drivers, as ethnicity and education are correlated with each other. We thus suspect multicollinearity and drop the additional control variables, namely shares of white, black, and Hispanic population shares[9], share of high-school graduates, college graduates, married and high income. Once the additional control variables are no longer included, Asian population shares and Asian car sales are again positively correlated as hypothesized. The regressions also confirm casual observation that it is predominantly the middle aged, employed none home owner with above lower income that tends to buy Asian cars.

	All Clusters	**Sales > 0 only**	**Sales > 0 only**
LHS Share A-Sales	**Extra Controls**	**Extra Controls**	**E.C. Dropped**
Share of Asian Pop.	1.08***	-0.020	0.37**
	(7.05)	(-0.084)	(2.46)
Share of Young	0.012	-0.86***	-1.12***
	(0.063)	(-2.62)	(-4.22)
Share of Old	0.094	-0.39*	-0.68***
	(0.76)	(-1.80)	(-3.91)
Share of Unemployed	-0.21	-1.16***	-0.67**
	(-1.30)	(-3.02)	(-2.18)
Share of Owner O. Houses	-0.069	-0.18*	-0.27***
	(-1.40)	(-1.78)	(-3.14)
Share of Lower Income	-0.25***	0.15	0.13*
	(-4.12)	(1.25)	(1.90)
Constant	0.77***	0.96***	0.91***
	(6.06)	(4.18)	(7.25)
Observations	3,526	1,216	1,216
R-squared	0.251	0.033	0.024

*** indicate significance at the 0.01 level, ** indicate significance at the 0.05 level, and * indicate significance at the 0.10 level

There are some important caveats that need to be noted in terms of the results. First, the motivation for this study was methodological in terms of data construction. So certain issues with data quality, regression methods and regression results deserve further consideration. In particular, the zeros deserve more explicit treatment econometrically, such as through probit analysis or a Heckman (1979) correction would clearly help improve the reliability of the results.

[9] Note that mixed race and pacific islander where exclueded from the regression to begin with since otherwise multicollinearity would have been built into the regession.

Moreover, results deserve to be evaluated for all tiers, origins, or both, with the respective clusters that result from alternate cluster assignments when only the dealerships of interest (such as luxury car dealerships) are considered. The issue of which are the right controls to be included to keep the model both parsimonious and reasonably predictive is far from settled for example, prior vehicle availability, the share of construction workers and drive times may all prove influential as controls or even decision variables in terms of what origin - or tier - car to buy. One also needs to be aware of some severe data availability issues: all we know is ethnicity, but not broken down by country: Japanese may be much more inclined to buy Japanese car brands than, for example, Vietnamese who had much more exposure to, for example French or American culture and products, even if it was far from voluntary. Moreover, there is no information on whether the Asian population shares are mostly formed by recent immigrants or second and third generation immigrants, which implies that they were born in the US and are much more familiar with American culture. The data is not broken down by new and used cars or online purchases, which can introduce substantial errors since car dealers are frequently willing to accept other car brands as trade-ins at the time of purchase. Other data quality issues, are that the data provided by both InfoUSA and ESRI constitute estimates which may or may not be as accurate as one might hope. Specifically, some features of the InfoUSA data were less desirable as the correlation between sales and sales personnel was close to 1 across brands, which deserves more investigation. Last but not least, the data available to us stems from a crisis year: Car sales had dropped dramatically and the relationship found in this paper may either be too strong or too weak, depending on how the Asian American population both fared and reacted to the crisis. All this deserves much more detailed scrutiny. However, the method shows substantial promise and deserves further application, we will discuss in the next section.

5. Conclusion and Empirical Implications

This paper introduces a new method on how to estimate the effect of consumer characteristics on product or brand attributes. It employs the case of ethnicity and its influence on brand choice in the car industry as an example. It describes the steps of data generation with the help of a Geographic Information System (GIS) in detail and provides some preliminary results to show the method can be used and what kind of results to expect. In particular, it shows that the geographically diversified data generated with the help of GIS is consistent with the hypothesis that Asians tend to buy Asian cars, as well as some more characteristics of who tends to buy Asian cars. Given data availability, namely that for the sake of this study, only data from a crisis year was available, some of the results have to be taken with a grain of salt, which is why we do not dare to provide detailed interpretations about the strengths of the effects we are finding: data quality may be an issue, the regression method needs additional fine tuning, and a crisis year may not be a good choice for investigating this relationship as the results may be too specific to the overall economic climate. However, the results that we obtained still indicate that the method that we used allows for substantial new insights in the future. It overcomes many shortcomings of traditional survey analysis, and it is, once set up, considerably cheaper to execute and can be repeated at very low cost year after year or whenever new data becomes available. The newly developed method is applicable to a wide range of other products and research questions, some of which we intend to take up in future work.

References

American Community Survey (2009) via ESRI: http://www.esri.com/data/esri_data/demographic-overview/census-overview/american_community, accessed 8/10/2012

Cobalt (2007) “Local Automotive Dealership: the Role of Interactive Media in the Local Car Shopping Process”http://www.slideshare.net/AutomotiveSocial/role-interactive-media-car-shopping-process, accessed 10/5/2012

ESRI (2009): "2009 Methodology Statement: ESRI Data--Business Locations and Business Summary", *ESRI white paper*, http://www.esri.com/library/whitepapers/pdfs/infousa-business-database.pdf, accessed 8/10/2012

Forbes (2008) http://www.forbes.com/2008/11/21/car-buyer-types-identity08-forbeslife-cx_jm_1120cars.html, accessed 12/10/2010

Heckman, James (1979). "Sample selection bias as a specification error". *Econometrica*, **47** (1): 153–61.

Hsieh, M., Pan, S. and Setiona, R. (2004), “Product-, corporate-, and country-image dimensionsand purchase behavior: a multicountry analysis”, Academy of Marketing Science, Vol. 32No. 3, pp. 251-70.

Humphreys, Jeffrey M. (2009). "The multicultural economy 2009" *Georgia Business and Economic Conditions*, University of Georgia Selig Center for Economic Growth, 69 (3).

Lin, L. and Chen, C. (2006), "The influence of the country-of-origin image, product knowledge andproduct involvement on consumer purchase decisions: an empirical study of insurance andcatering services in Taiwan", *Journal of Consumer Marketing*, 23(5), pp. 248-265.

Ogden, D. T., Ogden, J. R., &Schau, H. J. (2004). "Exploring the impact of culture and acculturation on consumer purchase decisions: Toward a microcultural perspective."*Academy of Marketing Science Review*, *3*(2004), 1-22.

Roth, M.S. and Romeo, J.B. (1992), "Matching product category and country image perceptions: a framework for managing country-of-origin effects", *Journal of International Business Studies*, 23(3), pp. 477-97.

Saeed, S. (1994), "Consumer evaluation of products in a global market", *Journal of InternationalBusiness Studies*, 25(3), pp. 579-604.

Wang, X. and Yang, Z. (2008), "Does country-of-origin matter in the relationship between brand personality and purchase intention in emerging economies? Evidence from China's auto industry." *International Marketing Review*, 25(4), pp. 458-474.

Wu, G. (2011). "Country image, informational Influence, collectivism/individualism, and brand loyalty: exploring the automobile purchase patterns of Chinese Americans." *Journal of Consumer Marketing*, 28(3), 169-177

Appendix:

Table 1: Tier and Origin Assignment

Luxury				Middle Class				Economy		
Asian	**O. Foreign**	**US**		**Asian**	**O. Foreign**	**US**		**Asian**	**O. Foreign**	**US**
Lexus	RR	Cadillac		Subaru	Mini	Buick		Mitsubishi	VW	Chevrolet
Infiniti	Maybach	Lincoln		Nissan		Mercury		Toyota	Lada	Saturn
Acura	Lamborghini	Hummer		Isuzu		Chrysler		Honda		
	Ferrari					Ford		Mazda		
	Bentley					Dodge		Hyundai		
	Aston Martin					Oldsmobile		Scion		
	Maserati					Pontiac		Suzuki		
	Porsche					Jeep		Kia		
	Mercedes					GMC				
	Jaguar									
	Lotus									
	BMW									
	Audi									
	Saab									
	Volvo									
	Land Rover									

Brand	Units	Average Price	Total sales	Cum units	% units	unit market share	value market share
RR	28	340,000	9,520,000	28	0%	0.003%	0.050%
Maybach	10	339,000	3,390,000	38	0%	0.001%	0.018%
Lamborghini	59	190,600	11,245,400	97	0%	0.007%	0.059%
Ferrari	116	173,079	20,077,164	213	0%	0.014%	0.105%
Bentley	360	170,990	61,556,400	573	0%	0.044%	0.323%
Aston Martin	37	119,500	4,421,500	610	0%	0.005%	0.023%
Maserati	201	110,000	22,110,000	811	0%	0.025%	0.116%
Porsche	2,133	67,319	143,590,400	2,944	0%	0.261%	0.752%
Mercedes	16,738	55,447	928,079,100	19,682	2%	2.049%	4.863%
Jaguar	1,866	51,641	96,361,760	21,548	3%	0.228%	0.505%
Land / Range Rover	3,981	47,250	188,102,250	25,529	3%	0.487%	0.986%
Lotus	210	46,270	9,716,700	25,739	3%	0.026%	0.051%
Cadillac	12,976	46,264	600,326,275	38,715	5%	1.589%	3.146%
BMW	18,965	42,859	812,822,300	57,680	7%	2.322%	4.259%
Lexus	13,798	38,977	537,804,695	71,478	9%	1.689%	2.818%
Infiniti	8,538	36,027	307,594,350	80,016	10%	1.045%	1.612%
Audi	6,994	35,291	246,823,420	87,010	11%	0.856%	1.293%
Hummer	5,960	33,390	199,004,400	92,970	11%	0.730%	1.043%
Acura	12,842	33,311	427,784,870	105,812	13%	1.572%	2.242%
Oldsmobile	100	32,000	3,200,000	105,912	13%	0.012%	0.017%
Saab	2,666	30,859	82,270,360	108,578	13%	0.326%	0.431%
Lincoln-Mercury	20,228	30,318	613,273,080	128,806	16%	2.477%	3.214%
Volvo	7,471	29,634	221,395,399	136,277	17%	0.915%	1.160%
GMC	40,102	29,000	1,162,958,000	176,379	22%	4.910%	6.094%
Isuzu	1,670	27,149	45,338,830	178,049	22%	0.204%	0.238%
Buick	14,454	26,069	376,796,750	192,503	24%	1.770%	1.974%
Jeep	38,338	23,000	881,774,000	230,841	28%	4.694%	4.621%
Mini	3,851	22,374	86,161,950	234,692	29%	0.471%	0.451%
Subaru	16,655	21,171	352,607,725	251,347	31%	2.039%	1.848%
Chrysler	33,354	20,941	698,470,875	284,701	35%	4.084%	3.660%
Ford	76,337	20,647	1,576,132,390	361,038	44%	9.346%	8.259%
Nissan	44,215	19,922	880,831,070	405,253	50%	5.413%	4.616%
Dodge	34,735	19,770	686,696,810	439,988	54%	4.253%	3.598%
Pontiac	24,003	19,739	473,786,730	463,991	57%	2.939%	2.483%
chevrolet	77,982	19,584	1,527,169,920	541,973	66%	9.548%	8.002%
VW	19,571	19,504	381,707,385	561,544	69%	2.396%	2.000%
Mitsubishi	7,126	18,771	133,760,863	568,670	70%	0.872%	0.701%
Toyota	89,251	18,213	1,625,566,930	657,921	81%	10.927%	8.518%
Honda	56,365	17,852	1,006,224,390	714,286	87%	6.901%	5.273%
Mazda	18,884	17,566	331,723,385	733,170	90%	2.312%	1.738%
Hyundai	33,281	16,572	551,541,200	766,451	94%	4.075%	2.890%
Saturn	7,652	16,050	122,812,820	774,103	95%	0.937%	0.644%
Scion	14,248	16,036	228,476,000	788,351	97%	1.744%	1.197%
Suzuki	6,061	14,292	86,624,259	794,412	97%	0.742%	0.454%
Kia	12,350	14,272	176,253,370	806,762	99%	1.512%	0.924%
Daewoo	10,000	14,000	140,000,000	816,762	100%	1.224%	0.734%