

# Car Centres Placement Problem

## Problem presented by

Emad Khalawi and Hasan Zidan

*Abdul Latif Jameel Co. Ltd*

## Executive Summary

Deciding on where to locate the next retail outlet is a problem with a long and distinguished history beginning with the early work of Hotelling (1929) and extended by Huff (1964). The basic idea is to view potential customers as sources of purchasing power while a retail store possesses attractiveness thus creating an interacting particle model. Here, we address the issue of where to locate a new car center based on a limited dataset. A method for distilling aggregate population information down to sub-regions is developed to provide estimates that feed into the optimization algorithm. Two measures were used in the optimization: (i) total market share and (ii) total attractiveness. Total market share optimization is found to lead to placing the center close to competitors, while total attractiveness optimization is found to lead to placing the center closer to centroid of the population.

**Version 1.0**  
**July 13, 2011**  
iii+9 pages

## **Report coordinator**

S. Jaimungal

## **Contributors**

Bashair Aman (KAUST, Saudi Arabia)  
Gustavo Chavez (KAUST, Saudi Arabia)  
Lifang Chen (KAUST, Saudi Arabia)  
Anas Ismail (KAUST, Saudi Arabia)  
Sebastian Jaimungal (U. of Toronto, Canada)  
Kamil Kulesza (Centre for Industrial Applications of Mathematics and Systems  
Engineering, Poland)  
Anthony Lock (OCIAM, UK)  
Faraan Rashid (KAUST, Saudi Arabia)  
Antoine Vigneron (KAUST, Saudi Arabia)

**KSG 2011 was organised by**  
King Abdullah University of Science and Technology (KAUST)  
In collaboration with  
Oxford Centre for Collaborative Applied Mathematics (OCCAM)

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Problem Statement . . . . .	1
<b>2</b>	<b>Thinning the Population</b>	<b>2</b>
<b>3</b>	<b>Thinning the Growth Rate</b>	<b>3</b>
<b>4</b>	<b>Interacting Particle System</b>	<b>5</b>
<b>5</b>	<b>Conclusions</b>	<b>7</b>
	<b>Bibliography</b>	<b>9</b>

# 1 Introduction

## 1.1 Problem Statement

Over the next three to five years Abdul Latif Jameel Ltd (ALJ) would like to open a number of car showrooms and service centres. The challenge is to first assess the current methods for finding good locations (in terms of their soundness and accuracy), and then to find a more accurate approach. The assessment criteria includes how far customers will have to drive (15 minutes cap), the market share that the new centres would deliver, and their profitability.

This problem is related to the literature on market share and retail locations. One early model was proposed by Hotelling (1929) who assumes that customers patronize the closest facility and disregard quality. Huff (1964) introduced a better approach for estimating the market share of a retail facility. He assumes that the probability that a customer patronizes a given facility is proportional to the total area of the facility and inversely proportional to the square of the distance between the customer and the facility. As such, the model can be viewed as an interacting particle model. A number of extensions appear in the literature, where they mostly focus on modifying the assumption that the facilities area affects its attractiveness and instead replaces floor area with a number of other “attractiveness” coefficients (see e.g. Nakanishi and Cooper (1974), Jain and Mahajan (1979)). However, these parameters are all rather ambiguous and difficult to estimate. We take a different approach and infer it from the market. Locating multiple new facilities was considered in Achabal et al. (1982) while Ghosh and Craig (1991) view the new facility as part of a franchise and try to maximize sales while simultaneously minimizing impact on other franchise outlets.

One of the main challenges in the specific problem posed by ALJ is the lack of data. ALJ has partitioned the country into Primary Market Areas (PMAs), effectively a city, and these are further broken up into several Customer Driving Areas (CDA) based on: survey results, traffic flow, major roads as edges, and customers’ mobility. Data is available at the PMA level only (total population and expected growth rate), and can only be estimated or guessed at the CDA level. We developed a method for thinning the PMA information down to CDA level which preserves the total population as well as the total population growth rate (over a small enough time frame). These corrections help ALJ to more accurately predict demand in the various CDAs and consequently affects the prediction of where a new facility should be located first. Given this information, we adapt the Huff (1964) approach by introducing a representative customer within each CDA and infer the attractiveness of the facilities based on current market share information. Then, two criteria were used to find the best location for the next facility: (i) market share (ii) total attractiveness of the franchise.

As an illustration, the developed methodology was tried on a small area of Jeddah shown in Figure 1. It was found that J maximises the total market share, while F maximises the total attractiveness.



**Figure 1:** The Jeddah PMA showing representative customers (white circles - radius is proportional to population), competitors (yellow pins), current ALJ locations (green pins), and potential future locations (blue pins).



**Figure 2:** Density appraisal for a specific CDA.

## 2 Thinning the Population

In this section, we show how the total PMA population should be thinned to a specific CDA. Within a PMA, high, medium and low density areas within a CDA were identified using domain specific knowledge from the ALJ team coupled with

satellite images. Figure 2 shows an example for a specific CDA. Denote the total area designated as high, medium and low density in CDA  $k \in \{1, 2, \dots, K\}$  as  $H_k$ ,  $M_k$  and  $L_k$ . Further, let  $\omega_H$ ,  $\omega_M$  and  $\omega_L$  denote the relative densities between high, medium and low areas. For the specific PMA of Jeddah, we estimate that  $\omega_H = 2\omega_M$  and  $\omega_L = 0.5\omega_M$ , i.e. the high density areas are twice as dense as the medium density areas and the low density areas are half as dense as the medium density areas. Clearly, the population within a given CDA is

$$P_k = \omega_H H_k + \omega_M M_k + \omega_L L_k$$

and the individual CDA populations must sum up to the total population

$$P = \sum_{k=1}^K (\omega_H H_k + \omega_M M_k + \omega_L L_k)$$

$$\Rightarrow \omega_M = \frac{P}{\sum_{k=1}^K \left( \frac{\omega_H}{\omega_M} H_k + M_k + \frac{\omega_L}{\omega_M} L_k \right)}.$$

Thus, only the relative densities are important. Substituting back into the individual CDA, we find that

$$P_k = \frac{\frac{\omega_H}{\omega_M} H_k + M_k + \frac{\omega_L}{\omega_M} L_k}{\sum_{k=1}^K \left( \frac{\omega_H}{\omega_M} H_k + M_k + \frac{\omega_L}{\omega_M} L_k \right)} P.$$

In Table 1 we show the estimated population within the high, medium and low densities of a particular CDA using the suggested method and the current method used by ALJ. It is important to note that the suggested method always sums to the total population regardless of what relative densities are used.

Population Density	Area	Current Estimate	Suggested Estimate
High	1.12	21,532	14,841
Medium	4.90	47,013	32,404
Low	24.90	119,358	82,268

**Table 1:** Population estimates within a CDA.

### 3 Thinning the Growth Rate

The current population estimates within a CDA is one important ingredient, however, the future population within CDAs provides additional vital information on how quickly a region is growing and consequently the demand in that region. Here, we provide a framework for determining the growth rate within a specific CDA by matching the total population growth over a small amount of time with the population growth within high, medium and low density areas. ALJ assumes that the growth rate is inversely related to the density, so that high density areas have low

growth rates and low density areas have high growth rates. We will simply assume that  $\lambda_H$ ,  $\lambda_M$  and  $\lambda_L$  represent growth rates in high density, medium density and low density areas. With these notations, we have that the total population will grow to

$$P_H(t) = P_H(0) e^{\lambda_H t}, \quad P_M(t) = P_M(0) e^{\lambda_M t}, \quad \text{and} \quad P_L(t) = P_L(0) e^{\lambda_L t}.$$

Here,

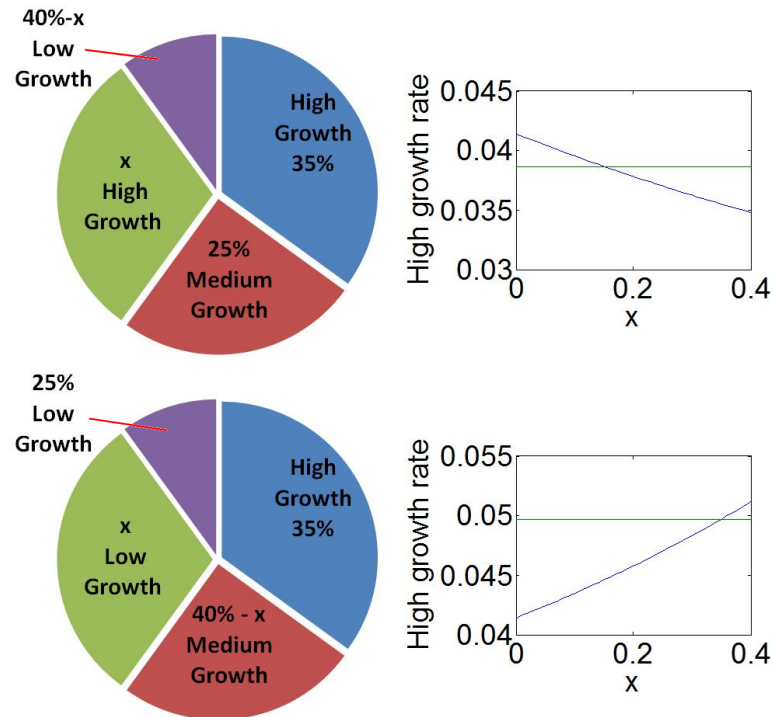
$$P_H(0) = \sum_{k:\omega_k=\omega_H} P_k, \quad P_M(0) = \sum_{k:\omega_k=\omega_M} P_k, \quad P_L(0) = \sum_{k:\omega_k=\omega_L} P_k,$$

and the notation  $\sum_{k:\omega=\omega_a}$  represents summing over the regions which have been identified as high, medium or low density. Only the total population growth  $\lambda$  is known and so  $P(t) = P(0)e^{\lambda t}$ . We assume that  $\lambda_L = a_L\lambda_M$  and  $\lambda_H = a_H\lambda_M$  with  $a_L$ , and  $a_H$  predefined relative rates of growth. If two populations grow at differing rates, then they cannot be represented as a single population growing at one fixed rate. However, over short enough time frames we can expand in  $t$  and equate the two population growths resulting in

$$\begin{aligned} P_H(t) + P_M(t) + P_L(t) &= P_H(0)(1 + \lambda_H t) + P_M(0)(1 + \lambda_M t) + P_L(0)(1 + \lambda_L t) + o(t) \\ &= P(0) + (P_H(0)\lambda_H + P_M(0)\lambda_M + P_L(0)\lambda_L)t + o(t) \\ &:= P(0)(1 + \lambda t) + o(t) \\ \Rightarrow \quad \lambda_M &= \frac{P(0)}{P_H(0)a_H + P_M(0) + P_L(0)a_L} \lambda. \end{aligned}$$

It is instructive to investigate the consequences in the two hypothetical situations. First, in the top panel of Figure 3, there are two high growth regions with a variable amount  $x$  in one region, and the growth rate in the high growth rate area is shown as a function of  $x$ . Note that as the high growth rate regions become larger, the implied high growth rate decreases. This is natural as the total population always grows with the fixed rate  $\lambda$  so if more of the population is growing fast, that rate must be low. Second, in bottom panel of Figure 3, there are two low growth rate regions with a variable amount  $x$  in one region, and the growth rate in the high growth rate area is shown as a function of  $x$ . In this panel, the high growth rate increases with  $x$  because a larger portion of the population growing at a low rate implies that there is a smaller proportion growing at the high rate. Further, since the total growth rate is fixed, the high growth rate area must compensate and increase with  $x$ . In contrast, both panels shows the (constant) estimate of the growth rate that the current method implies.

Finally, Table 2 shows the growth rate estimates based on the suggested method and the current method. Notice that the current method appears to underestimate the individual growth rates. Further, the total population projected at  $t = 5$  is 3,442,152 if the population growth rate is 2.9%. The current methods projection of 3,341,637 is over 100,000 too low. The suggested method on the other hand makes a prediction that is only 7,000 larger than the total population growth. If required, it would be possible to match the total growth exactly by equating the population and total individual growths at year 5.



**Figure 3:** The growth rate in the high growth rate region implied by thinning the total population in two scenarios. The growth rate factors  $a_H = 1.5$  and  $a_L = 0.5$ , while the total population growth rate is 2.9%.

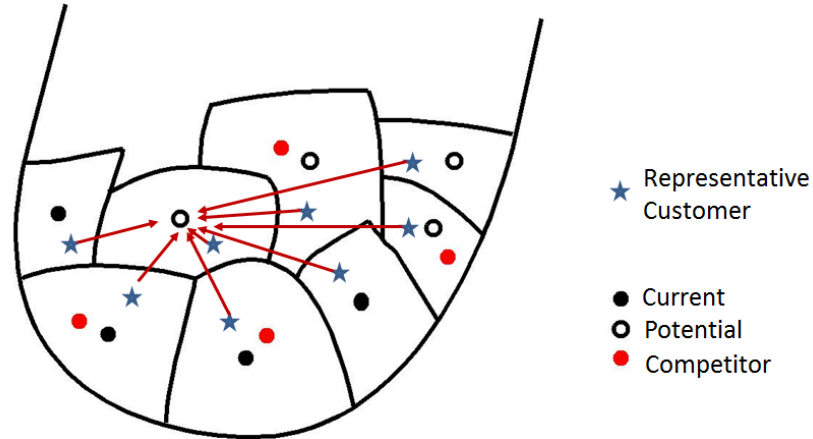
	Growth Rate		Pop. at $t = 0$		Pop. at $t = 5$	
	Suggested	Current			Suggested	Current
Low	2.17	1.7	2,179,990	2,429,766	2,374,090	
Medium	4.33	3.4	592,507	736,061	702,715	
High	6.5	5.1	205,042	283,904	264,832	
Total	-	-	2,977,539	3,449,731	3,341,637	

**Table 2:** The growth rate for high, medium and low growth areas implied by thinning the total growth rate. The growth rate factors  $a_H = 1.5$  and  $a_L = 0.5$ , while the total population growth rate is 2.9%.

## 4 Interacting Particle System

As described in the introduction, the Huff (1964) approach assumes that the customers patronize a retail location inversely proportional to the distance of the customer to the location and proportionally to the attractiveness (viewed as floor size) of the location. To simplify matters, since we have limited data on individual customers, we introduce a representative customer within each CDA of a PMA. This representative customer possesses buying/purchasing power  $B_i$  equal to the population in that CDA – the reason for this assumption is that ALJ has found that sales of vehicles and services are proportional to population within the CDA. We will further assume that there are  $Q$  competing franchises (with a total of  $M_q$ ,  $q \in \{1, \dots, Q\}$ ,





**Figure 4:** A franchise location possesses attractiveness inversely to distance to customer, while customers possess purchasing power.

outlets) and that each outlet of a franchise possesses an attractiveness coefficient  $S_q$  (i.e. all outlets of the same franchise have the same coefficient), but attractiveness decays as a function of distance to customer. Denote this decaying function by  $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$  – we will adopt an exponential decay function  $g(x) = e^{-\gamma x}$ , although any reasonable decreasing function can be used. Consequently, the market share of an individual outlet  $l$  from franchise  $q$  is given by

$$T_l^q = \sum_{i=1}^{N_{CDA}} B_i \frac{S_q g(d_{il}^q)}{\sum_{p=1}^Q S_p \sum_{k=1}^{M_p} g(d_{ik}^p)}. \quad (1)$$

Here,  $d_{ik}^q$  denotes the distance of representative customer  $i$  from outlet  $k$  of franchise  $q$ . Thus the total market share of the franchise  $q$  is

$$T^q = \sum_{l=1}^{M_q} T_l^q.$$

Current market share information can be used to infer the relative attractiveness of the various franchises. If we normalize the attractiveness factor and set, e.g.,  $S^q = \alpha^q S$ , then the following system of equations can be solved to determine the individual attractiveness of a franchise based only on current already open outlets:

$$T^q = \alpha^q \sum_{i=1}^{N_{CDA}} B_i \frac{\sum_{l=1}^{M_q} g(d_{il}^q)}{\sum_{p=1}^Q \alpha^p \sum_{k=1}^{M_p} g(d_{ik}^p)}, \quad \forall q \in \{1, \dots, Q\}.$$

The unknowns here are the  $\alpha^q$  (one of them can arbitrarily be set to unity) and this is a system of  $Q - 1$  polynomial equations of order  $N_{CDA}$ . Mild assumptions

on the function  $g$  will lead to the system having a unique solution. Based on ALJ estimates that Toyota and Hyundai have a 70%/30% split of market share, we find that the relative attractiveness of Toyota/Hyundai is  $\alpha^T = \frac{10}{3}\alpha^H$ . This result seems reasonable as it is clear that the Toyota is the front runner in Saudi Arabia, but it is somewhat surprising to see that although Toyota's market share is just under 2.5 times larger than Hyundai, Toyota's attractiveness is more than 3 times larger.

If we include a new location at  $l^*$  for franchise  $q$  and split (1) into terms corresponding to the new location and all others one finds

$$T^q = \sum_{i=1}^{N_{CDA}} B_i \frac{S_q g(d_{il^*}) + \sum_{l=1}^{M_q} S_q g(d_{il}^q)}{S_q g(d_{il^*}) + \sum_{p=1}^Q S_p \sum_{k=1}^{M_p} g(d_{ik}^p)}. \quad (2)$$

By optimizing (2) over the possible locations, the franchise will find the location which maximizes market share. Alternatively, the franchise may wish to instead maximize total attractiveness of the firm by ignoring competitors. This leads to optimizing the expression

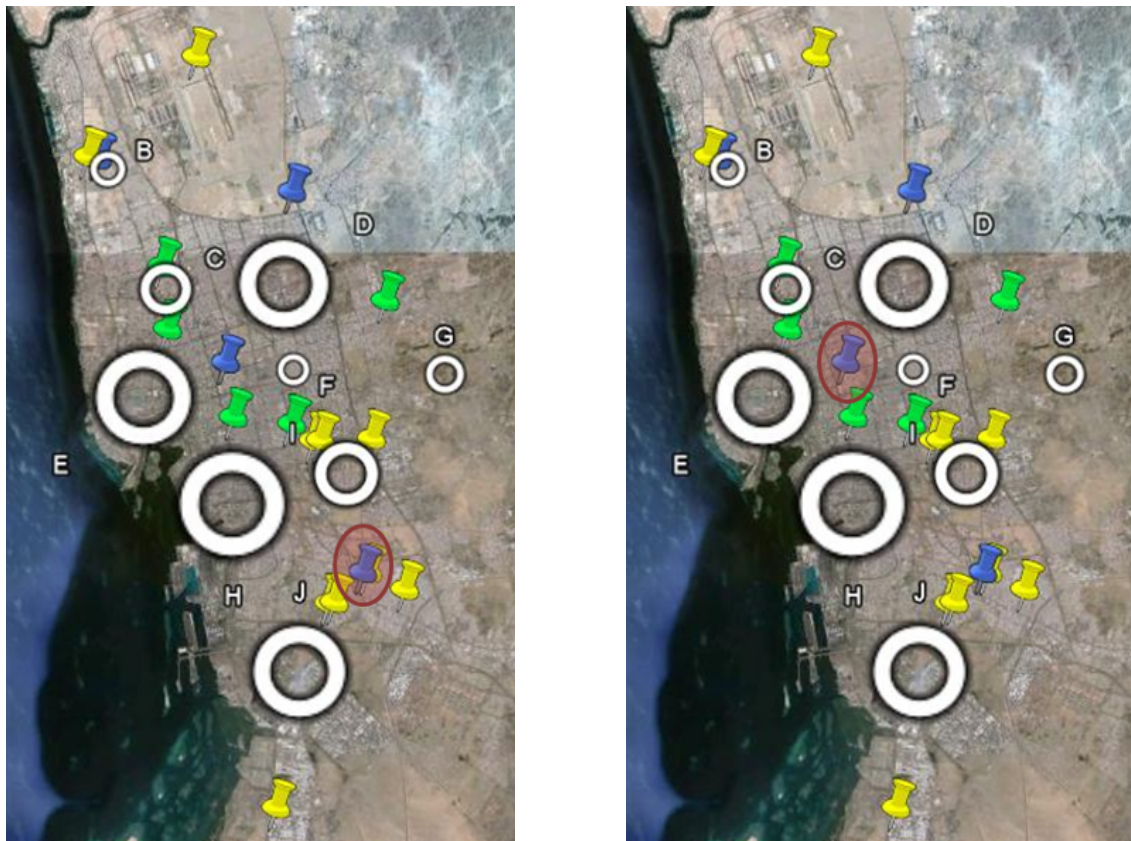
$$S_q^{total} = S_q \sum_{i=1}^{N_{CDA}} B_i \left( g(d_{il^*}) + \sum_{l=1}^{M_q} g(d_{il}^q) \right). \quad (3)$$

Both measures lead to differing results when applied to the specific data set provided by ALJ (note that this data set is obfuscated to protect proprietary information). Figure 5 provides the final results. **Notice that when optimizing total market share, it is optimal for the franchise to place a new location where the cluster of competitors are, while optimizing over total attractiveness leads to a location that is more central to the surrounding population. Both are intuitive results.**

## 5 Conclusions

By combining the current approach used by ALJ with insights from the existing literature and analysis by the Study Group an algorithm for the optimal placement of new retail outlet was developed.

Much further work could also be done, including: using network optimisation to enhance service; stochastic modelling of future population growth through, e.g., doubly stochastic Poisson processes; improving the attractiveness model to account for variation in showrooms; data set improvements (e.g. work with telecom companies, government health records, vehicle tracking agencies, introducing proactive loyalty programs); combining real options with location optimisation.



**Figure 5:** Optimal location based on (left) total market share (2) and (right) total attractiveness (3). Legend: representative customers (white circles - radius is proportional to population), competitors (yellow pins), current ALJ locations (green pins), and potential future locations (blue pins).

## Bibliography

- Achabal, D., W. L. Gorr, and V. Mahajan (1982). MULTILOOC: A multiple store location model. *Journal of Retailing* 58(2), 5–25.
- Ghosh, A. and C. Craig (1991). FRANSYS: A franchise distribution system location model. *Journal of Retailing* 67, 467–495.
- Hotelling, H. (1929). Stability in competition. *Economics Journal* 39, 41–57.
- Huff, D. L. (1964). Defining and estimating a trading area. *Journal of Marketing* 28, 34–38.
- Jain, A. K. and V. Mahajan (1979). Evaluating the competitive environment in retailing using multiplicative competitive interactive models. *Research in Marketing* 1, 217–235.
- Nakanishi, M. and L. G. Cooper (1974). Parameter estimate for multiplicative interactive choice model: Least squares approach. *Journal of Marketing Research* 11, 303–311.