Final Project

Finding the Best Location for Business

Tianyu Yang

Background:

Location is one of the most important factors that will influence the business. A good location will directly make the business successful. Thus, choosing a good location for business is very important. There are many factors that we usually consider about how to find the best location for one business.

1. Is the facility located in an area zoned for your type of business?
2. Is the facility large enough for your business? Does it offer room for all the, office, storage or workroom space you need?
3. Does it meet your layout requirements?
4. Do the existing utilities lighting, heating and cooling-meet your needs or will you have to do any rewiring or plumbing work? Is ventilation adequate?
5. Are the lease terms and rent favorable?
6. Is the location convenient to where you live?
7. Do people you want for customers live nearby? Is the population density of the area sufficient for your sales needs?
8. Is the trade area heavily dependent on seasonal business?
9. If you choose a location that's relatively remote from your customer base, will you be able to afford the higher advertising expenses?
10. Is the facility consistent with the image you'd like to maintain?
11. Is the facility located in a safe neighborhood with a low crime rate?
12. Is exterior lighting in the area adequate to attract evening shoppers and make them feel safe?
13. Are neighboring businesses likely to attract customers who will also patronize your business?
14. Are there any competitors located close to the facility? If so, can you compete with them successfully?
15. Is the area served by public transportation?

A lot more factors…

Objective:

In this project, I am aiming to build a roughly model to calculate the best location for business in Seattle area based on the data I collected. Since there are many factors will make influences on how to find the best location and the limitation of time and data sources, I decided to use only three factors in my project as an example to show the general idea about the project. Thus, the results I get may have a large error. Since this is a practical problem, I will use the gas stations business as an example. Therefore, the problem is how to find the best location for the gas station business in Seattle area. The three factors I used are the number of gas stations in this certain area, the populations in this certain area, and the price of land in this certain area.

Methodology:

In my project, four program languages are used. R, JavaScript, Html, and Sage.
R is using for calculation. Based on the data I typed in, the R will calculate the confidence interval for the gas prices in order to give a hint about how the gas station sets up the gas price.
JavaScript is using for drawing the maps based on the data. There are four kinds of maps, Seattle area map, population distribution map, gas stations distribution map and land price distribution map. Each map is a roughly map that using grids method.
Html has a best visual effect among all the methods I know. I use Html to display each map, and the information of the area will pop out when you move the mouse on the area.
Sage, I am using sage as a platform to run three methods above together.

Results:

Using the data and R programming the 95% confidence interval of the gas price in Seattle:
Regular Gas:

norm	3.862	4.059
basic	3.859	4.062
perc	3.858	4.061
bca	3.859	4.061

Premium Gas:

norm	4.120	4.345
basic	4.120	4.346
perc	4.123	4.350
bca	4.118	4.347

Midgrade Gas:

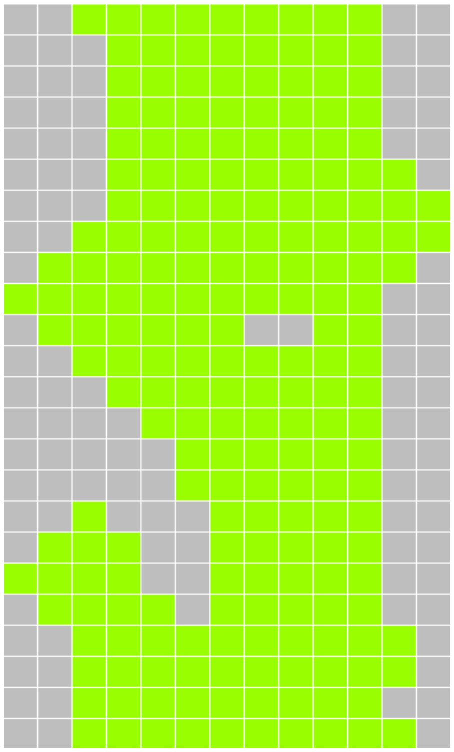
norm	4.020	4.229
basic	4.019	4.226
perc	4.024	4.231
bca	4.024	4.231

Diesel Fuel:

norm	4.019	4.235
basic	4.023	4.233
perc	4.028	4.238
bca	4.021	4.232

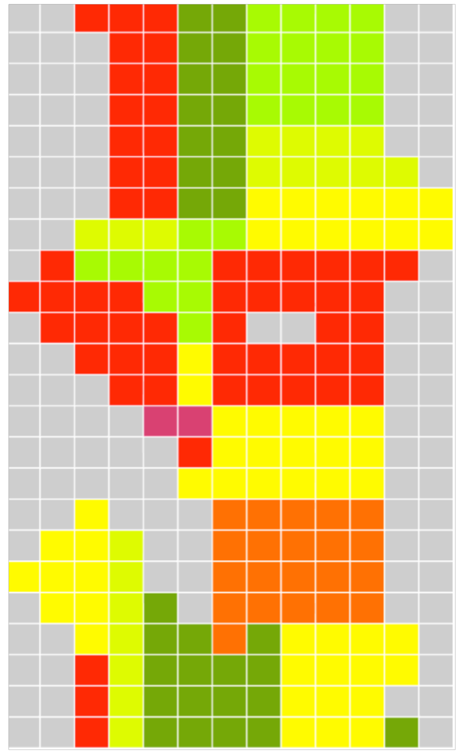
The table above gives us 4 kind of the confidence interval. We can see they are slightly different. That is because they are using 4 ways to resample the data in order to get the best one. The main method to calculate the price is using bootstrap, which is one of the most common resample methods to calculate the confidence interval.

The map of Seattle area:



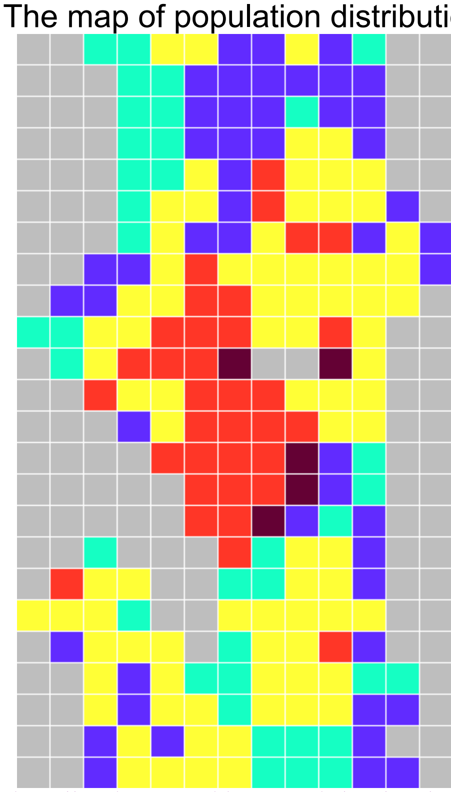
http://students.washington.edu/ty9/test/map.html

The map of land price distribution:



http://students.washington.edu/ty9/test/zola.html

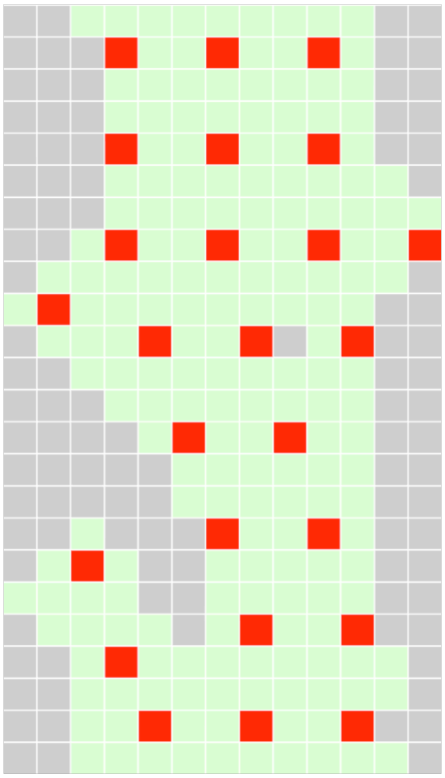
The map of population distribution:



http://students.washington.edu/ty9/test/population.html

The link under the picture could direct you to the actual html files (with animation).

The map of gas station distribution:



http://students.washington.edu/ty9/test/gasStation.html

The map above is about how many gas stations in the certain area. So I pick points in different location in the map, and the numbers showed on the red points are the numbers of gas stations in the red point area.

Formula:

$$\lambda = A\alpha + B\beta + C\gamma + \varepsilon$$

λ: Final score that we calculate by the equation, and the higher the score is, the better the location is good for gas station.

α: The populations in the certain area.

β: The number of gas stations in the certain area.

γ: The land price in the certain area.

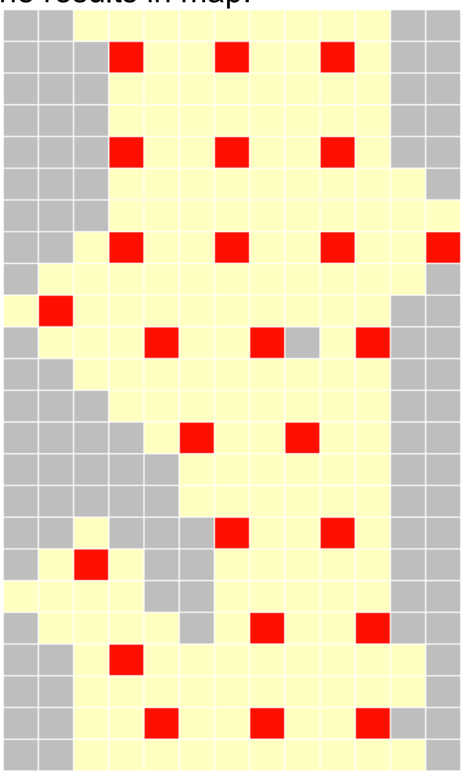
ε: Since these only three factors are considered, the model must exist an error that will be caused by other factors. Thus, ε represents the errors and other factors' influences.

A: the coefficient of the population in the certain area, since the project is about the problem of the locations of gas stations, the larger population implies more cars, and the more cars implies larger demand of the. But since not everyone owns a car, the population will not have a big effect on the location, thus in this project we consider it as 35 percent.

B: The coefficient of the number of gas stations in the certain area. This will associate with the topic about competition. The more gas stations the more choices

people will have. Thus in this project we consider it as negative 25 percent.
C: The coefficient of the land price in the certain area, since people whoever live in the rich area are with a rich economic background. This points out two things, first, they may have more chance to own more cars, and the cars they owned may need more gas (low mpg car). Second, since most people are richer, they may have a high level of consumptions, and they will not concern about go farther to get cheaper gas. Thus, in this project we consider it as 70 percent.

The way to calculate the weighted coefficient: using the multi-linear regression model to regress λ on three factors, α, β and γ. Then we can get the coefficients of three factors. In this project, since I did not consider all or most factors that will influence the location, I set up an error is equal 20%. Thus, the sum of weighted coefficients is equal to 0.8 instead of 1.
The results in map:



http://students.washington.edu/ty9/test/result.html

Conclusion:

Based on the score of the λ, we find out that the highest score is around 3000. By looking back to the real Seattle map, the locations are Green Lake, Westlake and Kerry Park areas.

Consideration:

The model I built is only a roughly model. Since I only consider 3 factors in order to calculate the location, the result I got will contain a big error. However, if we can add more data in this model, the result would be much better and more correct.