Executive Summary

This report is aimed at improving sales and marketing performance for a national retail chain company by analysing its sales data and current shop distribution locations. For the purpose of reducing costs associated with logistics while enhancing total revenue as well as customer base, we introduce the following recommendations based on what have been found through the investigation of both sales and logistics.

- **Sales**

Q1. To determine the retail shops that demonstrate the best and worst performance respectively according to their total revenue and product mix, coloured circles and bar charts in ArcGIS are adopted to depict these extreme values:

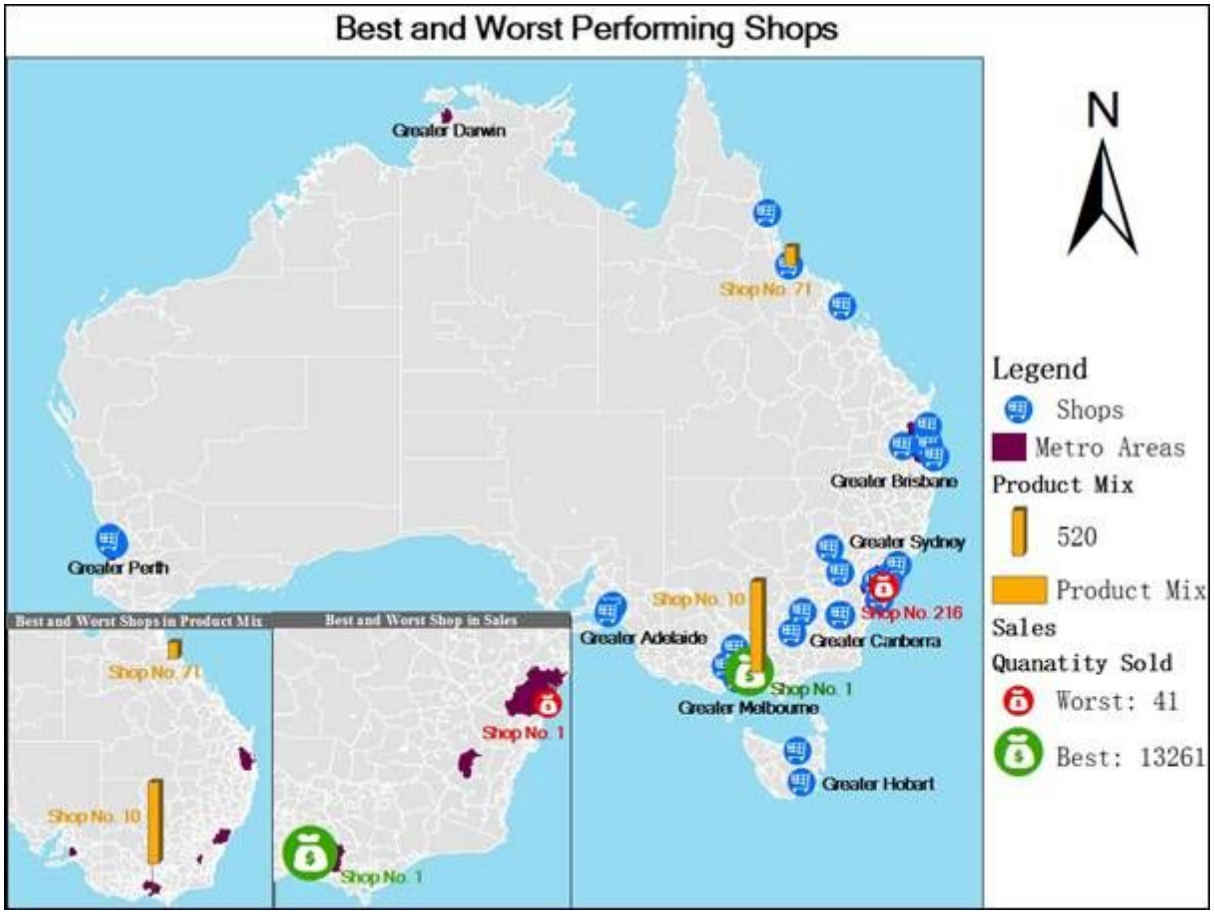
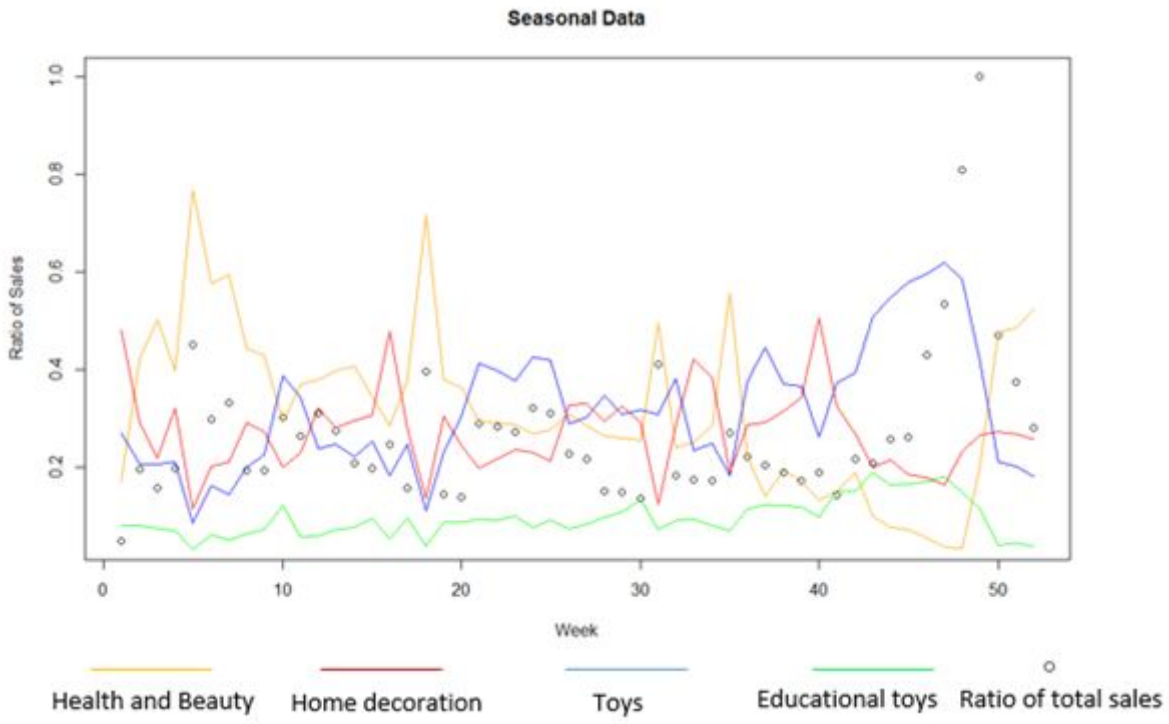


Figure 1.1 Best and Worst Performing Locations in terms of Sales and Product Mix

As can be seen from the map, Shop No.1, which is located in Melbourne, enjoys the highest revenue whereas Shop No.216, which sits in Sydney, experiences the least. With regards to product mix, Shop No.10 in Melbourne has the most variation while Shop No.71 in Cairns has the smallest number of varieties. Although it is acceptable that the best performing shop

is seated in a metropolis, it is surprised that the least performing store is also seated in a metropolis. Therefore, it should be kept in mind that we cannot choose expanding locations solely based on whether the city is a metropolis or not.

Q2. In order to generate the seasonal sales for the whole country, the data are plotted as below in a time series. The Black points represents the ratio of total sales for that week to the maximum of total sales for any given week over the course of the year. The line graphs represents the ratio of the sales from any given category to the total sales of that particular week.



-

Total sales shown in the graph seems fairly consistent throughout the year averaging at 27% of maximal sales however has some spikes that occur during Week 5 at 45% of maximal sales, Week 18 at 39.5% of maximal sales, Week 31 at 41% of maximal sales and then abnormally higher sales occurring around Christmas (Weeks 46-51) at 43%, 53%, 80%, 100%, 47% and 37% respectively.

Apart from seasonal variations, it is also crucial to investigate variations on geographic level. By plotting sales data for each state using pie charts and gradient ramp, the geographic variations can be shown on the following map:

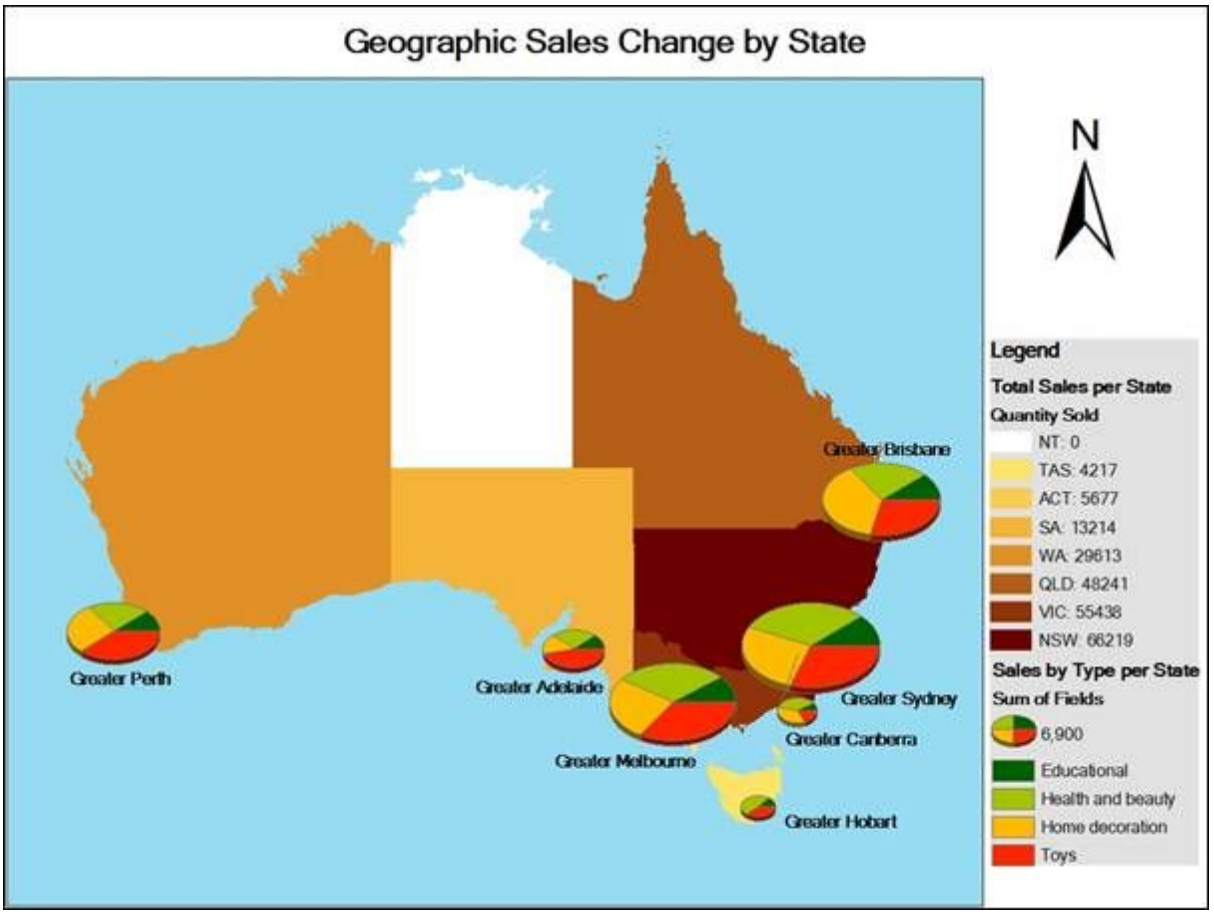
- 

Figure 1.2 Sales per Product Type categorized by States

The map shows that in every state, *Educational* products performed worse than the other categories suggesting that consumers are primarily not participating in commercial educational products. *Health and Beauty* proportions seems to be correlated with total sales suggesting that in larger metropolis areas, where higher levels of wealth and education reside, consumers are more willing (or able) to invest in their health and/or physical appearance. *Home Decoration* was consistent across many states, however Brisbane has a significantly higher proportion and Adelaide a significantly lower proportion. In Brisbane's case, the higher prioritisation of decoration may be due to high numbers of retirement or holiday properties where the property may be considered an art work or a project rather than merely a living environment. In Adelaide's case, this lack of decoration investment is convolved with a significantly larger proportion of *Toy* sales. This may indicate either that Adelaide has a higher density of younger families with higher requirements for toy consumption.

Q3. To find out the new shop locations, the most deterministic two factors, income level and population labour force, are relied on after computing the regression model for the sales of

the current shops. 14 places are chosen to be the candidate locations nationwide. Note that instead of putting candidate shops near the places where upper classes live, they have been allocated to the places where the bourgeoisie live (corresponding to the second income level illustrated in Figure 1.4 shown in the technical report). This is probably because these shops are not luxurious or high-level enough to attract upper classes and purchasing power of middle classes in Australia is increasing steadily. After further filtering, the final 5 locations are selected from the 14 candidates.

When considering which type of product should be sold in each newly-opened shop, Figure 1.2 containing past sales distribution patterns is referred to again. The new shop located in West Australia should stock with more goods that are categorised into toys whereas the three shops opened in Queensland should replenish more products associated with home decoration. More health and beauty products should be delivered to the shop opened in Tasmania.

- **Logistics and Distribution**

Q4. As for identifying the locations with potentially problematic inventory control, each shop's inventory level is compared with its sales level. It is found that more than half of the shops enjoy a normal inventory level. Besides, there are far more shops with too many stocks than those with too few stocks. Abnormally high inventory level may cause high storage costs while a low one may suffer shortage problems.

When analysing the stock control quality in terms of states or territories, it is found that all the shops in Australia Capital Territory and Tasmania are experiencing a regular inventory level while West Australia and Queensland are the only two states where shops only face the issue of having insufficient stocks. Shops in the remaining states are exposed to a mixture of both having too many and too little inventory stored.

Q5. When identifying the optimal number and locations of national distribution centres, the process is similar to choosing new shop locations except that the deterministic factor is changed to sales level. Also, based on the fact that most of the products are imported, the final locations are selected near the airport or docks in five major metropolis areas across Australia.

Q6. After taking into account several factors like vehicle capacity and fuel costs, a financially efficient routing plan for all the shops located in Sydney is generated, which requires two articulated trucks, one large rigid truck and one small rigid truck. The resulting total cost is $560,428 which is much lower than the plan with three articulated trucks.

Technical Report

❖ Data Justification and Assumptions

All the spatial data employed in this report are provided by Dr Adrian Ellison and Dr Richard Ellison on USYD Learning Management System under ITLS6107.

Apart from the assumptions given in the instruction, we make the following assumptions additionally. The first assumption that has been made is that the fuel cost per 100 kilometers is constant no matter how many kilograms of cargo the vehicle is bearing. However in reality, fuel consumption per hundred kilometers tends to increase when more and more cargo is carried on by the vehicle. What is more, when determining the seasonal patterns of the sales of this company across Australia, the year 2015 and 2016 are artificially divided into the following parts:

- ➢ Summer Sales 1: 2015.01.24 -- 2015.02.28
- ➢ Autumn Sales: 2015.03.07 -- 2015.05.30
- ➢ Winter Sales: 2015.06.06 -- 2015.08.29
- ➢ Spring Sales: 2015.09.05 -- 2015.11.28
- ➢ Summer Sales 2: 2015.12.05 -- 2016.01.16

The reason why we do not combine the two summers is because the analysis of seasonal changes is straightly based on time series. Besides, when checking the attributes table of the data given, although many of them have the total number column that can be used as an approximation for total population, they vary slightly because the datasets are collected by a number of organizations using a variety of different methods with potential errors. For this reason, it is necessary to demonstrate clearly that numbers in the total number column under 'SA2_AgeSex' is employed to approximate the total number of population. The fourth assumption is about choosing useful independent variables for the regression analysis. Specifically, when calculating the labor force (one of the inputs in regression), we assume

that it is obtained by subtracting people who are not in the labor force and who are not applicable in this area from the total number of population. When developing the routing plan for shops in Sydney, it has been found that all the shops in Sydney are located in metropolitan areas. That is to say, instead of using 70km/h as the average vehicle speed, all the vehicles should have an average speed of 40km/h. In addition, when designing the vehicle routing plan, it has been assumed that the cost per unit time is $0.20, which means that the truck driver is paid $0.20 per minute.

- **Map Creations**
- **Sales**

Q1

Similar to what have been done for geographic variation, we add up all the sales occurred in one season for one state or territory and repeat the process for the remaining seasons. Through the application of bar charts, the map regarding seasonal patterns in different states or territories is given by:

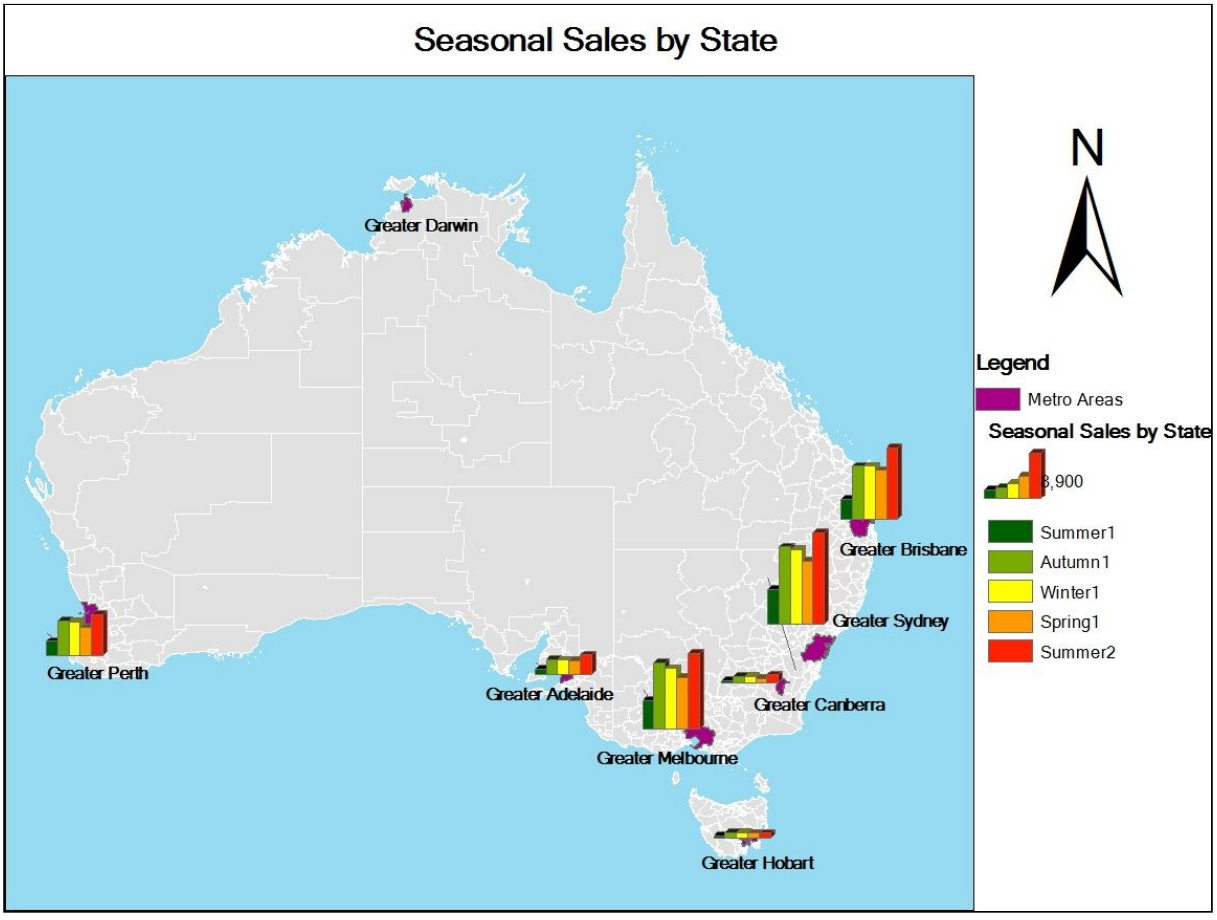
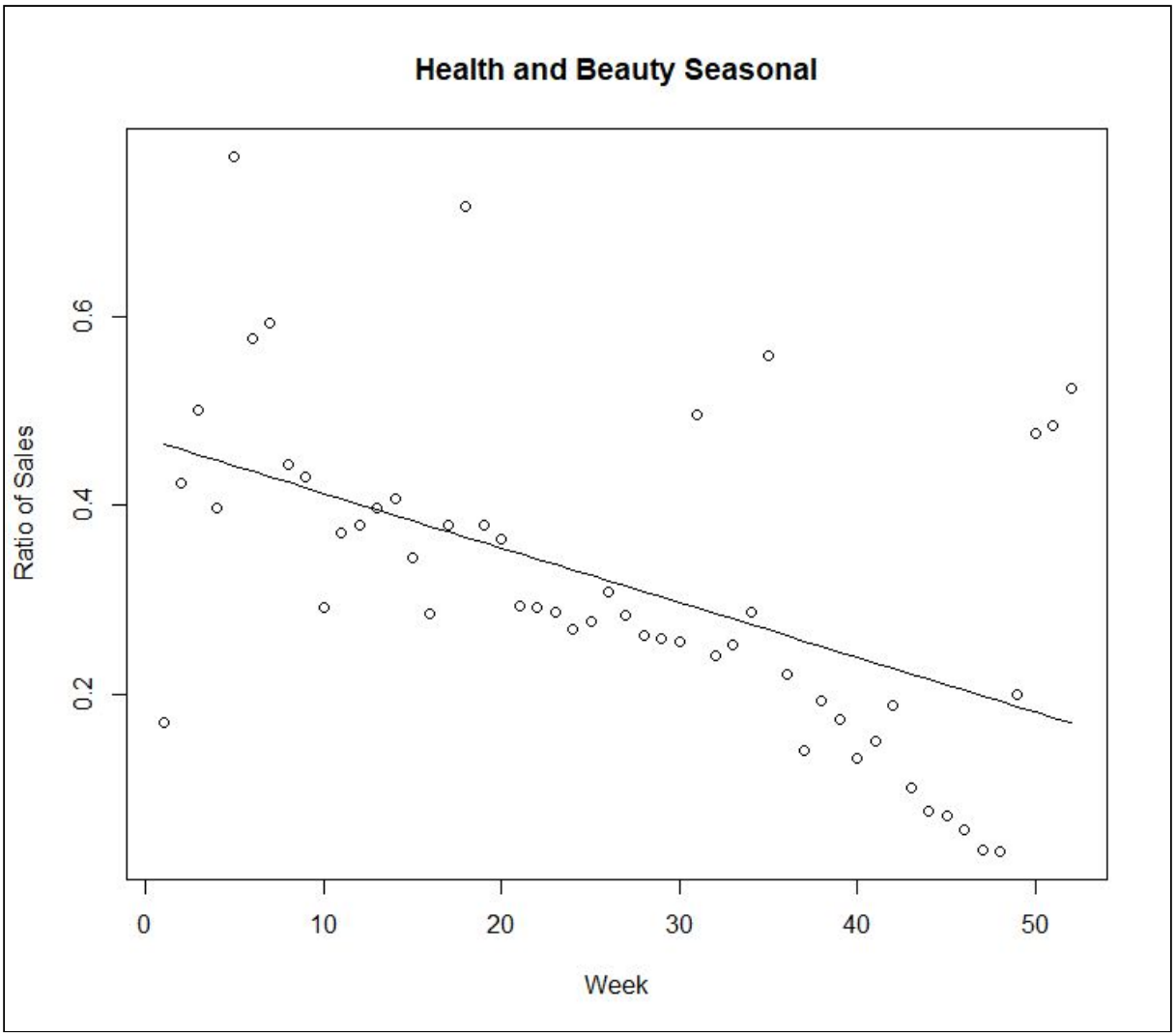
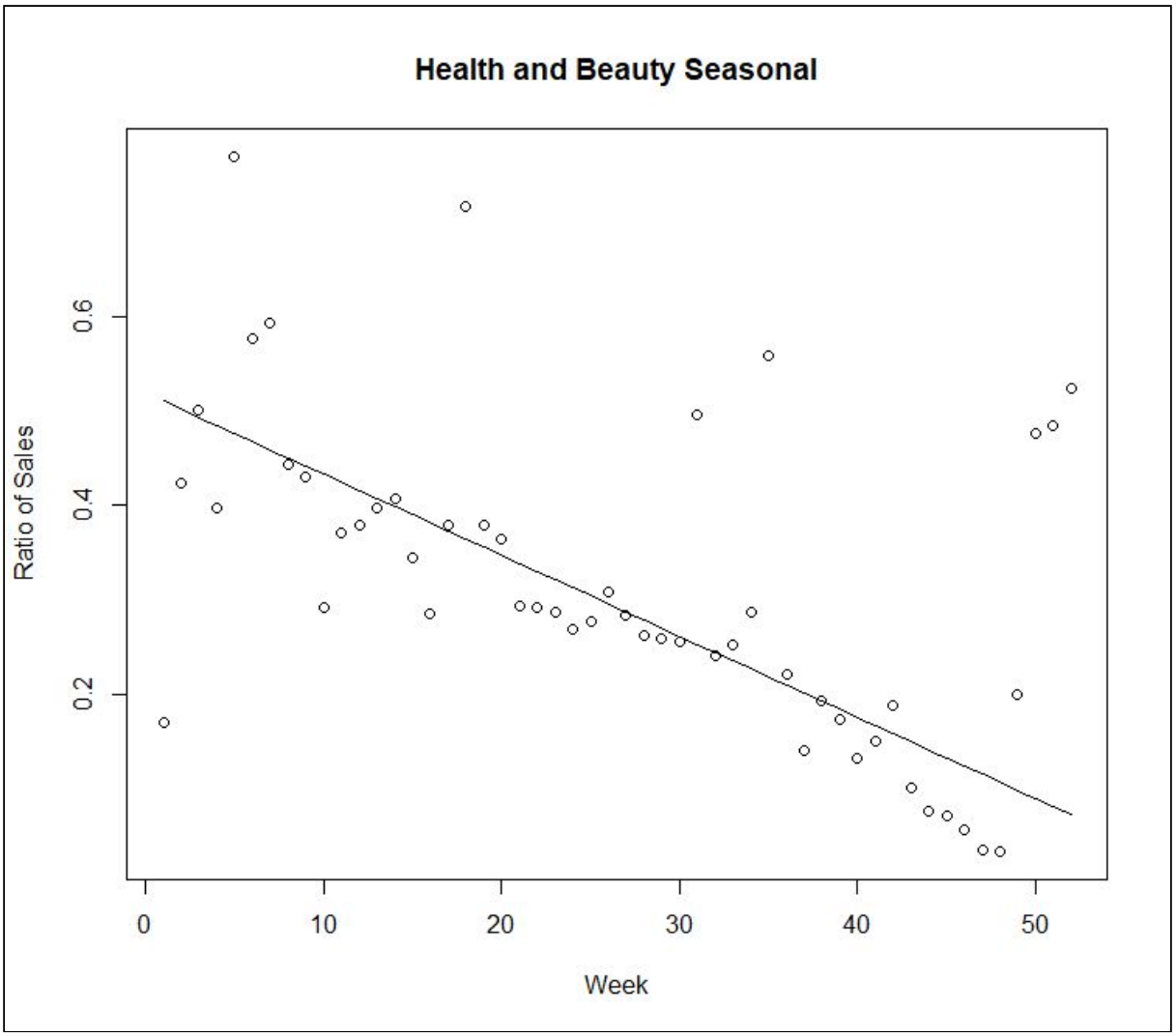


Figure 1.3 Seasonal Variation categorized by shops

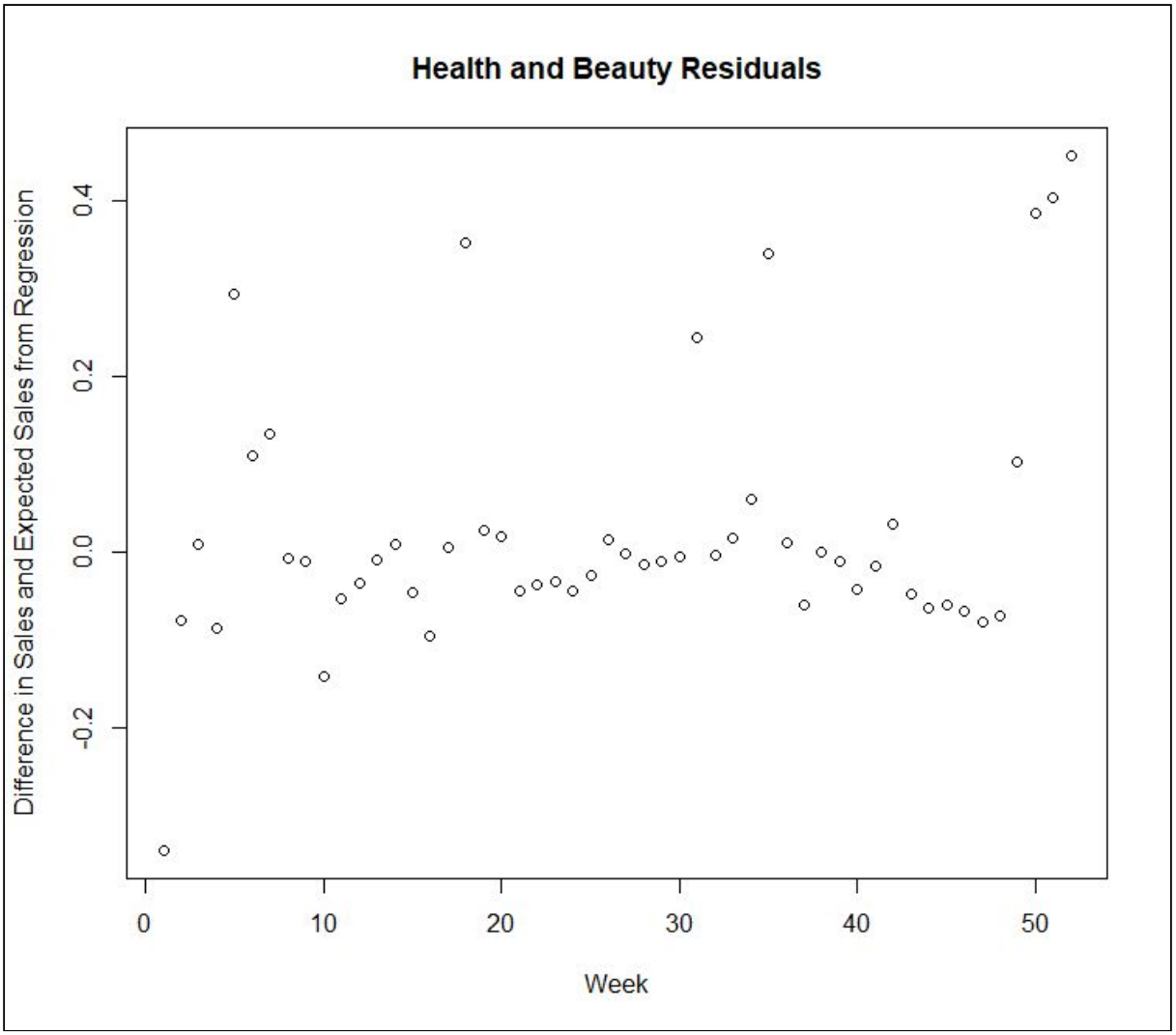
Like what we have expected, the general seasonal variations for different states or territories enjoy quite similar patterns. In particular, the highest sales revenue takes place in the second summer session, which starts from 2015.12.05. Interestingly, the lowest sales revenue also occurs in summer, which is the other session that starts from 2015.01.24. When considering the overall trend in response to seasonal changes, the pattern begins at the rock bottom and then rises dramatically followed by a downward tendency, which bounces and hits a peak at the end.

The spikes that occur at week 5, week 18 and week 31 are assumed to be specific discount offers that expands the number of items that are sold. What can be noted about these particular weeks is that the “Health and Beauty” categories spike simultaneously with week 5 attributing 76.9%, week 18 attributing 71.6% and week 31 attributing 49.6% of sales to the “Health and Beauty” products. These measures were then compared to a linear regression model.


Health and Beauty Seasonal
Ratio of Sales
0.2
0.4
0.6
0
10
20
30
40
50
Week



These graphs represent linear regressions of “Health and Beauty” with respect to the week. The first graph is taken over all weeks and the second graph has omitted weeks 49, 50, 51 and 52 in order to achieve a more realistic linear regression ignoring the christmas seasonal spike.



The residuals were plotted and measured and spikes can be seen occurring at the points of higher number of sales.

“Educational Toys” is the least popular category throughout the whole year except for “Health and Beauty” between week 44 and week 49.

“Toys” has a large and wide bump a few weeks out from christmas as expected because of purchasing presents for young children. Both “Toys” and “Home decoration” both sharply decline in a trough whenever both sales and “Health and Beauty” spike. This is suspected to be caused by a disproportionate increase in sales in the “Health and Beauty” sales relative to the other categories, rather than a decline in sales in the respective categories.

In terms of geographic variation, we divide the whole country into 6 states and 2 territories according to the administration partition. As a consequence, in order to determine the total sales for each state or territory, we first need to identify which shop belongs to which state or territory. After that, sales figure for each state or territory can be easily obtained by adding up all the sales of the shops located in that state or territory. We use gradient ramp to show how sales data vary in different geographic locations. By adding up all the sales of a particular product type in each state or territory, we can get the figure of sales for each product in one state or territory, which is visualised through the use of pie chart.

Based on what have been analysed so far, we are now able to determine where to open new shops and which types of products to sell there. To do so, a multiple regression is run in Excel with the following chosen independent variables which are considered to be relevant: total income, number of labor force, total population, number of males, number of females and average age. As demonstrated in the ANOVA table below, total income and number of labour force are proved to be the most relevant factors in explaining sales because their p-values are less than the level of significance, which is 0.05. Hence, new shops tend to be opened around places incurring both high level of income and large number of people being

employed.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.628846							
R Square	0.395447							
Adjusted R Squ	0.330564							
Standard Error	1827.477							
Observations	68							
ANOVA								
	df	*SS*	*MS*	*F*	*gnificance F*			
Regression	6	1.35E+08	22573492	8.11102871	1.85E-06			
Residual	62	2.07E+08	3339674					
Total	68	3.43E+08						
	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	1399.138	2788.49	0.501755	0.617617	-4174.97	6973.249	-4174.97	6973.249
total_in	-0.00054	0.000237	-2.27989	0.026066	-0.00102	-6.7E-05	-0.00102	-6.7E-05
labor_in_f	1.794472	0.413318	4.341623	0.000053	0.96826	2.620683	0.96826	2.620683
pop	-0.39031	0.466908	-0.83595	0.406391	-1.32365	0.543022	-1.32365	0.543022
male	-0.497	1.028521	-0.48322	0.630642	-2.55299	1.558983	-2.55299	1.558983
female	0	0	65535	#NUM!	0	0	0	0
average_ag	20.9216	76.08777	0.274967	#NUM!	-131.176	173.0189	-131.176	173.0189

14 places are chosen to be the new candidate shops based on the income level as well as the labour force density, which are shown in the map below. Note that instead of putting candidate stores near the places where upper classes live, they have been allocated to the places where the bourgeoisie live (corresponding to the second income level illustrated in Figure 1.4). This is probably because these shops are not luxurious or high-level enough to attract upper classes and purchasing power of middle classes in Australia is increasing steadily.

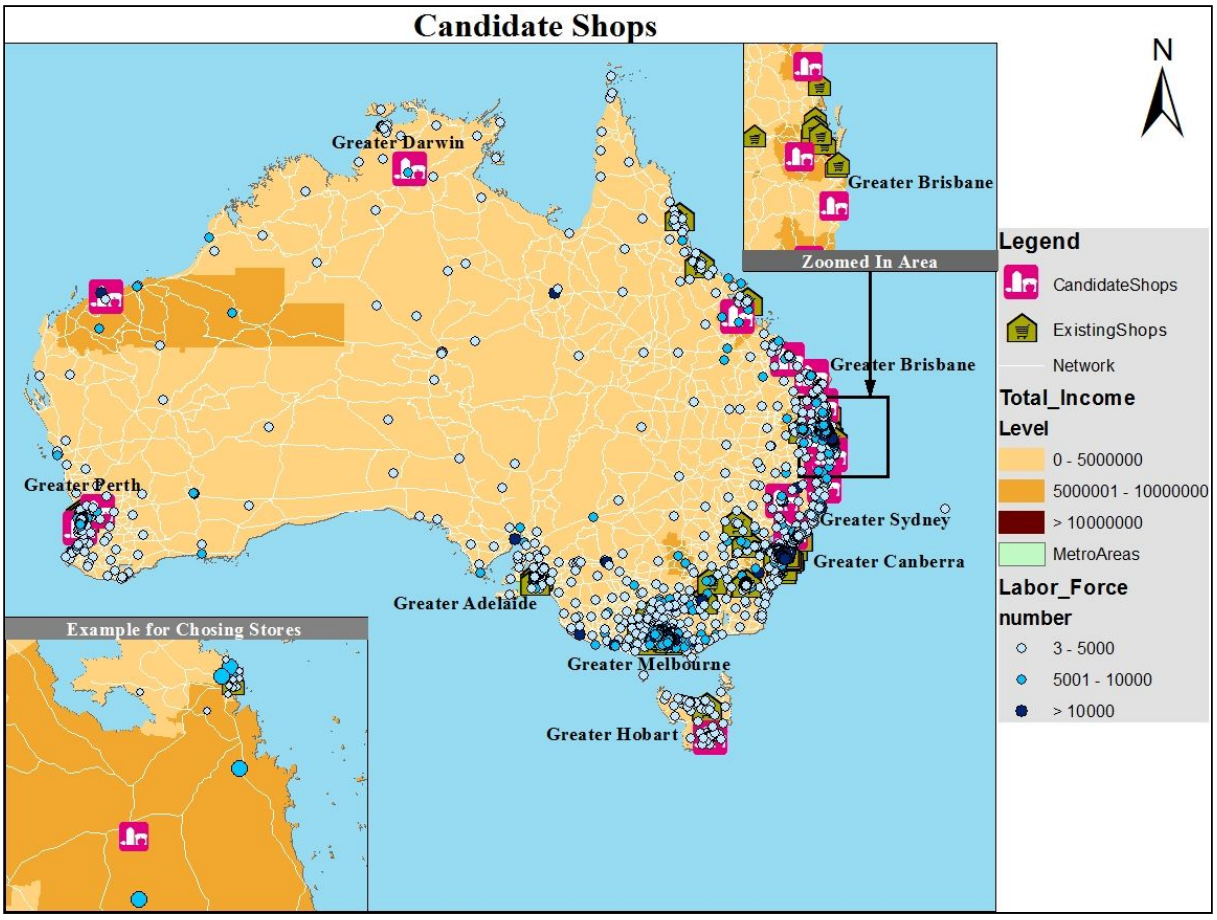


Figure 1.4 Fourteen New Candidate Shops Chosen

We further filter the candidate shops in order to determine the top best ones. Now only the number of labor force is taken into account since it has a far more significant p-value than that of total income from the regression table, meaning that labor force should account for more sales than average income. In addition, different points of labor forces shown on Figure 1.4 are treated as the approximation of potential customers. Since some of the candidate stores are located in metropolitan areas whereas some are not, we average the vehicle speed in metropolitan areas and that in other areas as the average speed utilized (i.e. (70 km/h + 40 km/h)/2 = 55km/h). Further, it has been artificially decided that shops that are going to be built should be arrived within 30 minutes from the points of labor force under a speed of 55km/h.

Figure 1.5 Five Best Locations for New Shop Opening

As for logistics and distribution aspect of this company. We first try to identify any locations with potentially problematic inventory control through the comparison of inventory level in each store with the total number of sales in that store. It has been artificially defined that any ratio of inventory level to sales level between 21 to 40 represents an unproblematic inventory control (labelled in green), whereas any ratio below 21 or above 40 indicates an irregular inventory level (labelled in red and blue respectively).

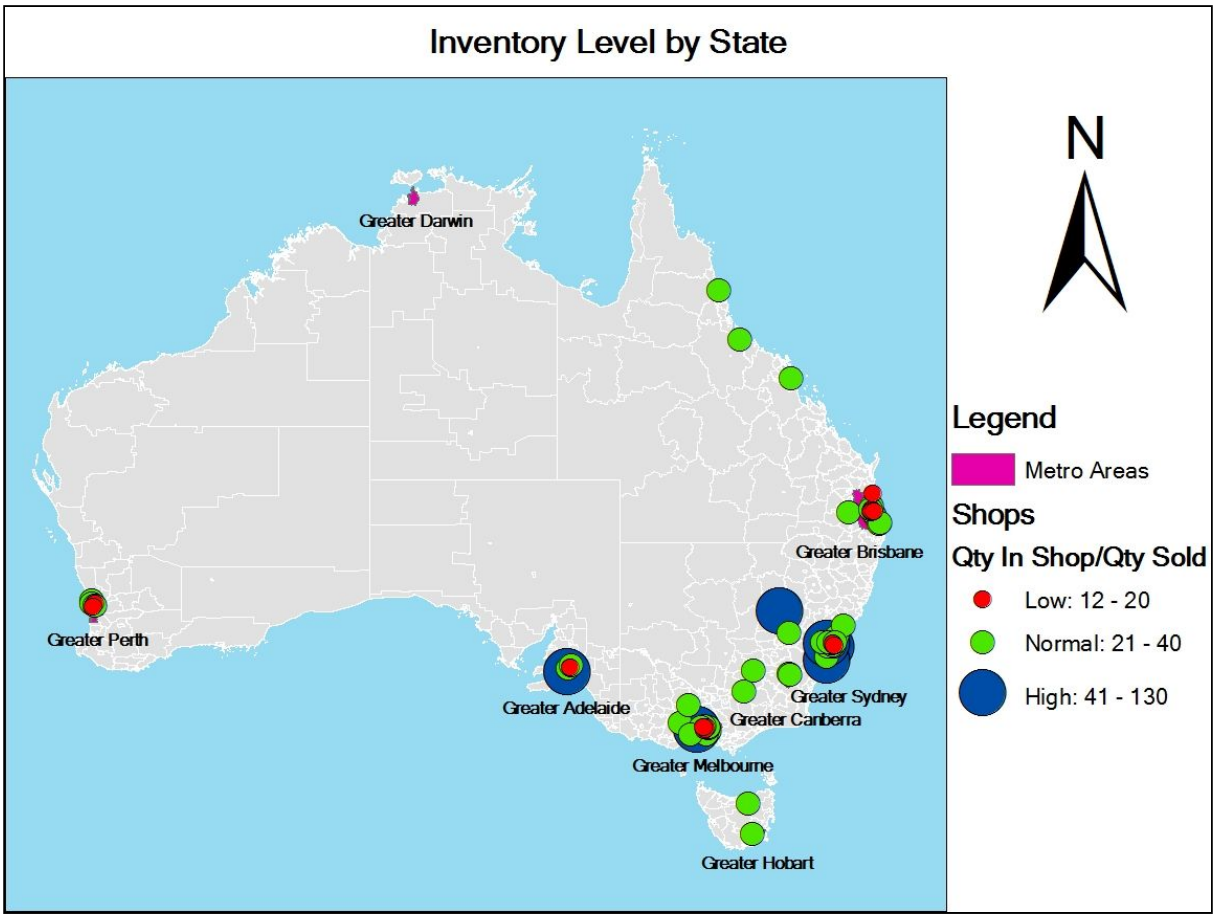


Figure 2.1 Inventory Level per Shop

Fortunately, the map above clearly illustrates that more than half of the shops in this company enjoy a normal inventory level. Besides, there are far more locations with too many stocks than locations with too little stocks, which may possibly reduce the storage costs but can also cause serious shortage in supply when comparing with the demand. As a result, if products cannot be replenished in time, this may depress the sales in those stores. When analyzing the stock control quality in terms of states or territories, we find that all the shops in Australia Capital Territory and Tasmania experience a regular inventory level while West Australia and Queensland are the only two states where shops do not face the issue of having too much stock. Shops in the remaining states are exposed to a mixture of both having too many and too little inventory stored.

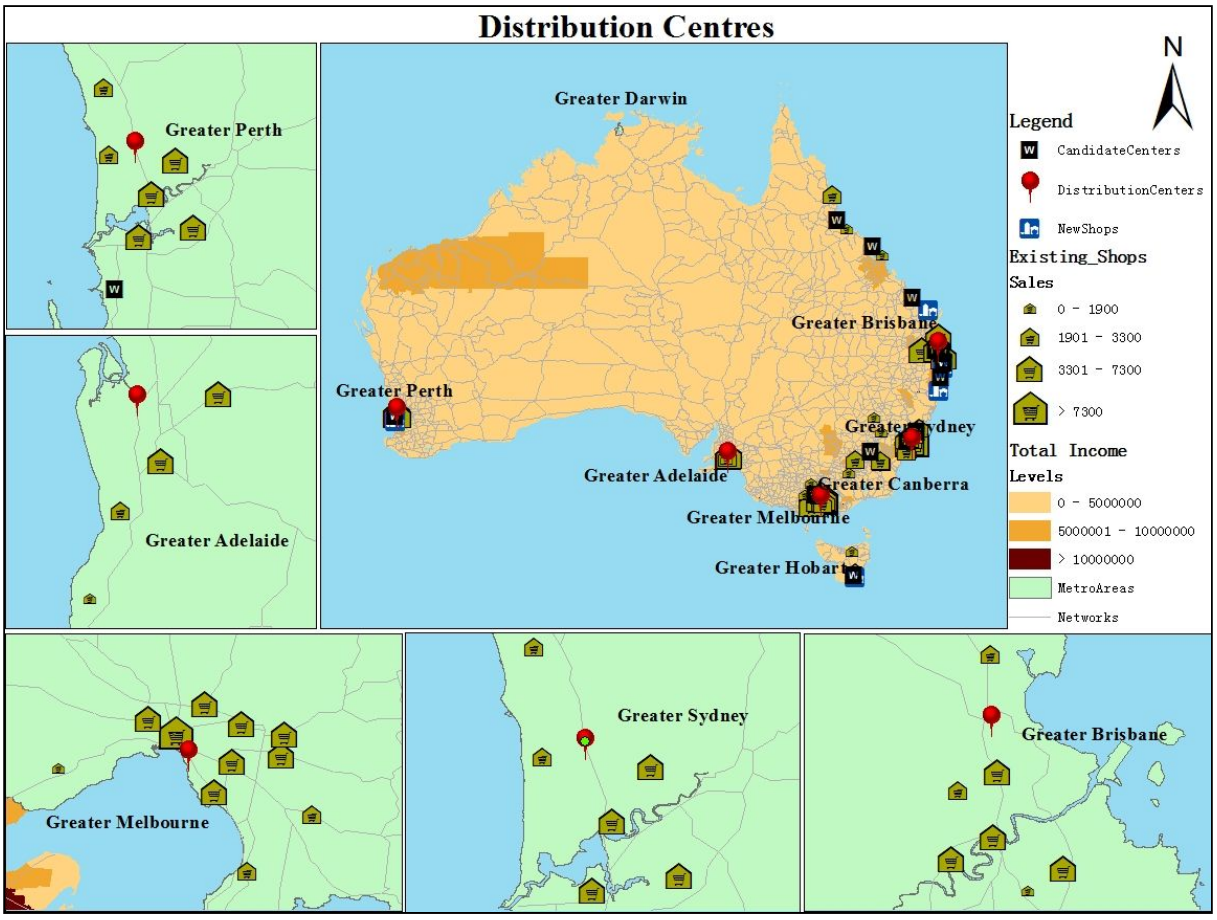
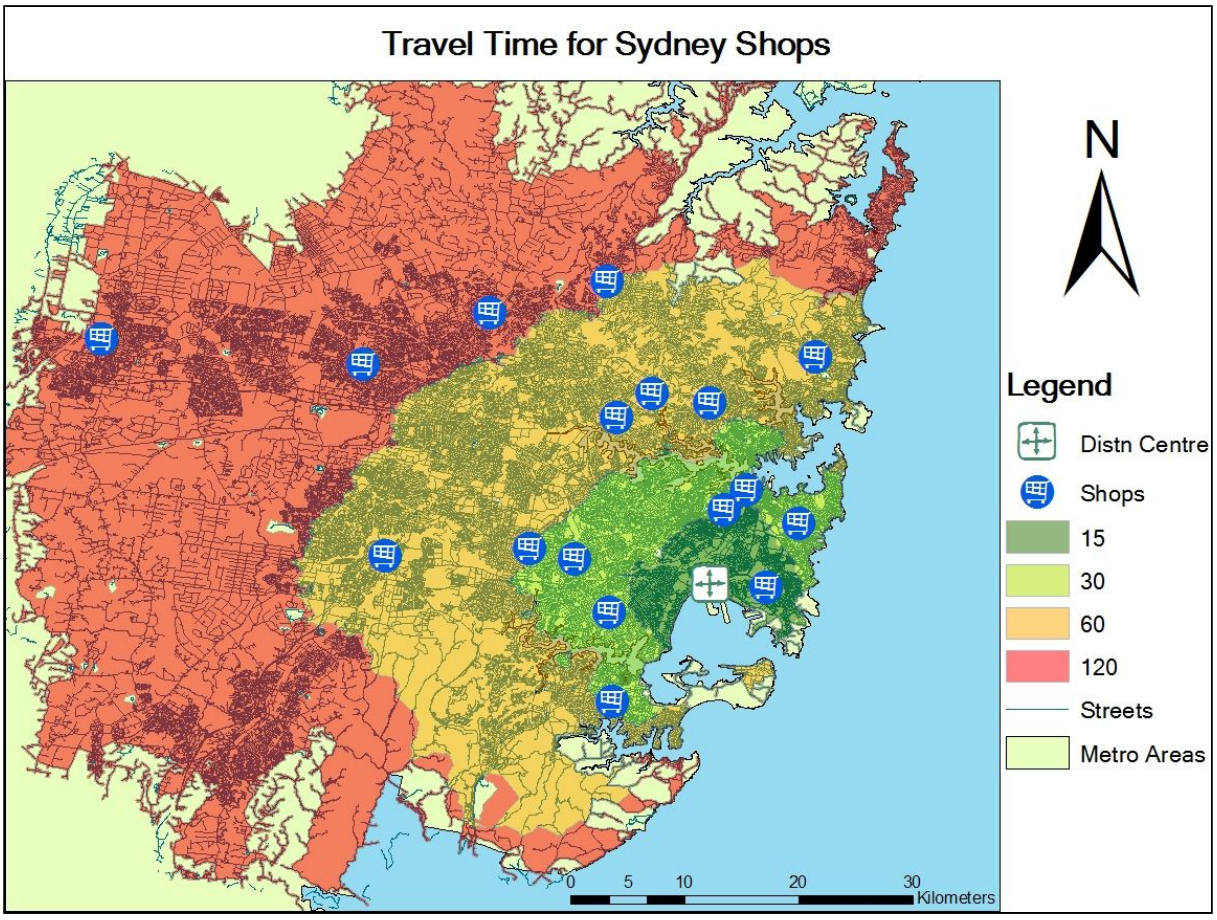


Figure 2.2 Locations of the Five Optimal National Distribution Centres



This maps potential first stop candidates in the inner sector for potential routes and provides the limitations on the chosen routes in the furthest sections.

The above map illustrates how long does it take for vehicles to arrive at each store from the distribution centre, which is useful for us to further determine the departure time of each truck. In particular, shops within the dark green areas can be arrived within 15 minutes whereas it takes around 30 minutes to reach those stores located in light green areas. Furthermore, shops seated in the yellow areas needs approximately one hour to be reached and it takes the longest time, which is 120 minutes to get to the shops located in the furthest sections which are labeled in red on the map.

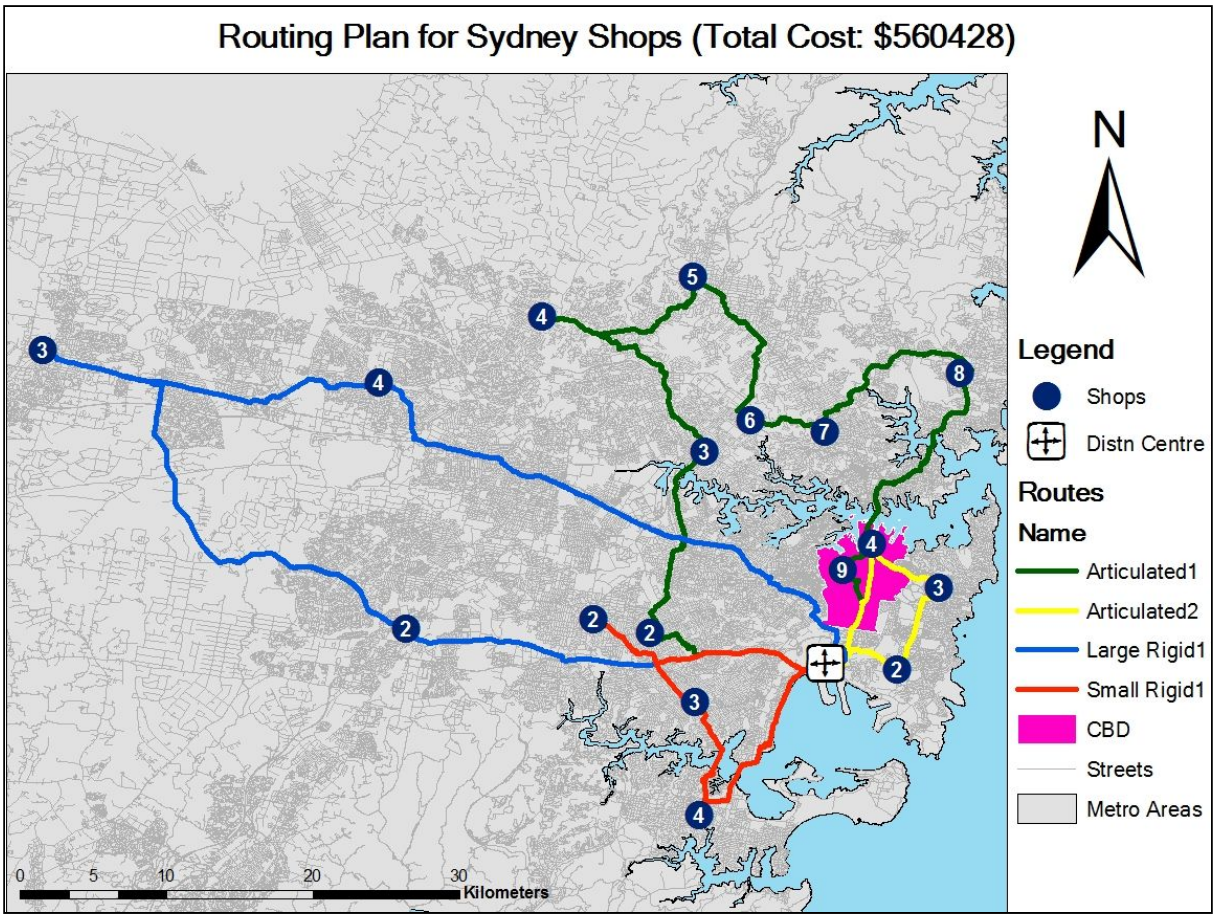


Figure 2.3 Routing Plan with Number of Vehicles Identified for Shops in Sydney

When developing the routing plan for all the shops located in Sydney, the following factors are taken into account:

- Capacity of each vehicle
- Maximum number of shops a truck can reach
- Cost per unit of time
- Cost per unit of distance (km)
- Service time (i.e. time consumed for loading and unloading)
- Fixed cost (i.e. purchase cost of vehicles)

As demonstrated from the map, two articulated trucks along with one large rigid and one small rigid truck are considered to be enough for serving all the Sydney shops. Particularly, the two articulated trucks serve 11 shops in total, while the large and the small rigid vehicle only provide cargo to 3 shops individually. The departure time of these 4 vehicles are all set to be 6 am, however, the actual departure time for Articulated 2 and Small rigid 1 have been changed to 7 am and 6:18 am automatically by ArcGIS, which may be due to the following reasons. As shown on the map, before reaching the shop within the CBD area, Articulated 2

transports the cargo to the other 2 shops located outside the CBD first. The possible reason for the deferred departure time of Articulated 2 and the choice of serving shops outside the CBD area first is that if Articulated 2 leaves the distribution center at 6 am, it will still arrive at the shop located in CBD area an hour earlier than the opening time, which is 10 am.

As for the reason why the departure time of small rigid 1 is set at 6:18 am by ArcGIS, we consider it might be that if the small rigid 1 leave at this time, it will be not too early and not too late to reach the first shop among the remaining 3 shops which all open at 7 am and haven't been covered by the other 3 vehicles.

Actually, another routing plan which consists of 3 articulated trucks can also serve all the shops in Sydney. Nonetheless, without considering other costs associated with delivery, purchasing three trucks in this type has already cost the company approximately $600,000. Yet, the plan that has been proposed first has a total cost of only $560,428, which makes it the final decision for vehicle type selection.

Conclusion

By utilizing RStudio to process raw data and ArcGIS to draw the main maps, this report has assisted a national retail chain company to improve its sales and distribution network based on the analysis of its current performance. In terms of the sales results, Shop No.1 in Melbourne and Shop No.216 in Sydney has experienced the best and the worst performance in 2015. Besides, Shop No.10 located in Melbourne and Shop No.71 seated in Cairns enjoy the most and the least variety of product mix respectively. When investigating the sales data state (territory) by state (territory), it has been found that New South Wales has accounted for the biggest share of this company's revenue whereas Tasmania has occupied the least. Seasonal variations are quite similar across different states or territories, where consumers are willing to purchase more products at the end of the year but are more reluctant to buy at the beginning of the year. As for where to open new shops, only five places have been considered to be the best locations in the end after investigating their demographic data, especially total income level and total number of labor force. Among which, three of them are located in Queensland with the other two located in West Australia and Tasmania separately. Dividing the total sales according to the type of products for each state or territory yields results suggesting that the three new shops in Queensland should replenish more toys to sell,

whereas the other two shops in West Australia and Tasmania need to import more products under home decoration as well as health and beauty separately.

With regard to logistics and distribution, northeastern Australia and Tasmania enjoy a normal level of inventory. Nevertheless, problematic inventory control has been found in the west and east of Australia where they got insufficient stocks. Shops located in southeastern Australia were found to be overstocked. Moreover, in view of the fact that most of the goods are imported and there are five metropolitan areas with high distribution density of shops, five locations have been selected out to be the optimal national distribution centres based on their distance from local transportation and current shops. Lastly, to serve all the shops in Sydney, two articulated, one large rigid and one small rigid trucks have been purchased after considering factors like total costs and vehicle capacity.

Technical Part (Sales and distribution center select)

For both parts to determine the optimal sales shops and distribution centers, the 'Network Analyst' is used to generate the locations.

Using multiple regression

References

aus5vgd_l.shp from GeoScience Australia National Data

shopsales.csv from Retail Sales Data with Locations

shoplocations.csv from Retail Sales Data with Locations

AusCapitalCBDs.shp from CBD Areas

MetroAreas.shp from Metropolitan Area Boundaries

SYD_streets fromSydneyInfrastructure.gdb

Sydeny_LGAs from CensusData.gdb

SA2_PersonalIncome from Australian Demographic Data by SA2

SA2_EmploymentStatus from Australian Demographic Data by SA2

SA2_AgeSex from Australian Demographic Data by SA2

Technical

SalesQ3

Using a multiple regression with the following chosen independent variables: total income, number of labor force, total population, number of males, number of females and average age, we find that total income and number of labour force are most relevant factors in explaining sales because their p-values are less than the level of significance, which is 0.05. Based on the result of multiple regression, we choose 14 candidate shops in areas with both the middle-income level and the high labour-force level. Then, we further filter the candidate shops. Now only the number of labour force is taken into account. With the aid of *Covert Polygons to Points*, we convert labour force in each SA2 area into demand point. Founded on 14 candidate shops and demand points created before, we use *New Location- Allocation* to find the top 5 shops .

Distribution Q2