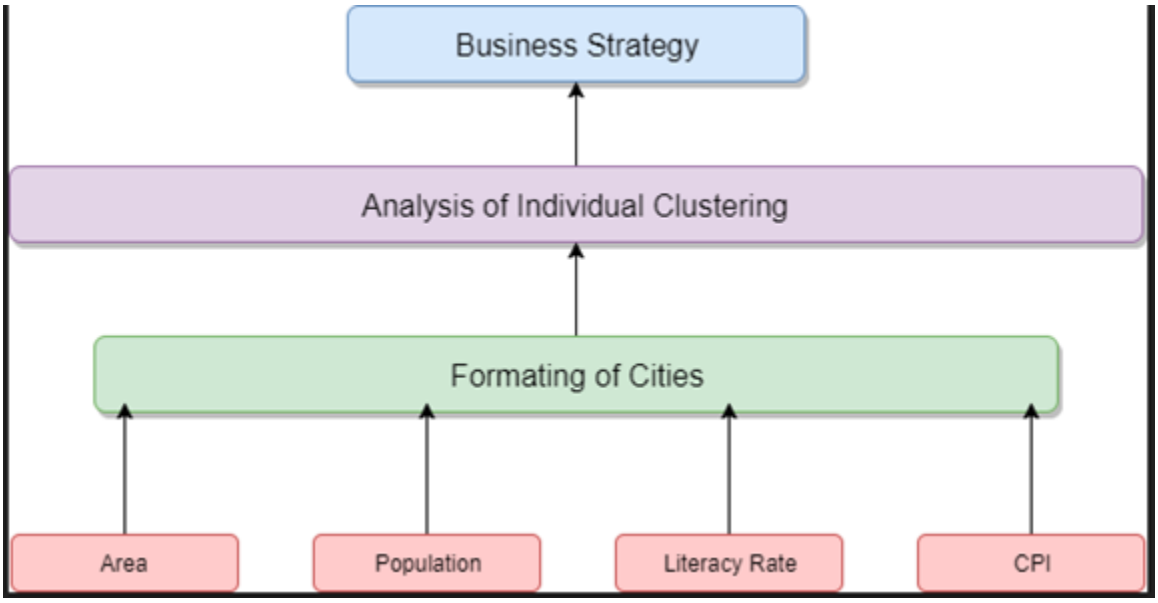
MARKET SEGMENTATION FOR STREET MART



The Idea of Behavioral Customer Segmentation

The Idea of Behavioral Customer Segmentation

In a Country like India with properties of diverse Cultures, differing Socio Economics Conditions, different psychological and cultural needs, the consumers can be segmented on various scales which ultimately correlates to the Choice i.e. the type of demand of the customers and also the the amount of products they consume.

The factors are highly influenced directly or indirectly by various quantifiable parameters of the levels of geographical, economic, education, growth and population which are statistically calculated and determined by the census data which comprises of Consumer price Index which tells about growth of a particular section of customers, Population and Area which give the mass dependent demand of a segment of consumers, the literacy rate which gives an idea about the choice and type of demand of the consumers and geographic data which classifies the region a consumer lives in, into various levels.

The Idea of Segmentation of Product market.

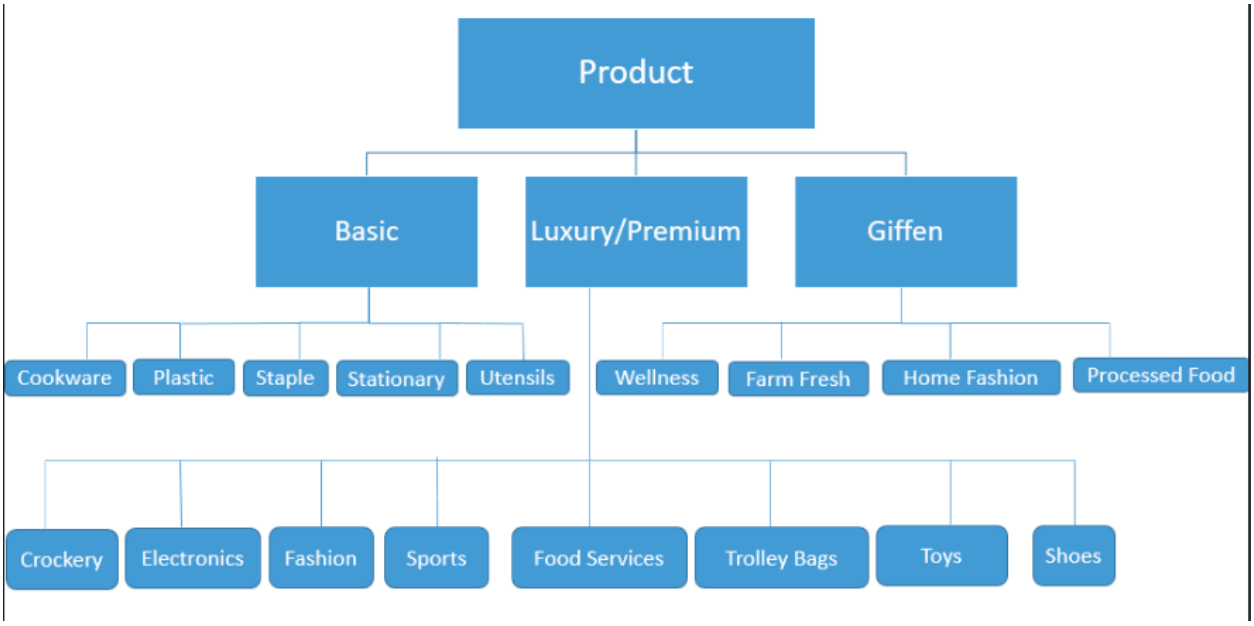
For multi-specialty stores and retail chains like Street mart whose market analysis is to be done the diversity in the type of the available goods or products which the chain sells, the consumer market can also be segmented on the basis of the Luxury and essentiality of a good.

The economically lower stratum of the consumer market majorly consumes only the basic amenities. Then there is a mediocre range of products which are majorly affected by price variations, are highly influenced by choice and real time economic condition of the buyers. Then coming to the luxury products which are bought by the economically higher stratum of the consumers, the demand of these products is not generally price driven.

Basic Products:
Cookware, Home essentials, Staples, Stationery, plastics and Utensils

Mediocre:
Crockery, Electronics, Fashion, Shoes, Trolley bags, sports.

Luxury:
Wellness, Farm fresh, Home fashion, Health care and Processed food.



Data Observation, Correlation and Missing Values

Describing The Data:-

We are given the sales data of different SKUs (Stock Keeping units) of a Retail Chain and econometric data of 23 cities where different stores of this Chain is Located.
Economic Data :- Population, Area, CPI (Consumer Price Index), Literacy Rate and City Type (Town or City).

SKUs :- Wellness, Cookware, Crockery, Electronics, Farm, Fresh, Fashion, Food, Services Shoes, Home Essentials, Home Fashion, Healthcare, Trolley Bags, Plastics, Processed Food, Sports, Staples, Stationery, Toys, Utensils.

Challenges

Solving Missing Data Problem:-

1. Curse of Bhagalpur :- Sales data of almost 50% SKUs was missing for Bhagalpur. We have described the method for handling that case in a section below.

2. After looking the given data thoroughly, we noticed that the SKU data for some cities were missing and some data seemed to be vague. All the data were reconfirmed and then corrected with the help of census data of 2011, e.g. the area of “Jorhat” in the provided data was 9 km2 which was later found out to be 924.6 km2 according to the census data of the year 2011.

Data Preprocessing

We were provided with the sales of different SKU’s in different cities. We observed a large number of correlated data, for example, the sales of all SKU’s were highly correlated. So the data used for clustering became highly redundant and it may suppress other features. It was better to use the data after removing redundant features. We further noticed that some cities had a large number of SKU data missing and some with vague city area. All the data was confirmed and vague values were corrected by matching it with the census data of 2011, for example, the city area of ‘Jorhat’ was given as 9 km2, which was later found out to be 924.6 km2 from the census data of 2011. We normalized all the data by subtracting it with its mean and dividing it with standard deviation so that the mean of final data was zero and variance was 1.

we noticed that some features like population has high variance so we decided to put them in buckets so that the minute changes in that feature will not affect clustering.

The Problem with Bhagalpur

It was observed that in case of bhagalpur some SKU’s drastically low values which was not possible and hence we inferred those to be garbage values.

To assign values to those garbage values we used the data of economically same cities which were identified by initial clustering analysis based on economic parameters. And then by linear regression on those parameters within the cluster, the missing values were found.

Scale the Data

To ensure equal weightage of parameters in the overall result, we normalized all the data by subtracting its mean and then dividing it by the standard deviation so that the mean of the final data comes out be zero and variance becomes one.
We've,

$$X=(x- \text{mean})/ \text{std.dev}$$

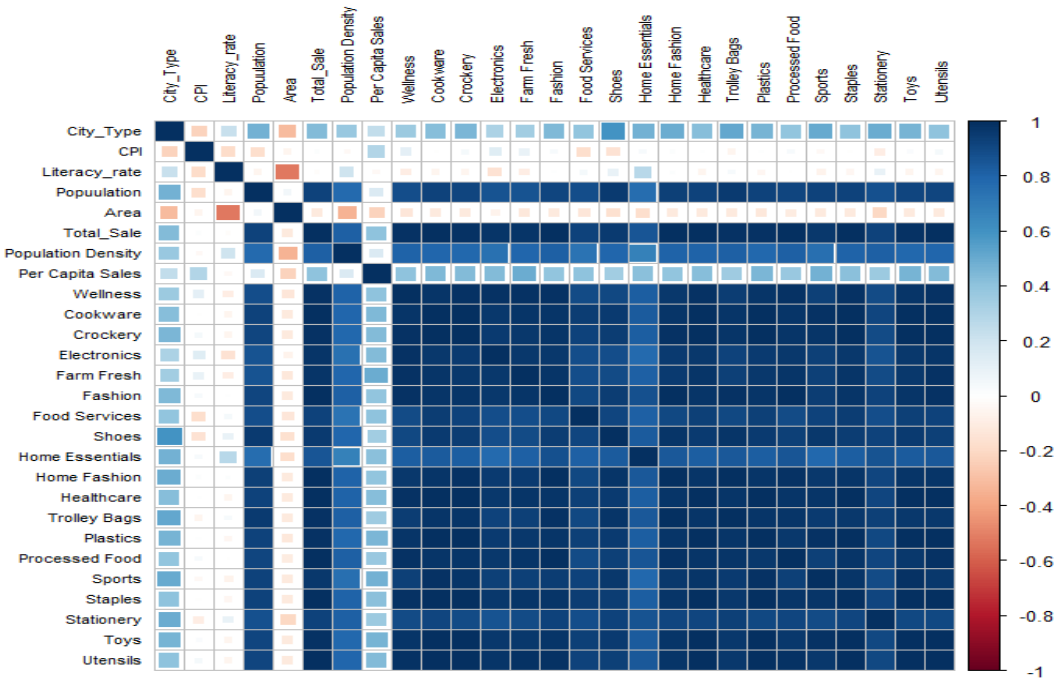
Our data were normalized by scaling between 0 and 1, such that all the features will be given equal weights.

$$X=(x-\min)/(\max-\min)$$

We found out that the data with zero mean and 1 variance was performing better. The reason for this may be stated as the features which are varying largely or have large variance will get lower weight for clustering which is obvious as they may cause interference in the clusters. Now, we noticed that some parameters like population had high variance so we segregated them by putting in buckets in order to ensure that the minute features do not affect clustering.

Scale the Data
To ensure equal weightage of parameters in the overall result, we normalized all the data by subtracting its mean and then dividing it by the standard deviation so that the mean of the final data comes out be zero and variance becomes one.
We've,
X=(x- mean)/ std.dev
Our data were normalized by scaling between 0 and 1, such that all the features will be given equal weights.
X=(x-min)/(max-min)
We found out that the data with zero mean and 1 variance was performing better. The reason for this may be stated as the features which are varying largely or have large variance will get lower weight for clustering which is obvious as they may cause interference in the clusters. Now, we noticed that some parameters like population had high variance so we segregated them by putting in buckets in order to ensure that the minute features do not affect clustering.

Correlation in Data:-
we observed that the SKU's are deeply correlated between themselves so we considered the total sale as a single parameter since taking all the SKU's data could result in redundancy of clustering and also it could suppress the effect resulting from other features



Cluster Analysis

Confirm data is metric

Clustering can only be performed on metric data. many of the statistical methods available for clustering require that the data are metric: this means not only that all data are numbers, but also that the numbers have an actual numerical meaning, that is, 1 is less than 2, which is less than 3 etc.

Hence, we put Town = 0 and City=1.

Selecting the Variables :-

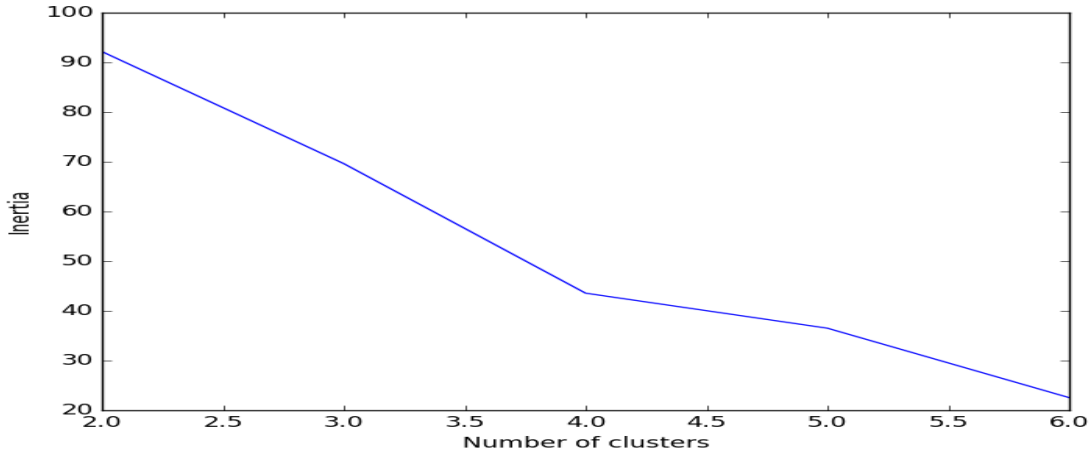
We chose City Type, CPI, Literacy Rate, Population Density and Per Capita Sales as the features used for clustering.

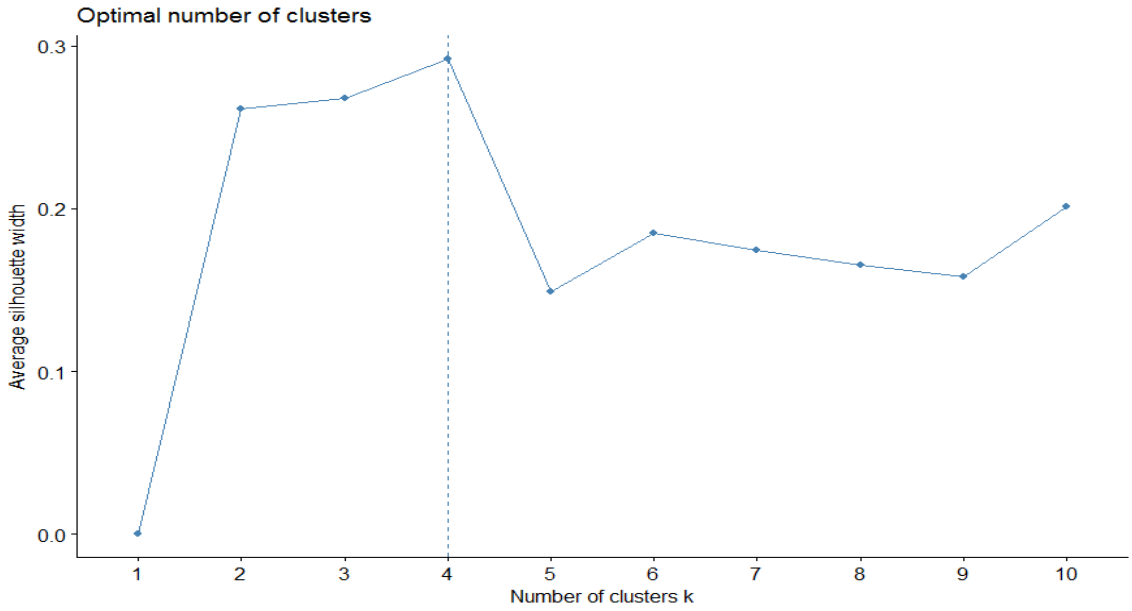
Important Parameters for Each Cluster

Finding the optimal No. of Clusters :-

It is a somewhat bigger challenge to find the optimal number of the cluster as we don't have any specific method in order to get it but we need to generate and analyze several plots in order to guess the optimal number of clusters. In our case, we generated inertia plot, silhouette plot and dendrogram and finally decided the optimum number of cluster to be four after rigorous analysis considering different parameters.

After the repeated Experiments we decided 4 as Cluster Numbers.

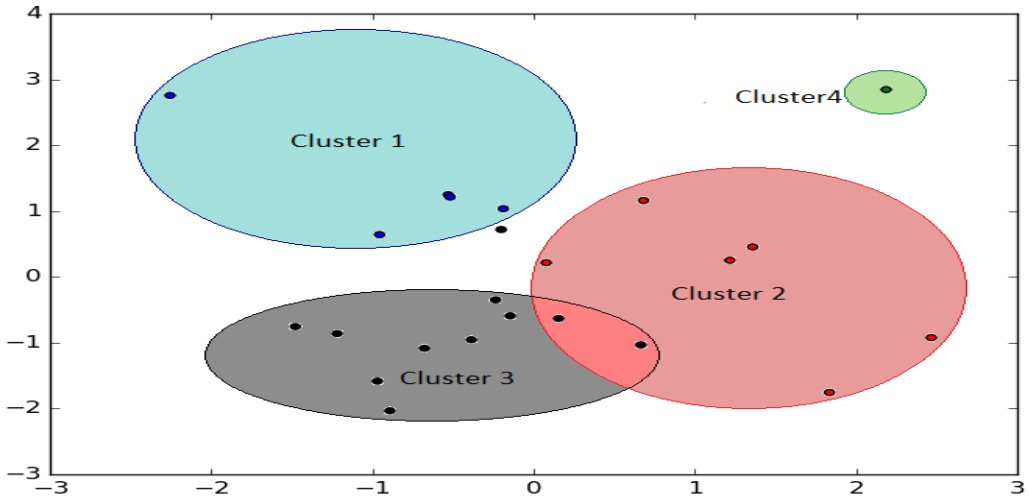




Final Clusters :-

Selecting the Variables :-

We chose City Type, CPI, Literacy Rate, Population Density and Per Capita Sales as the features used for clustering.



Important Parameters for Each Cluster

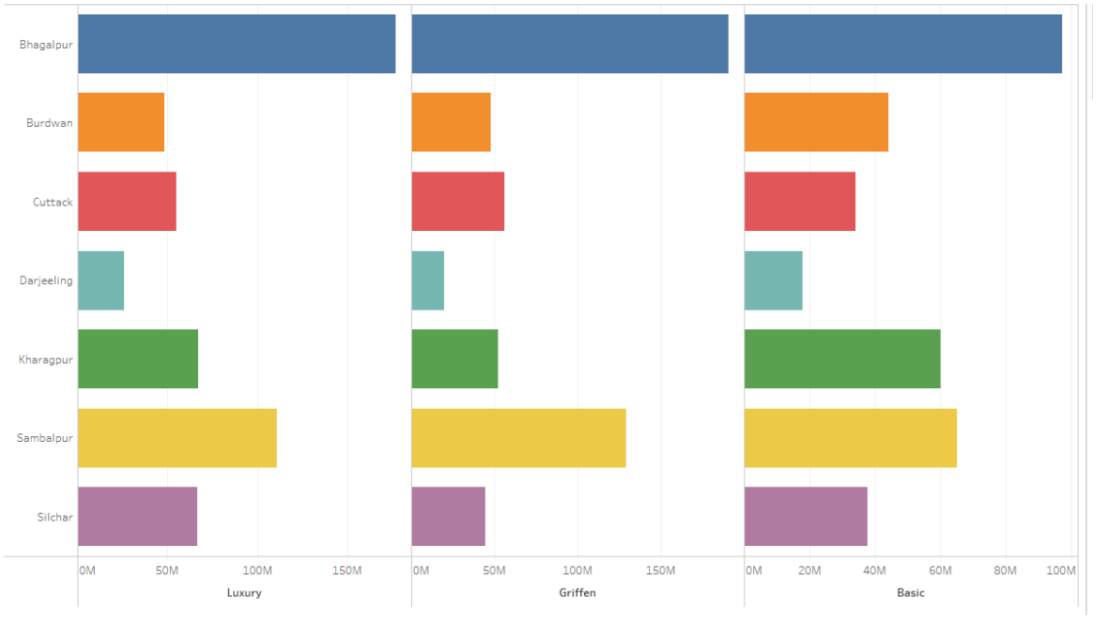
We used Neural Network based method for finding out the most important feature for each clusters.

For this we trained a neural network on the already existing clusters using all parameters available. We listed down the weights of each parameter and inferred which parameter are affecting the classification most.

Our Retail Chain should focus on following SKUs in the cities of each cluster.
Cluster 0 :- Cookware, Farm Fresh, Food Services, Toys
Cluster 1 :- Shoes, Home Fashion, Plastics, Sports
Cluster 2:- Wellness, Cookware, Fashion

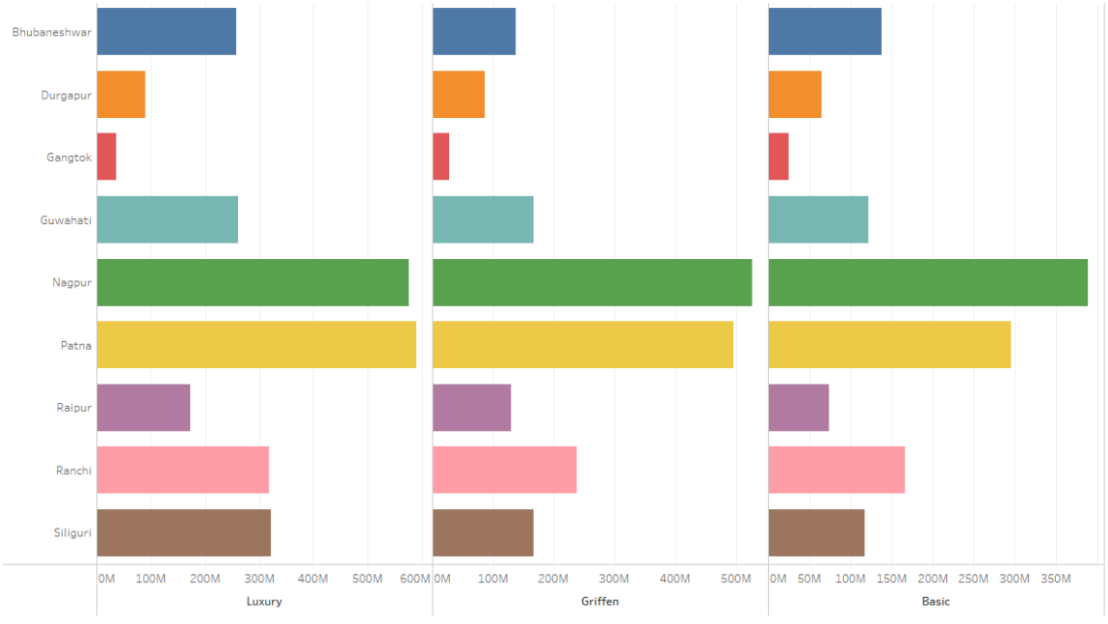
Business/Marketing Strategies For each Cluster

Tier 2 Cities



As we see that the store is a success in bhagalpur.
We suggest to use similar marketing strategies as bhagalpur in other cities of this group.
Kharagpur and Burdwan can focus on the sales of luxury goods as there is scope of improvement by decreasing their price which will call for more demand and hence more sales of luxury.
The basic sales and griffen sales can only be improved mainly by marketing strategies used as in Bhagalpur.

Tier 1 Cities



Here we see that Nagpur and Patna behave exceptionally well. The other cities should maily invest on luxury goods and implementing the strategies.

Tier 3 Cities

Asansol has shown good sales in this group. The marketing strategies and prices for luxury goods must be inspired from Jorhat.

Effect of Economic parameters

There are numerous factors which affect the retail sales of a region. The retail sales are highly influenced by price of goods which changes with inflation, variation of production costs, interest rates, changes of preferences of customers, entry of competitors in the market etc.

This is a general assumption that the income and buying power of people decreases with decreases with decreasing level(strata) of cities.

When there are changes in the economic factors like the inflation in a particular region and overall boom in the economy, the sales generally decrease prominently for the premium or

mediocre goods and only slightly for the necessities or the basic goods. The highly luxurious or griffin goods' sale remains the same.

The general effect of Consumer Price Index

The consumer price index or cpi of a region determines determines the amount of inflation a region is facing. It has effects on retail sales in the following ways.

1. Value of money: The value of money decreases due to which the consumers have to spend more for the same amount of products to be purchased, so the consumers in generally focus on spending only on necessary basic amenities and not on the premium goods. So the sale of the premium goods decreases and overall sales decreases even for the basic goods.

1. Increase in Interest rates: The rate of interest increases with increasing CPI as the inflation increases, due to this the manufacturing and production will increase and also the consumer spending on products will tend to decrease.

Effect of Consumer price index on various regions:

Tier 1 Cities: With the increase in inflation the spending certainly changes for premium goods, but it does not affect much the Tier 1 cities as these have generally high income so the sales do not decrease significantly for these cities for any type of goods, may i be giffin, luxury or basic goods.

Tier 2 Cities: There would be significant decrease in the sales of luxury goods in this type of cities as the people will not be able to spend on their wants and rather focus on spending on needs and necessities. Overall sales will also decrease for these cities.

Tier 3 Cities: The sales of luxury products are already less in case of these cities. The main spending is on basic necessities, people will start compromising and some amount of decrease will be observed in basic products.

The Effect of Gross State Domestic Product.

The Increase in GDP of a state can majorly seen as a deduction of increase in overall retail sales of a state but in long term can also be seen to affect the retail sales. As the increase in

GDP can be seen as an overall boom in the economy hence the buying power of consumers increases and the consumers move from satisfying needs to satisfying wants.

The effect of GDP to specific clusters

Tier1: the Buying power of consumers increases which in this case show a increase in sales of luxury or premium products but the increase will not be as significant as in case of tier2 cities. The sales of the basic necessities and Griffin goods will remain the same. These consumers will mainly move to investment rather than buying more products

Tier 2: Here a significant rise will be seen in the case of luxury goods as the consumers will move to buying luxury goods with the extra purchasing power which comes with the boom in economy, the sales of basic commodities remains same generally.

Tier 3: There will be some increase both in case of luxury and basic goods which comes with needs turning into wants of luxuries and also less compromise in case of basic needs which is in correlation with the buying power of customers.