# **Data Analysis and Visualization**

## Business Intelligence and Big Data Analytics

## Using Excel

### Submitted to

### Professor Sam Adhikari

### Report Prepared By:

### Mohit Joshi

### Rohini Bhandare

### Nitin Mali

**Introduction**

This report summarizes the data and analysis results associated with the transaction details of the super market stores in Mexico, Canada and USA, indicates the key points where the revenue could be boosted based upon the predictive big data analytics.

There are attributes like customer id, purchase date, gender, marital status, home owner, children, annual income, city, state, country, product family, product department, product category, number of units sold and revenue.

What predictive analytics does? It is the process of analysing historical data, understanding the present data, and using the connection between the historical data and the present data to infer what the future may be. So here in the given data set we will be analysing the past events to look for the insights on how to approach the future. We can analyse the trends or likelihood of a situation occurring. The techniques considered here are:

* Predicting market trends.
* Customer needs.
* Creating customized offers for each segment channel and to predict changes in the demand and supply across the entire supply chain.

**Data Analysis**

Our agenda is to analyse the supermarket transaction data of 3 different countries and provide substantial solutions to increase the revenue using big data predictive analytics and different marketing strategies. Below shown figure shows the geographical region of the different locations of the supermarkets, the brighter pink indicates more customer walk-ins and the black bars indicate the sum of revenue generated by the particular supermarket in different cities. So by this reference we can surely predict that we need to focus on the revenue generated by Canada and Mexico which is far less than The United States of America and study different aspects which affect the revenue and provide better solutions to boost it up.

Fig 1: World map locating Revenue sum and Number of Customers



Data set comprises of the purchase data of three major countries like Mexico, Canada and USA and the revenue generated from these countries are $11067.03, $48474.88 and $123288.52 respectively.focus is on various factors which directly affect the revenue to a certain extent and can be helpful in elevating the profits.

Firstly, affecting factor of the revenue can be depicted as the product family and product category. Pivot table, histograms and scatterplots we can observe that food is the contributing factor of the revenue in USA and Mexico as shown in the below figure.

Figure 2

* So focus is more on increasing the product sales in Canada and Mexico where the sales are moderately participating in the supermarket revenue.
* Moreover, we can predict the trend using linear or polynomial regression which is an approach for predicting a quantitative response Y on the basis of single predictor variable X. Here if we consider the annual income of the customer as the predictor variable on the revenue then we can easily predict future sales on the basis of the annual income.
* We can plot the points from the pivot table using a scatter plot and illustrate the equation of the regression and R square value. Below scatter plot illustrates the 3rd degree polynomial equation with a curved trend line and the R square value as 0.61 which represents how close the data is to the fitted line.

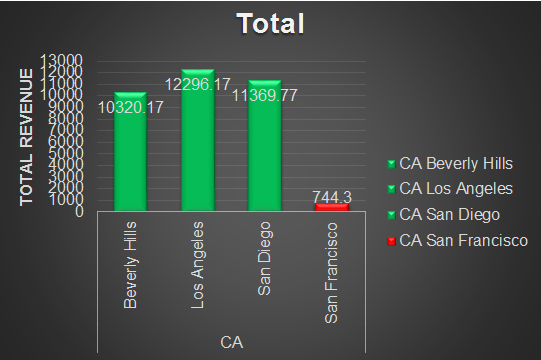
Figure 3

* So by segregating customers into privileged groups and giving add on points on their every purchase can attract them to buy more goods thus may cause our revenue to increase to some extent. Moreover, giving discount coupons to the customers having annual income in the range of $50k - $70k and $30k - $50k can have a positive effect on the revenue.

Secondly, we observed a trend in customer walk-ins in the supermarket. There are 4 cities in CA in USA namely Beverly Hills, Los Angeles, San Diego and San Francisco each generating more than $10k revenue except San Francisco.

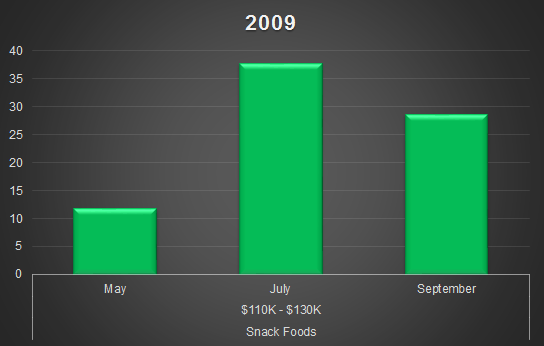
* Considering all the external and demographic factors same as that of the other cities, the store observed low customer walk-ins as compared to the other stores. There were only 13 customers who visited the store again due to this there was an abrupt decline in the revenue. This could be due to poor customer service or maybe poor ground staff management. To overcome these defects, we can give our customers some personalized shopping experience and increase our marketing capital to attract more customers to get attracting offers.

Figure 4



* Analysing the TPS data for Canada we found that the sales in Snacks Food for the year 2009 in the month of May, June and September was quite high compared to the other months of the year, this lead us to the fact that when the customers with the annual income range $110k - $130k have visited the store the sales were high compared to the months where they didn’t visit the store. So, to hike up the revenue to a certain extent we could include those customers to a privileged group and should be given certain discount coupons or promo code to attract them visit the store more often.

Figure 5



Lastly, analysing Mexico’s Yearly Transaction

Figure 6

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of Customer ID** | **Column Labels** |  |  |
| **Row Labels** | **2008** | **2009** | **Grand Total** |
| Acapulco | 3 | 380 | 383 |
| Camacho | 7 | 445 | 452 |
| Guadalajara |  | 75 | 75 |
| Hidalgo | 5 | 840 | 845 |
| Merida | 8 | 646 | 654 |
| Mexico City |  | 194 | 194 |
| Orizaba | 10 | 454 | 464 |
| San Andres | 4 | 617 | 621 |
| **Grand Total** | **37** | **3651** | **3688** |

The basic findings from the table:

* In 2007, none of the shops were opened.
* In 2008, all 6 shops were opened in the month of December.
* In 2009, Hidalgo has the highest number of customers.
* In 2009, Guadalajara and Mexico City has got least number of customers.

Hidalgo City with male and female ratio is 34.6% and 65.35% and for second highest city Merida the male and female ratio is 44.27% and 55.72%. Hence for these two cities try and concentrate more on female products by adding more customized female products of both high and low quality so that revenue could be increased and for the male customers we could try and provide more coupons to them in which we can attract male customers.

Alternatively, in the below graph the number of customers indicates the blue line so the vegetables and snack foods is the most sold ones hence clubbing the products which aren’t similar with these famous ones might increase the revenue for those products which aren’t sold.

Figure 7

Figure 8 and Figure 9

The above graph gives us the insights of two cities Guadalajara and Mexico City which are opened in the month of January 2009. Guadalajara shop is being shut down in the month of December due to low customers. Hence the number of customers for Mexico City is also low. Hence more customers can be attracted by giving a three day Sales with some amount of discount leading to inflow of huge number of customers which in turn results in revenue increase and shop not being closed forever.