Marketing plan for a Small Engineering Company,
through analysis on past sales and advertising data
Miriam Baumann, 2015 -11-08

Abstract

A variety of different statistical analyses on past sales data allows companies to develop an effective marketing plan, in an attempt to increase their sales in the future. This report analyses past sales data of a small company that does not yet have an effective marketing plan based on money spent on online ads. Data is organized, plotted and analyzed in the python programming language and is later converted to python scripts. The results show that there is no relationship between cost of ads and sales to new customers. The distribution of sales to new customers is constant across all years, while the distribution of sales to returning customers increases in numbers and amount in the last 2 business years.

Introduction

In the world of marketing there are many different approaches that can be taken to increase or maintain sales. Every company, whether they provide services or produce products/goods, is able to make sales forecasts, which is often a very important aspect of a company's success (Armstrong at al. , 1987). A sales forecast is the amount of products or services a company expects to sell during a specific period of time at a specified level of marketing activities (Pride & Ferrell, 2012). When making sales forecasts for marketing decisions, there are many factors to consider. Some factors to consider are; what situation exists (for example the state of the economy), what forecast horizon is appropriate (short, medium or long range), the value of the forecast to the company, what data are relevant and available, and what process should be used (such as how the data will be collected and analyzed) (Armstrong et al., 1987). Different companies working in diverse industries and fields, use different forecasting methods. Each company is unique and therefore different forecasting methods will work better for every company (Green, 2001).

The name of the company which will be analyzed in this project is Hanbay Incorporated. Hanbay is a very small company which does not yet have an effective sales forecasting method. They develop electric valve actuators which are then sold to various different industries. Valve actuators are industrial products, meaning they are sold to other businesses for consumption. Thus, sales do not rely on individual consumer behaviour, but rather on the state of the economy and the condition of particular industries.

Hanbay spends a certain amount of money on AdWords each month. Adwords is an advertising service provided by Google, where businesses can pay to display their ads on Google. Businesses only pay when potential customers click on the ads, which allows businesses to set an advertising budget. The purpose of this project is to determine whether previous marketing strategies at Hanbay were effective , and from this evaluating what type of marketing strategies should be carried out in the future. This will be done by analysing the relationship between sales over time based on money spent on AdWords, as well as by analyzing the

distribution of sales among new and returning customers over time . My hypothesis is that the more money is spent on Adwords, the higher the number of sales will be to new customers.

Methods

Choosing a dataset:
  The datasets used in this project are ones that I collected myself over the summer of 2015, while working for Hanbay. The first dataset includes customer sales data that was collected from the company's accounting program. This data includes Name of Customer, Date of Order, & Sales Amount. More columns were added to the dataset upon extensive research on the type of field customers worked in, whether the customers were new or returning, and the exchange rate on the given dates. The second dataset represents how much money was spent on Google AdWords per month, which was collected via the Google AdWords account of the company.

Working with the dataframes:
  To work with the data, a notebook in the python programming language was used. All code is stored in a repository on GitHub and can be accessed via the following link: https://github.com/msbaumann/Final_Project_on_Marketing.git   (The notebook with the code can be found in the src/ directory).
  This project uses two data frames in an excel file, in separate sheets. First, all required libraries, which include pandas, numpy, matplotlib.pyplot, and statsmodels.api were imported into the notebook. Both dataframes were loaded into python using pandas. Pandas is appropriate for these datasets because the data includes both numbers and strings. The data had some missing values in the 'exchange rate' column which were replaced by the number 1, because on these dates the value of the American and Canadian dollar was the same. In addition, the columns were renamed to include no whitespace, which makes it easier to work with in python. While most orders were in American dollars, some were in Canadian dollars. For this reason, an additional column of Sales in Canadian dollars was added (by multiplying the exchange rate column by the Total Sales column) to one of the sales dataframe to make all dollar values synonymous.

Plot 1
Merging the data frames:
  Before merging the data for plot 1, the Sales dataframe was down sampled to include only Sales from new customers, and dates after the 2012/ 2013 business year. The reason for this was because the Ads data frame only had data starting in the 2013/2014 business year. Once this was done, the two data frames were combined by their date column, to create one table that could easily be compared and plotted. The new table had four columns including Date of Order, New or Returning Customers, Ad Cost, and Sales in CAD.

Plotting the data:
  Two scatter plot were created from the merged data frame using matplotlib.pyplot. Both plots plotted the cost of ads on the x-axis and the y- axis differed, with one plotting Sales in

Canadian dollars to new customers and the other plotting Number of New Customers. Every data point is based on data collected in every month from August 2013 to July 2015.

Regression analysis:

      The data in plot 1 was analyzed using linear regression. In this method of analysis, the forecaster searches for a "relationship between past sales and one or more independent variables" which, in this case will be the amount of money spent on AdWords (Pride & Ferrell, 2012, p. 139). Plot 1 shows the relationship between Ad cost and Sales to new customers, where cost of ads is the dependant variable and the Sales to new customers is the independent variable.

Plot 2

Organizing the data frame:

      For plot 2 the Sales data frame had to be re - organized. Columns needed to plot this data were New or Returning Customer, Date of Order, and Sales in CAD. The groupby function was used to seperate New customers from returning customers, creating two new data frames.

Plotting and Analysing data:

      The data was plotted in two scatterplots (subplots), where the x- axis represents Date and y- axis represents Sales Amount. Each datapoint represents an individual sales order; one plot containing only sales orders from returning customers and the other only from new customers. Plot 2 was analysed by a visual comparison of the distribution and amount of sales between the two plots.

      Ultimately, all of the code in the python notebook was developed into general functions and exported to two python script: one to make plot 1, the other to make plot 2. The scripts are callable from the Bash shell and require several input arguments.
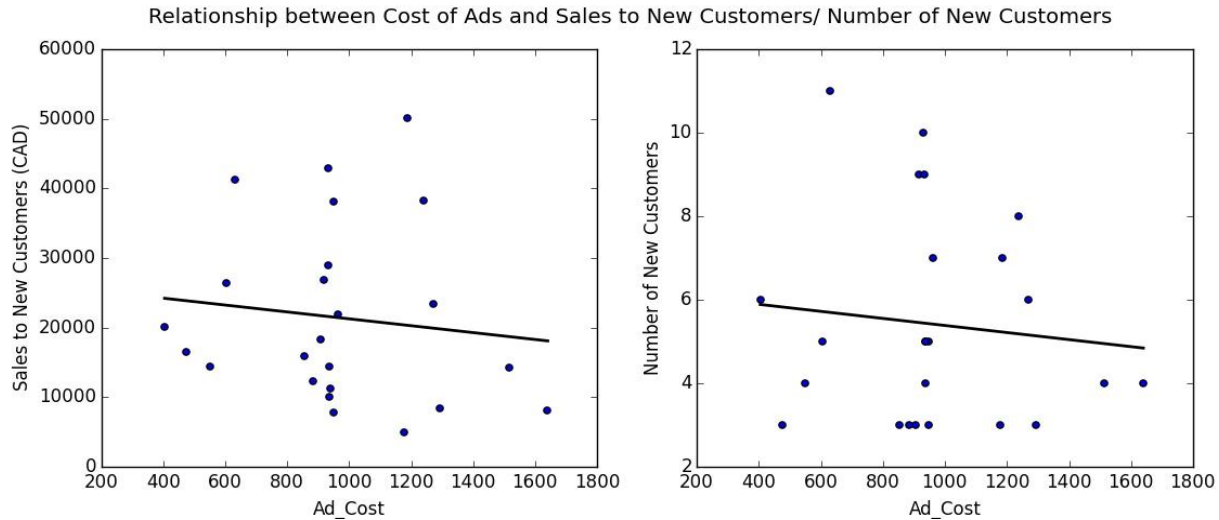
Results

Figure 1: The plot on the left shows the relationship between cost of ads and Sales to new customers, while the plot on the left shows the relationship between cost of ads and number of new customers. The linear regression line is plotted in both plots.

The R-squared values obtained from the linear regression lines in Figure 1, are 0.014 for the plot on the left, and 0.011 for the plot on the right. The slope of both regression lines is slightly negative, and almost identical in scope. When looking at both plots, the data points are extremely scattered and many are concentrated in the mid range of the x- axis at around $900. The scatter of points are fairly similar in both plots, however when looking more closely there are some differences especially in the lower numbers of the x-axis (around $500).
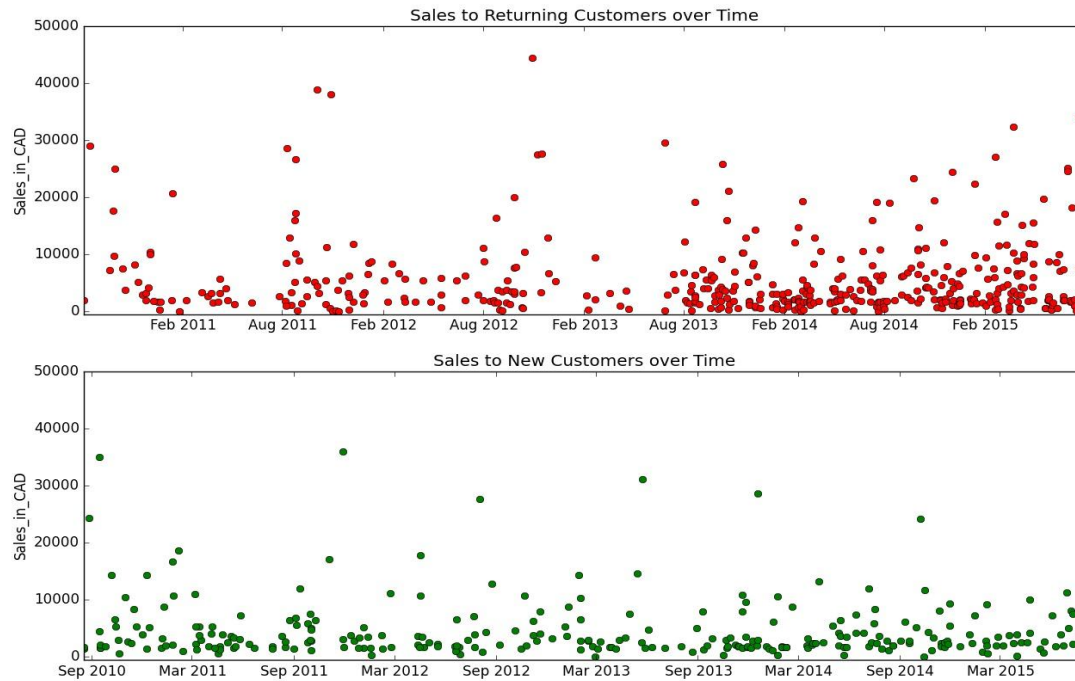
Figure 2: These two plots show the distribution and number of individual sales over 5 years (2010 - 2015). The top plot shows only sales to returning customers, while the bottom plot shows only sales to

The distribution and number of sales to returning and new customers seems to be similar in the first 3 business years, from August 2010 to July 2013. In August 2013, number of sales to returning customers noticeably increases (top plot), while sales to new customers stays the same throughout the full time period (bottom plot). In addition sales to returning customers seem to increase in amount, meaning returning customers are placing larger individual orders than they have in the past. In summary, distribution of sales over time is constant for new customers and varies for returning customers.

Discussion

The reason why both Sales to new customers and Number of new customers were plotted against the cost of ads in two subplots (Figure 1), is to recognize whether sales amounts are an accurate representation of the number of new customers. Some new customers may have placed orders for large amounts, which would somewhat skew the results. The amount of money spent on Adwords would no longer reflect how many new customers where obtained.

However our results show that all orders were around the same amount, which is why the linear regression lines are very similar in both plots. Both plots have an R-squared value of under 0.02. The R-squared value is a fraction between 0.0 and 1.0. The closer the R-squared value is to 1.0, the higher the linear relationship between the two variables. An R-squared value of 0.0 means that there is no relationship between the two variables. Therefore knowing the x-value does not help you predict the y-value. Since the R-squared values from the linear regression lines in both plots are very close to 0.0, we can conclude that there is no linear

relationship between the two variables (sales to new customers/ number of new customers and cost of ads). When the scatter in a plot is large, the R-squared value will be lower.

The reason why the plots do not show any relationship between sales to new customers and cost of ads, may be due to several factors. To begin with, the differences in money spent on ads per month are not very big, ranging only from 400CAD to 1650CAD. Sales numbers on the other hand are much higher and range from 5000CAD to 50000CAD per month. The relatively small amount of money spent on ads compared to sales amounts, may have no effect on attaining new customers. Another reason could be that AdWords is not an effective way of spending advertising dollars for this company. Perhaps Hanbay should find a different, more effective method of advertising their products that will show an actual increase in sales to new customers.

Figure 2 reinforces the results obtained from the linear regression analysis of Figure 1, by showing that the distribution and number of sales to new customers is constant over the past 5 business years. Any increases in yearly revenue in the years 2013/14 and 2014/15 for the company are almost entirely due to an increased number of sales to returning customers. Customers coming back and buying more products from Hanbay, are most probably satisfied with the product they bought the first time. However there may additional reasons why several customers are coming back and buying for larger amounts. One apparent reason could be that the customer wanted to test one product first before purchasing a larger amount. A more interesting reason for larger sales orders could be that the field or industry in which particular returning customers work in could be growing and expanding, and therefore in need of larger amounts of products from Hanbay. If this is the case, it may change the way that the company invests marketing dollars. Further analysis of which industries Hanbay has increased sales to, could help the company focus their advertising dollars to specific fields, by for example attending specialized business expositions.

This project shows that the company's current expenditures on AdWords are not increasing sales amounts. This may help the company realize that there are other, more effective ways of spending money to increase sales.

References

Green, Y.N.G (2001). Chapter 2: Literature Review. *An Exploratory Investigation of the Sales Forecasting Process in the Casual Theme* (pp. 9 - 35). VirginiaTech.
Armstrong J.S. , Brodie R.J. , McIntyre S.H. (1987). Forecasting Methods for Marketing: Review of Empirical Research. *International Journal of Forecasting, 3,* 355 - 376.
Pride W., Ferrell O.C. (2012). Part 2: Marketing Research and Target Markets. *Foundations of Marketing* (pp. 87 - 144). Nelson Education.