The first technique that will be used is 'regression analysis'. In this method, the forecaster searches for a "relationship between past sales and one or more independent variables" which, in this case, will be the industry in which the customer works in (Pride & Ferrell, 2012, p. 139). In order to produce a sales forecast using regression analysis, historical sales data must be available. Sales data of the company from the past 5 years (2010 - 2015) is readily available. This data includes date of invoice, customer name, industry in which the customer works in, invoice amount, and whether the customer is returning or new. Using functions in the Python programming language, this data will be sorted by industry for each year. Data for each industry will be compared over the years to observe whether there is a relationship. If there is a relationship between the two variables, closer analysis will follow and a sales forecast can be made.

A second method that will be applied is 'time series analysis'. This method also uses past sales data in an attempt to discover patterns in the company's sales over time (Pride & Ferrell, 2012). The results of four types of analyses (trend, cycle, seasonal, and random) are combined to develop a sales forecast (Pride & Ferrell, 2012). In other words, past sales data of Hanbay will be used to plot graphs in Python that will be analyzed in an attempt to find annual, seasonal, or random trends. This is an effective forecasting technique for products with reasonably stable demand, which is the case for valve actuators (Pride & Ferrell, 2012).

I predict that customers in industries which have increased in growth and success over the past few years, will show higher sales. On the other hand, customers in industries which have struggled in past years will show a decrease in sales. I believe that time series analysis will show that sales do not demonstrate obvious annual or seasonal trends. However, it will show random trends based on the state of the American economy. These two observations combined, will help the company decide how much money they should spend on marketing in specific industries.It will also help the company make predictions on future sales trends based on how much money and how effectively money is spent on marketing.

Abstract

Write a summary of the whole project with hypothesis and results.

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 create a sales forecast for Hanbay by analysing the relationship between sales over time based on money spent on AdWords. My hypothesis is that the more money is spent on Adwords per month, the higher the sales will be to new customers.

Methods (with link to Github repo):

*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 included 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 whether they paid in Canadian or American dollars. 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.

*Loading and organizing the dataframes:*

This project uses two data frames in an excel file, in separate sheets. The python notebook was used to work with the data.

First, all required libraries, which include pandas, 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. An additional column of Sales in Canadian dollars was added to one of the data frames to make all dollar values synonymous.

*Merging the data frames:*

Before merging the data, the Sales dataframe was downsampled to include only Sales from new customers, and dates after the 2012/ 2013 business year. The reason for this is because the Ads dataframe only has 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 can easily be compared and plotted. The new table had 4 columns including Date of Order, New or Returning Customers, Ad Cost, and Sales in CAD.

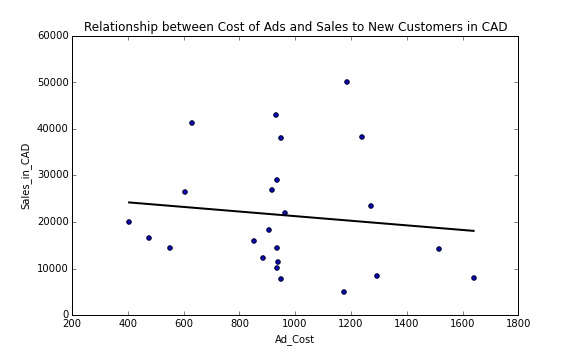
*Plotting the data:*

PLot1: A scatter plot was created using matplotlib.pyplot, where the x-axis is the cost of ads based on month and the y axis is the Sales in Canadian dollars to new customers. This plot 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.

*Statistical Analysis of the data:*

The data was analyzed using linear regression, because....the there is a true independent variable which is the Cost of Ads.

Results



The obtained R-squared value form the linear regression analysis on the two variables is 0.014 which corresponds to 1.4 %, and the adjusted R- squared value is -0.031. Both R- squared values obtained are very low. The slope of the line is negative and the y- intercept is at 26, 210. Many of the datapoints are concentrated around the 900

Discussion

The reason why both Sales to new customers and Number of new customers were plotted against the cost of ads in two subplots, 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.

"Experiments that observe a variable changing over time will produce time series data, where the independent variable is time and the dependent variable is the observation. " p.149 effective computation in physics.

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.