Marketing Data Analysis Proposal



# Task 1

***What information would you need/like to have to be able to investigate whether Fayes’ claims (both the advertising program is successful and factory production is ticking along) are correct?***

* Sales data
* Production data
* Marketing data, including costs, type of marketing campaign, campaign location
* Any new products introduced?
* Any new vendors?
* Logistics data, to account for any changes in costs.
* Supply Chain data
* Competition Data if available

We would like to have data above from the current and previous 4 quarters to be able to analyze if sales have been increasing before and after the promotion started, also to take into account any seasonal dependencies (i.e. compare the data to the same period last year). The data should include customers’ data, type of product sold, online vs. in-store

# Task 2

***Write a three-page proposal to management asking for the budget and approval to conduct analytical study to investigate Fayes’ claims as input to the decision whether to grant his request for a larger budget. Assume that Fayes needs his answer in four weeks.***

In order to determine the effects of the marketing promotion and evaluate its positive and negative effectiveness, we suggest to run an analytics study using the data. We need to determine if the sales have been increased and production is up to speed with the increase in sales and at the same time declare scientifically and statistically if the promotion had a direct impact on the sales.

To run a proper analysis of the data I suggest using a prediction model using the previous 2 years of data for sales, production, marketing, logistics, supply chain and other related data mentioned above. We are going to use the historical data for training the model and use the last 4 weeks as a test to see whether the same increase is sustainable or not.

The steps to follow are, first to analyze the data format and its amount to determine what types of software/hardware we are going to need, and then setup the system based on this study. Then we are going to acquire the data, store it on the system and start to learn and cleanse the data. In this stage we need to consult various business experts and domain expertise to learn the format and parameters of the data. In the same phase we may want to anonymize the data to make sure that we are not going to jeopardize customers and vendors privacy. We should also take into account the security of data, media and the channels. After this stage we can start digging deeper into the data and start the analysis process.

The conclusion we want to reach after the analysis is to see whether the marketing campaign had an effect on the revenue. In order to check out hypothesis we are going to break down the data into weeks, and plot the data (Chart 1). Because the data that we have is paired data, meaning that we have the same data for previous months, we are going to calculate the difference of values between consecutive weeks and plot the difference (Chart 2). Assume that production – sales = inventory and there is a minimum required “safety inventory” for the company, calculate the inventory level per week and also plot the inventory level per week (Chart 3). If inventory is consistently decreasing after the promotion started, then that means production has not been able to keep up to speed with sales, even if the inventory level has not fallen below the minimum safety level. For the difference between weeks for sales data, form a null hypothesis that the promotion does not have a direct impact for sales Assuming a normal distribution, compute the mean, standard error, Z score, and the corresponding p value for the Z score. By using the general rule of thumb of p < 0.05 to be significant, determine whether to reject or fail to reject the null hypothesis that the promotion has been increasing sales. Looking at Chart 1 and Chart 2, determine if there is a general trend that sales have been increasing, especially in the weeks after the promotion started. Then we are going to present a 2-part finding, from the general trend of sales in the charts and the null hypothesis testing, to the management and stakeholders.

The model we are going to use is a predictive modeling scheme called tree induction predictive modeling technique. We first must select one or more attributes/variables, which will provide us the most amount of information gain. We will have meeting with all the stakeholders and collect all the business attributes, then we will use J48 decision tree to calculate the information gain with different attribute combination. In sleepy industries, there are four attributes - time, factory production output, advertisement cost, cost of goods and two target variables – gross revenue, net profit for our model. Then we will use different training data to induce three different predicting model – one model with last quarter advertisement budget, one model with 50% more advertisement budget and third model will be no advertisement budget. We will follow up next four weeks revenue trend and see which predicting model is closer to the reality. This can provide some evidences of how necessary to increase advertisement budget.

In order to achieve the proper analytics and to be guarantee the correct results we need to prepare a data science team that includes the following people:

1. Project Manager (PM)
2. Data analyst for data analysis and report writing (DA)
3. Database administrator for maintaining the data (DBA)
4. Business Expert (BE) to coordinate communication between the data analyst and the marketing team and present findings converted from mathematical terminology to business terminology
5. In case software/hardware is required we need IT/Big Data tool expert support

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Activity - Team Role | DA | DBA | BE | PM | SPO |
| Feasibility & Scope Research (1 Day) | X |  | X | X |  |
| System Setup (2 Days) |  | X |  |  | X |
| Acquiring Data & Storage (2 Days) |  | X |  |  | X |
| Understanding & Cleaning Data (7 Days) | X |  |  |  |  |
| Running Analysis (5 Days) | X |  |  |  |  |
| Report Generation (3 Days) | X |  | X |  |  |
| Total Hours (8 Hours per day) | 128 | 32 | 32 | 8 | 32 |

1. Security and Privacy Officer (SPO) to determine the data security and privacy used in the study
2. Sales Manager
3. Production Manager
4. Marketing Manager

Time required to achieve results and to provide the results with a proper report would depend on the amount of data available, the availability of the hardware and software and the quality of the data available but a rough estimate would break down into the following phases, along with the expertise needed for each phase:

We are also going to provide frequent interim reports and updates to the CEO, board members and the project manager to make sure that everybody is aware of the progress and the direction of the study. In order to make the reports and the analysis more effective we recommend that Sales Manager, Production Manager and Marketing Manager also attend the report meeting

The estimated cost to perform the analysis based on $80 per hour for the core team members (DA, DBA, BE, PM and SPO) and the cost of buying data analysis software such as R (free) and Minitab ($2000), excluding the cost of maintaining the database and data warehouse (assume that the company is already paying for it for its daily operations) would be as follows:

Cost of the core team:

128 + 32 + 32 + 8 + 32 = 232 hours \* $80/hour = **$18560**

Total cost:

$18560 of core team + $2000 of software = **$20560**