# **Business Understanding**

## Determine business objectives

Background

In a saturated apparel market, it is imperative for businesses to employ various tactics in order to increase sales and grow their customer base, effectively. The clothing store chain, Cotton On, recognises this as a core element to their success and thus has exhausted an immense amount of time and effort into trialling different marketing strategies to promote sales. Following from this, Cotton On requires analysis of the effectiveness of each marketing strategy for the corresponding demographic group.

The ultimate objective for Cotton On is to increase the profitability of its overall operations. To do so, the chain needs to ensure that only cost-efficient strategies are implemented. Hence, it is imperative to build a cost-benefit table to detail the impact of using direct mail marketing.

|  |  |  |  |
| --- | --- | --- | --- |
| **Outcome** | **Classification** | **Actual response** | **Cost** |
| True negative | Nonresponse | Nonresponse |  |
| True positive | Response | Response |  |
| False negative | Nonresponse | Response |  |
| False positive | Response | Nonresponse |  |

As this project is a relatively new area for Cotton On, an organisational structure and a key individual to oversee the operations is lacking. This presents an opportunity for the clothing store to procure an in-house analyst team that will have the responsibility of being the steering committee. It would further benefit from commissioning a data analyst. Since this project is still in the conceptualisation stage, there is a lack of hierarchy and participants.

Currently, Cotton On has been able to establish large databases for customer and product information as it is a well-known clothing store chain. However, there are no procedures in place to clean and prepare the information collected for analysis. Thereby increasing the unreliableness of the data. Hence, a classification system can facilitate this problem whereby a customer is classified into different categories upon encountering and “purchasing” a product from the store. More specifically, if given the attributes of a customer observed through their purchasing behaviour (*e.g. type of items purchased*), it is possible to the classify their reaction (*e.g. response or nonresponse*) for specific marketing strategies.

As discussed previously, this project is still in its initial stages, therefore there is no evidence of solutions or alternatives to be implemented in order to analyse their customer base. There are obvious benefits to the success of this project such as more effective targeting of different consumer groups dependent on tailored marketing strategies, whilst maintaining, or reducing, costs. This will meet the objective of increasing profitability, and thus will be welcomed.

Defining business objectives

The project is commissioned with the following objectives:

* For all customers, the program is able to classify based on their characteristics what their response to direct mail marketing will be
* Identify whether customers require an additional form of marketing strategy to respond
* Promote sales for Cotton On through tailored marketing strategies
* Ultimately, reduce operational costs and increase profitability

Business success criteria

Tentatively, the project will be judged on its success if:

* Accurately identify customers’ responses to direct mail marketing with an error margin of less than 10%
* Minimise costs associated with Nonresponse customers

## Assessing the situation

Inventory of resources

Despite the presence of an existing in-house Database Management System that takes the form of operational databases, there are no database applications or hardware accessible to support the interaction between the data collected and Cotton On’s marketing department. Moreover, a data warehouse has yet to be built to facilitate efficient and quality analysis, therefore it still requires data extraction and cleansing.

**Personnel**

Evidently, there is internal expertise in gathering and simple processing of this data for simple marketing purposes. However, more sophisticated analysis of these relationships for various marketing and sales aspects is lacking. Employing a data analyst that is also proficient in database management would be cost effective.

**Data**

Availability of large data will facilitate better insights into consumer behaviours, allowing for improvements to marketing strategies to promote sales. Because the project is in its introductory stages, it will be beneficial to limit this study to consumer responses for direct mail marketing. The project can be expanded later.

**Risk**

The monetary outlays for the consultants and marketing expenses are the greatest concerns for this project since the ultimate goal is to increase profitability. Therefore, it is imperative to stay within budget. Aside from this, there is nothing noteworthy.

Requirements, assumptions and constraints

**Requirements**

There are no legal and security requirements in regard to this project’s results since information is collected anonymously and from basic marketing evaluation metrics that do not reveal consumers’ identities. As this is a new project, all stakeholders involved are willing and have approved of the project scheduling requirements. Results shall be implemented into Cotton On’s established databases and the created software to improve the classification of customers for enhanced marketing strategies.

**Assumptions**

*Economic:* The most prominent factor would be that Cotton On operates in a highly saturated retail market, denoting that it is subject to competitors’ ability to attract and retain customers.

*Data quality assumption:* This becomes a threat because this will compromise the quality of the data collected. There is the possibility of systematic error occurring in the data collection stage. One of them being sample-selection error because Cotton On does not operate in a monopoly whereby its customers represent the whole population. But rather it only captures a segment of the market and thus this ‘sample’ might not accurately detail how consumers respond. This is followed by data processing error to which incorrect data entries will manipulate the overall insights. However, it is assumed that these threats have not occurred and data collection has been conducted in a consistent manner that ensures these errors have not occurred. Additionally, data utilised is an accurate representation of the population.

Upon completion, final results must be presented to the key stakeholders in a business report. It needs to include model evaluation based on at least two substantively different models for classifying Cotton On’s customers. Along with including the project range (confidence interval) for the expected gross profit per customer contacted based on each model. The user would also require interpretation of the findings and final remarks on how to proceed with these insights.

**Constraints**

As discussed previously, there are no legal constraints of accessibility issues on the usage of this data since it is generic information that do not reveal consumers’ identities. All funding will be covered by Cotton On.

Risks and contingencies

This outlines the possible risks that may present itself over the course of the project and contingency plans:

|  |  |
| --- | --- |
| **Risk** | **Impact & Plan** |
| **Scheduling** | There are no issues with scheduling as there is no predetermined deadlines and this would be an ongoing project for Cotton On. |
| **Financial** | Again, no immediate issues present itself as the chain would be funding this project (assuming it remains profitable). |
| **Data** | There are existing procedures in place to prevent the likelihood of poor quality data. However, if this was an issue, Cotton On needs to be prepared, in regard to time and the financial costs, to gather more data that is relevant to the project. |
| **Results** | Even if the initial results are less dramatic than expected, this does not compromise the project as it signifies classification of customers is possible and fulfils the project objectives. It now provides scope for improvements on the classification algorithms and further iterations on other marketing strategies to be achieved. |

This discussion reveals there are no obvious risks that require immediate attention, and necessary contingency plans have already been outlined. If Cotton On no longer finds the project feasible, it shall be halted until the steering committee can resample their customer base and re-evaluate.

Terminology

There are no particular terminology required for this project.

Cost/benefit analysis

|  |  |
| --- | --- |
| **Benefits** | **Costs** |
| The classification algorithm shall reduce costs related to the execution of marketing strategies for Nonresponse customers | There are no costs required for data collection since this is already completed and no external databases are utilised. |
| More effective sales growth can be achieved by identifying which customers respond to which marketing strategy. | There will be minimal costs for result deployment as it can simply be installed into the hardware component |
| The advancement of knowledge regarding customers is highly valuable due to greater understanding of the behaviours of different customer groups. | There will be operational costs associated with the creation of the hardware for users to interact with the data |

## Determining data mining goals

Data mining goals

* Type of data mining problem: This is a Classification problem. We need to classify customers into their respective groups based on certain attributes
* Predictive capabilities: Predict and classify customers based on their response outcomes to specific marketing strategies following customer transactions
* Desired outcomes: Generating response predictions for all new customers

Data mining success criteria

The methods utilised for evaluation of these models will include:

* Confusion matrix
* Cost Matrix
* Metrics such as: Precision, Recall, and the F-measure
* Receiver Operating Characteristic Area Under Curve

Benchmarks for evaluating success are:

* 90% accuracy in identifying and classifying Cotton On customers
* Precision and Recall being each respectively at least above 80%
* ROC Area Under Curve being greater than at least 80%

Successful deployment of the model results is crucial to the success of the project.

## Producing a product plan

A product plan is constructed:

|  |  |  |  |
| --- | --- | --- | --- |
| Phase | Time | Resources | Risks |
| Business understanding | 3 days | All analysts | Alignment of goals and objectives |
| Data exploration | 1 week | All analysts | Data problems |
| Modelling & evaluation | 2 weeks | Data mining consultant, business analyst | Data problems. Constructing feasible models |
| Deployment & presentation of results | 1 week | Business analyst | Inability to implement results |

Assessing tools and techniques

This project can be structured as a binary classification problem. Therefore, there are a plethora of potential tools to use:

* KNN
* Logistic Regression
* Naïve Bayes
* Linear Discriminant Analysis
* Quadratic Discriminant Analysis
* Decision Trees
* Random Forests
* Extremely Random Forests
* Adaptive Boosting
* Gradient Boosting
* Ensemble voting

All these tools will be carried out on the training data and then based on their cross-validation scores, we then select a final model.