Descriptive Analytics for M&S*

MSIN0094 Case Study

Wei Miao

October 15, 2025

1 The Power of Descriptive Analytics

The amount of data created worldwide has been increasing exponentially over the past decade with some estimates placing the total at over 60 zettabytes as of 2030 (Source: Statista). Data without analytics, however, is of little value to business decision-makers aiming to improve performance and increase growth. It is therefore no surprise that top-tier consulting companies, analytics firms and business schools have been promoting the positive returns to greater usage of analytics technology. It also explains why an increasing number of data analytics enthusiasts (like Tom and you!) are eager to join the world's best MSc Business Analytics program at the UCL School of Management.

By identifying patterns and trends in massive amounts of data, business analytics enables organizations to make better decisions and improve performance. Descriptive analytics is the simplest and most widely used type of analytics; it is used to generate key performance indicators (KPIs) and metrics for business reports and dashboards. Research shows that, even with the adoption of very simple descriptive analytics, businesses can improve their performance by a large extent — Berman and Israeli (2022) analyze the staggered adoption of a retail analytics dashboard by more than 1,500 e-commerce websites,[1] and find an increase of 4%–10% in average weekly revenues post-adoption. The increase in revenue is not explained by price changes or advertising optimization. Instead, it is consistent with the improvement on customer relationship management, personalization, and prospecting technologies to retailer websites. The adoption and usage of descriptive analytics also increases the diversity of products sold, the number of transactions, the numbers of website visitors and unique customers, and the revenue from repeat customers. These findings are consistent with a complementary effect of descriptive analytics that serve as a monitoring device that helps business analysts control additional martech tools and amplify their value. Without using the descriptive dashboard, companies are unable to reap the benefits associated with these technologies.

In the remaining of this case, we will explore (1) how to conduct data wrangling and data cleaning using R and (2) how to conduct descriptive analytics for M&S to achieve marketing efficiency. You might ask, where is Tom? Well, M&S rewarded Tom with a few days off for his successful calculation of CLV last week, and he is now

*This case was prepared by Wei Miao, UCL School of Management, University College London for MSIN0094 Marketing Analytics module. This case was developed to provide material for class discussion rather than to illustrate either effective or ineffective handling of a business situation. Names and data may have been disguised or fabricated. Please do not circulate without permission. All copyrights reserved.

[1]We will cover the difference-in-differences technique to establish causal inference later in the module.

considering setting up his bubble tea business in Canary Wharf. You are the only one who can help M&S now until Tom's return!

2 Database Marketing at M&S

We learned in Week 2, how to compute the customer lifetime value for M&S. However, when computing the CLV, we used an "average" approach, which did not consider customer heterogeneity. That is, when considering each component in the CLV formula, such as customer spending in each period, their retention rate, etc., we took the average across all customers, and assumed customers are homogeneous. As a result, every customer would have the same CLV. Nevertheless, this is a strong assumption in practice — every customer is unique and should be treated differently.

The key to successful customer relationship management is to maintain a customer database that tracks detailed customer information, including their demographic information and past purchase history. This information would empower marketing analytics team to compute individual CLV for each customer, and conduct individualised targeted marketing.

2.1 Demographic Information

Knowing your consumer is a vital concept of running any business. Is the business selling fertiliser to farmers, apparel to teenage girls, or vacations to senior citizens? The distinctions are readily apparent in this comparison.

Demographics define the qualities of clients. To be successful, business owners must understand the demographics of their clients and the trends or changes that are occurring within those specific characteristics.

The following demographic information is usually of interest to business managers:

Age: Consumer behaviour is strongly influenced by age. Younger consumers are more affluent and willing to spend more on entertainment, fashion, and movies. Seniors spend less on these items; they are less active, spend more time indoors, and require more medical treatment. Additionally, market segments can be defined by age groups. For instance, digital devices such as iPhones are targeted more towards millennials than at seniors. While older adults are increasingly utilising technology, they remain less digitally savvy than millennials and purchase fewer digital products.

Gender: Gender identity is another key demographic factor. Consumer needs and preferences can differ based on gender, influencing purchasing decisions. Consequently, many products are marketed with specific gender identities in mind. For example, apparel retailers often organise sections by gendered clothing styles, and some brands, from personal care to accessories, may target their marketing campaigns toward specific gender groups.

Income: Income has a substantial influence on consumer behaviour and product purchases. Middle-income customers make purchases with due regard for the utility of money. They do not have unlimited money to spend, and hence the money spent on one item may be used on something else. On the other hand, consumers with higher incomes tend to be less price sensitive and have a higher willingness to pay.

Education: Consumers' level of education has an effect on their impressions of the world around them and on the amount of research they conduct prior to making a purchase. Individuals with a higher level of education will spend more time educating themselves before investing their money. Education has an impact on fashion, film, and television programming. Consumers with a higher level of education can be more distrustful of adverts and the facts offered.

M&S has collected rich customer demographic information through its loyalty program, M&S Sparks membership. The data scientist team has the following demographic variables:

- ID: Customer's unique identifier
- Year_Birth: Customer's birth year
- Education: Customer's education level
- Marital_Status: Customer's marital status
- Income: Customer's yearly household income
- Kidhome: Number of children in customer's household
- Teenhome: Number of teenagers in customer's household
- Dt_Customer: Date of customer's enrollment with the company

2.2 Purchase History

"History doesn't repeat itself, but it often rhymes."

This popular aphorism, frequently (and perhaps incorrectly) attributed to Mark Twain, is frequently invoked to demonstrate that, while past events do not always provide a clear indication of future events, they do provide valuable context. This sentiment is especially true for marketing managers, where a consumer's purchase history provides invaluable insight into their future purchasing habits.

M&S's data engineering team has assembled a **cross-sectional** customer purchase history data today, with variables including[2]:

- MntWines: Amount spent on wine in last 2 years
- MntFruits: Amount spent on fruits in last 2 years
- MntMeatProducts: Amount spent on meat in last 2 years
- MntFishProducts: Amount spent on fish in last 2 years
- MntSweetProducts: Amount spent on sweets in last 2 years
- MntGoldProds: Amount spent on gold and jewellery in last 2 years
- NumDealsPurchases: Number of purchases made with a discount
- NumWebPurchases: Number of purchases made through Ocado's web site
- NumCatalogPurchases: Number of purchases made using a catalogue
- NumStorePurchases: Number of purchases made directly in stores
- NumWebVisitsMonth: Number of visits to company's web site in the last month
- Complain: 1 if customer complained in the last 2 years, 0 otherwise
- Response: 1 if customer accepted the offer in the last campaign, 0 otherwise

[2]Cross-sectional data is data collected at a single point in time. It is a snapshot of customers at a particular moment. In contrast, **longitudinal** data is collected over time, allowing analysts to track changes in customer behaviour over time.

- Recency: Number of days since customer's last purchase

3 Data Wrangling

3.1 Data Loading

To work on the datasets, we first need to load the raw data into R. In R, we can use `read.csv(filepath)` to load the data into R environment.

For your convenience, I have stored the data file on my Dropbox and the download link will be provided. We can directly feed the url links to `read.csv()` function in R to download and create the dataset in R's environment.

The data is also provided to you in .csv format, which you can also load from your local disk.

> **Question 1**
> Load the dataset into your R environment. Name the dataset as `data_full`.

After running the above code blocks, you should see the dataset named `data_full` in your RStudio environment.

Now, click into the dataset, take a look, and get a sense of how it looks like.

3.2 Data Types

> **Question 2**
> Check all data types in `data_full` are correct and as expected

3.3 Data Cleaning

Data cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate records from a dataset. We often need to deal with missing values and outliers. In this case, we will focus on missing values.

> **Question 3**
> Check if any variables in `data_full` have missing values. If so, how many missing values are there in each variable? Use the variable average to replace missing values in the dataset.

4 Preliminary Customer Analysis

Next, once the final dataset is ready, we can proceed to use descriptive analytics to conduct preliminary customer analysis using `dplyr`.

Descriptive analytics is concerned with summarising and highlighting patterns in current and historical data in order to assist businesses in comprehending what has occurred thus far. However, it makes no attempt to

explain why something occurred or to forecast what may occur in the future. "why something occurred" belongs to the scope of causal inference; "forecast what may occur in the future" falls in the scope of predictive analytics. To answer those questions, analysts must combine descriptive analytics with other types of analytics, which we will learn in later weeks.

Your task is to 'get to know' the `data_full` by conducting some descriptive analytics.

Question 4

- Report the summary statistics of the cleaned data, discuss your business insights (open question; from central tendency and dispersion perspectives)

Question 5

- Mutate a new variable called `total_spending`, which means the sum of spending on all products by customers.
  - total_spending = MntWines + MntFruits + MntMeatProducts + MntFishProducts + MntSweetProducts + MntGoldProds
- Mutate a new variable called `Frequency`, which means the sum of the number of purchases made through different channels.
  - Frequency = NumStorePurchases + NumWebPurchases + NumCatalogPurchases
- Provide a correlation table between `total_spending` and other continuous variables, including Frequency, Complain, Response, Income, Kidhome, Teenhome, Recency.

Question 6

- Compare the average total spending across demographic groups. Which group has the highest average spending?
  - marital status
  - education

Question 7

- What is the average total £ spent on wine and total spending by customers with and without kids?
  - tip: first mutate a variable called `has_kid`, which equals 1 if the number of kids in the household is larger than 0, and otherwise 0; then group by this `has_kid` variable

5 After-class Exercise

After the class, you can further explore the dataset and conduct more descriptive analytics.

Please try to solve the questions on your own first, and then check the answers provided.

The solutions can be found on the html version of this case study.

Question 8

- Which education group accounts for the largest percentage of customers?
  - tip: in dplyr, there is a function called `n()`, which counts the number of rows in the group after `group_by()`

Question 9

For complainers and non-complainers who have Master's degrees, find the total number and also the percent of customers who responded to the offer.

Question 10

Group customers by Marital Status and compare the average spending in each group. What insights can you draw? - Tips: use `group_by()` + `summarise()`

Berman, Ron, and Ayelet Israeli. 2022. "The Value of Descriptive Analytics: Evidence from Online Retailers." *Marketing Science*, March, mksc.2022.1352. https://doi.org/10.1287/mksc.2022.1352.