

Background

The company wishes to assess the impact of its marketing strategy over the past 2 years for its flagship drug NZT-48, which is used to treat chronic Lyme disease. Your job as a data scientist is to estimate a dollar-value of sales attributable to each strategy. Your findings will be used to guide future budget-setting.

Additionally, the company wants to know the extent of loss of potential sales due to a new competitor drug entrance into the market since mid-2015. While the new drug is not a perfect substitute for NZT-48, the company is concerned about its impact on sales and is considering the costly option of switching marketing strategies.

Dataset description

Variable	Type	Description and values
accID	Numeric	Unique identifier of an account
accType	Numeric	Account type. Values: "Hospital"; "Private Clinic"; "Pharmacy"; "Polyclinic"
accSize	Nominal	Estimated number of patients seen in a month
accTargets	Numeric	Number of targeted doctors in account
district	Nominal	Integer identifier corresponding to 9 districts
month	Nominal	Date at the start of the month
sales	Numeric	Sales in SGD
qty	Numeric	Sales in units
strategy1	Numeric	Expenditure on marketing strategy 1
strategy2	Numeric	Expenditure on marketing strategy 2
strategy3	Numeric	Expenditure on marketing strategy 3
salesVisit1	Numeric	Expenditure on type 1 visits made by sales representative
salesVisit2	Numeric	Expenditure on type 2 visits made by sales representative
salesVisit3	Numeric	Expenditure on type 3 visits made by sales representative
salesVisit4	Numeric	Expenditure on type 4 visits made by sales representative
salesVisit5	Numeric	Expenditure on type 5 visits made by sales representative
compBrand	Numeric	Number of competitor brands in the market

What we will look for

- Well-documented programming scripts
- Thorough analysis (i.e. considered a wide range of possibilities)
- Suitable application of inference techniques
- Convincing results for the business-user with no specialized statistical knowledge
- Clear explanation of insights but not over-simplifying pertinent issues

Deliverables details

Programming script

You should make your code as readable as possible (e.g. comments, variable/object names should make sense). The programming script you used should be named MSD_<intern/FT>_<Name>.<program specific> where intern/FT indicates an internship position while FT indicates a full time position. For example if your name is James, using R and applying for an internship position it should be named “MSD_intern_james.R”.

Technical Report

The technical report details the rationale of your approach and each step of your analysis. Think of this as a log of the analyses you tried and their results. This is intended for your supervisor and fellow data scientists to understand the work you have done. The report should

- Document all your considerations and justify your selected approach
- Contain key estimation results, visual plots or any other relevant analysis outputs
- Highlight any caveats and assumptions of your analysis
- Think beyond the dataset provided and suggest additional resources (please be realistic) which will improve your analysis

In order to not to constrain your analysis, we impose no page limit for the technical report. However, the report should be concise, clear, and straight to the point as possible.

Business Report

The business report should be in the form of a slide deck. The target audiences are business decision-makers who have a layman’s understanding of statistics, so you should aim to convince the validity of your results without going into technical details. These are the people who will act on your results. In addition to any results you think are relevant to project goal, the deck should address the following key questions:

- What is the average marginal impact of each marketing strategy, assuming no concurrent strategies?
- What impact has the entrance of a new competitor drugs had on sales, if any?
- *[Optional challenge – this is your chance to impress us!]* The business suspects that different accounts may respond differently to different marketing strategies. Investigate this and report your findings, if any.

It is even more crucial here to be concise, clear and straight to the point. We will be grading your content rather than the aesthetics. However good slide design and information flow can help the decision-maker know what to focus on. Your deck should not exceed 12 slides in total (including title slide).