

ECoS project: using data analysis to improve marketing response

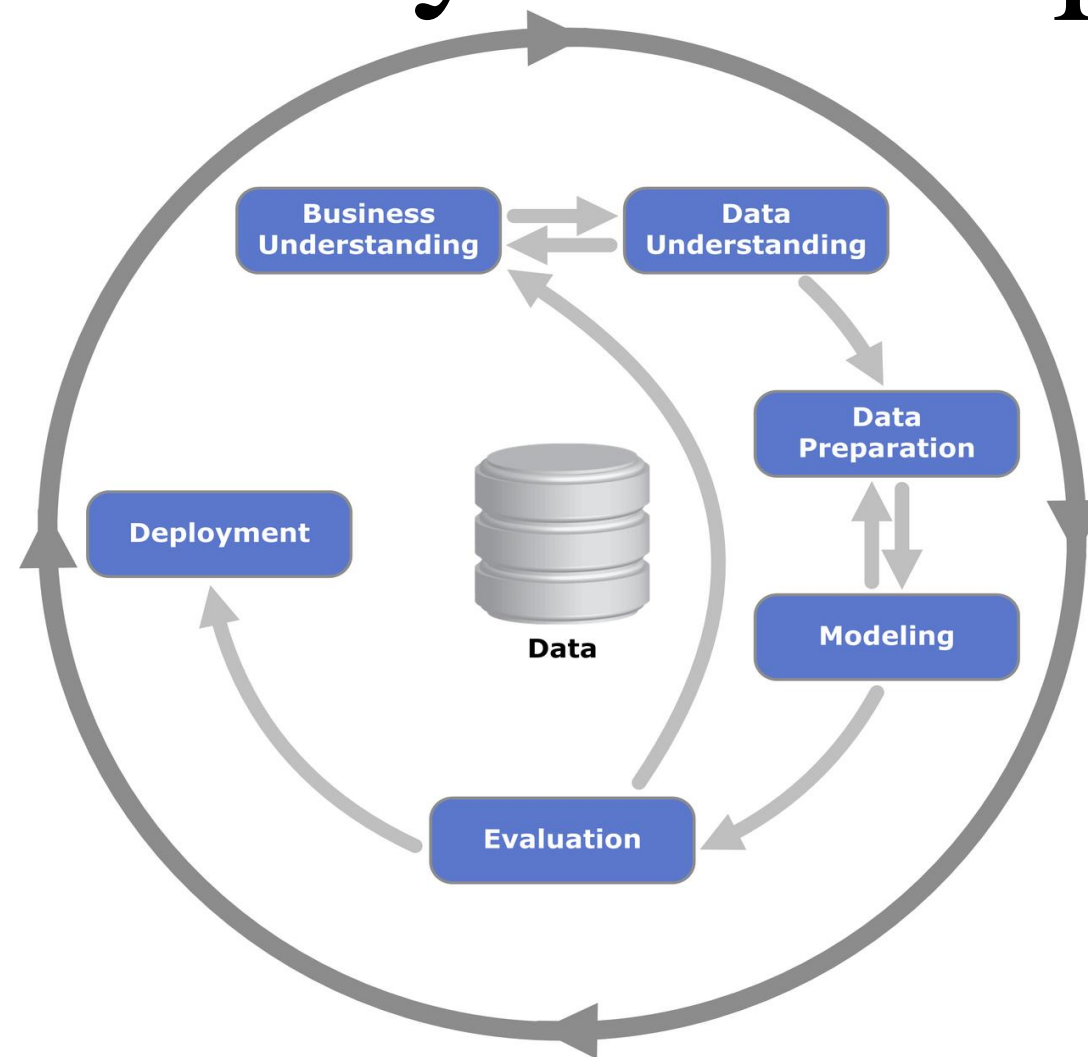
Problem and aims: The project analysed the response to Christmas emails sent by DC Thomson Shop in 2015. The aim was to identify the factors that caused emails to perform well so that future emails could be more effective. In order to do this two models were built: One examined emails, to see how factors such as the position of email links, size of pictures, price of links, number of days to Christmas had an influence on the success of an email.

The second model analysed a dataset of all users that were sent emails and how they responded. This set was merged with all the other information stored about the customer, such as their location, the age of their account, their gender and what time they opened the email amongst others. This model aimed to identify any common factors about users that influenced whether they purchased from the email.

The relationships identified by each model are very important. Targeted emails can be sent to customers likely to respond and emails can be structured more effectively according to what made past emails successful.

There were many challenges involved:

- Creating datasets for analysis
- Scraping data from emails
- Cleaning datasets
- Interpreting missing data
- Analysing datasets
- Visualising Results



Development stages of a big data project.

Methodology: Firstly, time was taken to understand the business, the data and the problem. Data was explored using R, and merged using Python's Pandas. Email data was scraped using regular expressions. Data was cleaned and modelled in R. Decision trees were used as they portray relationships in the data and could be applied to predict how future emails may perform.

Results: The results were presented to senior management. The email model suggested that links near the top and larger images encouraged users to click (see right). This meant future test emails could be sent that targeted these factors. The user model identified interesting areas and also where more data could be collected. For example, a user's location was important, but missing postcodes made inferring their location harder. It was suggested that the site could run engagement events such as competitions that would collect more postcodes to enrich the dataset to enable better analysis.



Christmas email with important factors

Acknowledgements – Chris Kelly, DC Thomson, Andrea Bracciali, The Datalab.