# Finomena Data Science Case Study

**Overview:** The case study is designed to simulate a project. You will be evaluated based upon:

- Business understanding
- Data understanding
- Data preparation
- Modeling
- Evaluation
- Code quality and reusability
- Presentation

The ordering above is not indicative of any prioritization for evaluation.

**Instructions**

1. Review this document in its entirety. Immediately ask any questions that come up.
2. Begin working on the case. You have up to 8 hours to complete the case, however we think 4-6 should be sufficient.
3. At no later than 8 hours from the start time (when you received the email containing this file), email back the entire set of case files and the files listed in the Output section, below.

**Points to Consider**

1. If you have questions, you may email sarvesh.pradhan@finomena.com. We'll do our best to respond ASAP. However, if we do not respond, then feel free to make a reasonable assumption. State this assumption in your presentation.
2. Please do not discuss this case with anyone else. However, you may use any Internet resources for syntactical assistance only.
3. Please complete this case using R, Python, Octave or any other open source tools (only) of your choice.
4. Ensure you can explain all of the steps of your process and code. You will be asked to present your process, code and results as part of the interview process.

**Scenario:** Finomena has pivoted from it's traditional business model and now is looking to sell insurance. The marketing team has a list of candidates they would like to contact, but they would also like to maximize profit across the insurance lifecycle. They turn to the Finomena data science team for help. Your objective is to determine which set of customers the marketing department should contact to maximize profit.

1. The cost of marketing to a particular customer is INR70. This cost is paid regardless of whether the customer responds to our marketing or not.
2. Only if a customer responds to our marketing, can we earn a profit.
3. Profit does NOT include the marketing cost.
4. Total Profit = (Average profit per responding Customer * Number of customers responding) – (Number of customers to whom you marketed * INR70)

**Data:** The marketing team provides data from a prior marketing campaign (training.csv) they had performed during the proof of concept phase of the business model. They also provide a list of potential customers to whom they wish to market (testingCandidate.csv). From this list of potential customers, you need to determine yes/no whether you wish to market to them.

| Type | Name | Description |
|---|---|---|
| Input Variables | custAge | The age of the customer (in years) |
| Input Variables | profession | Type of job |
| Input Variables | marital | Marital status |
| Input Variables | schooling | Education level |
| Input Variables | default | Has a previous defaulted account? |
| Input Variables | housing | Has a housing loan? |
| Input Variables | loan | Has a personal loan? |
| Input Variables | contact | Preferred contact type |
| Input Variables | month | Last contact month |
| Input Variables | day_of_weel | Last contact day of the week |
| Input Variables | campaign | Number of times the customer was contacted |
| Input Variables | pdays | Number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted) |
| Input Variables | previous | Number of contacts performed before this campaign and for this client |
| Input Variables | poutcome | Outcome of the previous marketing campaign |
| Input Variables | emp.var.rate | Employment variation rate - quarterly indicator |
| Input Variables | cons.price.id | Consumer price index - monthly indicator |
| Input Variables | cons.conf.idx | Consumer confidence index - monthly indicator |
| Input Variables | euribor3m | Euribor 3 month rate - daily indicator |
| Input Variables | nr.employed | Number of employees - quarterly indicator |
| Input Variables | pmonths | Number of months that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted) |
| Input Variables | pastEmail | Number of previous emails sent to this client |
| Target Variables | responded | Did the customer respond to the marketing campaign and purchase a policy? |
| Target Variables | profit | If the customer purchased a policy, how much profit (before marketing costs) did the company make on the policy? |

**Output:** Please email sarvesh.pradhan@finomena.com back the following files:

1. All associated code files you used to complete your analysis. Such that the code is in a state to be run.

2. Please add a column to the testingCandidate.csv file. In this column, for each observation indicate a 1 (yes) or a 0 (no) whether you wish to market to that candidate.
3. Please prepare a presentation to show to Finomena employees onsite. The presentation should be written as if you were presenting your results to a non-technical audience (including marketing) but one that is comfortable with graphs, and basic statistics.