**Uptake Data Science Case Study**

**Overview:** Enclosed is a case study that Uptake asks candidates for its Data Science team to complete. The case study is designed to test a real-world Data Science problem. You will be evaluated based upon:

1. Business understanding
2. Data understanding
3. Data preparation
4. Modeling
5. Evaluation
6. Code quality and reusability
7. Presentation, both to a non-technical audience and communication with other data scientists.

**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, email back the entire set of case files and the files listed in the “**Output You Are To Provide”** section.

**Points to Consider**

1. If you have questions, you may email recruiting-ds@uptake.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. There isn’t meant to be any tricks here. This is just a straightforward data analysis problem.
3. Please do not discuss this case with anyone else. However, you may use any Internet resources for syntactical assistance only.
4. Please complete this case using R, Python or Matlab/Octave.
5. Ensure you can explain all of the steps of your process and code. You will be asked to present your process, code and results to Uptake employees as part of the interview process.

**Scenario:** You are working for Uptake’s non-profit arm, [Beyond Uptake](https://beyond.uptake.com/) as a Data Scientist. Beyond Uptake has volunteered to assist a charity to develop a tool to optimize their outreach efforts. Your objective is to determine which set of potential donors the charity should contact to maximize the profitability of a marketing campaign to solicit donations for the charity.

1. The cost of marketing to a particular potential donor varies per zip code. This cost is paid regardless of whether the potential donor responds to our marketing or not. This cost is specified in the file zipCodeMarketingCosts.csv.
2. Only if a customer responds to our request, do we receive the donation amount.
3. The “amount” variable does not include the marketing cost.
4. Profitability of campaign = sum of all amounts received – sum of all marketing costs for customers to whom you marketed.

**Files You Are Provided:**

1. data.7z/train.csv: Historical data of potential donors. This data set contains a huge amount of data about the donors and that is defined in the dataDictionary.txt file. However, especially important fields are:
   1. id: A unique identifier for a given potential donor.
   2. amount: The dollar amount that was given by the donor for that donation only.
   3. responded: Whether or not the potential donor responded to the previous solicitation.
2. data.7z/test.csv: The company has also provided you with a list of potential customers to whom to solicit donations. From this list of potential donors, you need to determine yes/no whether you wish to solicit to them. As mentioned, the cost to solicit to a particular donor depends on their zip code. Further, since we do not know whether or not they will respond, this file does not contain the variables “amount” or “responded”. The column “market” initially contains all NA values. You will populate your results here in accordance with the instructions in the “Output You Are To Provide” section of the instructions.
3. zipCodeMarketingCosts.csv: The cost to market to a given zip code.
4. dataDictionary.txt: A list of all of the columns and their intuitive explanations.

**Output You Are To Provide:**

At the end of the 8 hour window, please email back the following files:

1. All associated code files you used to complete your analysis.
2. **IMPORTANT:** Email back the test.csv, with the “market” column populated with a 1 if you wish to solicit a donation from that customer and 0 if you do not wish to solicit a donation from that customer. We recommend that you save a few minutes to make sure that you:
   1. Did not add or remove any columns from this data set.
   2. Did not add or remove any rows from this data set.
   3. That you entered a 1 to the market column if you wish to solicit the potential donor and a 0 if you do not wish to solicit that potential donor.
3. Please prepare a presentation to show to Uptake employees during the final interview. The presentation should be written as if you were presenting your results to a non-technical audience from the non-profit that Beyond Uptake is working with. Plan on this presentation lasting no more than 10-15 minutes.

If you wish to make any additional changes after the 8 hour window, we’d be happy to look at them. Please send them as a separate email before your final interview. However, we’ll primarily score your initial submission.

**Frequently Asked Questions (FAQ)**

1. **Question**: I have a question about a certain variable. Can you tell me more about it? **Answer**: Unfortunately we do not have any more information about the data set. Feel free to make any reasonable assumptions.
2. **Question:** I found some really weird thing with one of the variables or observations! Why is that? **Answer**: We don’t know. This is a real-world data set. As such, it contains lots of real-world messiness. However, there are not any purposeful tricks to this data set.
3. **Question:** What does this data set have to do with what Uptake does? **Answer:** Uptake does not directly work in this domain. However, we do feel that this scenario is representative of a general data science type problem that one might encounter.
4. **Question:** I don’t know anything about non-profits or soliciting donations. How can I complete this case study? **Answer:** No domain knowledge is needed and would not really be helpful for a case study like this.
5. **Question:** I am running out of time or I am getting an unexpected error. What should I do? **Answer:** We understand that things happen. Feel free to include placeholders in either your presentation or code that indicate what you would do if you had more time.