**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 simulate a real project at Uptake. 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

**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 section.

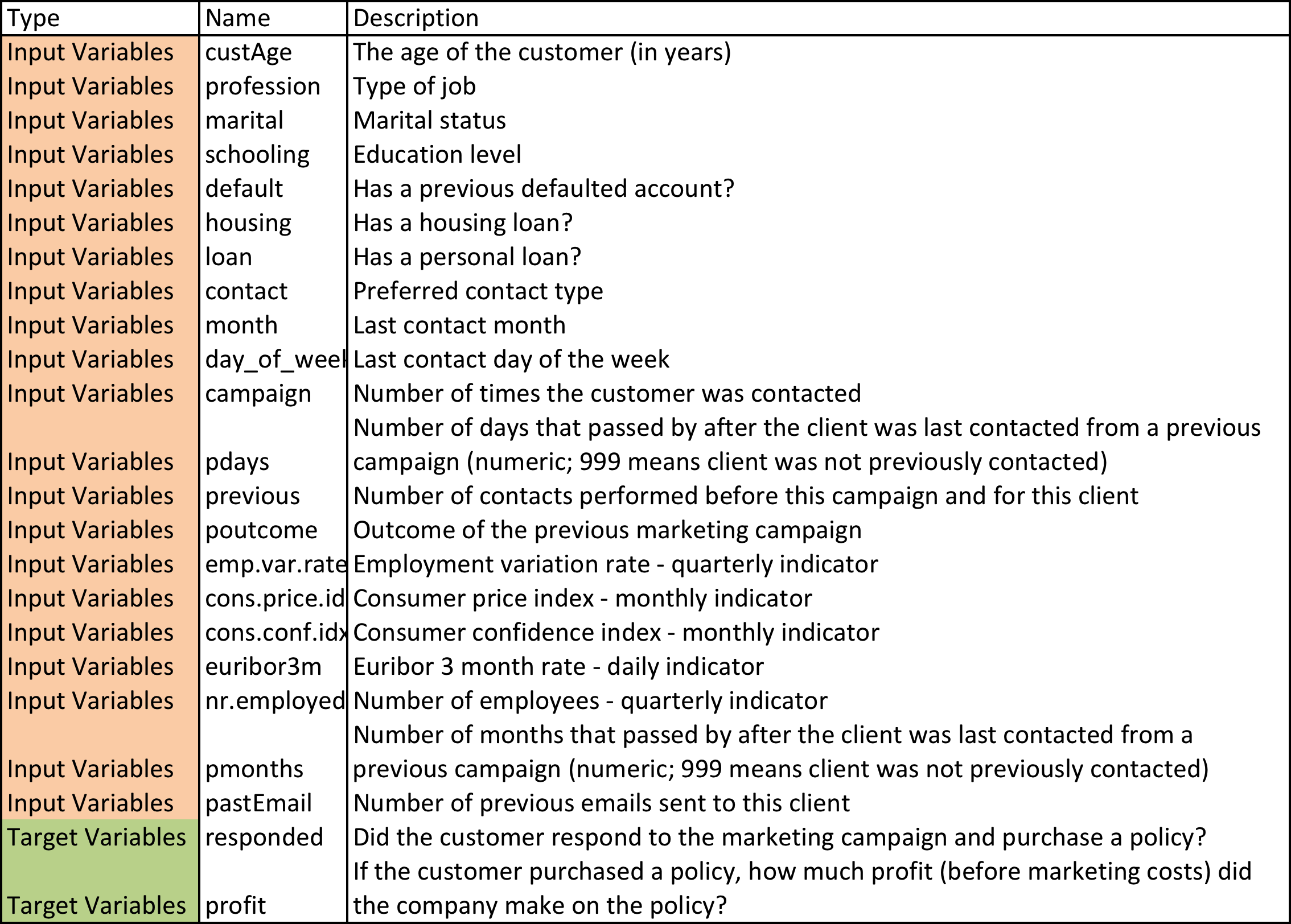
**Points to Consider**

1. If you have questions, you may email [adam.mcelhinney@uptake.com](mailto:adam.mcelhinney@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 as a Data Scientist. Uptake has been commissioned by an insurance company to develop a tool to optimize their marketing efforts. Your objective is to determine which set of customers the marketing firm should contact to maximize profit.

1. The cost of marketing to a particular customer is $30. This cost is paid regardless of whether the customer responds to our marketing or not.
2. Only if a customer responds to our marketing, do 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 \* $30

**Data:** The insurance company has provided you with a historical data set (training.csv). The company has also provided you with a list of potential customers to whom to market (testingCandidate.csv). From this list of potential customers, you need to determine yes/no whether you wish to market to them.



**Output:** Please email back the following files:

1. All associated code files you used to complete your analysis.
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 Uptake employees onsite. The presentation should be written as if you were presenting your results to a non-technical audience at the client.