Dear Data Science Candidate,

Congratulations on making it to this round of the data science interview!

Thank you for taking the time out of your schedule to work on this take-home project. For this next part, we want to see how you handle an end-to-end project. You will start with raw data, create a predictive model, present your work to our team, and have a conversation about your process.

We value your time, so please try to keep your work within 4 hours. Included in the zip file is training data and test data pulled from raw logs. The following page has more details on the project itself. Good luck and feel free to reach out with any questions you have!

Best regards,

The 6sense Data Science Team

6sense Data Science Interview Project

**Problem Statement:**

Using the training data, build a model that predicts which user\_id's will make a purchase in the FUTURE given the user's PAST activities.

**Deliverables:**

1. Answers to the following questions:
   1. How did you approach the time-based aspect of the problem?
   2. Describe which activity types you believe are most useful in predicting which user will purchase in the future.
   3. If you had more time, what else would you do?
2. Provide a file (csv, tsv, or similar) containing user\_id's from the test data and the probability that these users will make a purchase. Please ensure that the file only has these two columns.
3. Your code that you used to generate your answers.

**Evaluation:**

1. Score of your output file
2. EDA
3. Feature engineering
4. Model building
5. Model performance
6. Answers to the questions in Deliverable #1 above
7. Presentation and discussion

**Included fields:**

* user\_id: A hash that uniquely identifies the user
* activity\_date: The date of the activity
* activity\_type: The type of activity