**SLALOM New York  
Advanced Analytics**

Data Challenge Case Study

Data Challenge Case Study

Slalom Consulting NY

The goal of this exercise is to get a glimpse of how you would approach a typical data science problem end-to-end. We are providing you with data from an E-commerce website that would like to systematically identify fraudulent transactions on their platform. You’ll need to explore the data and build a model to predict whether a user’s activity is fraudulent. You only have information about the user’s first transaction on the site and based on that you must make a classification (Fraud/No Fraud).

Feel free to use your analysis techniques of choice, but please code in either **Python or R** and make sure to send the code in either **markdown or notebook** format. Below are a few helpful hints:

* Document your work clearly, we want to understand your thought process.
* This is not a modeling competition, so we're not looking for complex ensemble methods (e.g. stacking). We're more interested in understanding what methods you chose, why, and how you applied them to the problem.

**Please spend no more than 10 HOURS TOTAL to complete the analysis, and PLEASE SEND YOUR WORK BACK WITHIN 7 DAYS OF RECEIPT.**

The dataset can be found here:

<https://www.dropbox.com/s/i9efe59ct1u3v9q/Fraud.zip?dl=0>

Dataset:

"**Fraud\_Data**" - information about each user first transaction

Columns:

* **user\_id**: Unique user ID
* **signup\_time**: The time when the user created their account (GMT time)
* **purchase\_time**: The time when the user bought the item (GMT time)
* **purchase\_value**: The cost of the item purchased (USD)
* **device\_id**: The device id. You can assume that it is unique by device. (If two transactions have the same device ID that means that the same physical device was used to purchase)
* **source**: User marketing channel: ads, SEO, direct
* **browser**: The browser used by the user.
* **sex**: User sex: Male/Female
* **age**: user age
* **ip\_address**: user numeric IP address
* **class**: This is the target variable. Fraudulent activities is tagged as ‘1’, whereas non-fraudulent activities are tagged as ‘0’