

TECH ALPHARETTA

2972 Webb Bridge Rd, Alpharetta, GA 30009

**Fraud Prevention**

One of the most effective domains for machine learning is Fraud Prevention, especially in the e-commerce space.

E-commerce is a rapidly growing industry, but so is online fraud.

In this challenge, you'll implement a machine learning model that predicts whether a user transaction is potentially fraudulent.

**Background**

E-commerce fraud takes several forms, including:

* Using stolen credit cards
* Improper chargebacks
* Money laundering
* Identity theft

Some of these forms can be extremely costly, both for the consumer and for the e-commerce store. As a result, machine learning has been used heavily for preventative measures.



Offline money laundering?

**Data**

Our first table is **ecom\_txns.csv**.

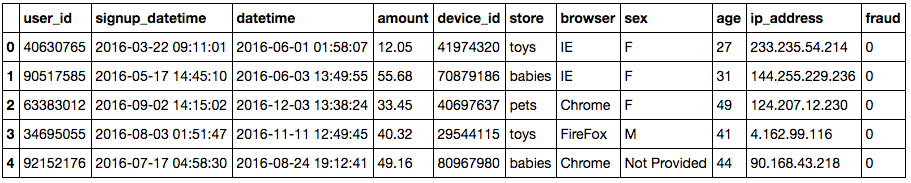
It contains past user transactions across three different e-commerce stores owned by the same parent company.

*Data Dictionary:*

* **user\_id -**Unique ID for the user.
* **signup\_datetime** - First time the user signed up for the site.
* **datetime** - First time the user made a purchase.
* **amount** - Amount of purchase.

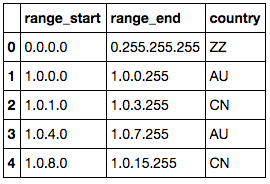
*Data Dictionary (continued):*

* **device\_id** - Unique ID of the user's device.
* **store** - E-commerce property (babies, toys, or pets)
* **browser** - User's browser.
* **sex** - User's gender.
* **age** - User's age.
* **ip\_address** - User's IP address.
* **fraud** - Was the transaction fraudulent or not?



Our second table is **ip\_mappings.csv**.

It contains IP address ranges for different countries.



*Data Dictionary:*

* **range\_start** - Start of IP range.
* **range\_end** - End of IP range.
* **country** - Country.

**Objectives**

This challenge requires 2 specific deliverables.

* *Deliverable 1:* A table of fraud rate by country.
  + You'll need to way to map users to their country by their IP addresses.
  + Which countries have the highest fraud rates?
* *Deliverable 2:* A classification model to predict whether a transaction is fraudulent.
  + Discuss how you could use this model to help the business.
  + How might you use this model in a new product?
* Provide actionable insights to the business. What have we learned from this analysis?