**train.zip**

* **Client\_train.csv** - Client information in the train population
* **Invoice\_train.csv** - Clients invoice in the train set

**test.zip**

* **Client\_test.csv** - Client information for the test population
* **Invoice\_test.csv** - Clients invoice in the test set
* **SampleSubmission.csv** - is an example of what your submission file should look like. The order of the rows does not matter, but the names of the IDs must be correct. The column "target" is your prediction.

**Variable definitions**

**Client:**

* Client\_id: Unique id for client
* District: District where the client is
* Client\_catg: Category client belongs to
* Region: Area where the client is
* Creation\_date: Date client joined
* Target: fraud:1 , not fraud: 0

**Invoice data**

* Client\_id: Unique id for the client
* Invoice\_date: Date of the invoice
* Tarif\_type: Type of tax
* Counter\_number:
* Counter\_statue: takes up to 5 values such as working fine, not working, on hold statue, ect
* Counter\_code:
* Reading\_remarque: notes that the STEG agent takes during his visit to the client (e.g: If the counter shows something wrong, the agent gives a bad score)
* Counter\_coefficient: An additional coefficient to be added when standard consumption is exceeded
* Consommation\_level\_1: Consumption\_level\_1
* Consommation\_level\_2: Consumption\_level\_2
* Consommation\_level\_3: Consumption\_level\_3
* Consommation\_level\_4: Consumption\_level\_4
  + New\_Index – Old\_Index = Total Consumption (Level 1 + 2 + 3 + 4)
  + Each level has an upper limit, then the consumption accumulates on the next level, and so on
* Old\_index: Old index – **Looks like the previous meter reading (ie. New\_Index from previous invoice\_date)**
* New\_index: New index – **Looks like the next meter reading (ie. Reading at current invoice\_date)**
* Months\_number: Month number
* Counter\_type: Type of counter