**Project Jarvis**

Projects end goal is to build our own fraud system that will be activated in cases where investigators cannot handle the amount of transactions in the fraud queue\*.

At this point we check fraud with even less data than the data provided. We are working with IT to help stream ON-SITE data (hits, time on site, mouse movement map etc..). This will be the second stage of the analysis.

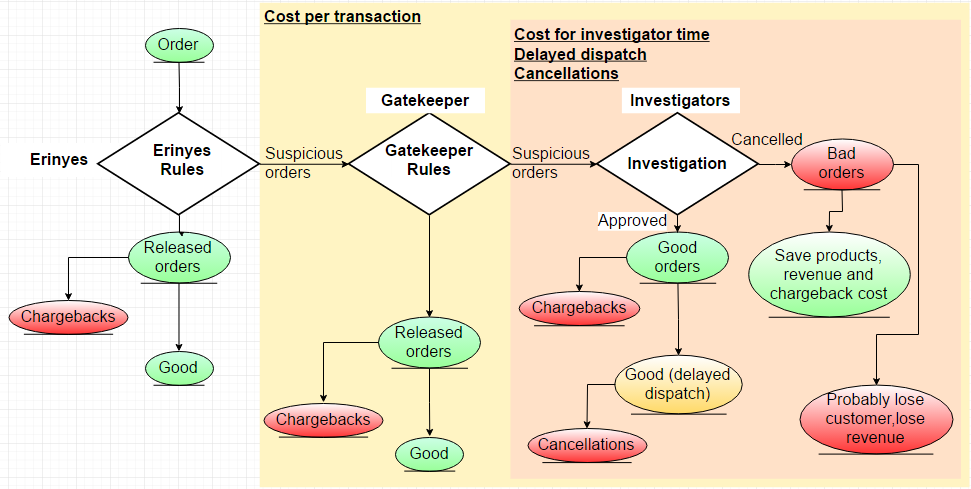
The first stage involves an initial analysis to check if we need to provide extra variables to the fraud platform that will help us identify fraud easier.

**Documents included:**

List of Data sets + Read me file

* MAIN\_customer\_data.csv
* MAIN\_transaction\_data.csv
* MAIN\_chargeback\_data.csv
* delivery\_option\_lookup.csv
* country\_code\_lookup.csv
* payment\_provider\_lookup.csv
* payment\_status\_lookup.csv
* medium\_lookup.csv
* locale\_lookup.csv
* payment\_method\_lookup.csv
* Read me.docx

**Transaction flow throughout Fraud checks**



\*Fraud queue: Amount of suspicious orders that need to be checked by our investigators

**MAIN\_customer\_data.csv**

|  |  |
| --- | --- |
| Account\_Key | Unique identifier for a user account on a particular site  Customer can have an account only for particular site. If he orders from several our sites, he has several different accounts. |
| Registered\_Date | Date when customer registered account with us |
| Country | Country customer provided at registration |
| PostCode | Postcode customer provided at registration |
| First\_Order\_Placed | Date of first order |
| Site\_Key | Unique identifier for every different site (regardless particular Locale) |
| Locale | Subsite, e.g. Amazon**.com**, Amazon**.co.uk**  In the format (the site domain)\_(language of the site) |
| SCV\_Key | Hash of the email customer provided at registration  Could be used to link the accounts created by the same customer.  E.g. if a user creates accounts on sites X and Y with the same email address, SCV key ties those accounts to the same user |
| EDomain | Domain of email address used at registration |
| X | Empty column |

**MAIN\_transaction\_data.csv**

|  |  |
| --- | --- |
| Order\_Number | Unique identifier of a transaction |
| Account\_Key | Unique identifier for a user account on a particular site |
| Delivery\_Option\_Type\_Key | Key of the delivery type  (Lookup table provided) |
| Order\_Date\_Key | Date of the transaction |
| Payment\_Method\_Key | Unique identifier of different payment methods  (Lookup table provided) |
| Locale\_Key | Key for Locale (see above)  (Lookup table provided) |
| Order\_Sequence\_No | The number assigned to the order to count the current amount of orders made from that account.  Example  Account number: 1234  Orders history on that account:  01.03.2014 – order 001 (allocated order sequence No “1”)  05.07.2015 – order 565 (allocated order sequence No “2”)  12.01.2017 – order 788 (allocated order sequence No “3”) |
| Medium\_Key | Unique identifier of different mediums  We record the page customer came to our site from.  DIRECT – the address of our site was explicitly entered in url string  ORGANIC – customer searched for our site by it’s name  EMAIL – link in promotion email used  SOCIAL – link on company’s social media sites used  AFFILIATE – link on paid personal site used  REFERRAL – link on personal site used  PPC (paid per click) – link in the advertisement used  (Lookup table provided) |
| Campaign\_Key | Some product or all products within some order can be considered to be sold within some particular campaign. Campaign – trading campaign on some site, like “January Sale” or “Penny Sale” |
| Order\_Payment\_Status\_Key | Unique identifier of payment status  0 Validated OK – released by Erynies or Gatekeeper  1 HELD – Marked as “Suspicious” by Gatekeeper and referred to investigators for check  6 In external checking – order is in process of being checked in Erynies or Gatekeeper  7 Failed external checking – order was Rejected by investigator  9 Failed – order was released but had Payment Problem (now customer has the opportunity to pay for that order again, but then his order will undergo checking process again)  (Lookup table provided) |
| Payment\_Provider\_Key | Unique identifier of different payment providers  (Lookup table provided) |
| Ordered\_Product\_Key | Unique key of one of the products in the order (the product is sold on one or several sites, it’s key is the same) |
| Category\_Level\_2 | Product category (e.g. snacks) |
| Category\_Level\_3 | Product sub-category (e.g. chocolate) |
| Product\_Charge\_Price | Price of the product (affected by any discounts/campaigns/promotional codes) **in GBP** |
| Ordered\_Qty | Amount of the units of that product in the order |
| Cancelled\_Qty | Number of units of that product which were cancelled  Example of customer order:  Product\_Key Ordered\_Qty Cancelled\_Qty  0001 3 2  0002 2 1  0003 1 1  0004 1 0  (Each different product in the order has his own row in the table) |
| Cancelled\_Date\_Key | Date of cancellation (Cancellation can be partial – only some products of the order were cancelled, or complete) |
| customer\_ip\_address | IP address used for placing that transaction |
| Postcode\_**Shipping\_Address** | Postcode provided in Shipping address.  Shipping and Billing Addresses are selected from the list or typed by customer each time he places the order; Shipping and Billing address can differ and addresses may change with every next order. |
| Country\_Code\_**Shipping\_Address** | Country provided in Shipping address  We use unique country\_code for each country  (Lookup table provided) |
| Postcode\_**Billing\_Address** | Postcode provided in Billing address |
| Country\_Code\_**Billing\_Address** | Country provided in Billing address  (Lookup table provided) |
| Empty1 | Empty field 1 |
| Empty2 | Empty field 2 |

**MAIN\_chargeback\_data.csv**

|  |  |
| --- | --- |
| Order.No. | Unique identifier of a transaction |
| Date.Logged | Date we received the chargeback notification and logged to our systems  We can receive chargeback during 3 months’ time after the order was placed (in very rare cases – even later) |
| Released\_by | **Bypass** – Released from Erinyes (our own system that uses a basic set of rules to release a transaction of send it to Gatekeeper)  **Gatekeeper** – A third party fraud platform we use to identify fraud. Uses a rules based model to give a score to each transaction.  Score of transaction = ∑ of the scores of all rules that triggered. If score is < some threshold (site-specific value), transaction is released; otherwise it is considered suspicious and referred to investigators.  **Investigators**: Part of the fraud team that check the transactions that are currently marked as “suspicious fraud” in Gatekeeper.  They can release or reject the transaction. |
| Internal.RC | Type of fraud |
| Defence.Status | In case we have a defence policy against a particular claim or we can support with evidence that customer is not telling the truth we defend the case. If we successfully defend against a claim we don’t refund the customer |
| GBP.Amount | Amount of the chargeback **in GBP**  Can be only part of the order price (if CB is requested only for some products ordered), can be higher than original order price, because we refund shipping cost. |

**Note:**

* Keys without lookup tables are not provided for security purposes
* Sensitive information like names, full address etc… couldn’t be provided even if we understand it’s important for fraud detection!
* Any comments or queries email directly
  + [Konstantinos.Tsirogiannis@thehutgroup.com](mailto:Konstantinos.Tsirogiannis@thehutgroup.com)
  + [Arina.Sinakova@thehutgroup.com](mailto:Arina.Sinakova@thehutgroup.com)
  + [Sean.Stock@thehutgroup.com](mailto:Sean.Stock@thehutgroup.com)