

**Financial Intelligence Unit - India (FIU-IND)**

**Approach Note – Transaction Based Relationship – AI-ML**

**CONFIDENTIALITY CLAUSE**

This document is the property of Financial Intelligence Unit - India under Project FINnet 2.0. All ideas and information contained within the document is the intellectual property of FIU-IND. These documents are not for general distribution and are meant for use solely by the person/persons to whom it is specifically issued to. Copying or unauthorized distribution of these documents, in any form or means including electronic, mechanical, photocopying or otherwise is illegal.

DOCUMENT CONTROL

|  |  |
| --- | --- |
| Document Version Control Note | |
| Name of the Document | Approach Note - RE Clustering |
| Document Version Number | 1.0 |
| Effective Date | 28-06-2022 |
| Approved By | FIU-IND |
| This Revision Supersedes | NA |
| Document Classification | Restricted |
| Distribution List | FIU-IND, PMU, MSP |
| Access Level | Read Only |

REVISION HISTORY

| Version | Release Date | Prepared By | Reviewed/  Approved By | Description of changes made | Section of document Impacted |
| --- | --- | --- | --- | --- | --- |
| 1.0 | 28-06-2022 | Nabeel Ansari | Robin Singh/Kuldeep Rana | Initial Draft |  |
|  |  |  |  |  |  |

REFERENCE DOCUMENTS

| S. No. | Document Name | Document Description |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

1. **Methodology of Transaction based relationship:**

* There are 4 Relationships which are to be created based on different Transactions:
  + Person to Person(P-P)
  + Organization to Organization(O-O)
  + Person to Organization (P-O)
  + Organization to Person(O-P)
* There are 5 transactions considered which are NEFT, RTGS, UPI, CBWTR, IMPS. Below are the FINGATE Tables which are used to fetch data.
  + NEFT -> Fingate\_TsOne and Fingate\_TsTen
  + RTGS -> Fingate\_TsOne and Fingate\_TsTen
  + UPI -> Fingate\_TsThree and Fingate\_TsTen
  + IMPS -> Fingate\_TsTwo and Fingate\_TsTen
  + CBWTR -> Fingate\_Cbwtrbank
* Scoring has been implemented using statistical technique of binning.
* Data is sorted, divided into three parts (binned) and a new label is assigned to each value in accordance with the bin they fall into.
  + Transaction\_Amount :- Total amount of transaction between two entities.
  + Number\_Of\_Transactions: Total number of transactions between two entities.
  + Frequency\_Of\_Transactions: Time period of all the transactions between two entities.
* We have 3 different columns and each column will have 3 different bin sizes B1, B2 and B3. We will be assigning Relationship strength of Low, Medium and High on respective bins.
  + Presence of Data Point in Third Bin for at least one Column will result in High
  + Presence of data point in Second bin for at least one Column will result in Medium
  + Presence of data point in First bin for all 3 Columns will result in Low.

|  |  |  |  |
| --- | --- | --- | --- |
| **Transaction Amount** | **Transaction Count** | **Frequency of Transaction** | **Relationship Strength** |
| B1 | B1 | B1 | Low |
| B1 | B1 | B2 | Medium |
| B1 | B2 | B3 | High |
| B2 | B2 | B1 | Medium |
| B2 | B3 | B2 | High |
| B2 | B3 | B3 | High |
| B3 | B1 | B1 | High |
| B3 | B1 | B2 | High |
| B3 | B2 | B3 | High |
| B1 | B2 | B1 | Medium |
| B1 | B3 | B2 | High |
| B1 | B3 | B3 | High |
| B2 | B1 | B1 | Medium |
| B2 | B1 | B2 | Medium |
| B2 | B2 | B3 | High |
| B3 | B2 | B1 | High |
| B3 | B3 | B2 | High |
| B3 | B3 | B3 | High |
| B1 | B1 | B1 | Low |
| B1 | B1 | B2 | Medium |
| B1 | B2 | B3 | High |
| B2 | B2 | B1 | Medium |
| B2 | B3 | B2 | High |
| B2 | B3 | B3 | High |
| B3 | B1 | B1 | High |
| B3 | B1 | B2 | High |
| B3 | B2 | B3 | High |

* Finally, this score is multiplied by the corresponding weights of their transaction type. Currently,

we have assigned below Weight to each Transaction Type which is configurable and can be changed as well:

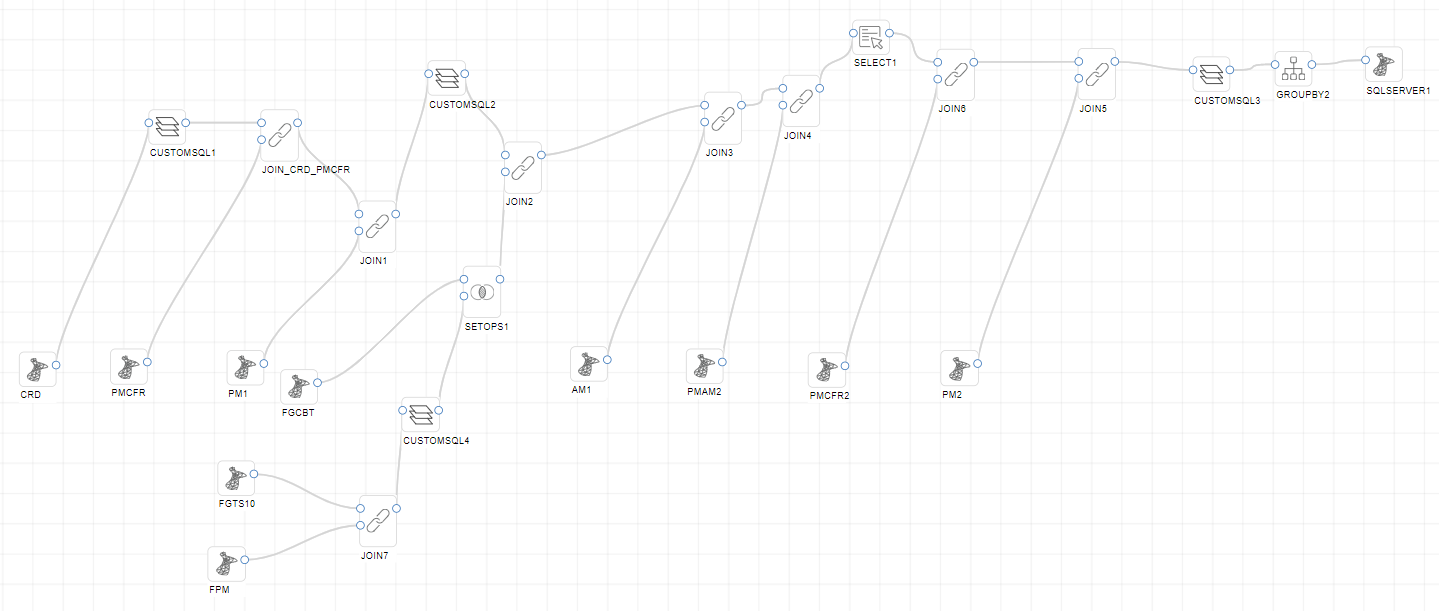
|  |  |
| --- | --- |
| **Transaction Type** | **Weightage** |
| CBWTR | 1 |
| NEFT | 0.9 |
| RTGS | 0.9 |
| IMPS | 0.8 |
| UPI | 0.7 |

1. **Assumptions:**

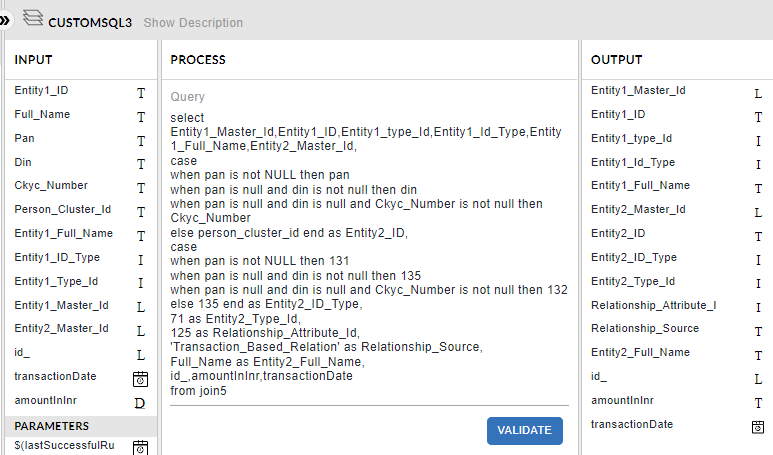
* Binning Exercise should run once in a Year wherein entire Binning exercise and Bins assignment for each Transaction Data would be done.
* The above point was discussed on 06/07. Sir wanted the Frequency of Binning Exercise to be flexible such that it could be changed later and we can lower the Frequency of complete Binning.
* 'Bin Assignment' would be run every month and only new Transaction Data would be considered and Bins will be assigned.
* Currently, Frequency of transactions between two entities is considered to be yearly. However, monthly Frequency can also be considered.

1. **Implementation Steps:**
2. A total of 16 FOSFOR Spectra flows are created to capture 4 different relationships
3. Data is retrieved from Cps\_ReportDetails having Processing\_Flag\_Cps2 value equals to 1 and Processing\_Flag\_Cps3 as 0.
4. This table is then joined with Person\_Master\_Cluster\_Fingate\_Report and Organization\_Master\_Cluster\_Fingate\_Report for person and organization respectively to get the Entity’s cluster Id.
5. This data is further joined with person\_master or organization\_master table to get the entity’s master id, Pan/din/gstin/ckyc number, Entity’s name, depending on the type of relationship.
6. Now, Pan/din/gstin/ckyc number is renamed to Entity1\_ID based on whichever is available in the following order of priority Pan>din>ckyc number>entity cluster id.
7. Entity1\_Id\_Type contains the mapped value of the entity1\_id e.g. 135 for pan.
8. Entity1\_Type\_id contains the mapped values for entity\_type e.g. 71 for person.
9. Now, the data is extracted from the corresponding tables based on the transaction type having is\_active\_flag equals to 1. Apart from the main table, data is also extracted from fingate\_TsTen. The information fetched from these tables includes the beneficiary details that acts as the second entity.
10. Further, Finnet\_paymentmode table is used to identify the type of transaction in fingate\_TsTen table.
11. Steps 6-10 are repeated for the second entity involved in the transaction.
12. The data is then grouped to get total number of transactions, amount of transactions, latest transaction and the last transaction.
13. Relationship\_attribute\_id column is added to the table and assigned a value as mentioned in the CPS document.
14. This data is then inserted in the resolved\_relationship table which then acts as the input for the transaction based relationship scoring using AI/ML.

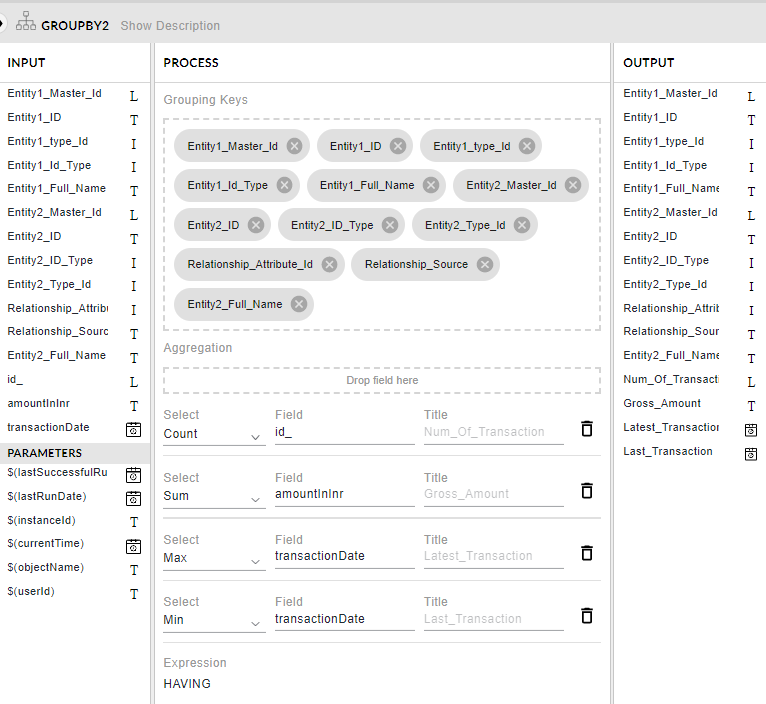
**Flow demonstrating person to person relationship for CBT transaction type**



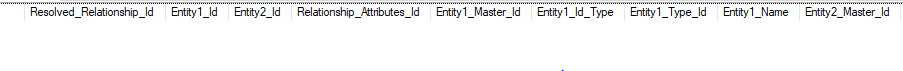
**Custom sql3**



**Groupby Expression**



**Resolved\_Relationship table**



Relationship Score Table

