**Fraud Alerts – Hackathon 2017**

**Requirements:**

Input Two Datasets – Master Data Set & Transaction Data Set

**Master Data Set Details :-**

100 rows of CSV File

Format:-

|  |  |  |  |
| --- | --- | --- | --- |
| S.No | Customer ID | Customer Name | Salary |

**Rules for Creating Master Data Set :**

1. Salary Need to be in the range of less than 5L, 5L to 10 L and greater than 10L
2. Out of 100 customers, there can be any probablities of low, medium or high salaried customers. This data is fixed

**Properties File for Alert Condition:**

1. Before Nov’2016 – Number of Credits & Numbers of Debits for each type of a customer.
2. After Nov’2016 - Number of Credits & Numbers of Debits for each type of a customer.
3. Highest Amount of Deposit for each type of customer

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Percentile for Transaction Data | Range in Salary | Before Nov’2016 | | After Nov’2016 | | High Deposit |
|  |  | Credit | Debit | Credit | Debit |  | |
| 30 % | <= 5,00,000 | <=2 | 5 | 4 | 7 | 10,00,000 | |
| 40% | 5,00,001 to 10,00,000 | <=4 | 12 | 6 | 12 | 15,00,000 | |
| 30% | >=10,00,000 | <=10 | 25 | 15 | 30 | 30,00,000 | |

**Transaction Data Set :**

1. Date Rage to be in 01-JAN-2016 to 31-MAR-2017.
2. 5,00,000 rows to be created randomly
3. Data set to meet the property file condition

**Format of Transaction Data Set:**

|  |  |  |  |
| --- | --- | --- | --- |
| Transaction Date | Customer ID | Transaction Amount | Debit/Credit |

**Required OUTPUT:**

1. Create Transaction Data Set & Master Date Set based on the conditions given
2. Property file should be configurable
3. Output the possible fraudulent customer details based on properties file.

Output should give – Customer ID, Number of Credit Transactions, Number of Debit Transactions and flag as abnormal

1. Predictive Analysis based on the properties file – Generate a sample dataset and feed it from particular folder path – Generate the output in some folder path.