Features in original invoice and client dataset:

| Attribute/ Features | Data Type | Allowed Values | Remarks |
| --- | --- | --- | --- |
| client\_id | string |  | 135493 Instances |
| district | int | {60,62,63,69} | 135493 Instances |
| client\_catg | int | {11,12,51} | 135493 Instances |
| region | int | {101, 107, 301, 105, 303, 103, 309, 311, 304, 104, 312, 305, 306, 308, 372, 307, 313, 310, 371, 302, 106, 379, 399, 206, 199} | Categorisable into 3 sub categories |
| creation\_date\_day | int |  |  |
| creation\_date\_month | int |  |  |
| creation\_date\_year | int |  |  |
| no\_months\_as\_client | float |  | Number of Months since creation date |
| target (Fraud/NFraud) | int | {0,1} | No Fraud: 127927  Fraud: 7566 |
| invoice\_data | object |  | REFER TO TABLE BELOW |
| services\_consumed | int | Electric: [11, 15, 10, 12, 14, 13, 29, 9, 8, 21, 27, 18, 24]  Gas: [40, 45, 30, 42] | Count of Services consumed |
| region\_group | int | 1,2,3 |  |

Additional features derived from existing ones:

| Feature | Data Type | Allowed Values | Remarks | Done by (JY-Jiaying,NG-Nikita Goh, ST-Shawn Tan ZX-Zhengxin) |
| --- | --- | --- | --- | --- |
| months\_of\_service | float |  | Number of Months of service | ZX |
| number\_of\_counter | int |  | Count of Unique counter for this service | ZX |
| number\_of\_instances | int |  | Count of Number of Invoice | ZX |
| number\_of\_person\_counting | int |  | Count of Unique person counting | ZX |
| min\_reading\_remark | int |  | Min Reading Remark | ZX |
| max\_reading\_remark | int |  | Max Reading Remark | ZX |
| mean\_reading\_remark | float |  | Average Reading Remark | ZX |
| mean\_index\_differenc | float |  | Average Index Difference | ST |
| sum\_index\_difference | int |  | Sum Index Difference | ST |
| min\_index\_difference | int |  | Min Index Difference | ST |
| max\_index\_difference | int |  | Max Index Difference | ST |
| min\_counter\_status | int |  | Minimum Counter Status | ST |
| max\_counter\_status | int |  | Maximum Counter Status | ST |
| mean\_counter\_status | float |  | Average Counter Status | ST |
| mean\_diff\_counter\_coefficient | float |  | Find the difference between each value and average the difference | ST |
| mean\_counter\_coefficient | float |  | Average | JY |
| min\_counter\_coefficient | int |  |  | JY |
| max\_counter\_coefficient | int |  |  | JY |
| sum\_counter\_coefficient | int |  |  | JY |
| sum\_total\_consumption | int |  |  | NG |
| mean\_total\_consumption | float |  |  | NG |
| min\_total\_consumption | int |  |  | NG |
| max\_total\_consumption | int |  |  | NG |
| tally\_check\_true | int |  |  | JY |
| tally\_check\_false | int |  |  | JY |
| sum\_tally\_value | int |  |  | JY |
| min\_tally\_value | int |  |  | JY |
| max\_tally\_value | int |  |  | JY |
| mean\_tally\_value | float |  |  | JY |
| counter\_type | int | {0,1,2} | 0-> GAS  1-> ELEC  2-> BOTH | NG |
| last\_year | int |  |  | NG |
| last\_month | int |  |  | NG |
| last\_day | int |  |  | NG |
| last\_day\_is\_weekday | int | {1,0} | 1-> True (Weekday)  0-> False (Weekend) | NG |

FOR EACH INVOICE ENTRY

FEATURES NOT IN FINAL DATASET BUT NEEDED TO COMPUTE THE STATS

| counter\_status | int | [0,5] | Counter Status |
| --- | --- | --- | --- |
| counter\_code | int |  | I am guessing is Serial Number of the person counting |
| reading\_remarque | int | [0,9] |  |
| counter\_coefficient | int |  | Counts the number of times it has overflow |
| total\_consumption | int |  | Sum of the 4 consumption level |
| old\_index | int |  |  |
| new\_index | int |  |  |
| difference\_index | int |  |  |
| tally\_check | boolean |  | Check if new\_index - old\_index - total\_consumption == 0 |
| tally\_value | int |  |  |
| month | int |  | Month of the Year |
| counter\_type | string | ‘ELEC’ or ‘GAS’ |  |