**Alert 1:**

Run our program every *1 minute* since data for each device is coming every one minute and calculate the current status of each device and update it in csv file every one minute.

New Files created to refrainfrom **precomputation on the whole data set:**

**current\_alert\_1\_status.csv:** This file store the current/ live alert status of every device and is updated every 1 minute.

|  |  |  |
| --- | --- | --- |
| house\_id | household\_id | alert\_status |
| 0 | 0 | 0 |
| 0 | 2 | 0 |
| 1 | 1 | 1 |

**alert\_1\_previous\_record.csv:**

Stores mean and standard deviation count for each hour, house\_id, household\_id

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| house\_id | household\_id | record\_hour | mean | standard\_deviation | count |
| 0 | 0 | 22 | 0 | 0 | 0 |
| 0 | 0 | 23 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 0 | 2 | 0 | 0 | 0 |
| 0 | 0 | 3 | 0 | 0 | 0 |
| 0 | 0 | 4 | 0 | 0 | 0 |
| 0 | 0 | 5 | 0 | 0 | 0 |
| 0 | 0 | 6 | 0 | 0 | 0 |
| 0 | 0 | 7 | 0 | 0 | 0 |
| 0 | 0 | 8 | 0 | 0 | 0 |
| 0 | 0 | 9 | 0 | 0 | 0 |
| 0 | 0 | 10 | 0 | 0 | 0 |
| 0 | 0 | 11 | 0 | 0 | 0 |
| 0 | 0 | 12 | 0 | 0 | 0 |
| 0 | 0 | 13 | 0 | 0 | 0 |
| 0 | 0 | 14 | 0 | 0 | 0 |
| 0 | 0 | 15 | 0 | 0 | 0 |
| 0 | 0 | 16 | 0 | 0 | 0 |
| 0 | 0 | 17 | 0 | 0 | 0 |
| 0 | 0 | 18 | 0 | 0 | 0 |
| 0 | 0 | 19 | 0 | 0 | 0 |
| 0 | 0 | 20 | 0 | 0 | 0 |
| 0 | 0 | 21 | 0 | 0 | 0 |

**Data Cleaning:**

* **Consumption value as 0:** if consumption\_value is zero, then drop that row since consumption value can’t be zero and also my score increases from **0.90856** to **0.98898**. Hence dropping zero consumption value is good choice.
* **Consumption value as empty:** if consumption\_value is empty, then drop that row since consumption value can’t be empty.
* **Consumption value is too large:**

Consumption\_value > mean + 3 \* standard\_deviation

In above case also drop that row since value is too large. It is outliers.

**Imputation Method Used:**

If Consumption\_value > mean + standard\_deviation then status of device is 1 otherwise 0.

Now, we have to *update mean and standard deviation:*

From file alert\_1\_previous\_record, we get mean, standard deviation upto last record. We can name them as

**Mold**: mean up to last record

**Sold:** standard deviation up to last record

**Cold:** count up to last record

Similarly, **Mnew, Snew, Cnew** are updated values after new consumption value **Xnew**.

And these values in alert\_1\_previous\_record.csv every *1 minute.*