**GSTN Data Analysis**

**Flow Of Postgres Model**

**CITY WISE PROCESS**

**RAW DATA TABLE STRUCTURE**

|  |  |  |
| --- | --- | --- |
| **Table Name Format:** “raw\_”+city\_name\_+”state\_code” | | |
| **Column Name** | **Type** | **Description** |
| raw\_name | text | raw area name |
| Pincode | integer | Area name pincode |

**MASTER DB TABLE STRUCTURE**

|  |  |  |
| --- | --- | --- |
| **Table Name Format:** city\_name+”\_ROAD\_NETWORK” | | |
| **Column Name** | **Type** | **Description** |
| ROAD\_NME | text | Road name |
| ROAD\_ALT | text | Alternative road name |
| ROAD\_BSE | text | Road base name |

**OUTPUT TABLE STRUCTURE**

|  |  |  |
| --- | --- | --- |
| **Table Name Format:** ”gstn\_output\_”+city\_name\_+”state\_code” | | |
| **Column Name** | **Type** | **Description** |
| srno | integer | Serial number |
| match | integer | Match code |
| MAIN\_STREET\_NAM | text | Raw street name |
| row\_id | integer | Row id by split road |
| STREET\_NAM | text | Filter street name by using of main street name |
| raw\_street\_nme | text | Match street name |
| ADMIN\_ROAD\_NME | text | Admin road name |
| ADMIN\_ALT\_NME | text | Admin alternative road name |
| ADMIN\_BSE\_NME | text | Admin road base name |
| M\_LOC\_NME | text | Master Locality Name |
| M\_LOC\_ID | integer | Master Locality ID |
| SUBL\_NME | character varying | Sub locality name |
| SUBL\_ID | integer | Sub locality id |
| SSLC\_NME | character varying | Sub sub locality name |
| SSLC\_ID | integer | Sub sub locality id |
| fuzzy\_ratio | integer | Fuzzy match ratio |
| Unmatch\_String | text | Remaining raw area name after all match |
| Match\_String | text | Matched raw area name |
| status | text | Match Status |
| N\_LOC | text | Near by locality name |
| N\_LOC\_MATCHED | text | Near by locality matched |
| N\_LOC\_MATCHED\_ID | integer | Near by locality matched id |
| N\_SUBL\_MATCHED | text | Near by sub locality matched |
| N\_SUBL\_MATCHED\_ID | integer | Near by sub locality matched id |
| N\_SSLC\_MATCHED | text | Near by sub sub locality matched |
| N\_SSLC-MATCHED\_ID | integer | Near by sub sub locality matched id |
| M\_N\_LOC | text | Main near by locality |
| M\_N\_LOC\_MATCHED\_ID | integer | Main near by locality matched id |

**GSTN DATA ANALYSIS ON STREET**

**Steps For New Approach**

**Step1:** find exact match with road name(direct match).

**Step2:** find exact match with road alt name(direct match).

**Step3:**Split by comma then match with road name.

**Step4:** Split by comma then match with road alt name.

**Step5:** Remove Road type words in raw string then match with road bse name.

**Step6:** Remove Road type words in raw string then match with road bse name by using of fuzzy match.

**Step7:** Add comma in the raw street name where find the road,gali,street,marg,….

this type of word.

**Step8:** then match with (Split by comma then match with road name).

**Step9:** then match with (Split by comma then match with road alt name).

**Step10:** then remaining string match with half split part for road name and road alt name by using of fuzzy match

**Step11:** another remaining string match by using of fuzzy match (split string in two parts (1,2),(2,3)).

**Step12:**unmatch string match with road name and road alt name by using of fuzzy match

**Step13:**Find not match road.

**Flow diagram of GSTN Street data.**

