Data_Wrangling

May 16, 2018

0.1 WHAT IS DATA WRANGLING OR DATA MUNGING?

DATA WRANGLING is the art of dealing with and or converting missing or ill-formatted data into a format that more easily lends itself to analysis.

BUT, HOW TO ORGANIZE MESSI DATA?

FIRST, TRY TO UNDERSTAND THE STRUCTURE OF THE DATA ITSELF.

BUT, before we can make this analysis, we need first to get the data we want to analyze. Three of the most common sources to get the data are from FILES, DATABASES AND WEBSITES THROUGH WEB APIs.

ACQUIRING DATA Acquiring data is not a fancy thing. It is just a matter of finding the right file in the internet and downloading it. The most of the data are stored in text files, particularly on government websites. Normally, files can come in one of these formats.

COMMON DATA FORMATS CSV ---------- XML ---------- JSON

Note: it is recommended to take the course Data Wrangling with mongoDB. CSV: it has a series of rows. Each row corresponding to an entry. The first row of every file tells us what each entry means. The values of the first row can be called as the identifiers.

XML: the benefit in comparison to CSV is that it supports nested structures. A document element can have a series of tables, which can have a number of children (every row/column).

```
    <Column1> Value <Column1/>
    <Column2> Value <Column2/>
    <Column3/> #Value missing
<Table/>
```

JSON: the benefit in comparison to CSV is that it supports nested structures. Json documents have a number of json objects. It looks like a python dictionary. We have keys, which corresponds to what we see in the header row in a CSV file, followed by values.

{ 'Key1': value1, 'Key2': value2, 'Key3': value3, }

Note: data format is not a matter of the file extension, being .csv or .json. A file format really has to do with how the data is organized inside the file. We could have data that is formatted in json or csv, but that comes in a file whose extension is .txt

SOURCES OF DATA FILE ---------- DATABASES ---------- WEB APIs

Once we ACQUIRE the data, it is always good to make a SANITY CHECKING because our data could have bad or missing values.

What is SANITY CHECKING? With sanity checking we answer questions like, does our data make sense? Is there a problem? Does the data look like I expected to?

How to do it? you can draw plots to visualize the data or run some analysis somehow. But pandas has a method for that, the method describe.

Note: it is recommended to take udacitys course: exploratoy data and analysis course.

```
In [1]: import pandas as pd
        from pandasql import sqldf
        import numpy as np
        import csv
```

HANDLING CSV FILES

GETTING DATA FROM CSV FILES

```
In [2]: def add_full_name(full_path, new_full_path):
            master.to_csv(new_full_path)
```

```
In [3]: if __name__ == "__main__":
            file_path="/home/camilo/Documents/Data_Science_Courses/Intro_to_Data_Science_Udacity
            new_file_path='/home/camilo/Documents/Data_Science_Courses/Intro_to_Data_Science_Uda
            master=pd.read_csv(file_path)
            master['nameFull']=master['nameFirst'] + ' ' + master['nameLast']
            add_full_name(file_path, new_file_path)
```

```
/usr/local/lib/python2.7/dist-packages/pandas/core/computation/check.py:17: UserWarning: The ins
The minimum supported version is 2.4.6

  ver=ver, min_ver=_MIN_NUMEXPR_VERSION), UserWarning)
```

```
In [4]: master.head(5)
```

```
Out[4]:     playerID  birthYear  birthMonth  birthDay birthCountry birthState  \
        0  aardsda01     1981.0        12.0      27.0          USA         CO
        1  aaronha01     1934.0         2.0       5.0          USA         AL
        2  aaronto01     1939.0         8.0       5.0          USA         AL
        3   aasedo01     1954.0         9.0       8.0          USA         CA
        4   abadan01     1972.0         8.0      25.0          USA         FL

            birthCity  deathYear  deathMonth  deathDay      ...            nameGiven  \
        0      Denver        NaN         NaN       NaN      ...          David Allan
        1      Mobile        NaN         NaN       NaN      ...          Henry Louis
```

```
2      Mobile     1984.0         8.0      16.0      ...            Tommie Lee
3      Orange        NaN         NaN       NaN      ...         Donald William
4  Palm Beach        NaN         NaN       NaN      ...          Fausto Andres

  weight height bats throws       debut   finalGame   retroID    bbrefID  \
0  215.0   75.0    R      R  2004-04-06  2015-08-23  aardd001  aardsda01
1  180.0   72.0    R      R  1954-04-13  1976-10-03  aaroh101  aaronha01
2  190.0   75.0    R      R  1962-04-10  1971-09-26  aarot101  aaronto01
3  190.0   75.0    R      R  1977-07-26  1990-10-03  aased001   aasedo01
4  184.0   73.0    L      L  2001-09-10  2006-04-13  abada001   abadan01

        nameFull
0  David Aardsma
1     Hank Aaron
2   Tommie Aaron
3       Don Aase
4      Andy Abad

[5 rows x 25 columns]
```

WANGLING DATA USING SQL (RELATIONAL DATABASES)

Commands:

1 select everything from the aadhaar data

SELECT * FROM Aadhaar_data;

2 select the first 20 lines of the data

SELECT * FROM Aadhaar_data LIMIT 20;

3 select columns district and subdisctrict from the data

SELECT district, subdistrict FROM Aadhaar_data LIMIT 20;

```
In [5]: data = pd.read_csv('/home/camilo/Documents/Data_Science_Courses/Intro_to_Data_Science_Ud

In [6]: data.rename(columns = lambda x: x.replace(' ', '_').lower(), inplace=True)

In [7]: data.head(10)

Out[7]:         registrar             enrolment_agency            state    district  \
        0  Allahabad Bank           Tera Software Ltd        Jharkhand      Ranchi
        1  Allahabad Bank           Tera Software Ltd        Jharkhand      Ranchi
        2  Allahabad Bank  Vakrangee Softwares Limited          Gujarat       Surat
```

```
3  Allahabad Bank  Vakrangee Softwares Limited  Himachal Pradesh      Kangra
4  Allahabad Bank  Vakrangee Softwares Limited    Madhya Pradesh  Chhindwara
5  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra  Ahmadnagar
6  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra  Ahmadnagar
7  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra  Ahmadnagar
8  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra  Ahmadnagar
9  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra  Ahmadnagar

  sub_district pin_code gender  age  aadhaar_generated  enrolment_rejected  \
0       Namkum   834003      M   63                 0                   1
1       Ranchi   834004      F   36                 0                   1
2        Nizar   394380      M   10                 1                   0
3     Baijnath   176081      M   44                 1                   0
4    Pandhurna   480334      M   35                 1                   0
5       Nevasa   414105      M   28                 1                   0
6        Rahta   413711      M   31                 0                   1
7        Rahta   423107      F    9                 0                   1
8        Rahta   423107      F   35                 0                   1
9        Rahta   423107      M   33                 0                   1

   residents_providing_email  residents_providing_mobile_number
0                          0                                  1
1                          0                                  1
2                          0                                  0
3                          1                                  1
4                          0                                  0
5                          0                                  0
6                          0                                  0
7                          0                                  0
8                          0                                  0
9                          0                                  1
```

```
In [8]: sqldf('SELECT * FROM data LIMIT 50;', globals())
```

```
Out[8]:          registrar             enrolment_agency             state  \
0   Allahabad Bank            Tera Software Ltd         Jharkhand
1   Allahabad Bank            Tera Software Ltd         Jharkhand
2   Allahabad Bank  Vakrangee Softwares Limited           Gujarat
3   Allahabad Bank  Vakrangee Softwares Limited  Himachal Pradesh
4   Allahabad Bank  Vakrangee Softwares Limited    Madhya Pradesh
5   Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
6   Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
7   Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
8   Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
9   Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
10  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
11  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
12  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
```

```
13  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
14  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
15  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
16  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
17  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
18  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
19  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
20  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
21  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
22  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
23  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
24  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
25  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
26  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
27  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
28  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
29  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
30  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
31  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
32  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
33  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
34  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
35  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
36  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
37  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
38  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
39  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
40  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
41  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
42  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
43  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
44  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
45  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
46  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
47  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
48  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra
49  Allahabad Bank  Vakrangee Softwares Limited       Maharashtra

       district sub_district pin_code gender  age  aadhaar_generated  \
0        Ranchi       Namkum   834003      M   63                  0
1        Ranchi       Ranchi   834004      F   36                  0
2         Surat        Nizar   394380      M   10                  1
3        Kangra     Baijnath   176081      M   44                  1
4    Chhindwara    Pandhurna   480334      M   35                  1
5    Ahmadnagar       Nevasa   414105      M   28                  1
6    Ahmadnagar        Rahta   413711      M   31                  0
7    Ahmadnagar        Rahta   423107      F    9                  0
8    Ahmadnagar        Rahta   423107      F   35                  0
```

```
9    Ahmadnagar        Rahta   423107      M   33                0
10   Ahmadnagar        Rahta   423107      M   56                0
11   Ahmadnagar        Rahta   423107      M   64                0
12   Ahmadnagar       Rahuri   413706      F   55                0
13   Ahmadnagar       Rahuri   413715      F   31                0
14   Ahmadnagar       Rahuri   413715      F   35                0
15   Ahmadnagar       Rahuri   413715      F   49                0
16   Ahmadnagar       Rahuri   413715      F   65                0
17   Ahmadnagar       Rahuri   413715      M   19                0
18   Ahmadnagar       Rahuri   413715      M   22                0
19   Ahmadnagar       Rahuri   413715      M   47                0
20   Ahmadnagar       Rahuri   413715      M   65                0
21   Ahmadnagar    Sangamner   413714      F   28                0
22  Ahmed Nagar       Karjat   414402      F   19                0
23  Ahmed Nagar    Kopargaon   413708      F   10                0
24  Ahmed Nagar    Kopargaon   413708      F   14                0
25  Ahmed Nagar    Kopargaon   413708      F   15                0
26  Ahmed Nagar    Kopargaon   413708      F   16                0
27  Ahmed Nagar    Kopargaon   413708      F   17                0
28  Ahmed Nagar    Kopargaon   413708      F   20                0
29  Ahmed Nagar    Kopargaon   413708      F   23                0
30  Ahmed Nagar    Kopargaon   413708      F   28                0
31  Ahmed Nagar    Kopargaon   413708      F   30                0
32  Ahmed Nagar    Kopargaon   413708      F   37                0
33  Ahmed Nagar    Kopargaon   413708      F   39                0
34  Ahmed Nagar    Kopargaon   413708      F   45                0
35  Ahmed Nagar    Kopargaon   413708      F   68                0
36  Ahmed Nagar    Kopargaon   413708      M   12                0
37  Ahmed Nagar    Kopargaon   413708      M   13                0
38  Ahmed Nagar    Kopargaon   413708      M   15                0
39  Ahmed Nagar    Kopargaon   413708      M   17                0
40  Ahmed Nagar    Kopargaon   413708      M   20                0
41  Ahmed Nagar    Kopargaon   413708      M   22                0
42  Ahmed Nagar    Kopargaon   413708      M   23                0
43  Ahmed Nagar    Kopargaon   413708      M   27                0
44  Ahmed Nagar    Kopargaon   413708      M   38                0
45  Ahmed Nagar    Kopargaon   413708      M   40                0
46  Ahmed Nagar    Kopargaon   413708      M   43                0
47  Ahmed Nagar    Kopargaon   413708      M   44                0
48  Ahmed Nagar    Kopargaon   413708      M   46                0
49  Ahmed Nagar    Kopargaon   413708      M   53                0

    enrolment_rejected  residents_providing_email  \
0                    1                          0
1                    1                          0
2                    0                          0
3                    0                          1
4                    0                          0
```

5	0	0
6	1	0
7	1	0
8	1	0
9	1	0
10	1	0
11	1	0
12	1	0
13	1	0
14	1	0
15	1	0
16	1	0
17	1	0
18	1	0
19	1	0
20	1	0
21	1	0
22	1	0
23	1	0
24	1	0
25	1	0
26	1	0
27	1	0
28	1	0
29	1	0
30	2	0
31	1	0
32	1	0
33	1	0
34	1	0
35	1	0
36	1	0
37	3	0
38	1	0
39	1	0
40	1	0
41	1	1
42	1	0
43	1	0
44	1	0
45	1	0
46	1	0
47	1	0
48	1	0
49	1	0

	residents_providing_mobile_number
0	1

1	1
2	0
3	1
4	0
5	0
6	0
7	0
8	0
9	1
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	1
22	0
23	1
24	1
25	0
26	0
27	0
28	0
29	0
30	1
31	0
32	0
33	0
34	1
35	0
36	0
37	0
38	1
39	1
40	0
41	1
42	0
43	1
44	0
45	0
46	1
47	0
48	0

```
        49                                  0
```

```
In [9]: sqldf('SELECT registrar, enrolment_agency FROM data LIMIT 50', globals())
```

Out[9]:

	registrar	enrolment_agency
0	Allahabad Bank	Tera Software Ltd
1	Allahabad Bank	Tera Software Ltd
2	Allahabad Bank	Vakrangee Softwares Limited
3	Allahabad Bank	Vakrangee Softwares Limited
4	Allahabad Bank	Vakrangee Softwares Limited
5	Allahabad Bank	Vakrangee Softwares Limited
6	Allahabad Bank	Vakrangee Softwares Limited
7	Allahabad Bank	Vakrangee Softwares Limited
8	Allahabad Bank	Vakrangee Softwares Limited
9	Allahabad Bank	Vakrangee Softwares Limited
10	Allahabad Bank	Vakrangee Softwares Limited
11	Allahabad Bank	Vakrangee Softwares Limited
12	Allahabad Bank	Vakrangee Softwares Limited
13	Allahabad Bank	Vakrangee Softwares Limited
14	Allahabad Bank	Vakrangee Softwares Limited
15	Allahabad Bank	Vakrangee Softwares Limited
16	Allahabad Bank	Vakrangee Softwares Limited
17	Allahabad Bank	Vakrangee Softwares Limited
18	Allahabad Bank	Vakrangee Softwares Limited
19	Allahabad Bank	Vakrangee Softwares Limited
20	Allahabad Bank	Vakrangee Softwares Limited
21	Allahabad Bank	Vakrangee Softwares Limited
22	Allahabad Bank	Vakrangee Softwares Limited
23	Allahabad Bank	Vakrangee Softwares Limited
24	Allahabad Bank	Vakrangee Softwares Limited
25	Allahabad Bank	Vakrangee Softwares Limited
26	Allahabad Bank	Vakrangee Softwares Limited
27	Allahabad Bank	Vakrangee Softwares Limited
28	Allahabad Bank	Vakrangee Softwares Limited
29	Allahabad Bank	Vakrangee Softwares Limited
30	Allahabad Bank	Vakrangee Softwares Limited
31	Allahabad Bank	Vakrangee Softwares Limited
32	Allahabad Bank	Vakrangee Softwares Limited
33	Allahabad Bank	Vakrangee Softwares Limited
34	Allahabad Bank	Vakrangee Softwares Limited
35	Allahabad Bank	Vakrangee Softwares Limited
36	Allahabad Bank	Vakrangee Softwares Limited
37	Allahabad Bank	Vakrangee Softwares Limited
38	Allahabad Bank	Vakrangee Softwares Limited
39	Allahabad Bank	Vakrangee Softwares Limited
40	Allahabad Bank	Vakrangee Softwares Limited
41	Allahabad Bank	Vakrangee Softwares Limited
42	Allahabad Bank	Vakrangee Softwares Limited

```
43  Allahabad Bank  Vakrangee Softwares Limited
44  Allahabad Bank  Vakrangee Softwares Limited
45  Allahabad Bank  Vakrangee Softwares Limited
46  Allahabad Bank  Vakrangee Softwares Limited
47  Allahabad Bank  Vakrangee Softwares Limited
48  Allahabad Bank  Vakrangee Softwares Limited
49  Allahabad Bank  Vakrangee Softwares Limited
```

```
In [10]: sqldf("SELECT * FROM data WHERE state='Gujarat'", globals())
```

```
Out[10]:                                registrar  \
         0                           Allahabad Bank
         1                            Bank Of India
         2                            Bank Of India
         3                            Bank Of India
         4                            Bank Of India
         5                            Bank Of India
         6                            Bank Of India
         7                            Bank Of India
         8                            Bank Of India
         9                            Bank Of India
         10                           Bank Of India
         11                           Bank Of India
         12                           Bank Of India
         13                           Bank Of India
         14                           Bank Of India
         15                           Bank Of India
         16                           Bank Of India
         17                           Bank Of India
         18                           Bank Of India
         19                           Bank Of India
         20                           Bank Of India
         21                           Bank Of India
         22                           Bank Of India
         23                           Bank Of India
         24                           Bank Of India
         25                           Bank Of India
         26                           Bank Of India
         27                           Bank Of India
         28                           Bank Of India
         29                           Bank Of India
         ..                                     ...
         55                     Bank of Maharashtra
         56                     Bank of Maharashtra
         57                     Bank of Maharashtra
         58                     Bank of Maharashtra
         59                     Bank of Maharashtra
         60                     Bank of Maharashtra
```

61	Bank of Maharashtra
62	Bank of Maharashtra
63	Bank of Maharashtra
64	Bank of Maharashtra
65	Bank of Maharashtra
66	Bank of Maharashtra
67	Bank of Maharashtra
68	Bank of Maharashtra
69	Bank of Maharashtra
70	Bank of Maharashtra
71	Bank of Maharashtra
72	CSC e-Governance Services India Limited
73	CSC e-Governance Services India Limited
74	CSC e-Governance Services India Limited
75	CSC e-Governance Services India Limited
76	CSC e-Governance Services India Limited
77	Central Bank of India
78	Central Bank of India
79	Central Bank of India
80	Central Bank of India
81	Central Bank of India
82	Central Bank of India
83	Central Bank of India
84	Central Bank of India

	enrolment_agency	state	district	\
0	Vakrangee Softwares Limited	Gujarat	Surat	
1	CHESSY CONSULTANTS PVT LTD	Gujarat	Ahmedabad	
2	CHESSY CONSULTANTS PVT LTD	Gujarat	Amreli	
3	CHESSY CONSULTANTS PVT LTD	Gujarat	Amreli	
4	CHESSY CONSULTANTS PVT LTD	Gujarat	Junagadh	
5	CHESSY CONSULTANTS PVT LTD	Gujarat	Kachchh	
6	CHESSY CONSULTANTS PVT LTD	Gujarat	Kachchh	
7	Frontech Systems Pvt Ltd	Gujarat	Junagadh	
8	Frontech Systems Pvt Ltd	Gujarat	Surendra Nagar	
9	MANTRA SOFTTECH (INDIA) PVTLTD	Gujarat	Ahmedabad	
10	MANTRA SOFTTECH (INDIA) PVTLTD	Gujarat	Ahmedabad	
11	MANTRA SOFTTECH (INDIA) PVTLTD	Gujarat	Ahmedabad	
12	MANTRA SOFTTECH (INDIA) PVTLTD	Gujarat	Amreli	
13	Matrix Processing House	Gujarat	Ahmedabad	
14	Matrix Processing House	Gujarat	Jamnagar	
15	Matrix Processing House	Gujarat	Jamnagar	
16	Matrix Processing House	Gujarat	Jamnagar	
17	Matrix Processing House	Gujarat	Jamnagar	
18	Matrix Processing House	Gujarat	Jamnagar	
19	Matrix Processing House	Gujarat	Jamnagar	
20	Matrix Processing House	Gujarat	Jamnagar	
21	Matrix Processing House	Gujarat	Jamnagar	

```
22                Matrix Processing House  Gujarat         Kachchh
23                Matrix Processing House  Gujarat         Kachchh
24                Matrix Processing House  Gujarat         Kachchh
25                Matrix Processing House  Gujarat          Rajkot
26                Matrix Processing House  Gujarat          Rajkot
27                Matrix Processing House  Gujarat          Rajkot
28                Matrix Processing House  Gujarat  Surendra Nagar
29                Matrix Processing House  Gujarat  Surendra Nagar
..                                    ...      ...             ...
55                     Alankit Finsec Ltd  Gujarat          Valsad
56                     Alankit Finsec Ltd  Gujarat          Valsad
57                     Alankit Finsec Ltd  Gujarat          Valsad
58                     Alankit Finsec Ltd  Gujarat          Valsad
59                     Alankit Finsec Ltd  Gujarat          Valsad
60                     Alankit Finsec Ltd  Gujarat          Valsad
61                     Alankit Finsec Ltd  Gujarat          Valsad
62                     Alankit Finsec Ltd  Gujarat          Valsad
63                     Alankit Finsec Ltd  Gujarat          Valsad
64                     Alankit Finsec Ltd  Gujarat          Valsad
65                     Alankit Finsec Ltd  Gujarat          Valsad
66                     Alankit Finsec Ltd  Gujarat          Valsad
67                     Alankit Finsec Ltd  Gujarat          Valsad
68                     Alankit Finsec Ltd  Gujarat          Valsad
69           Micro Technologies India Ltd  Gujarat     Banaskantha
70           Micro Technologies India Ltd  Gujarat          Valsad
71           Micro Technologies India Ltd  Gujarat          Valsad
72  A I Soc for Electronics and Comp Tech  Gujarat       Ahmedabad
73  A I Soc for Electronics and Comp Tech  Gujarat         Bharuch
74  A I Soc for Electronics and Comp Tech  Gujarat        Vadodara
75                       CMS Computers Ltd  Gujarat          Valsad
76             Vakrangee Softwares Limited  Gujarat       Ahmedabad
77           CALANCE SOFTWARE PRIVATE LTD  Gujarat        Vadodara
78           CALANCE SOFTWARE PRIVATE LTD  Gujarat        Vadodara
79           CALANCE SOFTWARE PRIVATE LTD  Gujarat        Vadodara
80                   IAP COMPANY Pvt. Ltd  Gujarat     Banaskantha
81                   IAP COMPANY Pvt. Ltd  Gujarat     Banaskantha
82                   IAP COMPANY Pvt. Ltd  Gujarat     Banaskantha
83                   IAP COMPANY Pvt. Ltd  Gujarat         Kachchh
84                   IAP COMPANY Pvt. Ltd  Gujarat           Patan

      sub_district pin_code gender  age  aadhaar_generated  \
0            Nizar   394380      M   10                  1
1   Ahmadabad City   380050      F   61                  1
2           Rajula   365560      F   34                  1
3           Rajula   365560      M   38                  1
4    Patan Veraval   362268      F   35                  1
5            Rapar   370145      M    9                  1
6            Rapar   370165      M   20                  1
```

7	Kodinar	362720	M	25	1
8	Sayla	363430	F	27	1
9	Ahmadabad City	380004	M	7	1
10	Ahmadabad City	380061	F	23	0
11	Daskroi	380059	M	48	1
12	Amreli	365601	F	44	1
13	Ahmadabad City	382475	M	45	1
14	Dhrol	361001	F	26	1
15	Dhrol	361130	F	26	1
16	Dhrol	361130	M	29	1
17	Jamjodhpur	360515	F	26	1
18	Jodiya	361011	F	5	1
19	Jodiya	361011	F	27	1
20	Jodiya	361011	M	29	1
21	Kalyanpur	361320	M	21	1
22	Mandvi	370455	M	20	1
23	Mundra	370410	M	6	1
24	Mundra	370410	M	34	1
25	Morvi	363642	F	26	1
26	Morvi	363642	M	33	1
27	Paddhari	360110	M	31	1
28	Chotila	363520	M	29	1
29	Limbdi	363421	F	32	1
..	...	...	...	...	...
55	Pardi	396191	M	20	0
56	Pardi	396191	M	23	1
57	Pardi	396191	M	26	1
58	Pardi	396191	M	31	1
59	Pardi	396191	M	35	1
60	Pardi	396193	M	8	1
61	Pardi	396193	M	11	1
62	Pardi	396193	M	50	1
63	Pardi	396195	F	37	1
64	Pardi	396195	M	11	1
65	Pardi	396195	M	16	1
66	Pardi	396195	M	21	1
67	Pardi	396195	M	27	1
68	Umbergaon	396155	F	31	1
69	Danta	385120	F	25	1
70	Pardi	396191	M	56	1
71	Pardi	396191	M	62	1
72	Ahmadabad City	380016	M	29	1
73	Anklesvar	392011	M	59	1
74	Padra	390016	M	29	1
75	Pardi	396195	M	22	1
76	Ahmadabad City	380015	F	50	1
77	Padra	390012	F	9	1
78	Padra	390012	F	36	1

```
79           Padra   390012      M   12                1
80        Palanpur   385001      F    8                1
81        Palanpur   385001      F   13                1
82          Vadgam   385210      M   30                1
83           Rapar   370165      M   26                1
84         Sidhpur   384151      F   12                1

    enrolment_rejected  residents_providing_email  \
0                    0                          0
1                    0                          0
2                    0                          0
3                    0                          0
4                    0                          0
5                    0                          0
6                    0                          0
7                    0                          0
8                    0                          0
9                    0                          0
10                   1                          0
11                   0                          0
12                   0                          0
13                   0                          0
14                   0                          0
15                   0                          0
16                   0                          0
17                   0                          0
18                   0                          0
19                   0                          0
20                   0                          0
21                   0                          0
22                   0                          0
23                   0                          0
24                   0                          0
25                   0                          0
26                   0                          0
27                   0                          0
28                   0                          0
29                   0                          0
..                 ...                        ...
55                   1                          0
56                   0                          0
57                   0                          0
58                   0                          0
59                   0                          0
60                   0                          0
61                   0                          0
62                   0                          0
63                   0                          0
```

64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0

	residents_providing_mobile_number
0	0
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	0
12	1
13	1
14	1
15	1
16	1
17	1
18	0
19	1
20	1
21	1
22	1
23	1
24	1

```
25                                  1
26                                  1
27                                  1
28                                  1
29                                  1
..                                ...
55                                  1
56                                  1
57                                  1
58                                  1
59                                  1
60                                  1
61                                  1
62                                  1
63                                  1
64                                  1
65                                  1
66                                  0
67                                  1
68                                  1
69                                  1
70                                  1
71                                  1
72                                  1
73                                  1
74                                  1
75                                  1
76                                  1
77                                  1
78                                  1
79                                  1
80                                  1
81                                  1
82                                  1
83                                  1
84                                  1

[85 rows x 12 columns]
```

```
In [11]: sqldf("SELECT district, sum(aadhaar_generated) FROM data GROUP BY district", globals())
```

Out[11]:

	district	sum(aadhaar_generated)
0	Adilabad	1
1	Agra	9
2	Ahmadnagar	552
3	Ahmed Nagar	0
4	Ahmedabad	9
5	Ajmer	527

6	Akola	55
7	Alappuzha	1
8	Aligarh	7
9	Alirajpur	2965
10	Allahabad	137
11	Almora	8
12	Alwar	303
13	Ambala	70
14	Ambedkar Nagar	13
15	Amethi	1
16	Amravati	11
17	Amreli	4
18	Amritsar	269
19	Amroha	7
20	Ananthapur	5
21	Angul	5
22	Anuppur	1938
23	Araria	3
24	Arwal	6
25	Ashok Nagar	72
26	Auraiya	10
27	Aurangabad	533
28	Azamgarh	16
29	Bagalkot	33
..	...	...
498	Ujjain	913
499	Umaria	1992
500	Una	56
501	Unakoti	6
502	Unnao	2
503	Uttar Dinajpur	2
504	Uttara Kannada	10
505	Vadodara	4
506	Vaishali	16
507	Valsad	30
508	Varanasi	13
509	Vellore	8
510	Vidisha	887
511	Viluppuram	85
512	Virudhunagar	5
513	Visakhapatnam	6
514	Vizianagaram	5
515	Warangal	4
516	Wardha	8
517	Washim	282
518	West Champaran	9
519	West Delhi	157
520	West Garo Hills	1

```
521    West Godavari                     4
522       West Siang                     1
523   West Singhbhum                  4170
524     West Tripura                   392
525           Yadgir                    69
526     Yamuna Nagar                  1586
527          Yavatmal                  469

[528 rows x 2 columns]
```

```
In [12]: sqldf("SELECT district, sub_district, sum(aadhaar_generated) FROM data GROUP BY distric
```

Out[12]:

	district	sub_district	sum(aadhaar_generated)
0	Adilabad	Lokeswaram	1
1	Agra	Bah	3
2	Agra	Etmadpur	1
3	Agra	Fatehabad	1
4	Agra	Kheragarh	2
5	Agra	Kiraoli	2
6	Ahmadnagar	Akola	3
7	Ahmadnagar	Jamkhed	2
8	Ahmadnagar	Karjat	3
9	Ahmadnagar	Kopargaon	4
10	Ahmadnagar	Nagar	4
11	Ahmadnagar	Nevasa	139
12	Ahmadnagar	Parner	5
13	Ahmadnagar	Pathardi	114
14	Ahmadnagar	Rahta	1
15	Ahmadnagar	Rahuri	123
16	Ahmadnagar	Sangamner	2
17	Ahmadnagar	Shevgaon	120
18	Ahmadnagar	Shrigonda	32
19	Ahmed Nagar	Karjat	0
20	Ahmed Nagar	Kopargaon	0
21	Ahmed Nagar	Nagar	0
22	Ahmed Nagar	Pathardi	0
23	Ahmed Nagar	Rahta	0
24	Ahmed Nagar	Rahuri	0
25	Ahmedabad	Ahmadabad City	8
26	Ahmedabad	Daskroi	1
27	Ajmer	Ajmer	1
28	Ajmer	Beawar	29
29	Ajmer	Bhinay	19
...	...	...	...
2440	West Singhbhum	Majhgaon	289
2441	West Singhbhum	Manjhari	1
2442	West Singhbhum	Manoharpur	227
2443	West Singhbhum	Noamundi (barajamda)	827

```
2444  West Singhbhum              Sonua                    109
2445  West Singhbhum           Tantnagar                     7
2446  West Singhbhum              Tonto                    406
2447    West Tripura       Agartala Sadar                    112
2448    West Tripura              Dukli                     24
2449    West Tripura            Hezamara                    42
2450    West Tripura             Jirania                   101
2451    West Tripura              Mandai                     44
2452    West Tripura            Mohanpur                     69
2453          Yadgir             Shahpur                     42
2454          Yadgir            Shorapur                     25
2455          Yadgir              Yadgir                      2
2456   Yamuna Nagar         Chhachhrauli                    148
2457   Yamuna Nagar             Jagadhri                   1438
2458       Yavatmal                 Arni                     13
2459       Yavatmal            Babulgaon                     49
2460       Yavatmal              Digras                      4
2461       Yavatmal             Ghatanji                     43
2462       Yavatmal               Kalamb                    189
2463       Yavatmal              Kelapur                      9
2464       Yavatmal             Mahagaon                     36
2465       Yavatmal                Pusad                     33
2466       Yavatmal             Ralegaon                     24
2467       Yavatmal             Umarkhed                     34
2468       Yavatmal             Yavatmal                     34
2469       Yavatmal          Zari-Jamani                      1

[2470 rows x 3 columns]
```

```
In [13]: sqldf("SELECT district, sub_district, sum(aadhaar_generated) FROM data WHERE age>60 GRO
```

Out[13]:

	district	sub_district	sum(aadhaar_generated)
0	Ahmadnagar	Nagar	0
1	Ahmadnagar	Nevasa	18
2	Ahmadnagar	Parner	1
3	Ahmadnagar	Pathardi	13
4	Ahmadnagar	Rahta	0
5	Ahmadnagar	Rahuri	5
6	Ahmadnagar	Shevgaon	23
7	Ahmadnagar	Shrigonda	3
8	Ahmed Nagar	Kopargaon	0
9	Ahmed Nagar	Nagar	0
10	Ahmedabad	Ahmadabad City	1
11	Ajmer	Beawar	3
12	Ajmer	Bhinay	1
13	Ajmer	Kekri	17
14	Ajmer	Kishangarh	3
15	Ajmer	Peesangan	1

16	Ajmer	Sarwar	16
17	Akola	Balapur	2
18	Akola	Telhara	1
19	Alirajpur	Bhavra	32
20	Alirajpur	Jobat	16
21	Allahabad	Allahabad	9
22	Allahabad	Bara	1
23	Allahabad	Koraon	1
24	Alwar	Alwar	1
25	Alwar	Behror	5
26	Alwar	Kathumar	1
27	Alwar	Kishangarh Bas	6
28	Alwar	Kotkasim	1
29	Alwar	Rajgarh	0
..	...	...	...
879	West Singhbhum	Chakradharpur	15
880	West Singhbhum	Goilkera	8
881	West Singhbhum	Jagannathpur	26
882	West Singhbhum	Jhinkpani	2
883	West Singhbhum	Khuntpani	13
884	West Singhbhum	Kumardungi	14
885	West Singhbhum	Majhgaon	12
886	West Singhbhum	Manoharpur	9
887	West Singhbhum	Noamundi (barajamda)	37
888	West Singhbhum	Sonua	10
889	West Singhbhum	Tantnagar	0
890	West Singhbhum	Tonto	16
891	West Tripura	Agartala Sadar	6
892	West Tripura	Dukli	1
893	West Tripura	Hezamara	2
894	West Tripura	Jirania	6
895	West Tripura	Mohanpur	2
896	Yadgir	Shahpur	9
897	Yamuna Nagar	Chhachhrauli	12
898	Yamuna Nagar	Jagadhri	131
899	Yavatmal	Arni	2
900	Yavatmal	Babulgaon	11
901	Yavatmal	Digras	0
902	Yavatmal	Ghatanji	3
903	Yavatmal	Kalamb	24
904	Yavatmal	Mahagaon	5
905	Yavatmal	Pusad	2
906	Yavatmal	Ralegaon	1
907	Yavatmal	Umarkhed	6
908	Yavatmal	Yavatmal	3

[909 rows x 3 columns]