ds-csv-file

August 29, 2024

[1]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
df
```

[1]:

	SNO	REGNO	NAME	DOB	GENDER	ADDRESS	M1	M2	M3	M4	TOTAL	AVG
0	1	1220121	ARUN	2000-02-10	MALE	THANDALAM	82.0	81.0	90.0	NaN	NaN	NaN
1	2	1220122	BABU	1999-01-25	MALE	KANCHIPURAM	56.0	61.0	80.0	56.0	253.0	84.333333
2	3	1220123	CHARAN	2000.09.21	MALE	THANDALAM	NaN	59.0	60.0	70.0	NaN	0.000000
3	4	1220124	DEVA	2000-11-09	MALE	POONAMALEE	74.0	79.0	80.0	74.0	307.0	102.333333
4	5	1220125	ESTER	2000-11-21	FEMALE	CHITHUR	92.0	95.0	96.0	92.0	375.0	125.000000
5	6	1220126	FARHANA	1999-03-05	FEMALE	THANDALAM	91.0	88.0	90.0	91.0	360.0	120.000000
6	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	49.0	219.0	73.000000
7	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	49.0	219.0	73.000000
8	8	1220128	HEMA	1999-01-25	FEMALE	POONAMALEE	95.0	96.0	90.0	95.0	376.0	125.333333
9	9	1220129	INDRA	2000.09.21	FEMALE	KANCHIPURAM	64.0	NaN	NaN	NaN	NaN	NaN
10	10	1220130	JAHITH	2000-11-09	MALE	THANDALAM	34.0	45.0	50.0	NaN	NaN	NaN
11	11	1220131	KANI	2000-11-21	FEMALE	CHITHUR	96.0	95.0	96.0	NaN	NaN	NaN
12	12	1220132	LATHESSH	1999-03-05	MALE	THANDALAM	NaN	68.0	70.0	NaN	NaN	NaN
13	13	1220133	MANI	2000-10-02	MALE	KANCHIPURAM	71.0	76.0	NaN	NaN	NaN	NaN
14	14	1220134	NANI	20001109	MALE	POONAMALEE	79.0	77.0	80.0	NaN	NaN	NaN
15	15	1220135	NaN	19990125	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16	16	1220136	PRATHAP	20000921	MALE	KANCHIPURAM	86.0	84.0	90.0	NaN	NaN	NaN
17	17	1220137	RAGHU	20001109	MALE	POONAMALEE	67.0	64.0	70.0	NaN	NaN	NaN
18	18	1220138	RATHI	20001121	FEMALE	KANCHIPURAM	81.0	86.0	90.0	NaN	NaN	NaN
19	19	1220139	SARVESH	19990305	MALE	THANDALAM	84.0	87.0	NaN	NaN	NaN	NaN
20	20	1220140	SANTHOSH	20001002	MALE	KANCHIPURAM	76.0	69.0	80.0	NaN	NaN	NaN

```
9   64.0    NaN    0.000000
10  34.0  163.0   54.333333
11  96.0  383.0  127.666667
12  70.0  208.0   69.333333
13  71.0    NaN    0.000000
14  79.0  315.0  105.000000
15   NaN    0.0    0.000000
16  86.0  346.0  115.333333
17   NaN  201.0   67.000000
18  81.0  338.0  112.666667
19  84.0    NaN    0.000000
20  76.0  301.0  100.333333
```

[2]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
x=df.dropna(how='any')
x
```

[2]:

	SNO	REGNO	NAME	DOB	GENDER	ADDRESS	M1	M2	M3	\
1	2	1220122	BABU	1999-01-25	MALE	KANCHIPURAM	56.0	61.0	80.0	
3	4	1220124	DEVA	2000-11-09	MALE	POONAMALEE	74.0	79.0	80.0	
4	5	1220125	ESTER	2000-11-21	FEMALE	CHITHUR	92.0	95.0	96.0	
5	6	1220126	FARHANA	1999-03-05	FEMALE	THANDALAM	91.0	88.0	90.0	
6	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	
7	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	
8	8	1220128	HEMA	1999-01-25	FEMALE	POONAMALEE	95.0	96.0	90.0	
10	10	1220130	JAHITH	2000-11-09	MALE	THANDALAM	34.0	45.0	50.0	
11	11	1220131	KANI	2000-11-21	FEMALE	CHITHUR	96.0	95.0	96.0	
14	14	1220134	NANI	20001109	MALE	POONAMALEE	79.0	77.0	80.0	
16	16	1220136	PRATHAP	20000921	MALE	KANCHIPURAM	86.0	84.0	90.0	
18	18	1220138	RATHI	20001121	FEMALE	KANCHIPURAM	81.0	86.0	90.0	
20	20	1220140	SANTHOSH	20001002	MALE	KANCHIPURAM	76.0	69.0	80.0	

	M4	TOTAL	AVG
1	56.0	253.0	84.333333
3	74.0	307.0	102.333333
4	92.0	375.0	125.000000
5	91.0	360.0	120.000000
6	49.0	219.0	73.000000
7	49.0	219.0	73.000000
8	95.0	376.0	125.333333
10	34.0	163.0	54.333333
11	96.0	383.0	127.666667
14	79.0	315.0	105.000000
16	86.0	346.0	115.333333
18	81.0	338.0	112.666667
20	76.0	301.0	100.333333

[3]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
df.nunique()
```

[3]:
```
SNO        20
REGNO      20
NAME       19
DOB        13
GENDER      2
ADDRESS     4
M1         17
M2         17
M3          6
M4         16
TOTAL      15
AVG        15
dtype: int64
```

[4]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
x2=df.dropna(how="all").shape
x2
```

[4]:
```
(21, 12)
```

[7]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
tot=df.dropna(subset=['TOTAL'],how='any')
tot
```

[7]:

	SNO	REGNO	NAME	DOB	GENDER	ADDRESS	M1	M2	M3	\
1	2	1220122	BABU	1999-01-25	MALE	KANCHIPURAM	56.0	61.0	80.0	
3	4	1220124	DEVA	2000-11-09	MALE	POONAMALEE	74.0	79.0	80.0	
4	5	1220125	ESTER	2000-11-21	FEMALE	CHITHUR	92.0	95.0	96.0	
5	6	1220126	FARHANA	1999-03-05	FEMALE	THANDALAM	91.0	88.0	90.0	
6	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	
7	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	
8	8	1220128	HEMA	1999-01-25	FEMALE	POONAMALEE	95.0	96.0	90.0	
10	10	1220130	JAHITH	2000-11-09	MALE	THANDALAM	34.0	45.0	50.0	
11	11	1220131	KANI	2000-11-21	FEMALE	CHITHUR	96.0	95.0	96.0	
12	12	1220132	LATHESSH	1999-03-05	MALE	THANDALAM	NaN	68.0	70.0	
14	14	1220134	NANI	20001109	MALE	POONAMALEE	79.0	77.0	80.0	
15	15	1220135	NaN	19990125	NaN	NaN	NaN	NaN	NaN	
16	16	1220136	PRATHAP	20000921	MALE	KANCHIPURAM	86.0	84.0	90.0	
17	17	1220137	RAGHU	20001109	MALE	POONAMALEE	67.0	64.0	70.0	
18	18	1220138	RATHI	20001121	FEMALE	KANCHIPURAM	81.0	86.0	90.0	
20	20	1220140	SANTHOSH	20001002	MALE	KANCHIPURAM	76.0	69.0	80.0	

```
      M4  TOTAL         AVG
1   56.0  253.0   84.333333
3   74.0  307.0  102.333333
4   92.0  375.0  125.000000
5   91.0  360.0  120.000000
6   49.0  219.0   73.000000
7   49.0  219.0   73.000000
8   95.0  376.0  125.333333
10  34.0  163.0   54.333333
11  96.0  383.0  127.666667
12  70.0  208.0   69.333333
14  79.0  315.0  105.000000
15   NaN    0.0    0.000000
16  86.0  346.0  115.333333
17   NaN  201.0   67.000000
18  81.0  338.0  112.666667
20  76.0  301.0  100.333333
```

[8]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
tot=df.dropna(subset=['M1','M2','M3','M4'],how='any')
tot
```

[8]:
```
    SNO    REGNO      NAME         DOB  GENDER      ADDRESS    M1    M2    M3  \
1     2  1220122      BABU  1999-01-25    MALE  KANCHIPURAM  56.0  61.0  80.0
3     4  1220124      DEVA  2000-11-09    MALE   POONAMALEE  74.0  79.0  80.0
4     5  1220125     ESTER  2000-11-21  FEMALE      CHITHUR  92.0  95.0  96.0
5     6  1220126   FARHANA  1999-03-05  FEMALE    THANDALAM  91.0  88.0  90.0
6     7  1220127      GANI  2000-10-02    MALE  KANCHIPURAM  49.0  51.0  70.0
7     7  1220127      GANI  2000-10-02    MALE  KANCHIPURAM  49.0  51.0  70.0
8     8  1220128      HEMA  1999-01-25  FEMALE   POONAMALEE  95.0  96.0  90.0
10   10  1220130    JAHITH  2000-11-09    MALE    THANDALAM  34.0  45.0  50.0
11   11  1220131      KANI  2000-11-21  FEMALE      CHITHUR  96.0  95.0  96.0
14   14  1220134      NANI    20001109    MALE   POONAMALEE  79.0  77.0  80.0
16   16  1220136   PRATHAP    20000921    MALE  KANCHIPURAM  86.0  84.0  90.0
18   18  1220138     RATHI    20001121  FEMALE  KANCHIPURAM  81.0  86.0  90.0
20   20  1220140  SANTHOSH    20001002    MALE  KANCHIPURAM  76.0  69.0  80.0

      M4  TOTAL         AVG
1   56.0  253.0   84.333333
3   74.0  307.0  102.333333
4   92.0  375.0  125.000000
5   91.0  360.0  120.000000
6   49.0  219.0   73.000000
7   49.0  219.0   73.000000
8   95.0  376.0  125.333333
```

```
10  34.0  163.0   54.333333
11  96.0  383.0  127.666667
14  79.0  315.0  105.000000
16  86.0  346.0  115.333333
18  81.0  338.0  112.666667
20  76.0  301.0  100.333333
```

[10]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
df.isna().sum()
```

[10]:
```
SNO        0
REGNO      0
NAME       1
DOB        0
GENDER     1
ADDRESS    1
M1         3
M2         2
M3         4
M4         3
TOTAL      5
AVG        1
dtype: int64
```

[11]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
df['M1']
```

[11]:
```
0     82.0
1     56.0
2      NaN
3     74.0
4     92.0
5     91.0
6     49.0
7     49.0
8     95.0
9     64.0
10    34.0
11    96.0
12     NaN
13    71.0
14    79.0
15     NaN
16    86.0
17    67.0
```

```
18    81.0
19    84.0
20    76.0
Name: M1, dtype: float64
```

[14]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
df.isnull()
```

[14]:

	SNO	REGNO	NAME	DOB	GENDER	ADDRESS	M1	M2	M3	M4	TOTAL	AVG
0	False	False	False	False	False	False	False	False	False	True	True	True
1	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	True	False	False	False	True	False
3	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False	False	False	False	False	False
6	False	False	False	False	False	False	False	False	False	False	False	False
7	False	False	False	False	False	False	False	False	False	False	False	False
8	False	False	False	False	False	False	False	False	False	False	False	False
9	False	False	False	False	False	False	False	True	True	False	True	False
10	False	False	False	False	False	False	False	False	False	False	False	False
11	False	False	False	False	False	False	False	False	False	False	False	False
12	False	False	False	False	False	False	True	False	False	False	False	False
13	False	False	False	False	False	False	False	False	True	False	True	False
14	False	False	False	False	False	False	False	False	False	False	False	False
15	False	False	True	False	True	True	True	True	True	True	True	True
16	False	False	False	False	False	False	False	False	False	False	False	False
17	False	False	False	False	False	False	False	False	False	True	True	False
18	False	False	False	False	False	False	False	False	False	False	False	False
19	False	False	False	False	False	False	False	False	True	False	True	False
20	False	False	False	False	False	False	False	False	False	False	False	False

```
14  False  False
15  False  False
16  False  False
17  False  False
18  False  False
19   True  False
20  False  False
```

[15]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
~df.isnull()
```

[15]:

	SNO	REGNO	NAME	DOB	GENDER	ADDRESS	M1	M2	M3	M4	TOTAL	AVG
0	True	True	True	True	True	True	True	True	True	False	False	False
1	True	True	True	True	True	True	True	True	True	True	True	True
2	True	True	True	True	True	True	False	True	True	True	False	True
3	True	True	True	True	True	True	True	True	True	True	True	True
4	True	True	True	True	True	True	True	True	True	True	True	True
5	True	True	True	True	True	True	True	True	True	True	True	True
6	True	True	True	True	True	True	True	True	True	True	True	True
7	True	True	True	True	True	True	True	True	True	True	True	True
8	True	True	True	True	True	True	True	True	True	True	True	True
9	True	True	True	True	True	True	True	False	False	True	False	True
10	True	True	True	True	True	True	True	True	True	True	True	True
11	True	True	True	True	True	True	True	True	True	True	True	True
12	True	True	True	True	True	True	False	True	True	True	False	True
13	True	True	True	True	True	True	True	True	False	True	True	True
14	True	True	True	True	True	True	True	True	True	True	True	True
15	True	True	False	True	False	False	False	False	False	False	False	False
16	True	True	True	True	True	True	True	True	True	True	True	True
17	True	True	True	True	True	True	True	True	True	False	True	True
18	True	True	True	True	True	True	True	True	True	True	True	True
19	True	True	True	True	True	True	True	True	False	True	True	True
20	True	True	True	True	True	True	True	True	True	True	True	True

```
11   True   True
12   True   True
13  False   True
14   True   True
15   True   True
16   True   True
17   True   True
18   True   True
19  False   True
20   True   True
```

[16]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
df.notnull()
```

[16]:
```
      SNO  REGNO   NAME   DOB  GENDER  ADDRESS     M1     M2     M3     M4  \
0    True   True   True  True    True     True   True   True   True  False
1    True   True   True  True    True     True   True   True   True   True
2    True   True   True  True    True     True  False   True   True   True
3    True   True   True  True    True     True   True   True   True   True
4    True   True   True  True    True     True   True   True   True   True
5    True   True   True  True    True     True   True   True   True   True
6    True   True   True  True    True     True   True   True   True   True
7    True   True   True  True    True     True   True   True   True   True
8    True   True   True  True    True     True   True   True   True   True
9    True   True   True  True    True     True   True  False  False   True
10   True   True   True  True    True     True   True   True   True   True
11   True   True   True  True    True     True   True   True   True   True
12   True   True   True  True    True     True  False   True   True   True
13   True   True   True  True    True     True   True   True  False   True
14   True   True   True  True    True     True   True   True   True   True
15   True   True  False  True   False    False  False  False  False  False
16   True   True   True  True    True     True   True   True   True   True
17   True   True   True  True    True     True   True   True   True  False
18   True   True   True  True    True     True   True   True   True   True
19   True   True   True  True    True     True   True   True  False   True
20   True   True   True  True    True     True   True   True   True   True

    TOTAL    AVG
0   False  False
1    True   True
2   False   True
3    True   True
4    True   True
5    True   True
6    True   True
7    True   True
```

```
8    True   True
9   False   True
10   True   True
11   True   True
12   True   True
13  False   True
14   True   True
15   True   True
16   True   True
17   True   True
18   True   True
19  False   True
20   True   True
```

[17]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
xl=df.dropna(axis=0)
xl
```

[17]:

	SNO	REGNO	NAME	DOB	GENDER	ADDRESS	M1	M2	M3	M4	TOTAL	AVG
1	2	1220122	BABU	1999-01-25	MALE	KANCHIPURAM	56.0	61.0	80.0	56.0	253.0	84.333333
3	4	1220124	DEVA	2000-11-09	MALE	POONAMALEE	74.0	79.0	80.0	74.0	307.0	102.333333
4	5	1220125	ESTER	2000-11-21	FEMALE	CHITHUR	92.0	95.0	96.0	92.0	375.0	125.000000
5	6	1220126	FARHANA	1999-03-05	FEMALE	THANDALAM	91.0	88.0	90.0	91.0	360.0	120.000000
6	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	49.0	219.0	73.000000
7	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	49.0	219.0	73.000000
8	8	1220128	HEMA	1999-01-25	FEMALE	POONAMALEE	95.0	96.0	90.0	95.0	376.0	125.333333
10	10	1220130	JAHITH	2000-11-09	MALE	THANDALAM	34.0	45.0	50.0	34.0	163.0	54.333333
11	11	1220131	KANI	2000-11-21	FEMALE	CHITHUR	96.0	95.0	96.0	96.0	383.0	127.666667
14	14	1220134	NANI	20001109	MALE	POONAMALEE	79.0	77.0	80.0	79.0	315.0	105.000000
16	16	1220136	PRATHAP	20000921	MALE	KANCHIPURAM	86.0	84.0	90.0	86.0	346.0	115.333333
18	18	1220138	RATHI	20001121	FEMALE	KANCHIPURAM	81.0	86.0	90.0	81.0	338.0	112.666667
20	20	1220140	SANTHOSH	20001002	MALE	KANCHIPURAM	76.0	69.0	80.0			

```
20  76.0  301.0  100.333333
```

[18]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
df.duplicated()
```

[18]:
```
0     False
1     False
2     False
3     False
4     False
5     False
6     False
7      True
8     False
9     False
10    False
11    False
12    False
13    False
14    False
15    False
16    False
17    False
18    False
19    False
20    False
dtype: bool
```

[19]:
```
import pandas as pd
df=pd.read_csv("SAMPLEIDs.csv")
m=df.drop_duplicates(inplace=False)
m
```

[19]:

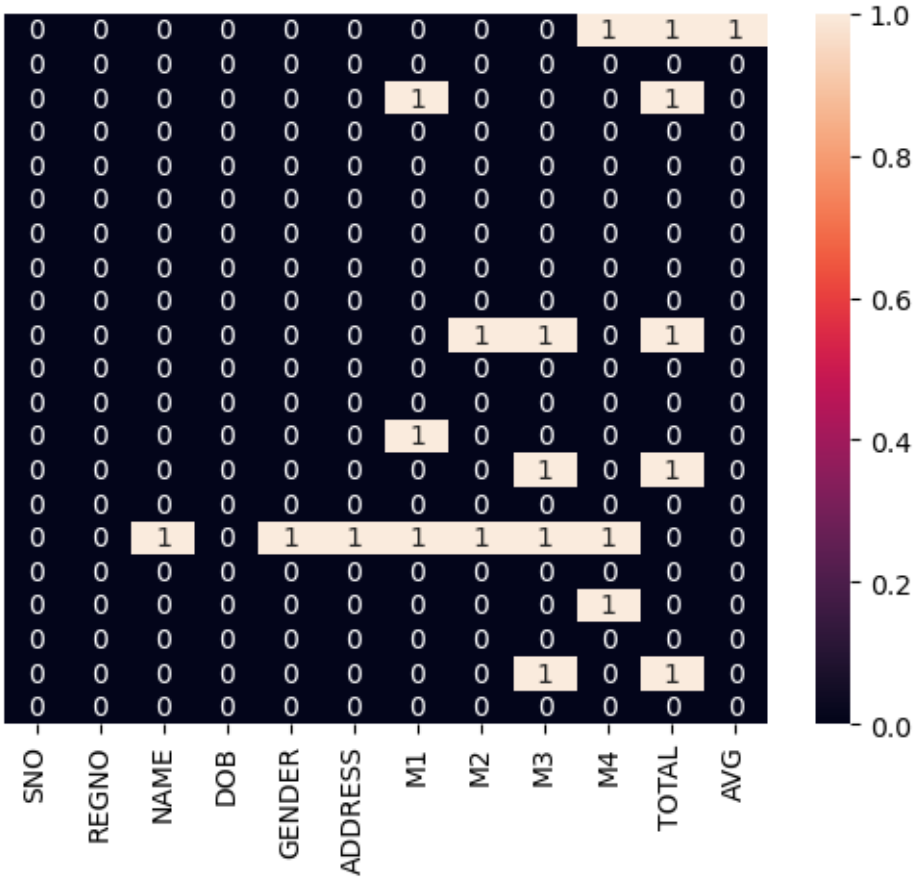
	SNO	REGNO	NAME	DOB	GENDER	ADDRESS	M1	M2	M3	\
0	1	1220121	ARUN	2000-02-10	MALE	THANDALAM	82.0	81.0	90.0	
1	2	1220122	BABU	1999-01-25	MALE	KANCHIPURAM	56.0	61.0	80.0	
2	3	1220123	CHARAN	2000.09.21	MALE	THANDALAM	NaN	59.0	60.0	
3	4	1220124	DEVA	2000-11-09	MALE	POONAMALEE	74.0	79.0	80.0	
4	5	1220125	ESTER	2000-11-21	FEMALE	CHITHUR	92.0	95.0	96.0	
5	6	1220126	FARHANA	1999-03-05	FEMALE	THANDALAM	91.0	88.0	90.0	
6	7	1220127	GANI	2000-10-02	MALE	KANCHIPURAM	49.0	51.0	70.0	
8	8	1220128	HEMA	1999-01-25	FEMALE	POONAMALEE	95.0	96.0	90.0	
9	9	1220129	INDRA	2000.09.21	FEMALE	KANCHIPURAM	64.0	NaN	NaN	
10	10	1220130	JAHITH	2000-11-09	MALE	THANDALAM	34.0	45.0	50.0	
11	11	1220131	KANI	2000-11-21	FEMALE	CHITHUR	96.0	95.0	96.0	
12	12	1220132	LATHESSH	1999-03-05	MALE	THANDALAM	NaN	68.0	70.0	

13	13	1220133	MANI	2000-10-02	MALE	KANCHIPURAM	71.0	76.0	NaN
14	14	1220134	NANI	20001109	MALE	POONAMALEE	79.0	77.0	80.0
15	15	1220135	NaN	19990125	NaN	NaN	NaN	NaN	NaN
16	16	1220136	PRATHAP	20000921	MALE	KANCHIPURAM	86.0	84.0	90.0
17	17	1220137	RAGHU	20001109	MALE	POONAMALEE	67.0	64.0	70.0
18	18	1220138	RATHI	20001121	FEMALE	KANCHIPURAM	81.0	86.0	90.0
19	19	1220139	SARVESH	19990305	MALE	THANDALAM	84.0	87.0	NaN
20	20	1220140	SANTHOSH	20001002	MALE	KANCHIPURAM	76.0	69.0	80.0

	M4	TOTAL	AVG
0	NaN	NaN	NaN
1	56.0	253.0	84.333333
2	70.0	NaN	0.000000
3	74.0	307.0	102.333333
4	92.0	375.0	125.000000
5	91.0	360.0	120.000000
6	49.0	219.0	73.000000
8	95.0	376.0	125.333333
9	64.0	NaN	0.000000
10	34.0	163.0	54.333333
11	96.0	383.0	127.666667
12	70.0	208.0	69.333333
13	71.0	NaN	0.000000
14	79.0	315.0	105.000000
15	NaN	0.0	0.000000
16	86.0	346.0	115.333333
17	NaN	201.0	67.000000
18	81.0	338.0	112.666667
19	84.0	NaN	0.000000
20	76.0	301.0	100.333333

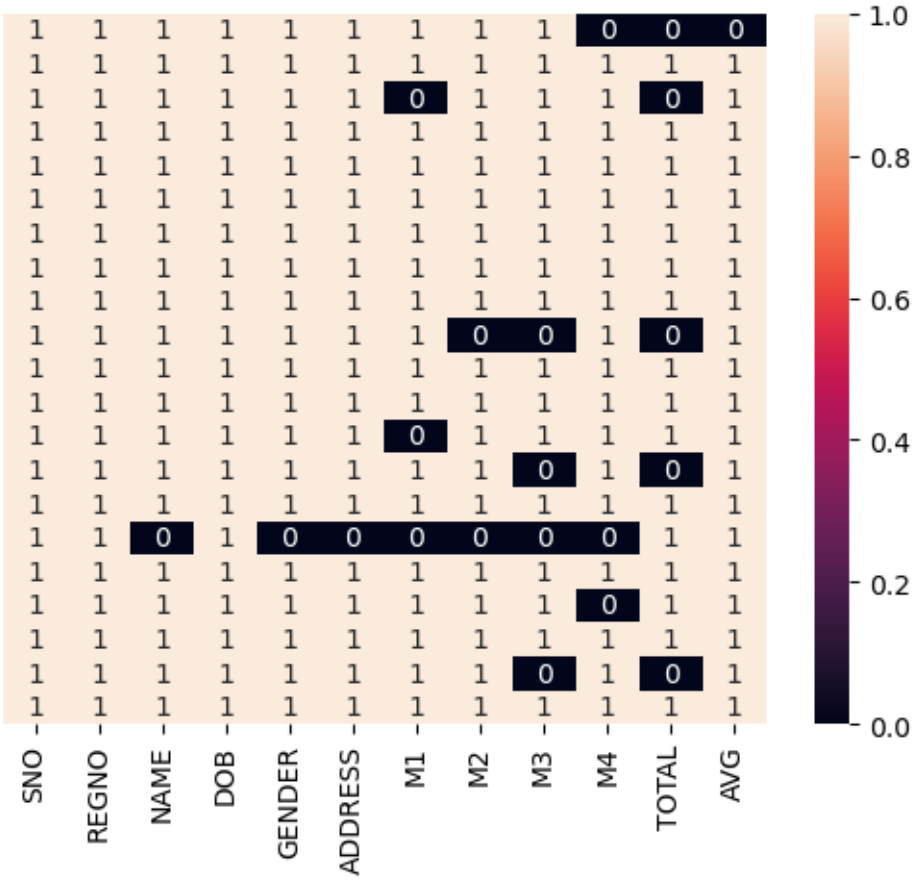
[25]:
```
import seaborn as sns
sns.heatmap(df.isnull(),yticklabels=False,annot=True)
```

[25]: <Axes: >



[26]:
```
import seaborn as sns
sns.heatmap(df.notnull(),yticklabels=False,annot=True)
```

[26]: <Axes: >



[28]:
```
import seaborn as sns

df.dtypes
```

[28]:
```
SNO          int64
REGNO        int64
NAME        object
DOB         object
GENDER      object
ADDRESS     object
M1         float64
M2         float64
M3         float64
M4         float64
TOTAL      float64
AVG        float64
dtype: object
```

[ ]:
```

```