```
import pandas as pd
data={
    "name":["alice","bob","Charlie"],
    "age":[24,25,26],
    "salary":[10000,None,3000],
    "gender":["F","M","F"],
    "height":[1.8,1.7,None]
}
df=pd.DataFrame(data)
df.dropna(how="all",inplace=True)
df
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 3,\n  \"fields\": [\n
{\n      \"column\": \"name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 3,\n
\"samples\": [\n          \"alice\",\n          \"bob\",\n
\"Charlie\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"age\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 1,\n        \"min\": 24,\n        \"max\": 26,\n
\"num_unique_values\": 3,\n        \"samples\": [\n          24,\n
25,\n          26\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"salary\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 4949.747468305833,\n        \"min\": 3000.0,\n
\"max\": 10000.0,\n        \"num_unique_values\": 2,\n
\"samples\": [\n          3000.0,\n          10000.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"gender\",\n      \"properties\":
{\n        \"dtype\": \"string\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"M\",\n          \"F\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"height\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
0.07071067811865482,\n        \"min\": 1.7,\n        \"max\": 1.8,\n
\"num_unique_values\": 2,\n        \"samples\": [\n          1.7,\n
1.8\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df"}
```


```
import pandas as pd
data={
    "name":["alice","bob","Charlie"],
    "age":[24,25,26],
    "salary":[10000,None,3000],
    "gender":["F","M","F"],
    "height":[1.8,1.7,None]
}
df=pd.DataFrame(data)
```

```
df.dropna(subset="salary",inplace=True)
df
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 2,\n  \"fields\": [\n
{\n      \"column\": \"name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"Charlie\",\n          \"alice\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"age\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 1,\n        \"min\": 24,\n        \"max\": 26,\n
\"num_unique_values\": 2,\n        \"samples\": [\n          26,\n
24\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"salary\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 4949.747468305833,\n        \"min\": 3000.0,\n
\"max\": 10000.0,\n        \"num_unique_values\": 2,\n
\"samples\": [\n          3000.0,\n          10000.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"gender\",\n      \"properties\":
{\n        \"dtype\": \"string\",\n        \"num_unique_values\": 1,\n
\"samples\": [\n          \"F\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"height\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": null,\n
\"min\": 1.8,\n        \"max\": 1.8,\n        \"num_unique_values\":
1,\n        \"samples\": [\n          1.8\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"df"}
```

```
import pandas as pd
data={
    "name":["alice","bob","Charlie"],
    "age":[24,25,26],
    "salary":[10000,None,3000],
    "gender":["F","M","F"],
    "height":[1.8,1.7,None]
}
df=pd.DataFrame(data)
df.dropna(how="any",inplace=True)
df
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 1,\n  \"fields\": [\n
{\n      \"column\": \"name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 1,\n
\"samples\": [\n          \"alice\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"age\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": null,\n        \"min\": 24,\n
\"max\": 24,\n        \"num_unique_values\": 1,\n        \"samples\":
```

[\n   24\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"salary\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": null,\n   \"min\": 10000.0,\n   \"max\": 10000.0,\n   \"num_unique_values\": 1,\n   \"samples\": [\n   10000.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"gender\",\n   \"properties\": {\n   \"dtype\": \"string\",\n   \"num_unique_values\": 1,\n   \"samples\": [\n   \"F\"\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"height\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": null,\n   \"min\": 1.8,\n   \"max\": 1.8,\n   \"num_unique_values\": 1,\n   \"samples\": [\n   1.8\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe","variable_name":"df"}

```
import pandas as pd
import numpy as np
data={
    "name":["alice","bob","Charlie","dave","eve","bob","Charlie"],
    "age":[24,np.nan,35,41,np.nan,np.nan,85],
    "salary":[10000,np.nan,2000,np.nan,3000,np.nan,4000]
}
df=pd.DataFrame(data)
~df.duplicated()
#
```

```
0     True
1     True
2     True
3     True
4     True
5    False
6     True
dtype: bool
```

```
import pandas as pd
import numpy as np
data={
    "name":["alice","bob","Charlie","dave","eve","bob","Charlie"],
    "age":[24,np.nan,35,41,np.nan,np.nan,85],
    "salary":[10000,np.nan,2000,np.nan,3000,np.nan,4000]
}
df=pd.DataFrame(data)
df_filled=df.fillna(10,inplace=True)
df
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 7,\n \"fields\": [\n{\n   \"column\": \"name\",\n   \"properties\": {\n

\"dtype\": \"string\",\n   \"num_unique_values\": 5,\n
\"samples\": [\n   \"bob\",\n   \"eve\",\n
\"Charlie\"\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"age\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 27.090676660149374,\n   \"min\": 10.0,\n   \"max\":
85.0,\n   \"num_unique_values\": 5,\n   \"samples\": [\n
10.0,\n   85.0,\n   35.0\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"salary\",\n   \"properties\":
{\n   \"dtype\": \"number\",\n   \"std\":
3588.5439413566637,\n   \"min\": 10.0,\n   \"max\":
10000.0,\n   \"num_unique_values\": 5,\n   \"samples\": [\n
10.0,\n   4000.0,\n   2000.0\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe","variable_name":"df"}

```
import pandas as pd
import numpy as np
data={
    "name":["alice","bob","Charlie","dave","eve","bob","Charlie"],
    "age":[24,np.nan,35,41,np.nan,np.nan,85],
    "salary":[10000,np.nan,2000,np.nan,3000,np.nan,4000]
}
df=pd.DataFrame(data)
df_filled=df.fillna(10)
df_filled
```

{"summary":"{\n \"name\": \"df_filled\",\n \"rows\": 7,\n
\"fields\": [\n   {\n   \"column\": \"name\",\n
\"properties\": {\n   \"dtype\": \"string\",\n
\"num_unique_values\": 5,\n   \"samples\": [\n   \"bob\",\
n   \"eve\",\n   \"Charlie\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"age\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 27.090676660149374,\n
\"min\": 10.0,\n   \"max\": 85.0,\n   \"num_unique_values\":
5,\n   \"samples\": [\n   10.0,\n   85.0,\n
35.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"salary\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 3588.5439413566637,\n   \"min\": 10.0,\n   \"max\":
10000.0,\n   \"num_unique_values\": 5,\n   \"samples\": [\n
10.0,\n   4000.0,\n   2000.0\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe","variable_name":"df_filled"}

```
import pandas as pd
import numpy as np
data={
```

```
    "name":["alice","bob","Charlie","dave","eve","bob","Charlie"],
    "age":[24,np.nan,35,41,np.nan,np.nan,85],
    "salary":[10000,np.nan,2000,np.nan,3000,np.nan,4000]
}
df=pd.DataFrame(data)
df_filled=df.fillna(method="ffill")
df_filled
```

```
<ipython-input-25-7cc257af65c2>:9: FutureWarning: DataFrame.fillna
with 'method' is deprecated and will raise in a future version. Use
obj.ffill() or obj.bfill() instead.
  df_filled=df.fillna(method="ffill")
```


```
{"summary":"{\n  \"name\": \"df_filled\",\n  \"rows\": 7,\n
\"fields\": [\n    {\n      \"column\": \"name\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 5,\n        \"samples\": [\n          \"bob\",\
n          \"eve\",\n          \"Charlie\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"age\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 20.606286604311986,\n
\"min\": 24.0,\n        \"max\": 85.0,\n        \"num_unique_values\":
4,\n        \"samples\": [\n          35.0,\n          85.0,\n
24.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"salary\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 3579.0395093549623,\n        \"min\": 2000.0,\n
\"max\": 10000.0,\n        \"num_unique_values\": 4,\n
\"samples\": [\n          2000.0,\n          4000.0,\n
10000.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df_filled"}
```


```
import pandas as pd
import numpy as np
data={
    "name":["alice","bob","Charlie","dave","eve","bob","Charlie"],
    "age":[24,np.nan,35,41,np.nan,np.nan,85],
    "salary":[10000,np.nan,2000,np.nan,3000,np.nan,4000]
}
df=pd.DataFrame(data)
df_filled=df.fillna(method="bfill")
df_filled
```

```
<ipython-input-26-b0061b5aa1c5>:9: FutureWarning: DataFrame.fillna
with 'method' is deprecated and will raise in a future version. Use
obj.ffill() or obj.bfill() instead.
  df_filled=df.fillna(method="bfill")
```

{"summary":"{\n \"name\": \"df_filled\",\n \"rows\": 7,\n \"fields\": [\n   {\n   \"column\": \"name\",\n   \"properties\": {\n   \"dtype\": \"string\",\n   \"num_unique_values\": 5,\n   \"samples\": [\n   \"bob\",\n   \"eve\",\n   \"Charlie\"\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"age\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 27.849083083136303,\n   \"min\": 24.0,\n   \"max\": 85.0,\n   \"num_unique_values\": 4,\n   \"samples\": [\n   35.0,\n   85.0,\n   24.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"salary\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 2768.874620972692,\n   \"min\": 2000.0,\n   \"max\": 10000.0,\n   \"num_unique_values\": 4,\n   \"samples\": [\n   2000.0,\n   4000.0,\n   10000.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe","variable_name":"df_filled"}

```
import pandas as pd
import numpy as np
data={
    "name":["alice","bob","Charlie","dave","eve","bob","Charlie"],
    "age":[24,np.nan,35,41,np.nan,np.nan,85],
    "salary":[10000,np.nan,2000,np.nan,3000,np.nan,4000]
}
df=pd.DataFrame(data)
df_filled=df.fillna(df["salary"].mean())
df_filled
```

{"summary":"{\n \"name\": \"df_filled\",\n \"rows\": 7,\n \"fields\": [\n   {\n   \"column\": \"name\",\n   \"properties\": {\n   \"dtype\": \"string\",\n   \"num_unique_values\": 5,\n   \"samples\": [\n   \"bob\",\n   \"eve\",\n   \"Charlie\"\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"age\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 2514.3314173335375,\n   \"min\": 24.0,\n   \"max\": 4750.0,\n   \"num_unique_values\": 5,\n   \"samples\": [\n   4750.0,\n   85.0,\n   35.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"salary\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 2541.325113662818,\n   \"min\": 2000.0,\n   \"max\": 10000.0,\n   \"num_unique_values\": 5,\n   \"samples\": [\n   4750.0,\n   4000.0,\n   2000.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe","variable_name":"df_filled"}

```
import pandas as pd
df=pd.read_csv("/content/SAMPLEIDS.csv")
df.fillna(0)
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 21,\n  \"fields\": [\n
{\n      \"column\": \"SNO\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 5,\n        \"min\": 1,\n
\"max\": 20,\n        \"num_unique_values\": 20,\n        \"samples\":
[\n          1,\n          18,\n          16\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"REGNO\",\n      \"properties\": {\
n        \"dtype\": \"number\",\n        \"std\": 5,\n        \"min\":
1220121,\n        \"max\": 1220140,\n        \"num_unique_values\":
20,\n        \"samples\": [\n          1220121,\n          1220138,\n
1220136\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"NAME\",\n      \"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 20,\n        \"samples\": [\n
\"ARUN\",\n          \"RATHI\",\n          \"PRATHAP\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"DOB\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 13,\n
\"samples\": [\n          \"19990305\",\n          \"20000921\",\n
\"2000-02-10\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"GENDER\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 3,\n        \"samples\":
[\n          \"MALE\",\n          \"FEMALE\",\n          0\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"ADDRESS\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 5,\n        \"samples\":
[\n          \"KANCHIPURAM\",\n          0,\n          \"POONAMALEE\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M1\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 30.990782039641584,\n        \"min\": 0.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 18,\n        \"samples\": [\n
82.0,\n          56.0,\n          64.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"M2\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 26.933074020439555,\n
\"min\": 0.0,\n        \"max\": 96.0,\n        \"num_unique_values\":
18,\n        \"samples\": [\n          81.0,\n          61.0,\n
0.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M3\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 34.05066253463534,\n        \"min\": 0.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 7,\n        \"samples\": [\n
90.0,\n          80.0,\n          0.0\n        ],\n
```

\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"M4\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 30.763847056476628,\n
\"min\": 0.0,\n   \"max\": 96.0,\n   \"num_unique_values\":
17,\n   \"samples\": [\n   0.0,\n   56.0,\n
91.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"TOTAL\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 148.25809220667148,\n   \"min\": 0.0,\n   \"max\":
383.0,\n   \"num_unique_values\": 15,\n   \"samples\": [\n
208.0,\n   346.0,\n   0.0\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"AVG\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 49.41936406671593,\n
\"min\": 0.0,\n   \"max\": 127.6666667,\n
\"num_unique_values\": 15,\n   \"samples\": [\n
69.33333333,\n   115.3333333,\n   0.0\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe"}

```
df.head(10)
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 21,\n \"fields\": [\n
{\n   \"column\": \"SNO\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 5,\n   \"min\": 1,\n
\"max\": 20,\n   \"num_unique_values\": 20,\n   \"samples\":
[\n   1,\n   18,\n   16\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"REGNO\",\n   \"properties\": {\
n   \"dtype\": \"number\",\n   \"std\": 5,\n   \"min\":
1220121,\n   \"max\": 1220140,\n   \"num_unique_values\":
20,\n   \"samples\": [\n   1220121,\n   1220138,\n
1220136\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"NAME\",\n   \"properties\": {\n   \"dtype\": \"string\",\n
\"num_unique_values\": 19,\n   \"samples\": [\n
\"ARUN\",\n   \"FARHANA\",\n   \"LATHESSH\"\n   ],\
n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   },\n   {\n   \"column\": \"DOB\",\n   \"properties\": {\
n   \"dtype\": \"string\",\n   \"num_unique_values\": 13,\n
\"samples\": [\n   \"19990305\",\n   \"20000921\",\n
\"2000-02-10\"\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"GENDER\",\n   \"properties\": {\n   \"dtype\":
\"category\",\n   \"num_unique_values\": 2,\n   \"samples\":
[\n   \"FEMALE\",\n   \"MALE\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"ADDRESS\",\n   \"properties\":

{\n   \"dtype\": \"category\",\n   \"num_unique_values\":
4,\n   \"samples\": [\n   \"KANCHIPURAM\",\n
\"CHITHUR\"\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"M1\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 17.5800689284602,\n   \"min\": 34.0,\n   \"max\":
96.0,\n   \"num_unique_values\": 17,\n   \"samples\": [\n
82.0,\n   56.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"M2\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 15.836149334070898,\n   \"min\": 45.0,\n   \"max\":
96.0,\n   \"num_unique_values\": 17,\n   \"samples\": [\n
81.0,\n   61.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"M3\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 13.010177011953102,\n   \"min\": 50.0,\n   \"max\":
96.0,\n   \"num_unique_values\": 6,\n   \"samples\": [\n
90.0,\n   80.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"M4\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 17.426315462203576,\n   \"min\": 34.0,\n   \"max\":
96.0,\n   \"num_unique_values\": 16,\n   \"samples\": [\n
56.0,\n   70.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"TOTAL\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 102.04868119350358,\n   \"min\": 0.0,\n   \"max\":
383.0,\n   \"num_unique_values\": 15,\n   \"samples\": [\n
315.0,\n   346.0\n   ],\n   \"semantic_type\": \"\",\
n   \"description\": \"\"\n   }\n   },\n   {\n
\"column\": \"AVG\",\n   \"properties\": {\n   \"dtype\":
\"number\",\n   \"std\": 48.017126572152904,\n   \"min\":
0.0,\n   \"max\": 127.6666667,\n   \"num_unique_values\":
15,\n   \"samples\": [\n   69.33333333,\n
115.3333333\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   }\n]\
n}","type":"dataframe","variable_name":"df"}

```
df.tail(10)
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 10,\n \"fields\": [\n
{\n   \"column\": \"SNO\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 3,\n   \"min\": 11,\n
\"max\": 20,\n   \"num_unique_values\": 10,\n   \"samples\":
[\n   19,\n   12,\n   16\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"REGNO\",\n   \"properties\": {\
n   \"dtype\": \"number\",\n   \"std\": 3,\n   \"min\":
1220131,\n   \"max\": 1220140,\n   \"num_unique_values\":
10,\n   \"samples\": [\n   1220139,\n   1220132,\n
1220136\n   ],\n   \"semantic_type\": \"\",\n

```
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"NAME\",\n      \"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 9,\n        \"samples\": [\n
\"SARVESH\",\n          \"LATHESSH\",\n          \"RAGHU\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"DOB\",\n      \"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 9,\n        \"samples\": [\n
\"19990305\",\n          \"1999-03-05\",\n          \"20000921\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"GENDER\",\n      \"properties\":
{\n        \"dtype\": \"category\",\n        \"num_unique_values\":
2,\n        \"samples\": [\n          \"MALE\",\n          \"FEMALE\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"ADDRESS\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          \"THANDALAM\",\n          \"POONAMALEE\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"M1\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 9.071147352221454,\n
\"min\": 67.0,\n        \"max\": 96.0,\n        \"num_unique_values\":
8,\n        \"samples\": [\n          71.0,\n          81.0\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M2\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 10.309111396128076,\n        \"min\": 64.0,\n        \"max\":
95.0,\n        \"num_unique_values\": 9,\n        \"samples\": [\n
87.0,\n          68.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M3\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 10.160614904526952,\n        \"min\": 70.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 4,\n        \"samples\": [\n
70.0,\n          90.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M4\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 8.501050355271232,\n        \"min\": 70.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 8,\n        \"samples\": [\n
70.0,\n          81.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"TOTAL\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 123.68277394782416,\n        \"min\": 0.0,\n        \"max\":
383.0,\n        \"num_unique_values\": 8,\n        \"samples\": [\n
208.0,\n          201.0\n        ],\n        \"semantic_type\": \"\",\
n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"AVG\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 51.69874750910483,\n        \"min\":
0.0,\n        \"max\": 127.6666667,\n        \"num_unique_values\":
8,\n        \"samples\": [\n          69.33333333,\n          67.0\n
```

```
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    }\n  ]\n}","type":"dataframe"}
```


```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21 entries, 0 to 20
Data columns (total 12 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   SNO      21 non-null     int64
 1   REGNO    21 non-null     int64
 2   NAME     20 non-null     object
 3   DOB      21 non-null     object
 4   GENDER   20 non-null     object
 5   ADDRESS  20 non-null     object
 6   M1       18 non-null     float64
 7   M2       19 non-null     float64
 8   M3       17 non-null     float64
 9   M4       18 non-null     float64
 10  TOTAL    16 non-null     float64
 11  AVG      20 non-null     float64
dtypes: float64(6), int64(2), object(4)
memory usage: 2.1+ KB
```

```
df.describe()
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 8,\n  \"fields\": [\n
{\n      \"column\": \"SNO\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 7.072545821569434,\n
\"min\": 1.0,\n        \"max\": 21.0,\n        \"num_unique_values\":
8,\n        \"samples\": [\n          10.333333333333334,\n
10.0,\n          21.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"REGNO\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 564804.4123841431,\n        \"min\": 5.816642788871715,\n
\"max\": 1220140.0,\n        \"num_unique_values\": 8,\n
\"samples\": [\n          1220130.3333333333,\n          1220130.0,\n
21.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M1\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 30.891874149417365,\n        \"min\": 17.5800689284602,\n
\"max\": 96.0,\n        \"num_unique_values\": 8,\n
\"samples\": [\n          73.66666666666667,\n          77.5,\n
18.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M2\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 30.12601293384157,\n        \"min\": 15.836149334070898,\n
\"max\": 96.0,\n        \"num_unique_values\": 8,\n
\"samples\": [\n          74.3157894736842,\n          77.0,\n
```

```
19.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M3\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 32.080827687721516,\n        \"min\": 13.010177011953102,\n
\"max\": 96.0,\n        \"num_unique_values\": 8,\n
\"samples\": [\n          79.52941176470588,\n          80.0,\n
17.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M4\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 30.70064611063876,\n        \"min\": 17.426315462203576,\n
\"max\": 96.0,\n        \"num_unique_values\": 8,\n
\"samples\": [\n          73.16666666666667,\n          75.0,\n
18.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"TOTAL\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 149.00876048154265,\n        \"min\": 0.0,\n        \"max\":
383.0,\n        \"num_unique_values\": 8,\n        \"samples\": [\n
272.75,\n          304.0,\n          16.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"AVG\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 44.06480665714355,\n
\"min\": 0.0,\n        \"max\": 127.6666667,\n
\"num_unique_values\": 8,\n        \"samples\": [\n
72.7333333295,\n          78.66666666500001,\n          20.0\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\n}","type":"dataframe"}
```


```
df.shape
```

```
(21, 12)
```

```
df.isnull().sum()
```

```
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
```

```
import pandas as pd
df=pd.read_csv("/content/SAMPLEIDS.csv")
df.nunique()
```

```
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
```

```
df.shape
```

```
(21, 12)
```

```
df['GENDER'].value_counts()
```

```
GENDER
MALE      14
FEMALE     6
Name: count, dtype: int64
```

```
df.dropna(how="any").shape
```

```
(13, 12)
```

```
x=df.dropna(how="any")
x
```


```
{"summary":"{\n  \"name\": \"x\",\n  \"rows\": 13,\n  \"fields\": [\n{\n      \"column\": \"SNO\",\n      \"properties\": {\n\"dtype\": \"number\",\n        \"std\": 5,\n        \"min\": 2,\n\"max\": 20,\n        \"num_unique_values\": 12,\n        \"samples\": [\n          18,\n          16,\n          2\n        ],\n\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"REGNO\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 5,\n        \"min\": 1220122,\n        \"max\": 1220140,\n        \"num_unique_values\": 12,\n        \"samples\": [\n          1220138,\n          1220136,\n1220122\n        ],\n        \"semantic_type\": \"\",\n\"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"NAME\",\n      \"properties\": {\n        \"dtype\": \"string\",\n\"num_unique_values\": 12,\n        \"samples\": [\n\"RATHI\",\n          \"PRATHAP\",\n          \"BABU\"\n        ],\n\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"DOB\",\n      \"properties\": {\n\"dtype\": \"string\",\n        \"num_unique_values\": 9,\n\"samples\": [\n          \"20001121\",\n          \"2000-11-09\",\n
```

\"20001109\"\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"GENDER\",\n   \"properties\": {\n   \"dtype\":
\"category\",\n   \"num_unique_values\": 2,\n   \"samples\":
[\n   \"FEMALE\",\n   \"MALE\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"ADDRESS\",\n   \"properties\":
{\n   \"dtype\": \"category\",\n   \"num_unique_values\":
4,\n   \"samples\": [\n   \"POONAMALEE\",\n
\"THANDALAM\"\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"M1\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 20.274715844225845,\n   \"min\": 34.0,\n   \"max\":
96.0,\n   \"num_unique_values\": 12,\n   \"samples\": [\n
81.0,\n   86.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"M2\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 18.063988826050917,\n   \"min\": 45.0,\n   \"max\":
96.0,\n   \"num_unique_values\": 11,\n   \"samples\": [\n
96.0,\n   61.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"M3\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 12.828254073103736,\n   \"min\": 50.0,\n   \"max\":
96.0,\n   \"num_unique_values\": 5,\n   \"samples\": [\n
96.0,\n   50.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"M4\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 20.274715844225845,\n   \"min\": 34.0,\n   \"max\":
96.0,\n   \"num_unique_values\": 12,\n   \"samples\": [\n
81.0,\n   86.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"TOTAL\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 70.47003363860159,\n   \"min\": 163.0,\n   \"max\":
383.0,\n   \"num_unique_values\": 12,\n   \"samples\": [\n
338.0,\n   346.0\n   ],\n   \"semantic_type\": \"\",\
n   \"description\": \"\"\n   }\n   },\n   {\n
\"column\": \"AVG\",\n   \"properties\": {\n   \"dtype\":
\"number\",\n   \"std\": 23.490011213604618,\n   \"min\":
54.33333333,\n   \"max\": 127.6666667,\n
\"num_unique_values\": 12,\n   \"samples\": [\n
112.6666667,\n   115.3333333\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe","variable_name":"x"}

```
x2=df.dropna(how="all").shape
x2
```

```
(21, 12)
```

```
tot=df.dropna(subset=["TOTAL"],how="any").shape
tot
```

```
(16, 12)
```

```
tot=df.dropna(subset=["TOTAL"],how="any")
tot
```


```
{"summary":"{\n  \"name\": \"tot\",\n  \"rows\": 16,\n  \"fields\": [\n    {\n      \"column\": \"SNO\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 5,\n        \"min\": 2,\n        \"max\": 20,\n        \"num_unique_values\": 15,\n        \"samples\": [\n          14,\n          16,\n          2\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"REGNO\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 5,\n        \"min\": 1220122,\n        \"max\": 1220140,\n        \"num_unique_values\": 15,\n        \"samples\": [\n          1220134,\n          1220136,\n          1220122\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"NAME\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 14,\n        \"samples\": [\n          \"NANI\",\n          \"RAGHU\",\n          \"BABU\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"DOB\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 10,\n        \"samples\": [\n          \"20001121\",\n          \"2000-11-09\",\n          \"20001109\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"GENDER\",\n      \"properties\": {\n        \"dtype\": \"category\",\n        \"num_unique_values\": 2,\n        \"samples\": [\n          \"FEMALE\",\n          \"MALE\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"ADDRESS\",\n      \"properties\": {\n        \"dtype\": \"category\",\n        \"num_unique_values\": 4,\n        \"samples\": [\n          \"POONAMALEE\",\n          \"THANDALAM\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"M1\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 19.561259088894488,\n        \"min\": 34.0,\n        \"max\": 96.0,\n        \"num_unique_values\": 13,\n        \"samples\": [\n          81.0,\n          86.0\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"M2\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 17.048111192683347,\n        \"min\": 45.0,\n        \"max\": 96.0,\n        \"num_unique_values\": 13,\n        \"samples\": [\n          86.0,\n          84.0\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"M3\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 12.569047393535932,\n        \"min\": 50.0,\n        \"max\":
```

```
96.0,\n        \"num_unique_values\": 5,\n        \"samples\": [\n
96.0,\n          50.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M4\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 19.504296511732548,\n        \"min\": 34.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 13,\n        \"samples\": [\n
81.0,\n          79.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"TOTAL\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 102.04868119350358,\n        \"min\": 0.0,\n        \"max\":
383.0,\n        \"num_unique_values\": 15,\n        \"samples\": [\n
315.0,\n          346.0\n        ],\n        \"semantic_type\": \"\",\
n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"AVG\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 34.016227063541415,\n        \"min\":
0.0,\n        \"max\": 127.6666667,\n        \"num_unique_values\":
15,\n        \"samples\": [\n          105.0,\n          115.3333333\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    }\n  ]\n}","type":"dataframe","variable_name":"tot"}
```


```
tot=df.dropna(subset=['M1','M2','M3','M4'],how="any")
tot
```


```
{"summary":"{\n  \"name\": \"tot\",\n  \"rows\": 13,\n  \"fields\": [\
n    {\n      \"column\": \"SNO\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 5,\n        \"min\": 2,\n
\"max\": 20,\n        \"num_unique_values\": 12,\n        \"samples\":
[\n          18,\n          16,\n          2\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"REGNO\",\n      \"properties\": {\
n        \"dtype\": \"number\",\n        \"std\": 5,\n        \"min\":
1220122,\n        \"max\": 1220140,\n        \"num_unique_values\":
12,\n        \"samples\": [\n          1220138,\n          1220136,\n
1220122\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"NAME\",\n      \"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 12,\n        \"samples\": [\n
\"RATHI\",\n          \"PRATHAP\",\n          \"BABU\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"DOB\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 9,\n
\"samples\": [\n          \"20001121\",\n          \"2000-11-09\",\n
\"20001109\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"GENDER\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          \"FEMALE\",\n          \"MALE\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"ADDRESS\",\n      \"properties\":
{\n        \"dtype\": \"category\",\n        \"num_unique_values\":
```

```
4,\n        \"samples\": [\n          \"POONAMALEE\",\n
\"THANDALAM\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M1\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 20.274715844225845,\n        \"min\": 34.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 12,\n        \"samples\": [\n
81.0,\n          86.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M2\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 18.063988826050917,\n        \"min\": 45.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 11,\n        \"samples\": [\n
96.0,\n          61.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M3\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 12.828254073103736,\n        \"min\": 50.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 5,\n        \"samples\": [\n
96.0,\n          50.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M4\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 20.274715844225845,\n        \"min\": 34.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 12,\n        \"samples\": [\n
81.0,\n          86.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"TOTAL\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 70.47003363860159,\n        \"min\": 163.0,\n        \"max\":
383.0,\n        \"num_unique_values\": 12,\n        \"samples\": [\n
338.0,\n          346.0\n        ],\n        \"semantic_type\": \"\",\
n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"AVG\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 23.490011213604618,\n        \"min\":
54.33333333,\n        \"max\": 127.6666667,\n
\"num_unique_values\": 12,\n        \"samples\": [\n
112.6666667,\n          115.3333333\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"tot"}
```


```
tot=df.dropna(subset=['M1','M2','M3','M4'],how="any").shape
tot
```

```
(13, 12)
```

```
s=df.fillna(method="ffill")
s
```

```
<ipython-input-33-b8fa547bb146>:1: FutureWarning: DataFrame.fillna
with 'method' is deprecated and will raise in a future version. Use
obj.ffill() or obj.bfill() instead.
  s=df.fillna(method="ffill")
```

```
{"summary":"{\n  \"name\": \"s\",\n  \"rows\": 21,\n  \"fields\": [\n
{\n      \"column\": \"SNO\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 5,\n        \"min\": 1,\n
\"max\": 20,\n        \"num_unique_values\": 20,\n        \"samples\":
[\n          1,\n          18,\n          16\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"REGNO\",\n      \"properties\": {\
n        \"dtype\": \"number\",\n        \"std\": 5,\n        \"min\":
1220121,\n        \"max\": 1220140,\n        \"num_unique_values\":
20,\n        \"samples\": [\n          1220121,\n          1220138,\n
1220136\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"NAME\",\n      \"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 19,\n        \"samples\": [\n
\"ARUN\",\n          \"FARHANA\",\n          \"LATHESSH\"\n        ],\
n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"DOB\",\n      \"properties\": {\
n        \"dtype\": \"string\",\n        \"num_unique_values\": 13,\n
\"samples\": [\n          \"19990305\",\n          \"20000921\",\n
\"2000-02-10\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"GENDER\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          \"FEMALE\",\n          \"MALE\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"ADDRESS\",\n      \"properties\":
{\n        \"dtype\": \"category\",\n        \"num_unique_values\":
4,\n        \"samples\": [\n          \"KANCHIPURAM\",\n
\"CHITHUR\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M1\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 17.44788157423621,\n        \"min\": 34.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 17,\n        \"samples\": [\n
82.0,\n          56.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M2\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 15.753155390648082,\n        \"min\": 45.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 17,\n        \"samples\": [\n
81.0,\n          61.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M3\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 12.271531126574066,\n        \"min\": 50.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 6,\n        \"samples\": [\n
90.0,\n          80.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M4\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 16.77058828959042,\n        \"min\": 34.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 16,\n        \"samples\": [\n
56.0,\n          70.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
```

\"TOTAL\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 96.06931434071306,\n   \"min\": 0.0,\n   \"max\":
383.0,\n   \"num_unique_values\": 15,\n   \"samples\": [\n
315.0,\n   346.0\n   ],\n   \"semantic_type\": \"\",\
n   \"description\": \"\"\n   }\n   },\n   {\n
\"column\": \"AVG\",\n   \"properties\": {\n   \"dtype\":
\"number\",\n   \"std\": 48.017126572152904,\n   \"min\":
0.0,\n   \"max\": 127.6666667,\n   \"num_unique_values\":
15,\n   \"samples\": [\n   69.33333333,\n
115.3333333\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   }\n]\
n}","type":"dataframe","variable_name":"s"}

```
df.isna().sum()
```

```
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
```

```
df['M1']
```

```
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
```

```
df.isnull()
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 21,\n  \"fields\": [\n
{\n      \"column\": \"SNO\",\n      \"properties\": {\n
\"dtype\": \"boolean\",\n        \"num_unique_values\": 1,\n
\"samples\": [\n          false\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"REGNO\",\n      \"properties\": {\
n        \"dtype\": \"boolean\",\n        \"num_unique_values\": 1,\n
\"samples\": [\n          false\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"NAME\",\n      \"properties\": {\n
\"dtype\": \"boolean\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          true\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"DOB\",\n      \"properties\": {\n        \"dtype\":
\"boolean\",\n        \"num_unique_values\": 1,\n        \"samples\":
[\n          false\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"GENDER\",\n      \"properties\": {\n        \"dtype\": \"boolean\",\
n        \"num_unique_values\": 2,\n        \"samples\": [\n
true\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"ADDRESS\",\n      \"properties\": {\n        \"dtype\":
\"boolean\",\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          true\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M1\",\n      \"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n          true\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"M2\",\n      \"properties\": {\n
\"dtype\": \"boolean\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          true\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"M3\",\n      \"properties\": {\n        \"dtype\":
\"boolean\",\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          true\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M4\",\n      \"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"TOTAL\",\n      \"properties\":
{\n        \"dtype\": \"boolean\",\n        \"num_unique_values\": 2,\
n        \"samples\": [\n          false\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"AVG\",\n      \"properties\": {\n
```

\"dtype\": \"boolean\",\n   \"num_unique_values\": 2,\n
\"samples\": [\n   false\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe"}

```
df.notnull()
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 21,\n \"fields\": [\n
{\n   \"column\": \"SNO\",\n   \"properties\": {\n
\"dtype\": \"boolean\",\n   \"num_unique_values\": 1,\n
\"samples\": [\n   true\n   ],\n   \"semantic_type\":
\"\",\n   \"description\": \"\"\n   }\n   },\n   {\n
\"column\": \"REGNO\",\n   \"properties\": {\n   \"dtype\":
\"boolean\",\n   \"num_unique_values\": 1,\n   \"samples\":
[\n   true\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"NAME\",\n   \"properties\": {\n   \"dtype\": \"boolean\",\n
\"num_unique_values\": 2,\n   \"samples\": [\n   false\n
],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   },\n   {\n   \"column\": \"DOB\",\n   \"properties\": {\
n   \"dtype\": \"boolean\",\n   \"num_unique_values\": 1,\n
\"samples\": [\n   true\n   ],\n   \"semantic_type\":
\"\",\n   \"description\": \"\"\n   }\n   },\n   {\n
\"column\": \"GENDER\",\n   \"properties\": {\n   \"dtype\":
\"boolean\",\n   \"num_unique_values\": 2,\n   \"samples\":
[\n   false\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"ADDRESS\",\n   \"properties\": {\n   \"dtype\":
\"boolean\",\n   \"num_unique_values\": 2,\n   \"samples\":
[\n   false\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"M1\",\n   \"properties\": {\n   \"dtype\": \"boolean\",\n
\"num_unique_values\": 2,\n   \"samples\": [\n   false\n
],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   },\n   {\n   \"column\": \"M2\",\n   \"properties\": {\n
\"dtype\": \"boolean\",\n   \"num_unique_values\": 2,\n
\"samples\": [\n   false\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"M3\",\n   \"properties\": {\n
\"dtype\": \"boolean\",\n   \"num_unique_values\": 2,\n
\"samples\": [\n   false\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"M4\",\n   \"properties\": {\n
\"dtype\": \"boolean\",\n   \"num_unique_values\": 2,\n
\"samples\": [\n   true\n   ],\n   \"semantic_type\":
\"\",\n   \"description\": \"\"\n   }\n   },\n   {\n
\"column\": \"TOTAL\",\n   \"properties\": {\n   \"dtype\":
\"boolean\",\n   \"num_unique_values\": 2,\n   \"samples\":
[\n   true\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":

\"AVG\",\n   \"properties\": {\n   \"dtype\": \"boolean\",\n
\"num_unique_values\": 2,\n   \"samples\": [\n   true\n
],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   }\n]\n}","type":"dataframe"}

```
df.dropna(axis=0)
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 13,\n \"fields\": [\n
{\n   \"column\": \"SNO\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 5,\n   \"min\": 2,\n
\"max\": 20,\n   \"num_unique_values\": 12,\n   \"samples\":
[\n   18,\n   16,\n   2\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"REGNO\",\n   \"properties\": {\
n   \"dtype\": \"number\",\n   \"std\": 5,\n   \"min\":
1220122,\n   \"max\": 1220140,\n   \"num_unique_values\":
12,\n   \"samples\": [\n   1220138,\n   1220136,\n
1220122\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"NAME\",\n   \"properties\": {\n   \"dtype\": \"string\",\n
\"num_unique_values\": 12,\n   \"samples\": [\n
\"RATHI\",\n   \"PRATHAP\",\n   \"BABU\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"DOB\",\n   \"properties\": {\n
\"dtype\": \"string\",\n   \"num_unique_values\": 9,\n
\"samples\": [\n   \"20001121\",\n   \"2000-11-09\",\n
\"20001109\"\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"GENDER\",\n   \"properties\": {\n   \"dtype\":
\"category\",\n   \"num_unique_values\": 2,\n   \"samples\":
[\n   \"FEMALE\",\n   \"MALE\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"ADDRESS\",\n   \"properties\":
{\n   \"dtype\": \"category\",\n   \"num_unique_values\":
4,\n   \"samples\": [\n   \"POONAMALEE\",\n
\"THANDALAM\"\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"M1\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 20.274715844225845,\n   \"min\": 34.0,\n   \"max\":
96.0,\n   \"num_unique_values\": 12,\n   \"samples\": [\n
81.0,\n   86.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"M2\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 18.063988826050917,\n   \"min\": 45.0,\n   \"max\":
96.0,\n   \"num_unique_values\": 11,\n   \"samples\": [\n
96.0,\n   61.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"M3\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 12.828254073103736,\n   \"min\": 50.0,\n   \"max\":
96.0,\n   \"num_unique_values\": 5,\n   \"samples\": [\n

```
96.0,\n          50.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M4\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 20.274715844225845,\n        \"min\": 34.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 12,\n        \"samples\": [\n
81.0,\n          86.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"TOTAL\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 70.47003363860159,\n        \"min\": 163.0,\n        \"max\":
383.0,\n        \"num_unique_values\": 12,\n        \"samples\": [\n
338.0,\n          346.0\n        ],\n        \"semantic_type\": \"\",\
n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"AVG\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 23.490011213604618,\n        \"min\":
54.33333333,\n        \"max\": 127.6666667,\n
\"num_unique_values\": 12,\n        \"samples\": [\n
112.6666667,\n          115.3333333\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe"}
```


```
df.dropna(axis=1)
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 21,\n  \"fields\": [\n
{\n      \"column\": \"SNO\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 5,\n        \"min\": 1,\n
\"max\": 20,\n        \"num_unique_values\": 20,\n        \"samples\":
[\n          1,\n          18,\n          16\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"REGNO\",\n      \"properties\": {\
n        \"dtype\": \"number\",\n        \"std\": 5,\n        \"min\":
1220121,\n        \"max\": 1220140,\n        \"num_unique_values\":
20,\n        \"samples\": [\n          1220121,\n          1220138,\n
1220136\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"DOB\",\n      \"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 13,\n        \"samples\": [\n
\"19990305\",\n          \"20000921\",\n          \"2000-02-10\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    }\n  ]\n}","type":"dataframe"}
```


```
df.duplicated()
```

```
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
```

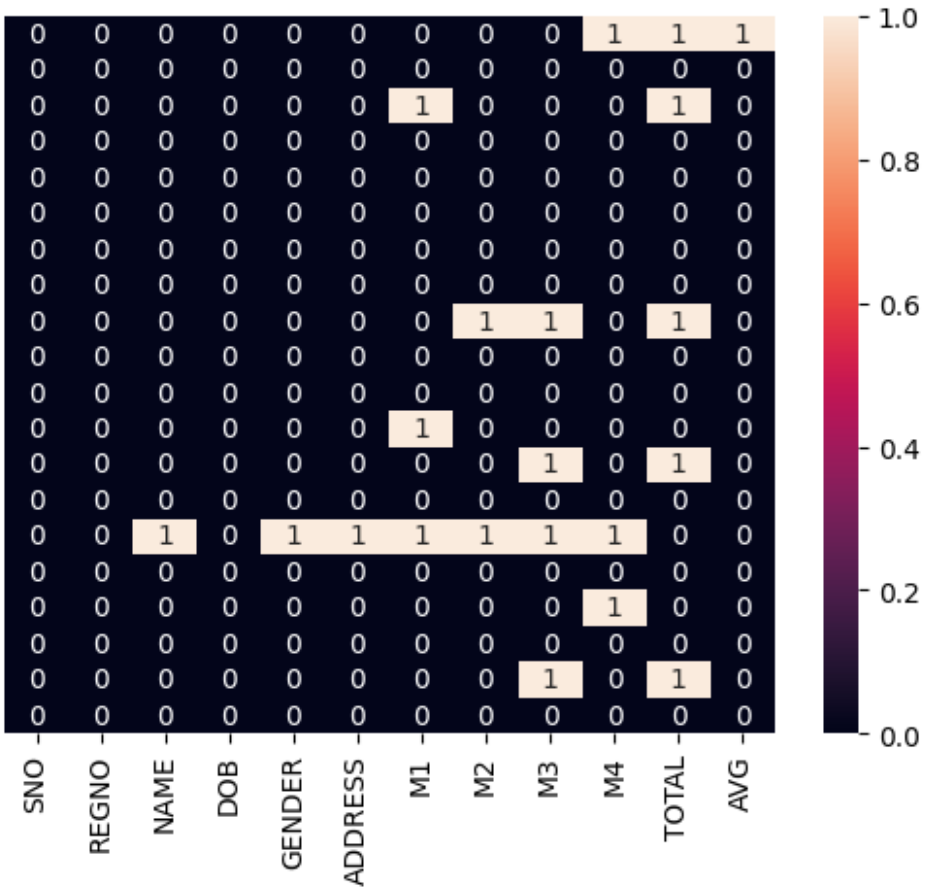
```
m=df.drop_duplicates(inplace=False)
m
```


```
{"summary":"{\n  \"name\": \"m\",\n  \"rows\": 20,\n  \"fields\": [\n
{\n      \"column\": \"SNO\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 5,\n        \"min\": 1,\n
\"max\": 20,\n        \"num_unique_values\": 20,\n        \"samples\":
[\n          1,\n          18,\n          16\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"REGNO\",\n      \"properties\": {\
n        \"dtype\": \"number\",\n        \"std\": 5,\n        \"min\":
1220121,\n        \"max\": 1220140,\n        \"num_unique_values\":
20,\n        \"samples\": [\n          1220121,\n          1220138,\n
1220136\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"NAME\",\n      \"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 19,\n        \"samples\": [\n
\"ARUN\",\n          \"FARHANA\",\n          \"LATHESSH\"\n        ],\
n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"DOB\",\n      \"properties\": {\
n        \"dtype\": \"string\",\n        \"num_unique_values\": 13,\n
\"samples\": [\n          \"19990305\",\n          \"20000921\",\n
\"2000-02-10\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"GENDER\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          \"FEMALE\",\n          \"MALE\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"ADDRESS\",\n      \"properties\":
{\n        \"dtype\": \"category\",\n        \"num_unique_values\":
4,\n        \"samples\": [\n          \"KANCHIPURAM\",\n
\"CHITHUR\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M1\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 16.97381200902281,\n        \"min\": 34.0,\n        \"max\":
```

```
96.0,\n        \"num_unique_values\": 17,\n        \"samples\": [\n
82.0,\n          56.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M2\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 15.224356989759338,\n        \"min\": 45.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 17,\n        \"samples\": [\n
81.0,\n          61.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M3\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 13.195327455833246,\n        \"min\": 50.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 6,\n        \"samples\": [\n
90.0,\n          80.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"M4\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 16.85251770333372,\n        \"min\": 34.0,\n        \"max\":
96.0,\n        \"num_unique_values\": 16,\n        \"samples\": [\n
56.0,\n          70.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"TOTAL\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 104.58330013279685,\n        \"min\": 0.0,\n        \"max\":
383.0,\n        \"num_unique_values\": 15,\n        \"samples\": [\n
315.0,\n          346.0\n        ],\n        \"semantic_type\": \"\",\
n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"AVG\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 49.33286575615893,\n        \"min\":
0.0,\n        \"max\": 127.6666667,\n        \"num_unique_values\":
15,\n        \"samples\": [\n          69.33333333,\n
115.3333333\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"m"}
```

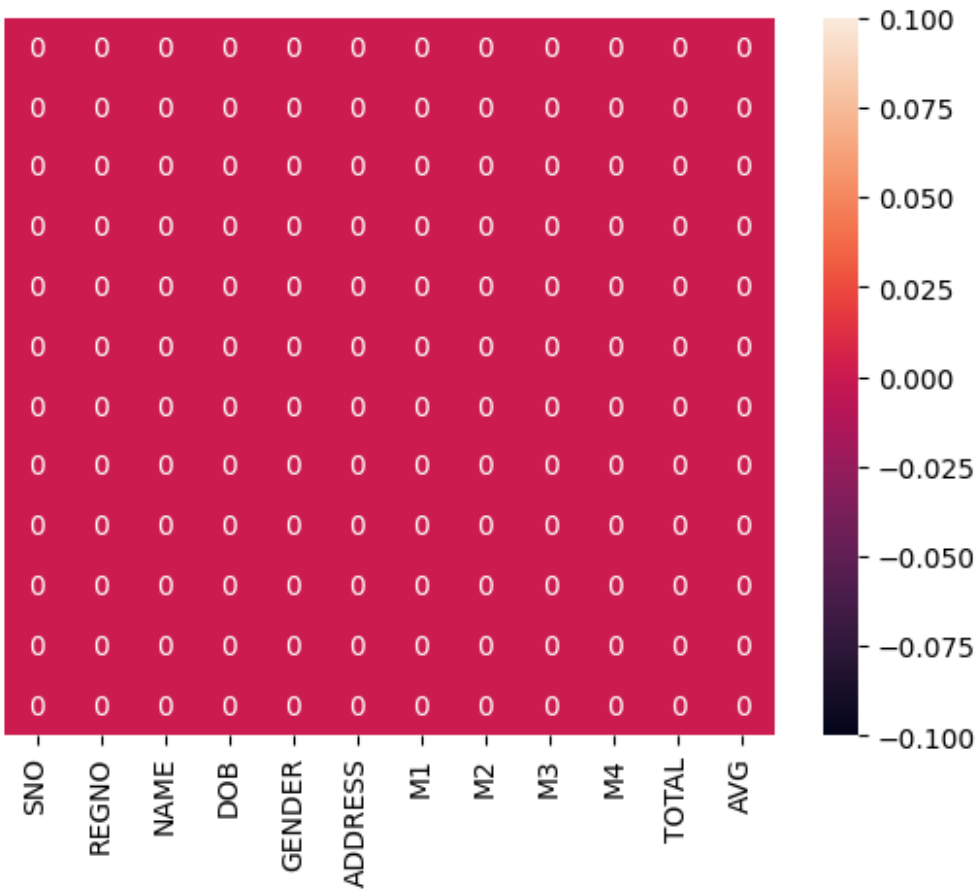

```
import seaborn as sns
sns.heatmap(df.isnull(),yticklabels=False,annot=True)
```

```
<Axes: >
```



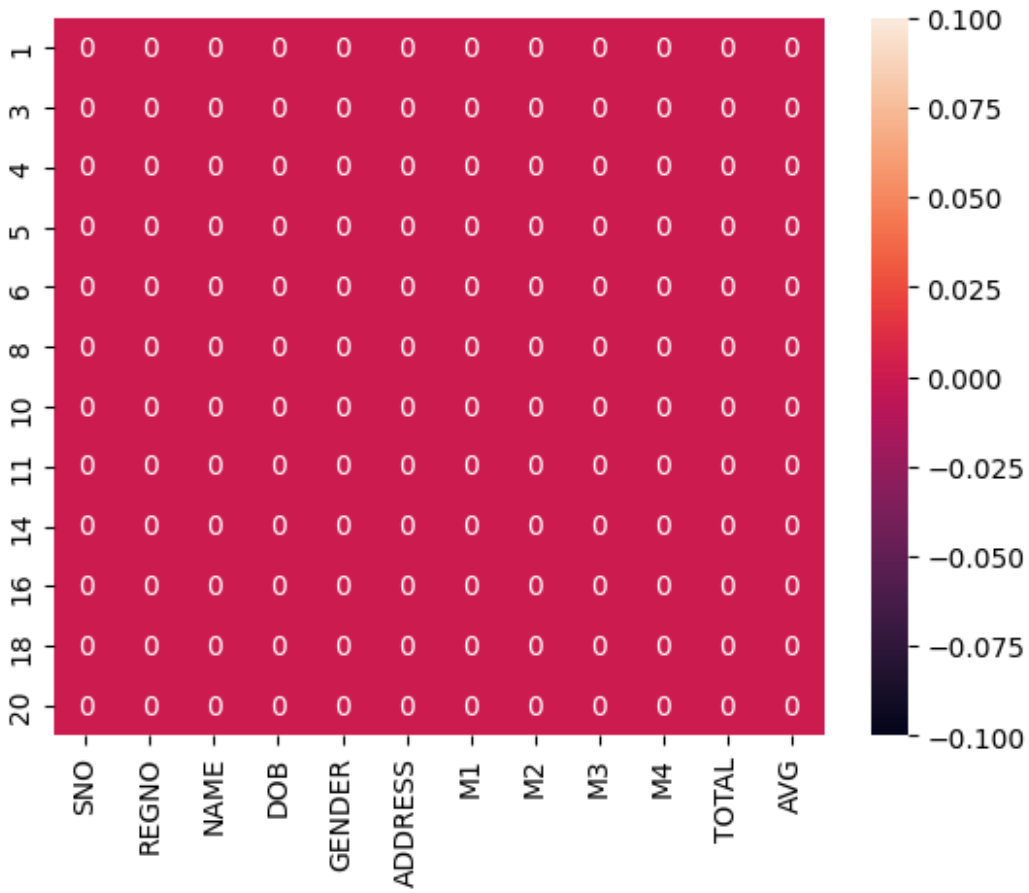

```
df.dropna(inplace=True)

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

```
<Axes: >
```




```
import seaborn as sns
sns.heatmap(df.isnull(),yticklabels=True,annot=True)
```

```
<Axes: >
```

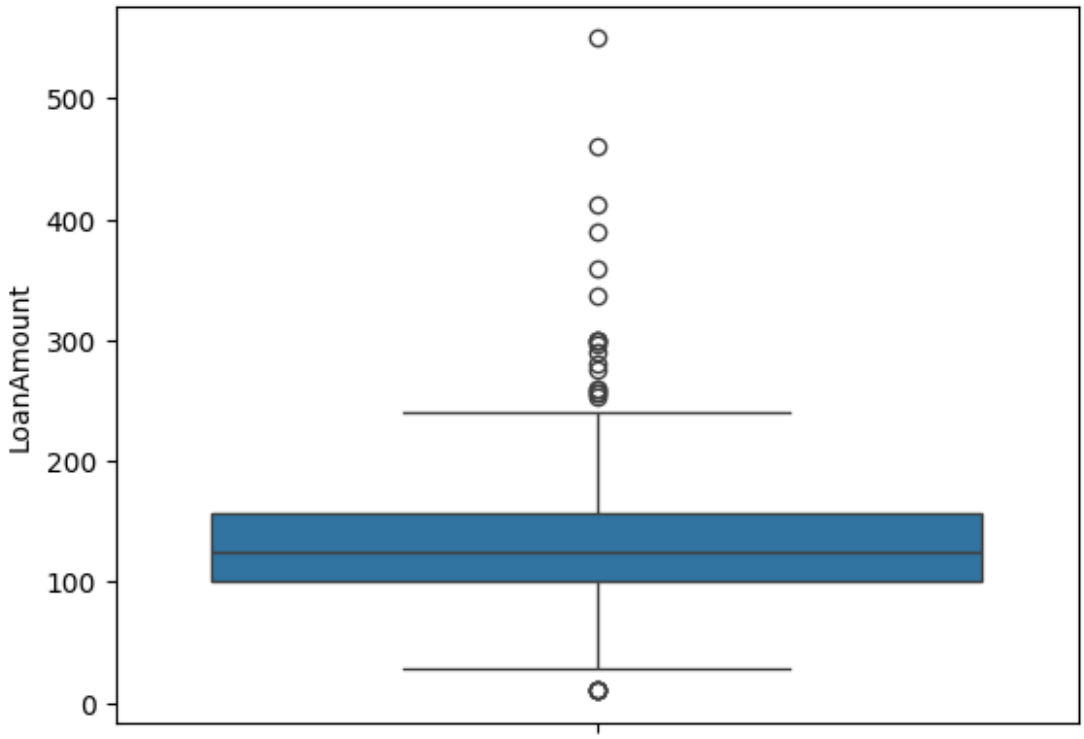



```
df.dtypes
```

```
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
```

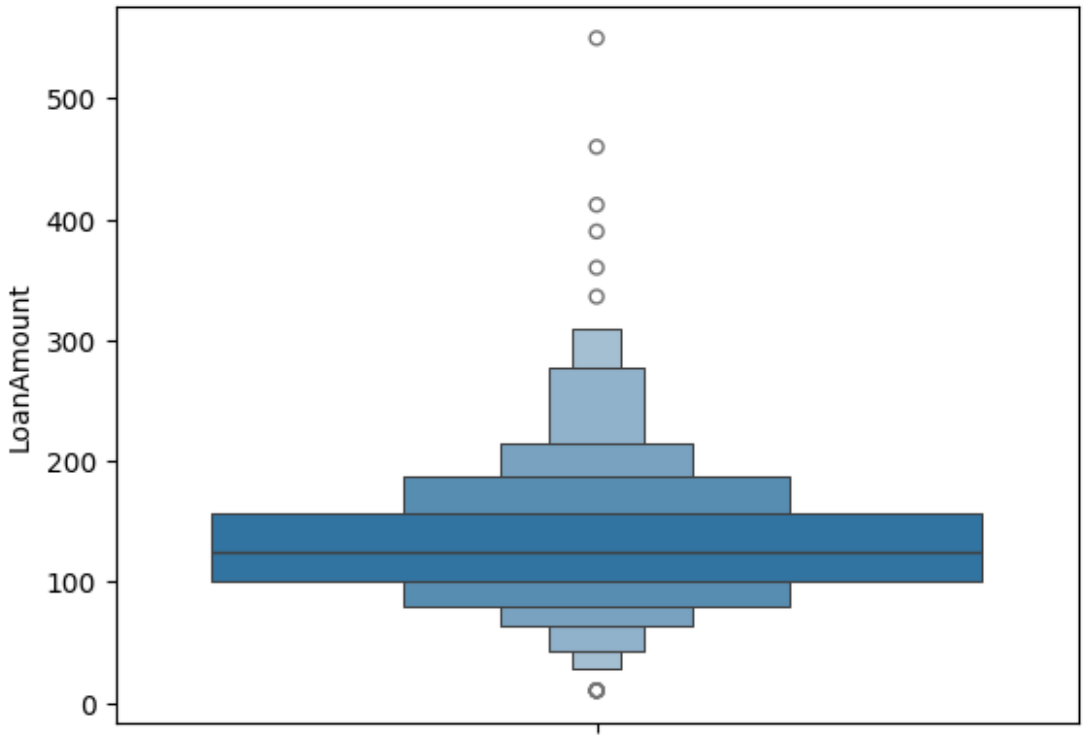
```
import pandas as pd
import numpy as np
import seaborn as sns
df=pd.read_csv("/content/Loan_data.csv")
df.fillna(10,inplace=True)
sns.boxplot(data=df['LoanAmount'])
```

```
<Axes: ylabel='LoanAmount'>
```



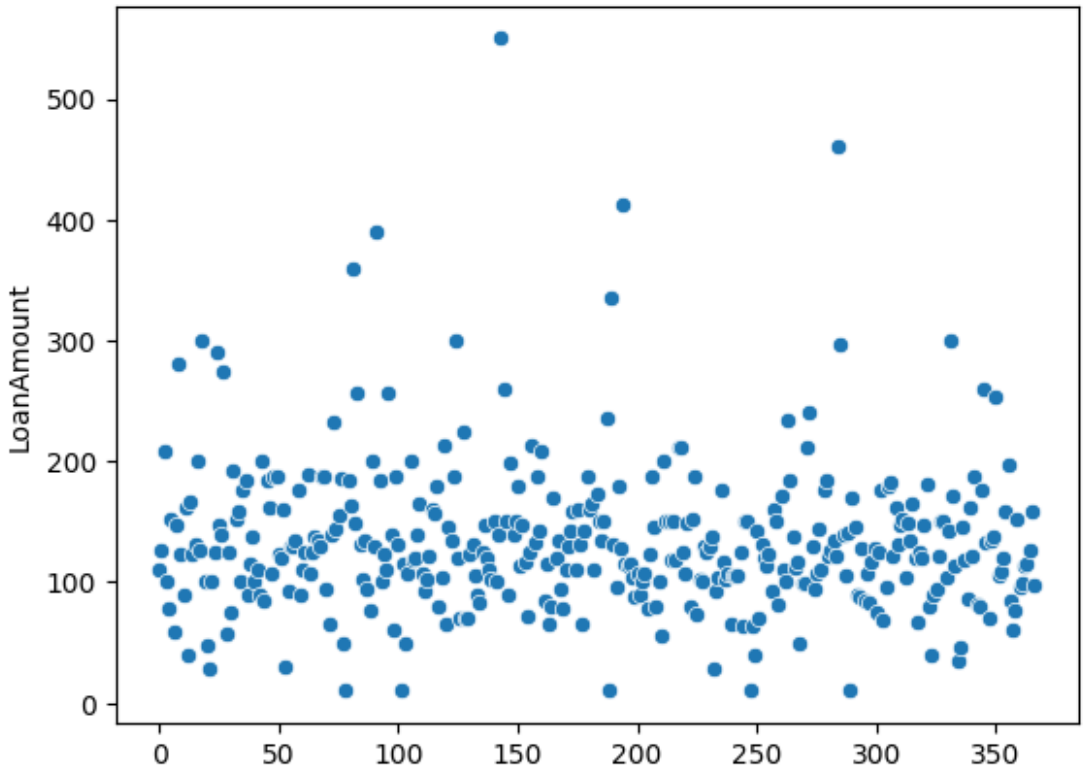

```
sns.boxenplot(data=df['LoanAmount'])
```

```
<Axes: ylabel='LoanAmount'>
```




```
sns.scatterplot(data=df['LoanAmount'])
```

```
<Axes: ylabel='LoanAmount'>
```




```
q1=np.percentile(df['LoanAmount'],25)
q3=np.percentile(df['LoanAmount'],75)
iqr=q3-q1
lower_bound=q1-1.5*iqr
upper_bound=q3+1.5*iqr
print("LOWER BOUND",lower_bound)
print("UPPERBOUND",upper_bound)
af=df[((df['LoanAmount']>=lower_bound)&(df['LoanAmount']<=upper_bound)
)]
print("AFTER REMOVING OUTLIERS",af['LoanAmount'])
sns.boxplot(data=af['LoanAmount'])
```

```
LOWER BOUND 13.75
UPPERBOUND 243.75
AFTER REMOVING OUTLIERS 0      110.0
1      126.0
2      208.0
3      100.0
4       78.0
       ...
362    113.0
363    115.0
364    126.0
365    158.0
```

```
366     98.0
Name: LoanAmount, Length: 344, dtype: float64

<Axes: ylabel='LoanAmount'>
```