```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns


from google.colab import files
import pandas as pd


df=pd.read_csv("/content/Sector_1.csv")


df.head()
```

	Academic_Year	State_Code	State_Name	District_Code	District_Name	Block_Code	Udise_Block_Name	School_Category_Id	School_Cat
0	2016-17	29	Karnataka	2902	BAGALKOT	290201	BADAMI	11	
1	2016-17	29	Karnataka	2902	BAGALKOT	290201	BADAMI	11	
2	2016-17	29	Karnataka	2902	BAGALKOT	290201	BADAMI	11	
3	2016-17	29	Karnataka	2902	BAGALKOT	290201	BADAMI	11	
4	2016-17	29	Karnataka	2902	BAGALKOT	290201	BADAMI	1	

5 rows × 45 columns

```
# Select only the necessary columns
selected_columns = [
    'District_Name',
    'Udise_Block_Name',
    'School_Category_Name',
    'School_Management_Name',
    'School_Type',
    'Total_Number_of_Schools',
    'Building',
    'Functional_Drinking_Water',
    'Functional_Electricity',
    'Functional_Boy_Toilet',
    'Functional_Girl_Toilet',
    'Library_or_Reading_Corner_or_Book_Bank',
    'Playground',
    'Computer_Available',
    'Internet',
    'Medical_Checkup',
    'Complete_Medical_Checkup',
    'Handwash',
    'Separate_Room_for_Headmaster',
    'Furniture',
    'Kitchen_Garden',
    'Water_Purifier',
    'Rain_Water_Harvesting',
    'Solar_Panel'
]

df_selected = df[selected_columns]
df_selected.head()
```

	District_Name	Udise_Block_Name	School_Category_Name	School_Management_Name	School_Type	Total_Number_of_Schools	Building	Fu
0	BAGALKOT	BADAMI	HSS (XI-XII)	Department of Education	Co-Ed	3	3	
1	BAGALKOT	BADAMI	HSS (XI-XII)	Private Unaided (Recognized)	Co-Ed	2	2	
2	BAGALKOT	BADAMI	HSS (XI-XII)	Government Aided	Co-Ed	1	1	
3	BAGALKOT	BADAMI	HSS (XI-XII)	Private Unaided (Recognized)	Co-Ed	4	4	
4	BAGALKOT	BADAMI	PS (I-V)	Government Aided	Girls	1	1	

5 rows × 24 columns

```
df_selected.isnull().sum()


df_selected = df_selected.drop_duplicates()
```

```
df_selected = df_selected.fillna({'Functional_Drinking_Water': 'No',
                                  'Functional_Electricity': 'No'})



# Save the cleaned dataset for dashboard use
df_selected.to_csv("Bagalkot_Education_Cleaned.csv", index=False)

# Download the cleaned file
files.download("Bagalkot_Education_Cleaned.csv")
```

```
import pandas as pd

# Load your cleaned dataset
df = pd.read_csv("Bagalkot_Education_Cleaned.csv")

# Show column names
print(df.columns.tolist())

# Show first 5 rows
print(df.head())
```

```
['District_Name', 'Udise_Block_Name', 'School_Category_Name', 'School_Management_Name', 'School_Type', 'Total_Number_of_Schools', 'E
  District_Name Udise_Block_Name School_Category_Name  \
0      BAGALKOT           BADAMI         HSS (XI-XII)
1      BAGALKOT           BADAMI         HSS (XI-XII)
2      BAGALKOT           BADAMI         HSS (XI-XII)
3      BAGALKOT           BADAMI         HSS (XI-XII)
4      BAGALKOT           BADAMI             PS (I-V)

         School_Management_Name School_Type  Total_Number_of_Schools  \
0       Department of Education       Co-Ed                        3
1  Private Unaided (Recognized)       Co-Ed                        2
2             Government Aided       Co-Ed                        1
3  Private Unaided (Recognized)       Co-Ed                        4
4             Government Aided       Girls                        1

   Building  Functional_Drinking_Water  Functional_Electricity  \
0         3                         3                       2
1         2                         2                       2
2         1                         1                       1
3         4                         4                       4
4         1                         1                       1

   Functional_Boy_Toilet  ...  Internet  Medical_Checkup  \
0                      3  ...         0                0
1                      2  ...         0                0
2                      1  ...         0                1
3                      4  ...         1                2
4                      1  ...         0                1

   Complete_Medical_Checkup  Handwash  Separate_Room_for_Headmaster  \
0                         0         1                             1
1                         0         1                             2
2                         0         1                             1
3                         0         4                             4
4                         0         0                             0

   Furniture  Kitchen_Garden  Water_Purifier  Rain_Water_Harvesting  \
0          3               0               0                      0
1          2               0               0                      0
2          1               0               0                      0
3          4               0               0                      0
4          1               0               0                      0

   Solar_Panel
0            0
1            0
2            0
3            0
4            0

[5 rows x 24 columns]
```

```
import pandas as pd

df = pd.read_csv("Bagalkot_Education_Cleaned.csv")
print(df.columns.tolist())
df.head()
```

```
['District_Name', 'Udise_Block_Name', 'School_Category_Name', 'School_Management_Name', 'School_Type', 'Total_Number_of_Schools', 'B
```

	District_Name	Udise_Block_Name	School_Category_Name	School_Management_Name	School_Type	Total_Number_of_Schools	Building	Fu
0	BAGALKOT	BADAMI	HSS (XI-XII)	Department of Education	Co-Ed	3	3	
1	BAGALKOT	BADAMI	HSS (XI-XII)	Private Unaided (Recognized)	Co-Ed	2	2	
2	BAGALKOT	BADAMI	HSS (XI-XII)	Government Aided	Co-Ed	1	1	
3	BAGALKOT	BADAMI	HSS (XI-XII)	Private Unaided (Recognized)	Co-Ed	4	4	
4	BAGALKOT	BADAMI	PS (I-V)	Government Aided	Girls	1	1	

5 rows × 24 columns

```
df.describe()
```

	Total_Number_of_Schools	Building	Functional_Drinking_Water	Functional_Electricity	Functional_Boy_Toilet	Functional_Gir
count	234.000000	234.000000	234.000000	234.000000	234.000000	23
mean	9.914530	9.478632	9.384615	9.277778	9.170940	
std	19.862866	19.629879	19.462673	19.137064	18.935985	1
min	1.000000	0.000000	0.000000	0.000000	0.000000	
25%	1.000000	1.000000	1.000000	1.000000	1.000000	
50%	3.000000	3.000000	3.000000	3.000000	3.000000	
75%	9.000000	8.000000	8.000000	8.000000	8.000000	
max	142.000000	142.000000	141.000000	141.000000	137.000000	13

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 234 entries, 0 to 233
Data columns (total 24 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   District_Name                           234 non-null    object
 1   Udise_Block_Name                        234 non-null    object
 2   School_Category_Name                    234 non-null    object
 3   School_Management_Name                  234 non-null    object
 4   School_Type                             234 non-null    object
 5   Total_Number_of_Schools                 234 non-null    int64
 6   Building                                234 non-null    int64
 7   Functional_Drinking_Water               234 non-null    int64
 8   Functional_Electricity                  234 non-null    int64
 9   Functional_Boy_Toilet                   234 non-null    int64
 10  Functional_Girl_Toilet                  234 non-null    int64
 11  Library_or_Reading_Corner_or_Book_Bank  234 non-null    int64
 12  Playground                              234 non-null    int64
 13  Computer_Available                      234 non-null    int64
 14  Internet                                234 non-null    int64
 15  Medical_Checkup                         234 non-null    int64
 16  Complete_Medical_Checkup                234 non-null    int64
 17  Handwash                                234 non-null    int64
 18  Separate_Room_for_Headmaster            234 non-null    int64
 19  Furniture                               234 non-null    int64
 20  Kitchen_Garden                          234 non-null    int64
 21  Water_Purifier                          234 non-null    int64
 22  Rain_Water_Harvesting                   234 non-null    int64
 23  Solar_Panel                             234 non-null    int64
dtypes: int64(19), object(5)
memory usage: 44.0+ KB
```