

Tamil Nadu PDS (Public Distribution System) Analysis - Dataset - 2019

Project Description:

This project focuses on analyzing the Public Distribution System (PDS) in Tamil Nadu, India. The PDS is a crucial government initiative to provide essential commodities to the public at subsidized rates. The analysis is performed using data from the Tamil Nadu PDS, which includes information about districts, taluks, villages, and PDS shops.

```
In [3]: import pandas as pd
import matplotlib.pyplot as plt
```

```
In [4]: df = pd.read_csv("tnpds.csv")
```

```
In [5]: df.head()
```

Out[5]:

	Sl.No	District Name	Taluk Name	Village Name	Shop Code	Shop Name
0	1	Ariyalur	Ariyalur	Ariyalur	15DA001PN	Amutham
1	2	Ariyalur	Ariyalur	Ariyalur	15DB001PN	Amaravathi-1
2	3	Ariyalur	Ariyalur	Ariyalur	15DB002PN	Amaravathi-2
3	4	Ariyalur	Ariyalur	Ariyalur	15DB004PN	Amaravathi-III
4	5	Ariyalur	Ariyalur	Ariyalur	15DE001PN	A.C.M.S.1

```
In [6]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 34773 entries, 0 to 34772
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype  
---  -
0   Sl.No           34773 non-null  int64  
1   District Name   34773 non-null  object  
2   Taluk Name      34773 non-null  object  
3   Village Name    34773 non-null  object  
4   Shop Code       34773 non-null  object  
5   Shop Name       34773 non-null  object  
dtypes: int64(1), object(5)
memory usage: 1.6+ MB
```

```
In [7]: district_counts = df["District Name"].nunique()
taluk_counts = df["Taluk Name"].nunique()
village_counts = df["Village Name"].nunique()
shop_counts = df["Shop Name"].nunique()

print(district_counts,"Districts")
print(taluk_counts,"Taluks")
print(village_counts,"Villages")
print(shop_counts,"PDS Shops")
```

```
32 Districts
288 Taluks
16349 Villages
31028 PDS Shops
```

```
In [8]: districts = df["District Name"].unique()
```

```
In [9]: for i in districts:
    district_name = df[df["District Name"]==i]
    taluk_counts = district_name["Taluk Name"].nunique()
    village_counts = district_name["Village Name"].nunique()
    pds_counts = district_name["Shop Name"].nunique()

    print(i)
    print(taluk_counts,"Taluks")
    print(village_counts,"Villages")
    print(pds_counts,"PDS Shops")
    print("")
```

```
770 PDS Shops
```

```
Salem
13 Taluks
722 Villages
1483 PDS Shops
```

```
Sivagangai
9 Taluks
534 Villages
808 PDS Shops
```

```
Thanjavur
9 Taluks
844 Villages
1161 PDS Shops
```

```
Theni
5 Taluks
96 Villages
```

```
In [10]: district_data = df.groupby('District Name').agg(
        {
            'Taluk Name': 'nunique'
        }
    )

    district_data = district_data.rename(columns={
        'Taluk Name': 'Taluks'
    })

    top_5_districts_taluks = district_data.sort_values(by=['Taluks'], ascending=False)

    print("Top 5 Districts with Maximum Taluks")
    print(top_5_districts_taluks)
```

Top 5 Districts with Maximum Taluks

District Name	Taluks
Chennai	17
Tirunelveli	15
Viluppuram	13
Vellore	13
Salem	13

```
In [11]: district_data = df.groupby('District Name').agg(
        {
            'Village Name': 'nunique',
        }
    )

    district_data = district_data.rename(columns={
        'Village Name': 'Villages',
    })

    top_5_districts_villages = district_data.sort_values(by=['Villages'], ascending=False)

    print("Top 5 Districts with Maximum Villages")
    print(top_5_districts_villages)
```

Top 5 Districts with Maximum Villages

District Name	Villages
Viluppuram	1411
Thiruvannamalai	1385
Cuddalore	1161
Vellore	1097
Kanchipuram	1017

In [12]:

```
district_data = df.groupby('District Name').agg(  
    {  
        'Shop Name': 'nunique'  
    }  
)  
  
district_data = district_data.rename(columns={  
    'Shop Name': 'PDS Shops'  
})  
  
top_5_districts_pds_shop = district_data.sort_values(by=['PDS Shops'], ascending=False)  
  
print("Top 5 Districts with Maximum PDS Shops")  
print(top_5_districts_pds_shop)
```

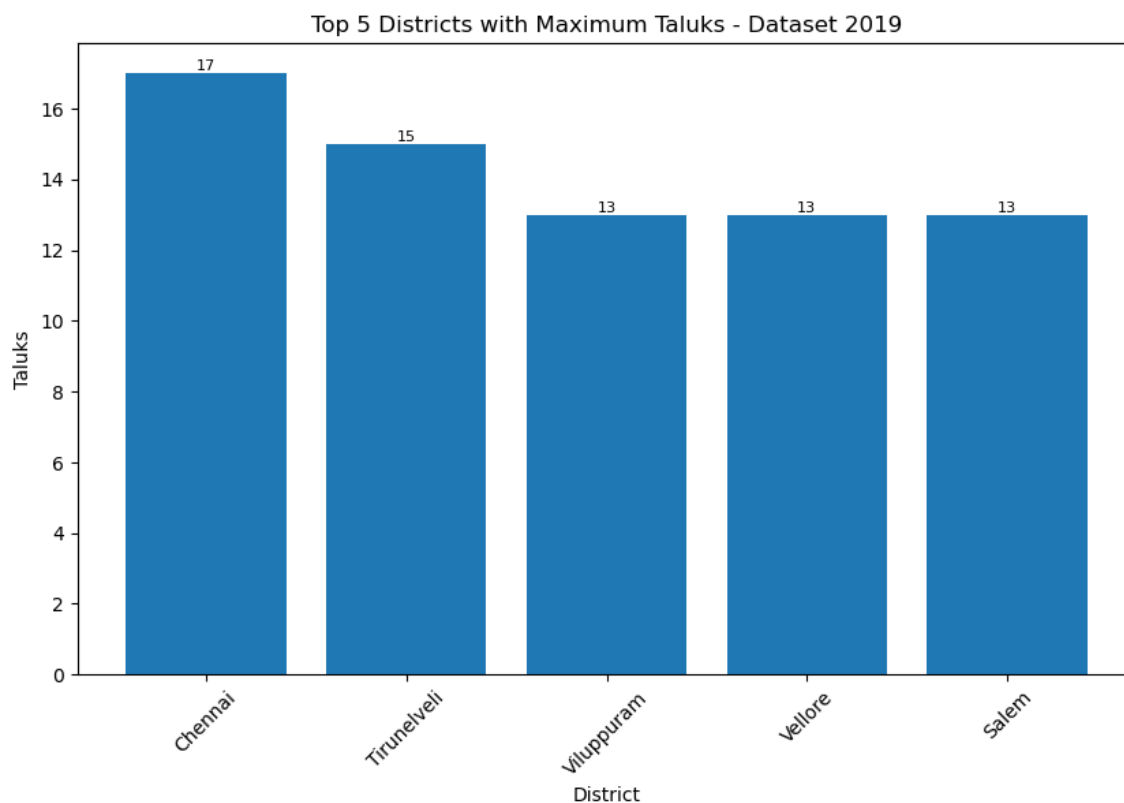
Top 5 Districts with Maximum PDS Shops

District Name	PDS Shops
Viluppuram	1968
Vellore	1735
Chennai	1702
Salem	1483
Tirunelveli	1398

```
In [13]: data = top_5_districts_taluks
plt.figure(figsize=(10, 6))
bars = plt.bar(data.index, data['Taluks'])
plt.title('Top 5 Districts with Maximum Taluks - Dataset 2019')
plt.xlabel('District')
plt.ylabel('Taluks')
plt.xticks(rotation=45)

for bar in bars:
    yval = bar.get_height()
    plt.text(bar.get_x() + bar.get_width() / 2, yval, int(yval), ha='center')

plt.show()
```



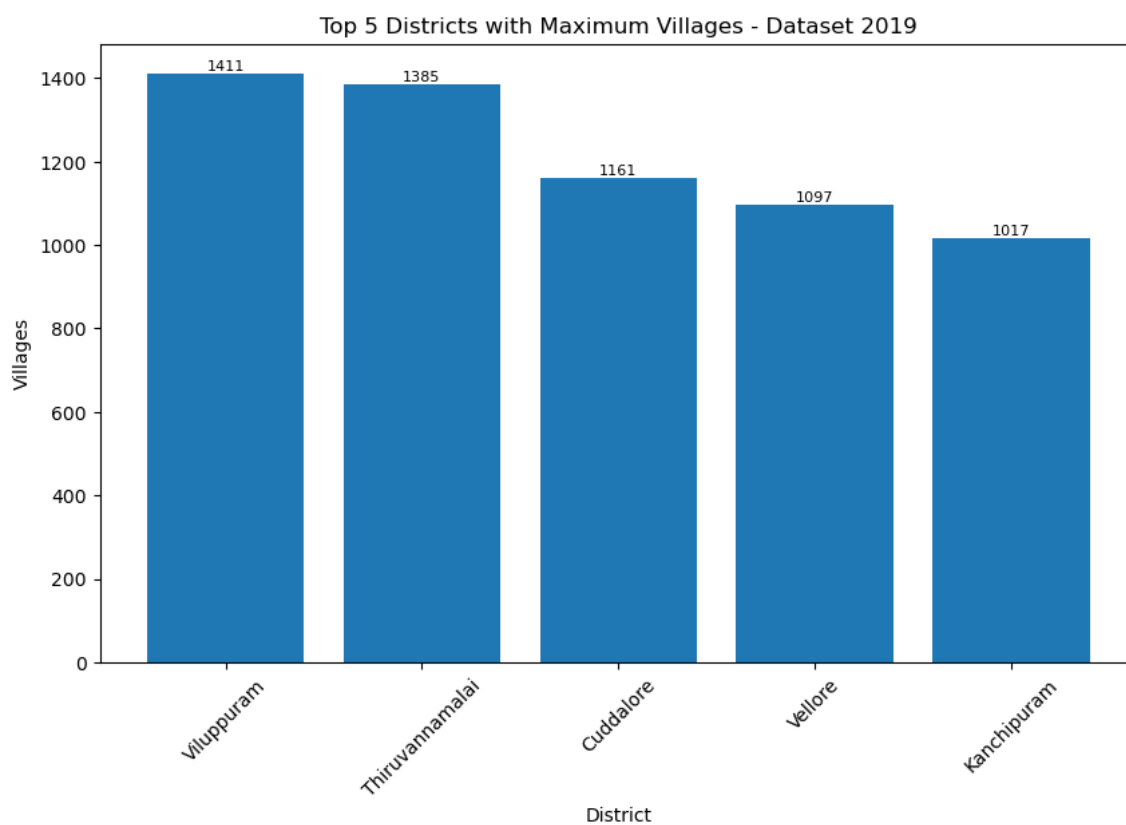
```

In [14]: data = top_5_districts_villages
plt.figure(figsize=(10, 6))
bars = plt.bar(data.index, data['Villages'])
plt.title('Top 5 Districts with Maximum Villages - Dataset 2019')
plt.xlabel('District')
plt.ylabel('Villages')
plt.xticks(rotation=45)

for bar in bars:
    yval = bar.get_height()
    plt.text(bar.get_x() + bar.get_width() / 2, yval, int(yval), ha='center')

plt.show()

```



```
In [15]: data = top_5_districts_pds_shop
plt.figure(figsize=(10, 6))
bars = plt.bar(data.index, data['PDS Shops'])
plt.title('Top 5 Districts with Maximum PDS Shops - Dataset 2019')
plt.xlabel('District')
plt.ylabel('PDS Shops')
plt.xticks(rotation=45)

for bar in bars:
    yval = bar.get_height()
    plt.text(bar.get_x() + bar.get_width() / 2, yval, int(yval), ha='center')

plt.show()
```

