Assessment of marginal workers in Tamilnadu

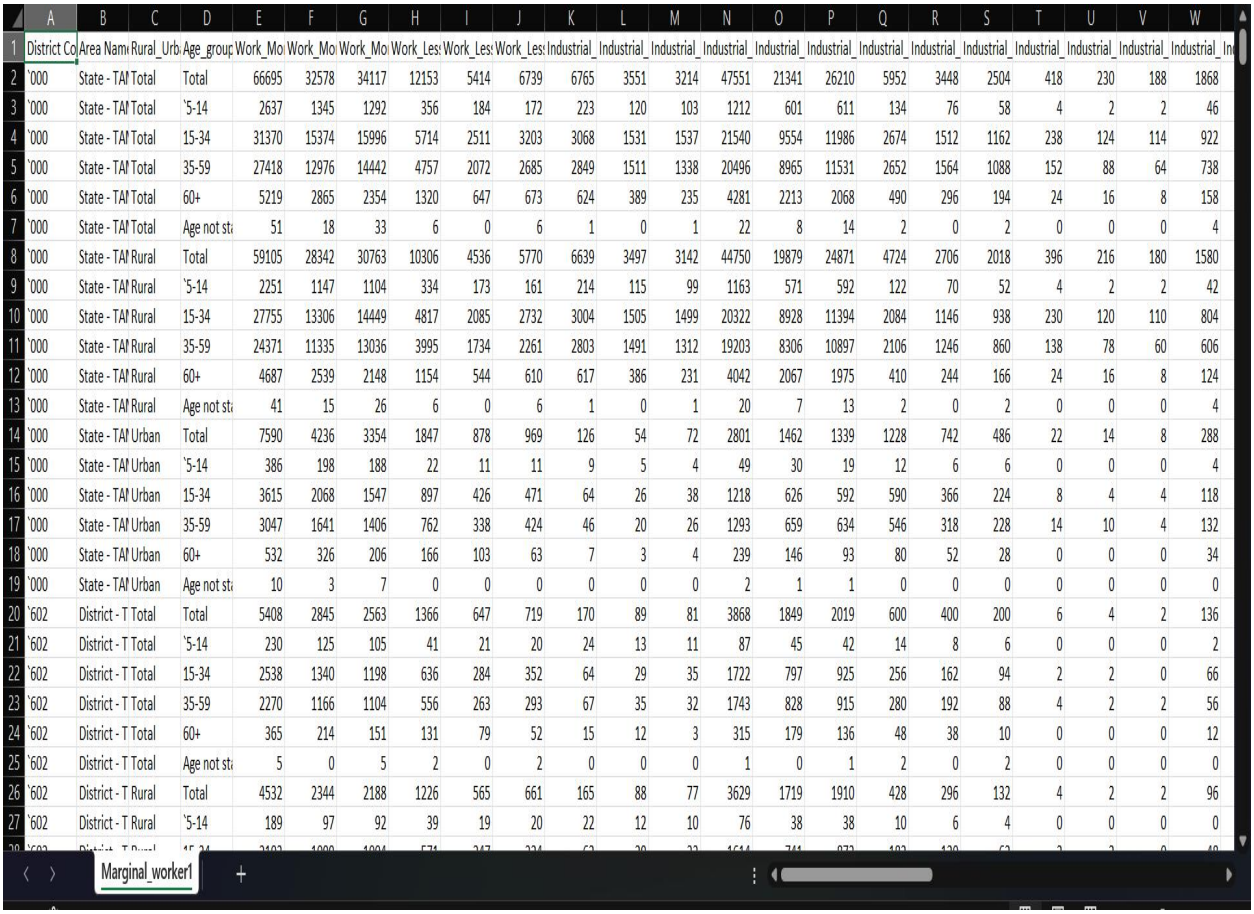
Team member : Karthikraja S

723721205021

Phase 4

Data collection:

The data is collected form government website and the data features are modified for my convenient.(website: https://tn.data.gov.in/catalog/marginal-workers-classified-age-industrial-category-and-sex-census-2011-india-and-states.)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	District Co	Area Nam	Rural_Urb	Age_group	Work_Mo	Work_Mo	Work_Mo	Work_Les	Work_Les	Work_Les	Industrial_	Industrial_	Industrial_	Industrial_	Industrial_	Industrial_	Industrial_	Industrial_	Industrial_	Industrial_	Industrial_	Industrial_	Industrial_
2	'000	State - TAI	Total	Total	66695	32578	34117	12153	5414	6739	6765	3551	3214	47551	21341	26210	5952	3448	2504	418	230	188	1868
3	'000	State - TAI	Total	'5-14	2637	1345	1292	356	184	172	223	120	103	1212	601	611	134	76	58	4	2	2	46
4	'000	State - TAI	Total	15-34	31370	15374	15996	5714	2511	3203	3068	1531	1537	21540	9554	11986	2674	1512	1162	238	124	114	922
5	'000	State - TAI	Total	35-59	27418	12976	14442	4757	2072	2685	2849	1511	1338	20496	8965	11531	2652	1564	1088	152	88	64	738
6	'000	State - TAI	Total	60+	5219	2865	2354	1320	647	673	624	389	235	4281	2213	2068	490	296	194	24	16	8	158
7	'000	State - TAI	Total	Age not st	51	18	33	6	0	6	1	0	1	22	8	14	2	0	2	0	0	0	4
8	'000	State - TAI	Rural	Total	59105	28342	30763	10306	4536	5770	6639	3497	3142	44750	19879	24871	4724	2706	2018	396	216	180	1580
9	'000	State - TAI	Rural	'5-14	2251	1147	1104	334	173	161	214	115	99	1163	571	592	122	70	52	4	2	2	42
10	'000	State - TAI	Rural	15-34	27755	13306	14449	4817	2085	2732	3004	1505	1499	20322	8928	11394	2084	1146	938	230	120	110	804
11	'000	State - TAI	Rural	35-59	24371	11335	13036	3995	1734	2261	2803	1491	1312	19203	8306	10897	2106	1246	860	138	78	60	606
12	'000	State - TAI	Rural	60+	4687	2539	2148	1154	544	610	617	386	231	4042	2067	1975	410	244	166	24	16	8	124
13	'000	State - TAI	Rural	Age not st	41	15	26	6	0	6	1	0	1	20	7	13	2	0	2	0	0	0	4
14	'000	State - TAI	Urban	Total	7590	4236	3354	1847	878	969	126	54	72	2801	1462	1339	1228	742	486	22	14	8	288
15	'000	State - TAI	Urban	'5-14	386	198	188	22	11	11	9	5	4	49	30	19	12	6	6	0	0	0	4
16	'000	State - TAI	Urban	15-34	3615	2068	1547	897	426	471	64	26	38	1218	626	592	590	366	224	8	4	4	118
17	'000	State - TAI	Urban	35-59	3047	1641	1406	762	338	424	46	20	26	1293	659	634	546	318	228	14	10	4	132
18	'000	State - TAI	Urban	60+	532	326	206	166	103	63	7	3	4	239	146	93	80	52	28	0	0	0	34
19	'000	State - TAI	Urban	Age not st	10	3	7	0	0	0	0	0	0	2	1	1	0	0	0	0	0	0	0
20	'602	District - T	Total	Total	5408	2845	2563	1366	647	719	170	89	81	3868	1849	2019	600	400	200	6	4	2	136
21	'602	District - T	Total	'5-14	230	125	105	41	21	20	24	13	11	87	45	42	14	8	6	0	0	0	2
22	'602	District - T	Total	15-34	2538	1340	1198	636	284	352	64	29	35	1722	797	925	256	162	94	2	2	0	66
23	'602	District - T	Total	35-59	2270	1166	1104	556	263	293	67	35	32	1743	828	915	280	192	88	4	2	2	56
24	'602	District - T	Total	60+	365	214	151	131	79	52	15	12	3	315	179	136	48	38	10	0	0	0	12
25	'602	District - T	Total	Age not st	5	0	5	2	0	2	0	0	0	1	0	1	2	0	2	0	0	0	0
26	'602	District - T	Rural	Total	4532	2344	2188	1226	565	661	165	88	77	3629	1719	1910	428	296	132	4	2	2	96
27	'602	District - T	Rural	'5-14	189	97	92	39	19	20	22	12	10	76	38	38	10	6	4	0	0	0	0

Marginal_worker1

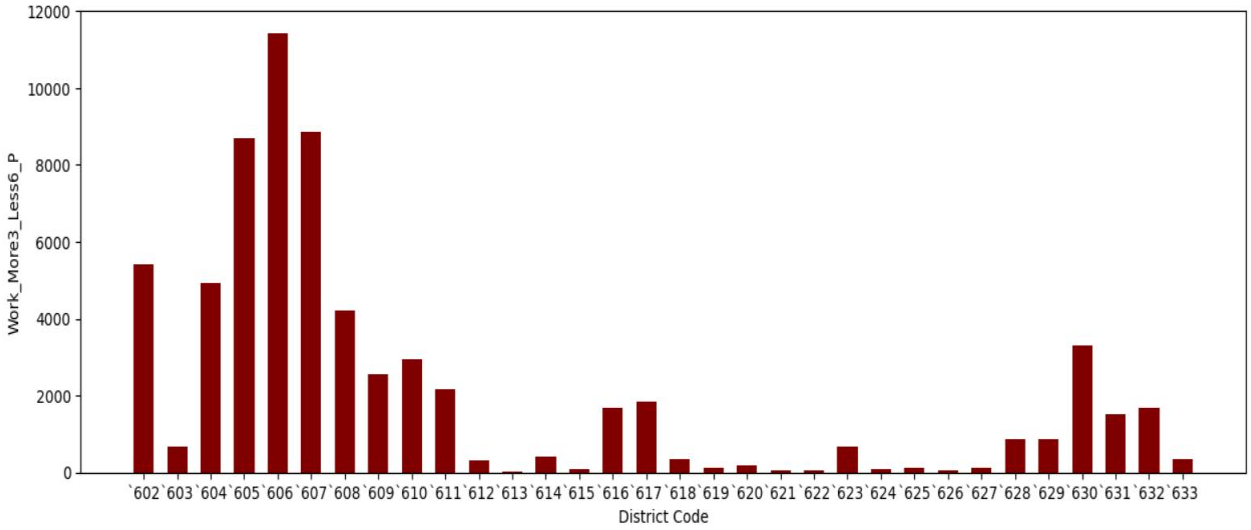
Program:

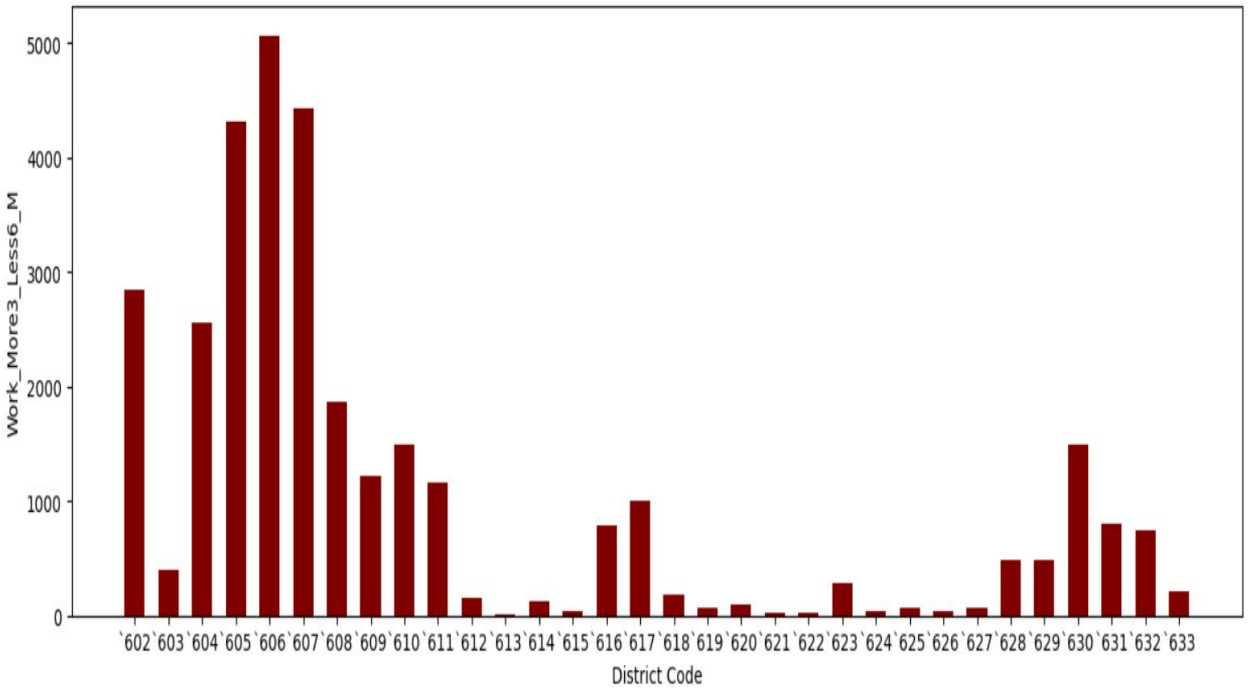
The program visualize the data form the given dataset. the visualization shows the relationship between the Area Code and number of workers worked in Tamil Nadu.

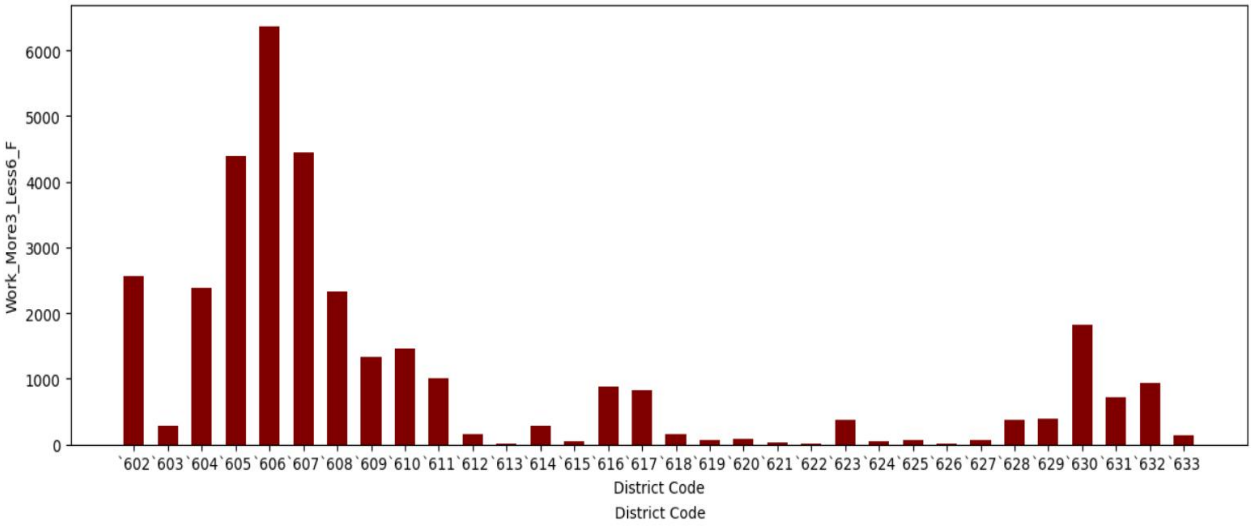
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

#Data collection and cleaning

df=pd.read_csv("/content/drive/MyDrive/Marginal_worker1.csv")
df=df[df["District Code"]!="`000"]
df=df.reset_index(drop=True)
tot_condition=df["Age_group"]=="Total"
Total_people=df[tot_condition]
con1=df["Rural_Urban"]=="Total"
con2=df["Age_group"]=="Total"
district_tot=Total_people[con1==con2]

#Data Visualization
cols=list(df.columns)
cols.remove("District Code")
cols.remove("Area Name")
cols.remove("Rural_Urban")
cols.remove("Age_group")
cols
for i in cols:
  plt.figure().set_figwidth(15)
  plt.bar(df["District Code"], df[i], color ='maroon', width = 0.6)
  plt.xlabel("District Code")
  plt.ylabel(i)
  plt.show()
```

Output:

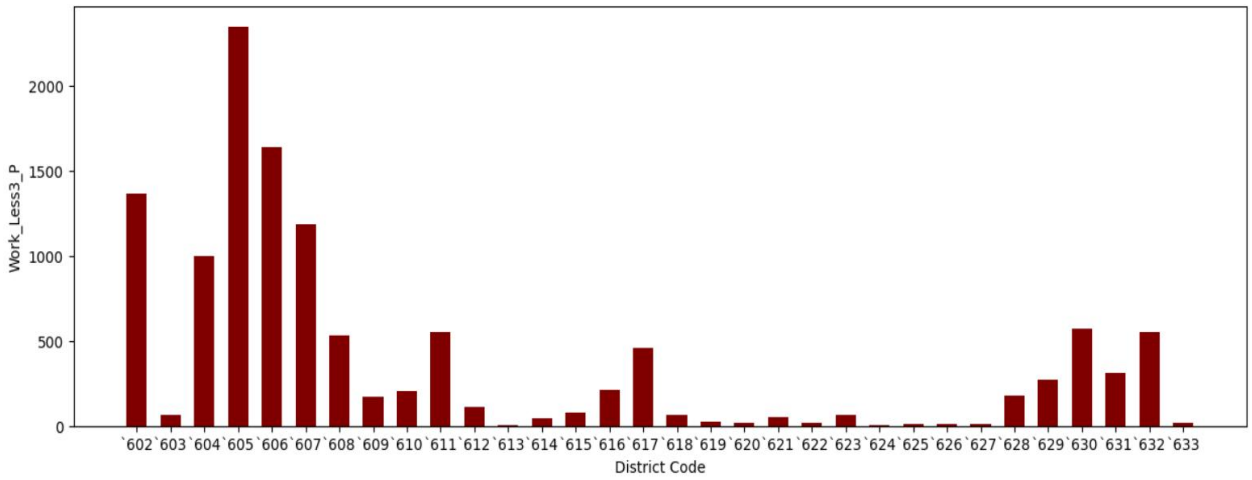
The output shows in Bar chart.

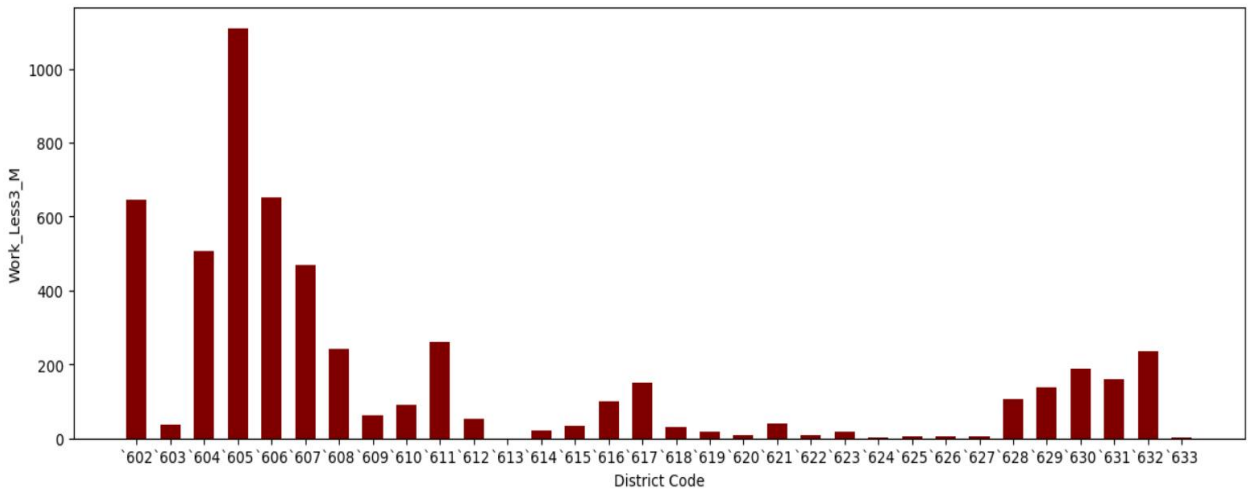


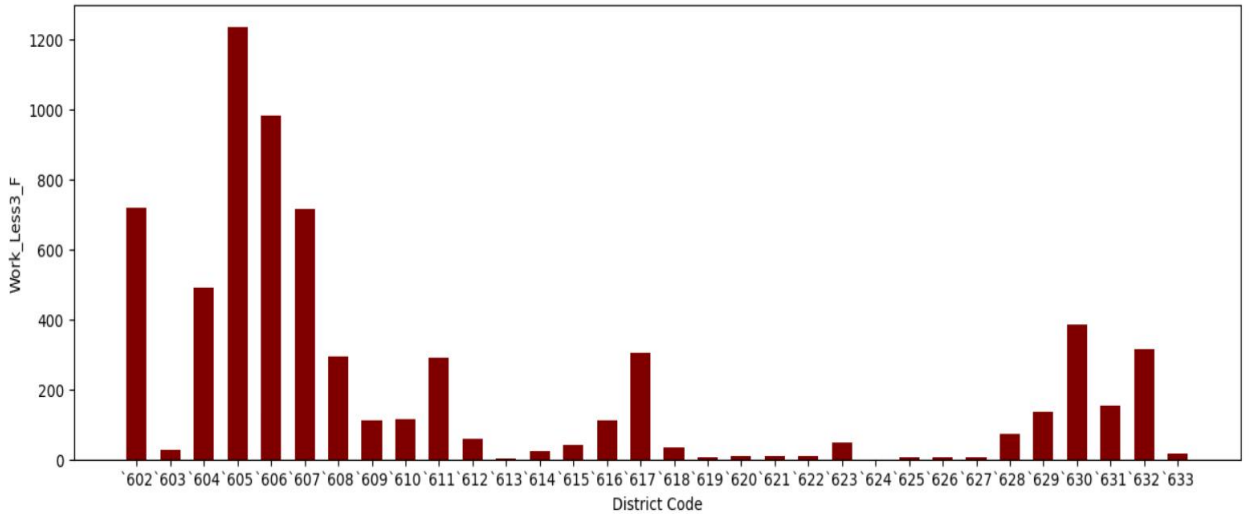


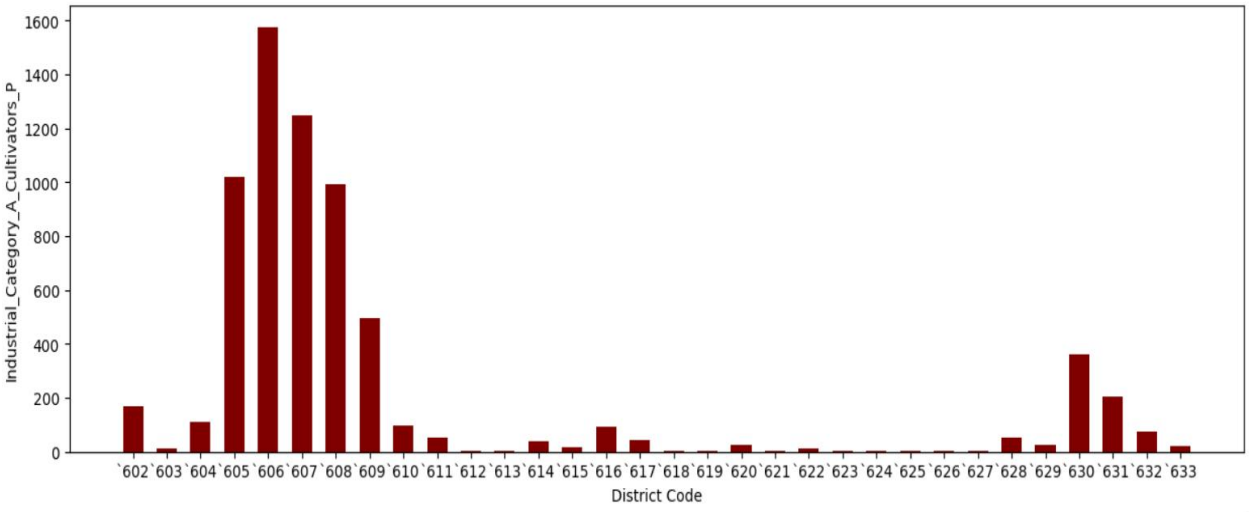


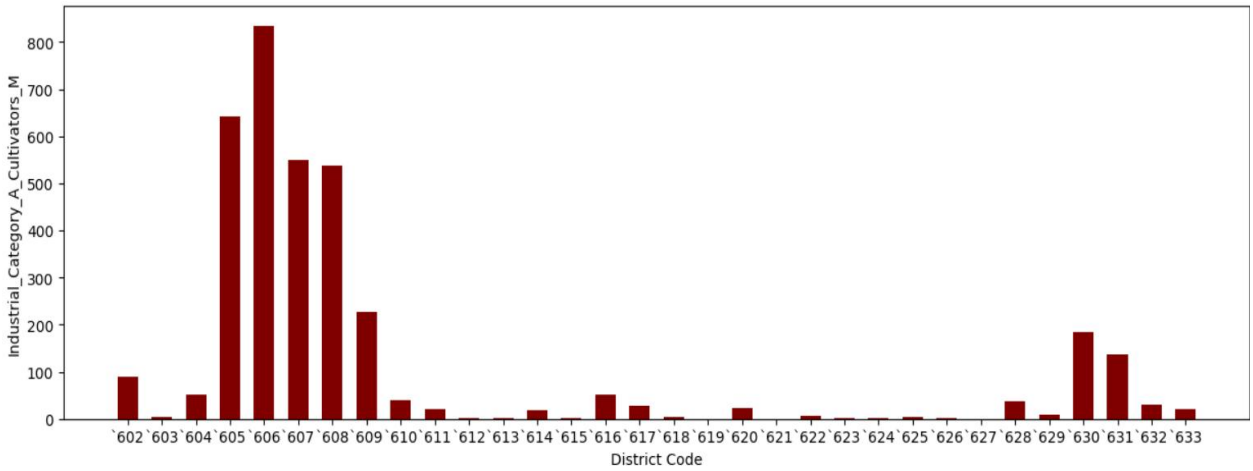
District Code

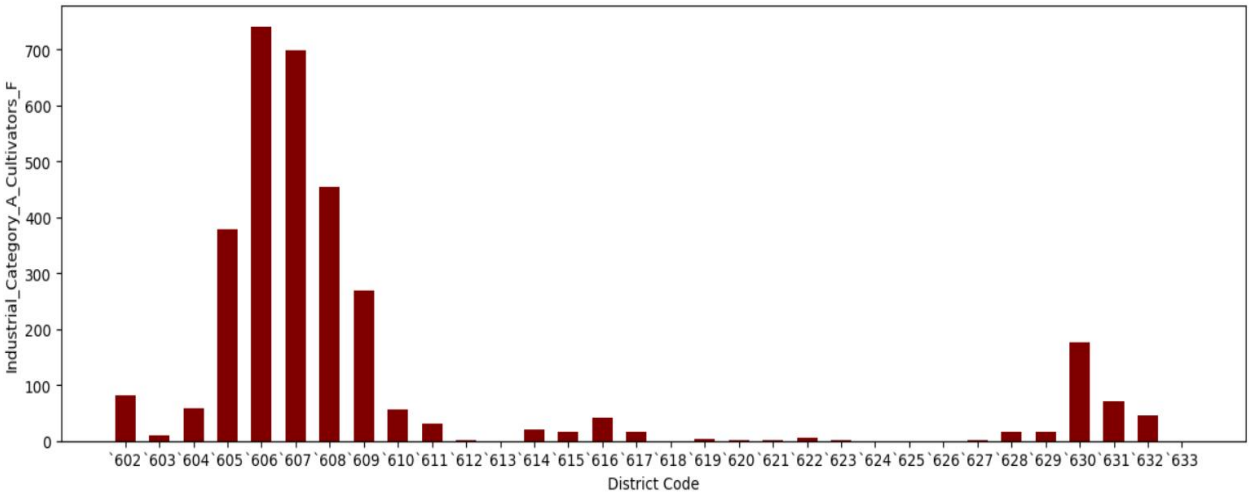


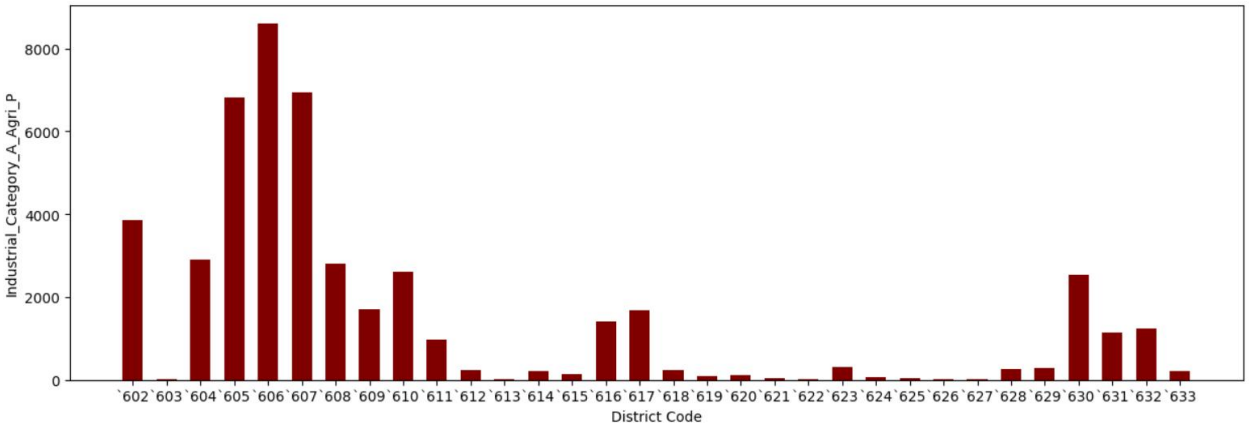


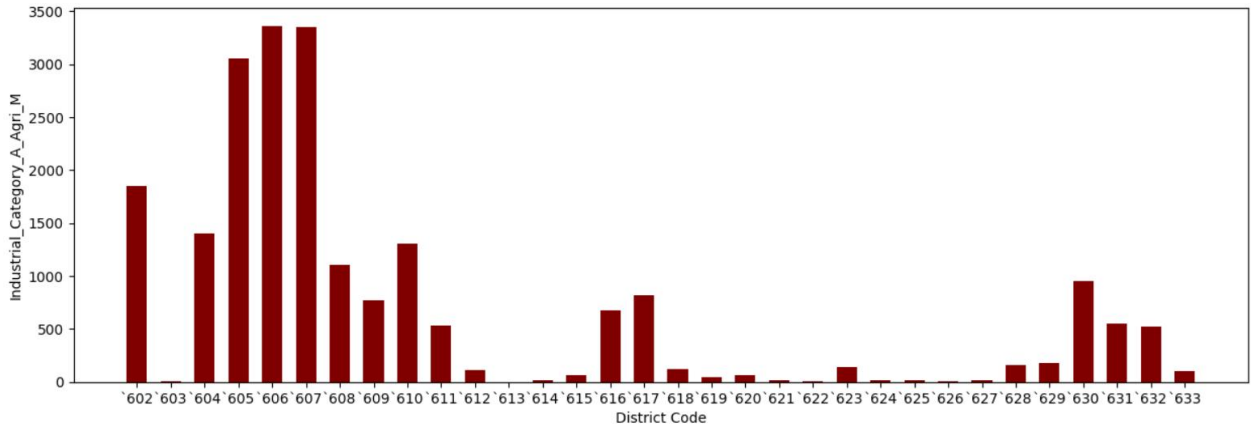

Work_Less3_F
0
200
400
600
800
1000
1200
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

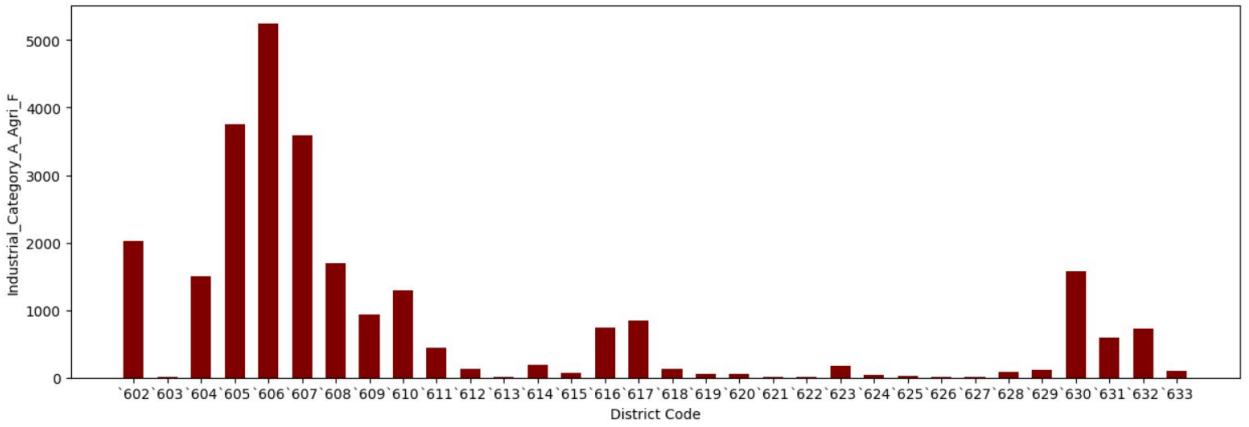

Industrial_Category_A_Cultivators_P
0
200
400
600
800
1000
1200
1400
1600
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

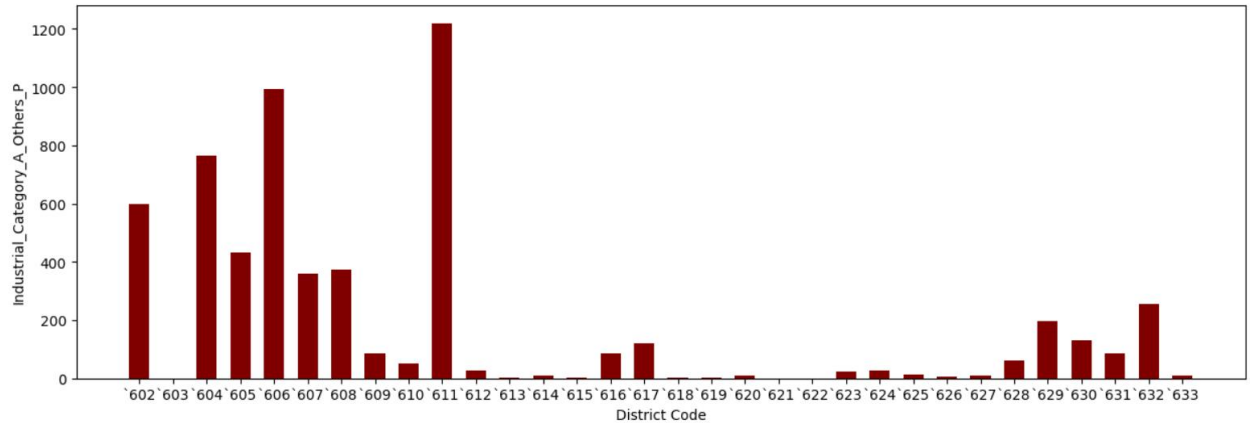

Industrial_Category_A_Cultivators_M
0
100
200
300
400
500
600
700
800
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

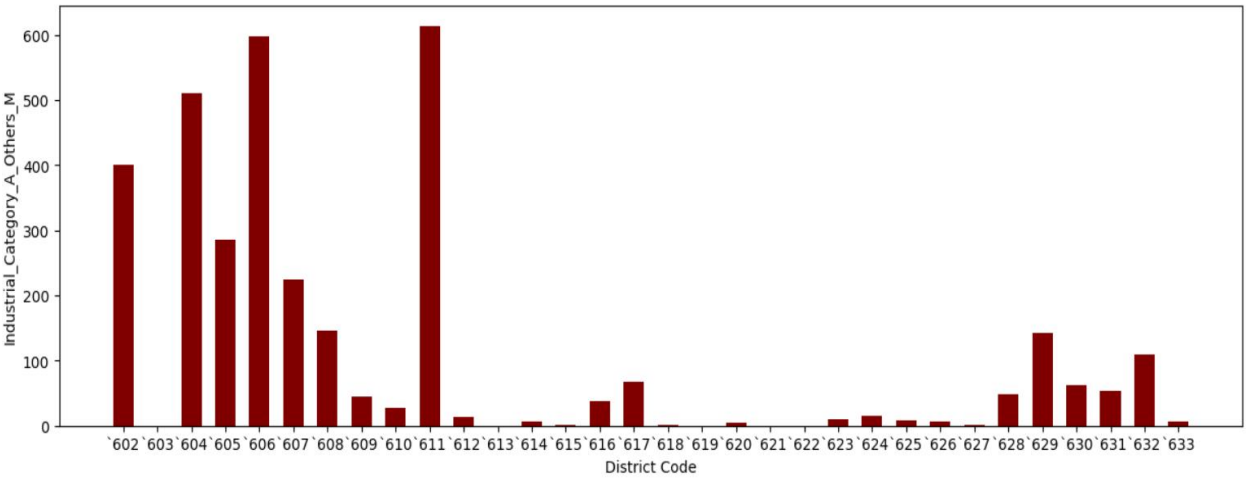

Industrial_Category_A_Cultivators_F
0
100
200
300
400
500
600
700
`602`603`604`605`606`607`608`609`610`611`612`613`614`615`616`617`618`619`620`621`622`623`624`625`626`627`628`629`630`631`632`633
District Code

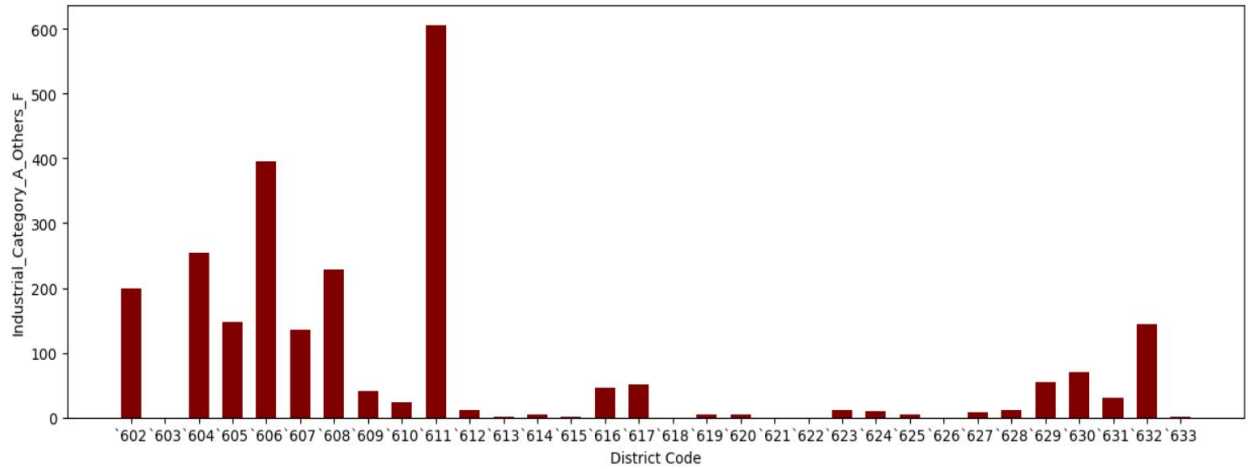

Industrial_Category_A_Agri_P
0
2000
4000
6000
8000
`602`603`604`605`606`607`608`609`610`611`612`613`614`615`616`617`618`619`620`621`622`623`624`625`626`627`628`629`630`631`632`633
District Code

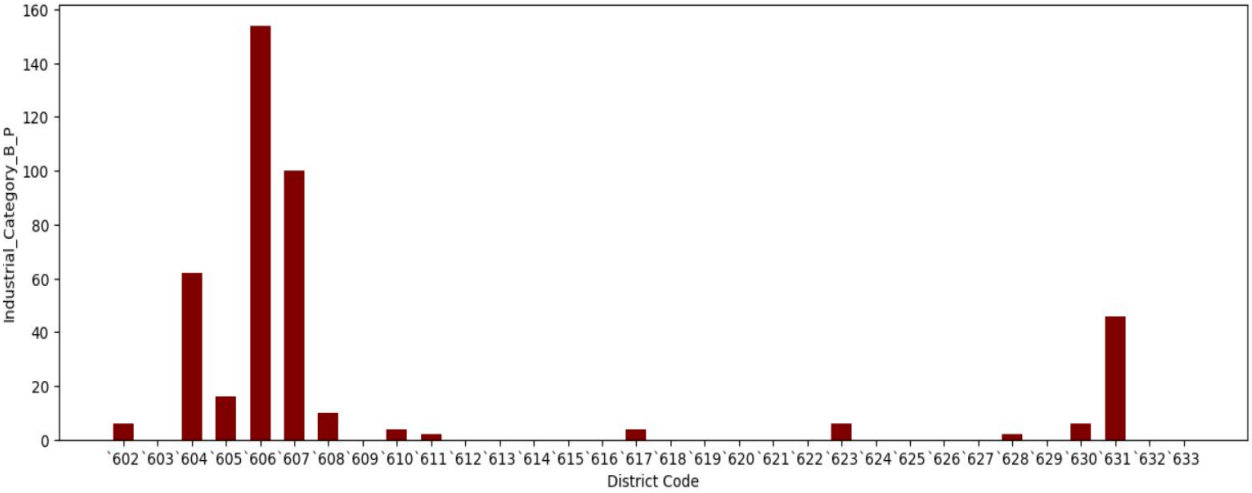

Industrial_Category_A_Agri_M
0
500
1000
1500
2000
2500
3000
3500
`602`603`604`605`606`607`608`609`610`611`612`613`614`615`616`617`618`619`620`621`622`623`624`625`626`627`628`629`630`631`632`633
District Code

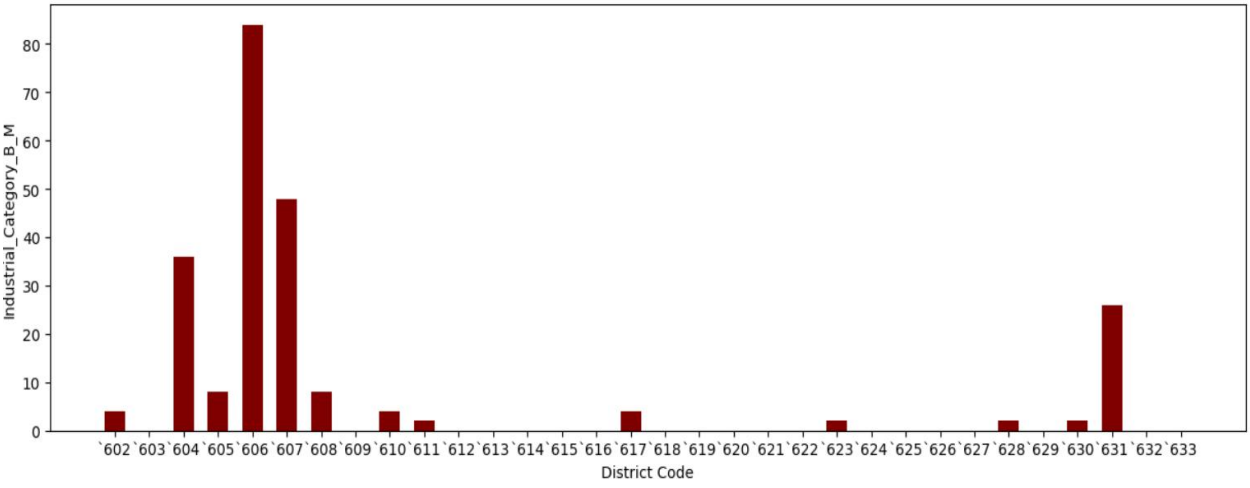

Industrial_Category_A_Agri_F
0
1000
2000
3000
4000
5000
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

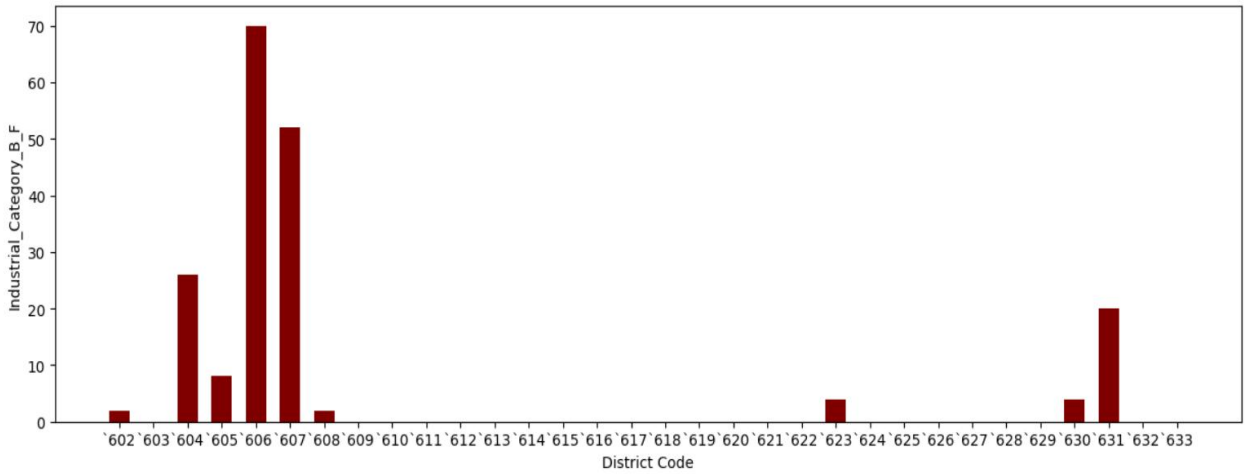

Industrial_Category_A_Others_P
0
200
400
600
800
1000
1200
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

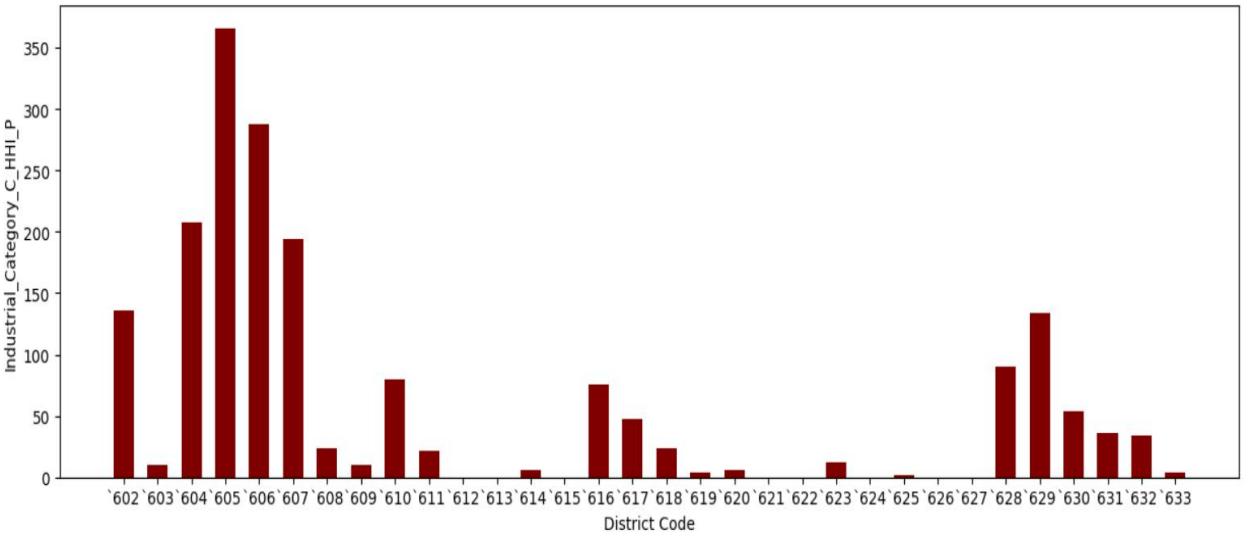

Industrial_Category_A_Others_M
0
100
200
300
400
500
600
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

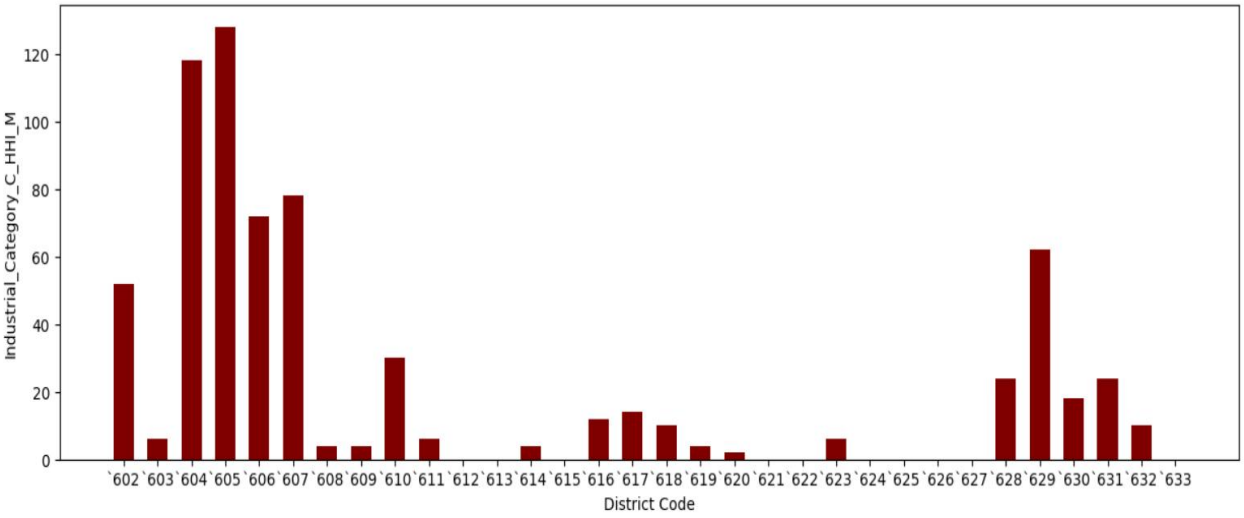

Industrial_Category_A_Others_F
0
100
200
300
400
500
600
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

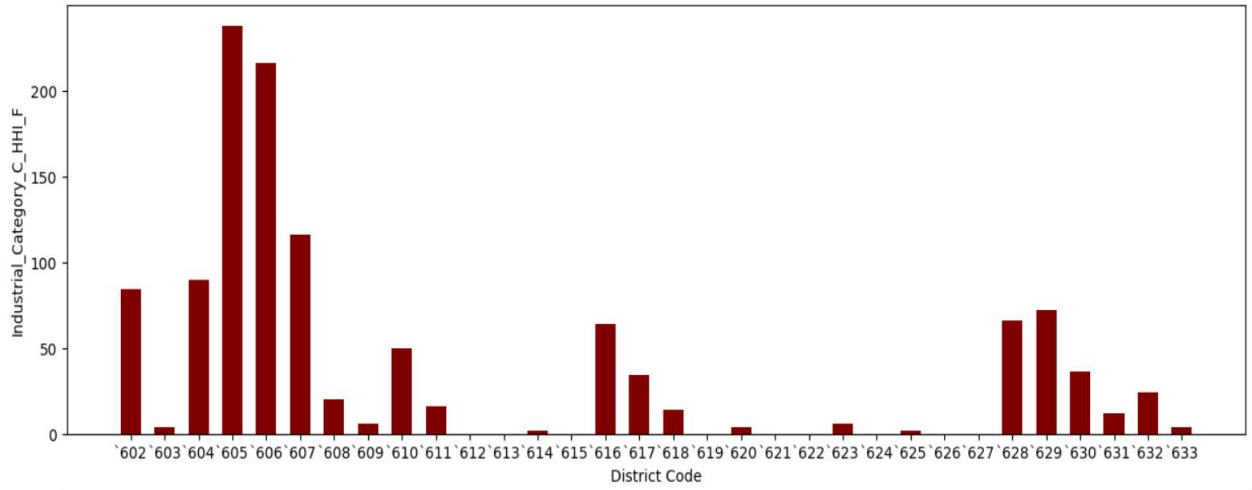

Industrial_Category_B_P
0
20
40
60
80
100
120
140
160
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

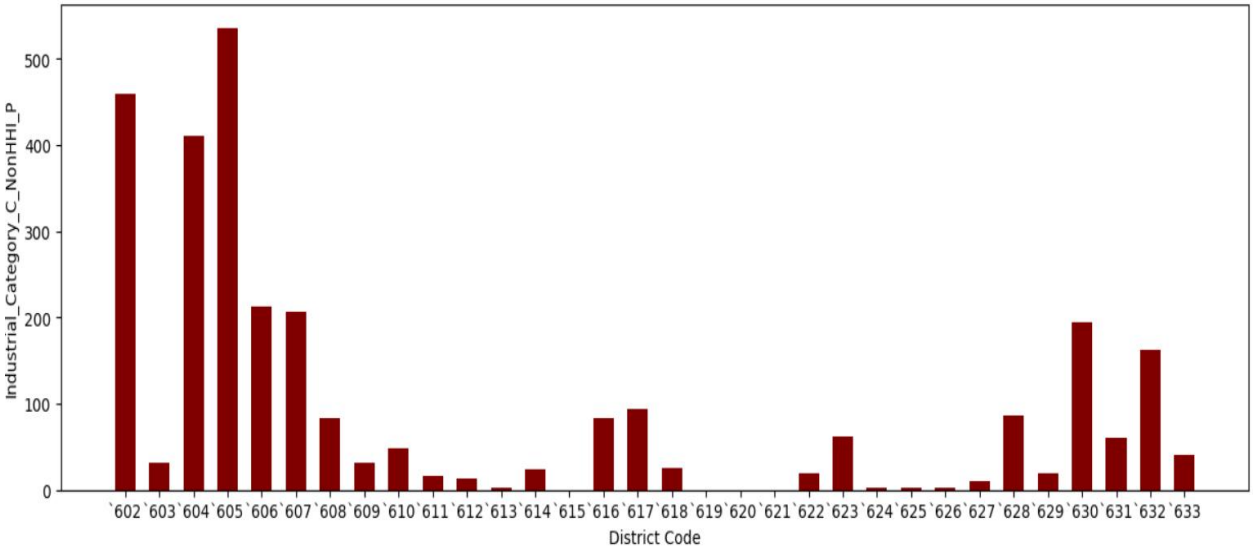

Industrial_Category_B_M
0
10
20
30
40
50
60
70
80
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

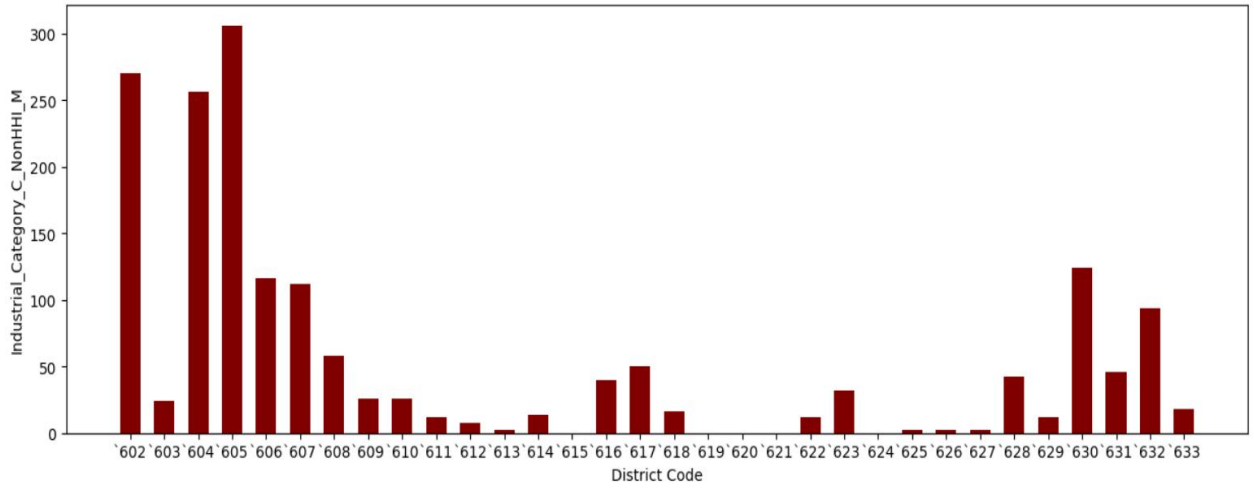

Industrial_Category_B_F
0
10
20
30
40
50
60
70
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

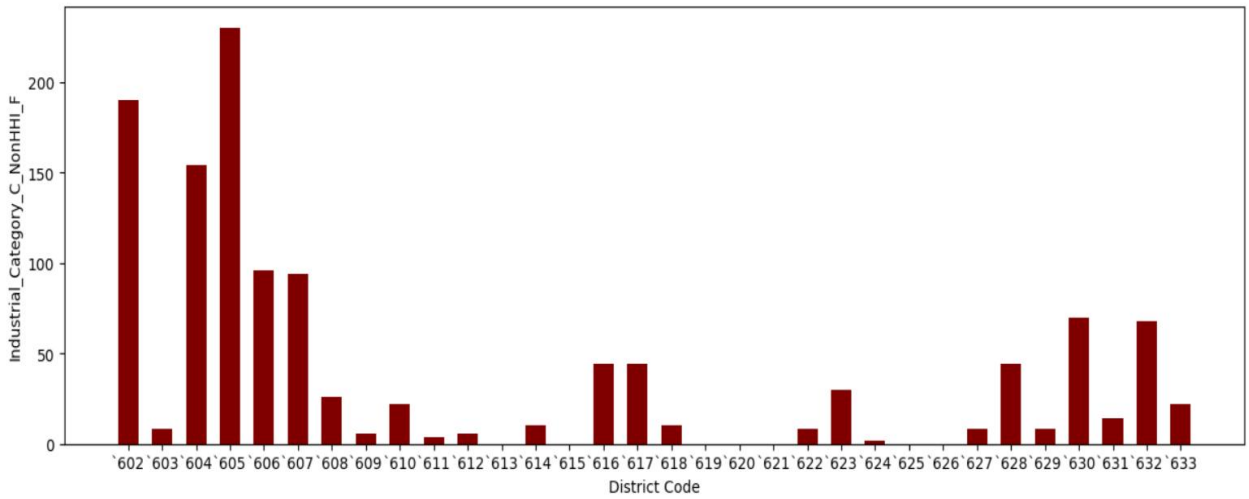

Industrial_Category_C_HHI_P
0
50
100
150
200
250
300
350
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

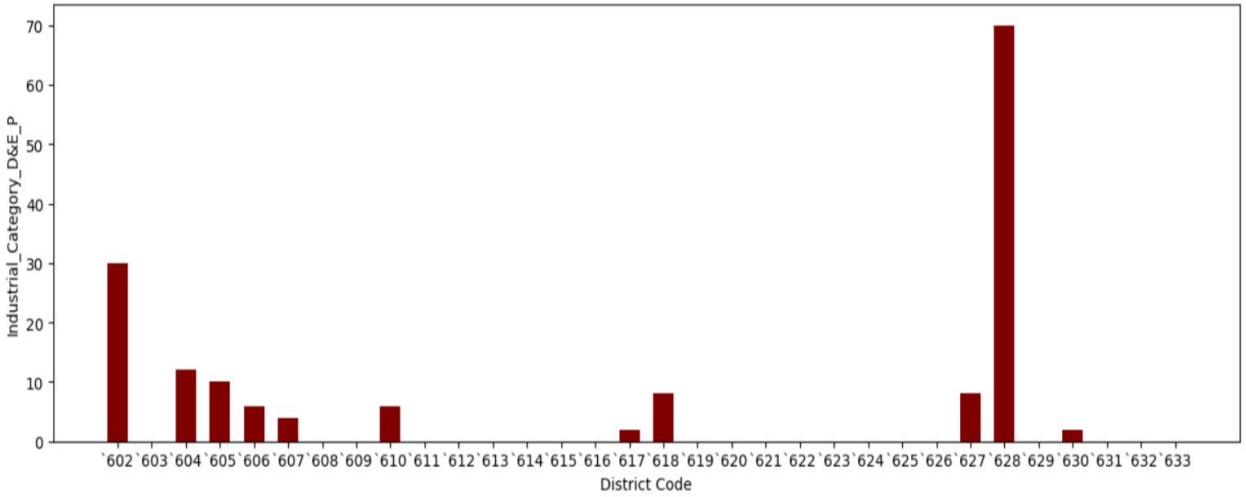

Industrial_Category_C_HHI_M
0
20
40
60
80
100
120
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

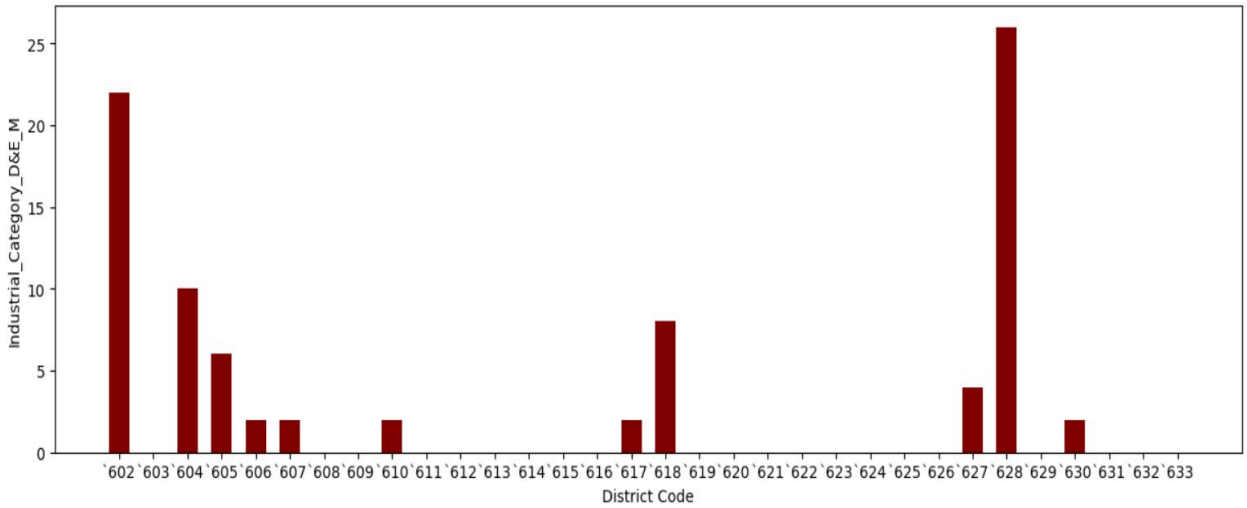

Industrial_Category_C_HHI_F
0
50
100
150
200
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

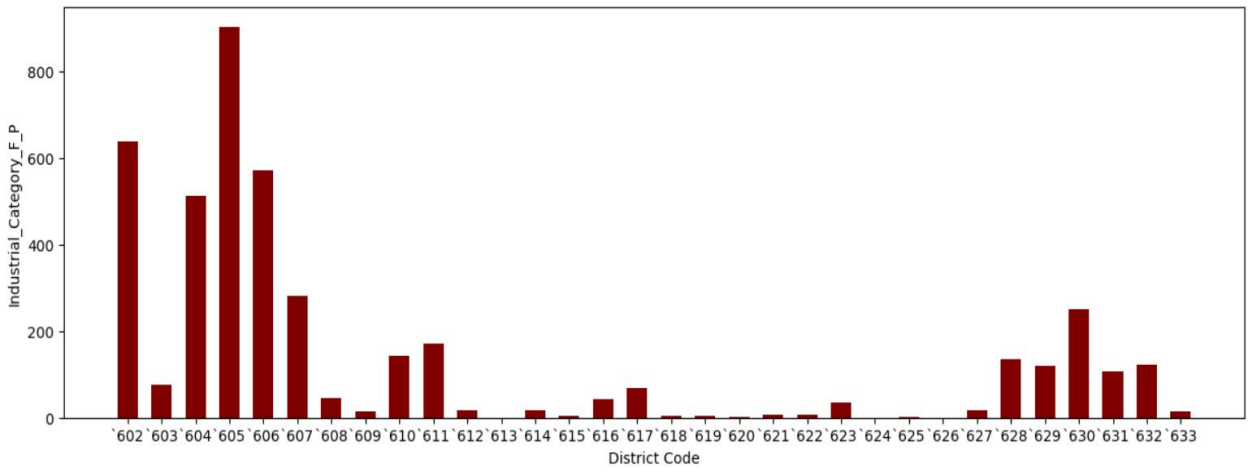

Industrial_Category_C_NonHHI_P
0
100
200
300
400
500
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

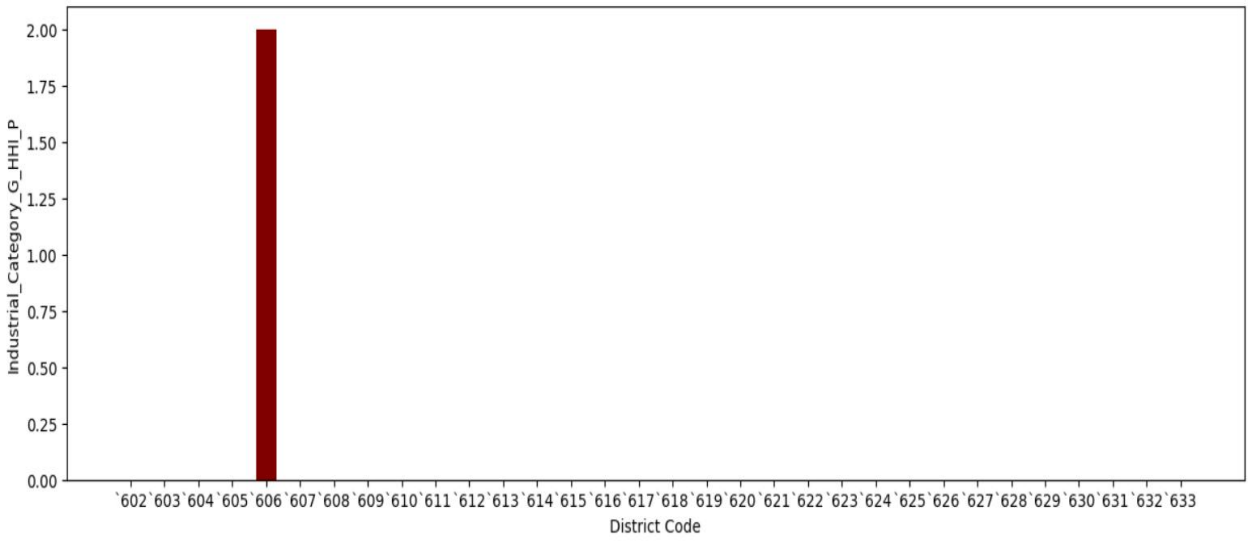

Industrial_Category_C_NonHHI_M
0
50
100
150
200
250
300
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

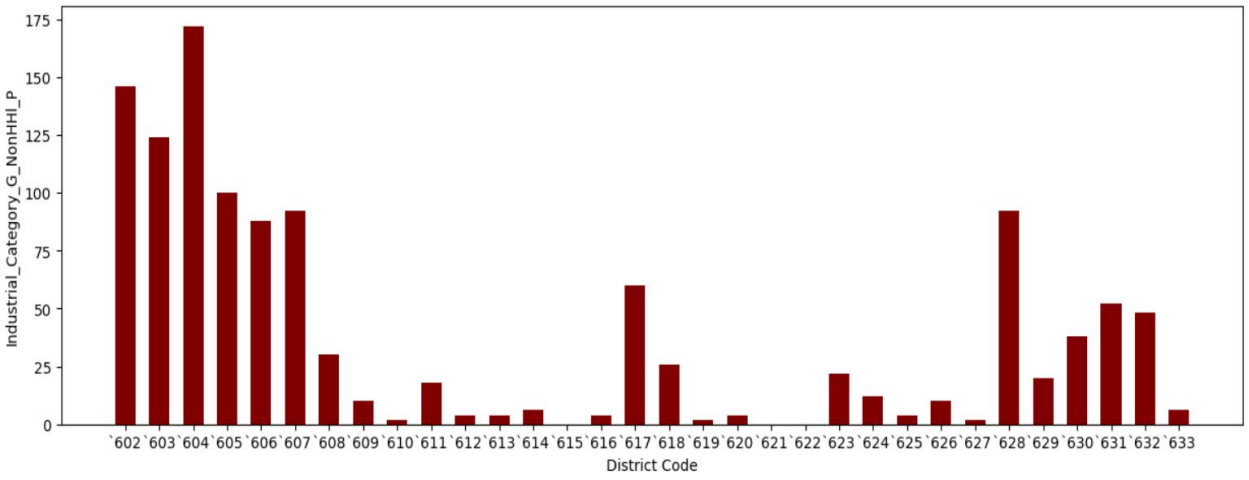

Industrial_Category_C_NonHHI_F
0
50
100
150
200
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

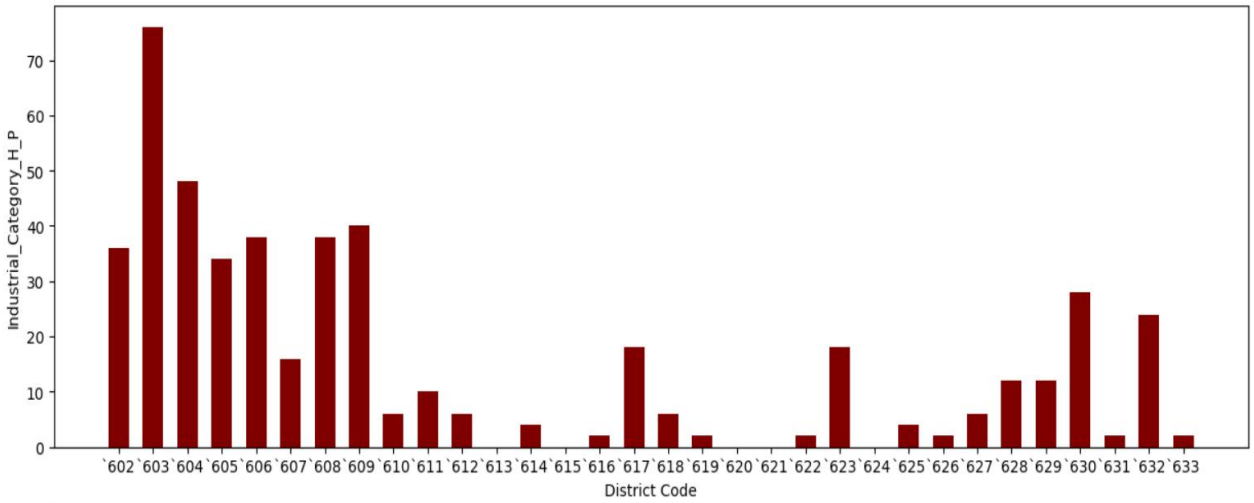

Industrial_Category_D&E_P
0
10
20
30
40
50
60
70
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

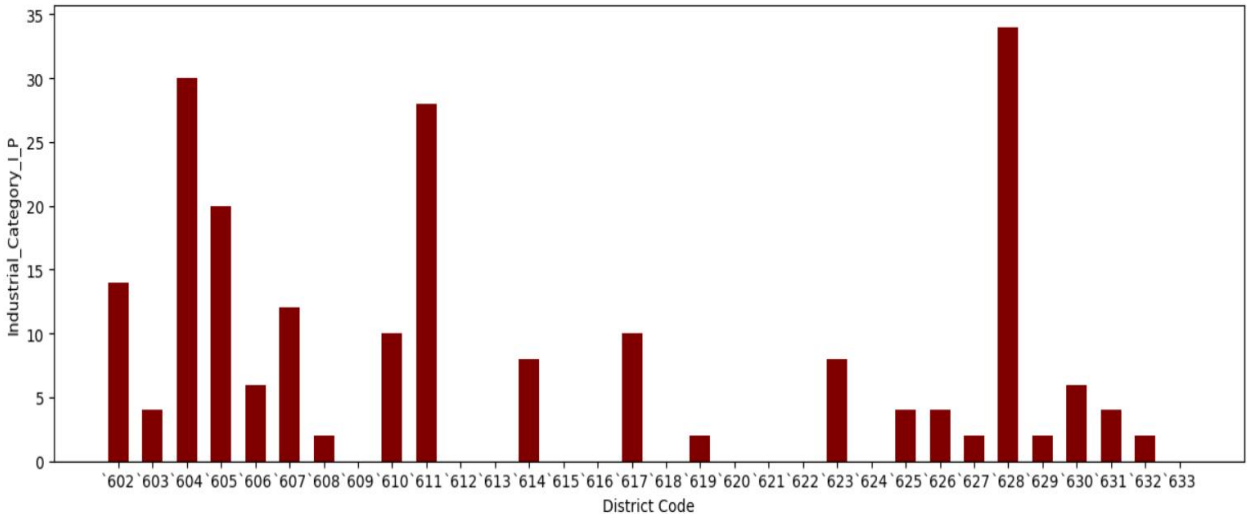

Industrial_Category_D&E_M
0
5
10
15
20
25
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

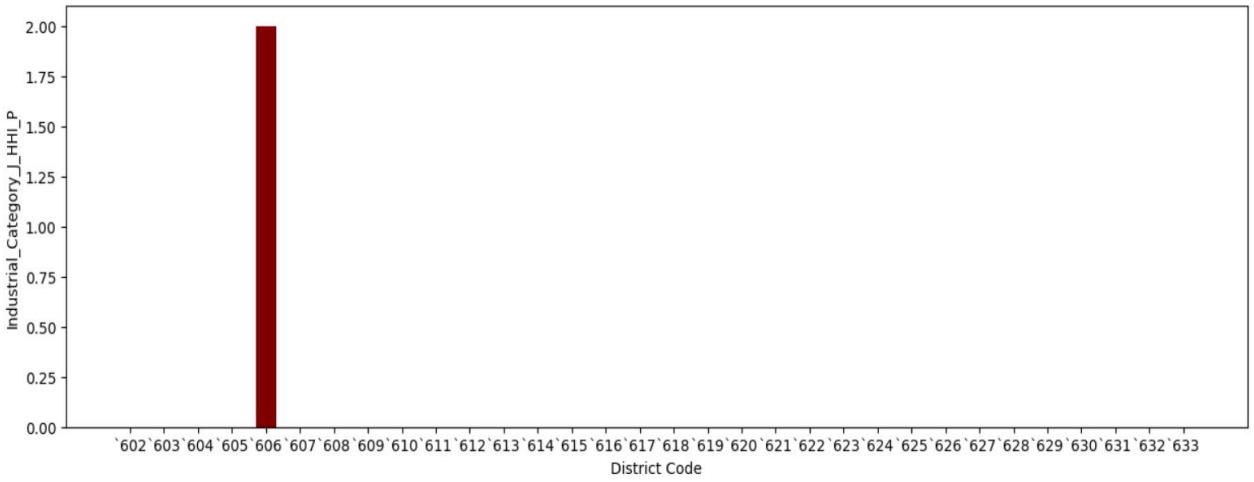

Industrial_Category_F_P
800
600
400
200
0
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

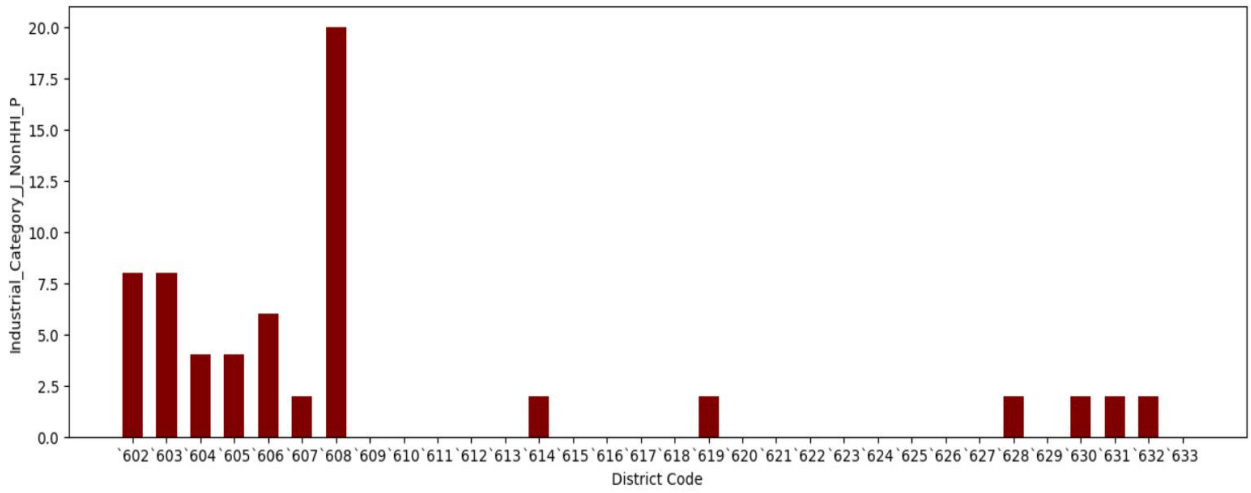

Industrial_Category_G_HHI_P
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

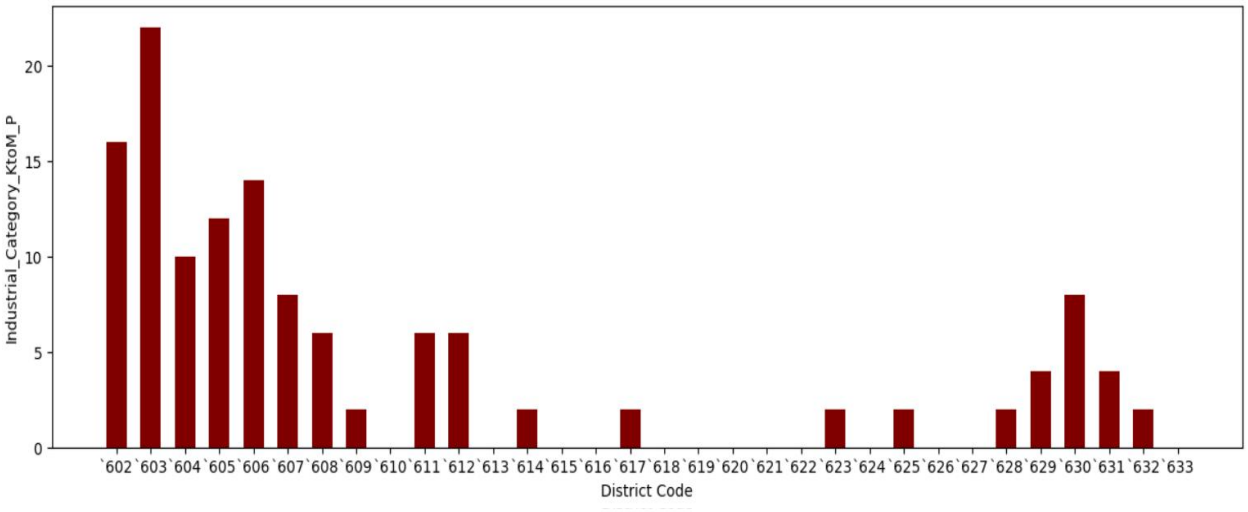

Industrial_Category_G_NonHHI_P
175
150
125
100
75
50
25
0
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

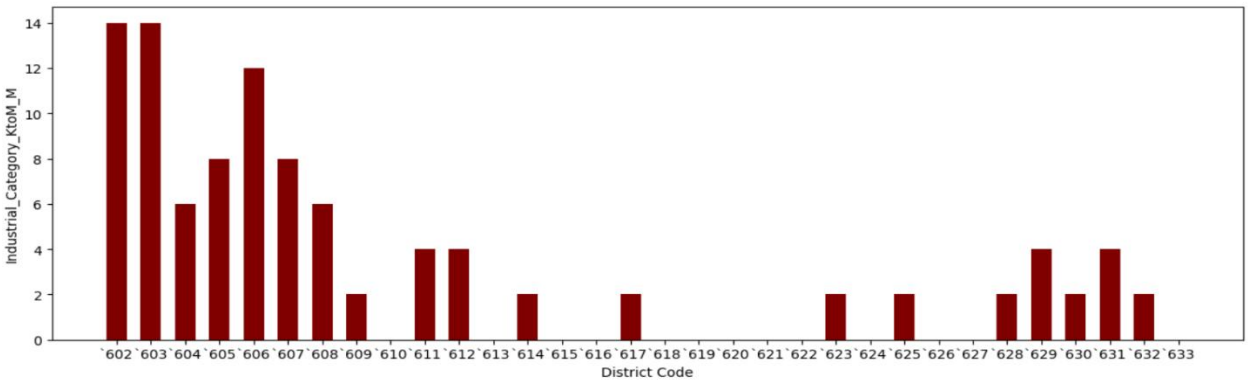

Industrial_Category_H_P
District Code

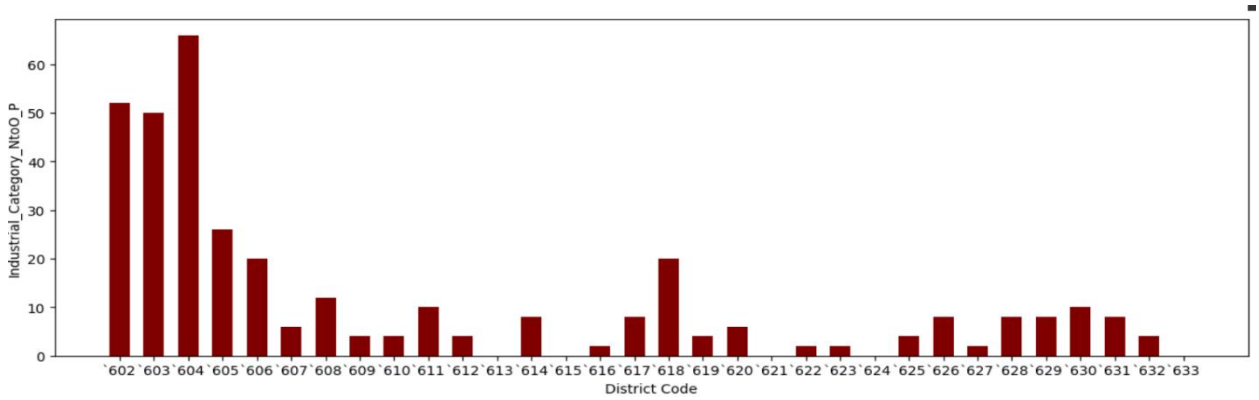

Industrial_Category_I_P
District Code

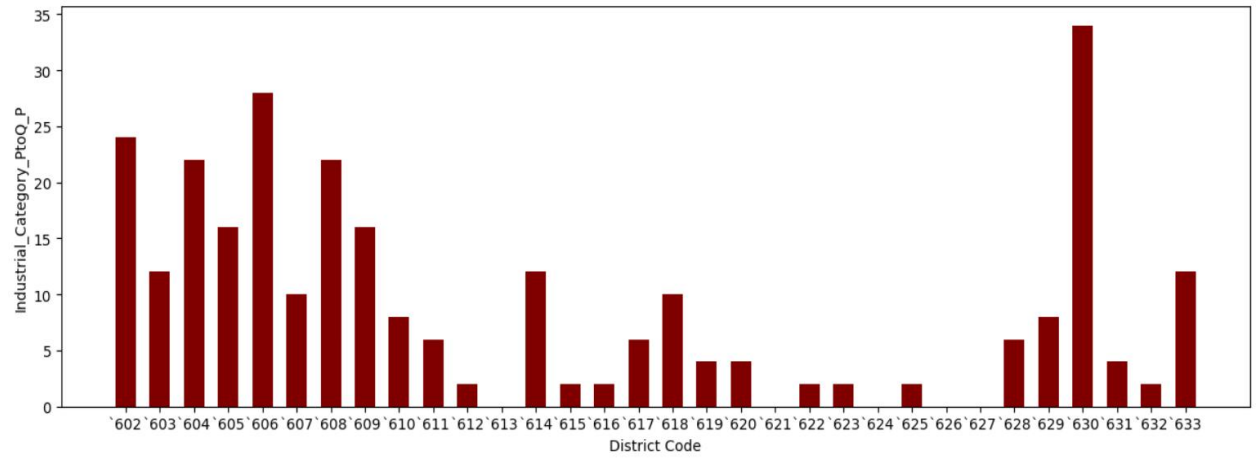

Industrial_Category_J_HHI_P
District Code

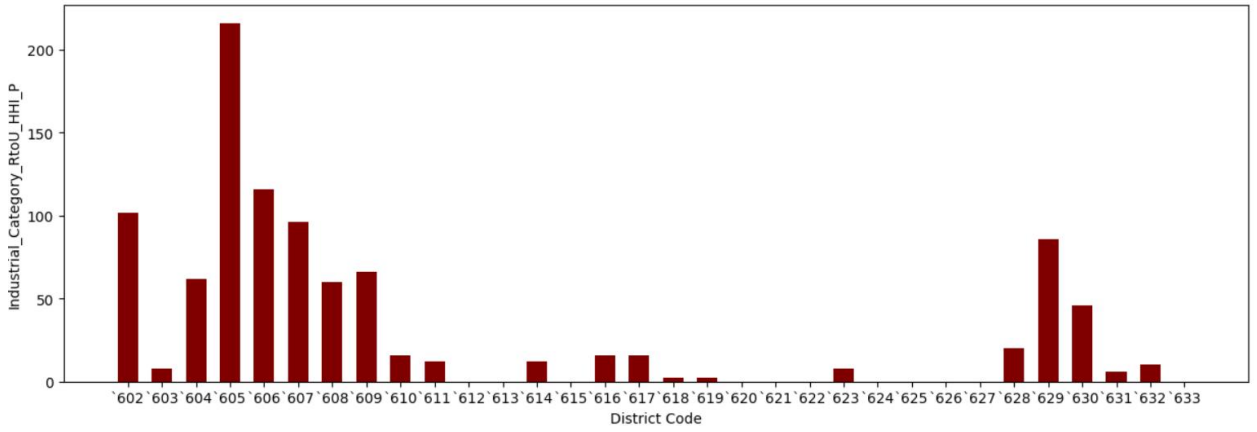

Industrial_Category_J_NonHHI_P
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

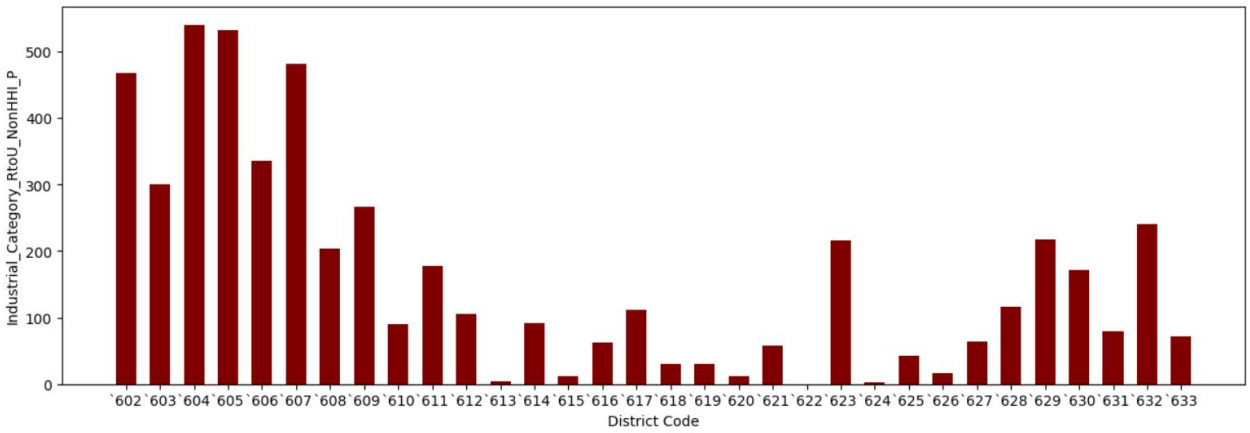

Industrial_Category_KtoM_P
20
15
10
5
0
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code


Industrial_Category_KtoM_M
14
12
10
8
6
4
2
0
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code


Industrial_Category_NtoO_P
0
10
20
30
40
50
60
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

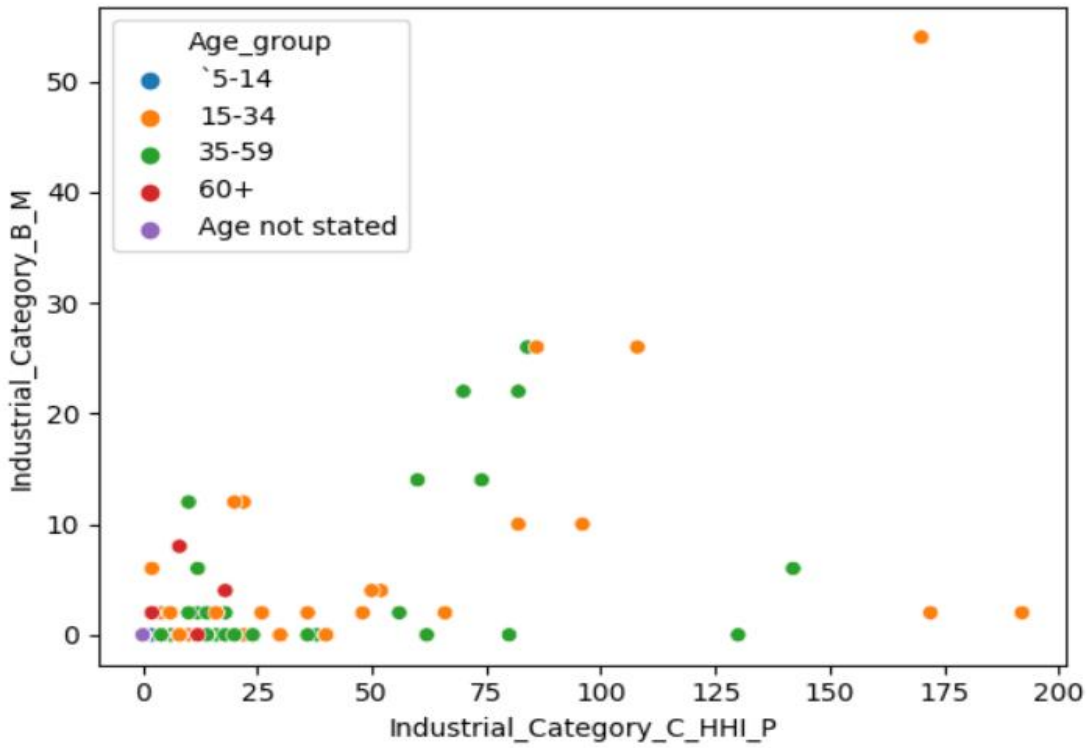

Industrial_Category_PtoQ_P
0
5
10
15
20
25
30
35
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

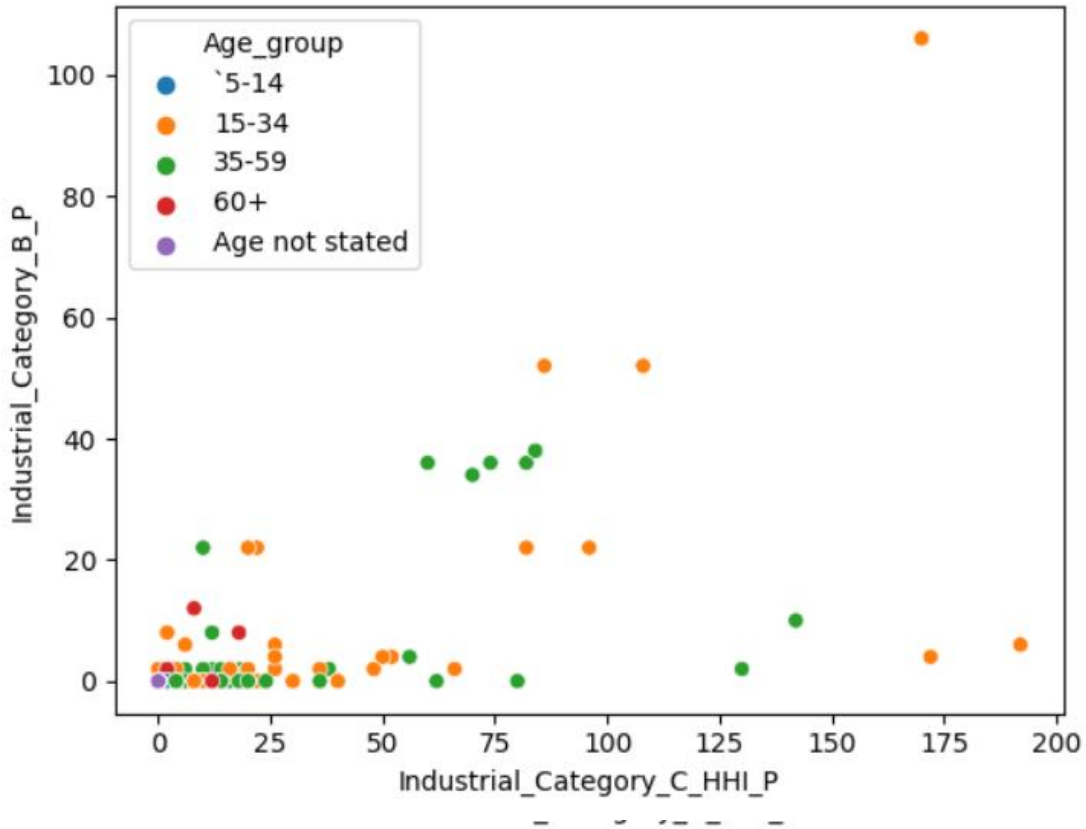

Industrial_Category_RtoU_HHI_P
0
50
100
150
200
`602 `603 `604 `605 `606 `607 `608 `609 `610 `611 `612 `613 `614 `615 `616 `617 `618 `619 `620 `621 `622 `623 `624 `625 `626 `627 `628 `629 `630 `631 `632 `633
District Code

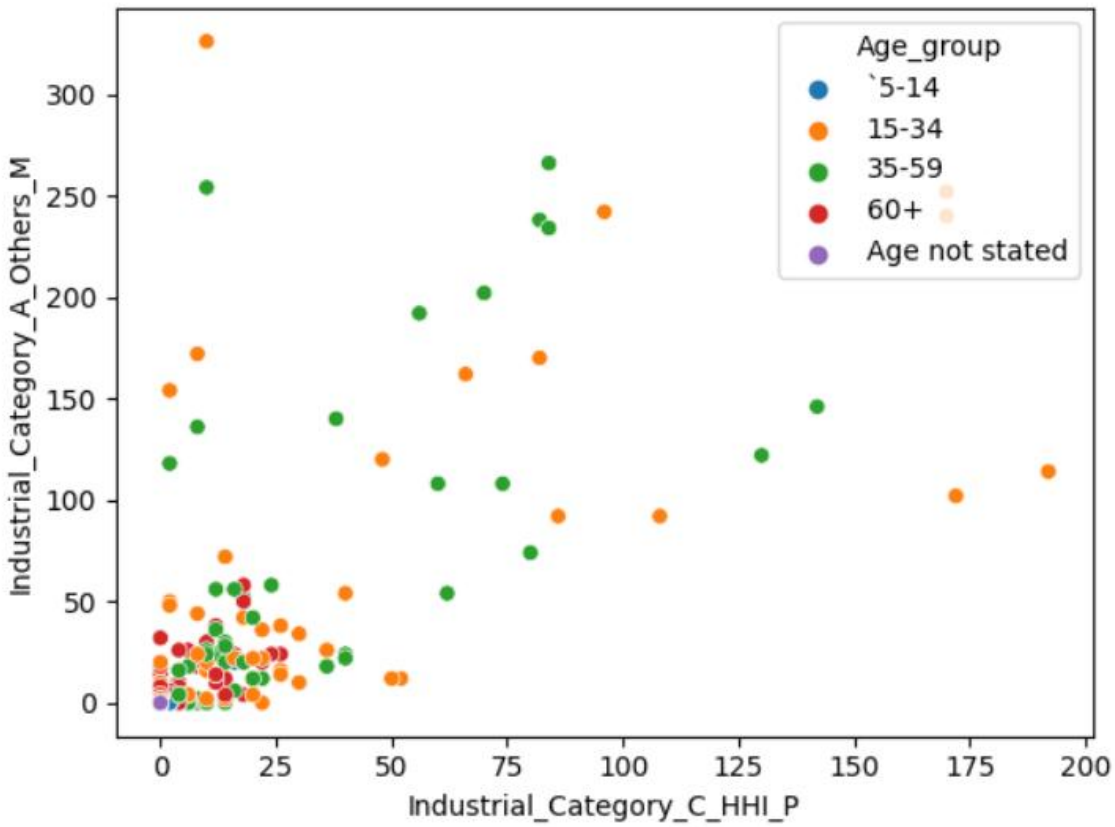


It shows the output and program in a form of scatterplot

```
Age_df=df["Age_group"] != "Total"
Age_data=df[Age_df]
cols1=list(df.columns)
cols1.remove("District Code")
cols1.remove("Area Name")
cols1.remove("Rural_Urban")
cols1.remove("Age_group")
for i in cols1:

  for j in cols1:
    sns.scatterplot(data=Age_data,x=i,y=j,hue=Age_data["Age_group"])
    plt.xlabel(i)
    plt.ylabel(j)
    plt.show()
```




Age_group
`5-14
15-34
35-59
60+
Age not stated
Industrial_Category_B_P
Industrial_Category_C_HHI_P
0
20
40
60
80
100
0
25
50
75
100
125
150
175
200


Age_group
`5-14
15-34
35-59
60+
Age not stated
Industrial_Category_A_Others_M
Industrial_Category_C_HHI_P
0
50
100
150
200
250
300
0
25
50
75
100
125
150
175
200

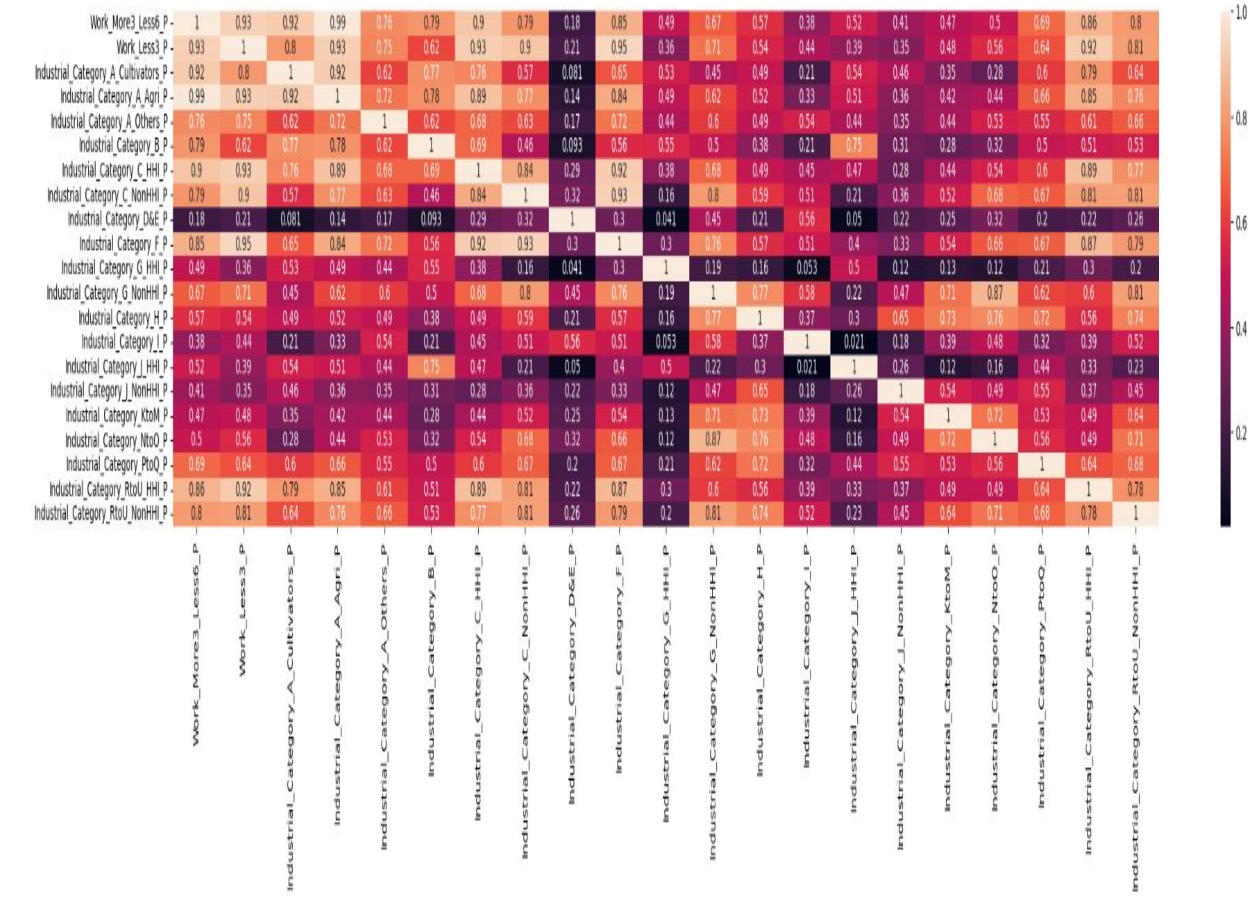
```
Age_df1=df["Age_group"] != "Total"
Age_data1=df[Age_df1]
cols1=list(df.columns)
cols1.remove("District Code")
cols1.remove("Area Name")
cols1.remove("Rural_Urban")
cols1.remove("Age_group")
cols2=[]
for i in cols1:

  if i.endswith("P"):
    cols2.append(i)
cols2
```

```
['Work_More3_Less6_P',
 'Work_Less3_P',
 'Industrial_Category_A_Cultivators_P',
 'Industrial_Category_A_Agri_P',
 'Industrial_Category_A_Others_P',
 'Industrial_Category_B_P',
 'Industrial_Category_C_HHI_P',
 'Industrial_Category_C_NonHHI_P',
 'Industrial_Category_D&E_P',
 'Industrial_Category_F_P',
 'Industrial_Category_G_HHI_P',
 'Industrial_Category_G_NonHHl_P',
 'Industrial_Category_H_P',
 'Industrial_Category_I_P',
 'Industrial_Category_J_HHI_P',
 'Industrial_Category_J_NonHHI_P',
 'Industrial_Category_KtoM_P',
 'Industrial_Category_NtoO_P',
 'Industrial_Category_PtoQ_P',
 'Industrial_Category_RtoU_HHI_P',
 'Industrial_Category_RtoU_NonHHI_P']
```

Heat map of the given dataset

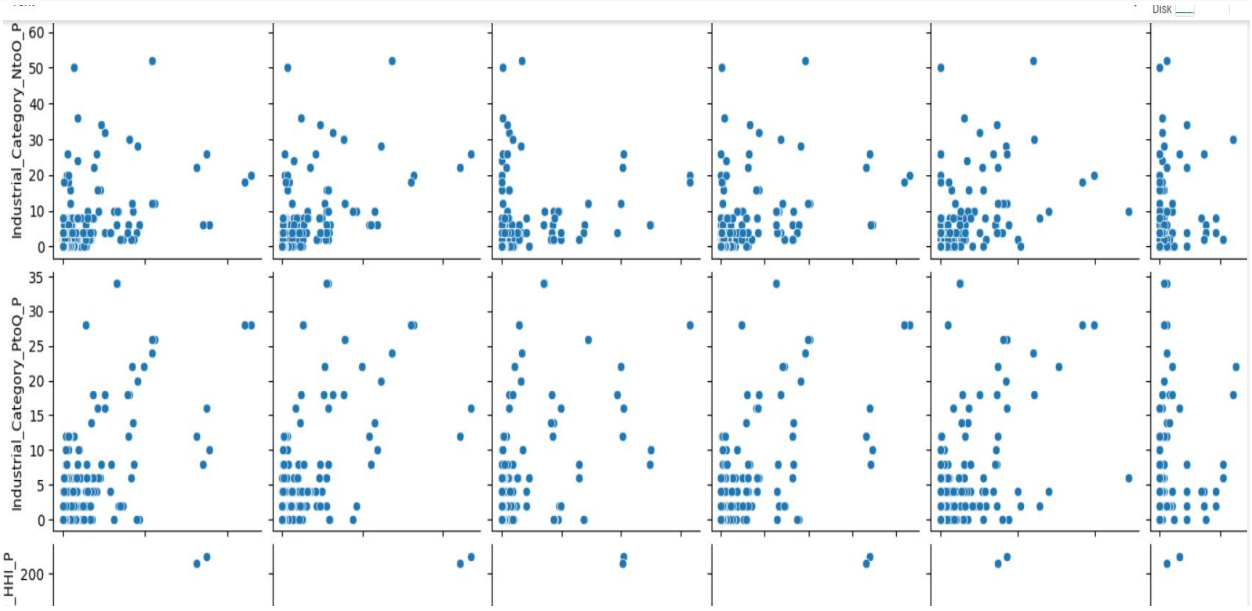
```
corr = df[cols2].corr()
plt.figure().set_figwidth(30)

sns.heatmap(corr, annot=True)


plt.show()
```

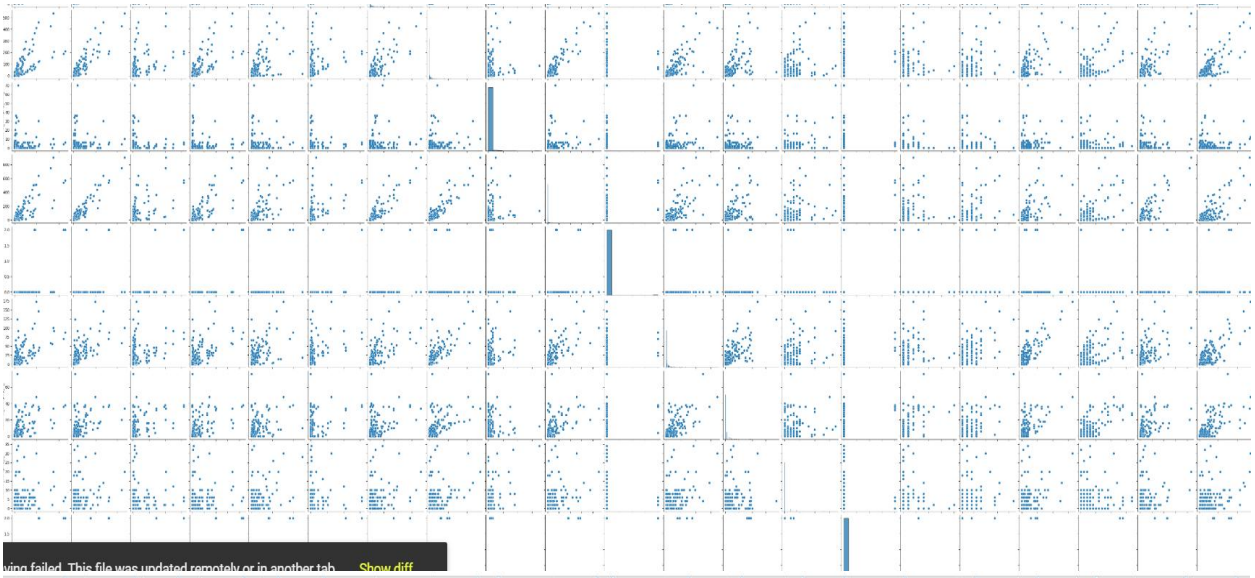
Output of heat map with correlation value



Pair plot of given dataset with different size of figure

```
data = df[cols2]

sns.pairplot(data)

plt.show()
```




ving failed. This file was updated remotely or in another tab.
Show diff