

# Minimal example of historical data parsing

## Input

Random historical table picked from 1951 UP census

25	Hamirpur—								
	Total .. ..	1,272	5,246	2,793	2,453	557	1,074	527	2,576
	Rural .. ..	821	3,241	1,710	1,534	372	700	342	1,664
	Urban .. ..	451	2,002	1,083	919	185	374	185	912
26	Banda—								
	Total .. ..	1,065	4,536	2,480	2,056	458	889	414	2,002
	Rural .. ..	882	3,662	2,013	1,619	387	758	346	1,663
	Urban .. ..	183	874	467	407	71	131	68	339
	Banaras Division								
	Total .. ..	13,077	64,020	32,968	31,052	4,841	10,230	5,160	25,605
	Rural .. ..	8,178	41,668	20,884	20,784	2,733	5,942	3,420	17,130
	Urban .. ..	4,899	22,352	12,084	10,268	2,108	4,288	1,740	8,475
27	Banaras—								
	Total .. ..	6,001	27,750	14,717	13,033	2,435	4,882	2,269	11,157
	Rural .. ..	2,896	14,575	7,401	7,174	952	1,927	1,236	6,166
	Urban .. ..	3,108	13,175	7,316	5,859	1,483	2,955	1,033	4,991
28	Mirzapur—								
	Total .. ..	1,565	7,041	3,675	3,366	650	1,414	648	3,204
	Rural .. ..	1,065	4,961	2,370	2,391	431	953	439	2,203
	Urban .. ..	500	2,080	1,105	975	219	461	209	1,001
29	Jaunpur—								
	Total .. ..	1,898	9,519	4,827	4,692	659	1,547	808	4,060
	Rural .. ..	1,532	7,780	3,859	3,921	514	1,248	675	3,409
	Urban .. ..	366	1,739	968	771	145	299	133	651
30	Ghazipur—								
	Total .. ..	1,599	8,556	4,389	4,167	500	1,011	623	3,123
	Rural .. ..	1,130	5,884	2,993	2,891	359	734	446	2,225
	Urban .. ..	469	2,672	1,396	1,276	141	277	177	898
31	Ballia—								
	Total .. ..	2,011	11,154	5,360	5,794	597	1,376	812	4,061
	Rural .. ..	1,555	8,468	4,061	4,407	477	1,080	624	3,127
	Urban .. ..	456	2,686	1,299	1,387	120	296	188	934
	Gorakhpur Division								
	Total .. ..	8,163	38,752	20,410	18,342	3,086	6,401	3,292	16,264
	Rural .. ..	5,697	27,047	13,988	13,059	2,098	4,388	2,384	11,816
	Urban .. ..	2,466	11,705	6,422	5,283	988	2,013	908	4,448
32	Gorakhpur—								
	Total .. ..	2,737	12,818	6,721	6,097	1,065	2,185	1,110	5,507
	Rural .. ..	1,547	7,414	3,763	3,651	551	1,172	682	3,398
	Urban .. ..	1,190	5,404	2,958	2,446	514	1,013	428	2,109
33	Deoria—								
	Total .. ..	1,256	5,924	3,204	2,720	519	1,103	465	2,313
	Rural .. ..	871	4,267	2,279	1,988	346	752	320	1,609
	Urban .. ..	385	1,657	925	732	173	351	145	704
34	Basti—								
	Total .. ..	1,690	7,369	3,992	3,377	718	1,456	676	3,350
	Rural .. ..	1,412	6,078	3,263	2,815	601	1,215	573	2,837
	Urban .. ..	278	1,291	729	562	117	241	103	513
35	Azamgarh—								
	Total .. ..	2,480	12,641	6,493	6,148	784	1,657	1,041	5,094
	Rural .. ..	1,867	9,288	4,683	4,605	600	1,249	809	3,972
	Urban .. ..	613	3,353	1,810	1,543	184	408	232	1,122
	Kumaon Division								
	Total .. ..	2,560	10,164	5,597	4,567	1,288	2,483	879	4,243
	Rural .. ..	1,618	6,504	3,429	3,075	787	1,582	590	2,825
	Urban .. ..	942	3,660	2,168	1,492	501	901	289	1,418
36	Naini Tal—								
	Total .. ..	997	3,499	2,125	1,374	595	1,052	273	1,334
	Rural .. ..	397	1,126	759	367	283	496	88	411
	Urban .. ..	600	2,373	1,366	1,007	312	556	185	923

## Processing

`DocParse.ipynb` runs `layoutparser` and extracts lists to produce a table. Here, we're interested in the 2nd numerical column (which is population).

## Output

Some errors, but much preferable to manual entry.

	index	0
0	Hamirpur_Total	5,246
1	Hamirpur_Rural	3,241
2	Hamirpur_Urban	2,002
3	Bande_Total	4,536
4	Bande_Rural	3,662
5	Bande_Urban	871
6	Banaras_Total	27,750
7	Banaras_Rural	14,575
8	Banaras_Urban	13,175
9	Mirzapur_Total	7,041
10	Mirzapur_Rural	4,961
11	Mirzapur_Urban	2,089
12	Jaunpur_Total	9,519
13	Jaunpur_Rural	7,780
14	Jaunpur_Urban	1,739
15	Ghazipur_Total	8,556
16	Ghazipur_Rural	5,894
17	Ghazipur_Urban	2,672
18	Ballia_Total	11,154
19	Ballia_Rural	8,468
20	Ballia_Urban	2,686
21	Gorakhpur_Total	12,818
22	Gorakhpur_Rural	7,414
23	Gorakhpur_Urban	5,404
24	Deoria_Total	5,924
25	Deoria_Rural	4,267
26	Deoria_Urban	1,657
27	Basti_Total	7,369
28	Basti_Rural	6,078
29	Basti_Urban	1,291

25	Hamirpur—				
	Total	..	..	1,272	5,246
	Rural	..	..	821	3,241
	Urban	..	..	451	2,002
26	Banda—				
	Total	..	..	1,065	4,536
	Rural	..	..	882	3,662
	Urban	..	..	183	871
	<b>Banaras Division</b>				
	Total	..	..	13,077	64,020
	Rural	..	..	8,178	41,668
	Urban	..	..	4,899	22,352
27	Banaras—				
	Total	..	..	6,001	27,750
	Rural	..	..	2,896	14,575
	Urban	..	..	3,108	13,175
28	Mirzapur—				
	Total	..	..	1,565	7,041
	Rural	..	..	1,065	4,961
	Urban	..	..	500	2,089
29	Jaunpur—				
	Total	..	..	1,898	9,519
	Rural	..	..	1,532	7,780
	Urban	..	..	366	1,739
30	Ghazipur—				
	Total	..	..	1,599	8,556
	Rural	..	..	1,130	5,894
	Urban	..	..	469	2,672
31	Ballia—				
	Total	..	..	2,011	11,154
	Rural	..	..	1,555	8,468
	Urban	..	..	456	2,686
	<b>Gorakhpur Division</b>				
	Total	..	..	8,163	38,752
	Rural	..	..	5,697	27,047
	Urban	..	..	2,466	11,705
32	Gorakhpur—				
	Total	..	..	2,737	12,818
	Rural	..	..	1,547	7,414
	Urban	..	..	1,190	5,404
33	Deoria—				
	Total	..	..	1,256	5,924
	Rural	..	..	871	4,267
	Urban	..	..	385	1,657
34	Basti—				
	Total	..	..	1,690	7,369
	Rural	..	..	1,412	6,078
	Urban	..	..	278	1,291