Task report

TOTAL TABLE: 2
- Enrollment Capacity and Utilization Report, Target calculation by school
- Enrollment Capacity and Utilization Report, Target calculation by Building

Table 1 details:
Pages: 215
Apprx time for 1 page: 19 minutes
Total time: 68 hours 5 minutes
Sample of manually copying the pdf table

Geo Dist	Org ID	Org Name	Bldg ID	Bldg Name	Enroll	Tareget Cap	Target Util	PK cap+	No. of cluster/Spec Rms Reported+	No. of cluster/Spec Rms Needed+
1	M015	P.S. 15-M	M015	P.S. 15-M	179	251	71	17	1	2
1	M019	P.S. 19-M	M019	P.S. 19-M	176	577	31	18	2	2
1	M020	P.S. 20-M	M020	P.S. 20-M	351	691	51	54	5	3
1	M034	P.S. 34-M	M034	P.S. 34-M	258	381	68	36	3	3
1	M063	P.S. 63-M	M063	P.S. 63-M	186	256	73	34	4	2
1	M064	P.S. 64-M	M064	P.S. 64 (OLD 71)- M	215	324	66	36	1	2
1	M080	MANHATTAN CHARTER SCHOOL II	M056	J.H.S. 56-M	110	391	28	0	0	2
1	M094	M094 SPED-M	M015	P.S. 15-M	41	96	43			
			M061	P.S. 61-M	45	84	54			
			M188	P.S. 188-M	103	122	84			
			M094 SPED-M TOTAL		189	302	63			
1	M110	P.S. 110-M	M110	M110	330	401	82	34	4	3
1	M134	P.S. 134-M	M134	M134	177	410	43	18	6	2
1	M140	P.S. 140-M	M140	M140	340	536	63	36	6	3
1	M142	P.S. 142-M	M142	M142	330	547	60	89	2	3
1	M184	P.S. 184-M	M137	P.S. 137-M	717	749	96	36	5	3
1	M188	P.S. 188-M	M188	P.S. 188-M	390	522	75	18	3	3
1	M292	J.H.S. 292 - M	M056	J.H.S. 56 - M	290	422	69		9	
1	M315	P.S. 315 - M	M061	P.S. 61-M	302	225	134	35	1	3
1	M320	MANHATTAN CHARTER - M	M142	P.S. 142-M	208	220	95	0	1	2

Table 2 details:
Pages: 175
Apprx time for 1 page: 18 minutes
Total time: 52 hours 30 imutes
Sample of manually copying the pdf table

Geo Dist	Bldg ID	Bldg Name	Bldg Enroll	Bldg Target Cap	Org ID	Org Name	Org Enroll	Org Target Cap	Org Target Util	PK Cap+	No. of Cluster/Spec Rms Reported+	No. of Cluster/Spec Rms Reported+
1	M015	P.S. 15-M	220	63	M015	P.S. 15-M	179	251	71	17	1	2
					M094	MO94 SPED-M	41	96	43			
					MADM	DOE ADMINISTRATIVE OFFICE						
					MCBO	COMMUNITY BASED ORGANIZATION -M						
1	M019	P.S. 19-M	176	31	M019	P.S. 19-M	176	577	31	18	2	2
1	M020	P.S. 20-M	351	51	M020	P.S. 20-M	351	691	51	54	5	3
					MCBO	COMMUNITY BASED ORGANIZATION -M						
					MSBH	SCHOOL BASED HEALTH CENTER-M						
					MSFS	FOOD SERVICES- M						
1	M022	N.E.S.T (OLD J22)- M	1640	114	M539	NEW EXPLORATIONS - M	1640	1444	114	0	12	5
1	M025	J.H.S. 25-M	631	40	M378	I.S. 378-M	167	597	28		5	
					M515	LOWER EAST SIDE PREP HS - M	464	970	48		11	
1	M034	P.S. 34-M	381	68	M034	P.S. 34- M	258	381	68	36	3	3
1	M056	J.H.S 56-M	1242	48	M080	MANHATTAN CHARTER SCHOOL- II	110	391	28	0	0	2
					M292	J.H.S 292-M	290	422	69		9	
					M323	I.S. 332-M	196	429	46		2	
					M973	RESTART- CITYWIDE	0					
					MADM	DOE ADMINISTRATIVE OFFICE						
					MCBO	COMMUNITY BASED ORGANIZATION -M						
					MSBH	HHC GOTHAM						
1	M060	J.H.S 60-M	1085	87	M330	GIRLS PREP CHARTER SCHOOL - M	246	204	121	16	4	3
					M450	EAST SIDE COMMUNITY SCHOOL	701	881	80		5	
					MCBO	BEACON PROGRAM-M						

Converting using Online Pdf to excel converters:
Best one:
https://www.adobe.com/in/acrobat/online/pdf-to-excel.html

Output:

Table 1

Geo Dist	Org ID	Org Name	Bldg ID	Bldg Name	Enroll	Target Cap	Target Util	PK Cap+	No. of Cluster/ Spec Rms Reported+	No. of Cluster Rms Needed+
1	M015	P.S. 15 - M	M015	P.S. 15 - M	179	251	71	17	1	2
1	M019	P.S. 19 - M	M019	P.S. 19 - M	176	577	31	18	2	2
1	M020	P.S. 20 - M	M020	P.S. 20 - M	351	691	51	54	5	3
1	M034	P.S. 34 - M	M034	P.S. 34 - M	258	381	68	36	3	3
1	M063	P.S. 63 - M	M063	P.S. 63 - M	186	256	73	34	4	2
1	M064	P.S. 64 - M	M064	P.S. 64 (OLD 71) - M	215	324	66	36	1	2
1	M080	MANHATTAN CHARTER SCHOOL II	M056	J.H.S. 56 - M	110	391	28	0	0	2
1	M094	M094 SPED - M	* M015	P.S. 15 - M	41	96	43			
			* M061	P.S. 61 - M	45	84	54			
			* M188	P.S. 188 - M	103	122	84			
				M094 SPED - M TOTAL	189	302	63			
1	M110	P.S. 110 - M	M110	P.S. 110 - M	330	401	82	34	4	3
1	M134	P.S. 134 - M	M134	P.S. 134 - M	177	410	43	18	6	2
1	M140	P.S. 140 - M	M140	P.S. 140 - M	340	536	63	36	6	3
1	M142	P.S. 142 - M	M142	P.S. 142 - M	330	547	60	89	2	3
1	M184	P.S. 184 - M	M137	P.S. 137 - M	717	749	96	36	5	3
1	M188	P.S. 188 - M	M188	P.S. 188 - M	390	522	75	18	3	3
1	M292	J.H.S. 292 - M	M056	J.H.S. 56 - M	290	422	69		9	
1	M315	P.S. 315 - M	M061	P.S. 61 - M	302	225	134	35	1	3
1	M320	MANHATTAN CHARTER - M	M142	P.S. 142 - M	208	220	95	0	1	2

Time taken: 11 seconds

Table2:

Bldg Geo Dist	Bldg Bldg ID	Bldg Name	Bldg Enroll	Target Cap	Target Util	Org ID	Org Name	Org Enroll	Org Target Cap	Org Target Util	PK Cap+	No. of Cluster/ Spec Rms Reported+	No. of Cluster Rms Needed+
1	M015	P.S. 15 - M	220	347	63	M015	P.S. 15 - M	179	251	71	17	1	2
						* M094	M094 SPED - M	41	96	43			
						MADM	DOE ADMINISTRATIVE OFFICE						
						MCBO	COMMUNITY BASED ORGANIZATION - M						
1	M019	P.S. 19 - M	176	577	31	M019	P.S. 19 - M	176	577	31	18	2	2
1	M020	P.S. 20 - M	351	691	51	M020	P.S. 20 - M	351	691	51	54	5	3
						MCBO	COMMUNITY BASED ORGANIZATION - M						
						MSBH	SCHOOL BASED HEALTH CENTER - M						
						MSFS	FOOD SERVICES - M						
1	M022	N.E.S.T (OLD J22) - M	1,640	1,444	114	M539	NEW EXPLORATIONS - M	1,640	1,444	114	0	12	5
1	M025	J.H.S. 25 - M	631	1,567	40	M378	I.S. 378 - M	167	597	28		5	
						M515	LOWER EAST SIDE PREP HS - M	464	970	48		11	
1	M034	P.S. 34 - M	258	381	68	M034	P.S. 34 - M	258	381	68	36	3	3
1	M056	J.H.S. 56 - M	596	1,242	48	M080	MANHATTAN CHARTER SCHOOL II	110	391	28	0	0	2
						M292	J.H.S. 292 - M	290	422	69		9	
						M332	I.S. 332 - M	196	429	46		2	
						M973	RESTART - CITYWIDE	0					
						MADM	DOE ADMINISTRATIVE OFFICE - M						
						MCBO	COMMUNITY BASED ORGANIZATION - M						
						MSBH	HHC GOTHAM						
1	M060	J.H.S. 60 - M	947	1,085	87	M330	GIRLS PREP CHARTER SCHOOL - M	246	204	121	16	4	3
						M450	EAST SIDE COMMUNITY SCHOOL	701	881	80		5	

‹ › ›| Table 1   +

Time taken: 11 seconds

Other links:
https://pdftables.com/
https://www.ilovepdf.com/pdf_to_excel
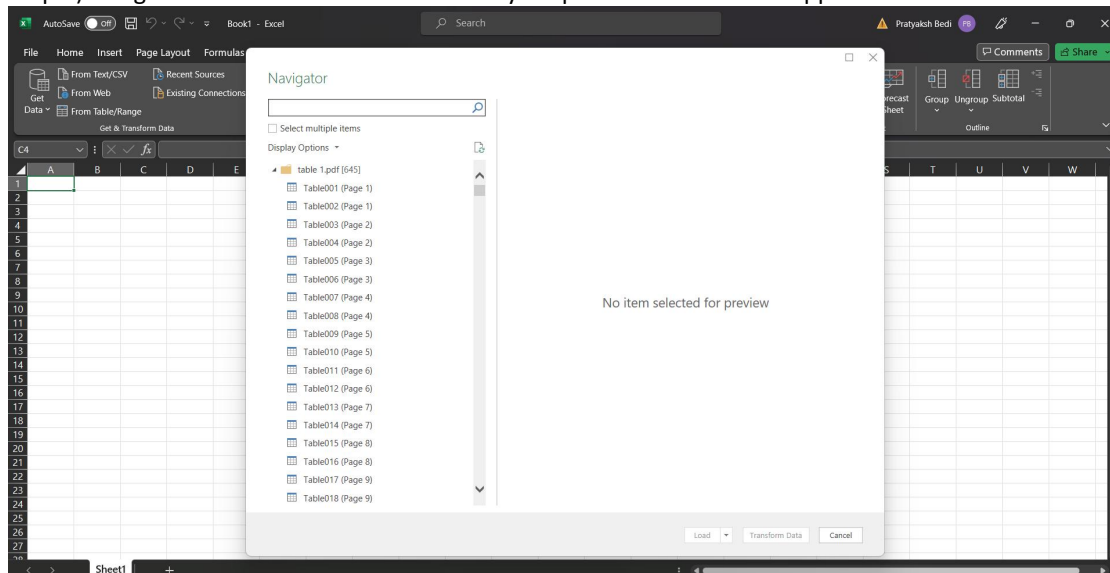
Other methods to convert:

Copy and Paste	**Pros** – In case you only have a couple of PDF files, it is the easiest method – No extra software or equipment is needed **Cons** – When you have a lot of files it is very time-consuming – Prone to errors – Extremely tedious process
PDF converters	**Pros** – Easy to use – Many online tools available – Cheap option

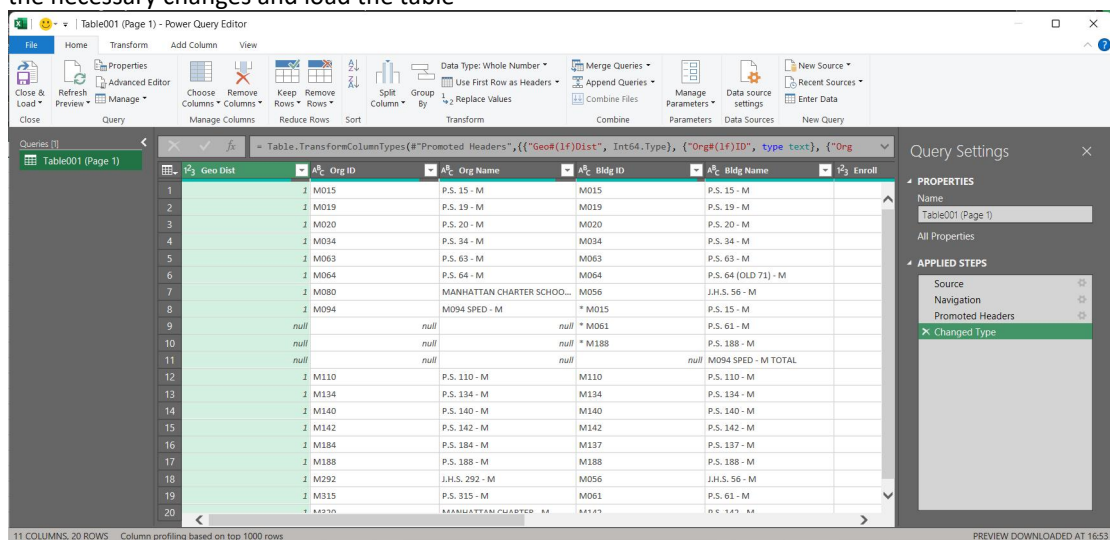|		
	Cons – Functionality is limited – Not possible to extract data in bulk – Only works with native PDFs	
Direct Import in Excel	**Pros** – Easy method – No extra software or equipment needed **Cons** – Impossible to extract data in bulk – Requires some knowledge of Excel – Errors in formatting	

Converting using excel features:

Step 1) go to data tab

Step 2) use get data. Select the destination of your pdf file. Window will appear

Step3) select the table you want among all the tables. power string query window will appear. Make the necessary changes and load the table

The pdf is converted into a table

GeoDist	OrgID	Org Name	BldgID	Bldg Name	Enroll	TargetCap	TargetUtil	PKCap+	No. ofCluster/Spec RmsReported+	No. ofClusterRm
1	M015	P.S. 15 - M	M015	P.S. 15 - M	179	251	71	17	1	
1	M019	P.S. 19 - M	M019	P.S. 19 - M	176	577	31	18	2	
1	M020	P.S. 20 - M	M020	P.S. 20 - M	351	691	51	54	5	
1	M034	P.S. 34 - M	M034	P.S. 34 - M	258	381	68	36	3	
1	M063	P.S. 63 - M	M063	P.S. 63 - M	186	256	73	34	4	
1	M064	P.S. 64 - M	M064	P.S. 64 (OLD 71) - M	215	324	66	36	1	
1	M080	MANHATTAN CHARTER SCHOOL II	M056	J.H.S. 56 - M	110	391	28	0	0	
1	M094	M094 SPED - M	* M015	P.S. 15 - M	41	96	43			
			* M061	P.S. 61 - M	45	84	54			
			* M188	P.S. 188 - M	103	122	84			
				M094 SPED - M TOTAL	189	302	63			
1	M110	P.S. 110 - M	M110	P.S. 110 - M	330	401	82	34	4	
1	M134	P.S. 134 - M	M134	P.S. 134 - M	177	410	43	18	6	
1	M140	P.S. 140 - M	M140	P.S. 140 - M	340	536	63	36	6	
1	M142	P.S. 142 - M	M142	P.S. 142 - M	330	547	60	89	2	
1	M184	P.S. 184 - M	M137	P.S. 137 - M	717	749	96	36	5	
1	M188	P.S. 188 - M	M188	P.S. 188 - M	390	522	75	18	3	
1	M292	J.H.S. 292 - M	M056	J.H.S. 56 - M	290	422	69		9	
1	M315	P.S. 315 - M	M061	P.S. 61 - M	302	225	134	35	1	
1	M320	MANHATTAN CHARTER - M	M142	P.S. 142 - M	208	220	95	0	1	

Converting using word:

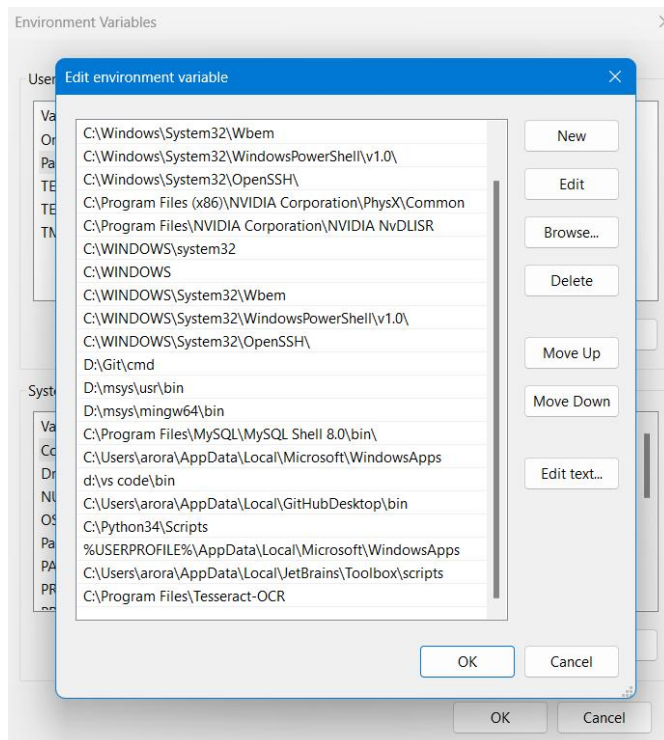Step1) open your pdf file into word.

Step2) copy the table from word and paste it directly on excel

Geo Dist	Org ID	Org Name	Bldg ID	Bldg Name	Enroll	Target Cap	Target Util	PK Cap+	No. of Cluster/ Spec Rms Reported+	No. of Cluster Rms Needed+
1	M015	P.S. 15 - M	M015	P.S. 15 - M	179	251	71	17	1	2
1	M019	P.S. 19 - M	M019	P.S. 19 - M	176	577	31	18	2	2
1	M020	P.S. 20 - M	M020	P.S. 20 - M	351	691	51	54	5	3
1	M034	P.S. 34 - M	M034	P.S. 34 - M	258	381	68	36	3	3
1	M063	P.S. 63 - M	M063	P.S. 63 - M	186	256	73	34	4	2
1	M064	P.S. 64 - M	M064	P.S. 64 (OLD 71) - M	215	324	66	36	1	2
1	M080	MANHATTAN CHARTER SCHOOL II	M056	J.H.S. 56 - M	110	391	28	0	0	2
1	M094	M094 SPED - M	* M015	P.S. 15 - M	41	96	43			
			* M061	P.S. 61 - M	45	84	54			
			* M188	P.S. 188 - M	103	122	84			
				M094 SPED - M TOTAL	189	302	63			
1	M110	P.S. 110 - M	M110	P.S. 110 - M	330	401	82	34	4	3
1	M134	P.S. 134 - M	M134	P.S. 134 - M	177	410	43	18	6	2

Conversion using tesseract ocr

Step1) convert the pdf to jpg

Step2) install tesseract on your pc. After installation go to system>advance system setting>environment variables>user variable>select path>click edit>add the path of tresseract. For ex: C:\Program Files\Tesseract-OCR

Press ok

Step3) open command prompt and cuse cd command to go to the location of the image file. Then type the following command: C:\Users\arora\Desktop\onelogica>tesseract name_of_file.jpg - -l eng

Output:

Copy this text into excel

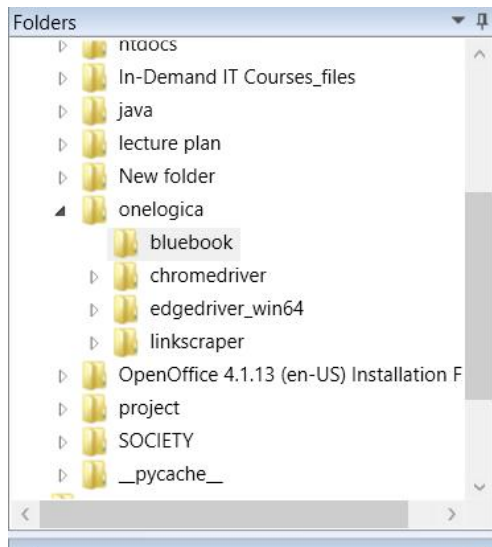it wasn't able to identify some of the characters efficiently. Other methods were more efficient

Converting using simple OCR

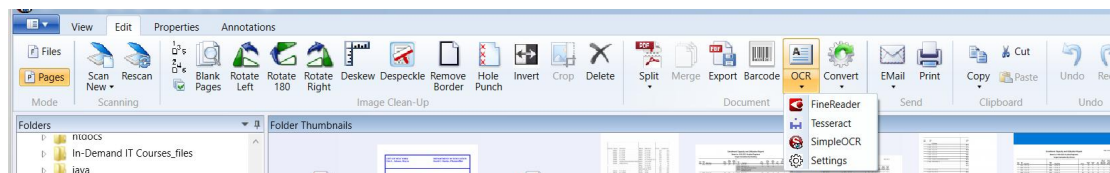Step1) install and set up simple ocr software

Step2)open and click on use free version
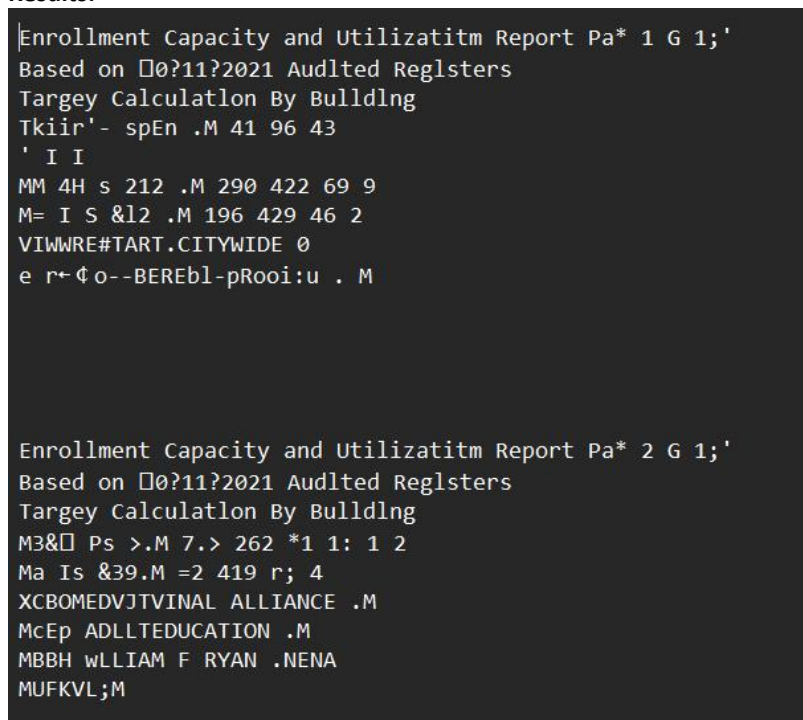
Step3) open the file by selecting it in the Folder section

Step 4) Go to edit tab. Click on ocr and select simple ocr

Step5) save the output.

Results:

```
Enrollment Capacity and Utilizatitm Report Pa* 1 G 1;'
Based on ▯0?11?2021 Audlted Reglsters
Targey Calculatlon By Bulldlng
Tkiir'- spEn .M 41 96 43
' I I
MM 4H s 212 .M 290 422 69 9
M= I S &l2 .M 196 429 46 2
VIWWRE#TART.CITYWIDE 0
e r←¢o--BEREbl-pRooi:u . M




Enrollment Capacity and Utilizatitm Report Pa* 2 G 1;'
Based on ▯0?11?2021 Audlted Reglsters
Targey Calculatlon By Bulldlng
M3&▯ Ps >.M 7.> 262 *1 1: 1 2
Ma Is &39.M =2 419 r; 4
XCBOMEDVJTVINAL ALLIANCE .M
McEp ADLLTEDUCATION .M
MBBH wLLIAM F RYAN .NENA
MUFKVL;M
```

it wasn't able to identify some of the characters efficiently too. Other methods were more efficient

Nanonets ocr online converter:

Step1) goto https://nanonets.com/blog/how-to-convert-pdf-file-to-excel-without-software/

Step2) click on try nanonets ocr and upload your file.

Step3) make the necessary changes and save your file

Output:

Using python to convert from pdf to excel
Step1) go to command prompt and type "pip install tabula"
Step2) open python idle and type the following code:

```
# Import Module
import tabula

# Read PDF File
# this contain a list
df = tabula.read_pdf("location_of_file", pages = 1)[0]

# Convert into Excel File
df.to_excel('output_ location/test.xlsx')
```

Output:

	A	B	C	D	E	F	G	H	I	J	K	L
		Unnamed:	Unnamed:	Unnamed:	Unnamed:	Unnamed:	Unnamed:	Unnamed:	Unnamed:	Unnamed:	No. of	No. of.1
0											Cluster/	Cluster
1	Geo	Org		Bldg			Target	Target	PK	Spec Rms	Rms
2	Dist	ID	Org Name	ID Bldg Na		Enroll	Cap	Util	Cap+	Reported+	Needed+
3	1	M015	P.S. 15 - M	M015 P.S.		179	251	71	17	1	2
4	1	M019	P.S. 19 - M	M019 P.S.		176	577	31	18	2	2
5	1	M020	P.S. 20 - M	M020 P.S.		351	691	51	54	5	3
6	1	M034	P.S. 34 - M	M034 P.S.		258	381	68	36	3	3
7	1	M063	P.S. 63 - M	M063 P.S.		186	256	73	34	4	2
8	1	M064	P.S. 64 - M	M064 P.S.		215	324	66	36	1	2
9	1	M080	MANHATT	M056 J.H.S		110	391	28	0	0	2
10	1	M094	M094 SPED	* M015 P.S		41	96	43			
11				* M061 P.S		45	84	54			
12				* M188 P.S		103	122	84			
13				M094 SPED		189	302	63			
14	1	M110	P.S. 110 - I	M110 P.S.		330	401	82	34	4	3
15	1	M134	P.S. 134 - I	M134 P.S.		177	410	43	18	6	2
16	1	M140	P.S. 140 - I	M140 P.S.		340	536	63	36	6	3
17	1	M142	P.S. 142 - I	M142 P.S.		330	547	60	89	2	3
18	1	M184	P.S. 184 - I	M137 P.S.		717	749	96	36	5	3
19	1	M188	P.S. 188 - I	M188 P.S.		390	522	75	18	3	3
20	1	M292	J.H.S. 292 -	M056 J.H.S		290	422	69		9	
21	1	M315	P.S. 315 - I	M061 P.S.		302	225	134	35	1	3
22	1	M320	MANHATT	M142 P.S.		208	220	95	0	1	2

It is more accurate as compared to OCR