ETL Pipeline for CPCB Data



ETL Flow diagram



Conversion of data

Extraction of data

- Attribute Selection
- Data appending
- Data Adding

Repeat the same process by another conversion software and subtract each corresponding cell and check for zero values



Data Validation

- Checking of extreme data
- Checking of Missing data
- Checking, fixing and removing outlier
- Mapping the data (Spatial data)

Data storing Format

- CSV file
- Excell file
- Shape file
- Tiff file
- Text file
- XML file

Open central pollution control board Website specifically NWMP DATA

https://cpcb.nic.in/nwmp-data/

Select specific year data i.e Year 2021

	country v Search	vise basin length - Google	
Central Pollution Control Board Ministry of Environment, Forest and Climate Change Government of India HOME ABOUT CPCB STANDARDS CPCB SU	ACTIVITIES		in Reader Skip to main content A · · · · · · · · · · · · · · · · · ·
Ore CPCB's Activities Water Quality Management Water Quality Data	NWMP Data	ory usage: 81.6 MB	
NWMP Data			Updated On : 22 Sep 2023
Quality Assurance/Quality + Year 2022 Control + Year 2021		Water Oual	ity Standards / Criteria
Vaste Management + Vear 2020 Contaminated Sites + Vear 2019 Vear 2019	Type of Water Body	Applic	ability of Standards/Criteria
Industrial Pollution + Year 2017 Noise Pollution + Year 2016	Rivers	Primary Water Quality https://cpcb.nic.in/wqn	Criteria for Outdoor Bathing n/Primary Water_Quality_Criteria.pdf
rban Pollution + Year 2015 + Year 2014	Lakes, Ponds & Tanks	Designated Best Use https://cpcb.nic.in/wqn	n/Designated Best Use Water Quality Criteria.pdf
ollution Control Planning + hformation Technology + Year 2013 Year 2012	Medium	Primary Water Quality https://cpcb.nic.in/wqn	Criteria for Outdoor Bathing n/Primary Water Quality Criteria.pdf
nvironmental Training + IGT/Court Cases	Rivers	Designated Best Use https://cpcb.nic.in/wqn	n/Designated_Best_Use_Water_Quality_Criteria.pdf
ublic Relations	Ground Water	BIS Drinking Water Spe https://cpcb.nic.in/wqn	cification IS 10500-2012 n/BIS Drinking Water_Specification.pdf Activat
nnual Reports	See Water/	Water Quality Standard	Go to Set

Download Water Quality of River-2021

Download Water Quality of Medium and Minor Rivers-2021

ne Sitemap RTI	Centra Ministry of I	al Pollution Control Board	Scree Apps by CPCB Jobs Ten	een Reader Skip to main ders Publication Tec	content A * •
оме авоит	срсв	STANDARDS - CPCB'S ACTIVITIES -	AIR WATER NOISE DATA	- LABORATORIES	 CONTACT US
me CPCB's Activitie	s 🔪 Water Quali	ty Management 🔪 Water Quality Data 🔪 NWMP Data 🔪 N	WMP Data 2020		
Environmental Acts	& Rules	NWMP Data 2020		Upo	lated On : 02 Sep 2022
nvironment Protect /ater Pollution	ion	WATER QUALITY I	DATA YEAR 2020		Water Quality Data
ir Pollution		Water Quality of Rivers -2020		Download	Year 2021
loise Pollution		Water Quality of Ground Water 2020 (State wise)		Download	Year 2020 Year 2019
	+	Water Quality of Lakes, Ponds and Tanks 2020 (State w	ise)	Download	Year 2018
		Water Quality of Canals - 2020		Download	Year 2017 Year 2016
		Water Quality of Creek Marine Sea Water Beach - 2020		Download	Year 2015
		Water Quality of Drains, STPs & WTPs-2020		Download	Year 2013
		Water Quality of Medium and Minor Rivers-2020		Download	rear 2012

Download both the fila and save in folder name WQ_2020

<mark></mark>	6/5/2024 2:12 PM	File folder	
Water Ouality data of River 2021	6/5/2024 1:04 PM	Microsoft Edge P	1.061 KB
wQuality_River-Data-2021	6/5/2024 1:03 PM	Microsoft Edge P	5,286 KB

Adobe

https://www.adobe.com > ... > Acrobat > Online Tools

Convert PDF to Excel online for free | Adobe Acrobat (India)

Convert **PDF to Excel** in just seconds. Adobe Acrobat **online** services turn your PDF content into an easily editable Microsoft Excel file. Try it for free!



Convert PDF to Excel

Drag and drop a PDF file to convert it into a Microsoft Excel spreadsheet.

Select a file 🗸 🗸

Select file from your folder WQ_2021

Down load the converted files in to your WQ_2021 folder

This software is paid

ILOVEPDF https://www.ilovepdf.com > pdf_to_excel

Convert PDF to Excel. PDF to XLS spreadsheets online



Select or drop Pdf file from your Directory and upload.

Click convert to excel button.

After few second or minute it convert to excel.

Now down load it in your desired directory.



There are two excel files

📴 Water_Quality_data_of_River_2021	6/5/2024 1:04 PM	Microsoft Edge P	1,061 KB
📴 WQuality_River-Data-2021	6/5/2024 1:03 PM	Microsoft Edge P	5,286 KB
Water_Quality_data_of_River_2021	6/5/2024 2:07 PM	Microsoft Excel W	52 KB
WQuality_River-Data-2021	6/5/2024 1:12 PM	Microsoft Excel W	295 KB

Open the excel file.

1		Wate	r Quality of N	/lediu	ım &	Mir	nor R	iver	s un	der N	WMI	P-20	21						
2	STN Code	Name of Monitoring Location	State Name	Tempe o	erature C	Diss Oxy (m	olved /gen g/L)	p	н	Condu (µmho	ctivity o/cm)	B((m)	DD g/L)	Nitra Ni N(n	te N + trite ng/L)	Fecal ((MPN/	Coliform (100ml)	Total G (MPN/:	Jiform LOOmi)
3				Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
4	1448	RIVER NAGAVALI AT THOTAPALLI REGULATOR, VIZIANAGARAM	ANDHRA PRADESH	23.0	32.0	5.9	8.2	7.0	8.2	242	400	1.4	2.0	0.72	1.66	7	15	93	150
5	4351	RIVER NAGAVALI IMMEDIATE BORDER BETWEEN ANDHRA PRADESH & ORISSA STATES, NEAR SIVALAYAM, KUNERU (V), KOMARADA (M)	ANDHRA PRADESH	24.0	32.0	5.2	8.3	6.7	8.5	240	612	1.2	2.5	0.92	3.02	11	21	150	210
6	4346	RIVER NAGAVALI NEAR NH- 16 BRIDGE, UPSTREAM (U/S) OF SRIKAKULAM TOWN, BEFORE CONFLUENCE OF TOWN SEWAGE	ANDHRA PRADESH	24.0	32.0	6.4	8.4	7.1	8.8	243	400	1.2	2.0	0.54	3.60	7	15	120	120
7	4347	RIVER NAGAVALI NEAR WATER PUMP HOUSE (KILLIPALEM) , DOWN STREAM (D/S) OF SRIKAKULAM TOWN, AFTER CONFLUENCE OF TOWN SEWAGE	ANDHRA PRADESH	24.0	32.0	6.0	7.5	7.1	8.4	330	648	1.4	2.3	0.33	4.02	11	15	150	210
8	4348	RIVER VAMSADHARA, IMMEDIATE BORDER BETWEEN ANDHRA PRADESH & ORISSA STATES, NEAR LALITHAMBA TEMPLE, BATTILI (V) BHAMINI (M)	ANDHRA PRADESH	22.0	28.0	5.0	8.3	6.9	8.5	221	410	1.5	2.2	0.42	2.46	9	21	120	210
9	2352	RIVER VAMSADHARA, KALINGAPATNAM, VIZIANAGARAM	ANDHRA PRADESH	24.0	32.0	5.8	8.8	7.0	8.3	228	420	1.5	2.2	0.42	3.05	7	15	120	150
10	1393	RIVER DAMANGANGA AT D/S OF MADHUBAN, DAMAN	DAMAN AND DIU, DADRA AND NAGAR HAVELI	26.2	30.0	6.6	7.9	7.6	8.5	104	251	1.0	1.0	0.30	0.30	5	9	24	39
11	2461	RIVER DAMANGANGA AT DAMAN JETTY, MOTI DAMAN	DAMAN AND DIU, DADRA AND NAGAR HAVELI	27.1	29.0	2.7	6.7	6.7	8.3	19214	25108	1.0	7.1	0.50	0.90	21	53	119	346
12	2460	RIVER DAMANGANGA AT DISCHARGE POINT OF DISTILLERY, DAMAN	DAMAN AND DIU, DADRA AND NAGAR HAVEL	27.1	30.0	4.8	6.1	6.7	8.0	15824	17943	2.6	6.1	0.30	0.50	36	110	297	427
13	2463	RIVER DAMANGANGA AT LAVACHA TEMPLE, SILVASSA	DAMAN AND DIU, DADRA AND NAGAR HAVELI	26.8	30.0	6.4	7.7	7.5	8.6	131	406	1.0	1.0	0.30	0.30	35	41	74	91
	2465	RIVER DAMANGANGA AT NAROLI BRIDGE, SILVASSA	DAMAN AND DIU,	26.4	30.5	6.6	7.5	7.4	8.9	112	277	1.0	1.0	0.30	0.50	12	18	31	38
	•	Table 1 🕀																	
Rea	dy															E	# E	—	

Delete column "Name of Monitoring Location" and "State Name"

Water Quality of Medium & Minor Rivers under NWMP-2021

STN Code	Tempe o	rature C	Diss Oxy (m	olved ygen g/L)	p	Н	Condu (µmho	ctivity o/cm)	B((m)	DD g/L)	Nitra Nit N(n	te N + trite ng/L)	Fecal ((MPN/	Coliform (100ml)	Total Co (MPN/1	oliform LOOml)
]	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
1448	23.0	32.0	5.9	8.2	7.0	8.2	242	400	1.4	2.0	0.72	1.66	7	15	93	150
4351	24.0	32.0	5.2	8.3	6.7	8.5	240	612	1.2	2.5	0.92	3.02	11	21	150	210

Replace top three rows by following rows

	Tmp	Tmp					Con	Con	Bod	Bod			Fa_	Fa_	To_	To_
	_Mi	_M	Do_	Do_	pH_	pH_	_Mi	_Ma	_Mi	_Ma	Nit_	Nit_	Co_	Co_	Co_	Co_
st_C	n_2	ax_	Min	Max	Min	Max	n_2	x_2	n_2	x_2	Min	Max	Min	Max	Min	Max
ode	1	21	_21	_21	_21	_21	1	1	1	1	_21	_21	_21	_21	_21	_21

Here Tmp_Min_21: Minimum Temperature of Year 2021

Similarly make appropriate column headings for other parameters

Now the top row looking like following excel sheet

	Tmp_	Tmp_	Do_	Do_	pH_	pH_	Con_	Con_	Bod_	Bod_	Nit_		Fa_Co		To_Co	
st_Co	Min_2	Max_	Min_	Max_	Min_	Max_	Min_2	Max_2	Min_	Max_	Min_	Nit_M	_Min_	Fa_Co_	_Min_2	To_Co_
de	0	21	21	21	21	21	1	1	21	21	21	ax_21	21	Max_21	1	Max_21
1448	23.0	32.0	5.9	8.2	7.0	8.2	242	400	1.4	2.0	0.72	1.66	7	15	93	150
4351	24.0	32.0	5.2	8.3	6.7	8.5	240	612	1.2	2.5	0.92	3.02	11	21	150	210
4346	24.0	32.0	6.4	8.4	7.1	8.8	243	400	1.2	2.0	0.54	3.60	7	15	120	120
4347	24.0	32.0	6.0	7.5	7.1	8.4	330	648	1.4	2.3	0.33	4.02	11	15	150	210
4348	22.0	28.0	5.0	8.3	6.9	8.5	221	410	1.5	2.2	0.42	2.46	9	21	120	210
2352	24.0	32.0	5.8	8.8	7.0	8.3	228	420	1.5	2.2	0.42	3.05	7	15	120	150
1393	26.2	30.0	6.6	7.9	7.6	8.5	104	251	1.0	1.0	0.30	0.30	5	9	24	39

Now save the file as "Water_Quality_Minor_River_2021"

Now for another file which have water quality data of rivers. It contains data in many tables.

E	יף ∃	E P = WQuality_River-Data-2021 - Excel (Product Activation Failed)	8	a –	
F	le Hon	ne New Tab Insert PageLayout Formulas Data Review View PowerPivot Q Tell me what you want to do			A Share
A4		× √ A 2			
1	A	B WATER QUALITY DATA OF RIVERS MONITORED UNDER NATIONAL WATER QUALITY MONITORING PROGRAMME (NWMP), 2021 INDEX	С	D	E
,	TABLE NO	. TABLE	PAGE NO.		
3	1	WATER QUALITY OF RIVER BEAS	1-3		
4	2	WATER QUALITY OF RIVERSUTLEJ	4-7		
5	2.1	WATER QUALITY OF TRIBUTARY STREAMS - RAVI, PARVATI, LARG, SIRSA, SIUEL, SUKETI KHAD, BINWA, NEUGAL, SPITI, BASPA, BANGANGA, BASANTER, CHENAB, DEVAK, TAWI, MANWAR, CHUNT KOL, GAWKDAL, JHELUM, LIDDER, SINDH, UJH, SARSA NADI, SWAN, TIRTHAN, PABBAR, ALHI NOG, HARABAGH, KUNNI PUL, ASHWANI, MANJHI KHAD, MOL KHAD, BANER KHAD, CHARAN KHAD, ASHWANI KHAD, GUMMA NAUTI KHA, DHAULI KHAD, SHIKARI KHAD, BATTA AND KALI BEIN	8-15		
6	3	WATER QUALITY OF RIVER GANGA	16-23		
7	4	WATER QUALITY OF RIVER YAMUNA	24-27		
8	4.1	WATER QUALITY OF TRIBUTARY STREAMS - TONS, GIRI, PABBAR, HINDON, BETWA, KALI SINDH, DHANARI DAM, BANAS, NEWTA DAM, CHAPPI DAM, UJAD DAM, CHAMBAL, PARVATI, KHAN AND KSHIPRA (2021)	28-32		
0	4.2	WATER QUALITY OF TRIBUTARY STREAMS - SUSWA, BHELLA, DHELLA, KOSHI, GOLA, KICCHA, PILKHAR, NANDOUR, KALYANI, RAMGANGA, KALI (E), KALINADI, BANGANGA, VARUNA, SAI, GOMTI, BANGANGA, RAPTI, GHAGHARA AND RIHAND (2021)	33-37		
10	4.3	WATER QUALITY OF TRIBUTARY STREAMS - SONE, TONS, GANDAK, SIKRAHNA, BURHI GANDAK, HARBORA, KAMALA, KOSHI, DAHA, DHOS, RAMREKHA, SIRSIYA, BAGMATI, MANUSMAR, POONFUN, KOHRA, HARHA, LAKHANDEI, GANGI AND PARMAR (2021)	38-42		
11	4.4	WATER QUALITY OF TRIBUTARY STREAMS - BICHIA, SANKH, SONE, VAISHALI, KARIYARI, NEWAJ, CHAMLA, CHOUPAN, GOUR, JAMUNI, SONE, BEEHAR, BETWA, KALIASOT, PARVATI, GOPAD, JOHILA, SINDH AND KOLAR DAM (2021)	43-45		
12	4.5	WATER QUALITY OF TRIBUTARY STREAMS - JUMAR, AJAY, KONAR, GARGA. BOKARO, SANKH, BARAKAR, DAMODAR, MATHABHANGA, MAYURAKSHI, CHURNI, JALANGI, RUPNARAYAN, DWARKA, DWARAKESHWAR, BINDYADHARI, KANSI, BARAKAR AND MAHANANDA (2021)	46-49		
13	4.6	WATER QUALITY OF TRIBUTARY STREAMS - KODRA DAM, BANAS, KELWARA, CHAMBAL, PIPLAD DAM, KOTHARI, GUWARDI RESERVOIR, BERECH, GAMBHIRI AND KANOTA DAM (2021)	50-52		
14	5	WATER QUALITY OF RIVER BRAHAMPUTRA	53		
	•	WATER QUALITY OF TRIBUTARY STREAMS - BARAK, BARALIA, BARNADI, BEKI, BERENG, BHARALU, BHCOCO, BOGINADI, BOLDAMGRE, BORSILA, BURADIA, BURHIDIHING, CITVAT CHATHE, CHINIMTIDIHI CHITTEYSTREAM DAMDIAL STREAM DAMBING, DEMOW DHANSIL DHATESHWARI DIRAXC, DICABLI, DIGOL DHINDO, DIKCHI DIKHWO DIKKONG, DISANG, Table 1 Table 2 Table 3 Table 4 Table 5 Table 6 Table 7 Table 8 Table 9 Table 10 Table 11 Table 12 Table 13 Table 14 Table 15 Table 16 Table 16 Table 17 Ta	e \ 54-67 IOV tings to activ ble 18 Tab	/S hate Windo le 19	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,

First table shows the index document, it need not to change anything.

Table 2 sheet has the water quality data of river beas

1	TABLE 1 -: WATER QUALITY DATA OF RIVER BEAS (2021) NAME OF MONITORING CODE NAME OF MONITORING LOCATION TEMPERATURE ('C) DISSOLVED OXYGEN ('MP) CONDUCTIVITY (µmhos/cm) BIO-CHEMICAL OXYGEN (µmhos/cm) NITRATE (mg/L) FECAL COLIFORM (MPN/100ML) TOTAL COLIFORM (MPN/100ML) STREPTOCC (MPN/100ML)																				
2	STATION CODE	NAME OF MONITORING LOCATION	STATE NAME	TEMPER/ (°C)	ATURE	DISSO OX1 (m	dlved Ygen g/l)	1	pH	CONDU (µmh	CTIVITY os/cm)	BIO- CHI OXY DEM (mg	emical 'gen Iand 9/L)	NIT (m	RATE g/L)	FECA COLIF((MPN/10	AL DRM IOML)	TOTAL CO (MPN/1	DLIFORM 00ML)	FE STREPT (MPN/	CAL TOCOCCI 100ML)
3				MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX
	PRIMARY W	ATER QUALITY CRIT	ERIA NOTIFIED			>	5.0	6.5	-8.5			<3	3.0			< 25	00			<	500
4	UNDER E(P)	RULES, 1986														MPN/10	OML			MPN/	100ML
	1001	RIVER BEAS AT U/S	HIMACHAL	4.0	14.0	7.9	9.7	7.2	8.0	80	270	1.0	1.0	0.30	0.72	2	110	25	540	2	2
5		MANALI	PRADESH																		

Delete column "Name of Monitoring Location" and "State Name" and replace top 5 rows by following row

Extraction of data

- Data appending
- Data Adding

		Tm												Fa_		To_	Fa_	Fa_
	Tm	p_		Do_		pH_	Con	Con	Bod	Bod	Nit	Nit	Fa_	Co_	To_	Co_	St_	St_
st_	p_	Ma	Do_	Ma	pH_	Ma	_Mi	_M	_Mi	_M	_Mi	_M	Co_	Ma	Co_	Ma	Mi	Ma
Cod	Min	x_2	Min	x_2	Min	x_2	n_2	ax_	n_2	ax_	n_2	ax_	Min	x_2	Min	x_2	n_2	x_2
e	_21	1	_21	1	_21	1	1	21	1	21	1	21	_21	1	_21	1	1	1

Open a new excel file and keep the top row as above mentioned

	금 ち・ご E P = Book1 - Excel (Product Activation Failed) 団 File Home New Tab Insert Page Layout Formulas Data Review View Power Pivot Q Tell me what you want to do															d —				
	File	Home	New Tab	Insert P	age Layout	Formulas	Data	Review	View	Power Pi	vot Q⊺	fell me what	you want to	o do					只 Sha	are
A	1	•	× 🗸	<i>f</i> ∗ st	_Code															~
	Α	В	С	D	E	F	G	н	1	J	К	L	М	N	0	Р	Q	R	S	
1	st_Code	Tmp_M 21	in_ Tmp_Max_ 21	Do_Min_2	Do_Max_2 1	pH_Min_2 1	pH_Max_2 1	Con_Min_2 1	Con_Max_ 21	Bod_Min_ 21	Bod_Max_ 21	Nit_Min_2 1	Nit_Max_2 1	Fa_Co_Min _21	Fa_Co_Ma x_21	To_Co_Min _21	To_Co_Ma x_21	Fa_St_Min _21	Fa_St_Max _21	

Copy the data from table 2 to new excel sheet

E	ا ا	e e - E	م 🗉	•			WQu	ality_Rive	er-Data-20	21 - Exce	l (Produ	ct Acti	vation Fa	iled)			œ	-		,	(8	ه. د	÷ - 🖽	<u>م</u>	•						Book2	Excel (roduct	Activatio	n Failed)						c		-	o x
F						Page													Я	Share																									유 Share
					e I	4.0	de															4		•	×	V 1	f _x	st_Cod	le																
1						SI_CO	lue															4 A	A	8	с	D		E	F	G		4	1	J	K	L		м	N	0	P		2	R	S
L.	t_Code	Tmp_M	f Tmp_N	1 Do_M	Do_Ma	a p	pH_	Con_	Con_Ma	Bod_Mi	Bod_M	Nit_	Nit_M	Fa_Co_	Fa_C	To_Co_	To_Co_	Fa_S	Fa_St	M	Ê	st_C	Cod Tr	np_M	Tmp_M	Do_M	lin Do	Ма р	H_Min	pH_M	a Con	_Mi C	on_M	Bod_M	li Bod	M Nit_	Mi Nit	Ma	Fa_Co_	Fa_Co	To_C	o_ To_	Co_ Fa	St_ F	Fa_St_
-	1001	4.0	ax 21 14.0	in 21 7.9	9.7	7.2	8.0	80 80	270	1.0	1.0	0.30	8x 21 0.72	2 2	110	Min 21 25	Max 21 540	2		1	1	e	in	20	ax 20	_20	× 2	0	20	x_20	n 2	D a	x_20	n_20	ax_20	n_20	x 2	1 05	Min 20	Max 2	0 Min	20 Ma	20 M	n_20 f	Max_20
2	2601	4.0	14.0	7.7	9.4	7.1	7.8	70	116	1.0	1.0	0.32	0.88	40	350	240	1600	2			3																								
3	4444	3.0	13.0	7.7	9.8	6.9	7.9	49	88	1.0	1.0	0.32	0.92	94	920	430	1600	2			4																								-
4	4037	4.0	13.0	7.8	9.9	6.9	7.9	45	94	1.0	1.0	0.32	0.82	33	140	170	430	2	-		6																								
5	<u> </u>	-		-	-	-	-					-							-	_	8							N																	
6	3866	4.0	12.0	7.9	9.8	7.1	8.4	37	92	1.0	1.0	0.32	0.62	26	94	120	350	2			1	2																							-
7	2602	6.0	16.0	5.8	9.1	7.3	8.1	67	127	1.0	1.0	0.32	1.22	70	140	350	920	2			1	1							1																
\$	4445	5.0	15.0	7.8	8.8	7.2	7.9	67	233	1.0	1.0	0.32	1.62	79	920	430	1600	2	2		1	3																							
•	1002	5.0	15.0	7.6	8.8	7.2	7.9	78	145	1.0	1.0	0.32	1.82	94	540	430	1600	2	2		1	4									1														_
10	1003	8.0	14.0	7.5	8.8	6.9	8.1	62	120	1.0	1.0	0.32	0.92	47	220	280	920	2	2		1	5									1	•													_
	1004	7.0	23.0	6.6	10.8	7.1	8.0	62	126	1.0	1.0	0.34	1.07	33	94	150	350	2	2		1	7 B								1															-
	2603	7.0	23.0	8.2	10.3	6.9	8.0	82	188	1.0	1.0	0.32	1.92	58	140	280	540	2	F		1	-		-						/															_
12	1005	12.0	22.0			73		75	264	10	10	0.22	0.97	42	140	220	540	,			2	1							/																
13	1550	7.0	23.0	8.4	11.2	7.3	8.2	71	180	1.0	1.0	0.40	1.73	21	70	110	280	2		_	2	2						V																	_
14	1006	8.0	23.0		10.8	7.2			224	10	10	0.13	2.61	70	170	240	920	2			2	4																							
15	2000		25.0		-	7.0			103	1.0	10	0.40	2.00				20		_		2	5																							
16	2604	16.0	26.0	8.4	3.5	7.3	8.2	118	392	1.0	1.0	0.40	3.10	2	2	"	39	2			2	7 B																							-
17	1007	15.0	28.5	8.6	9.6	7.3	8.2	109	380	1.0	1.0	0.40	z.40	2	2	21	39	2			- 2	9																							
	•	Tal	ble 1	Table	2 1	Table 3	Tal	ole 4	Table 5	Tab	le 6	Table	7 T	able 8	Table	9 Tab	le 10	Tab	. (+			-		She	et1	۲											4				1				•
Rea	\$y													III	(H)	m -		-	-+	1005	5	elect de	estinatio	n and pr	ress ENTE	R or cho	iose Pas	ste										Count	E 19	III 1	II) P				+ 100%

Similarly for table 3, delete two column and copy the data portion and append in new sheet where previously tables data was copied.

. 6	- د ا	¢- 🗉 .	₊ م					Bo	ook1 - Excel	(Product Ad	tivation Fail	ed)					٦	a –		×
	le He	ome Ne	ew Tab	Insert P	age Layout	Formula	is Data	Review	View	Power Pi	vot Qʻ	Tell me what	t you want to						₽ Sha	ire
				1															_	
A1		• E	×	∫x st_	Code															~
	A	В	С	D	E	F	G	н	1	J	К	L	м	N	0	Р	Q	R	S	
22	4436	10.3	29.0	7.7	9.5	7.0	8.1	131	561	1.0	1.0	0.30	1.60	2	2	24	49	2	2	
23	4437	10.2	29.0	7.8	9.8	7.3	8.0	122	307	1.0	1.0	0.30	1.10	2	2	24	39	2	2	
24	4438	10.2	29.0	7.8	9.7	7.2	8.1	138	326	1.0	1.0	0.34	2.50	2	2	26	49	2	2	
25	4443	5.0	15.0	7.5	9.3	7.0	8.3	71	118	1.0	1.0	0.32	1.07	32	110	210	430	2	2	
26	1693	8.0	36.0	6.9	8.3	7.5	8.3	158	244	1.0	1.2	0.40	1.00	18	140	84	260	2	2	1
27	1694	11.0	38.0	7.0	8.8	7.0	8.5	194	312	1.1	1.5	0.40	1.30	40	130	120	430	-	-	
28	1010	10.0	38.0	7.1	9.8	7.6	8.4	195	318	1.1	1.6	0.30	1.20	32	130	100	260	-	-	1
29	1695	11.0	38.0	6.8	8.1	7.5	8.2	200	302	1.1	1.7	0.50	1.00	40	130	140	470	-	-	1
30	1294	18.0	34.0	6.3	8.0	7.3	8.2	197	328	1.0	1.8	0.60	2.00	68	170	200	490	-	-	
31	1696	15.0	29.0	7.1	9.3	7.7	8.3	166	248	1.0	1.3	0.32	1.00	63	130	220	470	-	-	1
32	1697	15.0	29.0	7.2	9.5	7.7	8.0	184	245	1.0	1.4	0.50	1.10	78	260	330	490	-	-	
33	1011	15.0	34.0	6.6	9.1	7.3	8.1	185	254	1.0	1.7	0.60	1.20	70	210	280	840	-	-	
34	1012	15.0	29.0	7.0	9.2	7.7	8.2	169	274	1.0	1.4	0.40	1.10	70	170	280	490	-	-	
35	4164	18.0	33.0	6.9	9.0	7.6	8.2	189	1922	1.0	1.5	0.50	1.90	68	200	260	470	-		
36	1867	10.0	10.0	9.3	9.3	7.6	7.6	3538	3538	1.0	1.0	0.32	0.32	2	2	1600	1600	2	2	
37	2611	10.0	10.0	9.1	9.1	7.9	7.9	3637	3637	1.0	1.0	0.32	0.32	7	7	1600	1600	2	2	
38	3872	6.0	13.0	8.8	9.3	7.5	8.4	1107	5072	1.0	1.4	0.32	1.42	2	1600	79	1600	2	4	
39	4448	6.0	13.5	8.8	9.2	7.6	8.4	1166	5078	1.0	1.0	0.32	0.55	2	1600	46	1600	2	4	
40	4449	6.5	13.0	8.9	9.3	7.7	8.4	1164	5087	1.0	1.0	0.32	1.41	2	240	70	1600	2	4	
41	4452	7.0	13.5	8.7	9.3	7.6	8.3	1108	4635	1.0	1.0	0.32	1.22	2	1600	70	1600	2	2	
42	3871	6.5	13.0	8.7	9.1	7.5	8.4	1111	5046	1.0	1.0	0.32	1.22	2	240	70	1600	2	2	
43	3870	7.0	13.5	8.7	9.3	7.5	8.4	1159	4836	1.0	1.0	0.32	1.32	2	79	110	1600	2	2	
44	1389	7.0	13.0	8.7	9.2	7.6	8.4	1124	4518	1.0	1.0	0.32	1.82	2	1600	130	1600	2	7	
45	4453	7.0	13.5	8.7	9.2	7.4	8.3	1143	4559	1.0	1.0	0.32	1.22	2	110	130	1600	2	2	
46	3869	8.0	14.9	8.5	9.1	7.2	8.3	257	2602	1.0	1.0	0.32	2.32	5	1600	350	1600	2	7	
47	1086	8.0	14.5	8.6	9.1	7.5	8.5	1086	3840	1.0	1.0	0.32	1.32	8	1600	540	1600	2	94	
48	1087	8.0	14.5	8.6	9.1	7.0	8.2	1020	2681	1.0	1.2	0.32	1.92	9	1600	1600	1600	2	220	
49																				
50																				
51																				A
52																				19
	•	Sheet1	(+))																
Read	ly		_											Cou	unt: 19		─ -		+ 10	00%

Similarly do for all table.

vote		_																	
Do not copy abbr	eviat	tion a	and	not	e w	hic	h giv	/en	in di	ffer	ent	tal	bles	. Fe	w ex	am	ples	are	belo
PRIMARY WATER QUALITY CRITER 17 UNDER E(P) RULES, 1986	A NOTIFIED	0			> 5.0		6.5-8.5				< 3.0			< MPN	2500 /100ML				< 500 MPN/100N
	_			_				_		_	_	_				_	_		
TABLE - 16	1: WA	TER Q	UALI	TYO	F TRI	BUT	TARY	STR	EAMS	BUD	HAE	BALA	ANGA	AND	SONC	(202	1)	1	
STATION CODE	STATE NAME	TEMPER. (°C)	ATURE	OX (m	OLVED YGEN ng/L)		рн	CONI (µm	DUCTIVITY hos/cm)	CHEN OXY DEM	D- IICAL GEN AND	NIT (m	RATE g/ L)	(MPN/	OLIFORM	COL (MPN	IFORM 100 ML)	STREP (MPN	TOCOCCI 1/100ML)
		MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX
PRIMARY WATER QUALITY CRITERIA				>	5.0	6.	5-8.5			<:	0.0			<2 MDN/	500			< 500 MF	PN/100ML
NOTE: ABOVE DATA IS AS PER I THE INTER-STATE MONT BOL VALUES FOR DISSOL MPM/100 ML). ABBREVIATIONS	HE INFORM ORING LOC	MATION PR CATIONS AR EN (0.3 MG,	OVIDED I RE MONIT /L), CONE	BY THE ORED B DUCTIVI	SPCBS/PO Y CPCB T TY (5 µM	CCS TH HROUG HOS/C	HROUGH E GH HEAD CM), BOD (WQDES OFFICE [1 MG/I	PORTAL. F (DELHI) AN .), NITRATE	URTHER D REGION N + NIT	CLARIFI KAL DIR RITE-N	ICATION ECTOR	NS, IF AN ATES. G/L), FAE	Y MAY PL	EASE BE O	BTAINEI	FROM R	ESPECTIV	E SPCB/PCC
1. U/S - UP-STREAM					5. B	DL - B	ELOW DET	ECTION	LIMIT										
2. D/S - DOWNSTREAM					6. M	G/L -	MILLIGRA	M PER L	ITER										
3. A/C - AFTER CONFLUEN	CE				7. M	PN/ 1	00 ML - M	OST PR	OBABLE NU	MOCO / 4	00 MIL1	LILITER							

Save the file in excel and csv format.

Repeat the extraction of data proves from different software (small pdf)

Copy the converted data from both the file in one excel file in different sheet.

Subtract each corresponding numeric cell from sheet-1 to sheet -2 in sheet-3 using following Formula

='Table 1'!D4-'Table2'!D4

For string data ise following formula

=EXACT('Table 1'!D4,'Table 2'!C4)

Sheet 1 sheet2 and sheet 3 is shown below.

Table-1

2

	Water (Quality of M	Iediu	m &	Mir	nor R	iver	s un	der	NWN	/IP-2	021						
STN Cade	Hame of Munituring Lucation	State Hame	Tamp	eratu e C	Dirs 4 Ox (m	ralva 17 qan qfL)		H	Can vity (um)	ducti kufc	B(0D (/L)	Hitr + Hi H(=	ato H itrito 19/L)	Face Cali (MP	d farm H/100	Tatal Calif (MPI	
			His	Hez	His	Hez	His	He	His	Hez	Min	He	His	Hex	His	Hez	His	Hex
1448	RITER BACATALI AT TROTAPALLI RECULATOR, Tiziabacaram	ABPBEA PEAPESS	23.8	32.8	5.3	8.2	7.8	8.Z	242		1.4	2.8	8.72	1.66	7	15	33	158
4951	EITEE BAGATALI IMMEDIATE DORDER Detween andrea pradesn 6 orissa states, Near sitalatam, knnern [t],	ANDURA Pradesu	24.8	92.8	5.2	8.3	6.7	. .5	24 8	612	1.2	2.5	8.32	9.8Z	11	21	150	218
4346	RIYER MAGAYALI MEAR MM- 16 DRIDGE, Mpstream (M/S) of Srieaemlam Town, Defore compluence of Town	ABPBEA Peapesu	24.8	92. 8	5.4	1.4	7.1	8.8	243		1.2	2.8	8.54	3.68	,	15	128	128
4347	RITER BAGATALI BEAR WATER PAMP BOBSE EILLIPALEM] , DOWB STREAM (D/S) OF SRIEAEBLAM TOWB, AFTER	ABPBEA Peapesu	24.8	92. 8	6.8	7.5	7.1	8.4	338	648	1.4	2.3	8.33	4.8Z	11	15	158	218
4141	RITER TANSADBARA, IMMEDIATE DORDER		22.8	28.8	5.8	1.1	6.5	I. 5	221	418	1.5	2.2	1.42	2.46		21	128	218

Table-2

Water Quality of Medium & Minor Rivers under NWMP-2021

		-																
STH Cada	Name of Munituring Lucation	State Hame	Temp	eretu G	Din 4 Os (m	ralva 17 qan qfL)		H	Can vity (pm)	ducti ku/c	8 (m)	0D (/L)	Hitr + Hi H(=	ato H itrito 197L)	Fac Cali (MP	el Farm H/100	Tata Calif (MP)	 /1001
			His	Hex	Mis	Hez	His	He	His	Hex	Mis	He	His	Hex	His	Hex	His	Hex
1448	RITER BACATALI AT TROTAPALLI RECOLATOR, Tiziabacaram	ABDBEA PEADESB	23.8	32.8	5.5	8.Z	7.8	8.Z	242		1.4	2.8	8.72	1.66	'	15	33	158
4951	EIVEE BAGATALI IHHEDIATE DAEDEE Devveen andrea peadesn & aeissa states, Neae sivalatah, ennern ivi,	ANDURA PRADESU	24.B	32.8	5.2	8.3	6.7	8.5	248	61Z	1.2	2.5	8.32	3.8Z	11	21	158	218
4346	EIYEE BAGATALI NEAE NN- 15 DEIDGE, NPSTREAH (N/S) OF SRIEAENLAH TOWN, DEFORE CONFLUENCE OF TOWN	ANDURA Pradesu	24.8	32.8	5.4	8.4	7.1	8.8	243		1.2	2.8	8.54	3.68	7	15	128	128
	1			·														

Table-3

0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Repeat the same process by another conversion software and subtract each corresponding cell and check for zero values If all the cell has zero . it means the conversion is proper. If there is error in any cell then check both the cell. It may be there some special character like "-" "n.a.". Check the values from Original pdf and make correction.

Now the converted data is available. Sometime in the conversion process additional row may occurs. This type error is detected in the validation process. It may

Data Cleaning

- Duplicate removal
- Special character
- Noise removal
- Fixing Data type
- Standardize capitalization

possible, it is not detected in validation process if that is not detected. Sometime in the appending or adding the data in the table it may possible some duplicate row may be a added.

It is also possible in original file there may be some special character for foot note , Some time blank space is also represented by in the "*", "N.A","\$\$","!"

Duplicate Removal

- First find out which column has unique no.. In present case station id is unique for each row . here we compare by two consecutive row by following formula
- EXACT(A4, A5) if the result is Flse , it means these two rows are not dulicate .

	А	В	с		V
1		Water Q	uality of Me		
2	STN Code	Name of Monitoring Location	State Name		
3				_	FALSE
4	1448	RIVER NAGAVALIAT THOTAPALLI REGULATOR, VIZIANAGARAM	ANDHRA PRADESH	EXACT(A4, A5)	FALSE
÷	4351	RIVER NAGAVALI IMMEDIATE BORDER BETWEEN ANDHRA			THESE
5		PRADESH & ORISSA STATES, NEAR SIVALAYAM, KUNERU	ANDHRA PRADESH		FALSE
5	4346	RIVER NAGAVALI NEAR NH- 16 BRIDGE, UPSTREAM (U/S) OF SRIKAKUI AM TOWN, BEFORE CONFLUENCE OF TOWN	ANDHRA PRADESH	-	
6		SEWAGE		_	FALSE
_	4347	RIVER NAGAVALI NEAR WATER PUMP HOUSE (KILLIPALEM) , DOWN STREAM (D/S) OF SRIKAKULAM	ANDHRA PRADESH	-	
7	4348	IOWN, AFTER DIVED VAMSADHADA IMMEDIATE BODDED BETWEEN		-	FALSE
8	7540	ANDHRA PRADESH & ORISSA STATES, NEAR LALITHAMBA	ANDHRA PRADESH		

Special charterer Removal

Use replace statement to remove noise and special characters

Fixing data type

Before finalizing the data it is needed to check the data type of data.

Find and Replace	?	×
Fin <u>d</u> Re <u>p</u> lace <u>G</u> o To		
Find what: S		\sim
Replace with:		~
More >> Beplace Replace All Eind Next	Canc	el

Weather data is Integer, floating, Number of decimal place, Length of string.

In CPCB data base temperature is numeric with single decimal. Station Id as String of 10 character. The precision of data should be decided on the original data which has been transformed.

Checking of Extreme data

In excel, we can find out the extreme values of any row or any column by following formula

Subtotal (4, starting cell of selected Column: last cell of Selected Column)

It will give the maximum value selected column

Similarly

e Charling of outr

Data Validation

- Checking of extreme data
- Checking of Missing data
- Checking, fixing and removing outlier
- Mapping the data (Spatial data)

Subtotal (5, starting cell of selected Column: last cell of Selected Column) will give the minimum value

By finding out the maximum value and minimum value of column, one can make judgment based on the specific characteristic parameter. i.e the temperature of water bodies generally varies between 4°C to 40°C. of the max value and Min value beyond this range, the re check it from the original data/ PDF file.

We can also check that Max temperature should be more then minimum temperature . by subtractin t_max by t_min. similar analysis should also perform on other parameters

If the some values is unexplainable then it may remove.

Mapping the data

Sometime the data is spatial in nature in that case the values of X and Y or Latitude and longitude then compare the values weather it is coming in the targeted location. Location data (X, Y or latitude/longitude) should be projected and check whether the data point or shapes is projecting within the specific boundary. Here we can use open source GIS software (QGIS) or commercial software (ARCGIS)



Data storing format

🗄 🔎 Type here to search

There are so many formats for the Data storage. We have to select the data format such a manner that is acceptable by many software and simple. For CPCB data, we are saving the data in Coma separated Value (CSV) format

٨

o 🗄 💼 💽 🥥 🧮 🔍 🦧 📵 💽 🖬 🕼