Kausi Milestone		/ -O - IVI	lilestone 3	8							
	k Chattapa	adhyay									
• Rep	least 5 data transford lace Headers mat data into a			our website data. Fo	or example:						
• Ide	ntify outliers a d duplicates	nd bad data									
	casing or incon duct Fuzzy Match		ues								
# Load th	ill be scrapping data		ww.worldometers.ir	nfo/coronavirus/#cou	ntries to get the co	vid details for all	countries.				
<pre>import nu</pre>	, sys										
<pre>import ur from bs4 import ma</pre>	quests, io llib.request import BeautifulS tplotlib.pyplot a lotting packages	_									
<pre>import ma # advance import se # interac</pre>	tplotlib.pyplot a ed ploting aborn as sns etive visualization	n									
<pre>import pl # import</pre>	otly.express as protly.graph_objs a plotly.figure_fac ly.subplots impor	s go tory as ff	ots								
url = "ht response	the website data tps://www.worldom requests.get(ur	eters.info/co	oronavirus/#coun	tries"							
<pre>def p r s</pre>	TableParser: arse_url(self, ur esponse = request oup = BeautifulSo eturn [(table['id	s.get(url) up(response.t		ole)) for table i r	soup.find_all(table')]					
def p n n	<pre>### In parser method to the content of the con</pre>	_									
#	Find number of r we also find the or row in table.f # Determine th td_tags = row. if len(td_tags	column title ind_all('tr') e number of i find_all('td'	es if we can): rows in the tabl	i.e							
	n_rows+=1 if n_column # Set	ns == 0: the number or mns = len(td_		ır table							
	<pre>th_tags = row. if len(th_tags for th in</pre>	<pre>find_all('th') > 0 and ler th_tags: a = th.get_te a = colData.r</pre>	<pre>') n(column_names) ext() replace('/','').</pre>	== 0: replace(' ','').r	replace(',','').	replace('\n',''	').replace('	','').replace	e(' ','')	
	f len(column_name	s) > 0 and le	en(column_names)	<pre>!= n_columns: match the number</pre>	of columns")						
d	<pre>olumns = column_n f = pd.DataFrame(ow_marker = 0</pre>			_	_columns)						
f	<pre>or row in table.f column_marker columns = row. for column in df.iat[row] column mar</pre>	= 0 find_all('td' columns: _marker,colum		umn.get_text()							
	<pre>if len(columns row_marker * Convert to float</pre>) > 0: += 1									
	except ValueEr	df[col].astyr ror:	pe(float)								
# Parsing	eturn df Html data TableParser() p.parse_url(url)[0][1]									
# Grabbin table hea	g the table from d(10)	the tuple NewCases T						TotalTests Te	ests1Mpop P	opulation	Continent
1	\nNorth America\n \nAsia\n 212,704,671 nEurope\n 244,739,871	1 +109,988 1 +8,796	1,596,552 1,528,929 2,005,299	118,266,265 +322 197,200,461 +75 240,413,530	+90,358	3,466,210 13,975,281 2,321,042	8,730 15,996 6,542				North America Asia Europe
4 \n0	\nSouth America\n 67,686,626 Oceania\n 13,888,348 \nAfrica\n 12,773,223	3 +25	1,347,141 25,232 258,499	65,810,786 13,718,857 12,058,924	+48	528,699 144,259 455,800	10,233 77 547			Αι	South America ustralia/Oceania Africa
6 7 8 1 9 2	\n\n 721 World 675,122,487 USA 104,183,562 India 44,682,784	7 +118,809 2	15 6,761,667 1,132,719 530,740	706 +397 647,469,529 101,290,878 44,150,289	+136,811	0 20,891,291 1,759,965 1,755		1,159,772,464 915,265,788	3,464,021 33 ⁴ 650,679 1,40	4,805,269 06,631,776	All North America Asia
10 rows × 2								0.0120017.00	,,,,		, 1010
table.tai	ttom rows 1(10) CountryOther Total	Cases NewCase	es TotalDeaths Ne	ewDeaths TotalReco	vered NewRecover	d ActiveCases	SeriousCritical	TotalTests	Tests1Mpop	Population	Contir
237 230238 231239	Tokelau	5 3,302	5,272 1,596,552		9,053	5 118,977 3,466,210		160,000,000		<u> </u>	Australia/Oce
240241242243	Total: 212,70 Total: 244,73 Total: 67,68 Total: 13,88	39,871 +8,79 6,626		+322 197,20 +75 240,41 65,81	3,530 +45,79		15,996 6,542 10,233 77				Eur South Ame Australia/Oce
244 245 246	Total: 13,000 Total: 12,77 Total: 675,12	73,223 721	258,499 15	+397 647,46	3,924 706	455,800 0	547 0 42,125				Adstralia/Oce
	re some extra spe										
# So we con # Drop to df= table	to remove the extra drop the extra p buttom unwanted drop(table.index	top and bott	tom rows that we	e do not need for							
df.drop(d	<pre>1 unwanted rows f.tail(8).index,is 7 line '\n' charace</pre>	hter									
df.replac	e(['\n'], '', reg e([','], '', rege	-	· ·								
_	to format the tab			red using a loop a	us below:						
df[co	cial characters is wanted drop unwan n df.columns[0:20 1]=df[col].str.re	<pre>n the datafra ted special o]: place('+', ''</pre>	ame can be remove characters using ').str.replace('	,','').str.repla	ce('N/A', '').s						
<pre>/var/fold ersion. I df[col] # All the</pre>	cial characters in wanted drop unwan n df.columns[0:20	n the datafrated special of the spec	ame can be remove characters using ').str.replace('gm0000gn/T/ipykeregular expressistr.replace(','	rnel_49966/346314 ons will *not* be , '').str.replace	sce('N/A', '').s 5440.py:7: Futu: treated as lite ('N/A', '').str	reWarning: The eral strings wh replace(' ', '	default valu nen regex=Tru	e of regex wi	.ll change fro	om True to	False in a
<pre>df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(</pre>	ecial characters is wanted drop unwanted drop unwanted of columns[0:20 cl]=df[col].str.regers/bt/crg5s2yn41str.regers/bt/crg5s2yn41str.replanted data is extracted data is to rename some columns={''	n the datafrated special of ted special of ted special of the place ('+', '') and the polyments of the polym	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressistr.replace(','	rnel_49966/346314 ons will *not* be , '').str.replace olumn names are in	sce('N/A', '').s 5440.py:7: Futus treated as lite ('N/A', '').str proper for data eal': 'Serious_C	reWarning: The eral strings whe replace(' ', ' processing.	default valumen regex=Tru	e of regex wi e. ce(' ', '').s 'Tot_Cases_1	.ll change fro	om True to ', ''	False in a ') 'Deaths_1M_
<pre>df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Coun 1 2 2 3</pre>	ecial characters is wanted drop unwanted dro	n the datafrated special of ted special of ted special of ted special of the special of ted spec	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressistr.replace(','	rnel_49966/346314 ons will *not* be , '').str.replace olumn names are in	Sece('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data al': 'Serious_C NewRecovered A	reWarning: The eral strings where replace ('', '') processing. citical', 'Tot ctiveCases Serio 1759965 1755 96307	default valumen regex=Trus'').str.repla Cases1Mpop':	e of regex wi e. ce(' ', '').s 'Tot_Cases_1	.ll change fro str.replace('& .M_pop', 'Dea sts_1M_pop Po 3464021 33 650679 140	om True to ', '' ths1Mpop': opulation 84805269 No	False in a ') 'Deaths_1M_ Continent 10
<pre>df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Coun 1 2</pre>	cial characters is wanted drop unwanted drop	n the datafrated special of ted special of ted special of ted special of the place ('+', '') and the place of	ame can be remove characters using ').str.replace(' gm0000gn/T/ipyke regular expressi .str.replace(',' rmat and some co ': 'Country_Name 1132719 530740	rnel_49966/346314 ons will *not* be , '').str.replace olumn names are in e', 'SeriousCritic aths TotalRecovered 44150289	Sece('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data sal': 'Serious_C NewRecovered A	reWarning: The eral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755	default valumen regex=Trumen regex=Trumen'').str.replace Cases1Mpop': ous_Critical 3564 698	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188	.ll change from tr.replace('& str.replace('& str.re	om True to ', '' ths1Mpop': opulation 84805269 No	False in a ') 'Deaths_1M_ Continent 1C OrthAmerica Asia Europe Europe
df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Coun 0 1 1 2 2 3 3 4 4 5 5 rows × 22 # However df1.dtype	cial characters is wanted drop unwan n df.columns[0:20] l]=df[col].str.reg ers/bt/crg5s2yn41g n addition, single =df[col].str.repla extracted data is to rename some content rename(columns={'') atry_Name TotalCases USA 104183562 India 44682784 France 39517456 Germany 37758773 Brazil 36809608 columns c, it is still notes	n the datafrated special of ted spec	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressing str.replace(','remat and some comparts and some com	rnel_49966/346314 ons will *not* be , '').str.replace olumn names are in e', 'SeriousCritic aths TotalRecovered 101290878 44150289 39256968 37379500 35890413	Sece('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data al': 'Serious_C NewRecovered A 24700	reWarning: The eral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755 96307 213710 222303	default valumen regex=Tru '').str.repla Cases1Mpop': ous_Critical 3564 698 869 1281 8318	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188 122332384	.ll change from tr.replace('& str.replace('& str.re	om True to ', '' ths1Mpop': opulation 34805269 No 06631776 65584518 33883596	False in a ') 'Deaths_1M_ Continent 1C orthAmerica Asia Europe Europe
<pre>df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Coun 1 2 2 3 3 4 4 5 5 rows × 22 # However</pre>	n df.columns[0:20 l]=df[col].str.reg ers/bt/crg5s2yn41g n addition, single edf[col].str.repla extracted data into rename (columns={'') etry_Name TotalCases USA 104183562 India 44682784 France 39517456 Germany 37758773 Brazil 36809608 columns c, it is still not es ame object object object object object object object	n the datafrated special of ted special of ted special of the place ('+', '') and the place ('+', '') are considered as in text for column names. CountryOther' S NewCases T 2 4 1 1 3 3 enough for of the place ('+', '')	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressing str.replace(','remat and some comparts and some com	rnel_49966/346314 ons will *not* be , '').str.replace olumn names are in e', 'SeriousCritic aths TotalRecovered 101290878 44150289 39256968 37379500 35890413	Sece('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data al': 'Serious_C NewRecovered A 24700	reWarning: The eral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755 96307 213710 222303	default valumen regex=Tru '').str.repla Cases1Mpop': ous_Critical 3564 698 869 1281 8318	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188 122332384	.ll change from tr.replace('& str.replace('& str.re	om True to ', '' ths1Mpop': opulation 34805269 No 06631776 65584518 33883596	False in a ') 'Deaths_1M_ Continent 10 orthAmerica Asia Europe Europe
df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Coun 1	n df.columns[0:20 1]=df[col].str.reg ers/bt/crg5s2yn419 n addition, single edf[col].str.repla extracted data into rename some content of the columns of the	n the datafrated special of ted special of ted special of place('+', '') and the special of ted	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressing str.replace(','remat and some comparts and some com	rnel_49966/346314 ons will *not* be , '').str.replace olumn names are in e', 'SeriousCritic aths TotalRecovered 101290878 44150289 39256968 37379500 35890413	Sece('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data al': 'Serious_C NewRecovered A 24700	reWarning: The eral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755 96307 213710 222303	default valumen regex=Tru '').str.repla Cases1Mpop': ous_Critical 3564 698 869 1281 8318	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188 122332384	.ll change from tr.replace('& str.replace('& str.re	om True to ', '' ths1Mpop': opulation 34805269 No 06631776 65584518 33883596	False in a ') 'Deaths_1M_ Continent 1C OrthAmerica Asia Europe Europe
/var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# County	cial characters is wanted drop unwan n df.columns[0:20 1]=df[col].str.region ers/bt/crg5s2yn419 n addition, single edf[col].str.repla extracted data is to rename some content of the columns of the columns of the columns of the columns USA 104183562 India 44682784 France 39517457 Germany 37758773 Brazil 36809608 columns TotalCases USA 104183562 India 44682784 France 39517457 Germany 37758773 Brazil 36809608 columns TotalCases USA 104183562 India 44682784 France 39517457 Germany 37758773 Brazil 36809608 columns TotalCases USA 104183562 India 44682784 France 39517457 Germany 37758773 Brazil 36809608 columns TotalCases USA 104183562 India 44682784 France 39517457 Germany 37758773 Brazil 36809608 columns TotalCases Object	n the datafrated special of the special of the place ('+', '') and the special of the place ('+', ''). 99sy7_d3r0f0ge character race ('+', ''). s in text for solumn names. CountryOther's NewCases To a special of the place ('+', ''). s newCases To a special of the place ('+', '').	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressing str.replace(','remat and some comparts and some com	rnel_49966/346314 ons will *not* be , '').str.replace olumn names are in e', 'SeriousCritic aths TotalRecovered 101290878 44150289 39256968 37379500 35890413	Sece('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data al': 'Serious_C NewRecovered A 24700	reWarning: The eral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755 96307 213710 222303	default valumen regex=Tru '').str.repla Cases1Mpop': ous_Critical 3564 698 869 1281 8318	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188 122332384	.ll change from tr.replace('& str.replace('& str.re	om True to ', '' ths1Mpop': opulation 34805269 No 06631776 65584518 33883596	False in a ') 'Deaths_1M_ Continent 10 orthAmerica Asia Europe Europe
/var/foldersion. Idf[col] # All the # We need dfl = df. dfl. head(# Count 1 2 2 3 3 4 4 5 5 rows × 22 # However dfl. dtype # Country_N TotalCase NewCases TotalDeat NewDeaths TotalReco NewRecove ActiveCas Serious_C Tot Cases Deaths_1M TotalTest Tests_1M_Population Continent 1Casever 1Deatheve 1Testever NewCases1 NewDeaths ActiveCas dtype: ob	cial characters is wanted drop unwanted drop unwanted in df.columns[0:20] l]=df[col].str.resters/bt/crg5s2yn41str.resters/bt/crg5s2yn41str.replanted data is described and it to rename some content of the columns of t	n the datafrated special of limited special of limi	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressistr.replace(','rmat and some compared to the compared to t	rnel_49966/346314 ons will *not* be , '').str.replace plumn names are in eaths TotalRecovered 101290878 44150289 39256968 37379500 35890413 We need to check	Sece('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data al': 'Serious_C NewRecovered A 24700	reWarning: The eral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755 96307 213710 222303	default valumen regex=Tru '').str.repla Cases1Mpop': ous_Critical 3564 698 869 1281 8318	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188 122332384	.ll change from tr.replace('& str.replace('& str.re	om True to ', '' ths1Mpop': opulation 34805269 No 06631776 65584518 33883596	False in a ') 'Deaths_1M_ Continent 10 orthAmerica Asia Europe Europe
/var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# County	decial characters is wanted drop unwanted of columns [0:20] l]=df[col].str.region addition, single edf[col].str.replated of unwanted of un	n the datafrated special of the spec	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressionstr.replace(','rmat and some control of the control of	rnel_49966/346314 ons will *not* be , '').str.replace plumn names are in the strength of the	See('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data cal': 'Serious_C NewRecovered A 24700 A the data type	reWarning: The eral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755 96307 213710 222303	default valumen regex=Tru '').str.repla Cases1Mpop': ous_Critical 3564 698 869 1281 8318	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188 122332384	.ll change from tr.replace('& str.replace('& str.re	om True to ', '' ths1Mpop': opulation 34805269 No 06631776 65584518 33883596	False in a ') 'Deaths_1M_ Continent 10 orthAmerica Asia Europe Europe
/var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Coun 0 1 1 2 2 3 3 4 4 5 5 rows × 22 # However df1.dtype # Country_N TotalCase NewCases TotalDeat NewDeaths TotalReco NewRecove ActiveCas Serious_C Tot Cases Deaths_IM TotalTest Tests_IM_ Population Continent 1Caseever 1Deathever 1Testever NewCases Deaths_IM TotalTest Tests_IM_ Population Continent 1Caseever 1Deathever 1De	n df.columns[0:20 l]=df[col].str.re ers/bt/crg5s2yn419 n addition, single edf[col].str.repla extracted data i to rename some c rename(columns={'a lindia 44682784 France 39517456 Germany 37758773 Brazil 36809608 columns c, it is still not s object ame object object object object red object ryxppl object pop object pop object ryxppl	n the datafrated special of place ('+', '') 99sy7_d3r0f0ge character race ('+', ''). s in text for column names. CountryOther' NewCases T and the data types on the data types on the data types on the data types on the dataframe to the data	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressistr.replace(','rmat and some combined in the datafrest of the stotal properties of the	ral_49966/346314 ons will *not* be , '').str.replace rolumn names are in aths TotalRecovered 101290878 44150289 39256968 37379500 35890413 We need to check	See('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data cal': 'Serious_C NewRecovered A 24700 A the data type	reWarning: The eral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755 96307 213710 222303	default valumen regex=Tru '').str.repla Cases1Mpop': ous_Critical 3564 698 869 1281 8318	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188 122332384	.ll change from tr.replace('& str.replace('& str.re	om True to ', '' ths1Mpop': opulation 34805269 No 06631776 65584518 33883596	False in a ') 'Deaths_1M_ Continent 10 orthAmerica Asia Europe Europe
/var/fold ersion. I df[col] # All the # We need df1 = df. df1. head(# County	recial characters is wanted drop unwanted dr	n the datafrated special of place('+', '') 99sy7_d3r0f0ge character race('+', ''). s in text for olumn names. CountryOther' NewCases T and the data types on the special of the speci	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressistr.replace(','rmat and some combination of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the storage of the	ral_49966/346314 ons will *not* be , '').str.replace rolumn names are in aths TotalRecovered 101290878 44150289 39256968 37379500 35890413 We need to check	See('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data cal': 'Serious_C NewRecovered A 24700 A the data type	reWarning: The eral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755 96307 213710 222303	default valumen regex=Tru '').str.repla Cases1Mpop': ous_Critical 3564 698 869 1281 8318	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188 122332384	.ll change from tr.replace('& str.replace('& str.re	om True to ', '' ths1Mpop': opulation 34805269 No 06631776 65584518 33883596	False in a ') 'Deaths_1M_ Continent 10 orthAmerica Asia Europe Europe
/var/fold ersion. I df[col] # All the # We need df1 = df. df1. head(# Coun o 1 1 2 2 3 3 4 4 5 5 rows × 22 # However df1.dtype # Country_N TotalCase NewCases TotalDeath NewDeaths TotalReco NewRecove ActiveCas Serious_C Tot Cases Deaths_1M TotalTest Tests_1M_Populatio Continent 1Case very 1Deatheve 1Test ever NewCases! NewDeaths ActiveCas dtype: ob # Type co # Continent 1Case very 1 for col i df1.repla	recial characters is wanted drop unwan n df.columns[0:20] l]=df[col].str.region ers/bt/crg5s2yn41: n addition, single edf[col].str.repla extracted data is to rename some concent columns={'interpolation of the columns of the columns of the columns of the columns r, it is still not s object ame object	n the datafrated special of lead special of place ('+', '') and special of the control of the co	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressistr.replace(','rmat and some combination of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the storage of the	ral_49966/346314 ons will *not* be , '').str.replace rolumn names are in aths TotalRecovered 101290878 44150289 39256968 37379500 35890413 We need to check	See('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data cal': 'Serious_C NewRecovered A 24700 A the data type	reWarning: The eral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755 96307 213710 222303	default valumen regex=Tru '').str.repla Cases1Mpop': ous_Critical 3564 698 869 1281 8318	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188 122332384	.ll change from tr.replace('& str.replace('& str.re	om True to ', '' ths1Mpop': opulation 34805269 No 06631776 65584518 33883596	False in a ') 'Deaths_1M_ Continent 10 orthAmerica Asia Europe Europe
/var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Coun 1 2 2 3 3 4 4 5 5 rows × 22 # However df1.dtype # Country_N TotalCase NewCases TotalDeat NewDeaths TotalReco NewRecove ActiveCas Serious_C Tot Cases Deaths_1M TotalTest Tests_1M Populatio Continent 1Caseever 1Deatheve 1Testever NewCases1 NewDeaths ActiveCas dtype: ob # The dat # So we need # Convert df1.filln df1.repla df1.repla for col i df1[c df1.dtype # Country_N TotalCase NewCases TotalDeat NewDeaths ActiveCas dtype: ob # The dat # So we need # Convert df1.filln df1.repla for col i df1[c df1.dtype # Country_N TotalCase NewCases TotalTest Tests_1M	recial characters is wanted drop unwanted and unwanted an	lumn is object place('+', '') 99sy7_d3r0f0ge character race('+', ''). s in text for column names. CountryOther' NewCases T and the dataframe to the column names. CountryOther' and the dataframe to the column names. And the dataf	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressistr.replace(','rmat and some combination of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the storage of the	ral_49966/346314 ons will *not* be , '').str.replace rolumn names are in aths TotalRecovered 101290878 44150289 39256968 37379500 35890413 We need to check	See('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data cal': 'Serious_C NewRecovered A 24700 A the data type	reWarning: The eral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755 96307 213710 222303	default valumen regex=Tru '').str.repla Cases1Mpop': ous_Critical 3564 698 869 1281 8318	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188 122332384	.ll change from tr.replace('& str.replace('& str.re	om True to ', '' ths1Mpop': opulation 34805269 No 06631776 65584518 33883596	False in a ') 'Deaths_1M_ Continent 10 orthAmerica Asia Europe Europe
df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Coun 1	recial characters is wanted drop unwanted in df.columns[0:20] l]=df[col].str.region addition, single = df[col].str.replated in addition and in add	lumn is object place('+', '') 99sy7_d3r0f0ge character race('+', ''). s in text for olumn names. CountryOther NewCases T and the dataframe to the second of the secon	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressistr.replace(','rmat and some combination of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the sto appropriate object from one on numeric of the storage of the	ral_49966/346314 ons will *not* be , '').str.replace rolumn names are in aths TotalRecovered 101290878 44150289 39256968 37379500 35890413 We need to check	See('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data cal': 'Serious_C NewRecovered A 24700 A the data type	reWarning: The eral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755 96307 213710 222303	default valumen regex=Tru '').str.repla Cases1Mpop': ous_Critical 3564 698 869 1281 8318	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188 122332384	.ll change from tr.replace('& str.replace('& str.re	om True to ', '' ths1Mpop': opulation 34805269 No 06631776 65584518 33883596	False in a ') 'Deaths_1M_ Continent 10 orthAmerica Asia Europe Europe
/var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# County	recial characters is wanted drop unwanted in df.columns[0:20] l]=df[col].str.region addition, single = df[col].str.replated in addition and in add	n the datafrated special of the spec	ame can be remove characters using ').str.replace(' gm0000gn/T/ipyke regular expressi.str.replace(',' rmat and some condition of the condition	ral_49966/346314 ons will *not* be , '').str.replace rolumn names are in aths TotalRecovered 101290878 44150289 39256968 37379500 35890413 We need to check	See('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').str proper for data cal': 'Serious_C NewRecovered A 24700 A the data type	reWarning: The eral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755 96307 213710 222303	default valumen regex=Tru '').str.repla Cases1Mpop': ous_Critical 3564 698 869 1281 8318	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188 122332384	.ll change from tr.replace('& str.replace('& str.re	om True to ', '' ths1Mpop': opulation 34805269 No 06631776 65584518 33883596	False in a ') 'Deaths_1M_ Continent 1C OrthAmerica Asia Europe Europe
df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Count 0 1 1 2 2 3 3 4 4 5 5 rows × 22 # However df1.dtype # Country_N TotalCase NewCases TotalDeat NewDeaths TotalReco NewRecove ActiveCas Serious_C Tot Cases Deaths_1M TotalTest Tests_1M_ Populatio Continent 1Caseever 1Deatheve 1Testever NewCases1. NewDeaths ActiveCas dtype: ob # Type co # Convert df1.frepla df1.repla for col i df1[c df1.dtype # Country_N TotalCase NewCases TotalDeat NewDeaths ActiveCas dtype: ob # Type co # Convert df1.frepla df1.repla for col i df1[c df1.dtype # Country_N TotalCase NewCases TotalDeat NewDeaths TotalReco NewRecove ActiveCas dtype: ob # replace worldomet worldomet worldomet	mind columns[0:20]	n the datafrated special of ted special of ted special of the spec	ame can be remove characters using ').str.replace(' gm0000gn/T/ipyke regular expressi str.replace(',' rmat and some comparison of the comp	ralloop ,', '').str.repla rnel_49966/346314 ons will *not* be , '').str.replace plumn names are in aths TotalRecovered 101290878 44150288 39256968 37379500 35890413 We need to check) We need to check)	NewRecovered A the data type	reWarning: The tral strings whereplace('', 'processing. ritical', 'Tot ctiveCases Serio 1759965 1755 96307 213710 222303 of each data from the processing of the proce	default valuer regex=True value regex=True regex=True value regex=True regex	e of regex wie. ce(' ', '').s 'Tot_Cases_1 TotalTests Te 1159772464 915265788 271490188 122332384 63776166	Tests_1M_pop Pop	Om True to ', '' ths1Mpop': opulation 34805269 No 06631776 55584518 33883596 15353593 So	Talse in a Deaths_IM_ Continent 1C OrthAmerica Asia Europe Europe OuthAmerica
df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Coun 1	recial characters is wanted drop unwan in df.columns[0:20] al]=df[col].str.region is different in addition, single edf[col].str.replate is extracted data in the rename (columns={15	n the datafrated special of ted special of ted special of ted special of the spec	ame can be remove characters using ').str.replace(' gm0000gn/T/ipyke regular expressis.str.replace(',' rmat and some condition of the conditio	ralloop ,',','').str.repla rnel_49966/346314 ons will *not* be ,''').str.replace plumn names are in the data types in the data type to and we need to check the data type to and aths TotalRecovered 101290878.0 35890413 we need to check 101290878.0	NewRecovered A the data type	ctiveCases Serio 1759965 1759965.0 ctiveCases Serio 1759965.0 ctiveCases Serio 1759965.0	default valuen regex=Truen reg	TotalTests Te 1159772464 915265788 271490188 122332384 63776166	Tests_1M_pop Pop 3464021 33 650679 140 4139547 6 1458359 8 296146 21	Population 3.34805269 No 3.34805369 No 3.34805569 No 3.34805569 No 3.3480569 No 3	Continent 10 OrthAmerica Asia Europe Europe OuthAmerica
df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Coun 0 1 1 2 2 3 3 4 4 5 5 rows × 22 # However df1.dtype # Country_N TotalCase NewCases TotalDeat NewDeaths TotalReco NewRecove ActiveCas Serious_C Tot Cases Deaths_IM_ TotalTest_Tests_IM_ Populatio Continent 1Caseever 1Deatheve 1Testever NewCases! NewDeaths ActiveCas dtype: ob # The dat # So we n # Type co #convert df1.filln df1.repla df1.repla df1.repla for col i df1[c df1.dtype # Country_N TotalCase NewCases TotalDeat NewDeaths ActiveCas dtype: ob # The dat # So we n # Type co #convert df1.filln df1.repla df1.repla for col i colicate # Country_N TotalCase NewCases TotalDeat NewCases TotalDeat NewCases TotalDeat NewCases TotalDeat NewCases TotalCase NewCases Tota	cial characters i wanted drop unwan n df.columns[0:20] l=df[col].str.re ers/bt/crg5s2yn41:n addition, single-df[col].str.reple ers/bt/crg5s2yn41:n addition, single-df[col].str.reple extracted data i to rename some contename (columns={'a' to rename some contename some contename some columns} Columns	n the datafrated special of speci	ame can be remove characters using ').str.replace('gm0000gn/T/ipyke regular expressi.str.replace(','rmat and some combination of the combination o	rame. caths TotalRecovered 37379500 35890413 We need to check we need to check 101290878 44150289 37379500 35890413 35890413 And type to and type type type type type type type type	NewRecovered A the data type	ctiveCases Serio 1759965. 1759965 1755.0 96307 213710 222303 of each data from the company of t	default valuen regex=Truen reg	TotalTests Te 1159772464 915265788 271490188 122332384 63776166	Tests_1M_pop Pop 3464021 33 650679 140 4139547 6 1458359 8 296146 21	Population 3.34805269 No 3.34805369 No 3.34805569 No 3.34805569 No 3.3480569 No 3	Continent 10 OrthAmerica Asia Europe Europe outhAmerica
df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Coun 0 1 1 2 2 3 3 4 4 5 5 rows × 22 # However df1.dtype # Country_N TotalCase NewCases TotalDeat NewDeaths TotalReco NewRecove ActiveCas Serious_C Tot Cases Deaths_1M TotalTest Tests_1M_ Populatio Continent 1Caseever 1Deatheve 1Testever NewCases! NewDeaths ActiveCas dtype: ob # The dat # So we n # Type co #convert df1.filln df1.repla df1.repla df1.repla for col i df1[c df1.dtype # Country_N TotalCase NewCases TotalDeat NewDeaths ActiveCas dtype: ob # The dat # So we n # Type co #convert df1.filln df1.repla df1.repla for col i colice # Country_N TotalCase NewCases TotalDeath NewDeaths ActiveCas dtype: ob # replace worldomet worldomet worldomet worldomet worldomet worldomet # Coun 0 1 1 2 2 3 3 4 4 5	cial characters i wanted drop unwan n df.columns[0:20 l]=df[col].str.re ers/bt/crg5s2yn41: n addition, single edf[col].str.repla cextracted data i to rename some c rename(columns={'') try_Name TotalCases USA 104183562 India 44682784 France 3951745 Germany 37758773 Brazil 36809608 columns columns columns distributed by the columns	n the datafrated special of speci	ame can be remove characters using of the control o	rame. rame. rame. rame. rame. raths TotalRecovered 101290878 44150288 39256968 37379500 35890413 We need to check we need to check 101200878 00 44150289 00 39256968 00 37379500 00 35890413 ding'): ']>min_pop] nent', rame', r	NewRecovered A the data type	ctiveCases Serio 1759965. 1759965 1755.0 96307 213710 222303 of each data from the company of t	default valuen regex=Truen reg	TotalTests Te 1159772464 915265788 271490188 122332384 63776166	Tests_1M_pop Pop 3464021 33 650679 140 4139547 6 1458359 8 296146 21	Population 3.34805269 No 3.34805369 No 3.34805569 No 3.34805569 No 3.3480569 No 3	Continent 10 OrthAmerica Asia Europe Europe OuthAmerica
df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Coun 1	reial characters is wanted drop unwanted dro	lumn is object place('+', '') 9sy7_d3r0f0g e character race('+', '') s in text for olumn names. CountryOther' NewCases T A H H H H H H H H H H H H	ame can be remove characters using ').str.replace(').str.replace(').str.replace(', ').str.replace(', '	aths TotalRecovered 101290878 44150289 39256968 37379500 35890413 We need to check We need to check 10120878.00 0.0 37379500.00	NewRecovered A the data type	ctiveCases Serio 1759965. 1759965 1755 96307 213710 222303 of each data from the company of the company o	default valuen regex=Truen reg	TotalTests Te 1159772464 915265788 271490188 122332384 63776166	Tests_1M_pop Pop 3464021 33 650679 140 4139547 6 1458359 8 296146 21	Population 3.34805269 No 3.34805369 No 3.34805569 No 3.34805569 No 3.3480569 No 3	Continent 10 OrthAmerica Asia Europe Europe OuthAmerica
df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Count 0 1 1 2 2 3 3 4 4 5 5 rows × 22 # However df1.dtype # Country_N TotalCase NewCases TotalDeat NewDeaths TotalReco NewRecove ActiveCas Serious_C Tot Cases Deaths_IM TotalTest Tests_IM Populatio Continent 1Caseever 1Deatheve 1Testever NewCases1 NewDeaths ActiveCas dtype: ob # The dat # So we n # Type co # convert df1.filn df1.repla df1.repla df1.repla df1.repla df1.repla for col i df1[c df1.dtype # Country_N TotalCase NewCases1 NewDeaths ActiveCas dtype: ob # The dat # So we n # Type co # convert df1.filn df1.repla	cial characters is wanted drop unwan on df.columns[0:20] l=df[col].str.rectors is defected in addition, single edf[col].str.repleters/bt/crg5s2yn41; naddition, single edf[col].str.repleters/bt/crg5s2yn41; naddition, single edf[col].str.repleters/bt/crg5s2yn41; naddition, single edf[col].str.repleters is defected data is to rename some of the columns experience of the columns experience in the columns experience is defected in the columns experience is defected object experience is defected object experience is defected experience is defected experience in the columns experience is defected experience is defected experience in the columns experience in the columns experience is defected experience in the columns experience is defected experience in the columns experience in the	n the datafrated special of the datafrated special of place ("+", "") place ("", "") place ("") place ("", "") place ("") p	man can be remove characters using ').str.replace(', 'gm0000gn/T/ipyke regular expression str.replace(', 'rmat and some combination of the combina	### TotalRecovered ### To	NewRecovered A the data type	ctiveCases Serio 1759965. 1759965 1755 96307 213710 222303 of each data from the company of the company o	default valuen regex=Truen reg	TotalTests Te 1159772464 915265788 271490188 122332384 63776166	Tests_1M_pop Pop 3464021 33 650679 140 4139547 6 1458359 8 296146 21	Population 3.34805269 No 3.34805369 No 3.34805569 No 3.34805569 No 3.3480569 No 3	Palse in a in
df[co /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Coun 0 1 1 2 2 3 3 4 4 5 5 rows × 22 # However df1.dtype # Country_N TotalCase NewCases TotalDeat NewDeaths TotalReco NewRecove ActiveCas Serious_C Tot Cases Deaths_IM_ TotalTests_IM_ Populatio Continent 1Caseever 1Deatheve 1Testever NewCases1 NewDeaths ActiveCas dtype: ob # The dat # So we n # Type co #convert df1.filln df1.repla df1.repla df1.repla for col i df1[c df1.dtype # Country_N TotalCase NewCases1 NewDeaths ActiveCas dtype: ob # The dat # So we n # Type co #convert df1.filln df1.repla df1.repla for col i continent 1Caseever 1Deatheve 1Tests_IM_ Populatio Continent 1Caseever NewCases1 NewDeaths ActiveCas dtype: ob # Treplace worldomet worldomet worldomet # Coun 0 1 1 2 2 3 3 4 4 5 5 rows × 22 def plot_ df =	cial characters is wanted drop unwanted and unwanted drop unwanted and unwanted drop unwanted drop unwanted drop unwanted drop unwanted drop unwanted drop drop drop drop drop drop drop dro	n the datafrated special of speci	man can be remove characters using ').str.replace(', 'gm0000gn/T/ipyke regular expression str.replace(', 'rmat and some combination of the combina	### TotalRecovered ### To	NewRecovered A the data type	ctiveCases Serio 1759965 1755 96307 213710 222303 of each data for the each da	default valuen regex=Truen reg	TotalTests Te 1159772464 915265788 271490188 122332384 63776166	Tests_1M_pop Pop 3464021 33 650679 140 4139547 6 1458359 8 296146 21	Population 3.34805269 No 3.34805369 No 3.34805569 No 3.34805569 No 3.3480569 No 3	False in a ') 'Deaths_IM_ Continent 1C Continent 1C Asia Europe Europe outhAmerica
df[col /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Count 0 1 1 2 2 3 3 4 4 5 5 rows × 22 # However df1.dtype # Country_N TotalCases NewCases TotalDeat NewDeaths TotalReco NewRecove ActiveCas Serious_C Tot Cases Deaths_1M TotalTest Tests_1M_Populatio Continent 1Caseever 1Deatheve 1Testever NewCases1 NewDeaths ActiveCas dtype: ob # The dat # So we n # Type co # convert df1.filln df1.repla df1.repla df1.repla for col i df1[c df1.dtype # Country_N TotalCase NewCases1 NewDeaths ActiveCas Serious_C Tot Cases Deaths_1M Populatio Continent 1Caseever 1Deatheve 1Testever NewCases1 NewCases TotalDeat NewCases TotalTest Tests_1M Populatio Continent 1Caseever 1Deatheve 1Testever NewCases1 NewC	cial characters is wanted drop unwanted drop unwanted and unwanted drop unwanted in the color of	n the datafrated special of speci	man can be remove characters using ').str.replace(', 'gm0000gn/T/ipyke regular expression str.replace(', 'rmat and some combination of the combina	### TotalRecovered ### To	See('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').streated a	ctiveCases Serio 1759965 1755 96307 213710 222303 of each data for the each da	default valuen regex=Truen reg	TotalTests Te 1159772464 915265788 271490188 122332384 63776166	Tests_1M_pop Pop 3464021 33 650679 140 4139547 6 1458359 8 296146 21	Population 3.34805269 No 3.34805369 No 3.34805569 No 3.34805569 No 3.3480569 No 3	Palse in a Deaths_IM_ Continent 1C OrthAmerica Asia Europe Europe OuthAmerica
df[co /var/fold ersion. I df[co] # All the # We need df1 = df. df1.head(# Coun 0 1 1 2 2 3 3 4 4 5 5 rows × 22 # However df1.dtype # Country_N TotalCase NewCases TotalDeat NewDeaths TotalReco NewRecove ActiveCas Serious_C Tot Cases Deaths_IM Populatio Continent 1Caseever 1Deatheve 1Testeyer NewDeaths ActiveCas dtype: ob # The dat # So we n # Type co #convert df1.filln df1.repla	cial characters is wanted drop unwanted drop unwanted and unwanted drop unwanted in the color of	n the datafrated special of speci	man can be remove characters using ').str.replace(', 'gm0000gn/T/ipyke regular expression str.replace(', 'rmat and some combination of the combina	### TotalRecovered ### To	See('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').streated a	ctiveCases Serio 1759965 1755 96307 213710 222303 of each data for the each da	default valuen regex=Truen reg	TotalTests Te 1159772464 915265788 271490188 122332384 63776166	Tests_1M_pop Pop 3464021 33 650679 140 4139547 6 1458359 8 296146 21	Population 3.34805269 No 3.34805369 No 3.34805569 No 3.34805569 No 3.3480569 No 3	Palse in a in
df[col /var/fold ersion. I df[col] # All the # We need df1 = df. df1.head(# Count 0 1 1 2 2 3 3 4 4 5 5 rows x 22 # However df1.dtype # Country_N TotalCase NewBeaths TotalReco NewRecove ActiveCas Serious_C Tot Cases Deaths_Int TotalTase NewDeaths TotalTest Interpolatio Continent 1Caseever 1Deatheve 1Tests_IM_ Populatio Continent 1Caseever 1Deatheve 1Testever NewCases! NewDeaths ActiveCas dtype: ob # The dat # So we n # Type co #convert df1.filln df1.repla df1.repla df1.repla df1.repla for col i df1[c df1.dtype # Country_N TotalCase NewDeaths ActiveCas Serious_C TotalDeat NewDeaths TotalTests_IM_ Fopulatio Continent 1Caseever 1Deatheve 1Tests_IM_ TotalCase NewCases! NewDeaths TotalTest NewDeaths ActiveCas Serious_C TotalDeat NewDeaths TotalTest Interplace # Country_N TotalCase NewDeaths TotalCase NewCases! NewDeaths TotalCase NewConvert df1.filln df1.repla	wanted drop unwan n df.columns[0:20 l]=df[col].str.re ers/bt/crg5s2yn41: n addition, single edf[col].str.repl: def[col].str.repl: ers/bt/crg5s2yn41: n addition, single edf[col].str.repl: def[col].str.repl: def[col].str.rep	n the datafrated special of speci	man can be remove characters using ').str.replace(', 'gm0000gn/T/ipyke regular expression str.replace(', 'rmat and some combination of the combina	### TotalRecovered ### To	See('N/A', '').s 5440.py:7: Future treated as lite ('N/A', '').streated a	ctiveCases Serio 1759965 1755 96307 213710 222303 of each data for the each da	default valuen regex=Truen reg	TotalTests Te 1159772464 915265788 271490188 122332384 63776166	Tests_1M_pop Pop 3464021 33 650679 140 4139547 6 1458359 8 296146 21	Population 3.34805269 No 3.34805369 No 3.34805569 No 3.34805569 No 3.3480569 No 3	False in a sign of the continent of the