

First step was to install openpyx1 to make it possible to read excel files!pip install openpyxl==3.1.2

 $\label{eq:continuous} Requirement already satisfied: openpyxl==3.1.2 in /root/venv/lib/python3.9/site-packages (3.1.2)$ $Requirement \ already \ satisfied: \ et-xmlfile \ in \ /root/venv/lib/python 3.9/site-packages \ (from \ open pyxl == 3.1.2) \ (1.1.0)$

[notice] A new release of pip is available: $23.0.1 \rightarrow 24.0$ [notice] To update, run: pip install --upgrade pip

Import important libraries import numpy as np ${\color{red} \text{import pandas as pd}}$ import openpyxl

df = pd.read_excel("UPDATED_BMMS_overview.xlsx")

	road object	km float64	type object	LRPName object	name object	length float64	condition object	structureNr int64
	N1 3.9%	0.004 - 522.718	Box Culvert 44%	LRP003a 2.3%				
	N2 2.7% 706 others 93.4%		RCC Girder 19.1% 12 others 36.9%	LRP001a 2.2% 1492 others 95.4%				
0	N1	1.8	Box Culvert	LRP001a		11.3	А	11786
1	N1	4.925	Box Culvert	LRP004b		6.6	Α	11786
2	N1	8.976	PC Girder Bridge	LRP008b	Kanch pur Bridge.	394.23	А	11988
3	N1	10.88	Box Culvert	LRP010b	NOYAPARA CULV	6.3	Α	11253
4	N1	10.897	Box Culvert	LRP010c	ADUPUR CULVERT	6.3	А	11253
5	N1	11.296	Box Culvert	LRP011a	NAYABARI KASP	8.3	Α	10111
6	N1	12.239	Box Culvert	LRP012a	KHAS PARA BOX	9.3	A	10111
7	N1	12.253	Box Culvert	LRP012b	DAWAN BAG BOX	6.1	Α	10111
8	N1	12.66	PC Girder Bridge	LRP013a	Madanpur Bridge	27.5	Α	11989
9	N1	12.66	PC Girder Bridge	LRP013a	MADAN PUR (R)	26.3	Α	10984
4 4								

	road object	km float64 1.8 - 460.113	type object Box Culvert 46.5% RCC Girder 25.8% 7 others 27.7%	LRP031a 0.5% LRP013a 0.4% 637 others 99.1%	name object . 0.5% SOUTH MIT 0.5% 744 others 99%	length float64 0.8 - 1408.8	condition object A	structureNr int64 100403 - 121361
0	N1	1.8	Box Culvert	LRP001a		11.3	A	11786
1	N1	4.925	Box Culvert	LRP004b		6.6	А	11786
2	N1	8.976	PC Girder Bridge	LRP008b	Kanch pur Bridge.	394.23	Α	11988
3	N1	10.88	Box Culvert	LRP010b	NOYAPARA CULV	6.3	Α	1125
4	N1	10.897	Box Culvert	LRP010c	ADUPUR CULVERT	6.3	Α	1125
5	N1	11.296	Box Culvert	LRP011a	NAYABARI KASP	8.3	Α	1011
6	N1	12.239	Box Culvert	LRP012a	KHAS PARA BOX	9.3	A	1011
7	N1	12.253	Box Culvert	LRP012b	DAWAN BAG BOX	6.1	A	1011
8	N1	12.66	PC Girder Bridge	LRP013a	Madanpur Bridge	27.5	А	1198
9	N1	12.66	PC Girder Bridge	LRP013a	MADAN PUR (R)	26.3	Α	1098

Since we only need to take the road from Dhaka to Chittagong, we decided to look in the java file and select a LRP in Chittagong. This is the LRP with LRPName LRP249. Then, we only selected the rows with LRP's above this LRP.

```
#Select road until LRP249a
df_n1 = df_n1.sort_values(by = 'LRPName')
index_of_LRP249 = df_n1.index[df_n1['LRPName'] == 'LRP249a'][0]
filtered_df = df_n1.loc[:index_of_LRP249]
```

	road object N1	km float64 1.8 - 253.777	type object Box Culvert 53.7% PC Girder B 23.6% 6 others 22.6%	LRPName object LRP031a 1.4% LRP013a 1% 237 others 97.6%	name object	length float64 1.0 - 1408.8	condition object A	structureNr int64 100545 - 121361
0	N1	1.8	Box Culvert	LRP001a		11.3	A	11786
1	N1	4.925	Box Culvert	LRP004b		6.6	Α	11786
2	N1	8.976	PC Girder Bridge	LRP008b	Kanch pur Bridge.	394.23	Α	11988
127	N1	8.976	PC Girder Bridge	LRP008b	KANCHPUR PC G	397	С	10110
145	N1	10.543	Box Culvert	LRP010a	KATCHPUR BOX	8	В	10110
3	N1	10.88	Box Culvert	LRP010b	NOYAPARA CULV	6.3	Α	11253
4	N1	10.897	Box Culvert	LRP010c	ADUPUR CULVERT	6.3	Α	11253
5	N1	11.296	Box Culvert	LRP011a	NAYABARI KASP	8.3	A	10111
145	N1	11.808	Box Culvert	LRP011c	NAYABARI BOX C	10.6	В	10111
6	N1	12.239	Box Culvert	LRP012a	KHAS PARA BOX	9.3	A	1011

We decided to create a new column called LRP_filtered, where only the first 6 digits of the LRP are showed. Based on this row, we delete duplicate bridges if they are on the same LRP. The focus is only on the first 6 digits since there are bridges with the same name on for example LRPa and LRPb. When we would focus on the whole LRP, these would not get deleted

```
#Filter on the first 6 digits
filtered_df['LRP_filtered'] = filtered_df['LRPName'].str[:6]

/tmp/ipykernel_217/2857888424.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
filtered_df['LRP_filtered'] = filtered_df['LRPName'].str[:6]
```

```
filtered_df = filtered_df.drop_duplicates(subset=['LRP_filtered'])
```

	road object	km float64 1.8 - 253.777	type object Box Culvert 67.7% PC Girder B 12.2% 4 others 20.1%	LRP001a 0.6% LRP004b 0.6% 162 others 98.8%	name object 1.8% Kanch pur Br 0.6% 160 others 97.6%	length float64 1.0 - 1408.8	C ondition object A 61% B 27.4% C 11.6%	structureNr int64 100545 - 121361
0	N1	1.8	Box Culvert	LRP001a		11.3	Α	11786
1	N1	4.925	Box Culvert	LRP004b		6.6	Α	11786
2	N1	8.976	PC Girder Bridge	LRP008b	Kanch pur Bridge.	394.23	Α	11988
145	N1	10.543	Box Culvert	LRP010a	KATCHPUR BOX	8	В	10110
5	N1	11.296	Box Culvert	LRP011a	NAYABARI KASP	8.3	A	10111
6	N1	12.239	Box Culvert	LRP012a	KHAS PARA BOX	9.3	Α	10111
9	N1	12.66	PC Girder Bridge	LRP013a	MADAN PUR (R)	26.3	Α	10984
12	N1	13.574	Box Culvert	LRP014a	KAWTALA BOX C	11.9	Α	10979
145	N1	15.465	Box Culvert	LRP016a	MALIBAG BOX C	5.55	В	10980
145	N1	17.134	PC Girder Bridge	LRP017b	LANGOLBANDO	159.5	В	10980

#A new column is created where the type is shown, this is needed for the python files.

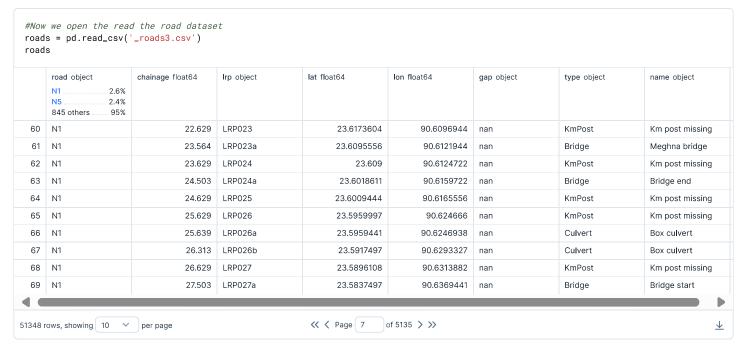
filtered_df['model_type'] = 'bridge'

/tmp/ipykernel_217/1403558431.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row_indexer,col_indexer] = value instead

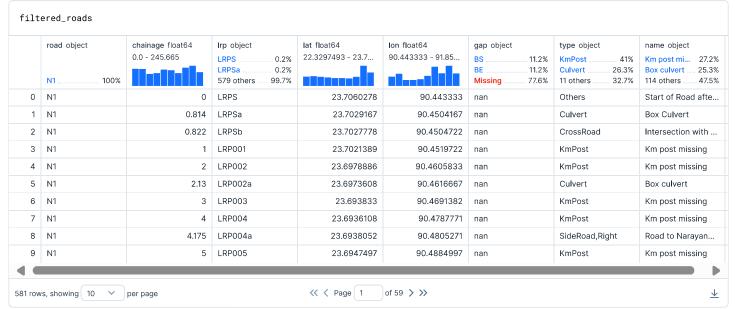
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
filtered_df['model_type'] = 'bridge'

filt	ered_df							
	road object	km float64	type object	LRPName object	name object Q baro	length float64	condition object	structureNr int64
95	N1	167.963	Box Culvert	LRP169a	BAROYARHAT BO	3	Α	101236
127	N1	205.615	RCC Girder Bridge	LRP208a	BAROCONDO BRI	22.25	С	100583
139	N1	206.866	Box Culvert	LRP209a	BARO-CUNDHO	10.2	A	100586
4 6								•
3 rows,	showing 10 v pe	er page		« < Page 1	of 1 > >>			\overline{ullet}





```
# Here we do the same as in the bridges dataset, only the LRPs above LRP249a are selected
roads_n1 = roads_n1.sort_values(by = 'lrp')
roads_n1 = pd.concat([roads_n1.iloc[-3:], roads_n1.iloc[:-3]])
index_of_LRP249_roads = roads_n1.index[roads_n1['lrp'] == 'LRP249a'][0]
filtered_roads = roads_n1.loc[:index_of_LRP249_roads]
```



A new column is added, and the type of the roads is "link"
filtered_roads['model_type'] = 'link'

/tmp/ipykernel_217/3636629436.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
filtered_roads['model_type'] = 'link'

#To calculate the length column, we calculate the difference between chainage of the previous point

 $filtered_roads['chainage'].shift(-1) - filtered_roads['chainage'] \\ filtered_roads['chainage'] \\ filtered_roads$

/tmp/ipykernel_217/4180423981.py:4: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy filtered_roads['length'] = filtered_roads['chainage'].shift(-1) - filtered_roads['chainage']

	road object	chainage float64 0.0 - 245.665	Irp object LRPS 0.2%	lat float64 22.3297493 - 23.7	Ion float64 90.443333 - 91.85	gap object BS 11.2%	type object KmPost	name object Km post mi 27.2%
	N1 100%		LRPSa 0.2% 579 others 99.7%			BE 11.2% Missing 77.6%	Culvert	Box culvert 25.39 114 others 47.59
0	N1	0	LRPS	23.7060278	90.443333	nan	Others	Start of Road afte.
1	N1	0.814	LRPSa	23.7029167	90.4504167	nan	Culvert	Box Culvert
2	N1	0.822	LRPSb	23.7027778	90.4504722	nan	CrossRoad	Intersection with
3	N1	1	LRP001	23.7021389	90.4519722	nan	KmPost	Km post missing
4	N1	2	LRP002	23.6978886	90.4605833	nan	KmPost	Km post missing
5	N1	2.13	LRP002a	23.6973608	90.4616667	nan	Culvert	Box culvert
6	N1	3	LRP003	23.693833	90.4691382	nan	KmPost	Km post missing
7	N1	4	LRP004	23.6936108	90.4787771	nan	KmPost	Km post missing
8	N1	4.175	LRP004a	23.6938052	90.4805271	nan	SideRoad,Right	Road to Narayan
9	N1	5	LRP005	23.6947497	90.4884997	nan	KmPost	Km post missing
4 =			_					

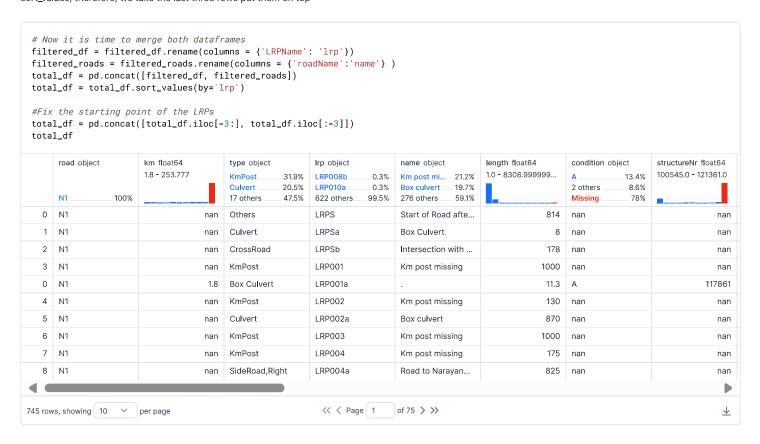
	N1 100%	km float64 1.8 - 253.777	box Culvert 67.7% Colored B 12.2% dothers 20.1%	LRP001a 0.6% LRP004b 0.6% 162 others 98.8%	name object	length float64 1.0 - 1408.8	Condition object A 61% B 27.4% C 11.6%	structureNr int64 100545 - 121361
145	N1	17.722	PC Girder Bridge	LRP018a	DARIKANDI BRID	20	В	109853
145	N1	18.742	PC Girder Bridge	LRP019a	MOLLIK PARA P.C	40.5	В	109810
16	N1	19.76	Box Culvert	LRP020a	KRIBORDI BOX C	6	Α	10981
19	N1	21.184	PC Girder Bridge	LRP021c	MARIKHALL BRID	89.2	Α	101103
145	N1	22.328	Box Culvert	LRP022a	SANGAKADI CUL	5.5	В	101112
20	N1	22.596	PC Girder Bridge	LRP023a	ASHIR CHAR (Rig	224.8	Α	109820
24	N1	24.393	PC Girder Bridge	LRP024a	Meghna Bridge	924.85	Α	11996
25	N1	26.115	Box Culvert	LRP026a	TATULTALA CUL	5.6	Α	10113:
27	N1	27.12	Box Culvert	LRP027a	MALUAKANDI CU	5.6	Α	101150
29	N1	28.315	PC Girder Bridge	LRP028a	Bhater char Bridg	173.25	Α	11997:

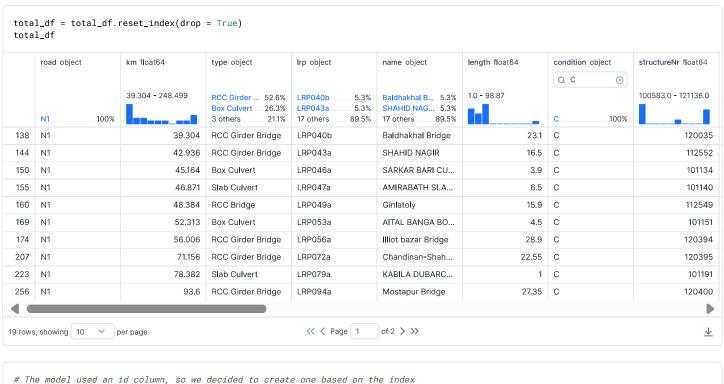
```
#Since length of the chainage was in kilometers, we needed to multiply the length by 1000
filtered_roads['length'] = filtered_roads['length']*1000

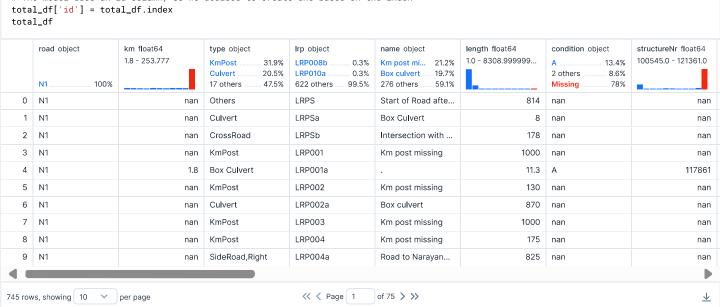
/tmp/ipykernel_217/2475612705.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy filtered_roads['length'] = filtered_roads['length']*1000
```

Now it is time to merge both dataframes. First, some column names needed to be renamed. In addition, the three LRPs for the start are now at the end due to the sort_values, therefore, we take the last three rows put them on top









#To be sure we did it correct, we decided to look at the length of the road, which is 252194meters, #Which corresponds to 252km. total_df.length.sum()

252194.270000000002