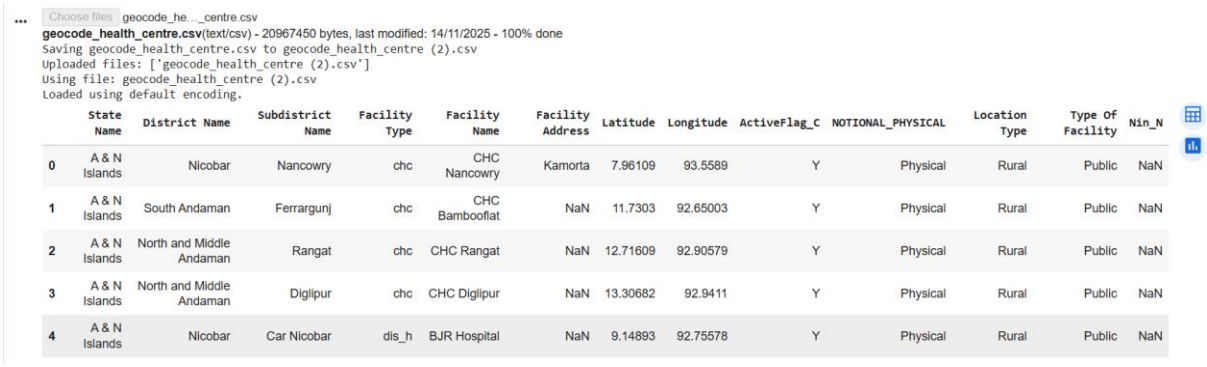
Output Screenshots

“This screenshot shows the raw government health-facility dataset loaded in Colab, displaying the first few records with key columns like State, District, Facility Type, Latitude, Longitude, and activity status.”



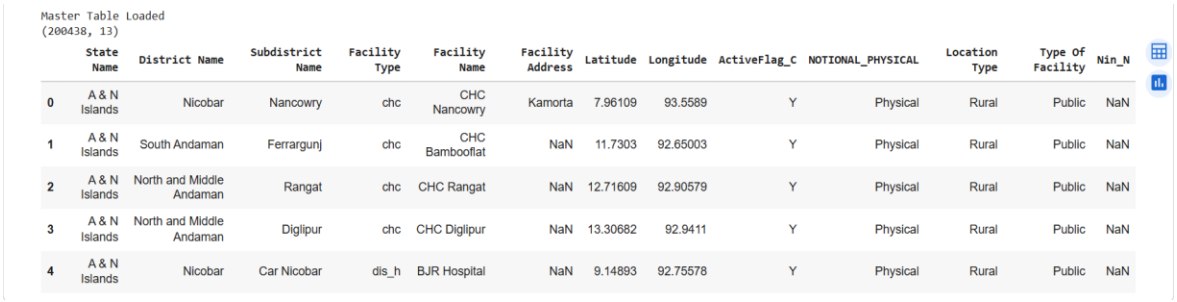
Choose files geocode_he..._centre.csv

geocode_health_centre.csv(text/csv) - 20967450 bytes, last modified: 14/11/2025 - 100% done

```
Saving geocode_health_centre.csv to geocode_health_centre (2).csv
Uploaded files: ['geocode_health_centre (2).csv']
Using file: geocode_health_centre (2).csv
Loaded using default encoding.
```

	State Name	District Name	Subdistrict Name	Facility Type	Facility Name	Facility Address	Latitude	Longitude	ActiveFlag_C	NOTIONAL_PHYSICAL	Location Type	Type Of Facility	Nin_N
0	A & N Islands	Nicobar	Nancowry	chc	CHC Nancowry	Kamorta	7.96109	93.5589	Y	Physical	Rural	Public	NaN
1	A & N Islands	South Andaman	Ferrargunj	chc	CHC Bambooflat	NaN	11.7303	92.65003	Y	Physical	Rural	Public	NaN
2	A & N Islands	North and Middle Andaman	Rangat	chc	CHC Rangat	NaN	12.71609	92.90579	Y	Physical	Rural	Public	NaN
3	A & N Islands	North and Middle Andaman	Diglipur	chc	CHC Diglipur	NaN	13.30682	92.9411	Y	Physical	Rural	Public	NaN
4	A & N Islands	Nicobar	Car Nicobar	dis_h	BJR Hospital	NaN	9.14893	92.75578	Y	Physical	Rural	Public	NaN

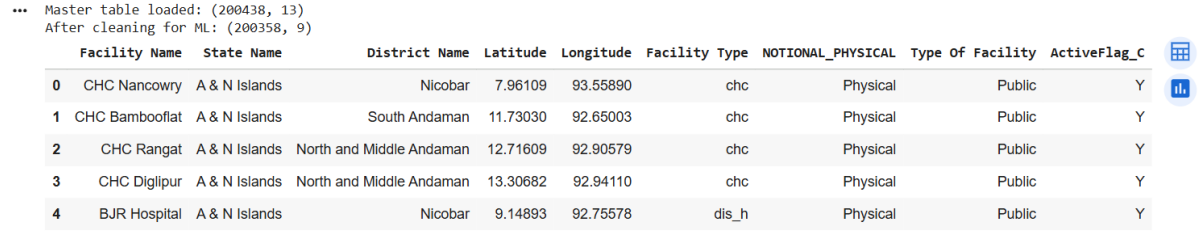
“This screenshot shows the fully merged master table (200,438 rows × 13 columns) containing all cleaned and combined health-facility information ready for analysis and clustering.”



```
Master Table Loaded
(200438, 13)
```

	State Name	District Name	Subdistrict Name	Facility Type	Facility Name	Facility Address	Latitude	Longitude	ActiveFlag_C	NOTIONAL_PHYSICAL	Location Type	Type Of Facility	Nin_N
0	A & N Islands	Nicobar	Nancowry	chc	CHC Nancowry	Kamorta	7.96109	93.5589	Y	Physical	Rural	Public	NaN
1	A & N Islands	South Andaman	Ferrargunj	chc	CHC Bambooflat	NaN	11.7303	92.65003	Y	Physical	Rural	Public	NaN
2	A & N Islands	North and Middle Andaman	Rangat	chc	CHC Rangat	NaN	12.71609	92.90579	Y	Physical	Rural	Public	NaN
3	A & N Islands	North and Middle Andaman	Diglipur	chc	CHC Diglipur	NaN	13.30682	92.9411	Y	Physical	Rural	Public	NaN
4	A & N Islands	Nicobar	Car Nicobar	dis_h	BJR Hospital	NaN	9.14893	92.75578	Y	Physical	Rural	Public	NaN

“This screenshot shows the cleaned master dataset reduced to 9 essential ML features after removing invalid and missing values, preparing the data for clustering.”



```
Master table loaded: (200438, 13)
After cleaning for ML: (200358, 9)
```

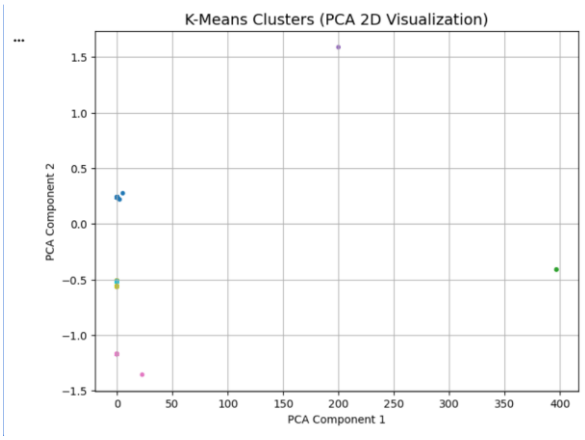
	Facility Name	State Name	District Name	Latitude	Longitude	Facility Type	NOTIONAL_PHYSICAL	Type Of Facility	ActiveFlag_C
0	CHC Nancowry	A & N Islands	Nicobar	7.96109	93.55890	chc	Physical	Public	Y
1	CHC Bambooflat	A & N Islands	South Andaman	11.73030	92.65003	chc	Physical	Public	Y
2	CHC Rangat	A & N Islands	North and Middle Andaman	12.71609	92.90579	chc	Physical	Public	Y
3	CHC Diglipur	A & N Islands	North and Middle Andaman	13.30682	92.94110	chc	Physical	Public	Y
4	BJR Hospital	A & N Islands	Nicobar	9.14893	92.75578	dis_h	Physical	Public	Y

“This screenshot shows the K-Means model assigning each health facility to a cluster, indicating the model has successfully grouped similar centres based on their features.”

K-Means model trained successfully!

	Facility Name	cluster
0	CHC Nancowry	4
1	CHC Bambooflat	4
2	CHC Rangat	4
3	CHC Diglipur	4
4	BJR Hospital	4

“This scatter plot shows the K-Means clusters projected into 2D using PCA, helping visualize how health facilities group together based on their features.”



“This table shows how many health facilities fall into each cluster, helping understand the size and distribution of the groups created by the K-Means model.”

And the second table shows the dominant facility type in each cluster, revealing what kind of health centres are most common within each group.”

Last but not least the table shows that 100% of the facilities in the dataset are active, indicating no inactive centers after cleaning.”

cluster	count
0	163062
3	29724
4	6318
5	1251
1	2
2	1

dtype: int64

cluster	Facility Type
0	sub_cen
1	sub_cen
2	sub_cen
3	phc
4	chc
5	s_t_h

dtype: object

ActiveFlag_C	proportion
Y	100.0

dtype: float64