2022-07-01:

Testing

DataTransformer\_v2

**Data**:

1. final\_extra.csv (9 GB)
2. row – 20M
3. cols - 46 (24 x float32 + 22x object)

.fit(**Data**, **Data**\_**info**)

Timing: 739.0 mins (12.3 hours)

>>> transformer = DataTransformer\_v2()

>>> start = time.time()

>>> transformer.fit(TEST\_DATA\_PATH, TEST\_DATA\_PATH[:-4]+"\_info.csv")

0.9661333560943604 mins to load PROPERTY\_TYPE

0.024533343315124512 mins to train encoder for PROPERTY\_TYPE

0.9964916626612346 mins to load BUILT\_FORM

0.029816627502441406 mins to train encoder for BUILT\_FORM

0.9742583314577739 mins to load TOTAL\_FLOOR\_AREA

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

32.69593328634898 mins to train encoder for TOTAL\_FLOOR\_AREA

0.9429332971572876 mins to load ENERGY\_TARIFF

0.024383366107940674 mins to train encoder for ENERGY\_TARIFF

0.9333832780520122 mins to load MAINS\_GAS\_FLAG

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:131: ConvergenceWarning: Number of distinct clusters (3) found smaller than n\_clusters (10). Possibly due to duplicate points in X.

cluster.KMeans(

1.8141916632652282 mins to train encoder for MAINS\_GAS\_FLAG

0.9099166472752889 mins to load FLOOR\_LEVEL

0.016316676139831544 mins to train encoder for FLOOR\_LEVEL

0.9565333485603332 mins to load FLAT\_TOP\_STOREY

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:131: ConvergenceWarning: Number of distinct clusters (3) found smaller than n\_clusters (10). Possibly due to duplicate points in X.

cluster.KMeans(

1.813000003496806 mins to train encoder for FLAT\_TOP\_STOREY

0.8598583300908407 mins to load FLAT\_STOREY\_COUNT

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

45.897783295313516 mins to train encoder for FLAT\_STOREY\_COUNT

0.9256333311398824 mins to load MAIN\_HEATING\_CONTROLS

0.028641712665557862 mins to train encoder for MAIN\_HEATING\_CONTROLS

0.8832333087921143 mins to load MULTI\_GLAZE\_PROPORTION

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

47.030224982897444 mins to train encoder for MULTI\_GLAZE\_PROPORTION

0.942175050576528 mins to load GLAZED\_TYPE

0.03167497316996257 mins to train encoder for GLAZED\_TYPE

0.980466651916504 mins to load GLAZED\_AREA

0.023416646321614585 mins to train encoder for GLAZED\_AREA

0.9443749904632568 mins to load EXTENSION\_COUNT

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:131: ConvergenceWarning: Number of distinct clusters (6) found smaller than n\_clusters (10). Possibly due to duplicate points in X.

cluster.KMeans(

1.832533355553945 mins to train encoder for EXTENSION\_COUNT

0.955608316262563 mins to load NUMBER\_HABITABLE\_ROOMS

25.36275834242503 mins to train encoder for NUMBER\_HABITABLE\_ROOMS

0.9476917107899984 mins to load NUMBER\_HEATED\_ROOMS

28.782933370272318 mins to train encoder for NUMBER\_HEATED\_ROOMS

0.8608333071072897 mins to load LOW\_ENERGY\_LIGHTING

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

38.66316666603088 mins to train encoder for LOW\_ENERGY\_LIGHTING

0.8953333775202433 mins to load NUMBER\_OPEN\_FIREPLACES

20.628633304437002 mins to train encoder for NUMBER\_OPEN\_FIREPLACES

0.9309009512265524 mins to load HOTWATER\_DESCRIPTION

0.029975032806396483 mins to train encoder for HOTWATER\_DESCRIPTION

0.9496333201726278 mins to load FLOOR\_DESCRIPTION

0.030924999713897706 mins to train encoder for FLOOR\_DESCRIPTION

0.9097166697184245 mins to load WINDOWS\_DESCRIPTION

0.026358318328857423 mins to train encoder for WINDOWS\_DESCRIPTION

0.9113083442052206 mins to load WALLS\_DESCRIPTION

0.03337499698003133 mins to train encoder for WALLS\_DESCRIPTION

0.9567416826883952 mins to load SECONDHEAT\_DESCRIPTION

0.024858335653940838 mins to train encoder for SECONDHEAT\_DESCRIPTION

0.8934083302815755 mins to load ROOF\_DESCRIPTION

0.032508357365926104 mins to train encoder for ROOF\_DESCRIPTION

0.9057250062624613 mins to load MAINHEAT\_DESCRIPTION

0.0350333293279012 mins to train encoder for MAINHEAT\_DESCRIPTION

0.9161749561627706 mins to load MAINHEATCONT\_DESCRIPTION

0.036400008201599124 mins to train encoder for MAINHEATCONT\_DESCRIPTION

0.9548750082651775 mins to load LIGHTING\_DESCRIPTION

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

41.82101662158966 mins to train encoder for LIGHTING\_DESCRIPTION

0.9337666710217793 mins to load MAIN\_FUEL

0.03888334433237712 mins to train encoder for MAIN\_FUEL

0.877049962679545 mins to load WIND\_TURBINE\_COUNT

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:131: ConvergenceWarning: Number of distinct clusters (9) found smaller than n\_clusters (10). Possibly due to duplicate points in X.

cluster.KMeans(

12.07329998811086 mins to train encoder for WIND\_TURBINE\_COUNT

0.8809749841690063 mins to load HEAT\_LOSS\_CORRIDOR

0.01548331578572591 mins to train encoder for HEAT\_LOSS\_CORRIDOR

0.8856500506401062 mins to load UNHEATED\_CORRIDOR\_LENGTH

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

34.74764163891474 mins to train encoder for UNHEATED\_CORRIDOR\_LENGTH

0.8903916557629903 mins to load FLOOR\_HEIGHT

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

34.96013333797455 mins to train encoder for FLOOR\_HEIGHT

0.8555999835332234 mins to load PHOTO\_SUPPLY

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

48.16043333212534 mins to train encoder for PHOTO\_SUPPLY

0.8821916659673055 mins to load SOLAR\_WATER\_HEATING\_FLAG

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:131: ConvergenceWarning: Number of distinct clusters (3) found smaller than n\_clusters (10). Possibly due to duplicate points in X.

cluster.KMeans(

2.3473166386286417 mins to train encoder for SOLAR\_WATER\_HEATING\_FLAG

0.9196264346440634 mins to load MECHANICAL\_VENTILATION

0.025591595967610677 mins to train encoder for MECHANICAL\_VENTILATION

0.9780166745185852 mins to load CONSTRUCTION\_AGE\_BAND

0.03258329629898071 mins to train encoder for CONSTRUCTION\_AGE\_BAND

0.8967000206311544 mins to load TENURE

0.02639164129892985 mins to train encoder for TENURE

0.876324991385142 mins to load FIXED\_LIGHTING\_OUTLETS\_COUNT

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

37.490400024255116 mins to train encoder for FIXED\_LIGHTING\_OUTLETS\_COUNT

0.9671416640281677 mins to load LOW\_ENERGY\_FIXED\_LIGHT\_COUNT

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

39.19987502495448 mins to train encoder for LOW\_ENERGY\_FIXED\_LIGHT\_COUNT

0.9173666596412658 mins to load CURRENT\_ENERGY\_RATING

0.021100016434987386 mins to train encoder for CURRENT\_ENERGY\_RATING

0.9391748944918314 mins to load POTENTIAL\_ENERGY\_RATING

0.021625030040740966 mins to train encoder for POTENTIAL\_ENERGY\_RATING

0.9100249846776326 mins to load CO2\_EMISSIONS\_CURRENT

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

34.17441954215368 mins to train encoder for CO2\_EMISSIONS\_CURRENT

0.9642000317573547 mins to load FLOOR\_ATT

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

36.81949164867401 mins to train encoder for FLOOR\_ATT

0.8743000308672587 mins to load ROOF\_ATT

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

31.773708351453145 mins to train encoder for ROOF\_ATT

0.8537499626477559 mins to load WALLS\_ATT

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

31.89322500228882 mins to train encoder for WALLS\_ATT

0.8895999948183696 mins to load latitude

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

32.74960837761561 mins to train encoder for latitude

0.9313916563987732 mins to load longitude

C:\Users\LukasCironis\anaconda3\envs\torchXtra\_git\lib\site-packages\sklearn\mixture\\_base.py:286: ConvergenceWarning: Initialization 1 did not converge. Try different init parameters, or increase max\_iter, tol or check for degenerate data.

warnings.warn(

33.80818332036336 mins to train encoder for longitude

>>> end = time.time()

>>> dif = end - start

>>> print(dif//60, "mins")

739.0 mins