```
C:\Users\Minh\anaconda3\envs\my-rdkit-env\python.exe D:\Minh\Books\
Research\Accelerated-Paper\Github\scripts\prepare_dataset.py
    rid         molecule  bond_index  ...        bdfe      bdscfe    set
0  1643  CC(=O)[C@H](C)O           2  ...   67.177976   85.596182  train
1  5327   CN1CCN(CCN)CC1           6  ...   74.432608   94.044523  train
2  5847    OCCc1ncc[nH]1           0  ...   85.799933  103.607444  train
3  5849    OCCc1ncc[nH]1           2  ...   97.257620  116.298432  train
4  7480      CCCC(=O)OCC           2  ...   79.632775   98.273105  train
5  8888  Nc1cc(N)cc(N)c1          11  ...  105.035594  120.115741  train
6  8891  Nc1cc(N)cc(N)c1          14  ...  105.035594  120.115741  train
7  8894  Nc1cc(N)cc(N)c1          17  ...  105.035594  120.115741  train
8  9204      CC(C)CC(C)C           2  ...   69.721270   91.806500  train
9  9205      CC(C)CC(C)C           3  ...   69.721270   91.806500  train

[10 rows x 9 columns]
Start sorting over 815395 rows.
        rid         molecule  bond_index  ...       bdfe      bdscfe    set
0  1102691  Br/C1=C\CCCCCC1           0  ...  70.665671   82.426866  train
1  1102692  Br/C1=C\CCCCCC1           9  ...  98.252849  113.905364  train
2  1102693  Br/C1=C\CCCCCC1          10  ...  77.981798   94.307472  train
3  1102694  Br/C1=C\CCCCCC1          11  ...  77.981798   94.307472  train
4  1102695  Br/C1=C\CCCCCC1          12  ...  86.847246  103.876175  train
5  1102696  Br/C1=C\CCCCCC1          13  ...  86.847246  103.876175  train
6  1102697  Br/C1=C\CCCCCC1          14  ...  86.099882  102.784938  train
7  1102698  Br/C1=C\CCCCCC1          15  ...  86.099882  102.784938  train
8  1102699  Br/C1=C\CCCCCC1          16  ...  85.890294  102.463496  train
9  1102700  Br/C1=C\CCCCCC1          17  ...  85.890294  102.463496  train

[10 rows x 9 columns]
Start indexing over 815395 rows.
Index: [(0, 'Br/C1=C\\CCCCCC1'),
 (14, 'Br/C=C/C1CCCCC1'),
 (29, 'Br/C=C/Cc1ccccc1'),
 (41, 'Br/C=C/[C@@H]1CC=CCC1'),
 (54, 'Br/C=C/[C@H]1CC=CCC1'),
 (67, 'Br/C=C/c1ccccn1'),
 (75, 'Br/C=C/c1cccnc1'),
 (82, 'Br/C=C\\c1ccccc1'),
 (91, 'Br/C=C\\c1ccccn1'),
 (99, 'BrC(Br)(Br)Br')]
Start filtering over 815395 rows with 63229 partitions.
Completed 3000 partitions with progress 4.74%.
Completed 6000 partitions with progress 9.49%.
Completed 9000 partitions with progress 14.23%.
Completed 12000 partitions with progress 18.98%.
Completed 15000 partitions with progress 23.72%.
Completed 18000 partitions with progress 28.47%.
Completed 21000 partitions with progress 33.21%.
Completed 24000 partitions with progress 37.96%.
Completed 27000 partitions with progress 42.70%.
Completed 30000 partitions with progress 47.45%.
```

```
Completed 33000 partitions with progress 52.19%.
Completed 36000 partitions with progress 56.94%.
Completed 39000 partitions with progress 61.68%.
Completed 42000 partitions with progress 66.43%.
Completed 45000 partitions with progress 71.17%.
Completed 48000 partitions with progress 75.91%.
Completed 51000 partitions with progress 80.66%.
Completed 54000 partitions with progress 85.40%.
Completed 57000 partitions with progress 90.15%.
Completed 60000 partitions with progress 94.89%.
Completed 63000 partitions with progress 99.64%.
Rows: 519758 rows --> Non-rows: 295637 rows.
Is filter works correct: True
Start exporting from ../BDE_data/20211201_bonds_for_neighbors.csv to ../BDE_data/source_dataset_v2_test.csv.
       rid         molecule  bond_index  ...       bdfe      bdscfe    set
0  1102691  Br/C1=C\CCCCCC1           0  ...  70.665671   82.426866  train
1  1102692  Br/C1=C\CCCCCC1           9  ...  98.252849  113.905364  train
2  1102693  Br/C1=C\CCCCCC1          10  ...  77.981798   94.307472  train
3  1102695  Br/C1=C\CCCCCC1          12  ...  86.847246  103.876175  train
4  1102697  Br/C1=C\CCCCCC1          14  ...  86.099882  102.784938  train
5  1102699  Br/C1=C\CCCCCC1          16  ...  85.890294  102.463496  train
6  1102701  Br/C1=C\CCCCCC1          18  ...  86.563612  103.491393  train
7  1102703  Br/C1=C\CCCCCC1          20  ...  80.646829   97.545021  train
8  1033498  Br/C=C/C1CCCCC1           0  ...  71.736829   83.205387  train
9  1033499  Br/C=C/C1CCCCC1           2  ...  84.133269  103.482457  train

[10 rows x 9 columns]

Process finished with exit code 0
```