Setup

```
!pip install chemprop
!pip install rdkit-pypi  # should be included in above after Chemprop v1.6 release

import chemprop
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib.offsetbox import AnchoredText
from sklearn.metrics import mean_absolute_error, mean_squared_error
from sklearn.decomposition import PCA
```

```
Requirement already satisfied: py4j in /usr/local/lib/python3.10/dist-packages (from hyperopt>=0.2.3->chemprop) (
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->c
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.0.3->chemp
Requirement already satisfied: xarray in /usr/local/lib/python3.10/dist-packages (from pandas-flavor>=0.2.0->chem
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from scikit-learn>=0.22.
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn
Requirement already satisfied: sphinxcontrib-applehelp in /usr/local/lib/python3.10/dist-packages (from sphinx>=3
Requirement already satisfied: sphinxcontrib-devhelp in /usr/local/lib/python3.10/dist-packages (from sphinx>=3.1
Requirement already satisfied: sphinxcontrib-jsmath in /usr/local/lib/python3.10/dist-packages (from sphinx>=3.1.
Requirement already satisfied: sphinxcontrib-htmlhelp>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from sph
Requirement already satisfied: sphinxcontrib-serializinghtml>=1.1.5 in /usr/local/lib/python3.10/dist-packages (f
Requirement already satisfied: sphinxcontrib-qthelp in /usr/local/lib/python3.10/dist-packages (from sphinx>=3.1.
Requirement already satisfied: Pygments>=2.0 in /usr/local/lib/python3.10/dist-packages (from sphinx>=3.1.2->chem
Requirement already satisfied: docutils<0.19,>=0.14 in /usr/local/lib/python3.10/dist-packages (from sphinx>=3.1.
Requirement already satisfied: snowballstemmer>=1.1 in /usr/local/lib/python3.10/dist-packages (from sphinx>=3.1.
Requirement already satisfied: babel>=1.3 in /usr/local/lib/python3.10/dist-packages (from sphinx>=3.1.2->chempro
Requirement already satisfied: alabaster<0.8,>=0.7 in /usr/local/lib/python3.10/dist-packages (from sphinx>=3.1.2
Requirement already satisfied: imagesize in /usr/local/lib/python3.10/dist-packages (from sphinx>=3.1.2->chemprop
Requirement already satisfied: requests>=2.5.0 in /usr/local/lib/python3.10/dist-packages (from sphinx>=3.1.2->ch
Requirement already satisfied: protobuf>=3.20 in /usr/local/lib/python3.10/dist-packages (from tensorboardX>=2.0-
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from torch>=1.4.0->chemprop)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages (from torch>=1.4.0->c
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from torch>=1.4.0->chemprop) (1.
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages (from torch>=1.4.0->chemprop) (2
Requirement already satisfied: triton==2.1.0 in /usr/local/lib/python3.10/dist-packages (from torch>=1.4.0->chemp
Collecting typing-inspect>=0.7.1 (from typed-argument-parser>=1.6.1->chemprop)
  Downloading typing_inspect-0.9.0-py3-none-any.whl (8.8 kB)
Collecting docstring-parser>=0.15 (from typed-argument-parser>=1.6.1->chemprop)
  Downloading docstring_parser-0.15-py3-none-any.whl (36 kB)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2>=3.0->flas
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.5.0->sph
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests>=2.5.
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.5.
Collecting mypy-extensions>=0.3.0 (from typing-inspect>=0.7.1->typed-argument-parser>=1.6.1->chemprop)
  Downloading mypy_extensions-1.0.0-py3-none-any.whl (4.7 kB)
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-packages (from sympy->torch>=1.4.0-
Building wheels for collected packages: typed-argument-parser
  Building wheel for typed-argument-parser (setup.py) ... done
  Created wheel for typed-argument-parser: filename=typed_argument_parser-1.9.0-py3-none-any.whl size=25615 sha25
  Stored in directory: /root/.cache/pip/wheels/f0/94/0f/9539f578bed7e1bd423c702e403712f5ee8989f831a71db000
Successfully built typed-argument-parser
Installing collected packages: tensorboardX, rdkit, mypy-extensions, docstring-parser, typing-inspect, typed-argu
Successfully installed chemprop-1.6.1 docstring-parser-0.15 mypy-extensions-1.0.0 pandas-flavor-0.6.0 rdkit-2023.
Collecting rdkit-pypi
  Downloading rdkit_pypi-2022.9.5-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (29.4 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 29.4/29.4 MB 14.6 MB/s eta 0:00:00
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from rdkit-pypi) (1.25.2)
Requirement already satisfied: Pillow in /usr/local/lib/python3.10/dist-packages (from rdkit-pypi) (9.4.0)
Installing collected packages: rdkit-pypi
Successfully installed rdkit-pypi-2022.9.5
```

```
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"


hiv_df = pd.read_csv("HIV.csv")
hiv_df.head()
```

	smiles	activity	HIV_active
0	CCC1=[O+][Cu-3]2([O+]=C(CC)C1)[O+]=C(CC)CC(CC)...	CI	0
1	C(=Cc1ccccc1)C1=[O+][Cu-3]2([O+]=C(C=Cc3ccccc3...	CI	0
2	CC(=O)N1c2ccccc2Sc2c1ccc1ccccc21	CI	0
3	Nc1ccc(C=Cc2ccc(N)cc2S(=O)(=O)O)c(S(=O)(=O)O)c1	CI	0
4	O=S(=O)(O)CCS(=O)(=O)O	CI	0

Next steps: View recommended plots

```
hiv_df.describe()
```

	HIV_active
count	41127.000000
mean	0.035086
std	0.184001
min	0.000000
25%	0.000000
50%	0.000000
75%	0.000000
max	1.000000

```
unique_values = hiv_df['HIV_active'].unique()
print(f"Unique values in 'HIV_active': {unique_values}")
```

```
Unique values in 'HIV_active': [0 1]
```

```
unique_values = hiv_df['smiles'].unique()
print(f"Unique values in 'smiles': {unique_values}")
print(f"length of uniqe value: {len(unique_values)}")
```

```
Unique values in 'smiles': ['CCC1=[O+][Cu-3]2([O+]=C(CC)C1)[O+]=C(CC)CC(CC)=[O+]2'
 'C(=Cc1ccccc1)C1=[O+][Cu-3]2([O+]=C(C=Cc3ccccc3)CC(c3ccccc3)=[O+]2)[O+]=C(c2ccccc2)C1'
 'CC(=O)N1c2ccccc2Sc2c1ccc1ccccc21' ...
 'Cc1ccc(N2C(=O)C3c4[nH]c5ccccc5c4C4CCC(C(C)(C)C)CC4C3C2=O)cc1'
 'Cc1cccc(N2C(=O)C3c4[nH]c5ccccc5c4C4CCC(C(C)(C)C)CC4C3C2=O)c1'
 'CCCCCC=C(c1cc(Cl)c(OC)c(-c2nc(C)no2)c1)c1cc(Cl)c(OC)c(-c2nc(C)no2)c1']
length of uniqe value: 41127
```

```
# Filter rows where 'your_column' is not equal to 1 or 0
filtered_df = hiv_df[(hiv_df['HIV_active'] != 1) & (hiv_df['HIV_active'] != 0)]
filtered_df
```

	smiles	activity	HIV_active

```
# Filter rows where 'target_column' is equal to 1h
hiv_df_filtered_active = hiv_df[hiv_df['HIV_active'] == 1]
hiv_df_filtered_active
```

	smiles	activity	HIV_active
11	O=C(O)Cc1ccc(SSc2ccc(CC(=O)O)cc2)cc1	CM	1
16	NNP(=S)(NN)c1ccccc1	CM	1
80	O=Nc1ccc(O)c(N=O)c1O	CM	1
203	Oc1ccc(Cl)cc1C(c1cc(Cl)ccc1O)C(Cl)(Cl)Cl	CM	1
234	NNC(=O)c1ccccc1SSc1ccccc1C(=O)NN	CM	1
...	...	...	...
41090	Cc1cn(COCCCOCC(=O)c2ccccc2)c(=O)[nH]c1=O	CM	1
41092	Cc1cn(C2CC3C(COC(CCC[Se]c4ccccc4)N3O)O2)c(=O)[...	CM	1
41093	Cc1cn(C2CC3C(COC(CCCC[Se]c4ccccc4)N3O)O2)c(=O)...	CM	1
41098	Cc1cn(C2CC3C(COC(CC[Se]C#N)N3O)O2)c(=O)[nH]c1=O	CM	1
41099	C[Se]CCC1OCC2OC(n3cc(C)c(=O)[nH]c3=O)CC2N1O	CA	1

1443 rows × 3 columns

Next steps: View recommended plots

```
# Filter rows where 'target_column' is equal to 1h
hiv_df_filtered_inactive = hiv_df[hiv_df['HIV_active'] == 0]
hiv_df_filtered_inactive = hiv_df_filtered_inactive.sample(n=1500, axis=0, replace=True)
hiv_df_filtered_inactive
```

	smiles	activity	HIV_active
3661	CN1CCc2c(sc(N)c2C#N)C1	CI	0
35554	CCOC(=O)C(NC(=O)c1cccc(Cl)c1)(Nc1ccc(S(=O)(=O)...	CI	0
10895	CCOC(=O)c1cncc(C#N)c1	CI	0
21007	CC(C)(C)OC(=O)NC(Cc1ccc(O)cc1)C(=O)N1CCCC1C(=O)O	CI	0
10519	Cc1ncc(CO)c2c1[OH+][Fe-3]13([OH+]S(=O)(=O)[OH+...	CI	0
...	...	...	...
39986	COC(=O)C=CC(C#N)(C#N)N=Cc1cccs1	CI	0
27018	COC(=O)c1ccccc1C#CC#Cc1ccccc1C(=O)OC	CI	0
4946	Nc1nc(N)c(-c2ccc(Cl)c(Cl)c2)c(C[PH](c2ccccc2)(...	CI	0
35085	CN(C)N=Nc1nc2c(c(=O)n(C)c(=O)n2C)n1CC1CO1	CI	0
7723	Cc1ccc2c(c1)SSc1ccc(C)cc1SS2	CI	0

1500 rows × 3 columns

Next steps: View recommended plots

```
hiv_df_sampled = pd.concat([hiv_df_filtered_active, hiv_df_filtered_inactive], axis=0, ignore_index=True)
hiv_df_sampled
```

	smiles	activity	HIV_active
0	O=C(O)Cc1ccc(SSc2ccc(CC(=O)O)cc2)cc1	CM	1
1	NNP(=S)(NN)c1ccccc1	CM	1
2	O=Nc1ccc(O)c(N=O)c1O	CM	1
3	Oc1ccc(Cl)cc1C(c1cc(Cl)ccc1O)C(Cl)(Cl)Cl	CM	1
4	NNC(=O)c1ccccc1SSc1ccccc1C(=O)NN	CM	1
...	...	...	...
2938	COC(=O)C=CC(C#N)(C#N)N=Cc1cccs1	CI	0
2939	COC(=O)c1ccccc1C#CC#Cc1ccccc1C(=O)OC	CI	0
2940	Nc1nc(N)c(-c2ccc(Cl)c(Cl)c2)c(C[PH](c2ccccc2)(...	CI	0
2941	CN(C)N=Nc1nc2c(c(=O)n(C)c(=O)n2C)n1CC1CO1	CI	0
2942	Cc1ccc2c(c1)SSc1ccc(C)cc1SS2	CI	0

2943 rows × 3 columns

Next steps: View recommended plots

```
# Randomly shuffle rows
hiv_df_sampled = hiv_df_sampled.sample(frac=1, random_state=42)

hiv_df_sampled.head()
```

1 to 5 of 5 entries Filter

index	smiles	activity	HIV_active
840	C[N+](C)(C)C.O=C(Nc1ccc(C=Cc2ccc(NC(=O)c3cc(S(=O)(O)=[OH+])ccc3O)cc2S(=O)(O)=[OH+])c(S(=O)(O)=[OH+])c1)c1cc(S(=O)(O)=[OH+])ccc1O	CM	1
1037	COC(=O)Cc1cc(O)c(CC=C(C)CCC=C(C)C(O)C(=O)C=C(C)C)cc1O	CM	1
2399	CC(C)OC1OC(CO)C(O)CC1n1cc(F)c(=O)[nH]c1=O	CI	0
678	Cc1cc(C)c(S(=O)(O)=[OH+])c(C)c1.N[S+]1Cc2nc3ccccc3n2C1c1c(F)cccc1F	CM	1
196	CCOc1cc(C2=NN3C(=S)NNC3=NN2)ccc1O	CM	1

Show 25 per page

Like what you see? Visit the data table notebook to learn more about interactive tables.

Next steps: View recommended plots

```
hiv_df_sampled.to_csv('HIV_2.csv', index=False)
# .drop(['activity'], axis=1).
hiv_df_sampled_2 = pd.read_csv("HIV_2.csv")
hiv_df_sampled_2.head()
hiv_df_sampled_2.tail()
```

	smiles	activity	HIV_active
0	C[N+](C)(C)C.O=C(Nc1ccc(C=Cc2ccc(NC(=O)c3cc(S(...	CM	1
1	COC(=O)Cc1cc(O)c(CC=C(C)CCC=C(C)C(O)C(=O)C=C(C...	CM	1
2	CC(C)OC1OC(CO)C(O)CC1n1cc(F)c(=O)[nH]c1=O	CI	0
3	Cc1cc(C)c(S(=O)(O)=[OH+])c(C)c1.N[S+]1Cc2nc3cc...	CM	1
4	CCOc1cc(C2=NN3C(=S)NNC3=NN2)ccc1O	CM	1

	smiles	activity	HIV_active
2938	N=c1ccn(C2CC(F)C(CO)S2)c(=O)[nH]1	CI	0
2939	COP(=O)(C=Cc1cc(C(=O)OCc2ccccc2)n(S(=O)(=O)c2c...	CM	1
2940	O=C(CCc1ccc(O)cc1)NCCc1ccc(O)cc1	CM	1
2941	CCc1c(Cc2cc(C)cc(C)c2)n(COCCCO)c(=O)[nH]c1=O	CA	1
2942	CCSCCCCCCCCCCC(=O)OCC1OC(n2cc(C)c(=O)[nH]c2=O)...	CA	1

```
arguments = [
    '--data_path', 'HIV_2.csv',
    '--dataset_type', 'classification',
    '--save_dir', 'test_checkpoints_multimolecule',
    '--epochs', '30',
    '--save_smiles_splits',
    '--quiet',
    '--batch_size', '64',
    '--ignore_columns', 'activity',
    '--depth', '5',
    '--hidden_size', '300'
]

args = chemprop.args.TrainArgs().parse_args(arguments)
```

```
mean_score, std_score = chemprop.train.cross_validate(args=args, train_func=chemprop.train.run_training)
```

```
 57%|██████    | 21/37 [00:02<00:01,  9.90it/s]
 59%|██████    | 22/37 [00:02<00:01,  9.61it/s]
 62%|██████    | 23/37 [00:02<00:01,  9.68it/s]
 65%|██████    | 24/37 [00:02<00:01,  9.29it/s]
 70%|███████   | 26/37 [00:02<00:01,  9.62it/s]
 73%|███████   | 27/37 [00:02<00:01,  9.33it/s]
 78%|███████   | 29/37 [00:02<00:00,  9.72it/s]
 81%|████████  | 30/37 [00:03<00:00,  9.58it/s]
 84%|████████  | 31/37 [00:03<00:00,  9.31it/s]
 86%|████████  | 32/37 [00:03<00:00,  9.11it/s]
 89%|████████  | 33/37 [00:03<00:00,  8.95it/s]
 92%|████████  | 34/37 [00:03<00:00,  8.89it/s]
 95%|████████  | 35/37 [00:03<00:00,  8.87it/s]
 97%|████████  | 36/37 [00:03<00:00,  8.36it/s]

  0%|          | 0/5 [00:00<?, ?it/s]
 20%|██        | 1/5 [00:00<00:00,  8.54it/s]
 40%|████      | 2/5 [00:00<00:00,  8.50it/s]
 60%|██████    | 3/5 [00:00<00:00,  8.94it/s]
100%|██████████| 5/5 [00:00<00:00, 11.25it/s]
 53%|█████     | 16/30 [01:22<01:06,  4.75s/it]
  0%|          | 0/37 [00:00<?, ?it/s]
  3%|▎         | 1/37 [00:00<00:04,  8.87it/s]
  5%|▌         | 2/37 [00:00<00:04,  7.88it/s]
  8%|▊         | 3/37 [00:00<00:04,  7.89it/s]
 11%|█         | 4/37 [00:00<00:04,  7.71it/s]
 14%|█▎        | 5/37 [00:00<00:04,  7.90it/s]
 16%|█▌        | 6/37 [00:00<00:03,  7.84it/s]
 19%|█▉        | 7/37 [00:00<00:03,  8.18it/s]
 22%|██▏       | 8/37 [00:00<00:03,  8.41it/s]
 24%|██▍       | 9/37 [00:01<00:03,  8.37it/s]
 30%|███       | 11/37 [00:01<00:02,  9.24it/s]
 32%|███▏      | 12/37 [00:01<00:02,  9.01it/s]
 35%|███▌      | 13/37 [00:01<00:02,  9.02it/s]
 38%|███▊      | 14/37 [00:01<00:02,  9.09it/s]
 41%|████      | 15/37 [00:01<00:02,  9.02it/s]
 43%|████▎     | 16/37 [00:01<00:02,  9.18it/s]
 49%|████▊     | 18/37 [00:02<00:01,  9.57it/s]
 51%|█████     | 19/37 [00:02<00:01,  9.44it/s]
 54%|█████▍    | 20/37 [00:02<00:01,  9.51it/s]
 59%|█████▉    | 22/37 [00:02<00:01,  9.67it/s]
 62%|██████▏   | 23/37 [00:02<00:01,  9.48it/s]
 65%|██████▍   | 24/37 [00:02<00:01,  9.47it/s]
 68%|██████▊   | 25/37 [00:02<00:01,  9.38it/s]
 70%|███████   | 26/37 [00:02<00:01,  9.43it/s]
 76%|███████▌  | 28/37 [00:03<00:01,  8.81it/s]
 78%|███████▊  | 29/37 [00:03<00:01,  7.74it/s]
 81%|████████  | 30/37 [00:03<00:01,  6.90it/s]
 84%|████████  | 31/37 [00:03<00:00,  6.36it/s]
 86%|████████  | 32/37 [00:03<00:00,  5.83it/s]
 89%|████████  | 33/37 [00:04<00:00,  5.66it/s]
 92%|████████  | 34/37 [00:04<00:00,  5.35it/s]
 95%|████████  | 35/37 [00:04<00:00,  5.24it/s]
 97%|████████  | 36/37 [00:04<00:00,  5.12it/s]
100%|██████████| 37/37 [00:04<00:00,  5.52it/s]

  0%|          | 0/5 [00:00<?, ?it/s]
 20%|██        | 1/5 [00:00<00:00,  4.78it/s]
```

```
mean_score, std_score
```

```
(0.8266132744177483, 0.0)
```

```
bp_df = pd.read_csv("BBBP.csv")
bp_df.head()
```

	num	name	p_np	smiles
0	1	Propanolol	1	[Cl].CC(C)NCC(O)COc1cccc2ccccc12
1	2	Terbutylchlorambucil	1	C(=O)(OC(C)(C)C)CCCc1ccc(cc1)N(CCCl)CCCl
2	3	40730	1	c12c3c(N4CCN(C)CC4)c(F)cc1c(c(C(O)=O)cn2C(C)CO...
3	4	24	1	C1CCN(CC1)Cc1cccc(c1)OCCCNC(=O)C
4	5	cloxacillin	1	Cc1onc(c2ccccc2Cl)c1C(=O)N[C@H]3[C@H]4SC(C)(C)...

Next steps: View recommended plots

```
bp_df.tail()
```

	num	name	p_np	smiles
2045	2049	licostinel	1	C1=C(Cl)C(=C(C2=C1NC(=O)C(N2)=O)[N+](=O)[O-])Cl
2046	2050	ademetionine(adenosyl-methionine)	1	[C@H]3([N]2C1=C(C(=NC=N1)N)N=C2)[C@@H]([C@@H](...
2047	2051	mesocarb	1	[O+]1=N[N](C=C1[N-]C(NC2=CC=CC=C2)=O)C(CC3=CC=...
				C1=C(OC)C(=CC2=C1C(=[N+](C(=C2CC)C)

```
bp_df.drop(['num', 'name', 'p_np'], axis=1).to_csv('BBBP_2.csv', index=False)

bp_df_2 = pd.read_csv("BBBP_2.csv")
bp_df_2.head()
bp_df_2.tail()
```

	smiles
0	[Cl].CC(C)NCC(O)COc1cccc2ccccc12
1	C(=O)(OC(C)(C)C)CCCc1ccc(cc1)N(CCCl)CCCl
2	c12c3c(N4CCN(C)CC4)c(F)cc1c(c(C(O)=O)cn2C(C)CO...
3	C1CCN(CC1)Cc1cccc(c1)OCCCNC(=O)C
4	Cc1onc(c2ccccc2Cl)c1C(=O)N[C@H]3[C@H]4SC(C)(C)...

	smiles
2045	C1=C(Cl)C(=C(C2=C1NC(=O)C(N2)=O)[N+](=O)[O-])Cl
2046	[C@H]3([N]2C1=C(C(=NC=N1)N)N=C2)[C@@H]([C@@H](...
2047	[O+]1=N[N](C=C1[N-]C(NC2=CC=CC=C2)=O)C(CC3=CC=...
2048	C1=C(OC)C(=CC2=C1C(=[N+](C(=C2CC)C)[NH-])C3=CC...
2049	[N+](=NCC(=O)N[C@@H]([C@H](O)C1=CC=C([N+]([O-]...

```
arguments = [
    '--test_path', 'BBBP_2.csv',
    '--preds_path', 'BBBP_preds.csv',
    '--checkpoint_dir', 'test_checkpoints_multimolecule'
]

args = chemprop.args.PredictArgs().parse_args(arguments)
preds = chemprop.train.make_predictions(args=args)
```

```
[16:40:27] Explicit valence for atom # 1 N, 4, is greater than permitted
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] Explicit valence for atom # 6 N, 4, is greater than permitted
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
```

```
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] Explicit valence for atom # 11 N, 4, is greater than permitted
[16:40:27] Explicit valence for atom # 12 N, 4, is greater than permitted
[16:40:27] Explicit valence for atom # 5 N, 4, is greater than permitted
[16:40:27] Explicit valence for atom # 5 N, 4, is greater than permitted
[16:40:27] Explicit valence for atom # 5 N, 4, is greater than permitted
[16:40:27] Explicit valence for atom # 5 N, 4, is greater than permitted
[16:40:27] Explicit valence for atom # 5 N, 4, is greater than permitted
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] Explicit valence for atom # 5 N, 4, is greater than permitted
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
[16:40:27] WARNING: not removing hydrogen atom without neighbors
/usr/local/lib/python3.10/dist-packages/torch/utils/data/dataloader.py:557: UserWarning: This DataLoader will cre
  warnings.warn(_create_warning_msg(
Test size = 2,039
  0%|          | 0/1 [00:00<?, ?it/s]Loading pretrained parameter "encoder.encoder.0.cached_zero_vector".
Loading pretrained parameter "encoder.encoder.0.W_i.weight".
Loading pretrained parameter "encoder.encoder.0.W_h.weight".
Loading pretrained parameter "encoder.encoder.0.W_o.weight".
```

```
bp_preds_df = pd.read_csv("BBBP_preds.csv")
bp_preds_df.head()
```

	smiles	HIV_active
0	[Cl].CC(C)NCC(O)COc1cccc2ccccc12	0.20634931325912476
1	C(=O)(OC(C)(C)C)CCCc1ccc(cc1)N(CCCl)CCCl	0.14107653498649597
2	c12c3c(N4CCN(C)CC4)c(F)cc1c(c(C(O)=O)cn2C(C)CO...	0.4731158912181854
3	C1CCN(CC1)Cc1cccc(c1)OCCCNC(=O)C	0.16012685000896454
4	Cc1onc(c2ccccc2Cl)c1C(=O)N[C@H]3[C@H]4SC(C)(C)...	0.3700324594974518

Next steps: View recommended plots

```
bp_preds_df.tail()
```

	smiles	HIV_active
2045	C1=C(Cl)C(=C(C2=C1NC(=O)C(N2)=O)[N+](=O)[O-])Cl	0.17658869922161102
2046	[C@H]3([N]2C1=C(C(=NC=N1)N)N=C2)[C@@H]([C@@H](...	0.30184003710746765
2047	[O+]1=N[N](C=C1[N-]C(NC2=CC=CC=C2)=O)C(CC3=CC=...	0.432624489068985
2048	C1=C(OC)C(=CC2=C1C(=[N+](C(=C2CC)C)[NH-])C3=CC...	0.3804272413253784
2049	[N+](=NCC(=O)N[C@@H]([C@H](O)C1=CC=C([N+]([O-]...	0.3450404107570648

```
bp_preds_df.describe()
```

	smiles	HIV_active
count	2050	2050
unique	2050	1994
top	[Cl].CC(C)NCC(O)COc1cccc2ccccc12	Invalid SMILES
freq	1	11

```
bp_preds_df = bp_preds_df[bp_preds_df['HIV_active'] != "Invalid SMILES"]
bp_preds_df.describe()
```

	smiles	HIV_active
count	2039	2039
unique	2039	1993
top	[Cl].CC(C)NCC(O)COc1cccc2ccccc12	0.1317506581544876
freq	1	3

```
bp_preds_df['HIV_active'] = bp_preds_df['HIV_active'].astype(float)
```

```
bp_preds_df['HIV_active_2'] = bp_preds_df['HIV_active'].apply(lambda x: 1 if x > 0.8 else 0)
bp_preds_df.head()
```

	smiles	HIV_active	HIV_active_2
0	[Cl].CC(C)NCC(O)COc1cccc2ccccc12	0.206349	0
1	C(=O)(OC(C)(C)C)CCCc1ccc(cc1)N(CCCl)CCCl	0.141077	0
2	c12c3c(N4CCN(C)CC4)c(F)cc1c(c(C(O)=O)cn2C(C)CO...	0.473116	0
3	C1CCN(CC1)Cc1cccc(c1)OCCCNC(=O)C	0.160127	0
4	Cc1onc(c2ccccc2Cl)c1C(=O)N[C@H]3[C@H]4SC(C)(C)...	0.370032	0

Next steps: View recommended plots

```
bp_preds_df.describe()
```

	HIV_active	HIV_active_2
count	2039.000000	2039.000000
mean	0.300141	0.005885
std	0.132512	0.076508
min	0.010219	0.000000
25%	0.208201	0.000000
50%	0.274841	0.000000
75%	0.366519	0.000000
max	0.981535	1.000000

```
# Filter rows where 'target_column' is equal to 1
bp_preds_df_filtered = bp_preds_df[bp_preds_df['HIV_active_2'] == 1]
bp_preds_df_filtered
```

1 to 12 of 12 entries Filter

index	smiles	HIV_active	HIV_active_2
11	CC1=CN([C@H]2C[C@H](F)[C@@H](CO)O2)C(=O)NC1=O	0.9438915848731995	1
147	OC[C@H]1O[C (C[C@@H]1O)n2cnc3[C@H](O)CN=CNc23	0.8226878046989441	1
150	N[S](=O)(=O)c1cc2c(NC=N[S]2(=O)=O)cc1Cl	0.9183015823364258	1
235	OC[C@H]1O[C@H](C[C@@H]1O)N2C=C(F)C(=O)NC2=O	0.920660674571991	1
289	OC[C@@H]1CC[C@@H](O1)n2cnc3C(=O)N=CNc23	0.8001607060432434	1
319	CC1=CN([C@@H]2O[C@H](CO)C=C2)C(=O)NC1=O	0.9433940649032593	1
346	NC1=NC(=O)N(C=C1)[C@H]2CC[C@@H](CO)O2	0.8727076053619385	1
426	N[S](=O)(=O)c1cc2c(NC(N[S]2(=O)=O)C(Cl)Cl)cc1Cl	0.8049882054328918	1
445	CC1=CN([C@H]2C[C@H](N=[N+]=[N-])[C@@H](CO)O2)C(=O)NC1=O	0.9764421582221985	1
1512	C1=C(C(=C(C(=N1)C)O)CO)CSSCC2=CN=C(C(=C2CO)O)C	0.8555630445480347	1
1559	C1=CC=CC2=C1C4(C3=C(OC2)C=CC=C3)OC(CN(C)C)CO4	0.8222217559814453	1
2044	[N+](=[N-])=O	0.9815350770950317	1

Show 25 per page

Like what you see? Visit the data table notebook to learn more about interactive tables.

Next steps: View recommended plots

```
smiles_to_check = bp_preds_df_filtered['smiles'].to_list()
```

```
hiv_df_sampled_2[hiv_df_sampled_2['smiles'].isin(smiles_to_check)]
```

smiles	activity	HIV_active

```
hiv_df[hiv_df['smiles'].isin(smiles_to_check)]
```

smiles	activity	HIV_active

```
bp_df[bp_df['smiles'].isin(smiles_to_check)]
```

	num	name	p_np	smile
11	12	alovudine	1	CC1=CN([C@H]2C[C@H](F)[C@@H](CO)O2)C(=O)NC1=(
147	149	Pentostatin	1	OC[C@H]1O[C@H](C[C@@H]1O)n2cnc3[C@H (O)CN=CNc2
150	152	chlorothiazide	0	N[S](=O)(=O)c1cc2c(NC=N[S]2(=O)=O)cc1(
235	237	floxuridine	0	OC[C@H]1O[C@H](C[C@@H]1O)N2C=C(F)C(=O)NC2=(
289	291	Didanosine	0	OC[C@@H]1CC[C@@H](O1)n2cnc3C(=O)N=CNc2
319	321	Stavudine	1	CC1=CN([C@@H]2O[C@H](CO)C=C2)C(=O)NC1=(
346	348	Zalcitabine	1	NC1=NC(=O)N(C=C1)[C@H]2CC[C@@H](CO)O
426	428	Trichlormethiazide	0	N[S](=O)(=O)c1cc2c(NC(N[S]2(=O)=O)C(Cl)Cl)cc1(
445	447	zidovudine	0	CC1=CN([C@H]2C[C@H](N=[N+]=[N-])[C@@H](CO)O2)C.
1512	1516	pyritinol	1	C1=C(C(=C(C(=N1)C)O)CO)CSSCC2=CN=C(C(=C2CO)O)
1559	1563	spiroxepin	1	C1=CC=CC2=C1C4(C3=C(OC2)C=CC=C3)OC(CN(C)C)CO
2044	2048	nitrous-oxide	1	[N+](=[N-])=(

```
sub_df = pd.read_csv("substances.csv")
sub_df.head()
```

	zinc_id	smiles
0	ZINC000000000027	N[C@@H](CCc1ccc(N(CCCl)CCCl)cc1)C(=O)O
1	ZINC000016090786	N[C@H](CCc1ccc(N(CCCl)CCCl)cc1)C(=O)O
2	ZINC000001763088	N[C@H](CCCc1ccc(N(CCCl)CCCl)cc1)C(=O)O
3	ZINC000002033385	N[C@@H](CCCc1ccc(N(CCCl)CCCl)cc1)C(=O)O
4	ZINC000000001673	N[C@@H](Cc1ccc(N(CCCl)CCCl)cc1)C(=O)O

Next steps: View recommended plots

```
sub_df.tail()
```

	zinc_id	smiles
46	ZINC000196349655	O=C(O)CCSc1ccc(N(CCCl)CCCl)cc1
47	ZINC000064454242	N=NCCCc1ccc(N(CCCl)CCCl)cc1
48	ZINC000005161807	O=C(O)C/C=C/c1ccc(N(CCCl)CCCl)cc1
49	ZINC000001682294	O=C(O)CCOc1ccc(N(CCCl)CCCl)cc1
50	ZINC000079564304	O=C(O)CNC(=O)c1ccc(N(CCCl)CCCl)cc1

```
sub_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 2 columns):
 #   Column   Non-Null Count  Dtype 
---  ------   --------------  ----- 
 0   zinc_id  51 non-null     object
 1   smiles   51 non-null     object
dtypes: object(2)
memory usage: 944.0+ bytes
```

```
arguments = [
    '--test_path', 'substances.csv',
    '--preds_path', 'substances_preds.csv',
    '--checkpoint_dir', 'test_checkpoints_multimolecule',
    '--smiles_columns', 'smiles'
]

args = chemprop.args.PredictArgs().parse_args(arguments)
preds = chemprop.train.make_predictions(args=args)
```

```
Loading training args
Setting molecule featurization parameters to default.
Loading data
51it [00:00, 56800.19it/s]
100%|██████████| 51/51 [00:00<00:00, 39041.71it/s]
/usr/local/lib/python3.10/dist-packages/torch/utils/data/dataloader.py:557: UserWarning: This DataLoader will cre
  warnings.warn(_create_warning_msg(
Validating SMILES
Test size = 51
  0%|          | 0/1 [00:00<?, ?it/s]Loading pretrained parameter "encoder.encoder.0.cached_zero_vector".
Loading pretrained parameter "encoder.encoder.0.W_i.weight".
Loading pretrained parameter "encoder.encoder.0.W_h.weight".
Loading pretrained parameter "encoder.encoder.0.W_o.weight".
Loading pretrained parameter "encoder.encoder.0.W_o.bias".
Loading pretrained parameter "readout.1.weight".
Loading pretrained parameter "readout.1.bias".
Loading pretrained parameter "readout.4.weight".
Loading pretrained parameter "readout.4.bias".
Moving model to cuda

  0%|          | 0/2 [00:00<?, ?it/s]
 50%|█████     | 1/2 [00:00<00:00,  2.44it/s]
100%|██████████| 1/1 [00:01<00:00,  1.37s/it]Saving predictions to substances_preds.csv
Elapsed time = 0:00:02
```

```
fda_df = pd.read_csv("fda_approved.csv")
fda_df.head()
```

	zinc_id	smiles
0	ZINC000001530427	C[C@@H]1O[C@@H]1P(=O)(O)O
1	ZINC000003807804	Clc1ccccc1C(c1ccccc1)(c1ccccc1)n1ccnc1
2	ZINC000000120286	Nc1nc(N)c2nc(-c3ccccc3)c(N)nc2n1
3	ZINC000242548690	C[C@H]1O[C@@H](O[C@H]2[C@@H](O)C[C@H](O[C@H]3[...
4	ZINC000000008492	Oc1cccc2cccnc12

Next steps: View recommended plots

```
arguments = [
    '--test_path', 'fda_approved.csv',
    '--preds_path', 'fda_approved_preds.csv',
    '--checkpoint_dir', 'test_checkpoints_multimolecule',
    '--smiles_columns', 'smiles'
]

args = chemprop.args.PredictArgs().parse_args(arguments)
preds = chemprop.train.make_predictions(args=args)
```

```
Loading training args
Setting molecule featurization parameters to default.
Loading data
892it [00:00, 161716.84it/s]
100%|██████████| 892/892 [00:00<00:00, 130929.80it/s]
/usr/local/lib/python3.10/dist-packages/torch/utils/data/dataloader.py:557: UserWarning: This DataLoader will cre
  warnings.warn(_create_warning_msg(
Validating SMILES
Test size = 892
  0%|          | 0/1 [00:00<?, ?it/s]Loading pretrained parameter "encoder.encoder.0.cached_zero_vector".
Loading pretrained parameter "encoder.encoder.0.W_i.weight".
Loading pretrained parameter "encoder.encoder.0.W_h.weight".
Loading pretrained parameter "encoder.encoder.0.W_o.weight".
Loading pretrained parameter "encoder.encoder.0.W_o.bias".
Loading pretrained parameter "readout.1.weight".
Loading pretrained parameter "readout.1.bias".
Loading pretrained parameter "readout.4.weight".
Loading pretrained parameter "readout.4.bias".
Moving model to cuda

  0%|          | 0/18 [00:00<?, ?it/s]
  6%|▌         | 1/18 [00:01<00:26,  1.57s/it]
 11%|█         | 2/18 [00:01<00:13,  1.21it/s]
 22%|██        | 4/18 [00:02<00:06,  2.16it/s]
 50%|█████     | 9/18 [00:02<00:01,  6.20it/s]
 67%|██████▋   | 12/18 [00:02<00:00,  8.62it/s]
100%|██████████| 1/1 [00:03<00:00,  3.08s/it]Saving predictions to fda_approved_preds.csv
Elapsed time = 0:00:03
```

```
fda_preds_df = pd.read_csv("fda_approved_preds.csv")
fda_preds_df.head()
```

	zinc_id	smiles	HIV_active
0	ZINC000001530427	C[C@@H]1O[C@@H]1P(=O)(O)O	0.009933
1	ZINC000003807804	Clc1ccccc1C(c1ccccc1)(c1ccccc1)n1ccnc1	0.596011
2	ZINC000000120286	Nc1nc(N)c2nc(-c3ccccc3)c(N)nc2n1	0.106520
3	ZINC000242548690	C[C@H]1O[C@@H](O[C@H]2[C@@H](O)C[C@H](O[C@H]3[...	0.734897
4	ZINC000000008492	Oc1cccc2cccnc12	0.092682

Next steps: View recommended plots

```
fda_preds_df = fda_preds_df[fda_preds_df['HIV_active'] != "Invalid SMILES"]
fda_preds_df.describe()
fda_preds_df['HIV_active'] = fda_preds_df['HIV_active'].astype(float)
fda_preds_df['HIV_active_2'] = fda_preds_df['HIV_active'].apply(lambda x: 1 if x > 0.8 else 0)
fda_preds_df.head()
```

	HIV_active	HIV_active_2
count	892.000000	892.000000
mean	0.297628	0.223094
std	0.232897	0.416555
min	0.002699	0.000000
25%	0.109553	0.000000
50%	0.230911	0.000000
75%	0.421475	0.000000
max	0.996994	1.000000

	zinc_id	smiles	HIV_active	HIV_active_2
0	ZINC000001530427	C[C@@H]1O[C@@H]1P(=O)(O)O	0.009933	0
1	ZINC000003807804	Clc1ccccc1C(c1ccccc1)(c1ccccc1)n1ccnc1	0.596011	0
2	ZINC000000120286	Nc1nc(N)c2nc(-c3ccccc3)c(N)nc2n1	0.106520	0
		C[C@H]1O[C@@H]		

Next steps: View recommended plots View recommended plots

```
# Filter rows where 'target_column' is equal to 1
fda_preds_df_filtered = fda_preds_df[fda_preds_df['HIV_active_2'] == 1]
fda_preds_df_filtered
```

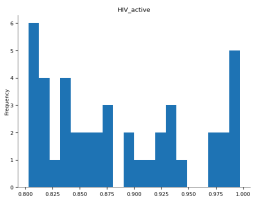
1 to 25 of 43 entries Filter

index	zinc_id	smiles
47	ZINC000003813010	O=c1[nH]c(=O)n([C@H]2C[C@H](O)[C@@H](CO)O2)cc1F
55	ZINC000000592419	O=C(Nc1c(Cl)cncc1Cl)c1ccc(OC(F)F)c(OCC2CC2)c1
81	ZINC000003818726	O=C(/C=C/c1cccc(S(=O)(=O)Nc2ccccc2)c1)NO
153	ZINC000245204949	C[N@+]1(CC2CC2)CC[C@]23c4c5ccc(O)c4O[C@H]2C(=O)CC[C@@]3(O)[C@H]1C5
158	ZINC000003830391	CC1=C(C(=O)O)N2C(=O)[C@@H](NC(=O)[C@H](N)c3ccc(O)cc3)[C@H]2SC1
197	ZINC000000005423	Cc1nc(-c2ccc(OCC(C)C)c(C#N)c2)sc1C(=O)O
228	ZINC000009302239	NC(=O)[C@@H]1CC[C@@H]2CN1C(=O)N2OS(=O)(=O)O
247	ZINC000003922770	C[C@@H](O)[C@H]1C(=O)N2C(C(=O)O)=C(S[C@@H]3CN[C@H](CNS(N)(=O)=O)C3)[C@H](C)[C@H]12
264	ZINC000040899447	CS(=O)(=O)c1ccc(C(=O)Nc2ccc(Cl)c(-c3ccccn3)c2)c(Cl)c1
283	ZINC000003955219	CC(C)CN(C[C@@H](O)[C@H](Cc1ccccc1)NC(=O)O[C@H]1CO[C@H]2OCC[C@@H]12)S(=O)(=O)c1ccc(N)cc1
313	ZINC000014210457	CC(C)(C)NC(=O)N[C@H](C(=O)N1C[C@H]2[C@@H]([C@H]1C(=O)N[C@H](CC1CCC1)C(=O)C(N)=O)C2(C)C)C(C)(C)C
314	ZINC000014879992	CN(C)c1ccc(O)c2c1C[C@H]1C[C@H]3[C@H](N(C)C)C(O)=C(C(N)=O)C(=O)[C@@]3(O)C(O)=C1C2=O
321	ZINC000013597823	O=c1[nH]cnc2c1ncn2[C@H]1CC[C@@H](CO)O1
324	ZINC000019632618	Cc1ccc(NC(=O)c2ccc(CN3CCN(C)CC3)cc2)cc1Nc1nccc(-c2cccnc2)n1
340	ZINC000001530621	CCN[C@H]1C[C@H](C)S(=O)(=O)c2sc(S(N)(=O)=O)cc21
410	ZINC000004097225	C[C@@H](O)[C@H]1C(=O)N2C(C(=O)O)=C(SCCNC=N)C[C@H]12
443	ZINC000000643114	C[C@@H]1Cc2ccccc2N1NC(=O)c1ccc(Cl)c(S(N)(=O)=O)c1
479	ZINC000003830264	C[C@H]1[C@H](NC(=O)/C(=N\OC(C)(C)C(=O)O)c2csc(N)n2)C(=O)N1S(=O)(=O)O
513	ZINC000058581064	C[C@@H]1CCO[C@H]2Cn3cc(C(=O)NCc4ccc(F)cc4F)c(=O)c(O)c3C(=O)N21
612	ZINC000003929508	CCOC(=O)C1=C[C@@H](OC(CC)CC)[C@H](NC(C)=O)[C@@H](N)C1
641	ZINC000038212689	O=c1[nH]cc(F)c(=O)[nH]1
662	ZINC000035342787	CCN(CC)C(=O)/C(C#N)=C/c1cc(O)c(O)c([N+](=O)[O-])c1
677	ZINC000000012346	Nc1ccn([C@@H]2CS[C@H](CO)O2)c(=O)n1
690	ZINC000000601305	C[C@H]1Cc2ccccc2N1NC(=O)c1ccc(Cl)c(S(N)(=O)=O)c1
712	ZINC000004095696	CC1(C)C[C@@H]1C(=O)N/C(=C\CCCCSC[C@H](N)C(=O)O)C(=O)O

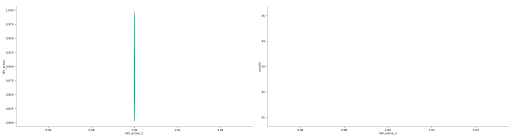
Show 25 per page 1 2

Like what you see? Visit the data table notebook to learn more about interactive tables.

Distributions



Time series



Values

Next steps: View recommended plots

```
!wget https://zinc15.docking.org/substances/subsets/named.csv
```

```
--2024-03-10 05:58:28--  https://zinc15.docking.org/substances/subsets/named.csv
Resolving zinc15.docking.org (zinc15.docking.org)... 169.230.75.4
Connecting to zinc15.docking.org (zinc15.docking.org)|169.230.75.4|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/csv]
Saving to: ‘named.csv.1’

named.csv.1             [ <=>                ]   9.28K  --.-KB/s    in 0.04s

2024-03-10 05:58:29 (242 KB/s) - ‘named.csv.1’ saved [9499]
```

```
zinc_df = pd.read_csv("named.csv")
zinc_df.head()
zinc_df.tail()
```

```
---------------------------------------------------------------------------
FileNotFoundError                         Traceback (most recent call last)
<ipython-input-123-5bc884f87412> in <cell line: 1>()
----> 1 zinc_df = pd.read_csv("named.csv")
      2 zinc_df.head()
      3 zinc_df.tail()

 6 frames 
/usr/local/lib/python3.10/dist-packages/pandas/io/common.py in
get_handle(path_or_buf, mode, encoding, compression, memory_map, is_text,
errors, storage_options)
    854         if ioargs.encoding and "b" not in ioargs.mode:
    855             # Encoding
--> 856             handle = open(
    857                 handle,
    858                 ioargs.mode,

FileNotFoundError: [Errno 2] No such file or directory: 'named.csv'
```

```
arguments = [
    '--test_path', 'named.csv',
    '--preds_path', 'named_preds.csv',
    '--checkpoint_dir', 'test_checkpoints_multimolecule',
    '--smiles_columns', 'smiles'
]

args = chemprop.args.PredictArgs().parse_args(arguments)
preds = chemprop.train.make_predictions(args=args)
```

```
Loading training args
Setting molecule featurization parameters to default.
Loading data
100it [00:00, 59764.95it/s]
100%|██████████| 100/100 [00:00<00:00, 66905.47it/s]
/usr/local/lib/python3.10/dist-packages/torch/utils/data/dataloader.py:557: UserWarning: This DataLoader will cre
  warnings.warn(_create_warning_msg(
Validating SMILES
Test size = 100
  0%|          | 0/1 [00:00<?, ?it/s]Loading pretrained parameter "encoder.encoder.0.cached_zero_vector".
Loading pretrained parameter "encoder.encoder.0.W_i.weight".
Loading pretrained parameter "encoder.encoder.0.W_h.weight".
Loading pretrained parameter "encoder.encoder.0.W_o.weight".
Loading pretrained parameter "encoder.encoder.0.W_o.bias".
Loading pretrained parameter "readout.1.weight".
Loading pretrained parameter "readout.1.bias".
Loading pretrained parameter "readout.4.weight".
Loading pretrained parameter "readout.4.bias".
```