Ideas bachelor feature engineering

Solution to sort lon and lat to one station with the algorithm known.  
  
Using a 2023 city as test set  
  
combine 3 citys in nrw and test with 4th  
  
schreib hyerparameter spezifizierung rein  
  
Starting hyperparameter tuning...  
Fitting 5 folds for each of 50 candidates, totalling 250 fits  
Best Hyperparameters: {'subsample': 0.6, 'n\_estimators': 200, 'max\_depth': 10, 'learning\_rate': 0.01}  
Best Cross-Validation Accuracy: 0.6989  
Total tuning time: 18755.91 seconds  
  
C:\Users\Flunch\miniconda3\python.exe C:\Users\Flunch\Desktop\Bachelor\_thesis\_nextbike\ready\_for\_modeling\data\no\_split\_test.py

Evaluating on 2022\_complete\_dresden.csv...

Results for 2022\_complete\_dresden.csv:

Classifier Accuracy: 0.6770

Regressor Mean Squared Error: 7.9820

Evaluating on 2022\_complete\_nürnberg.csv...

Results for 2022\_complete\_nürnberg.csv:

Classifier Accuracy: 0.8703

Regressor Mean Squared Error: 28.1098

Evaluating on 2022\_complete\_heidelberg.csv...

Results for 2022\_complete\_heidelberg.csv:

Classifier Accuracy: 0.8712

Regressor Mean Squared Error: 55.5019

Evaluating on 2022\_complete\_essen.csv...

Results for 2022\_complete\_essen.csv:

Classifier Accuracy: 0.7856

Regressor Mean Squared Error: 34.7673

Evaluating on 2022\_combined\_city\_data.csv...

Results for 2022\_combined\_city\_data.csv:

Classifier Accuracy: 0.7689

Regressor Mean Squared Error: 25.9542

Process finished with exit code 0

Evaluating on 2022\_complete\_dresden.csv...

Results for 2022\_complete\_dresden.csv:

Classifier Accuracy: 0.6783

Regressor Mean Squared Error: 7.9661

Evaluating on 2022\_complete\_nürnberg.csv...

Results for 2022\_complete\_nürnberg.csv:

Classifier Accuracy: 0.8696

Regressor Mean Squared Error: 28.1373

Evaluating on 2022\_complete\_heidelberg.csv...

Results for 2022\_complete\_heidelberg.csv:

Classifier Accuracy: 0.8718

Regressor Mean Squared Error: 55.5642

Evaluating on 2022\_complete\_essen.csv...

Results for 2022\_complete\_essen.csv:

Classifier Accuracy: 0.7865

Regressor Mean Squared Error: 34.7457

Evaluating on 2022\_combined\_city\_data.csv...

Results for 2022\_combined\_city\_data.csv:

Classifier Accuracy: 0.7697

Regressor Mean Squared Error: 25.9569

Best Parameters: {'colsample\_bytree': 0.5765133157615585, 'gamma': 2.8319073793101883, 'learning\_rate': 0.015406375121368878, 'max\_depth': 11, 'min\_child\_weight': 1, 'n\_estimators': 70, 'subsample': 0.9033844439161279}

Best Cross-Validation Accuracy: 0.5668  
XGBoost Classifier Accuracy with Tuned Hyperparameters: 0.7783

Evaluating on 2022\_combined\_city\_data.csv...

Results for 2022\_combined\_city\_data.csv:

XGBoost Classifier Accuracy: 0.7783

Evaluating on 2022\_complete\_dresden.csv...

Results for 2022\_complete\_dresden.csv:

XGBoost Classifier Accuracy: 0.6427

Evaluating on 2022\_complete\_essen.csv...

Results for 2022\_complete\_essen.csv:

XGBoost Classifier Accuracy: 0.8444

Evaluating on 2022\_complete\_heidelberg.csv...

Results for 2022\_complete\_heidelberg.csv:

XGBoost Classifier Accuracy: 0.8798

Evaluating on 2022\_complete\_nürnberg.csv...

Results for 2022\_complete\_nürnberg.csv:

XGBoost Classifier Accuracy: 0.9136

C:\Users\Flunch\miniconda3\python.exe C:\Users\Flunch\Desktop\Bachelor\_thesis\_nextbike\ready\_for\_modeling\data\xgbm\_hyper\_regressor.py

Fitting 5 folds for each of 500 candidates, totalling 2500 fits

Best Parameters: {'colsample\_bytree': 0.6977924525180372, 'gamma': 4.897754201908785, 'learning\_rate': 0.03899301447596341, 'max\_depth': 9, 'min\_child\_weight': 5, 'n\_estimators': 265, 'subsample': 0.6657489216128507}

Best Cross-Validation MSE: 18.6461

Evaluating on 2022\_combined\_city\_data.csv...

XGBoost Regressor MSE with Tuned Hyperparameters: 19.6229

Process finished with exit code 0

There is also a maintanacnce value in the json