# Feature Engineering in Forecasting

# Open source libraries

## for feature engineering in forecasting

# Traditional feature engineering methods



- fit() ➜ learns and stores parameters

- transform() ➜ transforms data

For engineering steps that are also used in tabular data

# Feature engineering for forecasting

- fit() ➔ learns and stores parameters

- transform() ➔ transforms data

For engineering steps that are also used in tabular data

**Category Encoders**

fit() and transform()

Lag and window features

# Feature engineering for forecasting

- fit() ➜ learns and stores parameters

- transform() ➜ transforms data

For engineering steps that are also used in tabular data

fit() and transform()

Lag and window features

Wrappers for compatibility with pipeline

Lag and window features

# Can I still use pipelines?

```python
# set up pipeline
our_pipe = Pipeline([
        ("step1", Imputation()),
        ("step2", CategoricalEncoding()),
        ("step3", WindowFeatures()),
        ("step4", LagFeatures()),
        ("model", Lasso()),
])

# train pipeline
our_pipe.fit(X_train, y_train)

# predict
our_pipe.predict(X_train)

our_pipe.predict(X_test)
```

# Feature engineering pipelines

```python
# set up pipeline
our_pipe = Pipeline([
        ("step1", Imputation()),
        ("step2", CategoricalEncoding()),
        ("step3", WindowFeatures()),
        ("step4", LagFeatures()),
        ("model", Lasso()),
])

# train pipeline
our_pipe.fit(X_train, y_train)

# predict
our_pipe.predict(X_train)

our_pipe.predict(X_test)
```

| t1 | t2 | t3 | t4 | t5 | t6 |
|----|----|----|----|----|----|

# Feature engineering pipelines

```
# set up pipeline
our_pipe = Pipeline([
        ("step1", Imputation()),
        ("step2", CategoricalEncoding()),
        ("step3", WindowFeatures()),
        ("step4", LagFeatures()),
        ("model", Lasso()),
])


# train pipeline
our_pipe.fit(X_train, y_train)


# predict
our_pipe.predict(X_train)


our_pipe.predict(X_test)
```

| t1 | t2 | t3 | t4 | t5 | t6 |

| t1 | t2 | t3 | t4 | t5 | t6 |

| t4 | t5 | t6 | t7 | t8 | t9 |

# Which data can I use as features?

**All data that is available at time of prediction.**

We need to be **very careful** not to accidentally use data that is not available at time of prediction.

# Summary

Some features can't be created after we split the data.

We can use all data available at time of prediction.

No unified approach to create features from past data.