PROJECT REPORT

Bakery sales prediction

Matthias Faust
Ngoc Linh Nguyen
Thorsten Köhler

VARIABLES by group

WEATHER

Weather Categories
Rolling Mean Temperature
Peak Temperature

HOLIDAY

Public holidays
School holidays
Kieler Woche festival

RESIDUAL

Oddities

TIME

Weekend
Day of the week
Month of the year

ECONOMY

GDP
CPI
Unemployment rate

Weather features

Varname	Original variable	Description
Weather_category	Wettercode (Weather code)	Rain, After rain, Showers & Thunderstorms, Snow & Ice, Fog, Others
Temperature_class	Temperatur (Temperature)	Cold, Cool, Mild, Warm, Unknown
Cloud_class	Bewoelkung (Cloud cover)	Sunny, Cloudy, Unknown
Wind_class	Windgeschwindigkeit (Wind speed)	Breeze, Wind, Storm, Unknown
Temperatur		Temperature
Rolling_temp_mean		Mean temperature of the last 5 Days
Temperature above/below mean		Significant difference to the rolling temp mean

MISSING VALUE IMPUTATION

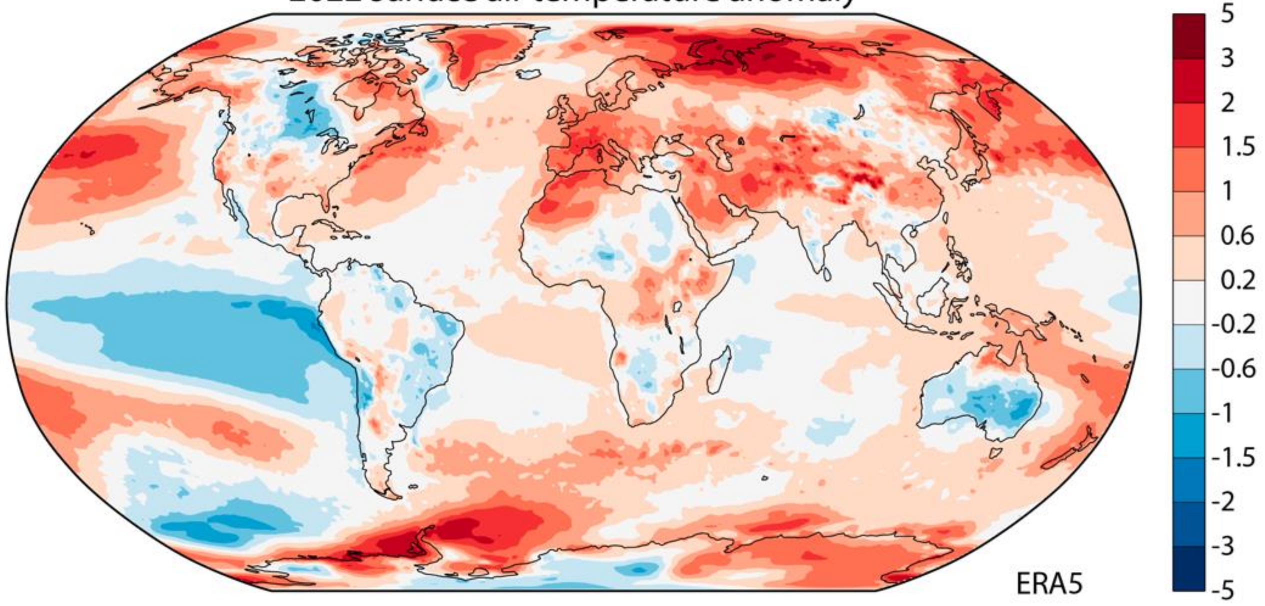
ERA5 Reanalysis Data

- Simulation of the past weather, prompted by measurements

- Data openly available from 1940 to present

- Hourly time resolution

2022 surface air temperature anomaly



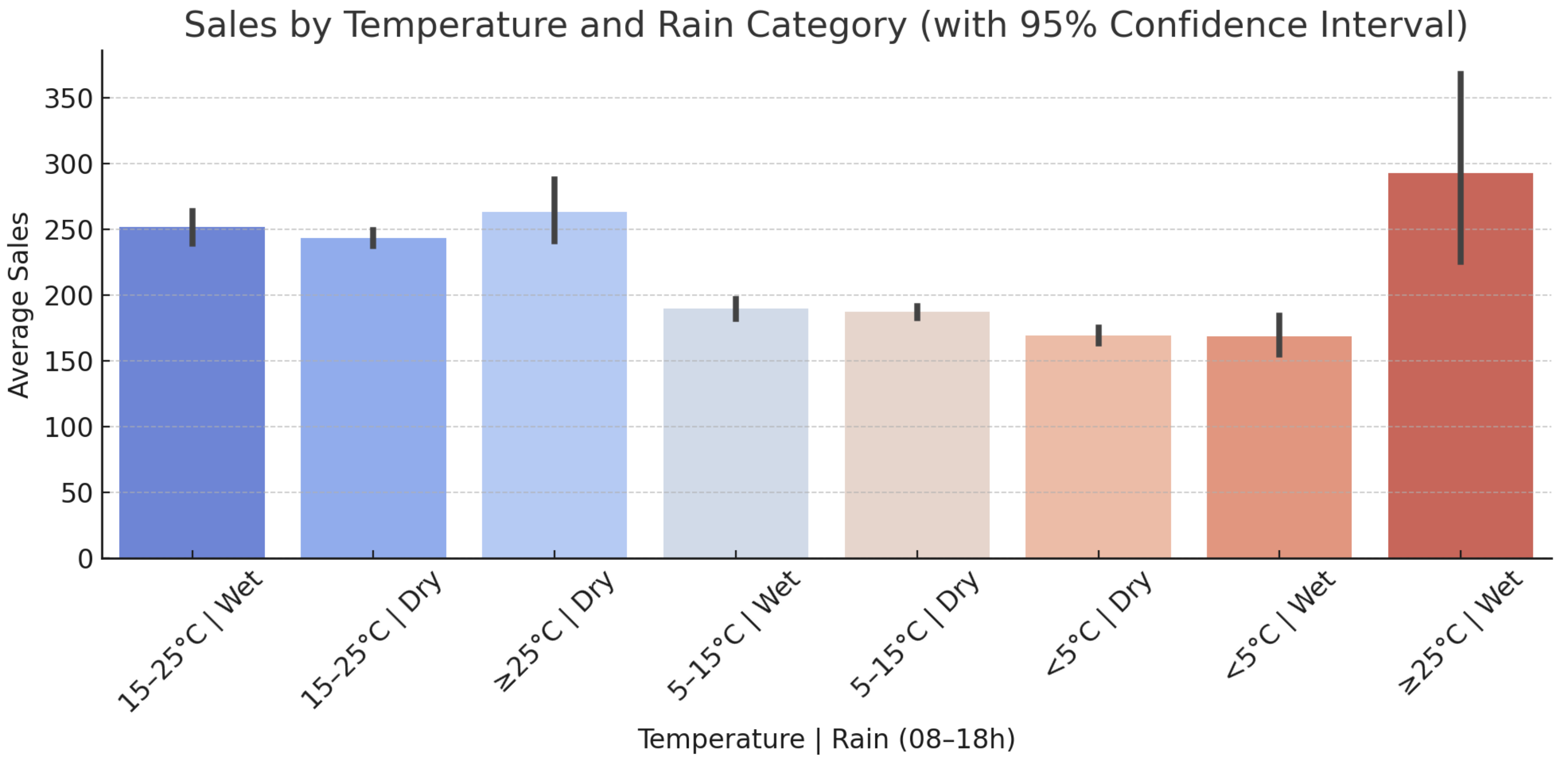
Copernicus Climate Change Service
Climate Indicators | 2022

PROGRAMME OF THE EUROPEAN UNION

Copernicus Europe's eyes on Earth

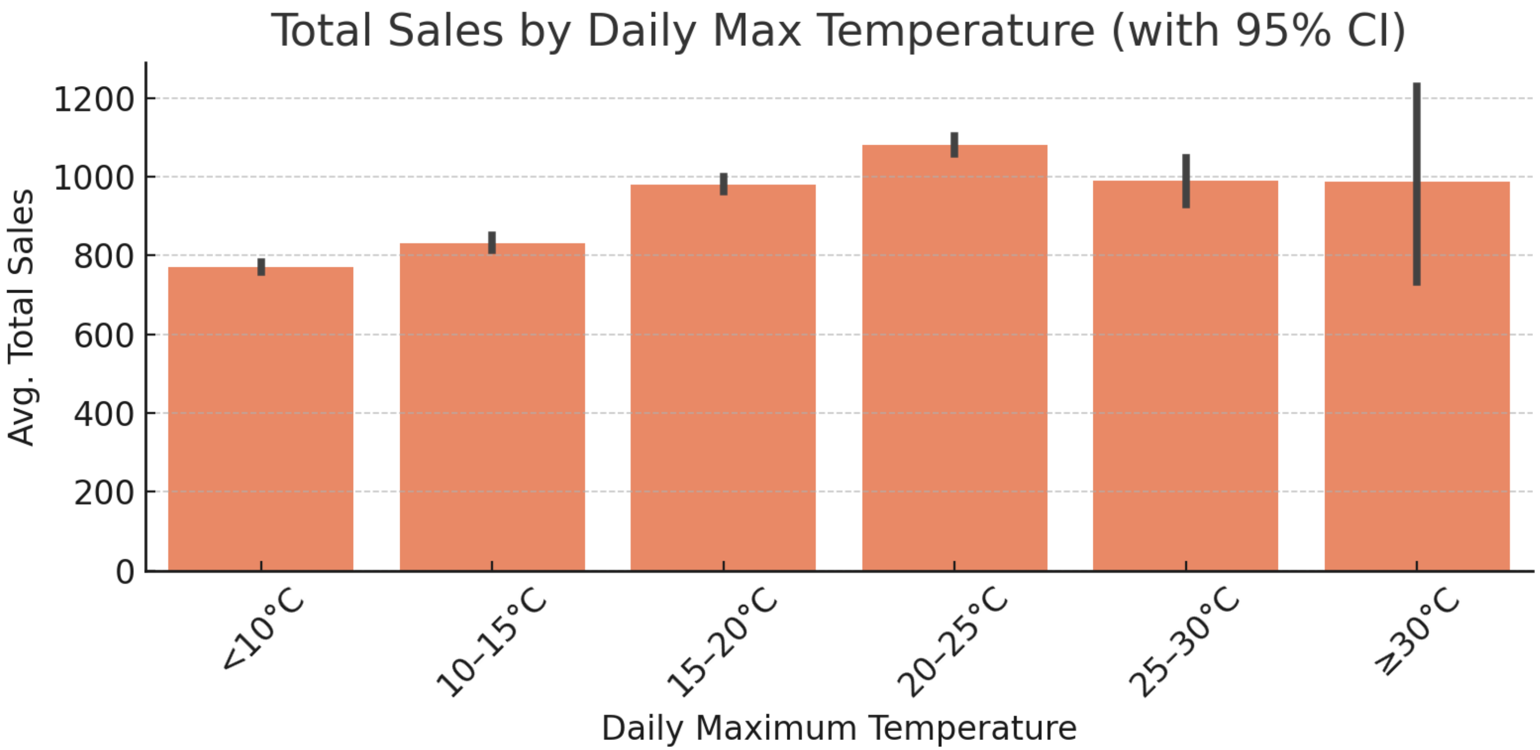
IMPLEMENTED BY ECMWF

Weather variable



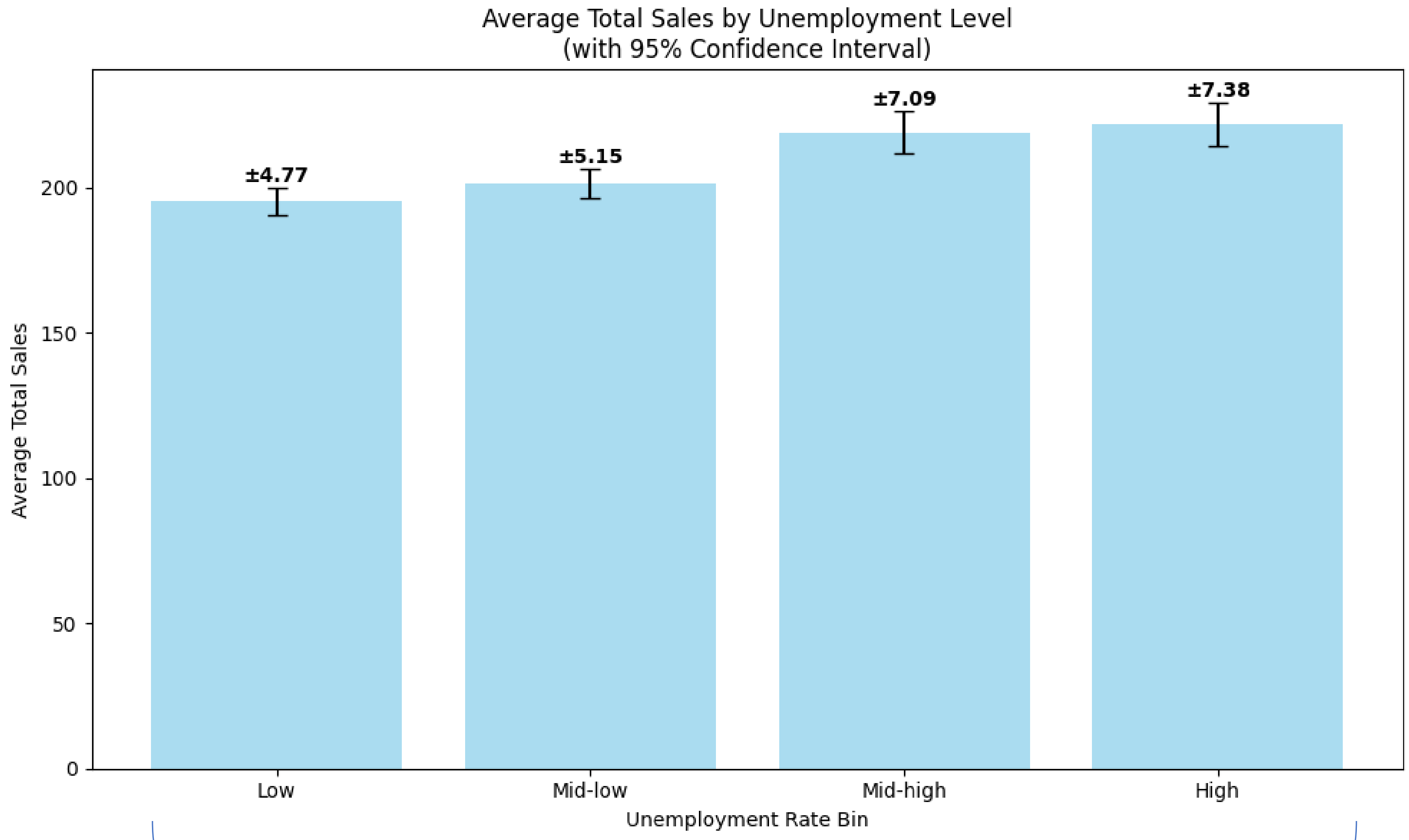
Combination of rain and high temperature increases sales

Weather variable



Sales increase with temperature but decrease if the maximum temperature is above 25°C

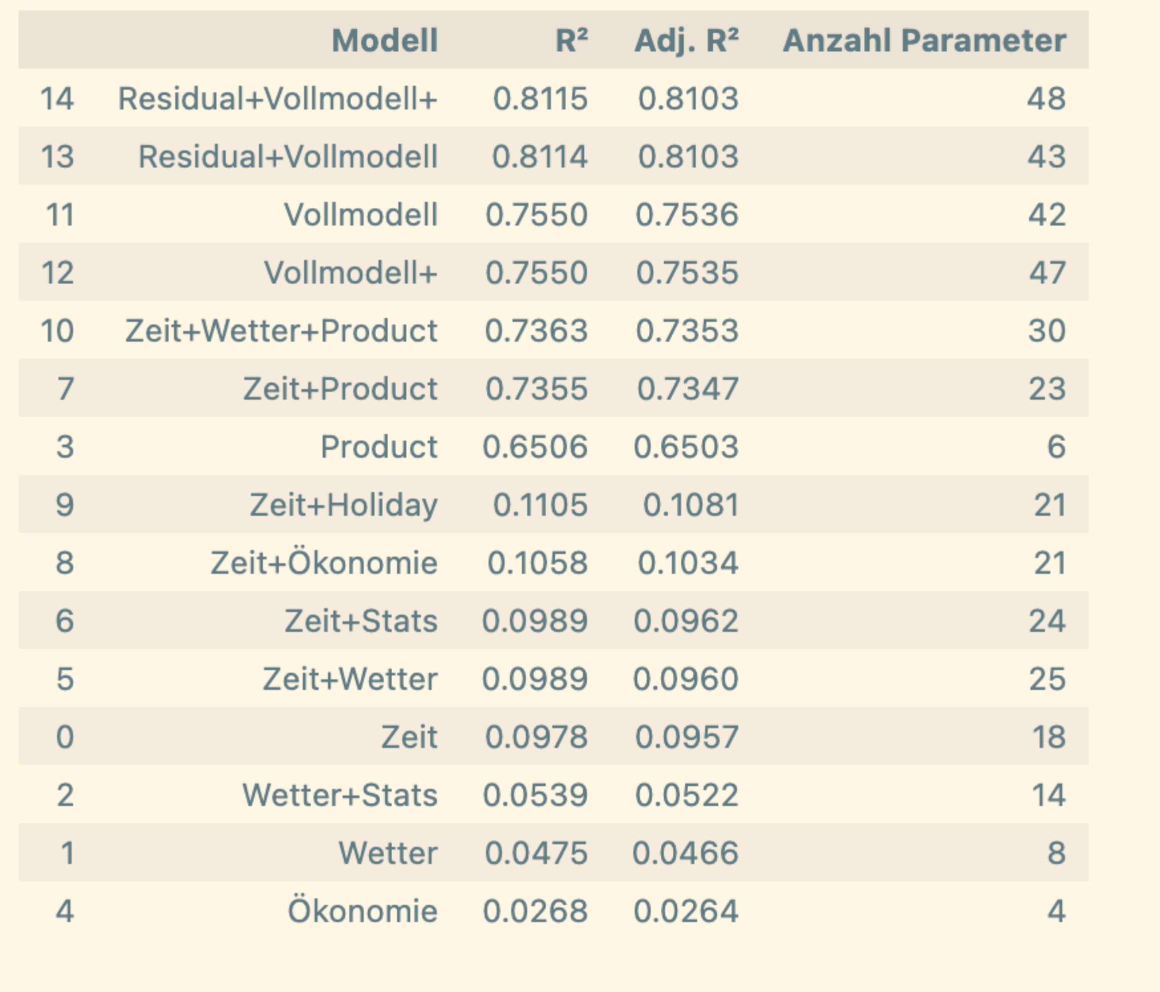
Unemployment rate



4 equally sized groups

Higher unemployment rate is associated with slightly higher average total sales

LINEAR REGRESSION MODEL



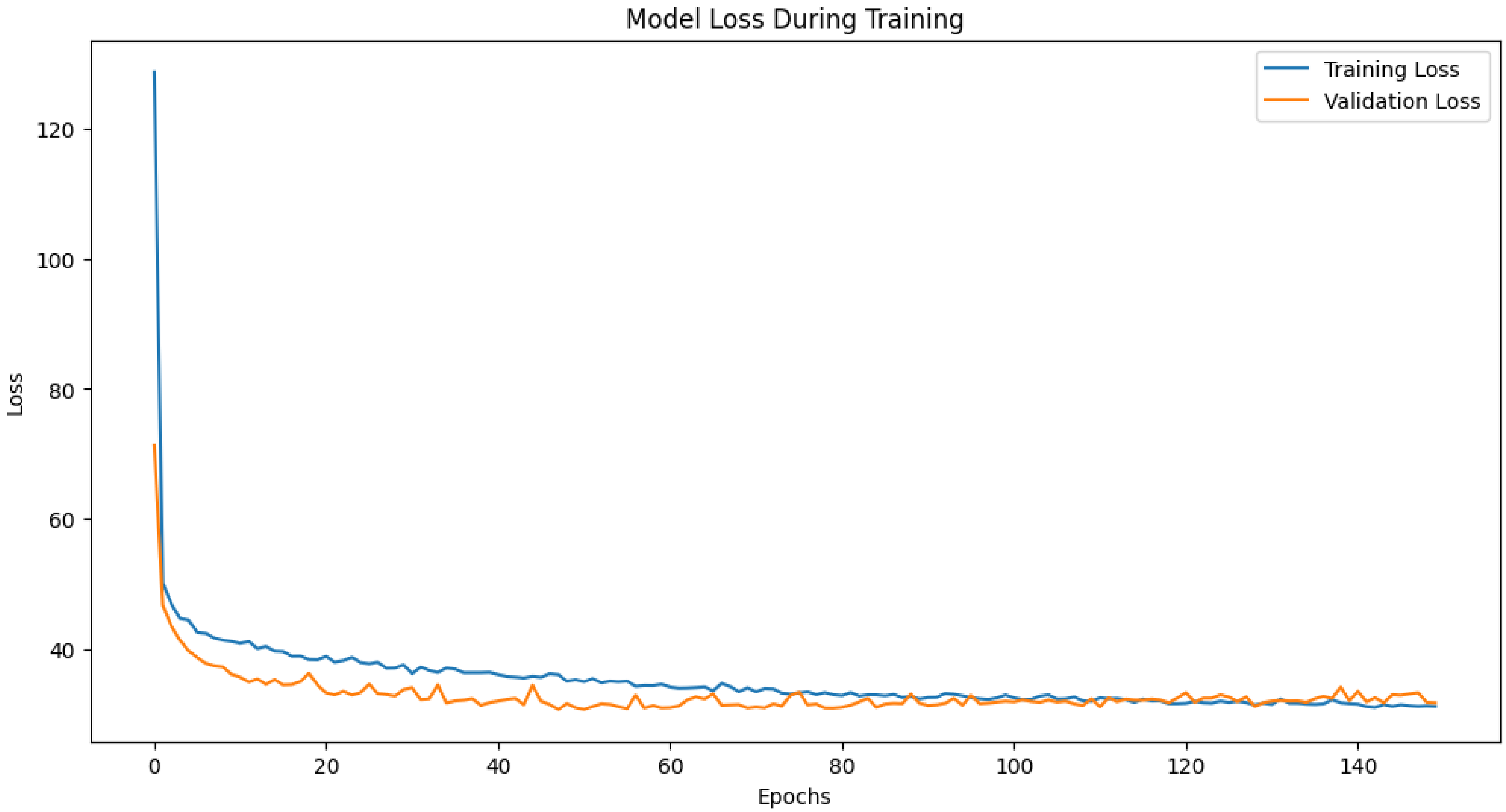
	Modell	R²	Adj. R²	Anzahl Parameter
14	Residual+Vollmodell+	0.8115	0.8103	48
13	Residual+Vollmodell	0.8114	0.8103	43
11	Vollmodell	0.7550	0.7536	42
12	Vollmodell+	0.7550	0.7535	47
10	Zeit+Wetter+Product	0.7363	0.7353	30
7	Zeit+Product	0.7355	0.7347	23
3	Product	0.6506	0.6503	6
9	Zeit+Holiday	0.1105	0.1081	21
8	Zeit+Ökonomie	0.1058	0.1034	21
6	Zeit+Stats	0.0989	0.0962	24
5	Zeit+Wetter	0.0989	0.0960	25
0	Zeit	0.0978	0.0957	18
2	Wetter+Stats	0.0539	0.0522	14
1	Wetter	0.0475	0.0466	8
4	Ökonomie	0.0268	0.0264	4

→ *Sales ~ Product + Time*
+ Weather + Economy
+ Holiday + Residual

NEURAL NETWORK

```
from tensorflow.keras.layers import Dropout
from tensorflow.keras.layers import Dense, InputLayer, BatchNormalization
model = Sequential([
    InputLayer(shape=(training_features.shape[1], )),
    BatchNormalization(),
    Dense(64, activation='relu'),
    Dropout(0.2),
    Dense(32, activation='relu'),
    Dense(1)
])
```

NEURAL NETWORK



Loss function: Huber

Learning rate: 0.001

Epochs: 150

Batch size: 32

MAPE

- *Training Data: 16.23%*
- *Validation Data: 17.99%*
  - *Bread: 18.97%*
  - *Roll: 11.98%*
  - Croissant*: 18.77%*
  - Confectionery*: 25.82%*
  - *Cake: 14.44%*
  - *Seasonal Bread: -- %*
    (not in the validation period)

HIGHLIGHT

Worst fails:

- *Working with Git codespaces and Tensorflow*
- *The quite often seemingly esoteric effects of Feature Engineering and Model Optimization*

Best improvement:

- *Encoding date and weather features*
- *Talking things through as a team – very important to get a better understanding*

Thank You