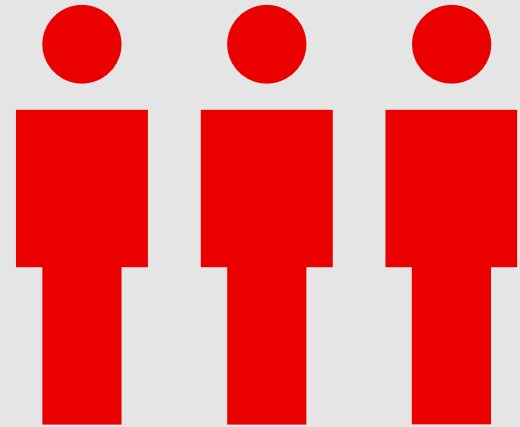
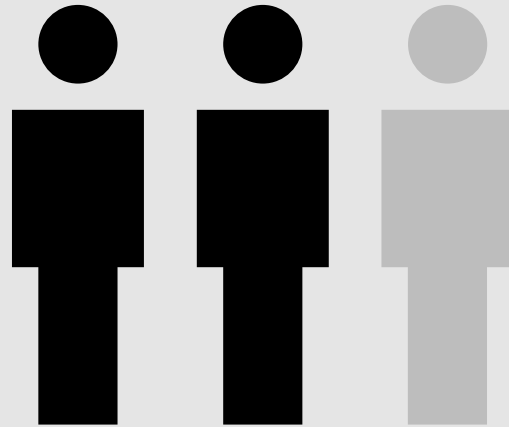
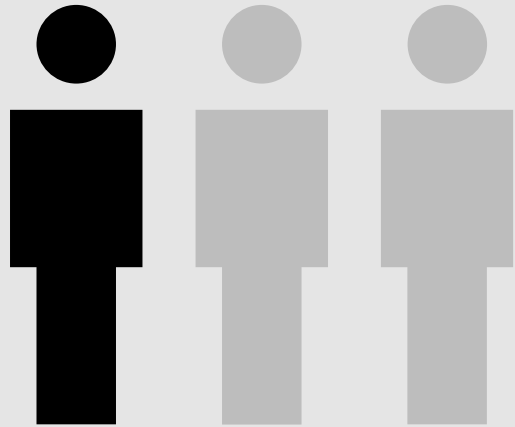


# The SBB Occupancy Forecast

SDS2024

Matthias Minder, Frederik Steiner







# The Landscape.

# The two flavors of the occupancy forecast.

**1. 2.**  **Class-specific forecast**

**S 26** Richtung Lenzburg  
**17:59** ————— **18:25**  
Gl. 2   1. 2.      26 min

Applications:

- Planning of rolling stock
- Supersaver tickets
- sbb.ch & SBB Mobile
- Open Data
- Emergency handling

**1** **Wagon-specific forecast**

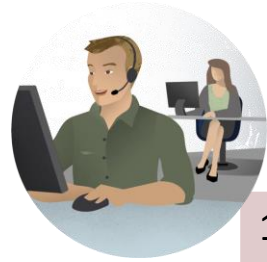
Sektor

1   2   2   2

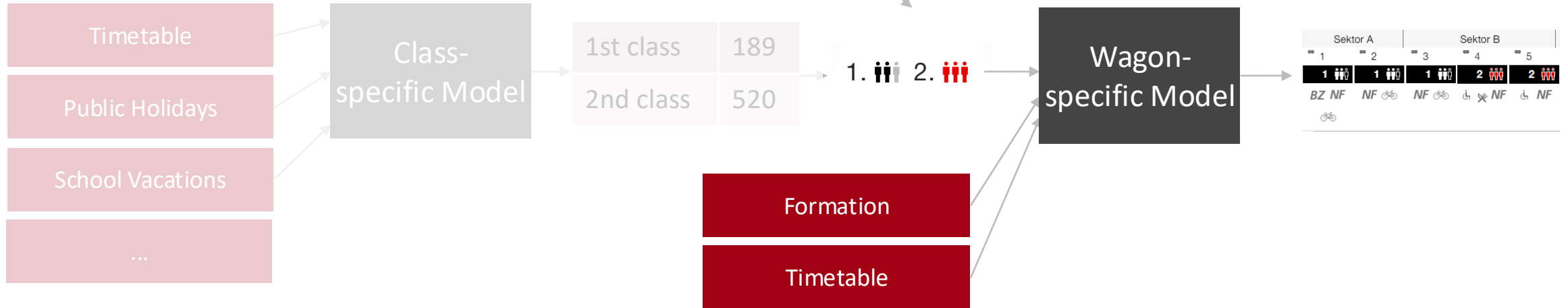
Applications:

- sbb.ch & SBB Mobile
- Display Boards

# The Process.



1st class	250
2nd class	900



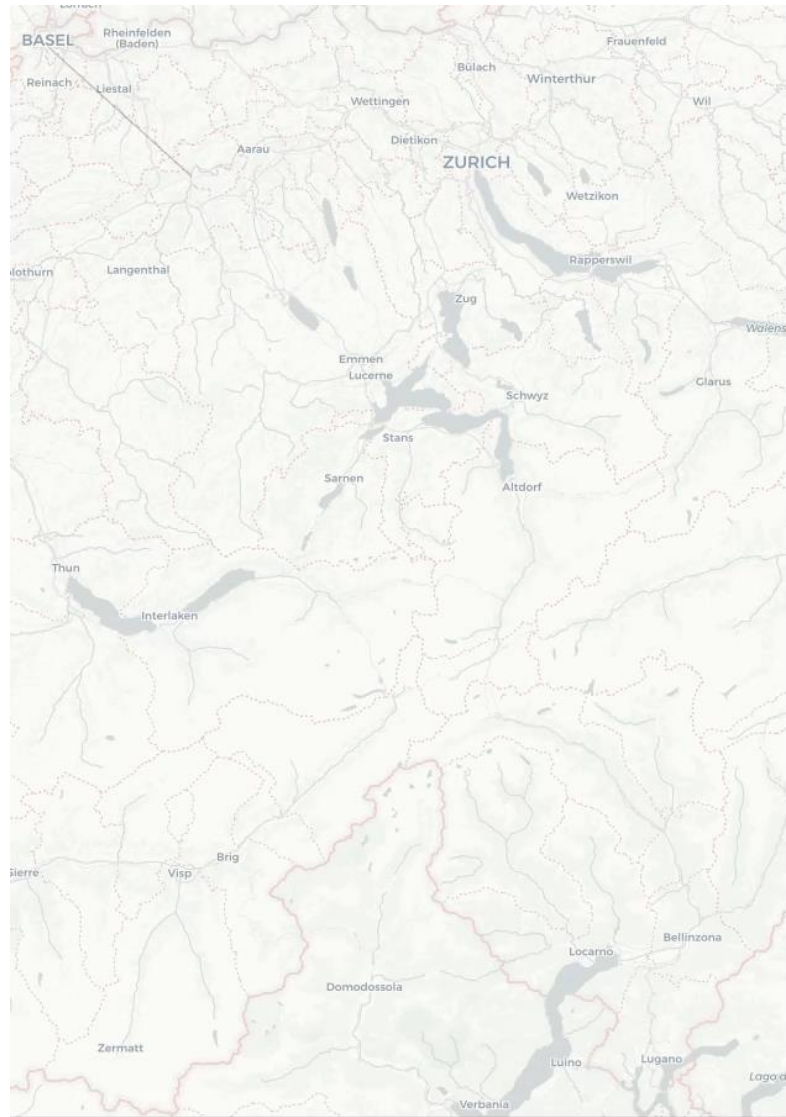


# The Data.

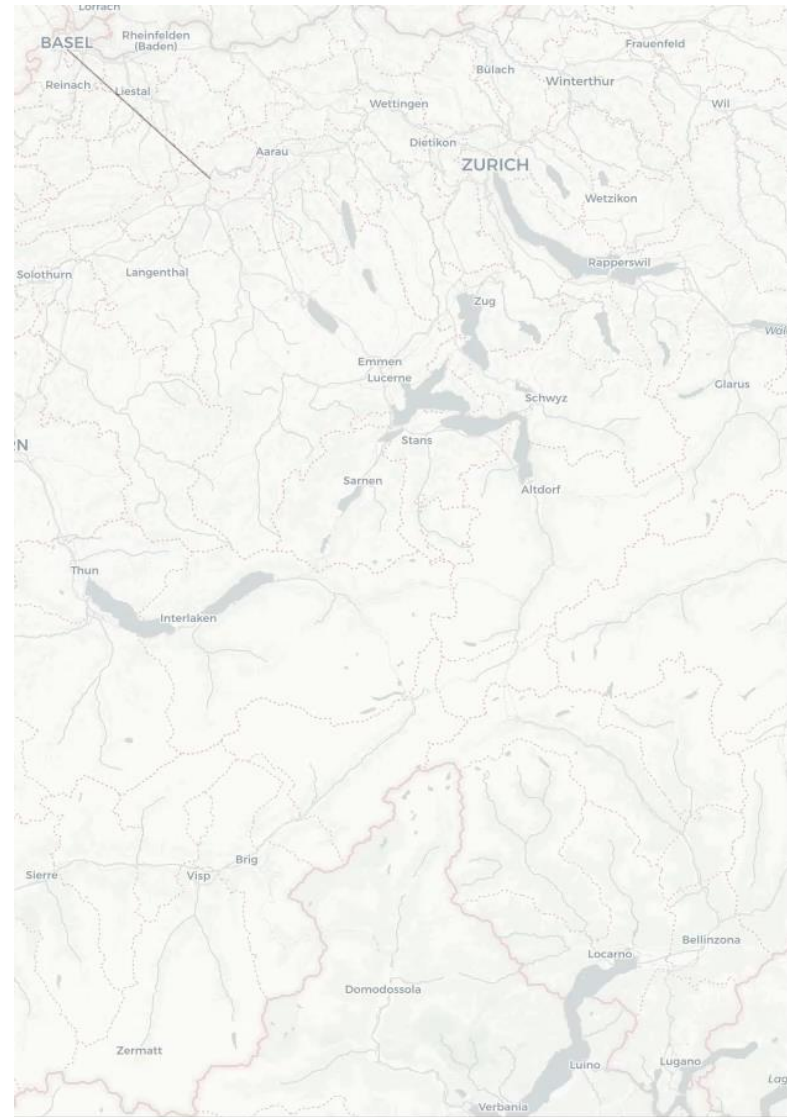
## **What is our «ground truth»?**

Long-distance services: Manual counts by train crew.

Regional services: Sensor data.



2024-05-02 05:24:07



2024-05-09 05:16:46

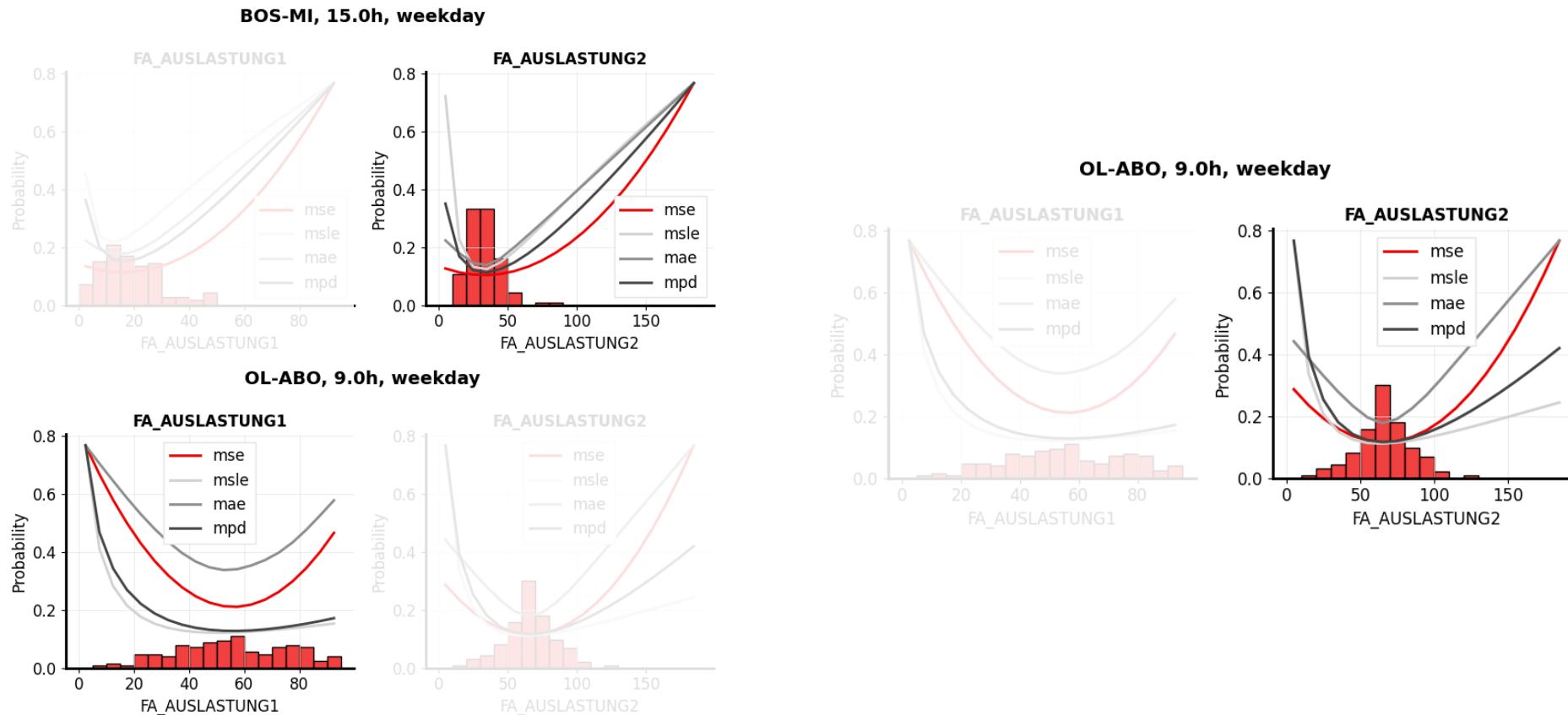


# The Model.



# Choosing our metric.

We measure model quality using the mean poisson deviance, since it best reflects business intuitions of what makes a good model.



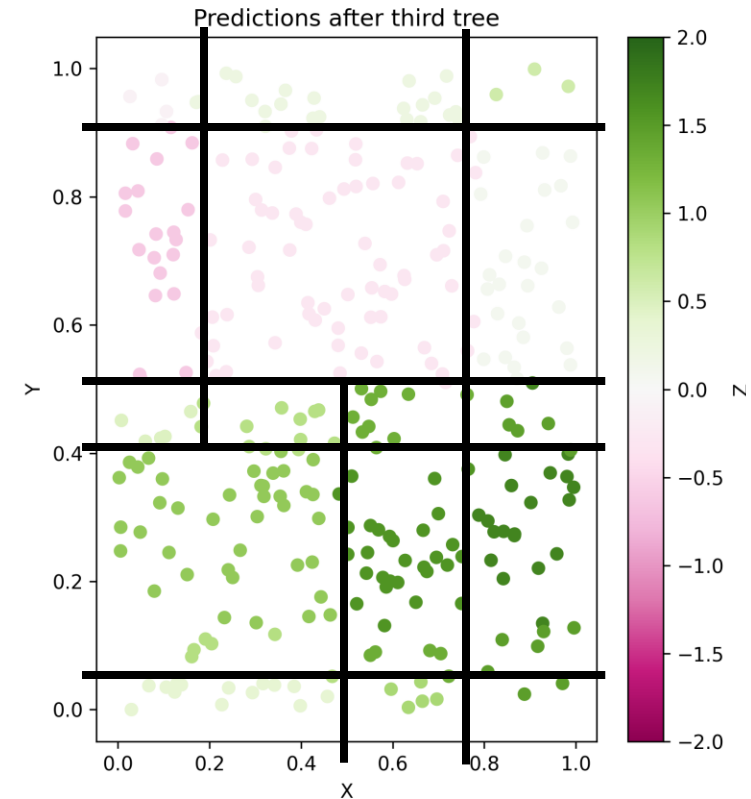
# Model Architecture: LightGBM.

## Why?

- Fast, easy to train, few hyperparameters to tune.
- It is not clear what neural network architecture would be well-suited.

## Challenges:

- We have many categorical features: Sections, public holidays and school vacations.
- Tree-based models are poorly suited for many interactions in high dimensions.



# Embedding our categorical data.

Datum	Hour	Section	2 <sup>nd</sup> class occupancy
Sa, 11.5.24	13	BR-VI	262
Sa, 4.5.24	13	BR-VI	204
Su, 12.5.24	19	SP-TH	590
Mo, 13.5.24	16	TH-SP	710

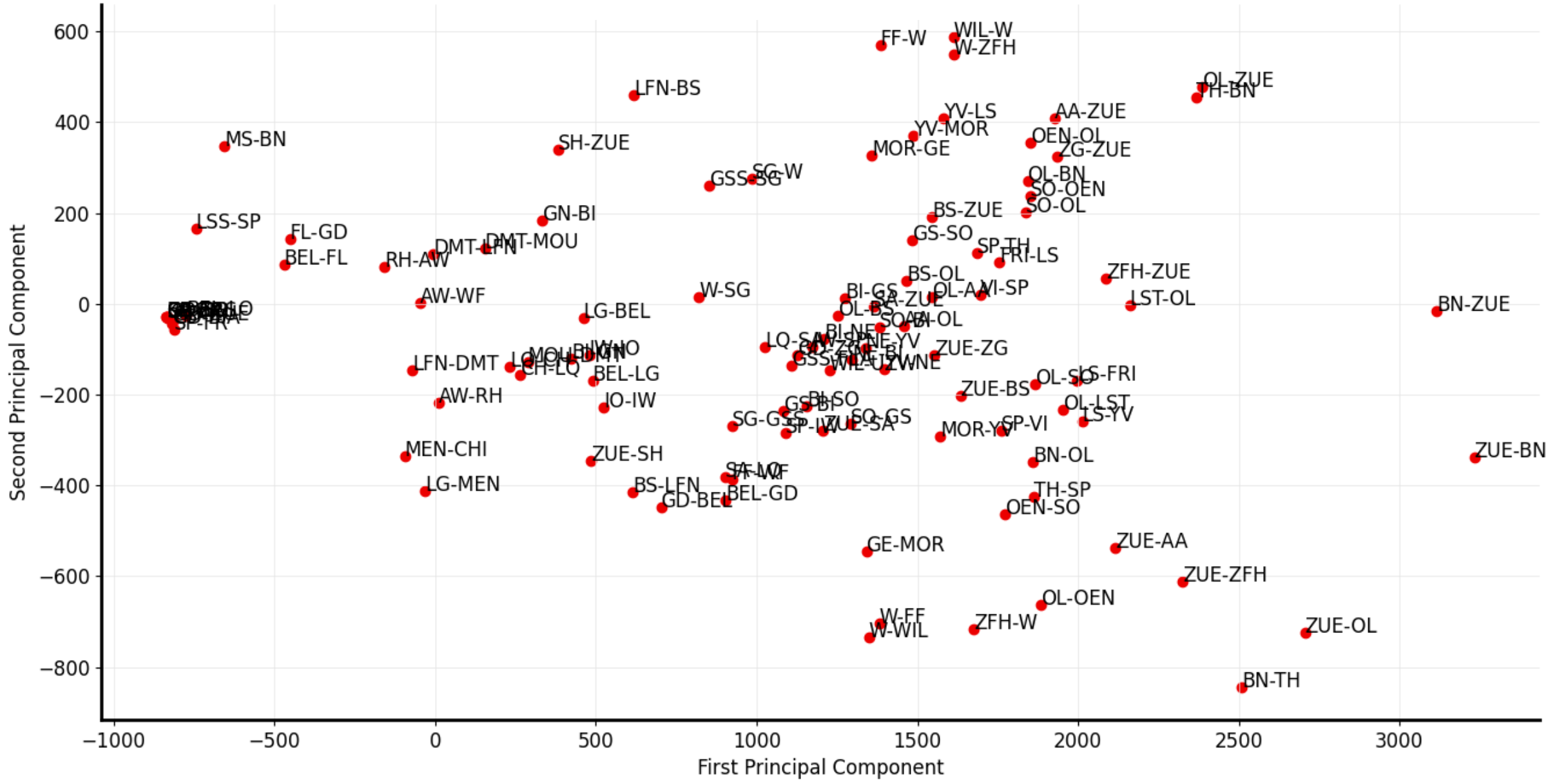


	Sa, 13	So, 17	Mo, 14	Mo, 15	Mo, 16	...
BR-VI	227	724	203	186	582	
VI-SP	293	436	385	461	564	
SP-TH	581	472	418	407	572	

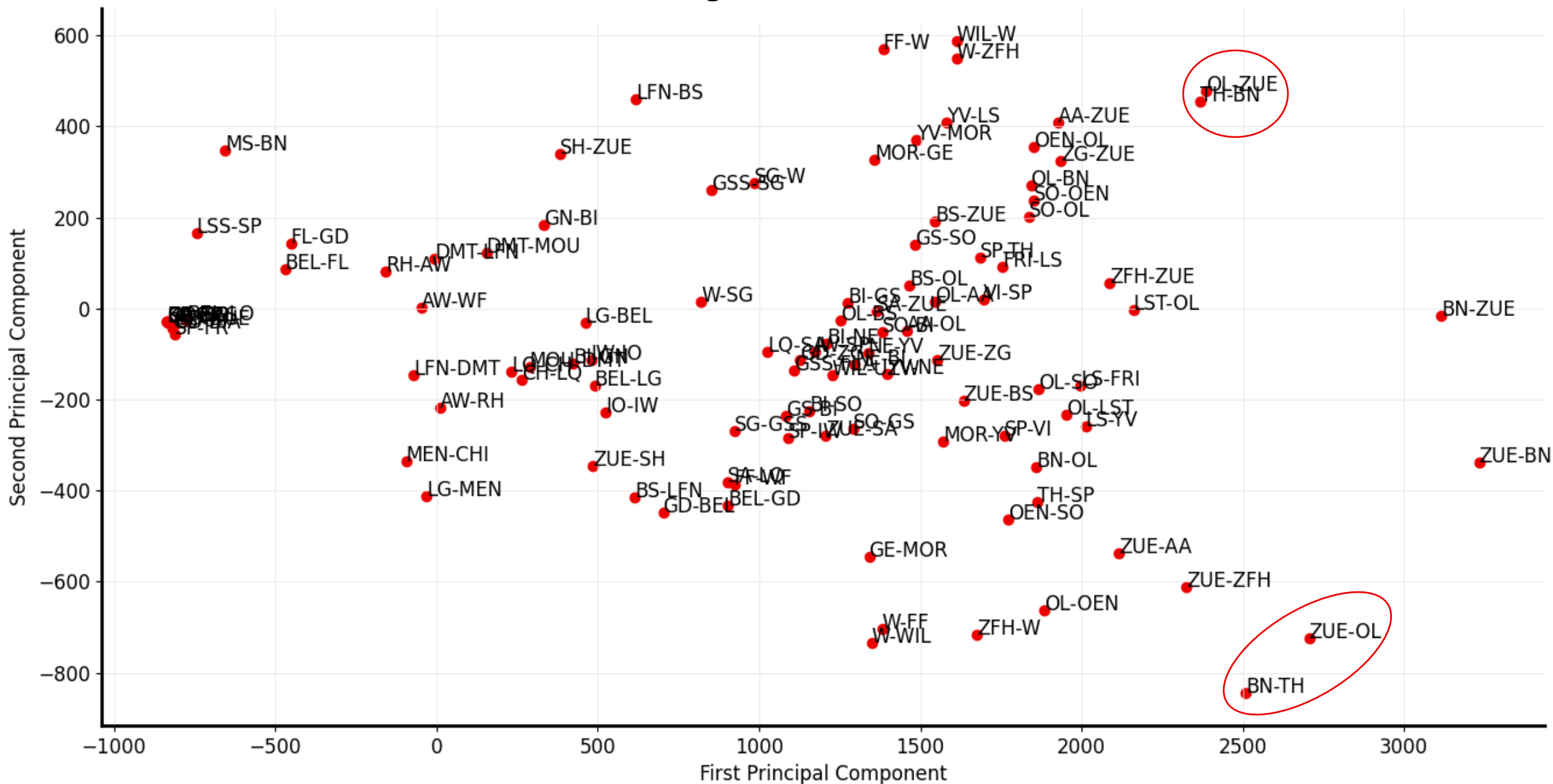


	PC1	PC2	PC3
BR-VI	227	724	203
VI-SP	293	436	385
SP-TH	581	472	418

# Long-Distance Sections



# Long-Distance Sections



# What we predict.



# What we're struggling with: Events



# What we're struggling with: The Weather

Even more interactions: What weather is relevant to which section, on which week-day, during which season?





 Quality assurance.

## Model Probing

Together with business experts, we have defined model probes which have to be done right for the model to be deployed automatically.

### **Examples:**

- On commuter routes, predicted occupancy is higher at 7 than 11 on a week-day.
- On leisure routes, predicted occupancy is higher on Saturday at 9 than Wednesday at 9.

# Model Interpretability with SHAP Partial Dependence Plots

