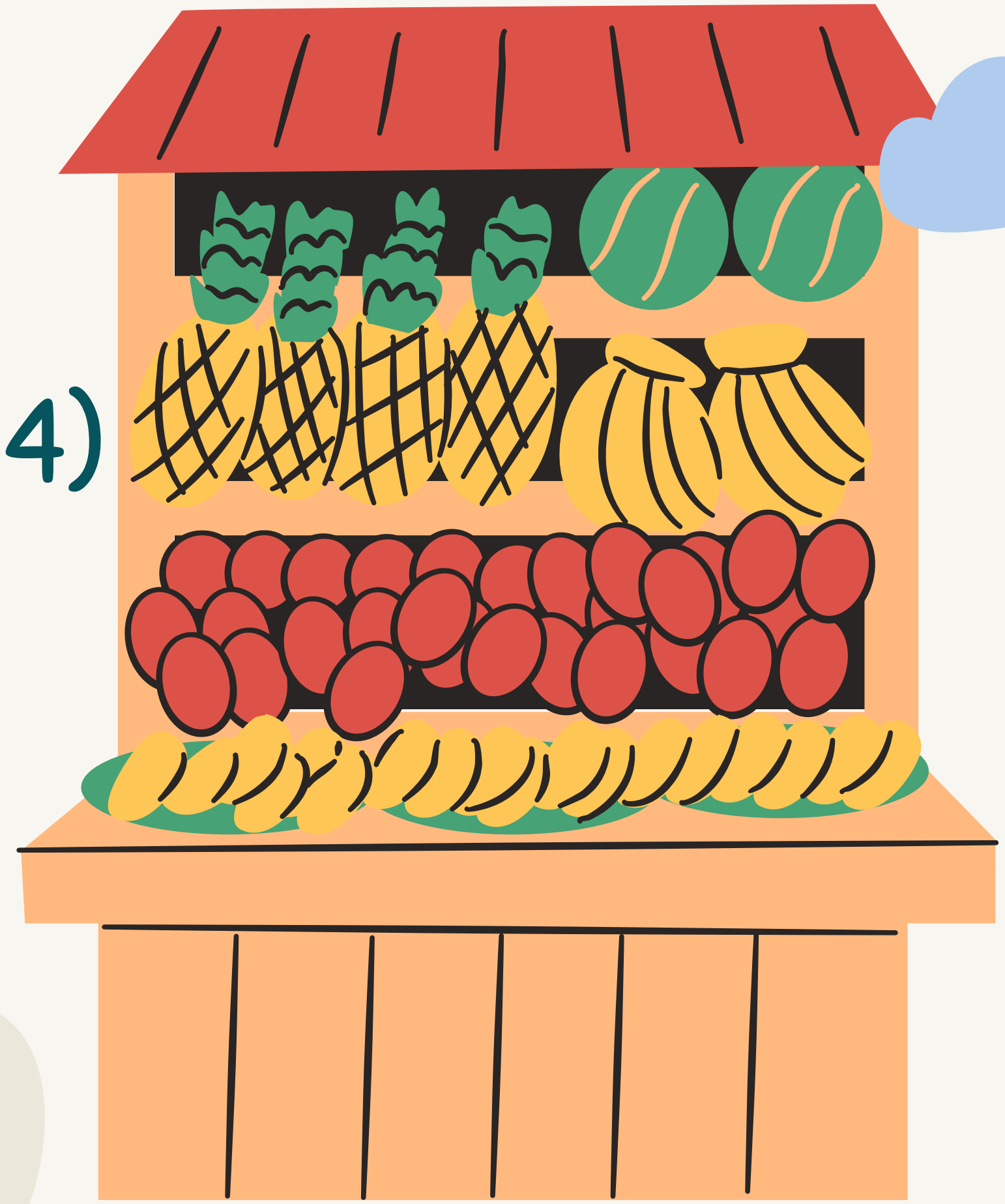


# Favorita – Grocery Sales Forecasting (Guayas, Jan–Mär 2014)

Julia Felgentreu – Data Analyst



# Problem & Zielgruppe

- Zielgruppe: Demand Planner in Guayas
- Use-Case: Tages-Forecasts für Store×Item
- Zeitraum: Train bis Feb 2014, Forecast März 2014
- Outcome: belastbarer Forecast + klare Doku (App optional)

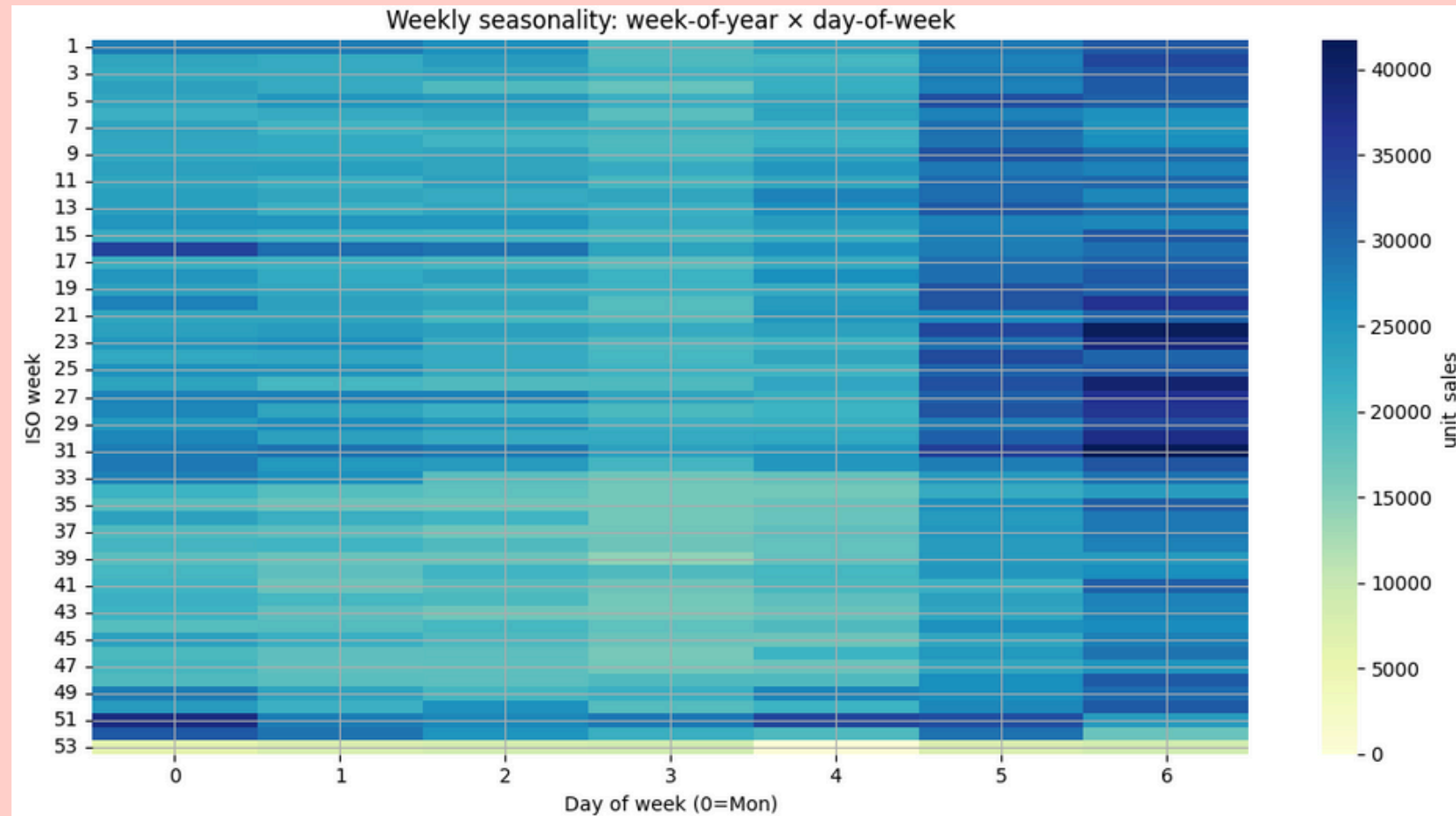


# Datenquellen & Cleaning

- train.csv (Store, Item, unit\_sales, onpromotion)  
→ nur Guayas
- Items/Stores (Meta: family, perishable, type/cluster)
- Transactions & Oil als externe Signale
- Fixes: negative Sales→0, vollständiger Tageskalender, Interpolation Oil, Q1-Export



# Was verraten die Daten?



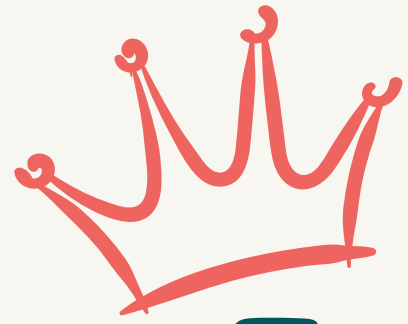
Starker Wochenrhythmus  
(Sa/So > Werktage)

Promo-Lift deutlich (onpromotion ↑)

Oil globaler Trend, schwache/neg.  
Korrelation zu Sales

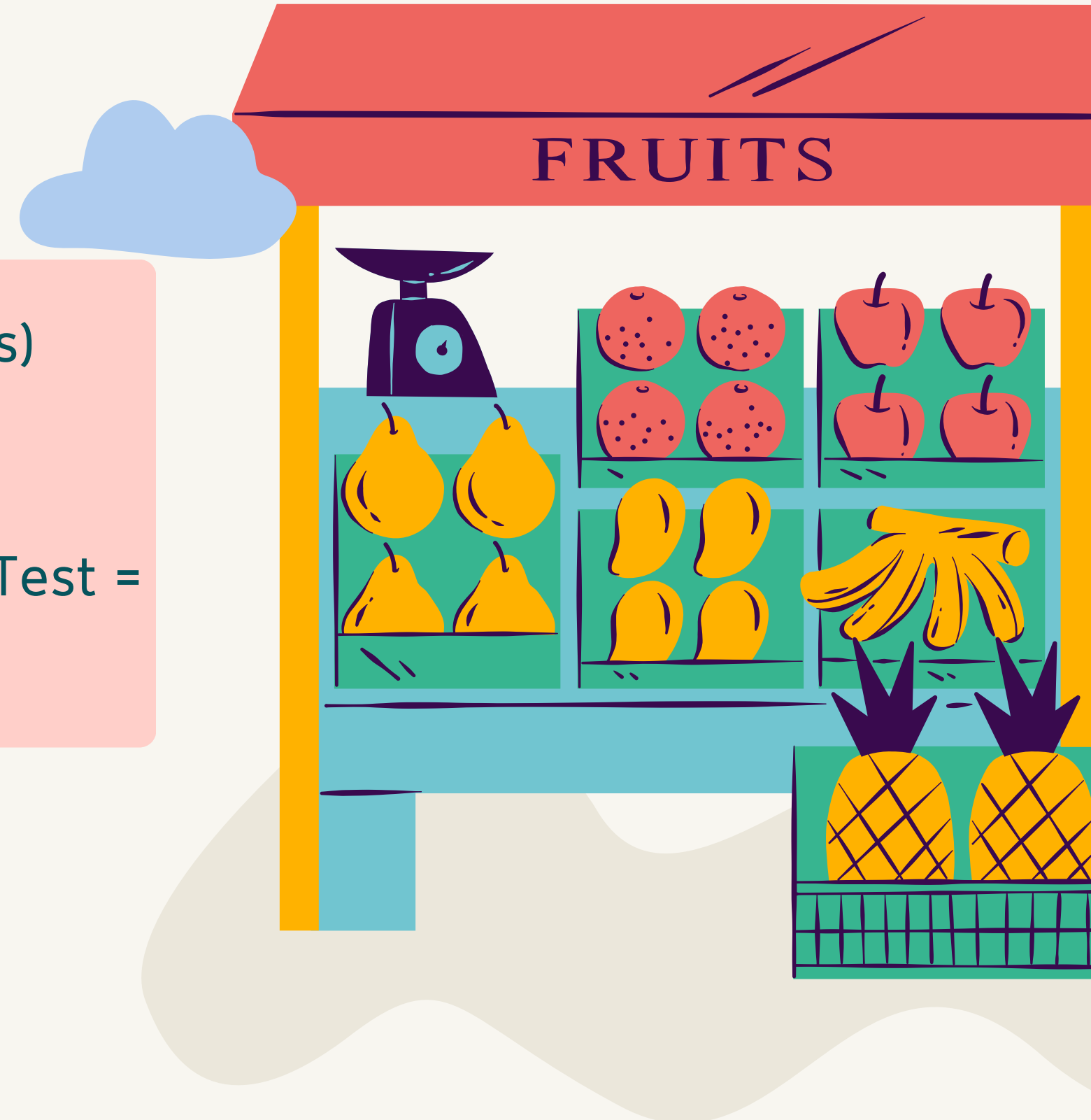
Top-Items/Stores treiben Großteil der Nachfrage





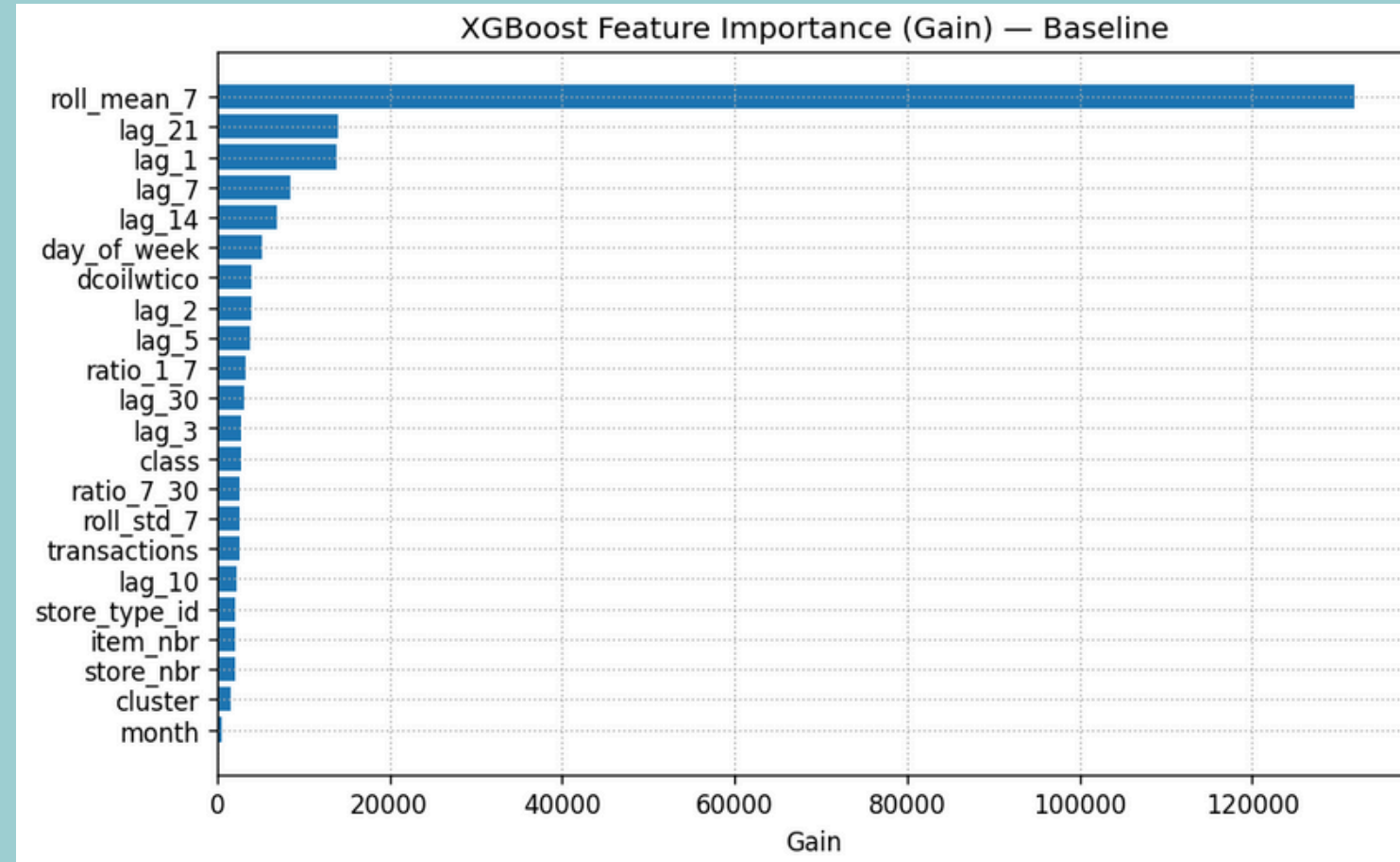
# Feature Engineering & zeitliche Splits

- Lags: 1/7/14/30 (+ roll\_mean/std\_7, ratio-Features)
- Kalender: day\_of\_week, month, is\_weekend
- Exogen: transactions, dcoilwtico, onpromotion
- Splits: Train bis 28. Feb, Val = letzte 7 Feb-Tage, Test = März





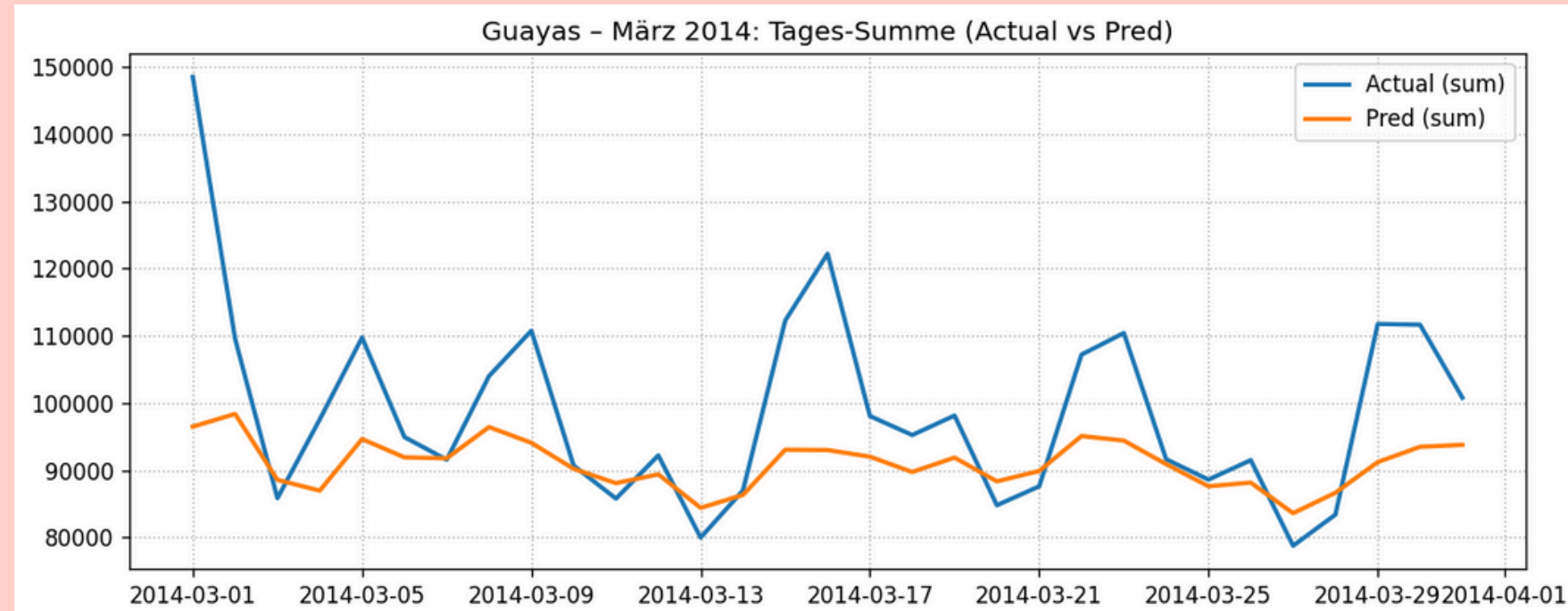
# Warum XGBoost? + Backtesting



- XGBoost (reg:squarederror), Early Stopping
- Schlankes Random-Search + Optuna-Fine-tuning
- Rolling-Origin Backtest (5-fold)
- LSTM nur als Referenz (deutlich schlechter)



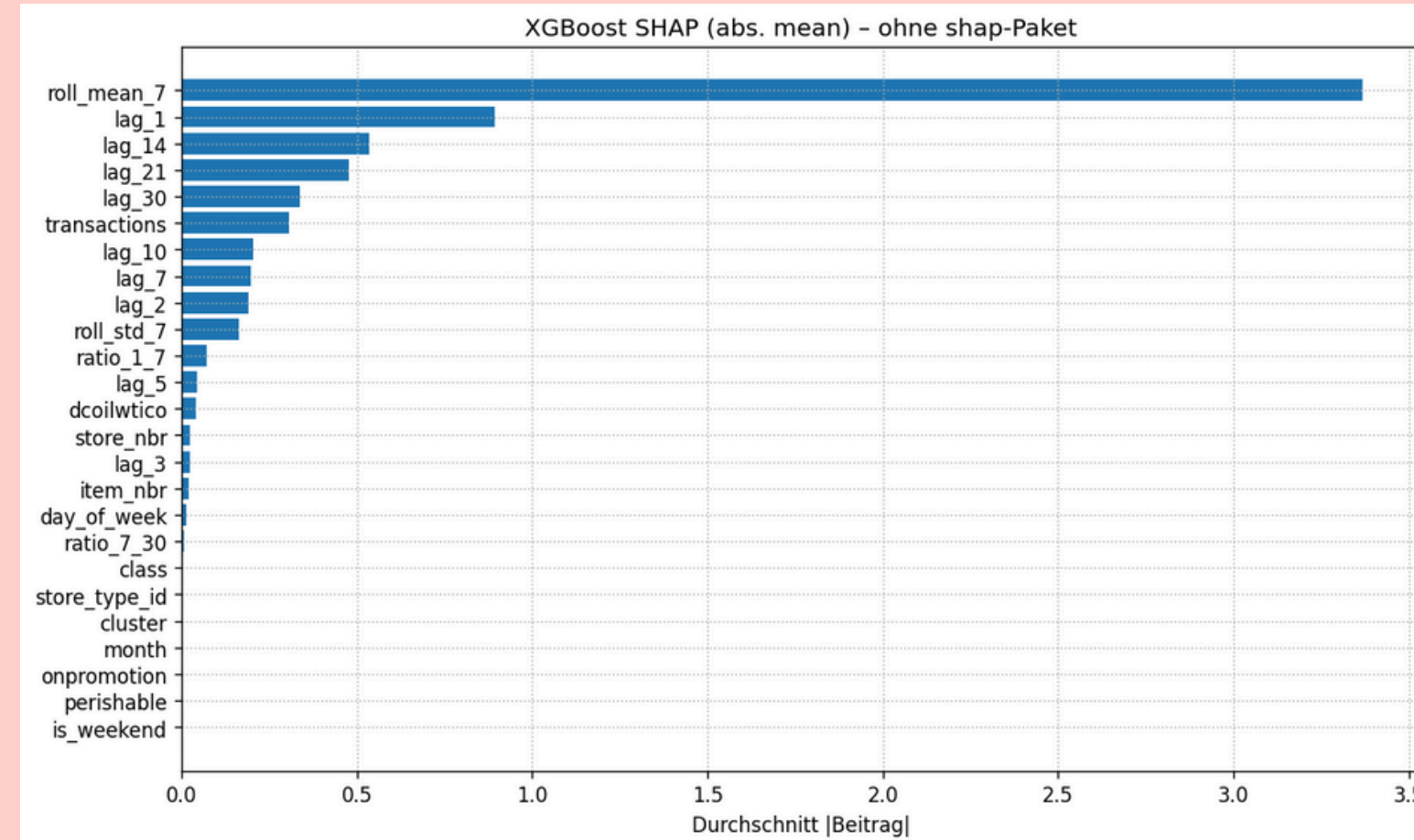
# Forecast-Güte: März 2014



- Trend sauber getroffen; Abweichungen an Promo-/Peak-Tagen
- Bias  $\sim 0$  (-0.1 bis -0.12)



# Explainability (SHAP) & Drift (PSI)







- zu PSI (aus psi\_latest.csv): Oil hoch (globaler Trend), Transactions moderat
- Wichtigste Treiber: roll\_mean\_7, Lag-14/21/30, Transactions, DOW
- PSI: erwartete Verschiebung bei Oil; Transactions variabel, aber im grünen Bereich



# Was funktioniert – was würde ich als Nächstes tun?



-  Saubere Pipeline, stabile XGB-Performance, reproduzierbare Artefakte
-  Peaks/Promos bleiben schwierig → feineres Promo-Modeling, Feiertags-Features
-  SOON Item-Cluster-Modelle / Hierarchische Forecasts (Store→Item)
-  SOON Optionale Streamlit-App als Forecast-Explorer (Upload, Filter, Intervalle)





# Danke! 🙌

- alle Infos und detaillierten Informationen findet ihr auf meinem [GitHub](#)
- Vielen Dank fürs Zuhören!

