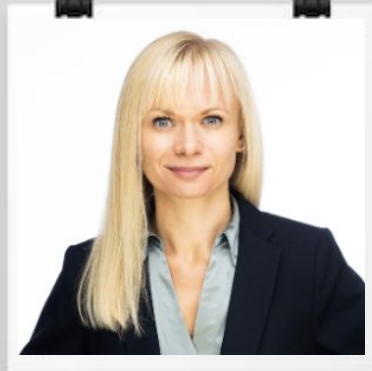

Superstore Sales Prediction

Capstone Project

Retail Riddle Crew

Angelina Neunzig, Klaus Finkbeiner, Ricarda Albers

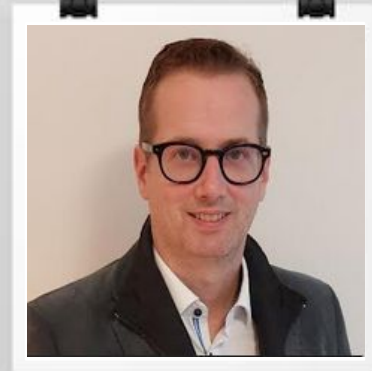
19.09.2023



Ricarda Albers



Angelina Neunzig

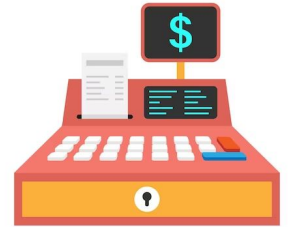
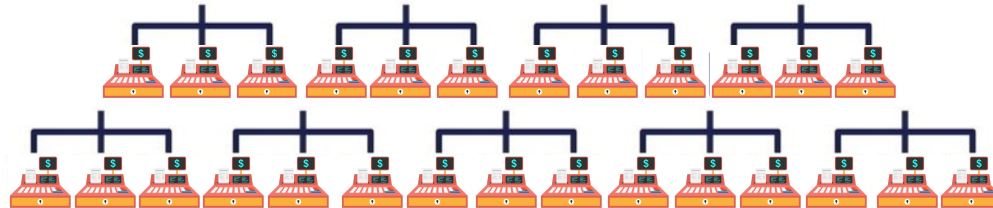
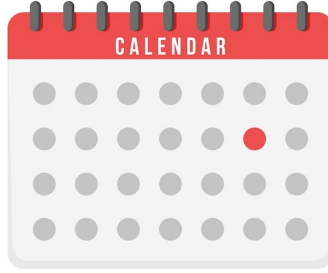


Klaus Finkbeiner











Our Target perspective

Mission:

- Predicting sales more precisely than the current way of predicting (baseline model)

Goals:

- Allow to better manage inventory
- Demand-oriented order management
- Ultimately increase sales and avoid money loss



comes in

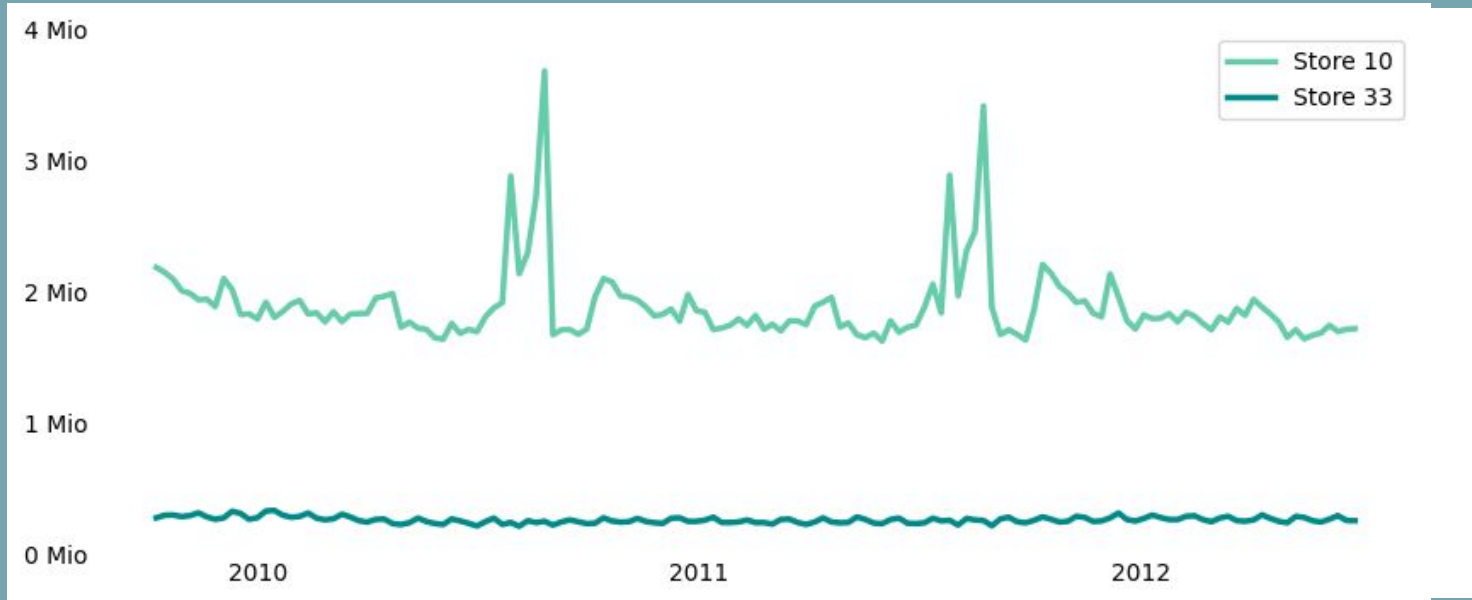
Many outliers

Maximum: 693.099\$

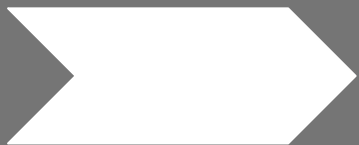
[illegible]

Seasonality

shown by weekly sales in total over time



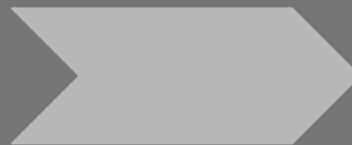
Our journey predicting sales



Deep dive into
data

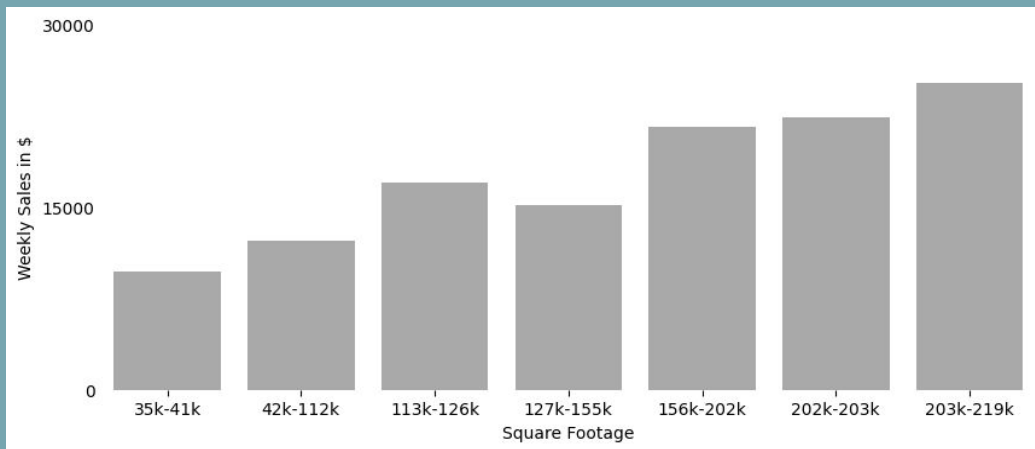


Feature
Engineering



Modelling &
Prediction

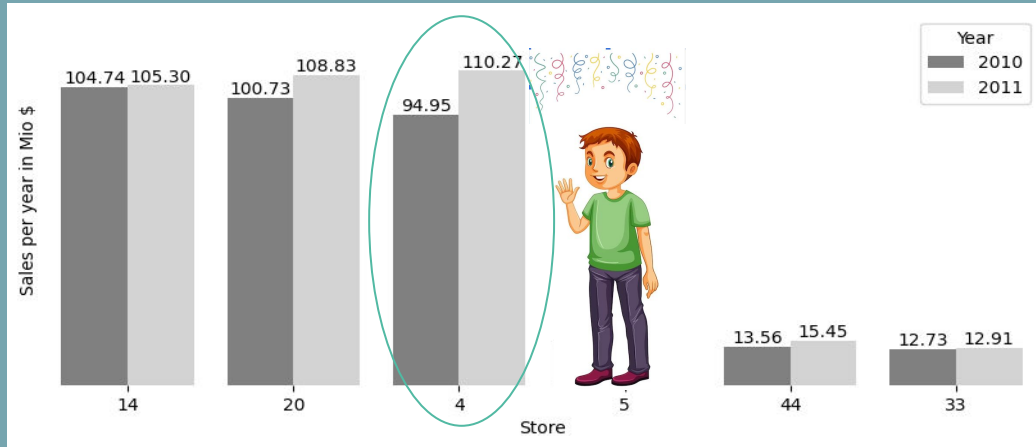
Weekly sales by store size



- Correlation between weekly sales and store size (pearson's r : 0.24)
- No correlation between other features (fuel price, temperature, ...)



Top and bottom stores' sales per year



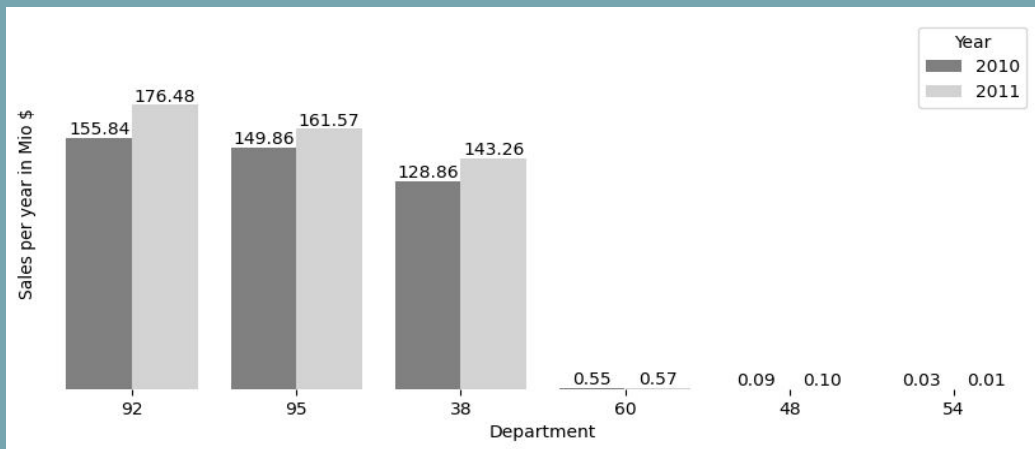
Main explanation:

- Size of stores
- Amount of departments



Top and bottom departments'

sales per year



Top performers:

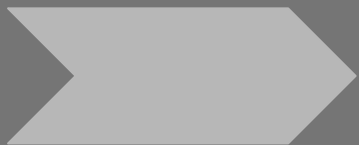
- 92 - Grocery Dry Goods
- 95 - Grocery & Snacks
- 38 - Pharmacy

Bottom performers:

- 48 - Firearms
- 54 - Jewelry
- 60 - Concept Stores and Stamps



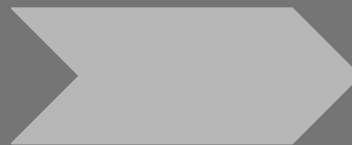
Our journey predicting sales



Deep dive into
data



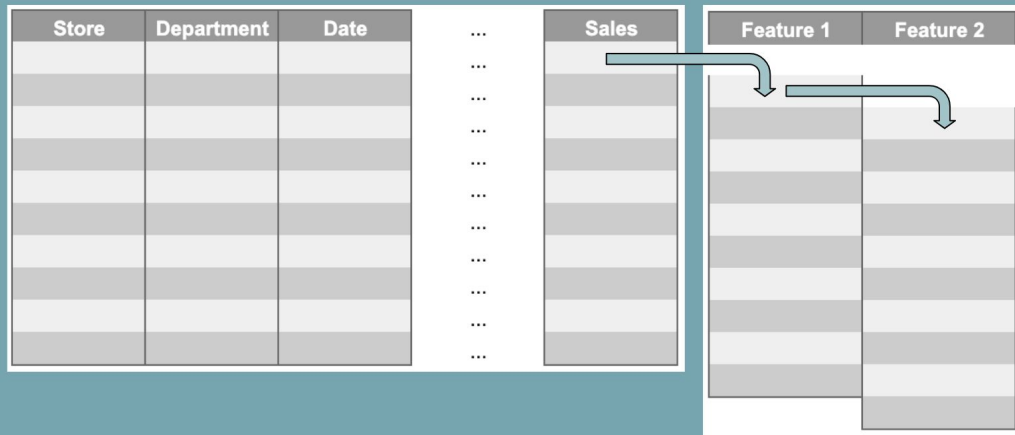
Feature
Engineering



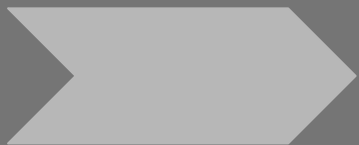
Modelling &
Prediction

Feature Engineering

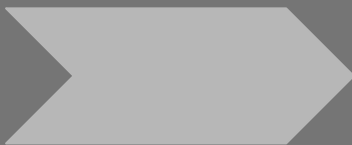
- Introduce new features to improve predictions
 - Sliding window technique for time series
- Introduce new features to consider seasonal patterns
 - Christmas
 - Thanksgiving



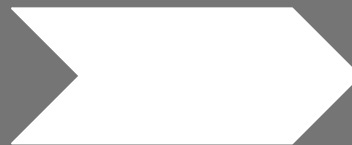
Our journey predicting sales



Deep dive into
data



Feature
Engineering



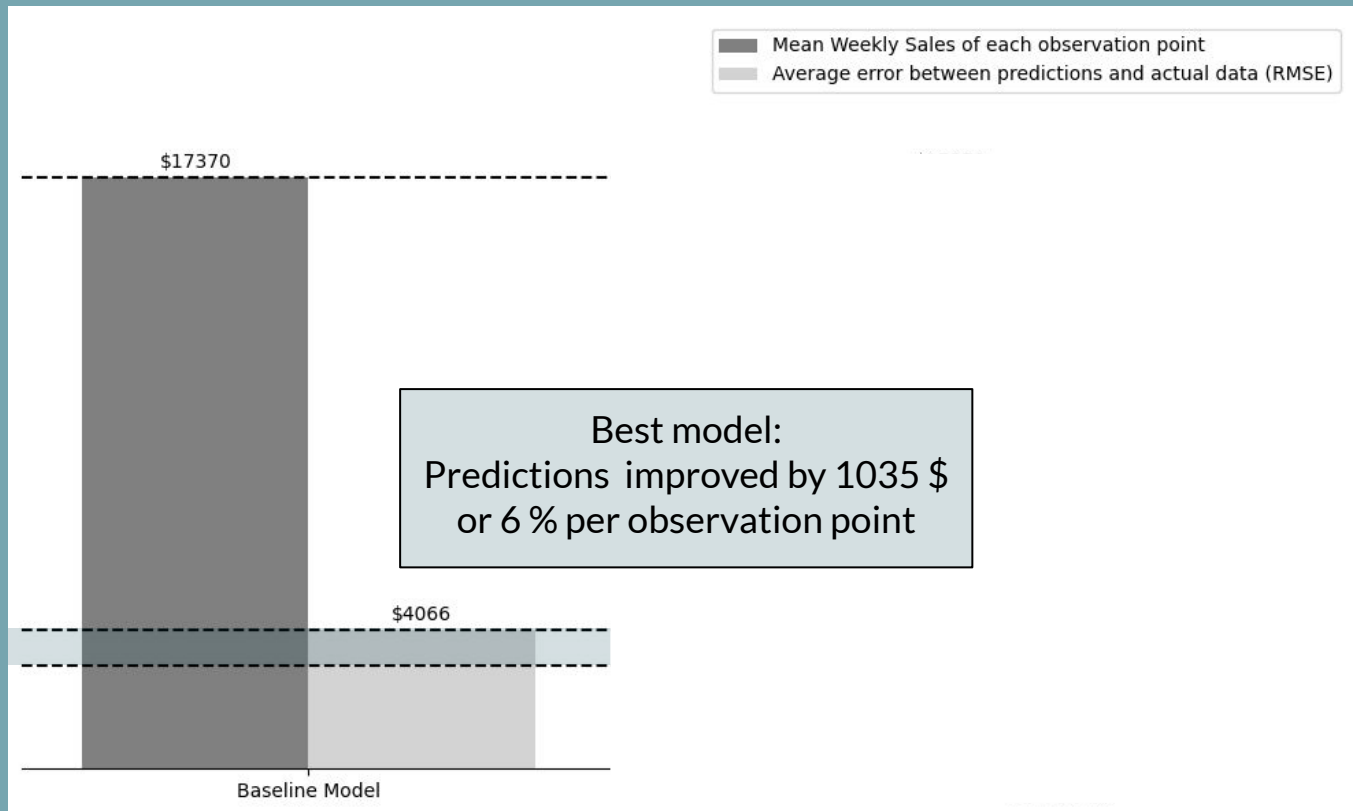
Modelling &
Prediction

Modelling

- Baseline model: consider what happened last year
- Evaluation metric: average error between predictions and actual data (RMSE)
- Improved modelling: by using different algorithms

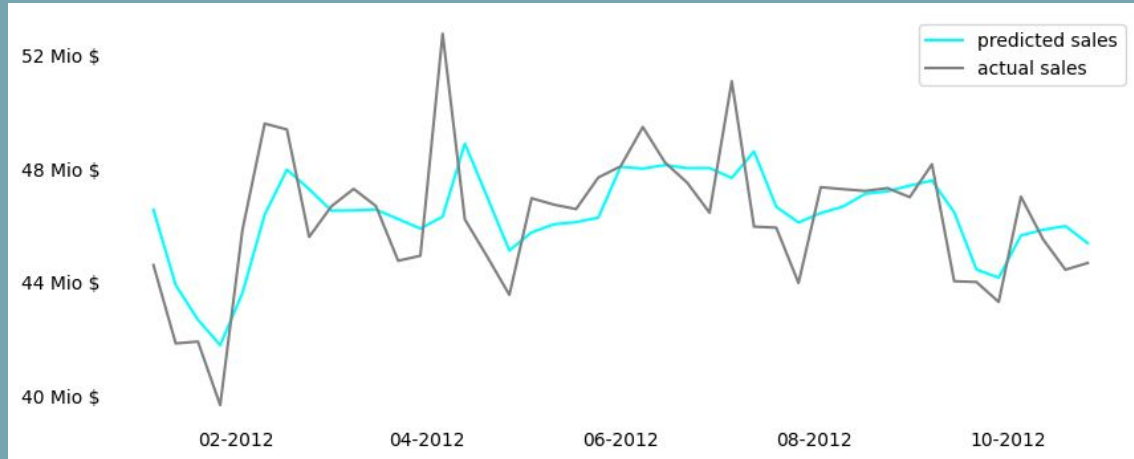


Best model (Extra Trees) compared to baseline model



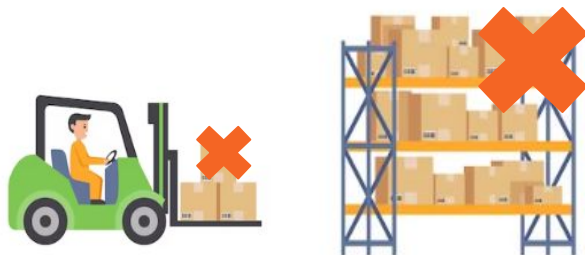
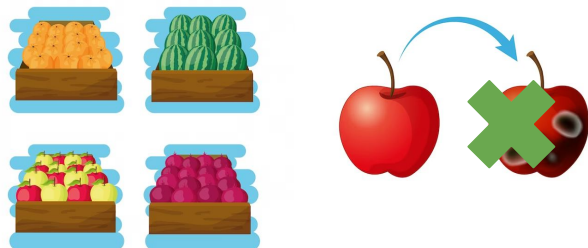
Prediction during test period

- Cumulative weekly sales during test period: ~ 2 Billion \$
- Error in predictions with baseline model: ~ 465 Million \$
- Error in predictions with best model: ~ 347 Million \$



Enhancement in test period predictions by \$118 Million.





Future Work

- Feature engineering to deal better with seasonal patterns
- Applying time series algorithm
- Filter by stores with outlying patterns and model separately



Thank you for your attention

Our journey predicting sales



Knowing our
data



Exploratory
Data Analysis



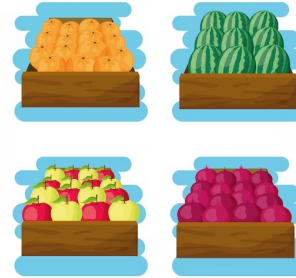
Feature
Engineering

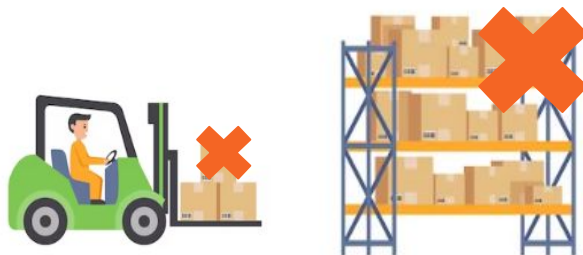
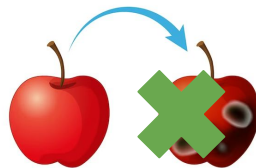
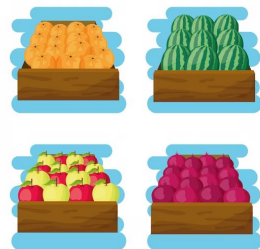


Modelling

Goals achieved

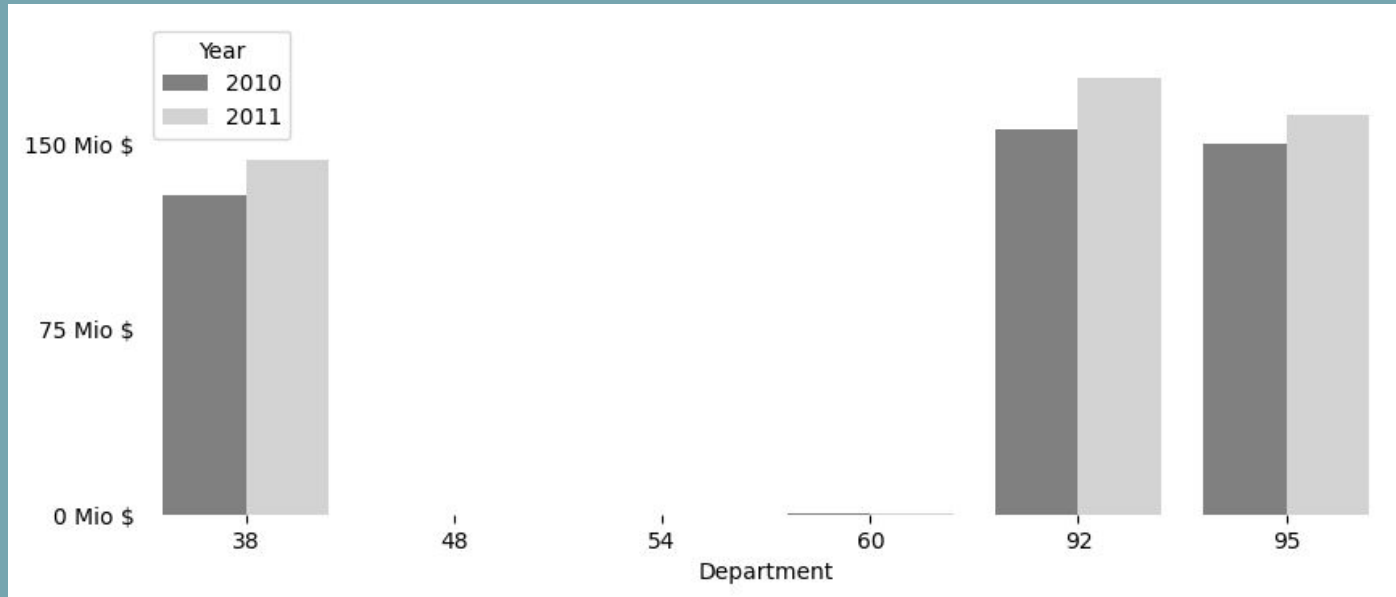
- Better planning of inventory
- Decrease of losses during the storage of perishable foods
- Reduction of the storage of currently not needed goods

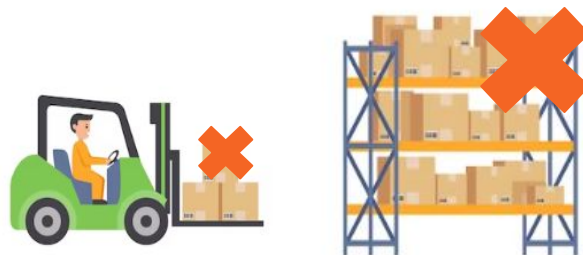
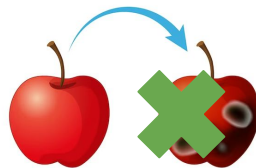
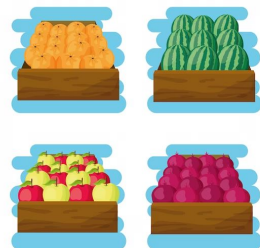




Total sales

per year by department
top 3 and bottom 3





Model comparison

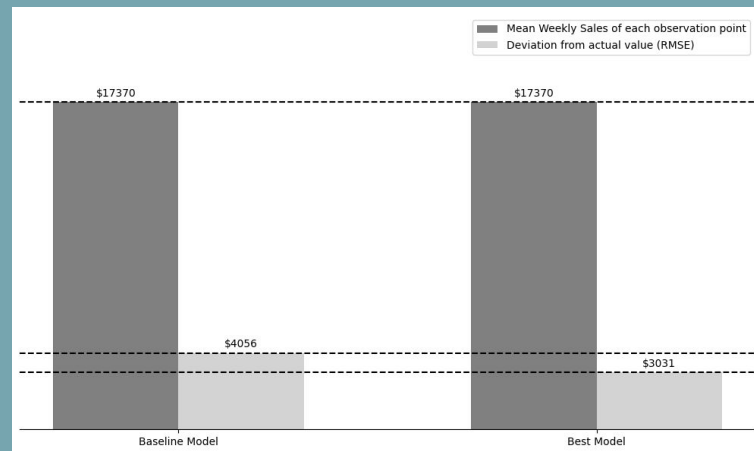
RMSE	Baseline model	Extra Tree	Mean	Standard deviation
train		799,85	17.639,94	23.533,86
validation	4.172,0	1703,86	17.342,50	22.893,48
Test	4.056	3.031	17.370	22.530,90

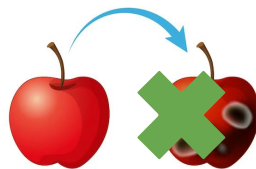
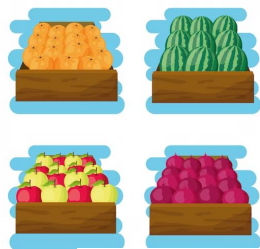


we can improve our prediction by about XXXUSD for the validation data, while an improvement of about XXXUSD was achieved for the test period. This means an improvement of the prediction in

Comparison of Baseline Model and best model comparison

Test period from Jan 2012 to Okt 2012	Baseline model	Best model Extra Tree	Mean Sales per week
RMSE	4.056 \$	3.031 \$	17.370 \$





Goals achieved

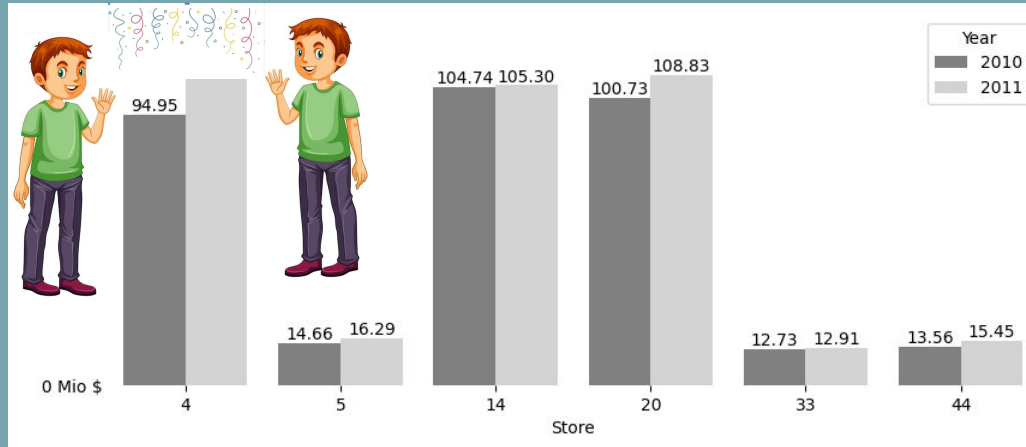
- Better planning of inventory
- Decrease of losses during the storage of perishable foods
- Reduction of the storage of currently not needed goods





Total sales

per year by store
top 3 and bottom 3



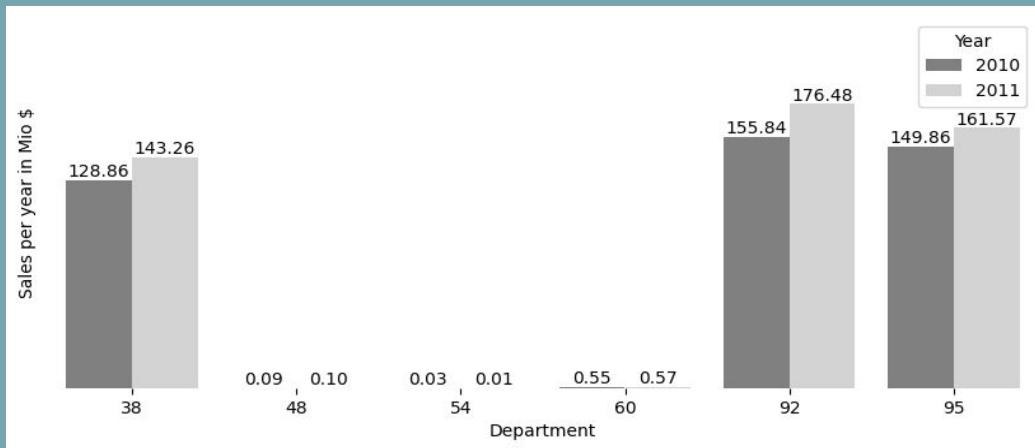
Main explanation:

- Size of stores
- Amount of departments



Top 3 and bottom 3 total sales per year by department

Unsorted,
with
decimals



Top performers:

- 92 - Grocery Dry Goods
- 95 - DSD Grocery & Snacks
- 38 - Pharmacy

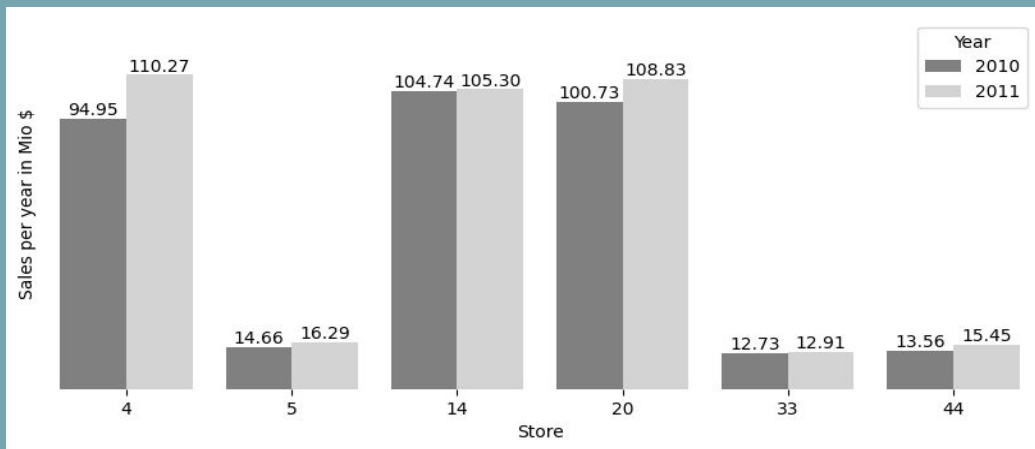
Bottom performers:

- 48 - Firearms
- 54 - Jewelry
- 60 - Concept Stores and Stamps



Top 3 and bottom 3 total sales per year by store

Unsorted,
with
decimals

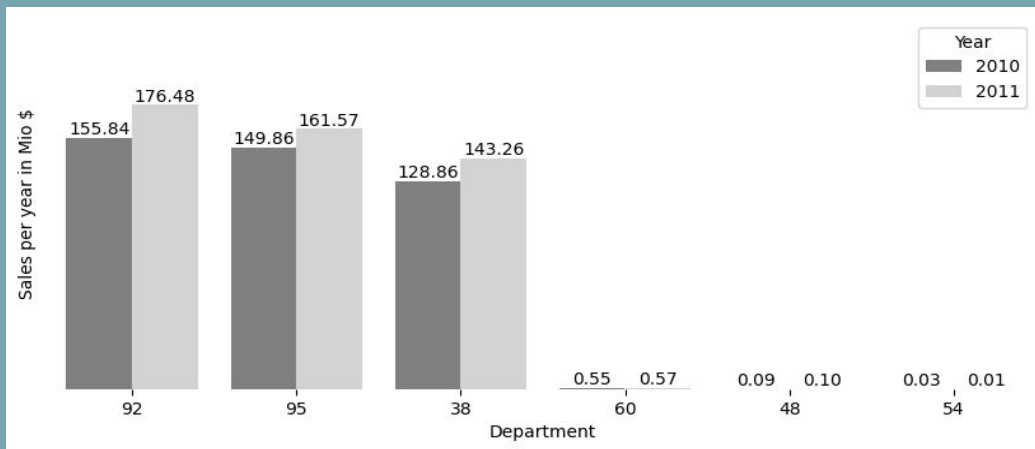


Main explanation:

- Size of stores
- Amount of departments



Top 3 and bottom 3 total sales per year by department



Top performers:

- 92 - Grocery Dry Goods
- 95 - DSD Grocery & Snacks
- 38 - Pharmacy

Bottom performers:

- 48 - Firearms
- 54 - Jewelry
- 60 - Concept Stores and Stamps



Model:
Extra Trees

Evaluation
Metric:
RMSE

Result:
6 % better
prediction
of weekly
sales

→
112 Mio \$

Superstore Sales Prediction

