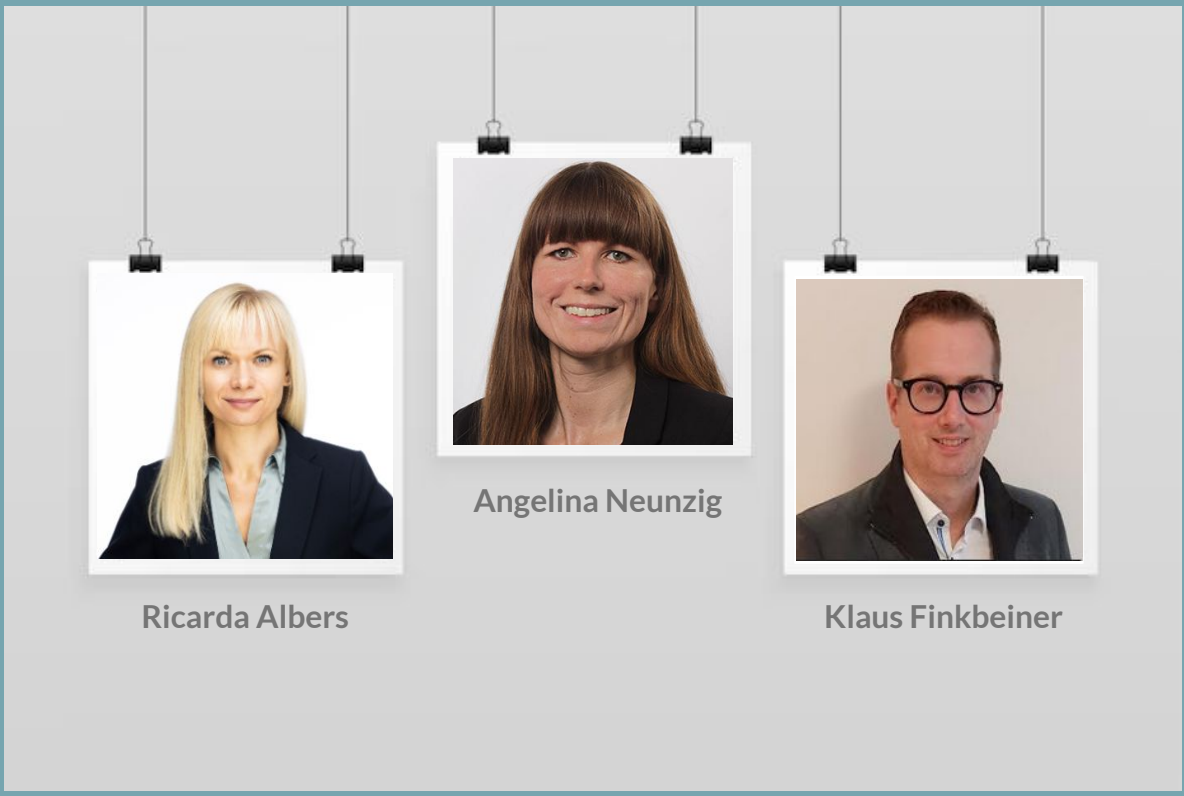
Superstore Sales Prediction

Capstone Project

Retail Riddle Crew
Klaus Finkbeiner, Ricarda Albers, Angelina Neunzig

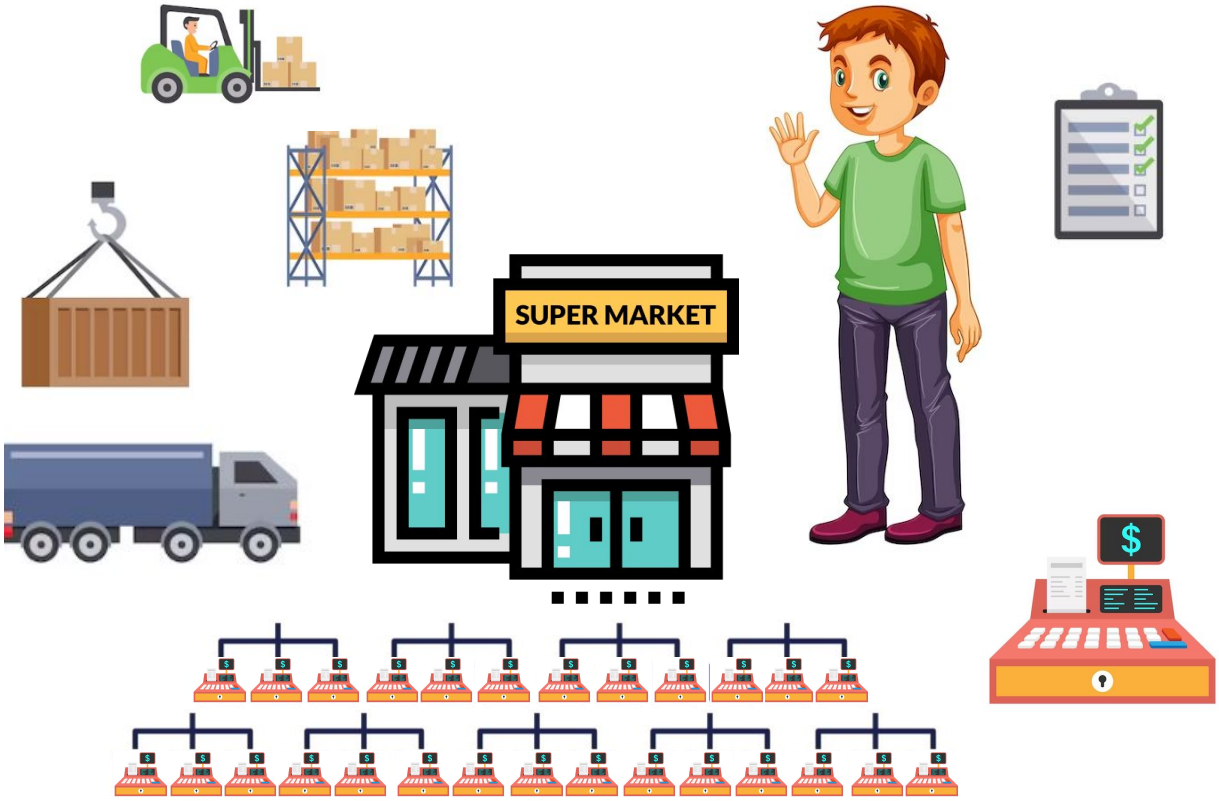
19.09.2023



This is Tom. Tom is managing a large superstore. This superstore has different departments. And there are quite a lot of them. In fact 81.

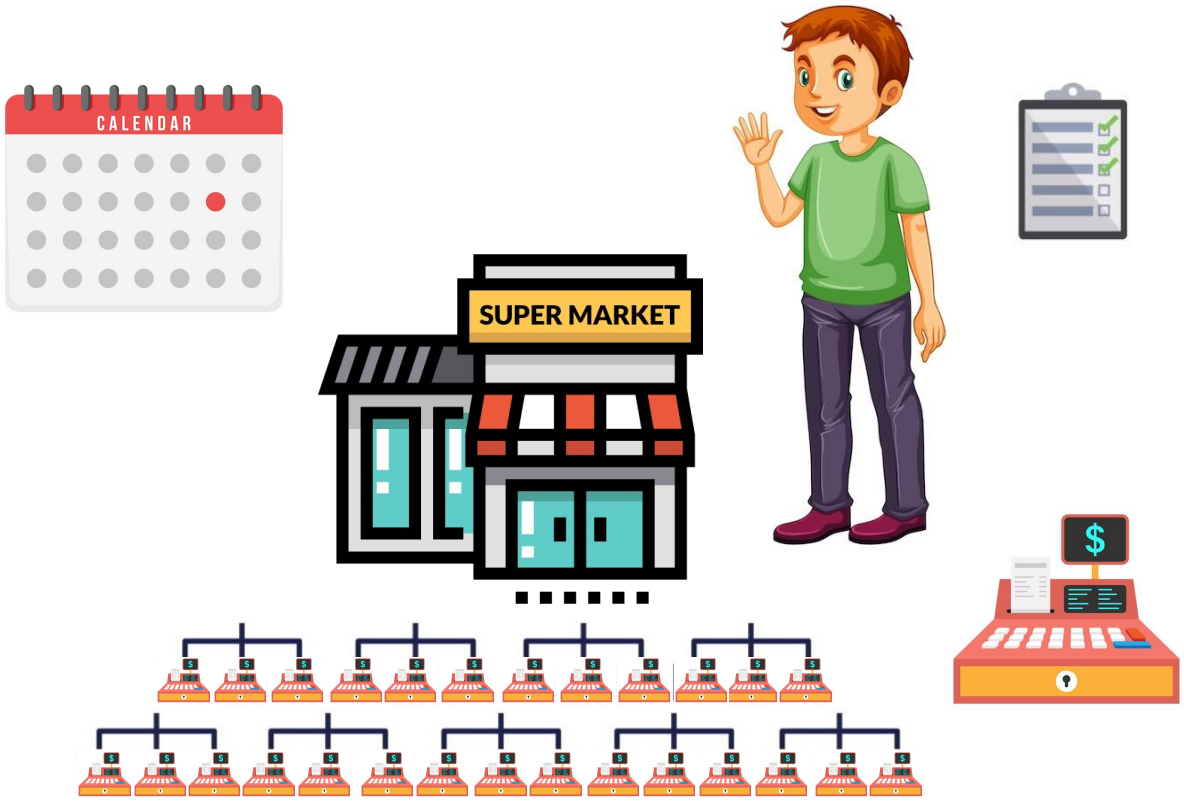
Speaker notes were added on the left side for the data storytelling part.
When presenting, the images of this part were animated.

Tom has many tasks. He has to order goods, keep stock and also take care of the accounting - for each department.



What Tom takes depends on various factors. For example the size of his store. But also on external factors, such as weather, big events or holidays like christmas.

To be able to plan his tasks better, it would help him to know his next week's sales more precisely. Currently he looks at the same weeks sale from the previous two years.



And ...
Tom is not alone. There are 44 other managers.

This is where we come into the game.

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

Data comes in

10 Features

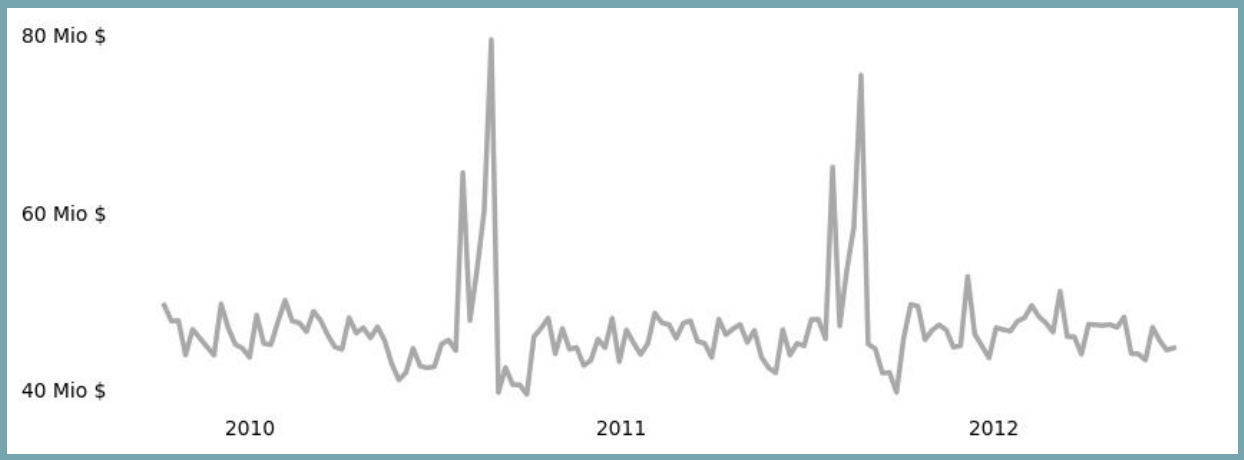
Mean: 17.639 $
Many outliers
Maximum: 693.099$

Feb 2010 - Oct 2012

Store	Department	Date	Storesize	Weather	...	Sales
					...	
					...	
					...	
					...	
					...	
					...	
					...	
					...	
					...	
					...	
					...	
					...	

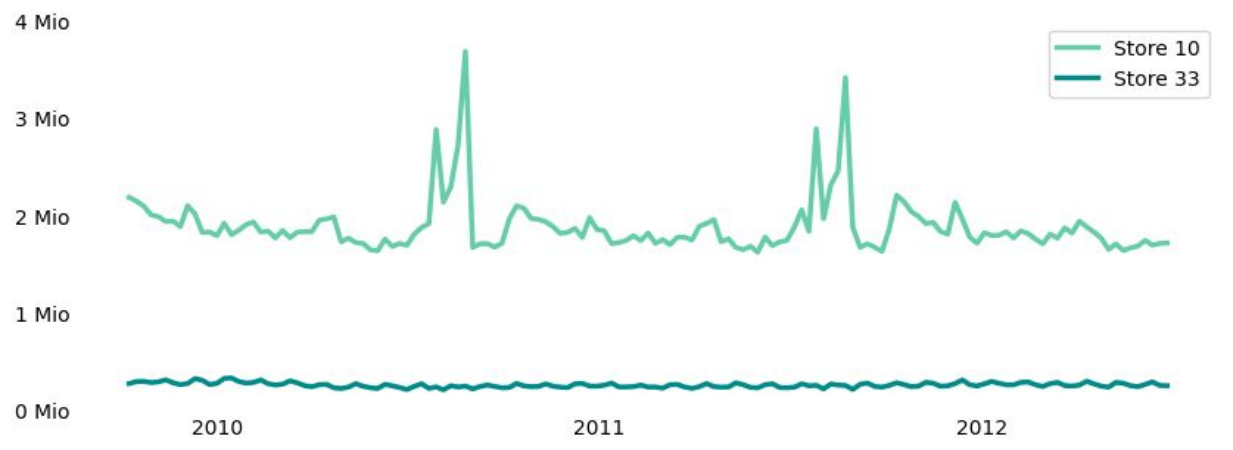
420.000 Obser-vations

Seasonality shown by weekly sales in total over time



The data show a clear seasonal pattern. There are 4 immense peaks that can be attributed to Christmas and Thanksgiving.

Seasonality extreme patterns over time



Not all stores show this seasonal pattern over time. This fact is a challenge for predictive modeling.

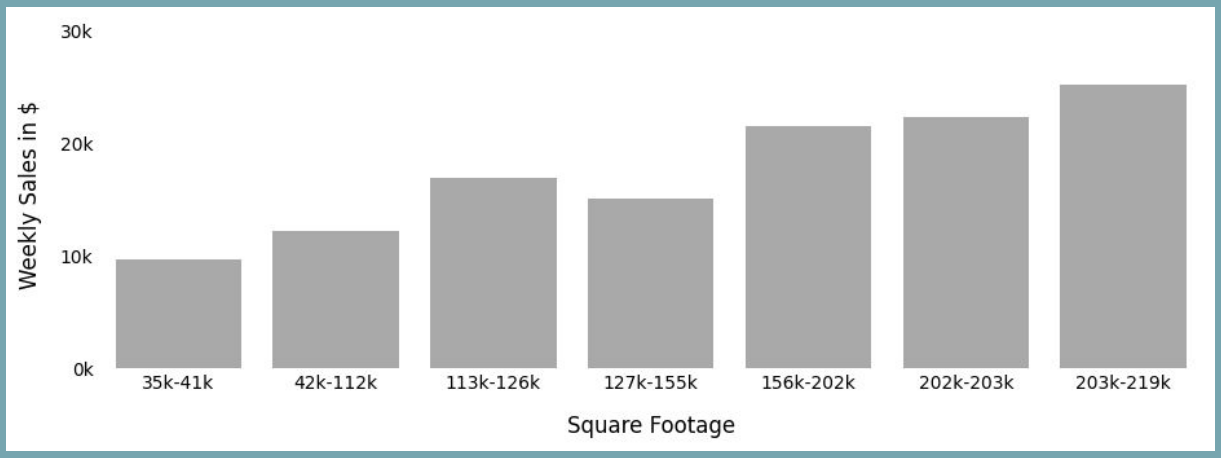
Our journey predicting sales

Deep dive into data

Feature Engineering

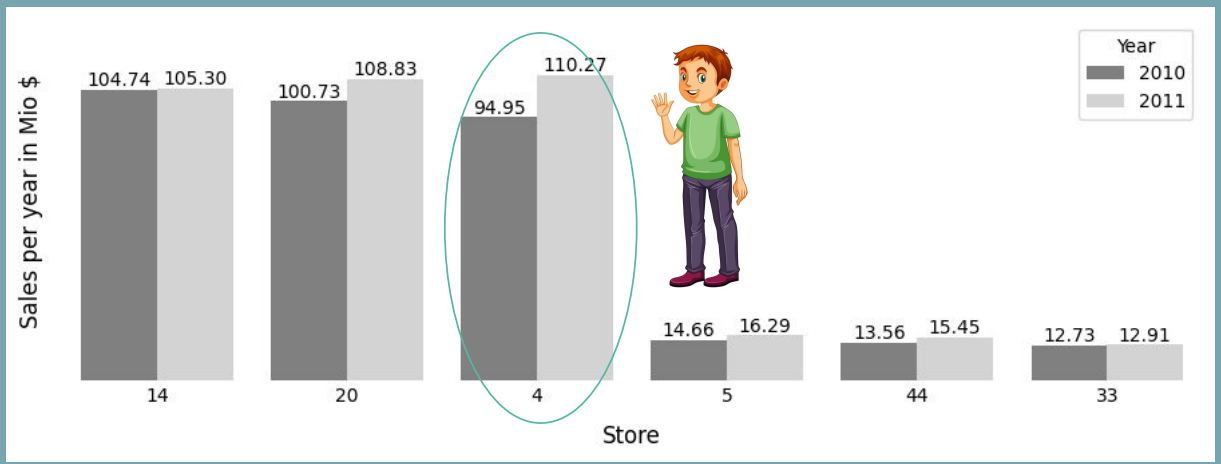
Modelling & Prediction

Weekly sales by store size



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

Top and bottom stores' sales per year

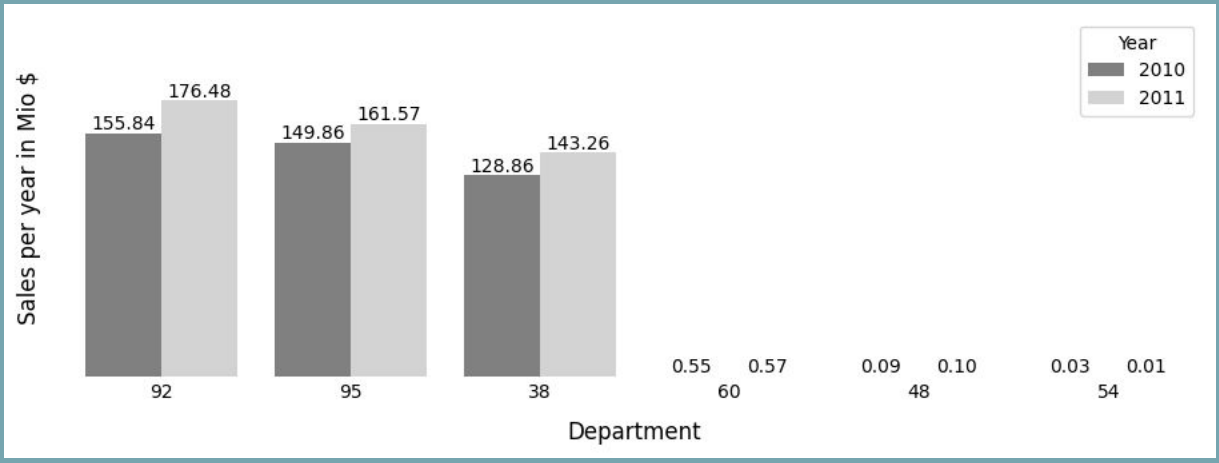


Main explanation of the large differences:

- 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 & Stamps

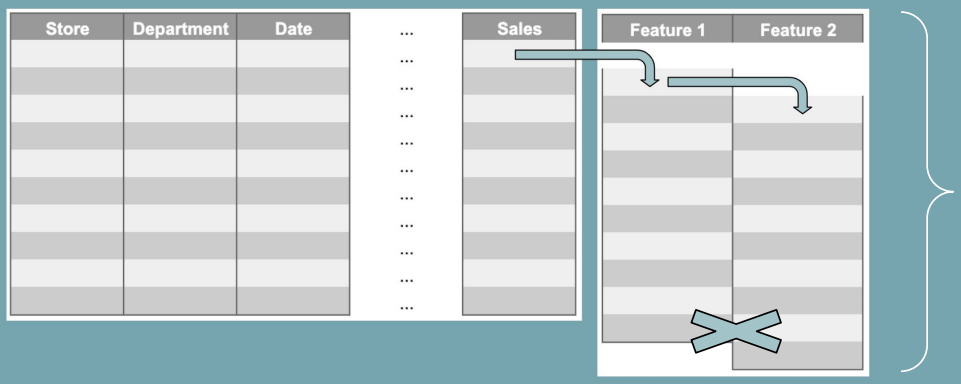
Our journey predicting sales

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 column
- Thanksgiving column

Introduce time related features:

- Year column
- Month column

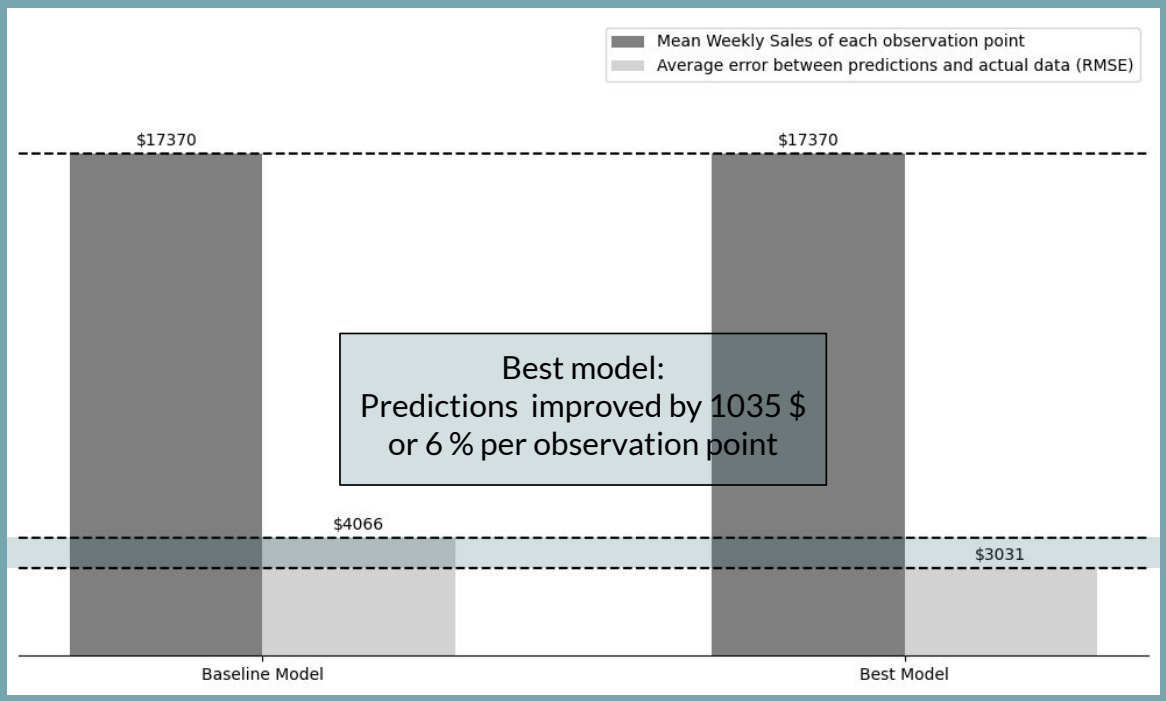
Our journey predicting sales

Deep dive into data → Feature Engineering → **Modelling & Prediction**

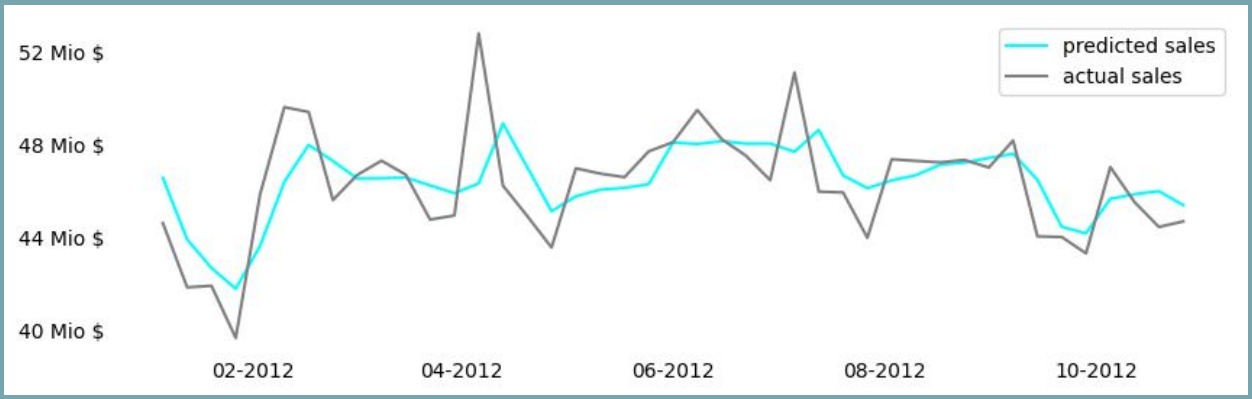
Modelling

- **Baseline model:** Consider what happened last year
- **Evaluation metric:** RMSE (Root Mean Squared Error)
  - Error can be expressed in $
  - Punishes bigger mistakes more
- **Improved modelling:** Using different algorithms, mainly
  - Extra Trees
  - Random Forest

Best model (Extra Trees) compared to baseline model



Prediction during test period



- 2 billion $ cumulative weekly sales
- Error with baseline model: 465 million $
- Error with best model: 347 million $

Enhancement in test period predictions by 118 million $

On average Tom now knows the next week's sales 155.000 Dollar more precisely.
Now he can better order perishable goods like fruit and vegetables on demand. So less spoils and he saves money.
He can also use his warehouse more efficiently by reducing the storage of currently not needed goods. This also saves money, because storage costs equipment, space and staff.
Tom can furthermore sell more continuously because his products don't sell out so quickly.
Using our model Tom is able to achieve the intended goals and he saves money.

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