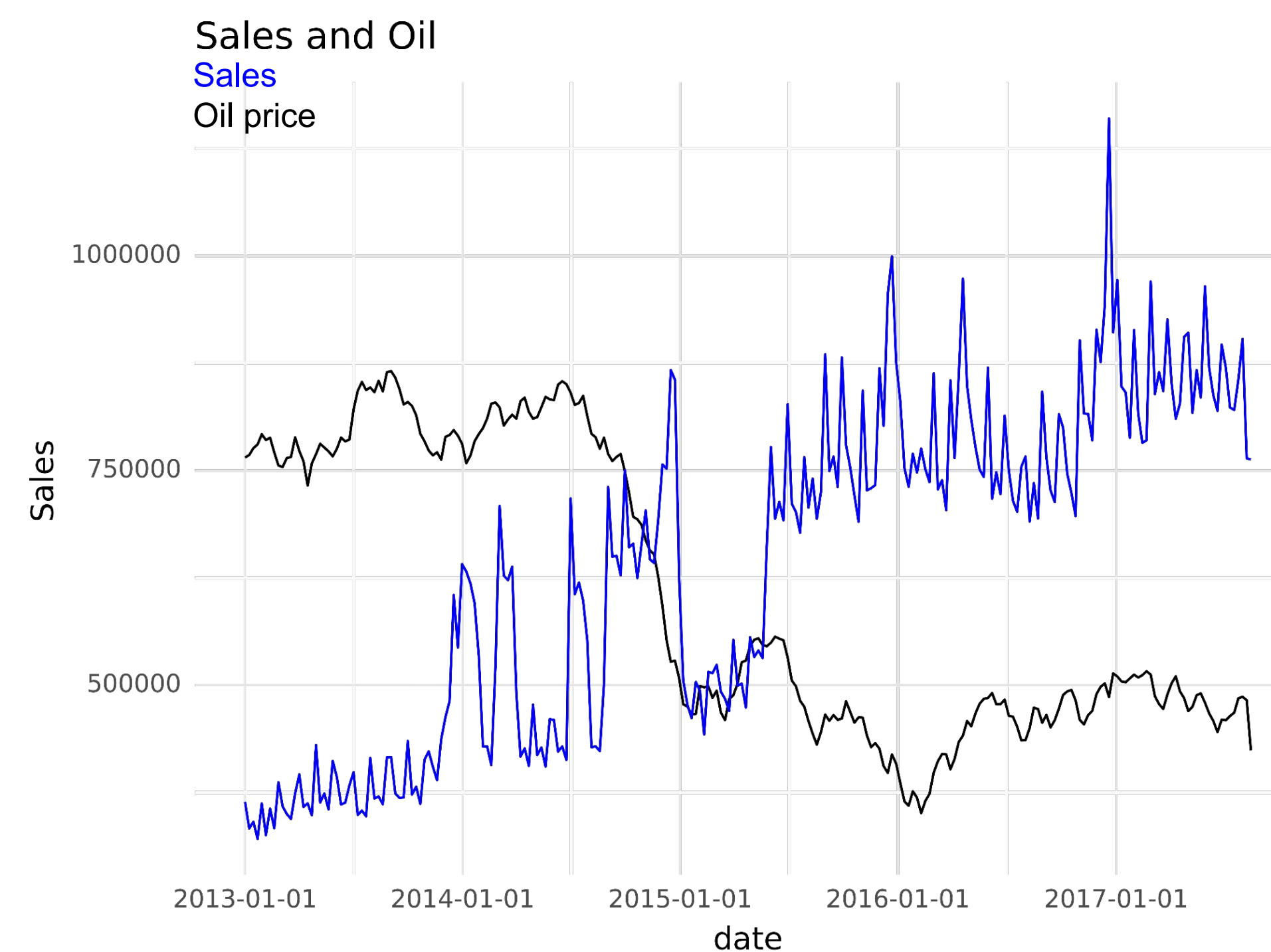


# Store Sales - Time Series Forecasting

Kristjan Radsin, Joonas Tiitson, Juhan Pauklin

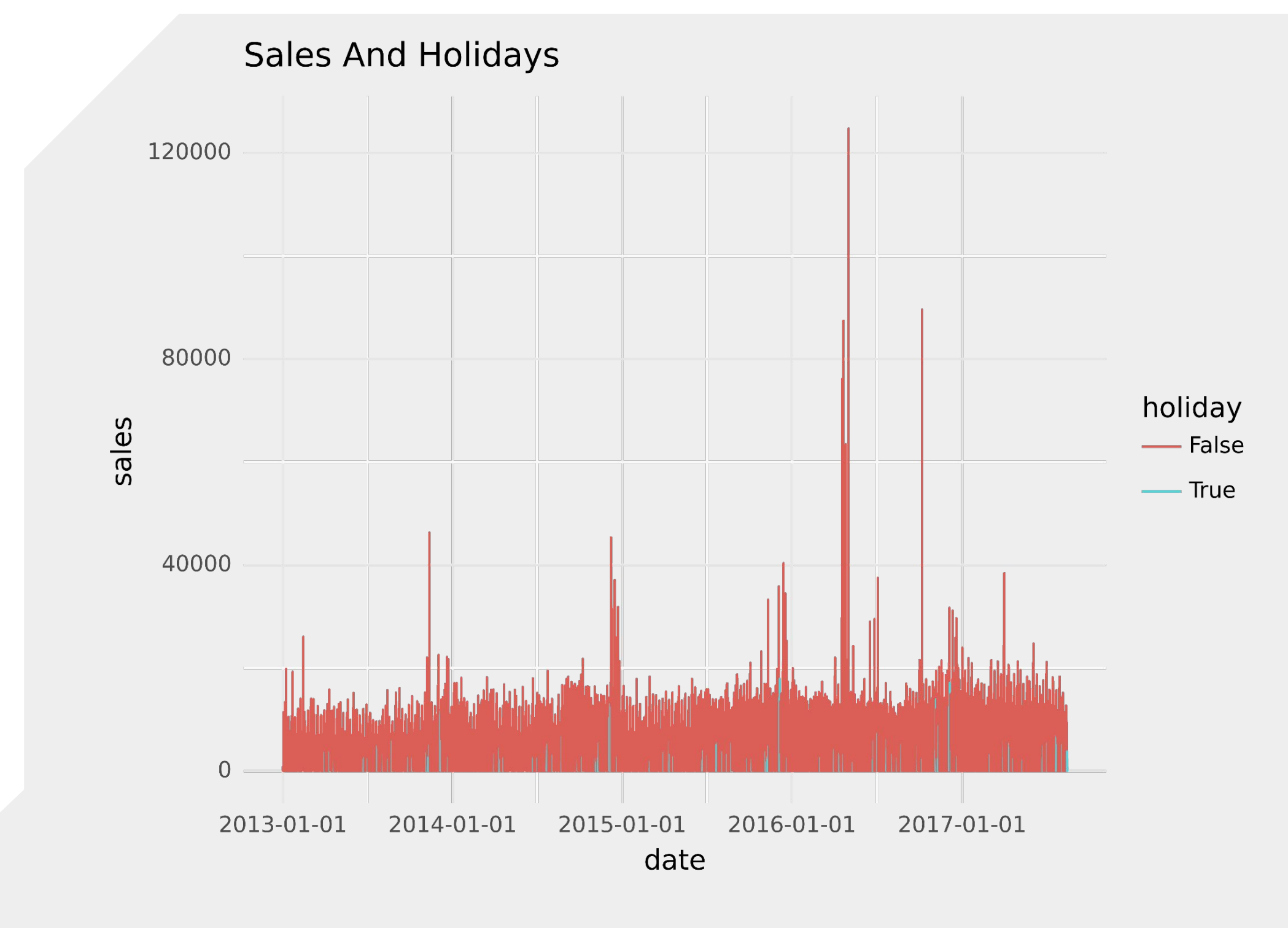
## Team-C5



### Data processing

First the data had to be cleaned, for example filling in missing values in the oil dataset.

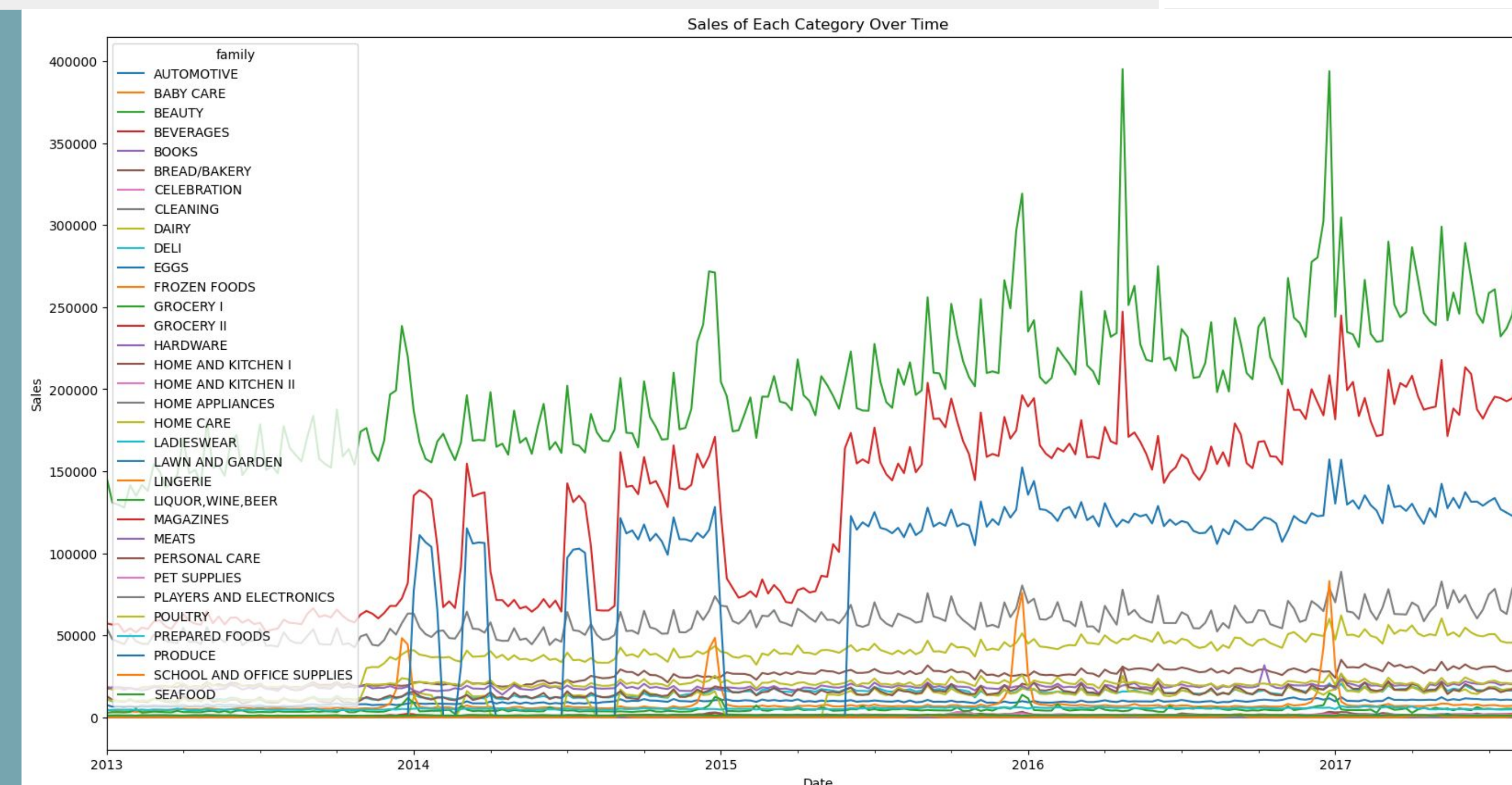
Then the different tables had to be merged so that if a sale was made on x date, the table included the oil price and if a holiday occurred on date x.



### Model training

Then different predictive models were tested, such as KNN, Random Forest Regression and ARIMA

As the merged dataset was large, requirement engineering was performed and unnecessary attributes removed.



### Introduction

Being able to predict sale numbers allows vendors to optimise inventory management and improve overall operational efficiency.

In our project we worked on a dataset from Kaggle that contains information regarding sales for the thousands of product families sold at Favorita stores located in Ecuador.

This data included info such as oil price, holidays, sold product family type as well as store type and location.

### Objective

The objective of this project was to use different machine learning techniques to forecast the number of sales of a product based on the specifications of its status and to identify how seasonality, promotions and oil prices influence the sale of products.

### Result

Interestingly KNN and Random Forest Regression gave better results than ARIMA, having lower Root Mean Square Error.

The data did show strong correlation between oil price and sales, but little correlation between change in sales on holidays.