

Project report: Retail sales prediction

Capstone 2

# 

# 

# Introduction

Making sales predictions is a crucial activity for retail companies to optimize inventory and promotional efforts. Major challenge however is limited historical data to make accurate predictions and limited understanding of factors affecting the demand such as promotions or holidays.

With growing uncertainty, managers are urged to more accurately capture the ‘true’ demand signal and optimize return on promotions.

# 

# Business opportunity

This solution provides 3 tangibles benefits to the retail company:

Short term forecasting solution

Long term sales prediction

Recommendation for markdowns optimization

The solution is targeting product managers with as tangible benefits:

Better understanding of true demand signal

Improved forecast accuracy

Improve return on markdowns

# Dataset

1. Dataset source: Kaggle ( <https://www.kaggle.com/manjeetsingh/retaildataset?select=sales+da>)
2. Description: Dataset is deriving from the aggregation of 3 tables
3. Our target feature is ‘Weekly\_Sales’
4. Data covers the period from Feb 2010 to Feb 2013

# Data wrangling

**Data cleaning and tidy:**

* Merging the 3 tables to form a single data set
* Addition of a new column to sum up all markdowns
* Correction of negative sales values by inputting zeros. This concerned 0.3% of data
* Imputing gaps in markdowns by zeros assuming this correspond to period where the promotional activity was not active.
* Size, Dept, Type: replacement of missing values by the mode
* Removal of outliers for employment and Temperature

# EDA (Exploratory Data Analysis)

**Approach**

* Univariate analysis of the distribution of each feature
* Correlation pair plot
* Heatmap
* Time series analysis through decomposition of target variable into its different components: trend, seasonal and residual.

**Key finding:**

* Weekly sales are highly influenced by the size of the store and the department . Other factors have little influence on sales performance
* Fuel price and year display some significant level of collinearity
* At the exception of Christmas, holidays do not have a significant impact on sales.
* Markdowns have typically low impact on sales however, since they only occur during specific periods of the year, further analysis is required to understand their impact.
* Size of stores matters for markdowns, the bigger the store, the more effective the markdown
* Markdown 2 and 5 impact is more volatile as affected by fuel price

# Pre-processing

**Data transformation:**

* For all numerical variables having high levels of skewness, a log transformation was applied to correct the distribution to normal.
* For categorical variables, binary standardardization has been applied
* Datetime variable has been encoded by splitting every component of year, month, week and day

**Train and test split**

The partition was done 70/30 of the data to train-test-split.

# Modelling

## Time series modelling

* Arima forecast using cross validation to tune parameters has been proposed for sales forecast and to improve the timing of markups.
* Focus markdowns only during the Christmas period where its have the biggest impact rather than spreading all year long

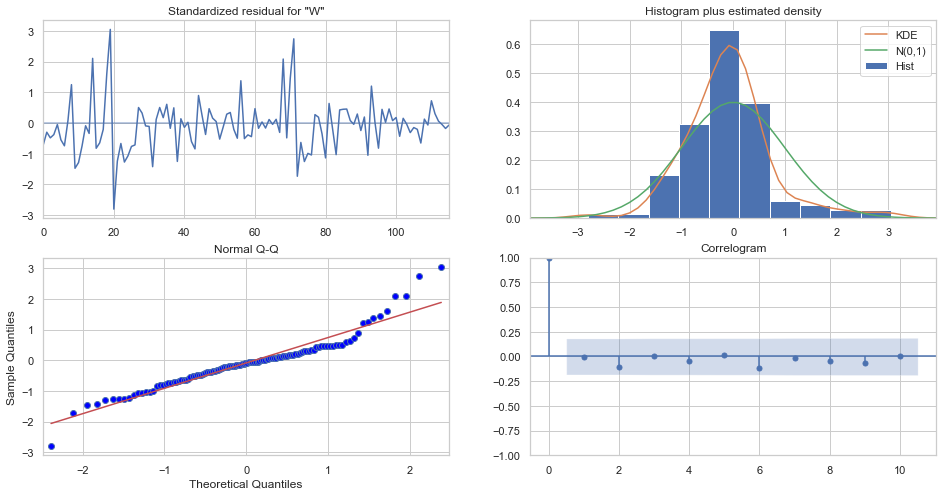
**Machine learning modelling**

3 types of machine learning models were tested including Linear regression, Random Forest and Gradient Boost. Prior to model selection, we also used Ridge and Lasso regressions to evaluate features of importance.

Random forest model was selected based on lower error metric (MSE)

# Model evaluation

**Time series model evaluation**



**Machine learning**

|  |  |
| --- | --- |
| Model | mse\_score |
| mse\_lm2 | 2.767948e+01 |
| mse\_gbr | 2.767948e+01 |
| mse\_rf | 1.683097e-04 |
| mse\_lm | 3.536717e-27 |
| mse\_ridge | -3.237305e-13 |
| mse\_lasso | -3.237305e-13 |

# 

# Further analysis

* Customizing the model to forecast at department level will provide more granularity to predictions
* Price elasticity study can also bring additional insight to understand sensitivity of each store/department to markdown and hence improve on depth of discount