

Forecasting sales

Background and project definition

Sales forecasting are is an important task in companies since it provides a great source of information for planning and decision making. Demand forecasting is the basis for a lot of managerial decisions budgeting, production planning, inventory control, and human resource planning. However, demand forecasting can be challenging as there are many uncertainties associated with them. There are many internal and external variables (e.g., advertisement, promotions, economic situation, competitor's behaviour, weather, special events) that may impact demand.

In this project, we will investigate a real-world problem for a company (Rossmann) that operates over 3,000 drug stores in 7 European countries. Store sales are influenced by many factors, including promotions, competition, school and state holidays, seasonality, and locality. Students will extract some information from the data as requested by managers.

Goals of the project:

Managers are interested to have descriptive statistics of the data and then use it for various purposes.

Some of the student's tasks include:

1. Gathering data from the web
2. Cleaning and wrangling data, i.e., taking top 20% of stores in terms of sales and putting that into a proper format
3. Providing descriptive statistics of data across the selected observation. This includes mean, standard deviation, min and max for each store and across all stores and interpreting the results.
4. Plotting series over time and describing trends and changes in the selected series. Using different plots to explore data, i.e., boxplot, pie chart.
5. Looking at the correlations between variables, measuring it and visualising it using scatter plot
6. Creating a simple model that can represent the relationship between sales and variables. Predicting sales for four weeks and then interpreting results.

Data:

You are provided with historical sales data for 1,115 Rossmann stores. Data and description can be found on <https://www.kaggle.com/c/rossmann-store-sales/data>.

You will only need to work with the train data set. The rest of the data set are for a forecasting competition and is not needed for the current project.