What is the problem

I need a program to predict outlet item sales based on a few parameters

T: Predict item sales in different outlets

E: A corpus of items, each one with the how much money it earned

P: RMSE error between predictions and actual ground truth sales.

Assumptions

Item identifier and weight don’t matter to sales

Healthier items are sold more

Visible items are sold more

People buy more items of a specific type

The year of outlet establishment doesn’t matter

Bigger outlets sell more

Location and outlet types effects the sales

Similar problems

Predict house prices, predict black Friday sales.

Why does the problem needs to be solved

This is a learning exercise.

Having the problem solved will allow me to enhance sales of a particular item by affecting some of the parameters. It will allow me to create ads and program to enhance the sales of undersold items. It will allow me to stop selling items which do not benefit me and focus on other items.

Solution Use

The solution will be valid for a few years, until the market changes and the sales are no longer valid. If some items are stopped being sold, new items are introduced or there is drastic change in the supply of items.

How would I solve the problem?

Clean the data and explore it. Check my assumptions on the data. Think which new features can be produced from the given data. Build models and test them.