Predicting total Sales in a given month on the basis of Historical data.

Kajetan Kubik 145451

Workflow: I want to build as universal model as possible, so I decided to look only at historical sales data. I decided to validate my prediction, firstly trying to predict last month that i have in data. I take data from Kaggle. Dataset contains no missing values. I choose from this dataset data on one company, drop all unnecessary columns, drop 2 outliers, and group all sales in a given day/month together(I need the 'day' for method I choose, and I need the month because my goal was predicting total sales in a given month, so for good visualization). Then I try to see if in this data set exist any patterns "by human eye" and explore a little the dataset.

Then I start predicting using fbProphet. This is a very fancy model, so I don't have to normalise data, because models do it for me. Model takes 2 columns: date and value at a given date. Model predicts total sales every day, so I sum up predictions for the next 30 days, and this, in total, gives me a desired prediction for total sales in November. I valide my model, predicting total sales in October, and comparing it with data, and in results it turns out that model is wrong, something around 200, so 6.5 sales per day. So, in the end I decided that my prediction is quite accurate.

Reflection: When I tried to predict without choosing only 1 shop, accuracy went slightly higher, but I thought that it is a rare case: some firm/shop wants to predict total sales from data which contains information from 50 different stories.

'Bonus' Section: Since I have this data, which is gathered to build a model which will predict no total sales, but sales of a given item in a given shop, but it was quite hard. I build a pivot table, where rows are items, columns are months, and cell's values are sales in this month of this item. The biggest issue here was the fact that 90% of items were sold in total 10 times, and meaningful rest in hundreds or even thousands, so all models that i try, tend to underestimate more than can someone imagine.