Prediction of Rossmann Store Sales

Yilin Wei, Tingting Gao, Jialu Yan
Instructor: Dr. German Creamer

STEVENS INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®
Business Intelligence & Analytics

Briefing

The goal of this project is to predict Rossmann store Sales with different periods data and different algorithms. This help us to learn the effects of promotion and arrange supply chain management, investment and workforce.

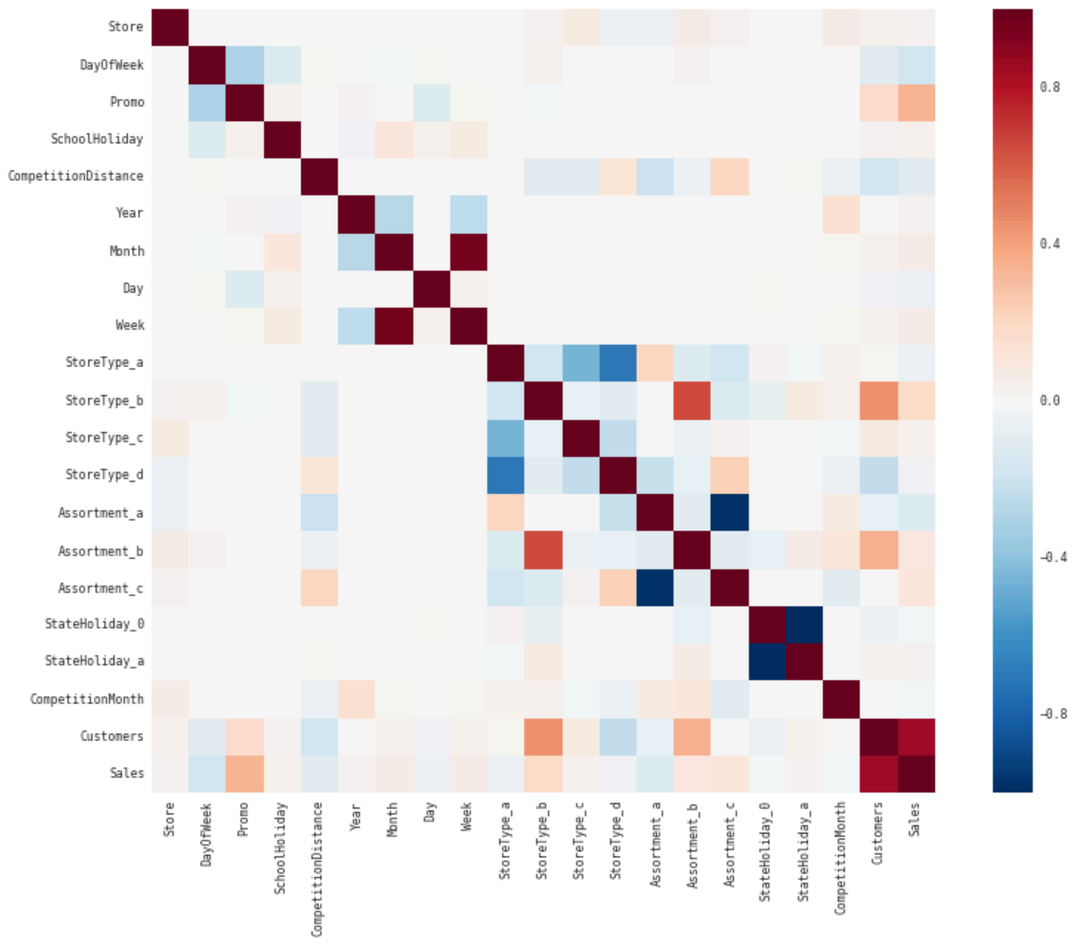
3,000 drug stores in 7 European countries

Methodology

Exploration

- Customers, Promo and StoreType_b positively relate to Sales
- DayofWeek and Assortment_a highly negately relate to Sales

Correlation Matirx for Train (Promo2==0)



Modeling Process

- Separately split train and test into two based on Promo2(0 or 1)
- Because Customer highly relate to Sales, predict Customer for test
- Compared with decision tree and KNN, random forest performs the best

Store table: information of 1115 stores

Sales table: sales during 2013-1-1 2015-7-31

Merge into one table

Features	
Store	SchoolHoliday
Day of week	HaveCompetitor
Promo	CompetitionDistance
Year	CompetitionMonth
Month	Dummies of StoreType
Day	Dummies of Assortment
Week	StateHoliday

Train: Promo2==0
...
Test:Promo2==0
Test:Promo2==1
Train: Promo2==1
...
Promo2Month
Promo2Week

Algorithm 1
Algorithm2
Predict Customers for test data

Prediction

Evaluation:

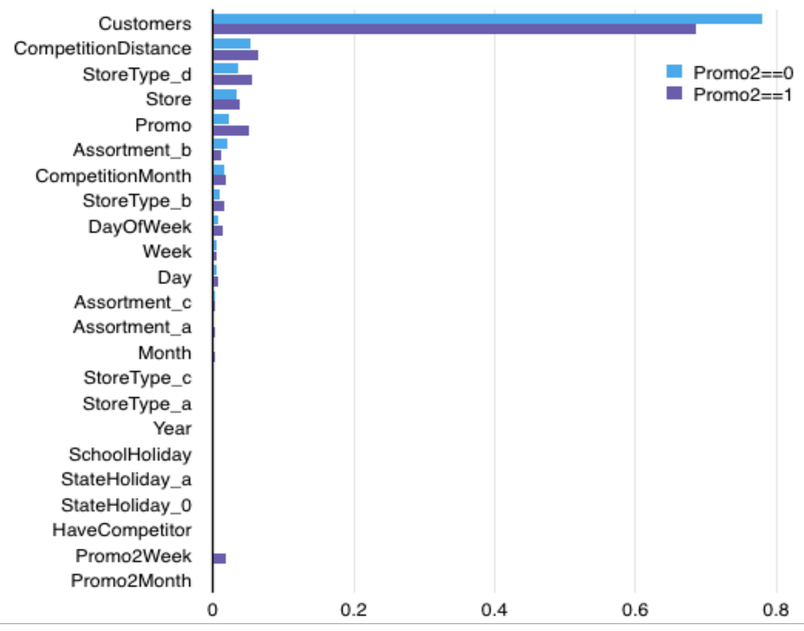
$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2}$$

RMSPE	Before predict customer			After predict customer		
Promo2	0	1	All	0	1	All
DT	0.1958	0.2112	0.2036	0.0893	0.1063	0.0981
KNN	0.2870	0.3359	0.3123	0.0797	0.0944	0.0873
RF	0.1518	0.1649	0.1585	0.0608	0.0690	0.0650

Methodology

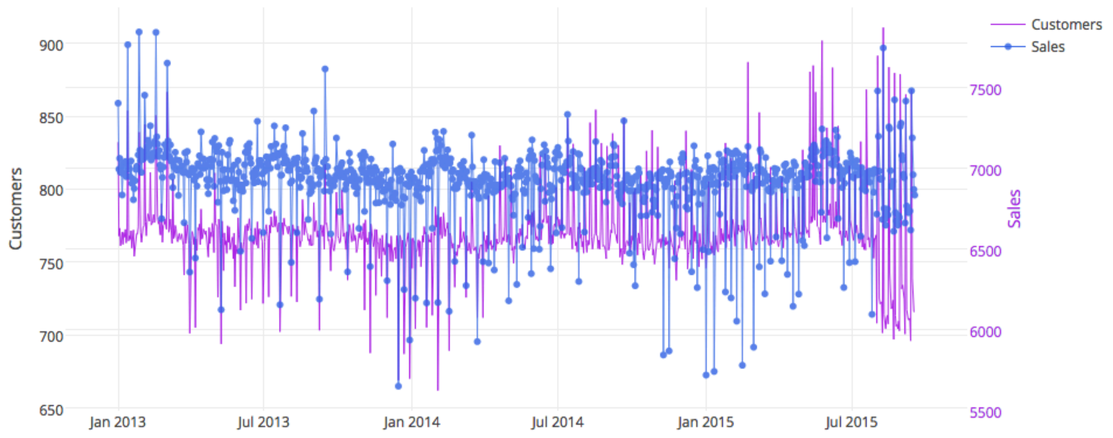
Feature Importance

- Customers, CompetitionDistance and StoreType_b are important features for both train datasets

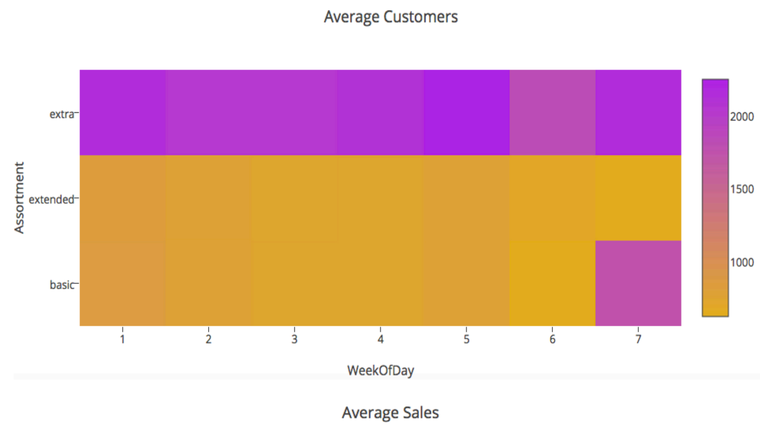


Analysis

- Predictions are effective at first and then fluctuate more than the real data. We should use time series analysis to improve accuracy



- Sunday and Monday have more customers and sales
- Extra assortment stores have more customers and sales



- Basic stores conduct more short-term promotion than email marketing
- Extended stores conduct more email marketing than short-term promotion
- The customers of extended stores are more profitable than that of basic store