Thank you for accepting the rules.

$35,000 • 3,031 teams

Rossmann Store Sales

Merger and 1st Submission Deadline

Wed 30 Sep 2015 Mon 14 Dec 2015 (14 days to go)

Dashboard

Home
- Data
- Make a submission

Information
- Description
- Evaluation
- Rules
- Prizes
- Timeline

Forum

Scripts
- New Script
- New Notebook

Leaderboard

My Team

My Submissions

Leaderboard

1. Gert
2. SDN
3. Tobias Wolfanger
4. FOols w/ ToOls
5. HassanAbdulQayyum
6. Objective Partners
7. MAX-CON DATA SCIENCE
8. Mr. Lucky
9. Oleg Agapkin & Sergey Makarevich
10. MMD

1,803 Scripts

XGBoost Feature Importance
24 Votes / 8 days ago / Python

A Journey through Rossmann

Competition Details » Get the Data » Make a submission

Data Files

File Name	Available Formats
sample_submission.csv	.zip (55.25 kb)
store.csv	.zip (8.33 kb)
test.csv	.zip (143.25 kb)
train.csv	.zip (5.66 mb)

You are provided with historical sales data for 1,115 Rossmann stores. The task is to forecast the "Sales" column for the test set. Note that some stores in the dataset were temporarily closed for refurbishment.

Files

- **train.csv** - historical data including Sales
- **test.csv** - historical data excluding Sales
- **sample_submission.csv** - a sample submission file in the correct format
- **store.csv** - supplemental information about the stores

Data fields

Most of the fields are self-explanatory. The following are descriptions for those that aren't.

- **Id** - an Id that represents a (Store, Date) duple within the test set
- **Store** - a unique Id for each store
- **Sales** - the turnover for any given day (this is what you are predicting)
- **Customers** - the number of customers on a given day
- **Open** - an indicator for whether the store was open: 0 = closed, 1 = open
- **StateHoliday** - indicates a state holiday. Normally all stores, with few exceptions, are closed on state holidays. Note that all schools are closed

Stores
24 Votes / 10 days ago / Python

Interactive Sales Visualization!
76 Votes / 31 days ago / RMarkdown

Interesting stores
8 Votes / 4 days ago / R

Exploratory Analysis Rossmann
177 Votes / 47 days ago / RMarkdown

Filling Gaps in the Training Set
50 Votes / 39 days ago / RMarkdown

Forum (184 topics)

How come we can use weather data?
3 hours ago

External data and Other information
3 hours ago

Am I overfitting??? Need help!!
6 hours ago

Use of Data for Academic Research
11 hours ago

Lots of Scores of 0.10361
14 hours ago

Store type
yesterday

teams

players

entries

on public holidays and weekends. a = public holiday, b = Easter holiday, c = Christmas, 0 = None

- **SchoolHoliday** - indicates if the (Store, Date) was affected by the closure of public schools
- **StoreType** - differentiates between 4 different store models: a, b, c, d
- **Assortment** - describes an assortment level: a = basic, b = extra, c = extended
- **CompetitionDistance** - distance in meters to the nearest competitor store
- **CompetitionOpenSince[Month/Year]** - gives the approximate year and month of the time the nearest competitor was opened
- **Promo** - indicates whether a store is running a promo on that day
- **Promo2** - Promo2 is a continuing and consecutive promotion for some stores: 0 = store is not participating, 1 = store is participating
- **Promo2Since[Year/Week]** - describes the year and calendar week when the store started participating in Promo2
- **PromoInterval** - describes the consecutive intervals Promo2 is started, naming the months the promotion is started anew. E.g. "Feb,May,Aug,Nov" means each round starts in February, May, August, November of any given year for that store