$35,000 • 392 teams

Rossmann Store Sales

Merger and 1st Submission Deadlin

Wed 30 Sep 2015

Mon 14 Dec 2015 (2 months to go)

Dashboard

Home
- Data
- Make a submission

Information
- Description
- Evaluation
- Rules
- Prizes
- Timeline

Forum

Scripts
- New Script

Leaderboard

My Team

My Submissions

Leaderboard

1. BreakfastPirate
2. Shize Su
3. GIVE ME A TOP1 SCRIPT ^_^
4. Jason Noriega
5. Sputnik 1 Replica
6. Sidhha
7. Ralph Neumann
8. Alexander Larko
9. Ivanhoe
10. vnneo & GM

110 Scripts

predict_sales_with_pandas.py
5 Votes / 19 hours ago / Python

XGB W/ RMSPE Eval
6 Votes / yesterday / R

H2O Random Forest Example
12 Votes / 3 days ago / R

Random Forest Example (0.12579)
8 Votes / 2 days ago / R

Sklearn RF
2 Votes / 1 hour ago / Python

predict_sales_with_pandas.py
2 Votes / 4 hours ago / Python

Forum (28 topics)

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 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

An issue with the RMSPE metric
18 minutes ago

predict_sales_with_pandas.py
4 hours ago

What is your best score using sklearn random forest?
4 hours ago

4 Promo data fields?
9 hours ago

XGBoost w/Custom Error Function
11 hours ago

CompetitionOpenSince data missing
11 hours ago

teams

players

entries

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