File descriptions

* sales\_train.csv - the training set. Daily historical data from January 2013 to October 2015.
* test.csv - the test set. You need to forecast the sales for these shops and products for November 2015.
* sample\_submission.csv - a sample submission file in the correct format.
* items.csv - supplemental information about the items/products.
* item\_categories.csv  - supplemental information about the items categories.
* shops.csv- supplemental information about the shops.

Data fields

* ID - an Id that represents a (Shop, Item) tuple within the test set
* shop\_id - unique identifier of a shop
* item\_id - unique identifier of a product
* item\_category\_id - unique identifier of item category
* item\_cnt\_day - number of products sold. You are predicting a monthly amount of this measure
* item\_price - current price of an item
* date - date in format dd/mm/yyyy
* date\_block\_num - a consecutive month number, used for convenience. January 2013 is 0, February 2013 is 1,..., October 2015 is 33
* item\_name - name of item
* shop\_name - name of shop
* item\_category\_name - name of item category

Number of categories: 84

Number of items: 22170

Time series decomposition involves thinking of a series as a combination of level, trend, seasonality, and noise components.

Decomposition provides a useful abstract model for thinking about time series generally and for better understanding problems during time series analysis and forecasting.

A useful abstraction for selecting forecasting methods is to break a time series down into systematic and unsystematic components.

* **Systematic**: Components of the time series that have consistency or recurrence and can be described and modeled.
* **Non-Systematic**: Components of the time series that cannot be directly modeled.

A given time series is thought to consist of three systematic components including level, trend, seasonality, and one non-systematic component called noise.

These components are defined as follows:

* **Level**: The average value in the series.
* **Trend**: The increasing or decreasing value in the series.
* **Seasonality**: The repeating short-term cycle in the series.
* **Noise**: The random variation in the series.