

Sales Forecast

Problem:

- Given 34 months of sales history (Jan 2013 – Oct 2015) of a Russian toy retailer by items and by shops, predict sales for Nov 2015

Clients:

- **Leadership of the toy company:** need a statistic reference on company sales potential based on historical trend in order to predict opportunities/risks for business to hit company sales target. Based on the study, the company will priorities marketing/investment resources to maximize opportunities/mitigate risks.
- **Toy product developers:** Study may reflect meaningful category trend in which developer can prioritize resources on innovation in those categories
- **Competitors:** leverage the result to develop market/product strategy to tap into market potential
- **Data Scientists/Researchers:** If the model generates reasonable prediction, it can be scaled up to include more company's data and to be used to predict total market share for the toy industry

Data:

- Kaggle competition (<https://www.kaggle.com/c/competitive-data-science-predict-future-sales/data>)
- Kaggle API command: kaggle competitions download -c competitive-data-science-predict-future-sales

Outline:

- Data Cleansing/Wrangling: Understand the data structures, checking for missing values or invalid records
- Exploratory: Finding historical sales patterns and identify potential correlation factors that could serve as good training features
- Statistical Test: evaluate the statistical significance of the features identified in the exploratory stage on predicting target (i.e. sales), narrow down features that matters most to the predictions
- Machine Learning: measure accuracy as well as root mean squared errors across 3-4 models to pick the winning models
- Draw recommendations/next steps

Deliverable

- Codes, presentation slides to summary findings. All materials will be uploaded to Github folder below:

https://github.com/sittingman/sales_forecast