Management, Technology & Entrepreneurship

MGT-432 Data science for business

Ekaterina Kryukova

Business Use Case Report

OZON. Sales predictions and Purchase order automation in retail

OZON is a Russian e-commerce company, analogue of Amazon. Recently, on the 27 of November, the company went public in on IPO and in the first minutes of trading its shares increased by more than 40% [1]

In 2019 OZON gross merchandise volume (GMV) increased by 93% to $1.1 billion, direct sales consisted of $0,8 billion and volume of orders more than doubled to 32.2 million. Gross merchandise volume (GMV) grew 142% year-over-year for the first nine months in 2020 and reached $ 1.5 billion by Q3 2020. [3]

As OZON has been growing very rapidly, it has been collecting more and more data about customers' behavior. In 2019 the company used to accumulate 250 million lines of user logs or 170-200 GB from the website and application per day. This big data should be used efficiently to analyze the ongoing business and find out opportunities to optimize processes. With this increase in 2018–2019 the company has expanded data scientist employees threefold. [2]

For example, the retailer uses data science in purchases: how many and which products should be ordered. With such a great number of suppliers and purchasers that OZON has, mistakes are inevitable. A mistake of purchases in excess of 1% will cost the company 10 million rubles (~1.2 mln $). [4]

Using sales prediction model, the company efficiently plans warehouse, replenishes stocks, avoids the overflow and estimates demand elasticity (price, temperature and other factors).

Below there is a demonstration of model that OZON has successfully implemented.

The goal was to predict sales for each product for a certain period. The further description is for 7 days.

As a metric mean absolute error (MAE) was selected because of unbalanced training sample. As a range of products is broad (product lists of 1.5 M), a certain product in a certain region is purchased for a few quantities. And if in sum, OZON sells dozens of green dresses, a certain green dress with kittens is purchased 2-3 times a day. As a result, a sample is biased to small values. On the other hand, OZON sells Iphones, hand spinners and other popular things that are sold in every town in large quantities. MAE enables to avoid large penalties for rare products and be efficient for the most of products.

The first step was a feature engineering. Data scientists generated features after communicating with business people, understanding the business processes and factors that could influence the sales. Some of the features were previous sales for 1,2,3,4 weeks, for a week 1,2,3 year ago; views, adding to cart for previous 1,2,3,4 weeks, for a week 1,2,3 year ago; conversion of views to adding to cart and to purchase, conversion of adding to cart to purchase, its change; ratio of sales for 4 weeks to sales for the previous 1 week (if the ratio is much more than 4, now the demand for the product is volatile); ratio of the product sales to the sales of the whole category (if the ratio is close to 1, the product is a monopolist); knowledge about competitors; price; day/month/year; availability; the market growth; reviews and ratings; the quality of description; product seasonality (the approach will be demonstrated below); cross-products features[1]. In sum, 170 features and 12M samples were obtained.

[1] Used features of 10 nearest neighbours in embedding dimension for each product. Not in production yet.

Looking ahead, the most important features were the sales for the previous week (for 2, 3 and 4); product availability on the previous week – percentage of time when a product was on the website; angle coefficient of the sales graph for the last 7 days; ratio of the previous price to the future (a product with a large discount is purchased quickly); the number of direct competitors on the website (if a pen is only one in the category, the sales are more or less static); the product size (if a product is narrow and long sales are more volatile, the example: umbrellas and fishing rods); day of the year (New Year, the Women Day – 8/03, the beginning of a new season)

The next step was a sample collection. Sample collection for long period was not simple task because the business processes changed and some features estimated differently than it had been before but the columns stayed the same. Moreover, there had been some cases when server had crashed but nobody had noted and the values zero did not imply the absence of sales.

Eventually, a sample of 15M was gathered for 4 weeks and 2 000 lines of code to process data were written on Spark. After feature cleaning such as avoidance of anomalies (3 sigma was used) and recovery of sales during the product absence in a warehouse, a sample of 10M was left.

Then, data scientists built Machine Learning models.

In the beginning, data scientists built simple models as baselines:

a. A random generator in a range from 0 to 1000. MAE was 496.
b. Average of weekly previous sales. MAE was 1,45
c. Sales for the previous week. MAE was 1,26.
d. ½* Average of weekly previous sales+ ½* Sales for the previous week. MAE was 1,2.

Afterwards, data scientists built more complex models and compared results:

a. Linear regression, MAE was 1,15
b. RandomForestRegressor, MAE was 1,1
c. XGBoost, MAE was 1,03 but too long (no GPU)
d. LightGBM, MAE was 1,01, quicker than XGBoost

All products were divided into 13 categories according to the categories on the website catalog: tables, notebooks, bottles and so on. For each category models were trained with different depth – from 5 to 16 days. For tuning parameters random search was used that gave the best 10 hyper parameters and then data scientists built metrics for different range of target, learning curves, other graphs and trained again. The training took 5 days and huge computer clusters, 130 models were built: 13 types of product and 10 depth of prediction. The mean MAE for 5-fold time series cv was 1.

The next goal was to deploy and automate purchase orders. To train a model to purchase the necessary amount, data scientists estimated the cost of every product overstock an understock and built a model that receives a sales prediction, adds a normally distributed noise (modeling of suppliers imperfection) and for every certain product learns to add number of sales to minimize loss of money.

Every night algorithm starts, pulls data (around 20 GB) into a local HDFS from different sources, chooses a supplier for every product, gathers product features, predicts sales and creates applications according to a delivery schedule. By 6-7 a.m. data scientists provide results to employees responsible for communication with suppliers, they check and push a button. The application is registered.

Product seasonality

The main problem of identifying product seasonality was that product history was required for 2 years minimum, but OZON launched 1000 new products a day, thus not all products have this prerequisite.

To solve this challenge data scientists clustered sales time series for products that had a history of 2 years. There were 8 clusters that separated very well. These clusters were interpretable: winter/summer, New Year, Women Day.

Then, data scientists chose some features that did not require 2 years history to understand seasonality (for example, product description) and classified data sample from clusterization in the previous step.

Finally, data scientists applied obtained classifier to products without history

Solution in details:

To clusterize sales time series for products, the team preprocessed data: fill nan, logarithm (because it increased the score), used standard scaler (to compare high scaling and low scaling products). The, data scientists generated features such as angle coefficients, autocorrelation coefficients, skewness of time series, top 3 absolute value of Fourier coefficients and corresponding frequencies, discrete wavelet transform coefficients and key dates (New Year, Women Day and others)

Afterwards, data scientists applied clusterization in feature dimensions. Some results are shown below: Figure 1 Clusterizarion result. New Year products and Figure 2 Clusterization result. School products

Then, the collected features that exist for all products, applied Tf-Idf for product descriptions, used products' history, if existed, applied products' words embedding (Word2Vec) and classifies data sample from clusterization using Random Forest Classifier.

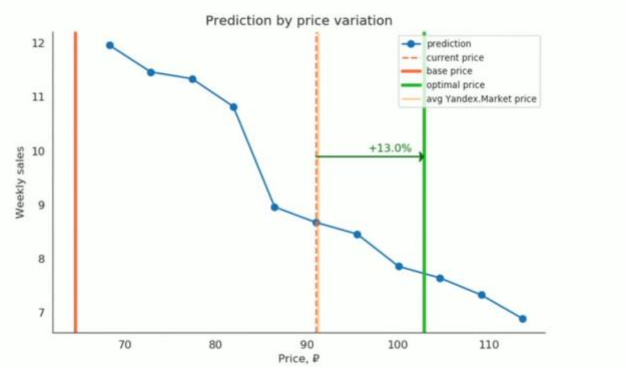
Finally, the team applied model on data without history and got precision ~ 0.7, recall ~ 0.7.

Due to information about product seasonality MAPE for sales prediction decreased by 10%. As for business results, the quantity of manual work reduced (e.g. for labeling winter/summer products), products procurement has become automated, unnecessary products procurement decreased.

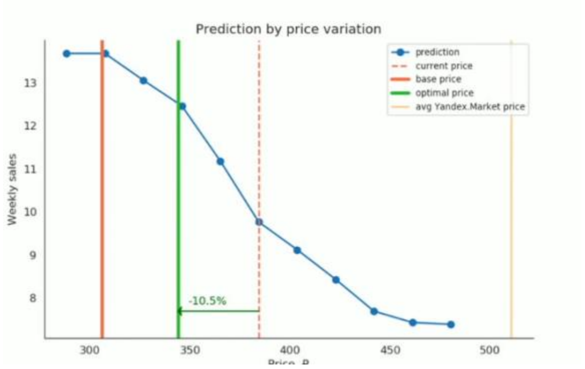
Price optimization

As a side effect from the sales prediction model data scientists could solve the problem of finding optimal prices that maximizes revenue with profit margin constraints [Figure 3 Price optimization task]. There was an assumption that price of one product doesn't influence prices of other products. For this optimization problem data scientists used scikit learn. Results can be seen on graphs:

- For goods with inelastic demand model showed that the price should be increased (the sales will not change significantly):



- For goods with elastic demand model showed that the price should be decreased (the sales will change):



The best result model showed for products from category “Pharmacy”: +3.5% sales while remaining the same margin.

As for business benefits:

- Products availability at the end of week increased by 20% and at the same time products, that are purchased less than 1 time in a month and are left, remained the same
- Manual work decreased
- Opportunity to illustrate dependences of demand on features is available now
- Opportunity to price, description and photos optimization is available now

References

1. https://www.themoscowtimes.com/2020/11/24/ozon-shares-jump-on-us-ipo-a72139
2. https://www.benzinga.com/news/20/11/18497391/ozon-ipo-what-investors-should-know-about-the-amazon-of-russia?utm_source=The+Bell+%28Eng%29&utm_campaign=8c6f5bbc5d-EMAIL_CAMPAIGN_2018_06_01_10_28_COPY_01&utm_medium=email&utm_term=0_cc8c2d1cde-8c6f5bbc5d-73689389
3. https://www.vedomosti.ru/management/articles/2019/09/25/811995-v-rossii-viros-spros-na-spetsialistov-po-dannim
4. https://www.e-xecutive.ru/management/marketing/1990848-kak-ozon-ispolzuet-big-data-v-marketinge
5. https://www.youtube.com/watch?v=WwNRRvYLGXI&feature=emb_logo
6. https://usedata.ru/2019/abstracts/5391
7. https://habr.com/ru/company/ozontech/blog/497682/

Appendix

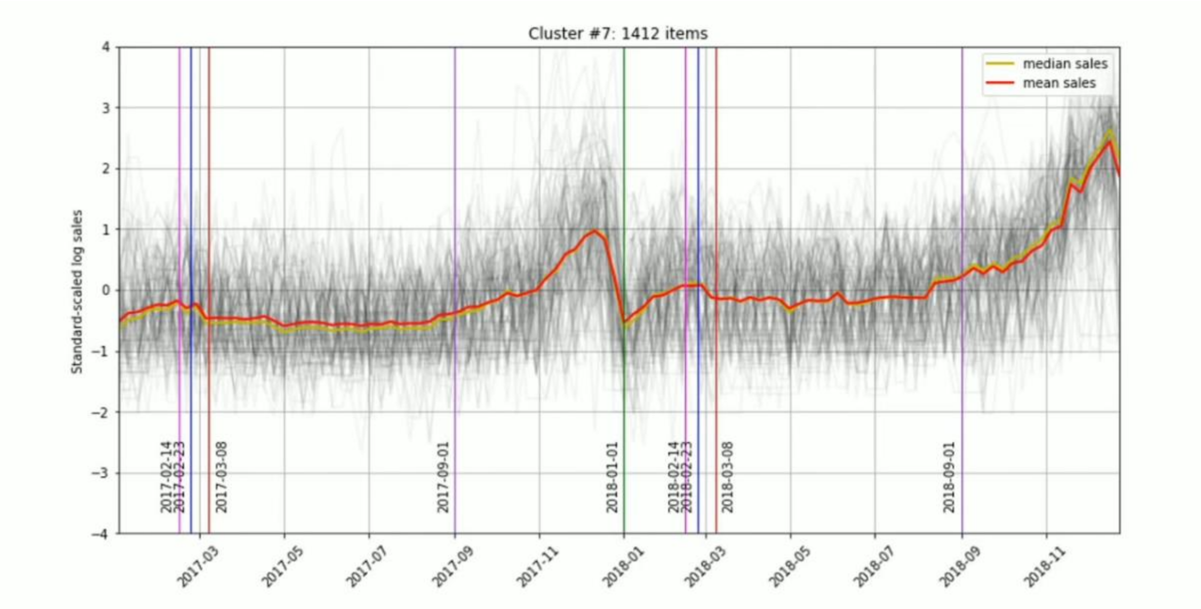


Figure 1 Clusterizarion result. New Year products

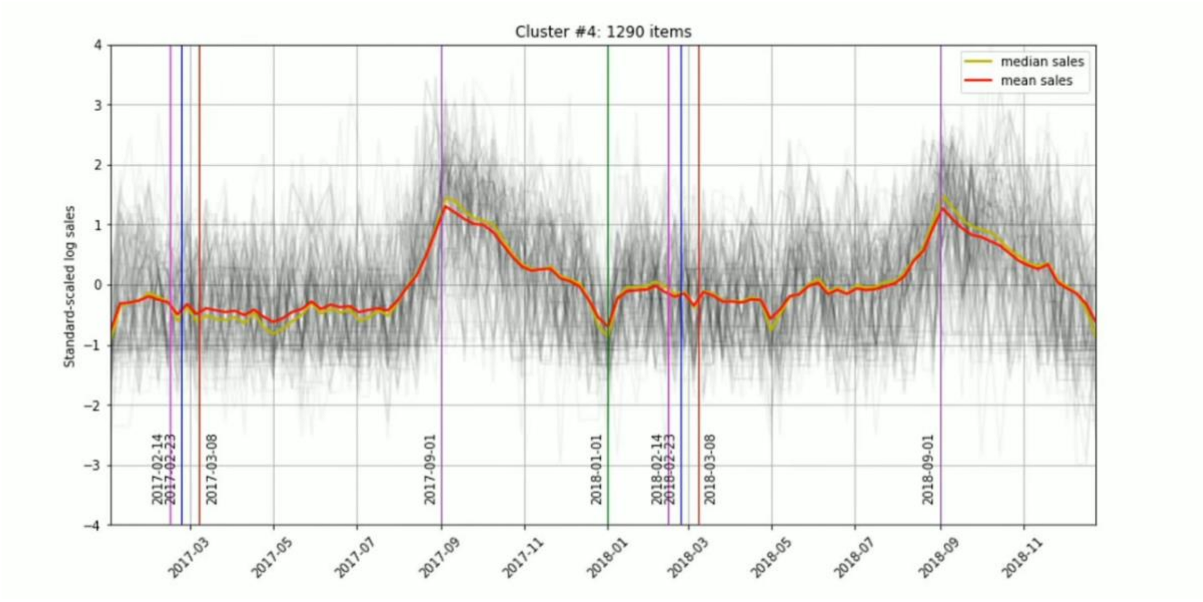


Figure 2 Clusterization result. School products

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \quad & f(\mathbf{X}) = -\big(Pr(\mathbf{X}),\ \mathbf{X}\big) = -\sum_{i=1}^{n} Pr_i(X_i)X_i \\ \text{subject to} \quad & C_i^{lower} \leqslant X_i \leqslant C_i^{upper}, \\ & \frac{\sum\limits_{i=1}^{n} Pr_i(X_i)(X_i - X_{iBase})}{\sum\limits_{i=1}^{n} Pr_i(X_i)X_i} \geqslant \alpha \quad \Rightarrow \quad \frac{\sum\limits_{i=1}^{n} Pr_i(X_i)X_{iBase}}{\sum\limits_{i=1}^{n} Pr_i(X_i)X_i} \leqslant 1 - \alpha, \end{aligned}$$

Figure 3 Price optimization task

Problems encountered during the price optimization problem and applied solution:

1. As data scientists chose random 30 moments of time in the past for train data, the ratios of features could be significantly different in train and test data.
2. Extremely seasonal products that are purchased several days a year (e.g. a New Year tree) could have zero values everywhere when choosing random 30 moments of time. Thus, during this season model breaks. Data scientists plot MAE from the most important features and understood that in few cases moving average or logistic regression is a better solution.
3. Model can’t predict the highest sales if the model didn’t encounter such high sales. Extrapolation should be improved.