

Lenta Hack

Team Latent DS

26 - 28 Sept. 2020

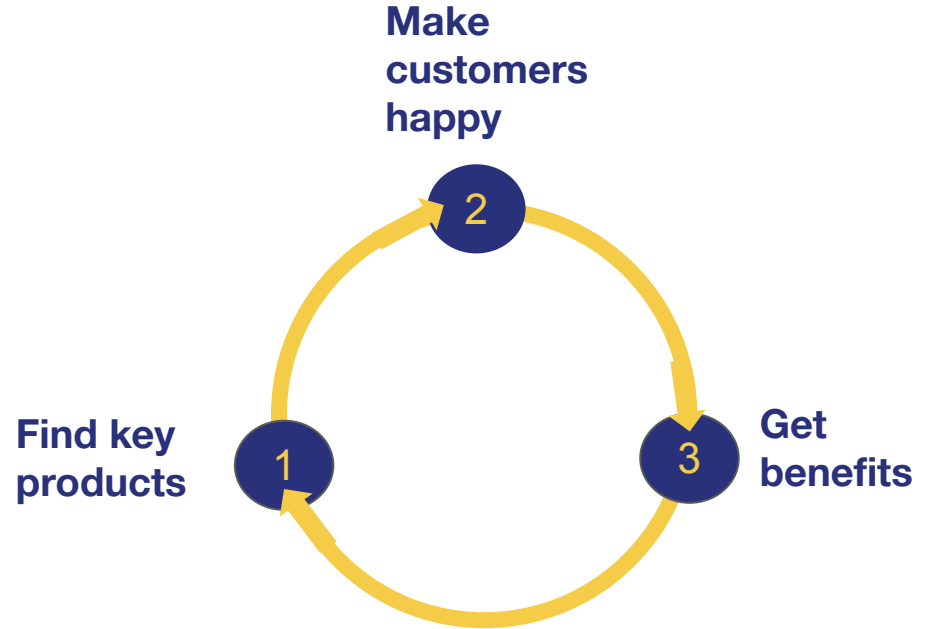
Business impact & customer focus

Customer focus

The feeling that **everything** can be found in one market

Business impact

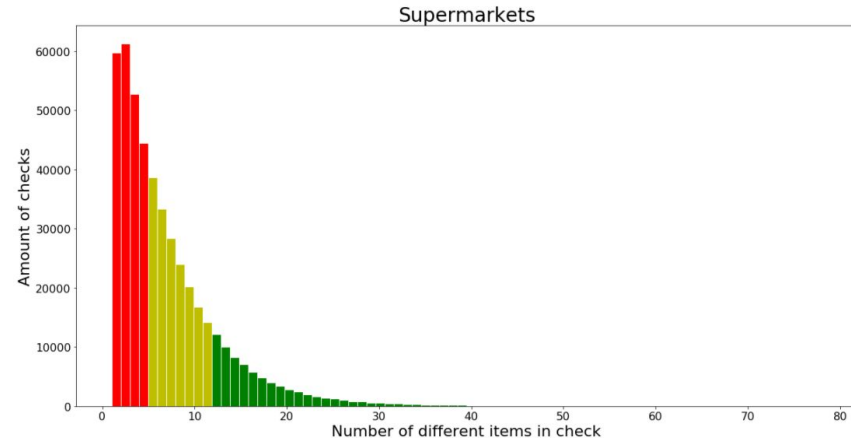
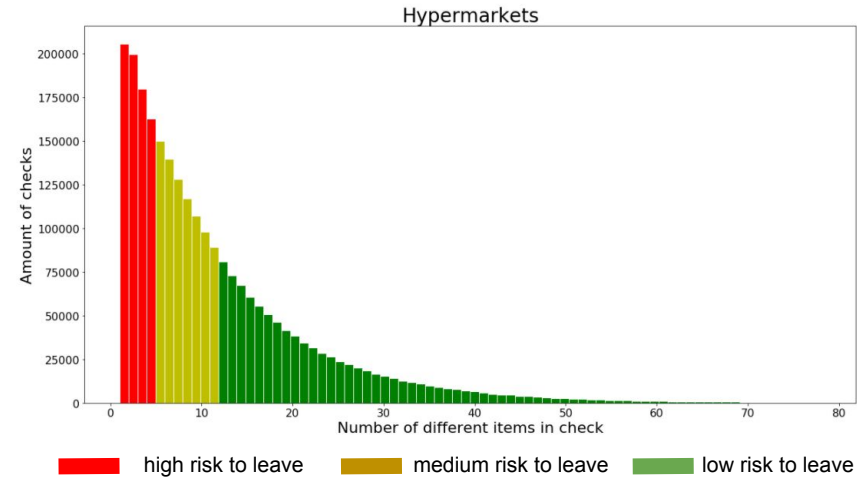
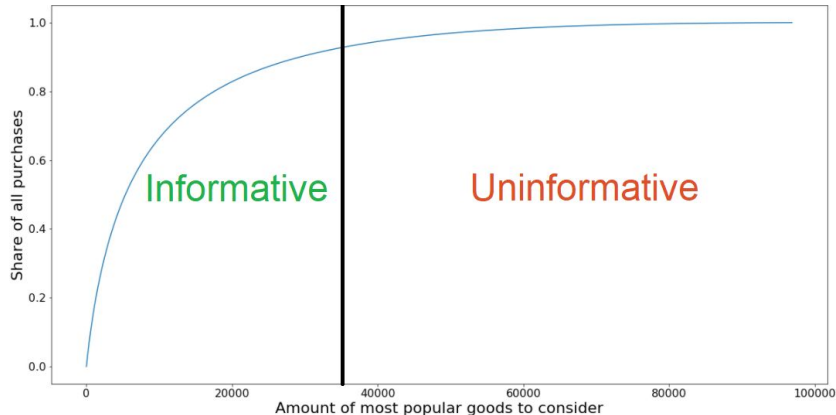
Emphasize key products which satisfy the needs of the customer, increase the total profit and attract new buyers to Lenta



Data preprocessing

We analyzed data, clean and split it for better algorithm work by following factors:

- Cheques with a big amount of different items
(big chance to leave if you haven't found what you need)
- Type of plant *(different customers behavior pattern)*
- A few most popular items and all low popular items *(first - outliers, second - not informative)*
- Transactions in which there was a wholesale purchase goods
(for example cafe owner buy 100 liters of Coca Cola - outliers)



Material links

We build relationships between products by analyzing the history of their joint purchases based on lift score* as follows:

*If customer **doesn't** buy  without , we say that*

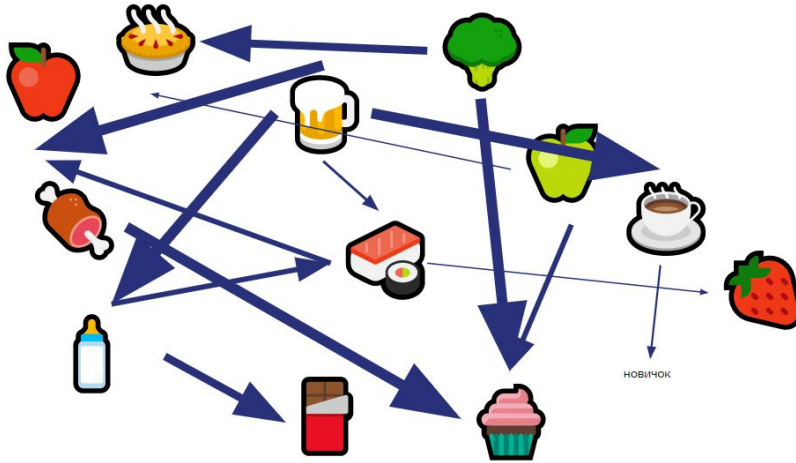
 *refers to* 



By this, we concentrate on such **key** products which important for increasing the likelihood of sales.

*See some math in Appendix

Building a graph



We find the most referenced (**key**) products and due to the nature of the relation between them

These key products have the greatest impact on the purchase of other products.



But now, which article is the most important one?

PageRank algorithm will help you, bro



Resulting graph

This cute guy's name is

b939325b0513f70814cf1c96cc130657

TOP 5 PRODUCTS FOR SM:

- 1) 45ab674261d963f7226278f2f8acb4fe
- 2) ad8aa31a2702ae04e5dd383fadb3392
- 3) 910371782b63f8b421cdc00c3b0829e9
- 4) c8451d871e1a1121cf427a8f43608403
- 5) b1d5888c2bc3c05abc3a9bcd9b3f8b0

TOP 5 PRODUCTS FOR HM:

- 1) b939325b0513f70814cf1c96cc130657
- 2) 6ddf883483679d4f737e369194ca3a26
- 3) 33cb45456bdf0993719773847cf8fc7e
- 4) 50ebcb2a3d645affe0c7dae19a12c5d4
- 5) 9ec0c4ea4ddc96175d2f39e614befc9a

The bigger the node,
the higher its
PageRank score

Model evaluation

- Cheque with **key** product gives **30%** more profit for SM and **14%** more for HM compared to a cheque without keys
- Mean cheque with a **key** is more on **3** products for SM and on **4** products for HM
- 100 **key** products are **not** in the 100 most popular products

Summary

- Model the relationship between products depending on the history of joint purchase
- Build a graph of items for specific case (market or clients)
- Find **key** products for increase the likelihood of sales by analyzing the graph
- We evaluate our model and find that it's strongly helpful for detection of **key** product

Q&A

Nikita Drobyshev
MsC-2, DS



Ruslan Aliev
MsC-2, IST



Daniil Svirskiy
MsC-2, PQM



Saveliy Galochkin
MsC-1, SES



Maxim Kan
MsC-2, DS



Appendix

We say that product A refers to B if the probability that product A will be bought, provided that product B is not on the stock tend to zero

$$p(x|\bar{y})$$

There x - event then product A is bought, y - event then product B is bought or event that B in the stock

Then A has link to B if:

$$p(x|\bar{y}) = \frac{N(A) - N(A \wedge B)}{|T| - N(B)} \rightarrow 0$$

there $N(Z)$ – number of transaction then product Z was bought and
 $|T|$ -total number of transactions

Lift

$$\text{lift}(X \Rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X) \times \text{supp}(Y)}$$

$$\text{supp}(X) = \frac{|\{t \in T; X \subseteq t\}|}{|T|}$$