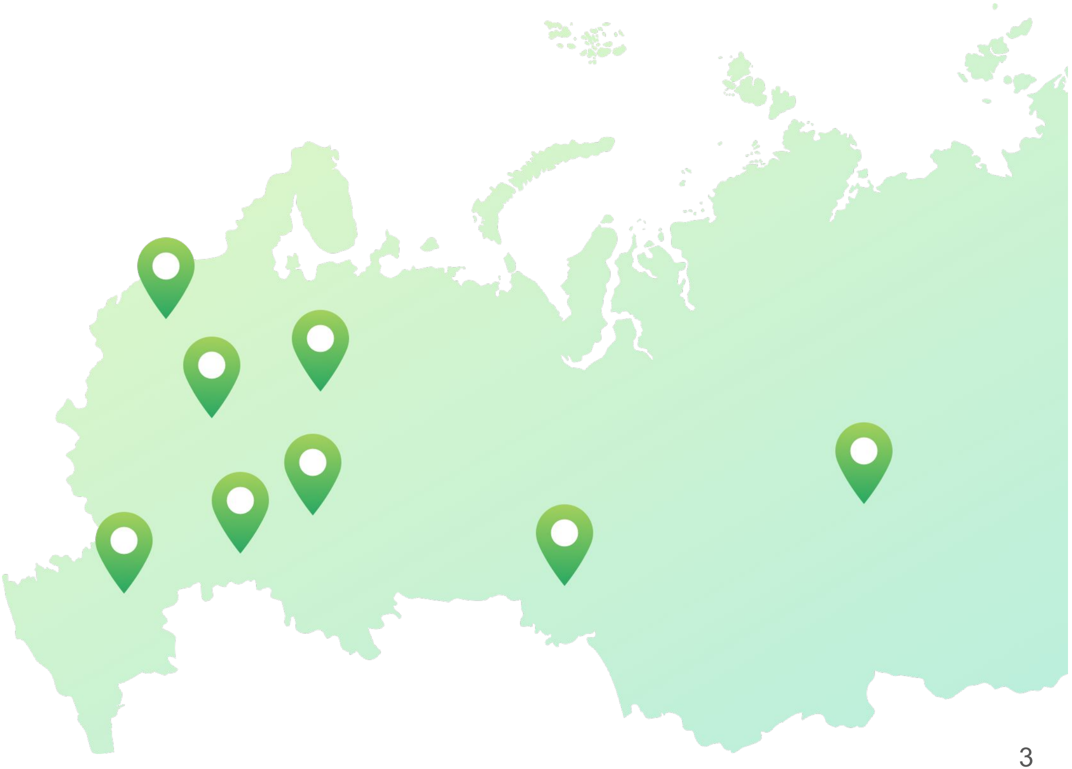
СБЕРМАРКЕТ

Predicting customers' next purchases

sbermarket.ru

October 2020

About us



Sbermarket operates in 157 cities across the country

Москва
Санкт-Петербург
Волгоград
Воронеж
Екатеринбург
Казань
Краснодар
Красноярск
Нижний Новгород
Новосибирск
Омск
Ростов-на-Дону
Самара
Уфа

Челябинск
Пермь
Тюмень
Калининград
Рязань
Иркутск
Сургут
Владимир
Тверь
Ярославль
Тула
Калуга
Благовещенск
Псков

Петрозаводск
Чита
Улан-Удэ
Чита
Владивосток
Петропавловск-Камчатский
Анадырь
Горно-Алтайск
Хабаровск
Магадан
Элиста
Южно-Сахалинск
Якутск
Биробиджан

Sbermarket is represented in every region of Russia

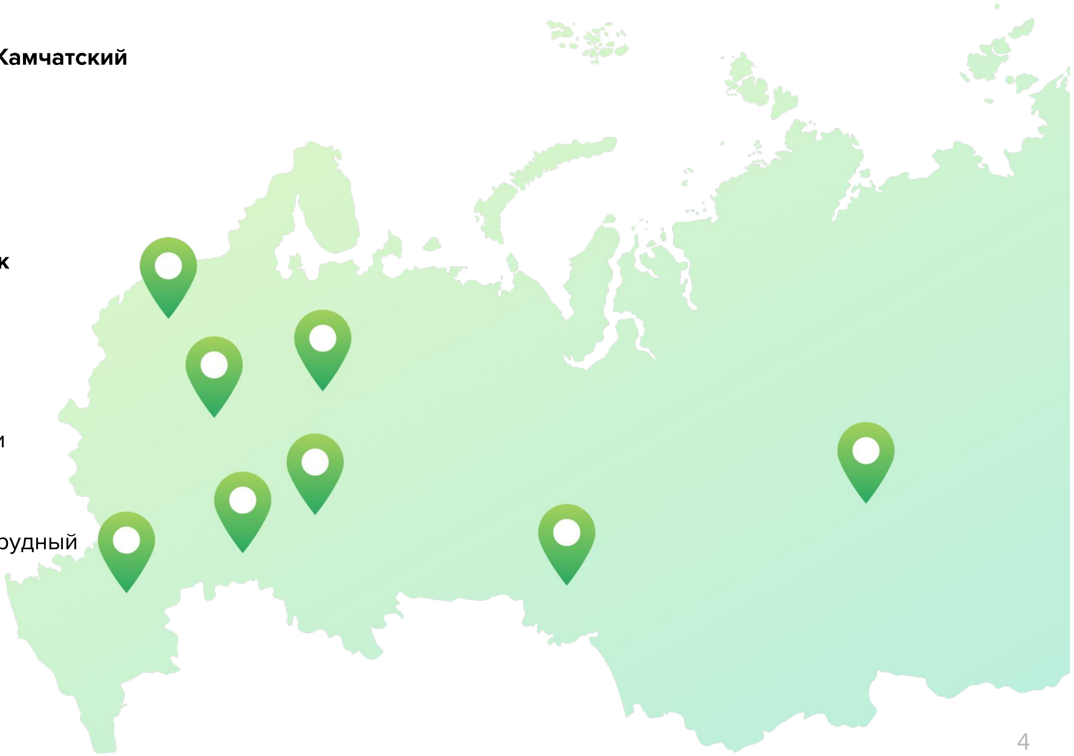
Города МО

Железнодорожный
Балашиха
Одинцово
Картмазово

Дзержинский
Зеленоград
Лобня
Ногинск

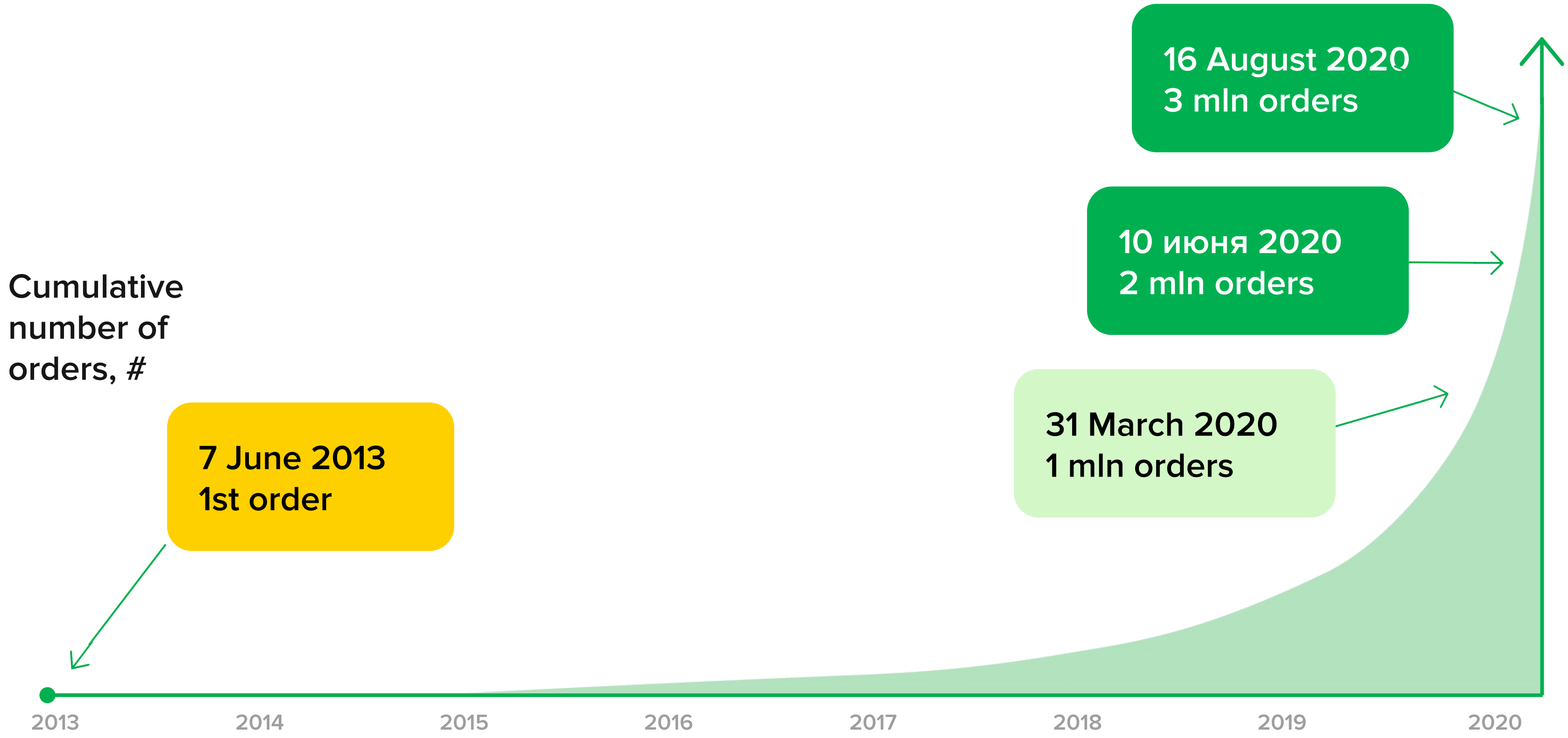
Электросталь
Подольск
Щербинка
Видное

Мытищи
Химки
Реутов
Долгопрудный



* Кроме Крыма и Севастополя

It took Sbermarket 7 years to deliver first million of orders and 71 day to deliver second million



Our goals

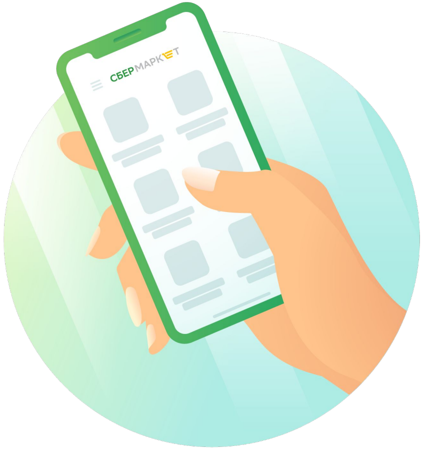
Our mission:

Save time, energy and money of our customers for something more important

Our vision:

Become an e-grocery market leader in Russia

СБЕРМАРКЕТ



Lowest prices

Widest assortment

Best quality

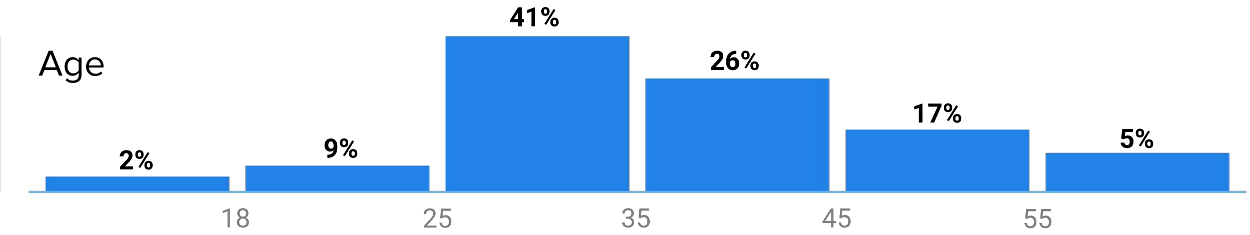
The service is focused on meeting the daily needs of a mass audience …

Moms / Housewives	Office staff	Other types of consumers	Corporate Clients / B2B
32%	19%	28%	21%
• Active Internet users • No time to shop in brick-and-mortar stores • Shopping for the family	• Accustomed to on-demand • Little free time • Wide range of purchases	• Mobility impaired and / or elderly • Customers who order products for loved ones • No time to visit hypermarkets	• Big and heavy shopping • Shopping for office and staff

Current client profile

22% male

78% female



… and covers 100% of the grocery basket and convenience goods

Average basket breakdown of by category

Fresh 49%


Fruits and vegetables 19%


Beverages 8%

Canned goods 5%


Dairy 15%

Sweet and Snacks 8%


Cheeses 4%


Grocery 9%

Personal Care 7%


Bakery 4%


Household products 8%


Meat and Seafood 9%


Frozen 4%

COVID accelerated not only GMV growth...

COVID accelerated not only GMV growth...

Перекрёсток
быстро

igooods

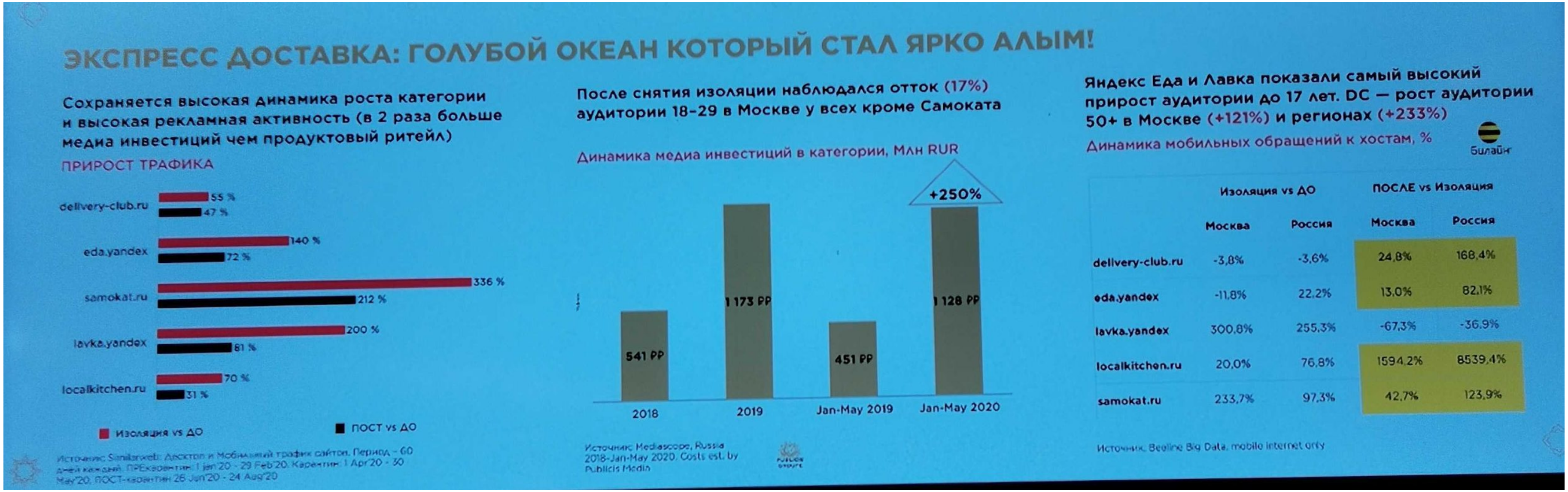
Перекрёсток
впрок

ВкусВилл
экспресс

Delivery
Club

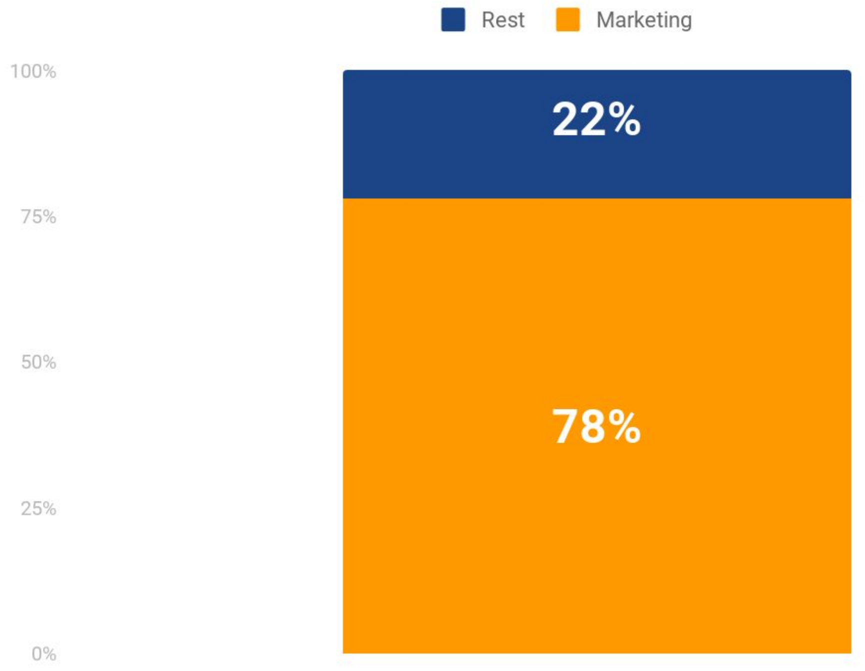
... but also a rapid spike in number of competitors

“Blue ocean turned extremely red”



Such competition causes huge burn of marketing budgets across industry

Sbermarket budget shares



- Budget is allocated on acquisition and reactivation
- **Each percentage gain in conversion from launch to order leads to drastic optimization of marketing budget (tens of millions $ per month)**

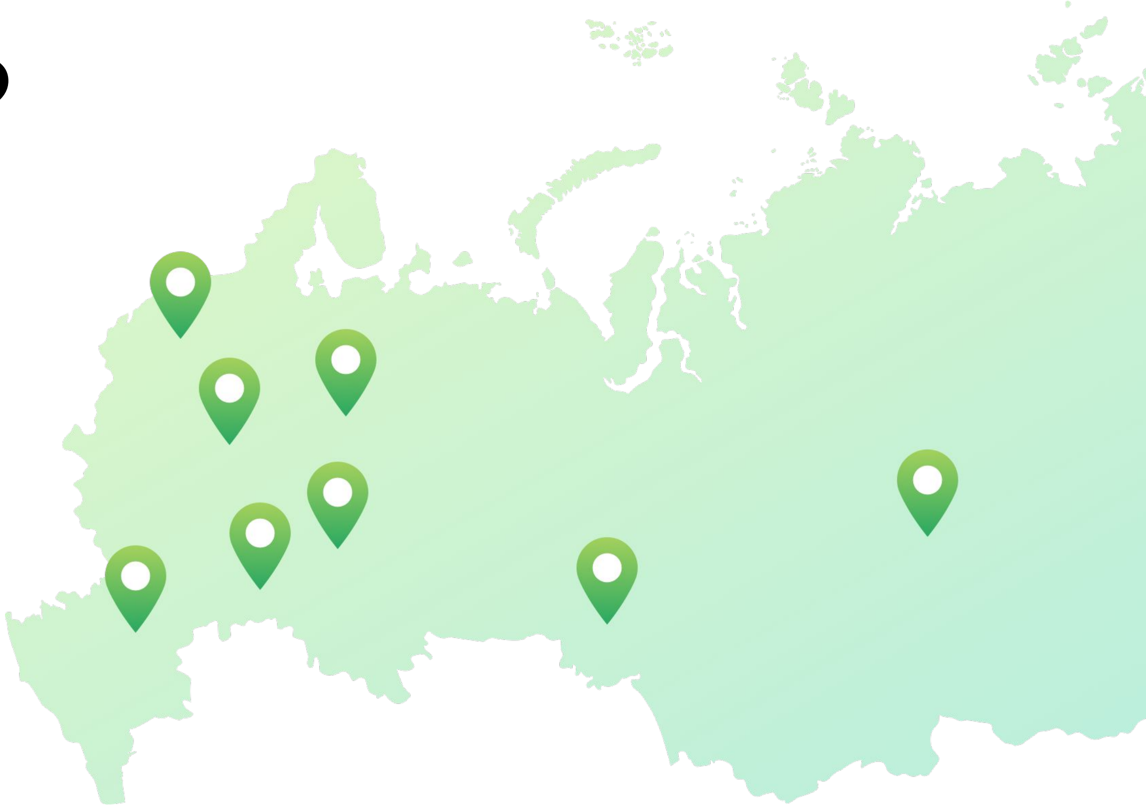
One way to increase conversion is to predict customers future basket and to offer it somehow

- Customer **doesn’t struggle with service interface**/ catalogue while collecting products one-after-another

- Customers loyalty (and retention) increases as they **feel that we know their taste and care about them**

- We are able to **communicate with customer in advance of his purchase**, delivering WOW-effect

What do we offer?



Problem:

- Using our dataset help us predict which products will customers order next time sorted by relevance

Tasks:

- Propose a solution (predictive model) that would allow you to **predict customers next purchase products sorted by relevance**

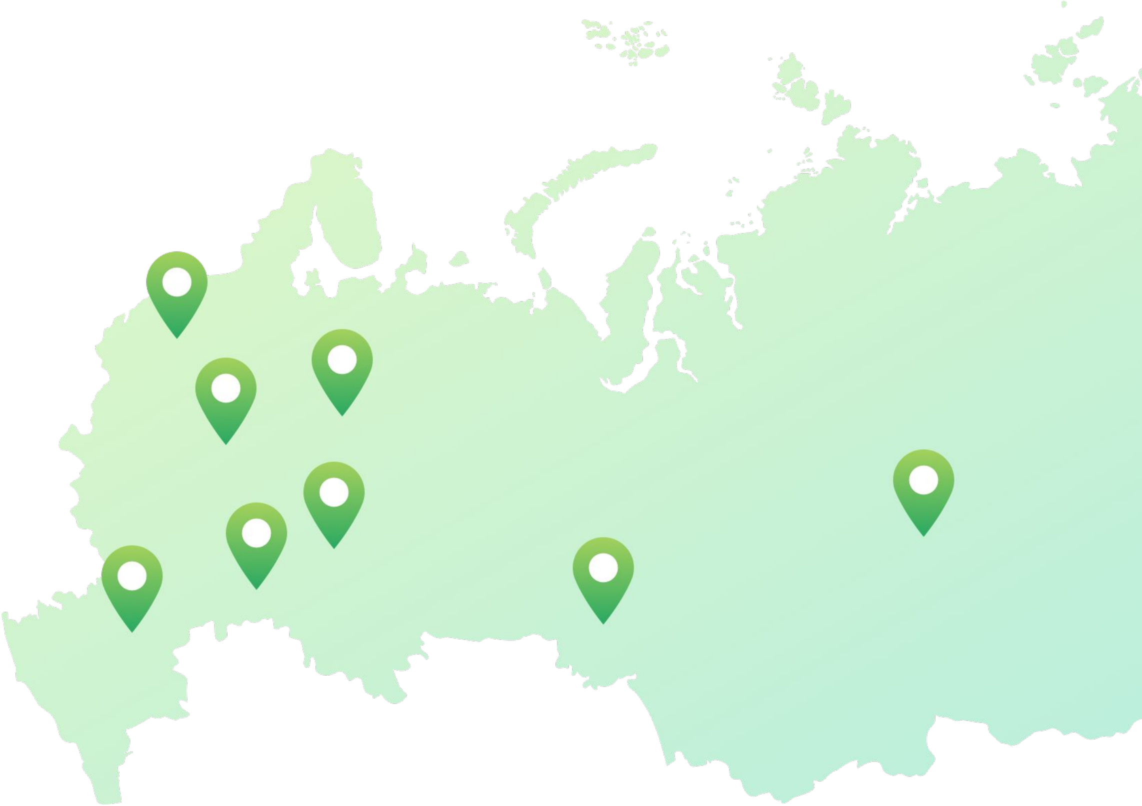
- What **business tasks** are being solved with the help of recommendations? Suggest **ways to change the MAP@k metric or supplement it with new metrics** in order to measure the performance of the model on historical data in relation to various business tasks

- While crunching data, prepare some ExploratoryDA and **share at least 5 your most interesting insights regarding clients/orders patterns**

How your solution will be evaluated?

Part of your solution	Metrics of evaluation	# of points
Quality of the model	MAP@k	**10**
Data insights	Jury decision	**5**
Any crazy business ideas on task/metric	Jury decision	**5**
Pitch & QA session	Jury decision	**5**

Data & Metrics



p@k

$$p@K(target, prediction) = \frac{1}{K}\sum_{k=1}^{K} r^{true}(\pi^{-1}(k)) = \frac{number\ of\ relevant}{K}$$

$r^{true}(e)$ can take values of 1 or 0 for relevant or irrelevant item e

$\pi^{-1}(k)$ represents the element with position k

p@k example

milk
bread
bananas

k = 6

algorithm 1

1	ice cream
2	tomatos
3	water
4	milk
5	bread
6	bananas

p@k = 3/6

algorithm 2

1	milk
2	bread
3	bananas
4	cookies
5	coffee
6	ice cream

p@k = 3/6

p@k example

milk
bread
bananas

k = 6

algorithm 1

1	ice cream
2	tomatos
3	water
4	milk
5	bread
6	bananas

p@k = 3/6

algorithm 2

1	milk
2	bread
3	bananas
4	cookies
5	coffee
6	ice cream

p@k = 3/6

AP@k

$$p@K(target, prediction) = \frac{1}{K}\sum_{k=1}^{K} r^{true}(\pi^{-1}(k)) = \frac{number\ of\ relevant}{K}$$

$$AP@K(target, prediction) = \frac{1}{min(K, M)}\sum_{k=1}^{K} r^{true}(\pi^{-1}(k)) \cdot p@k(target, prediction)$$

M - number of elements in real order

AP@k example

milk
bread
bananas

k = 6

algorithm 1

1	ice cream
2	tomatos
3	water
4	milk
5	bread
6	bananas

AP@k = 1/3 * (

0 * 0 +
0 * 0 +
0 * 0 +
1 * 1/4 +
1 * 2/5 +
1 * 3/6) = 0.383

$$AP@K(target, prediction) = \frac{1}{min(K, M)} \sum_{k=1}^{K} r^{true}(\pi^{-1}(k)) \cdot p@k(target, prediction)$$

AP@k example

milk
bread
bananas

k = 6

algorithm 2

1	milk
2	bread
3	bananas
4	cookies
5	coffee
6	ice cream

AP@k = 1/3 * (

1 *1/1 +
1 * 2/2 +
1 * 3/3 +
0 * 3/4 +
0 * 3/5 +
0 * 3/6) = 1

$$AP@K(target, prediction) = \frac{1}{min(K, M)} \sum_{k=1}^{K} r^{true}(\pi^{-1}(k)) \cdot p@k(target, prediction)$$

AP@k example

milk
bread
bananas

k = 6

algorithm 1

1	ice cream
2	tomatos
3	water
4	milk
5	bread
6	bananas

AP@k = 0.383

algorithm 2

1	milk
2	bread
3	bananas
4	cookies
5	coffee
6	ice cream

AP@k = 1

MAP@k

$$p@K(target, prediction) = \frac{1}{K}\sum_{k=1}^{K} r^{true}(\pi^{-1}(k)) = \frac{number\ of\ relevant}{K}$$

$$AP@K(target, prediction) = \frac{1}{min(K, M)}\sum_{k=1}^{K} r^{true}(\pi^{-1}(k)) \cdot p@k(target, prediction)$$

$$MAP@K(solution, submission) = \frac{1}{N}\sum_{n=1}^{N} AP@K_n(target_n, prediction_n)$$

Table 1 - orders

	user_id	order_id	order_created_time	retailer	store_id	platform
0	72	17431000	2020-09-26 10:48:57	METRO	21	app
1	83	9718154	2020-05-08 09:46:18	METRO	87	web
2	142	10056850	2020-05-14 15:06:03	METRO	320	app
3	187	15952443	2020-09-01 17:34:00	ВкусВилл	533	app
4	224	10409918	2020-05-20 06:32:50	Ашан	183	web

Table 2 - products

	user_id	order_id	line_item_id	price	quantity	discount	product_name	product_id	brand_name	master_category_id	parent_category_id
0	51	10717803	99293130	65.720001	1	0.0	Морковь мытая свежая	94333	Без бренда	85.0	84
1	51	10717803	99293227	127.330002	1	20.9	Помидоры	55133	Без бренда	85.0	84
2	51	10717803	99293243	99.900002	2	0.0	Помидоры черри Новиков 250 г	22035	Новиков	85.0	84
3	51	10717803	99293334	229.899994	1	0.0	Кукуруза сладкая в вакууме 230 г	6005183	Без бренда	85.0	84
4	51	10717803	99293366	69.900002	1	10.0	Бананы	709	Без бренда	91.0	90

Table 3 - categories

	id	name	parent_id
0	1	Продукты питания	0
1	2	Замороженные продукты	1
2	3	Замороженные овощи и фрукты	2
3	4	Замороженные полуфабрикаты	2
4	5	Рыба замороженная	2

Table 4 - user profiles

	user_id	gender	bdate
0	2224890	NaN	NaN
1	1683001	male	1987-10-11
2	2102480	NaN	NaN
3	2224895	NaN	NaN
4	930197	NaN	NaN

Table 5 - product properties

	product_id	proterty_name	property_value
0	1	Вес	100 г
1	2	Вид	Молочный напиток
2	2	Вес	100 г
3	2	Вкус	Черника
4	2	Сырье	Натуральное молоко

sbermarket.ru