



# КЛИЕНТСКАЯ АНАЛИТИКА ПО ДАННЫМ ИНТЕРНЕТ-МАГАЗИНА ЭЛЕКТРОНИКИ

Руководитель: Титова Наталия Николаевна

# ЦЕЛЬ И ЗАДАЧИ

## Цель:

Разделить клиентов на кластеры для более персонализированных предложений.

- 1 Исследование данных
- 2 Чистка данных
- 3 Сегментация клиентов

# НАША КОМАНДА



Кириянова Ольга Алексеевна  
DBSCAN, K-means



Бабак Артём Юрьевич  
Подготовка данных



Мэн Сыфэй  
RFM, K-means

# ДОРОЖНАЯ КАРТА ПРОЕКТА



Section 1

# ПОДГОТОВКА ДАННЫХ

# АНАЛИЗ ДАТАСЕТА

	Zero_Count	Unique_Count	NaN_Count	NaN_Percentage
Order_ID	0	166794	0	0.000000
Email_new	61516	99284	0	0.000000
Phone_new	7575	123135	0	0.000000
Source	0	4	0	0.000000
OrderDate	0	61	0	0.000000
время	0	58549	0	0.000000
месяц	0	2	0	0.000000
ChangeDate	0	155189	0	0.000000
DeliveryDate	0	100	0	0.000000
PaymentDate	0	105118	0	0.000000
Status	0	15	0	0.000000
Status_ID	0	15	0	0.000000
OneClick	326057	2	0	0.000000
CancelReason	0	34	230868	64.662387
Actions	0	41	261496	73.240794
DeliveryType	0	2	8285	2.320494
PaymentType	0	7	0	0.000000

Status	0	15	0	0.000000
Status_ID	0	15	0	0.000000
OneClick	326057	2	0	0.000000
CancelReason	0	34	230868	64.662387
Actions	0	41	261496	73.240794
DeliveryType	0	2	8285	2.320494
PaymentType	0	7	0	0.000000
Region	0	73	2	0.000560
Area	0	61	1631	0.456817
Store_ID	965	167	0	0.000000
FullSum	65	15995	20	0.005602
Discount	309994	1920	20	0.005602
IM_Rozn_Sum	5574	20622	20	0.005602
Row_ID	0	41	20	0.005602
Articul	0	26930	20	0.005602
Nom_Name	0	26888	58	0.016245
NomGroup	0	849	20	0.005602
Quant	5	43	20	0.005602
RowPrice	152336	9601	20	0.005602
RowDiscount	333347	2006	20	0.005602
RowSum	152813	12377	20	0.005602
Brand	0	1179	166141	46.533403
TN	0	17	166101	46.522199
TK	0	112	166101	46.522199
NomFullPath	0	878	20	0.005602
Week	0	9	20	0.005602
Nom_ID	3	26928	20	0.005602

# ЧИСТКА ДАННЫХ ЧАСТЬ 1

Строки с нулевыми номерами телефона удалены **(2716 строк)**

Чеки с отрицательными скидками удалены **(6 строк)**

Строки с нулевыми суммами удалены **(152815 строк)**

Строки с ошибками, дублированными и тестовыми заказами удалены **(30 строк)**

Отмененные заказы удалены **(70562 строк)**



# ЧИСТКА ДАННЫХ ЧАСТЬ 2

Заказы с скидками больше 90% удалены **(999 строк)**

Покупки повседневных товаров удалены **(7472 строк)**

Услуги удалены **(6849 строк)**

Аналогичные столбцы удалены **(17 столбцов)**

Некоторые строки TN и ТК восстановлены **(7222 строк)**



# СРАВНЕНИЕ ДАТАСЕТА ДО И ПОСЛЕ ЧИСТКИ ДАННЫХ

## ДО

- 357036 строк, 37 столбцов

## ПОСЛЕ

- 88147 строк, 34 столбцов

# РАЗМЕЩЕННЫЕ ЗАКАЗЫ

Месяц	Сумма_заказов	Кол_во_заказов	Кол_во_уникальных_клиентов	Средний_чек	Среднее_кол_во_товаров
Март	1,257,011,053	85,678	54,301	12,372	1
Апрель	1,197,676,943	79,132	47,603	12,546	1

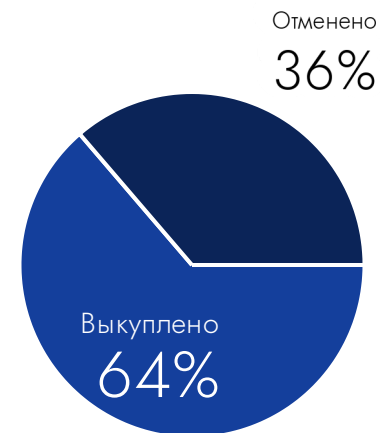
# ВЫКУПЛЕННЫЕ ЗАКАЗЫ

Месяц	Сумма_заказов	Кол_во_заказов	Кол_во_уникальных_клиентов	Средний_чек	Среднее_кол_во_товаров
Март	675,142,693	54,579	38,595	10,440	1
Апрель	652,107,969	51,372	34,370	10,546	1

# ПРОЦЕНТ ВЫКУПА



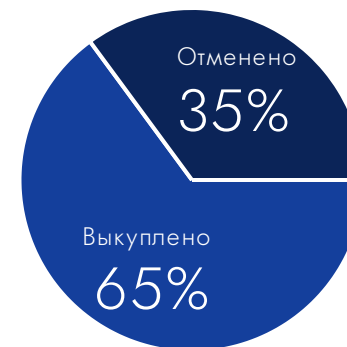
Март: Выкупленная сумма



Март: Выкупленное количество



Апрель: Выкупленная сумма



Апрель: Выкупленное количество

Section 2

# RFM СЕГМЕНТАЦИЯ

# RFM СЕГМЕНТАЦИЯ

## Recency

*Как давно клиенты приобретали товар?*

- "LastOrderDate"

## Frequency

*Как часто клиенты покупают?*

- "Кол-во чеков"




## Monetary

*Общая сумма трат*




- "Общая сумма чеков".

# ПРЕИМУЩЕСТВА И НЕДОСТАТКИ МЕТОДА

## Преимущества

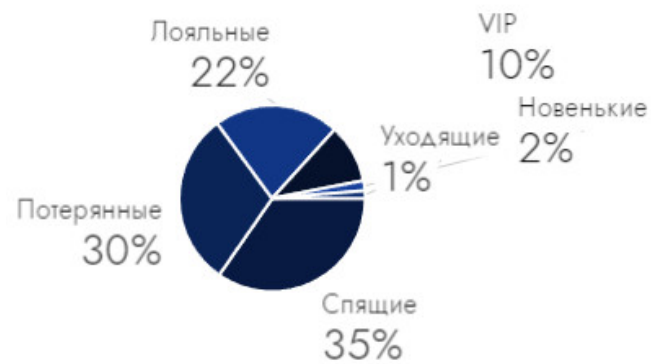
-  Просто и точно определяет целевую аудиторию
-  Позволяет снизить затраты на рекламные кампании и привлечение новых клиентов
-  Уменьшает отток клиентов

## Недостатки

-  Требуется дополнительных исследований для определения причин оттока клиентов
-  Не подходит для компаний с разовыми продажами.
-  Анализ необходимо проводить регулярно



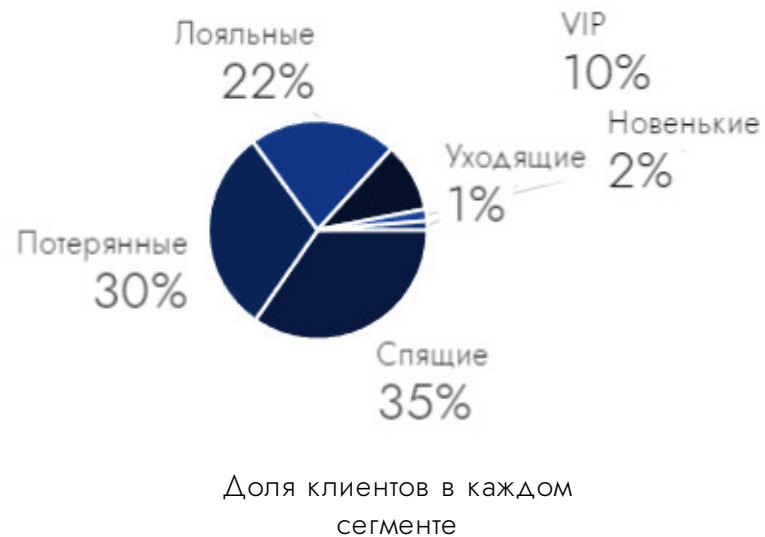
# АНАЛИЗ СЕГМЕНТОВ (ПО ТОВАРАМ %)



Segment	Авто- товары	Акces- суары	Детские товары	Игры, софт и раз- влечения	Климат	Крупная бытовая техника	Мебель	Мелкая бытовая техника	Сделай сам	Спорт и активный отдых	ТВ- Аудио	Товары для дома	Цифро- вая Техника
VIP	0	3	6	0	2	38	2	11	8	8	10	1	11
Лояльные	0	7	9	0	2	16	5	25	13	9	4	3	7
Новенькие	0	14	7	1	2	2	7	34	9	11	2	7	4
Потерянные	0	8	6	0	2	15	6	22	8	6	6	4	16
Спящие	0	10	7	1	2	14	7	18	10	10	6	4	11
Уходящие	0	18	5	0	1	12	7	21	8	7	4	8	10

VIP клиенты - любители крупной бытовой техники, остальные клиенты - любители мелкой бытовой техники.

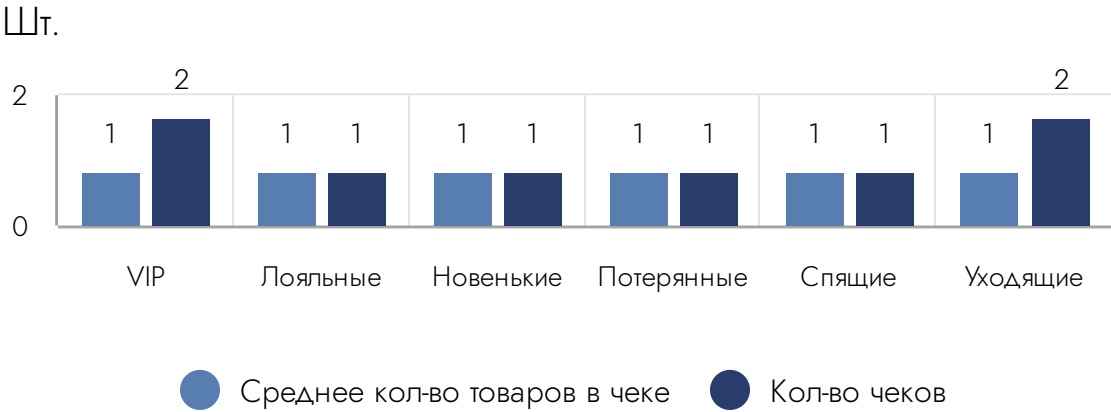
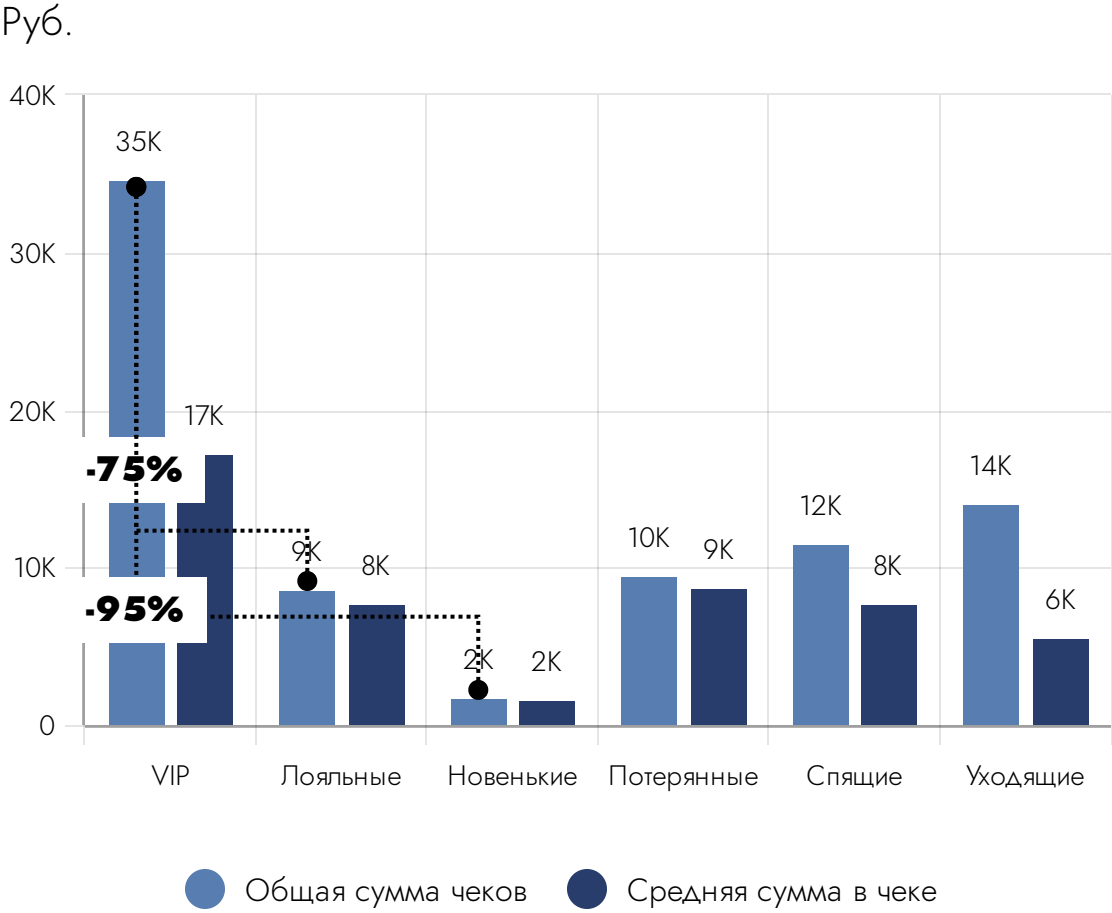
# АНАЛИЗ СЕГМЕНТОВ (ПО ФЕДЕРАЛЬНЫМ ОКРУГАМ %)



Segment	ДФО (Дальне- восточный)	ПФО (Привол- жский)	СЗФО (Северо- Западный)	СФО (Сибир- ский)	УФО (Ураль- ский)	ЦФО (Центра- льный)	ЮФО (Южный)
VIP	0.1	10.3	29.4	3.7	3.4	50.3	2.9
Лояль- ные	0.0	10.9	24.7	4.4	2.7	53.9	3.3
Новень- кие	0.2	10.4	23.8	5.3	2.7	54.1	3.5
Потерян- ные	0.0	10.5	23.5	4.8	2.9	55.2	3.2
Спящие	0.0	11.0	24.8	4.6	2.8	53.1	3.7
Уходя- щие	0.0	11.5	26.1	3.0	2.6	53.3	3.5

Замечание: Большинство клиентов покупали в центральном ФО

# АНАЛИЗ СЕГМЕНТОВ ПО ЧЕКАМ



# ПРОФИЛИРОВАНИЕ СЕГМЕНТОВ



## VIP

- Покупал 8 дней назад, в среднем совершил 2 покупки и потратил 35 тыс. руб.



## Спящие

- Покупал 33 дней назад, совершил 1 покупку и потратил 11.5 тыс. руб.



## Лояльные

- Покупал 13 дней назад, в среднем совершил 1 покупку и потратил 8.6 тыс. руб.



## Уходящие

- Покупал 50 дней назад, в среднем совершил 2 покупки и потратил 14 тыс. руб.



## Новенькие

- Покупал 18 дней назад, в среднем совершил 1 покупку и потратил 1.7 тыс. руб.



## Потерянные

- Покупал 52 дней назад, совершил 1 покупку и потратил 9.5 тыс. руб.

# ПРЕДЛОЖЕНИЯ ДЛЯ КЛИЕНТОВ



## **VIP**

- Награждать их. Сделать из них фокус для нового продукта



## **Лояльные**

- Продать им более дорогие товары. Просить оставить отзыв.



## **Новенькие**

- Создать устойчивое понимание магазина. Дать скидку на первую крупную трату.



## **Спящие**

- Предложить ограниченные по времени оферы. Рекомендовать товары на основе прошлых покупок.



## **Уходящие**

- Делиться с ними полезной информацией. Рекомендовать популярные товары. Предложить скидки и промокоды.



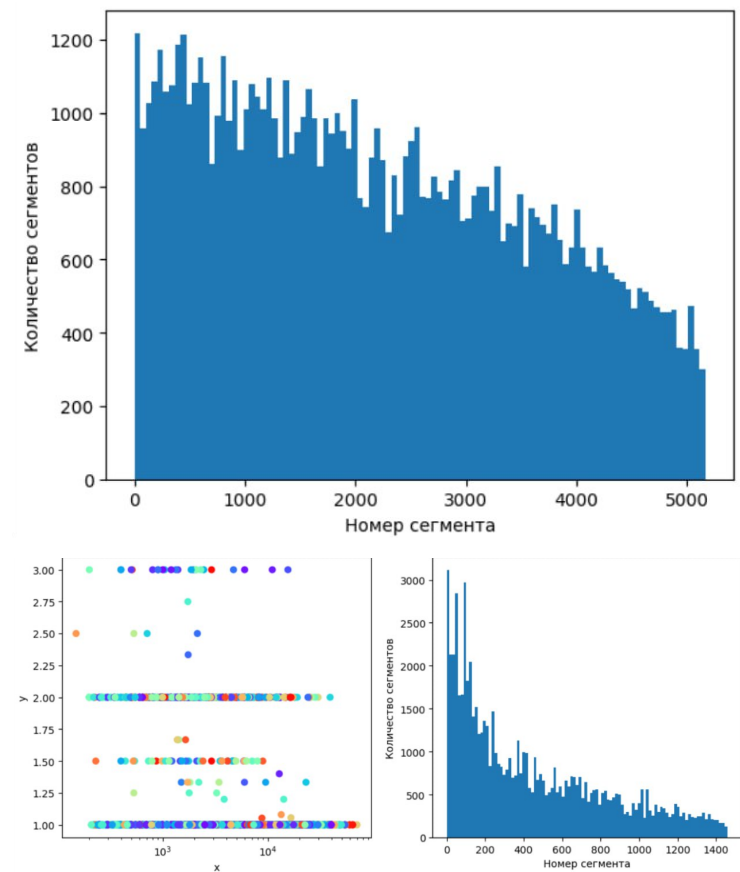
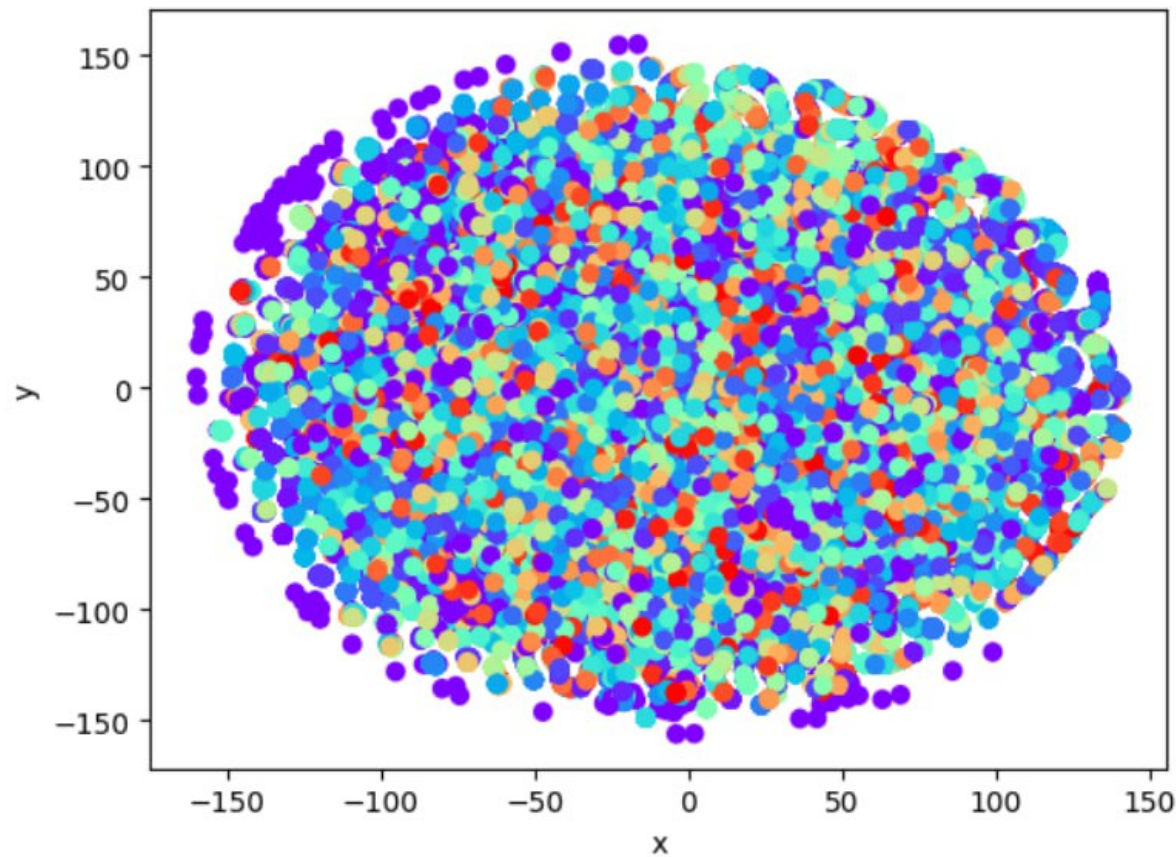
## **Потерянные**

- Отправить персонализированные имейлы, чтобы возобновить связь.

Section 3

# DBSCAN СЕГМЕНТАЦИЯ

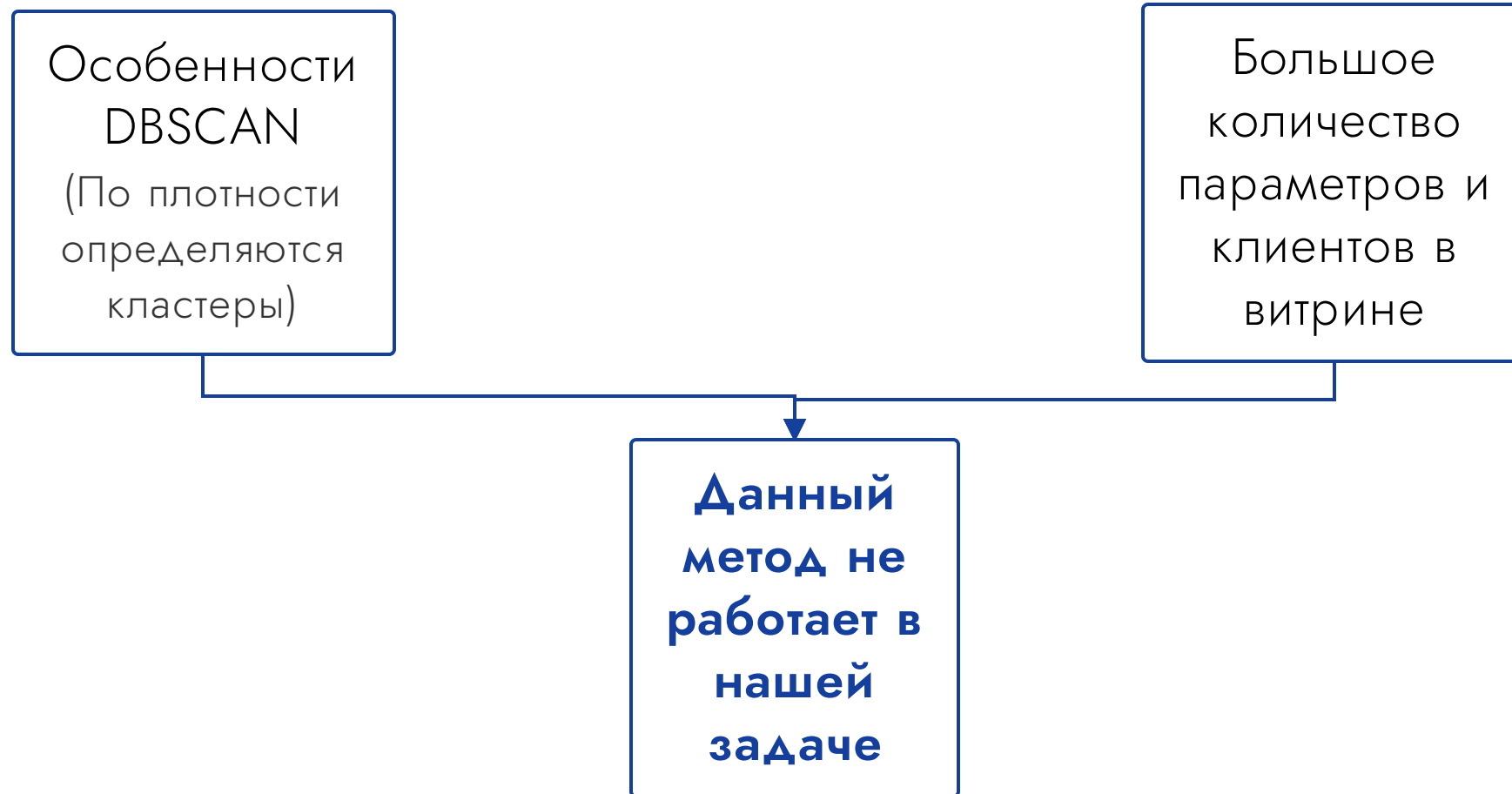
# РЕЗУЛЬТАТЫ DBSCAN



Замечание: Данные очень плотные



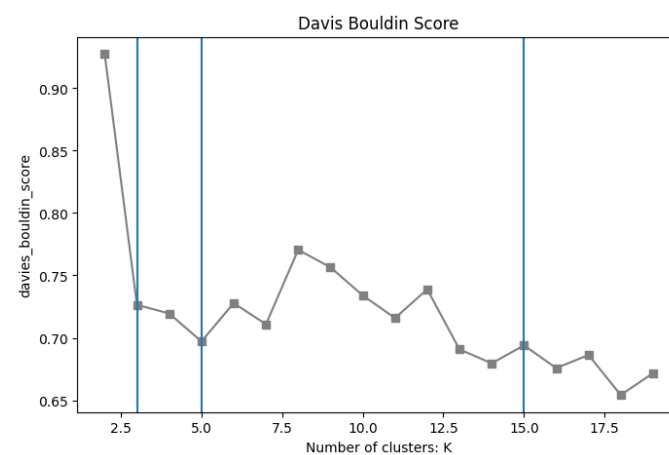
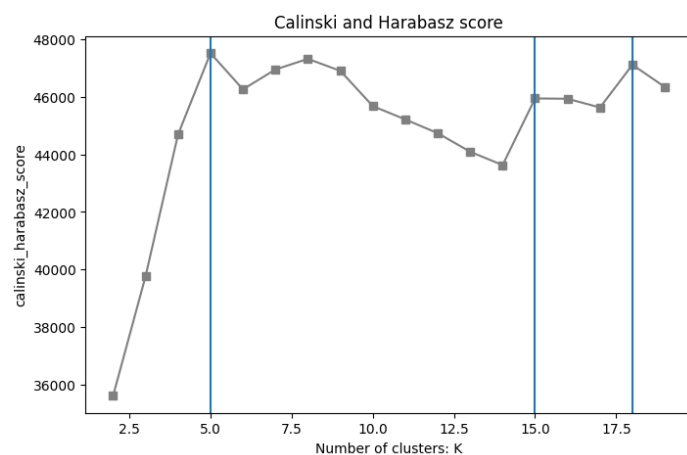
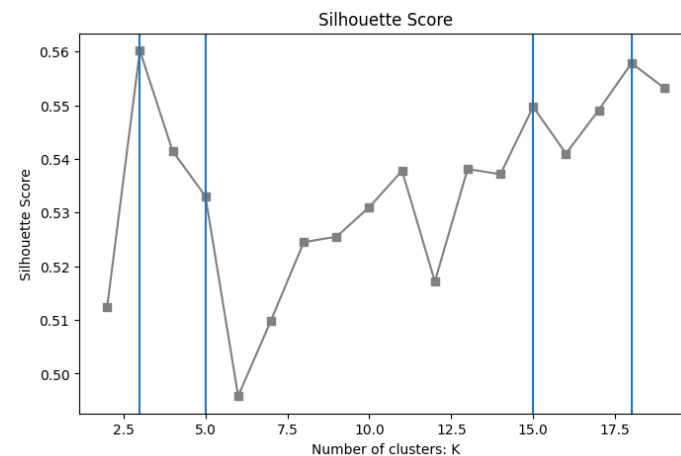
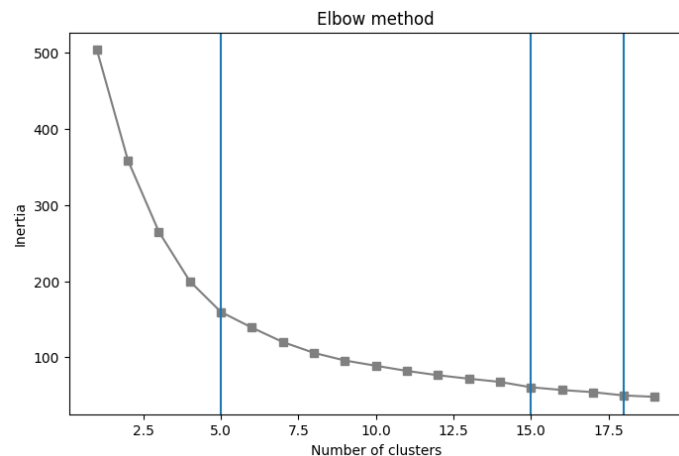
# ПОЧЕМУ НЕ ПОДОШЁЛ DBSCAN



Section 4

# KMEANS СЕГМЕНТАЦИЯ

# ОПРЕДЕЛЕНИЕ КОЛИЧЕСТВА КЛАСТЕРОВ

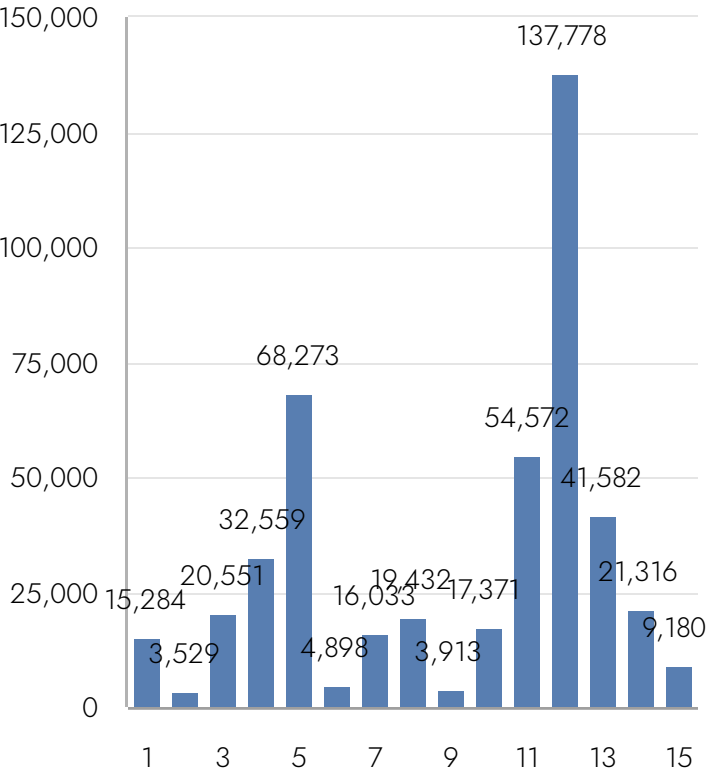


Количество кластеров = 15

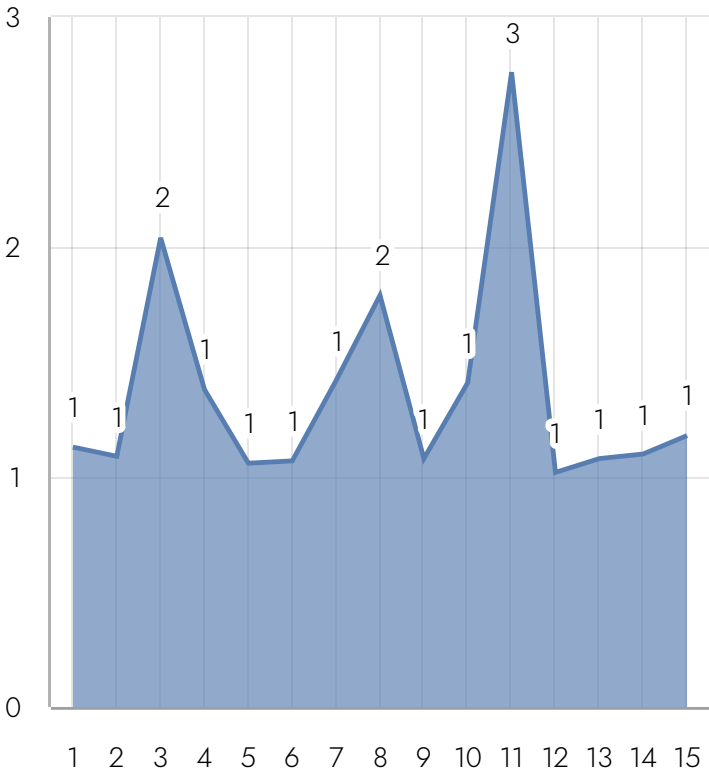
# АНАЛИЗ СЕГМЕНТОВ ПО ФЕДЕРАЛЬНЫМ ОКРУГАМ

	Дальневосточный		Приволжский		Северо-Западный		Сибирский		Уральский		Центральный		Южный	
Segment	Area_ДФО		Area_ПФО		Area_СЗФО		Area_СФО		Area_УФО		Area_ЦФО		Area_ЮФО	
1		0,00%		12,59%		24,51%		5,51%		3,53%		50,69%		3,17%
2		0,05%		9,97%		25,50%		4,45%		2,52%		54,19%		3,32%
3		0,00%		9,43%		19,30%		3,73%		2,63%		61,40%		3,51%
4		0,02%		12,94%		23,79%		5,53%		5,41%		48,36%		3,96%
5		0,15%		10,64%		24,82%		5,17%		1,92%		54,51%		2,81%
6		0,02%		7,62%		22,10%		3,57%		1,63%		62,24%		2,82%
7		0,05%		4,83%		19,58%		3,44%		2,21%		66,87%		3,01%
8		0,19%		12,23%		24,39%		5,17%		4,16%		50,14%		3,72%
9		0,00%		12,67%		25,63%		5,04%		3,78%		49,55%		3,32%
10		0,07%		11,48%		28,06%		4,35%		2,45%		50,45%		3,15%
11		0,00%		1,84%		22,70%		2,84%		1,67%		65,61%		5,34%
12		1,56%		12,50%		12,50%		1,56%		3,13%		59,38%		9,38%
13		0,07%		11,56%		24,81%		4,36%		3,23%		52,20%		3,77%
14		0,00%		8,62%		24,34%		3,63%		2,03%		58,47%		2,90%
15		0,00%		10,10%		25,63%		4,58%		2,43%		54,17%		3,09%

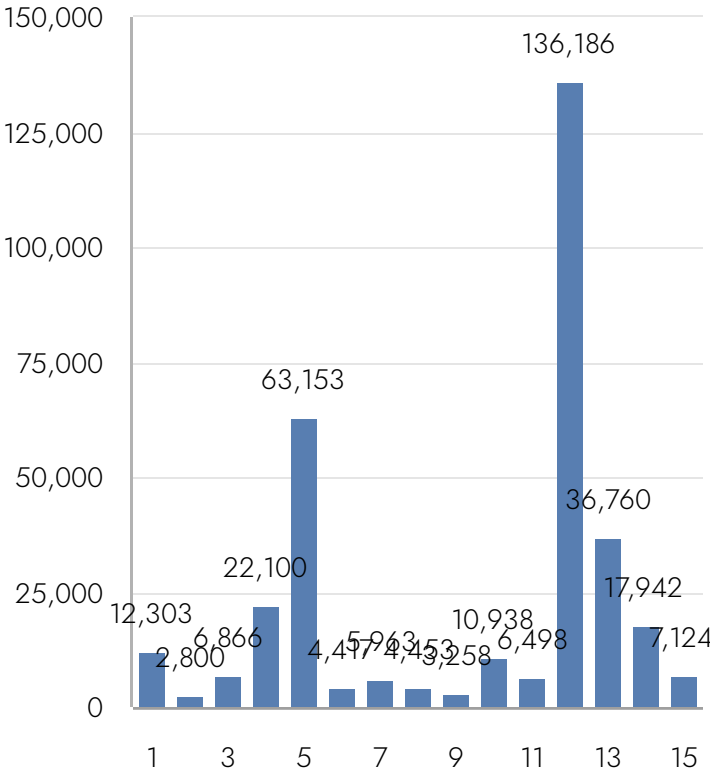
# АНАЛИЗ СЕГМЕНТОВ ПО ЧЕКАМ



● Общая сумма чеков

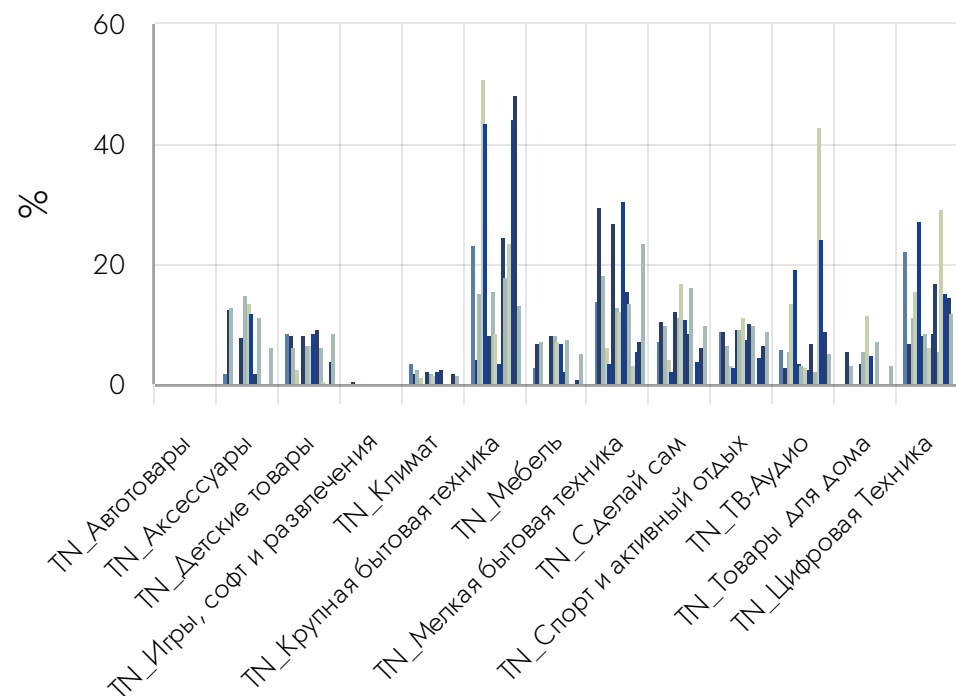


● Кол-во чеков



● Средняя сумма в чеке

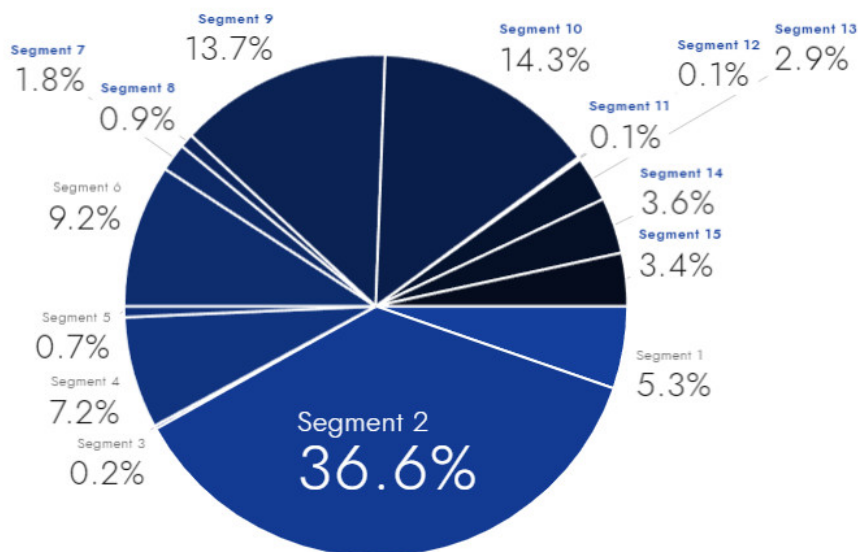
# АНАЛИЗ СЕГМЕНТОВ ПО ТОВАРАМ (%)



Segment	Автотовары	Аксессуары	Детские товары	Игры, софт и развлечения	Климат	Крупная бытовая техника	Мебель	Мелкая бытовая техника	Сделай сам	Спорт и активный отдых	ТВ-Аудио	Товары для дома	Цифровая Техника
1	0.00	1.90	8.80	0.10	3.60	23.40	3.10	13.90	7.50	8.90	6.10	0.40	22.20
2	0.10	12.80	8.50	0.60	1.90	4.20	6.90	29.60	10.70	8.90	3.10	5.70	6.90
3	0.00	13.00	6.20	0.00	2.80	15.30	7.40	18.20	10.10	6.70	5.70	3.30	11.20
4	0.00	0.50	2.80	0.10	1.40	50.90	0.50	6.30	4.30	3.50	13.80	0.10	15.80
5	0.00	0.40	0.00	0.00	0.00	43.60	0.20	3.80	2.30	3.00	19.20	0.00	27.40
6	0.00	7.90	8.20	0.40	2.50	8.30	8.20	27.10	12.40	9.30	3.80	3.70	8.20
7	0.00	15.10	6.60	0.20	2.10	15.80	8.30	13.10	11.30	9.50	3.50	5.70	8.60
8	0.00	13.80	6.60	0.20	1.80	8.80	7.30	12.40	17.00	11.20	3.00	11.80	6.20
9	0.10	12.10	8.60	0.50	2.40	3.70	6.90	30.60	10.90	7.60	2.80	5.10	8.70
10	0.00	1.90	9.30	0.00	2.80	24.60	2.40	15.80	8.60	10.20	7.00	0.40	16.90
11	0.00	11.50	6.30	0.40	0.30	18.10	7.60	13.70	16.40	9.90	2.40	7.50	5.80
12	0.00	0.00	0.80	0.00	0.00	23.80	0.00	3.20	0.00	0.00	42.90	0.00	29.40
13	0.00	0.30	0.20	0.30	0.30	44.40	0.10	5.80	4.00	4.80	24.40	0.00	15.40
14	0.00	0.50	3.90	0.00	1.90	48.40	1.10	7.30	6.30	6.60	9.00	0.10	14.80
15	0.00	6.50	8.70	0.50	1.70	13.40	5.50	23.80	9.90	9.10	5.20	3.50	12.00



# ОПИШЕМ ПОТРЕБИТЕЛЕЙ



1

## Обновление техники

- Средний чек 15300р.
- Покупают мелкую бытовую и при этом предпочитают безналичную оплату.

2

## Повседневные покупки

- Самый многочисленный сегмент со средним чеком 2800р.
- Покупают: мелкую бытовую технику, аксессуары, сделай сам
- Небольшая единичная покупка.

3

## Должники с повседневными покупками

- 6800р.
- Покупки совершены через сайт, оплата наличными, при этом много кредитов по сравнению с остальным, покупают различные категории товаров.

4

## Переезд. Необходимость в новой технике

- 22100р средний чек.
- Покупают через сайт, наличные. Категория товара: крупная бытовая техника.

5

## Должники с крупной бытовой техникой.

- Средний чек 63000р.
- Преимущественно покупают через онлайн резервирование, при этом много в кредит. Категории: крупная бытовая техника, цифровая.



# ОПИШЕМ ПОТРЕБИТЕЛЕЙ

## 6 Возрастное поколение - любители аксессуаров

- 4400р средний чек.
- В основном категория аксессуары, при этом покупка через кол-центр, оплата наличными.

## 7 Возрастное поколение - core-сегмент

- Средний чек 5963р.
- Покупки через кол-центр, наличная оплата. Покупают все категории.

## 8 Молодое поколение - core-сегмент

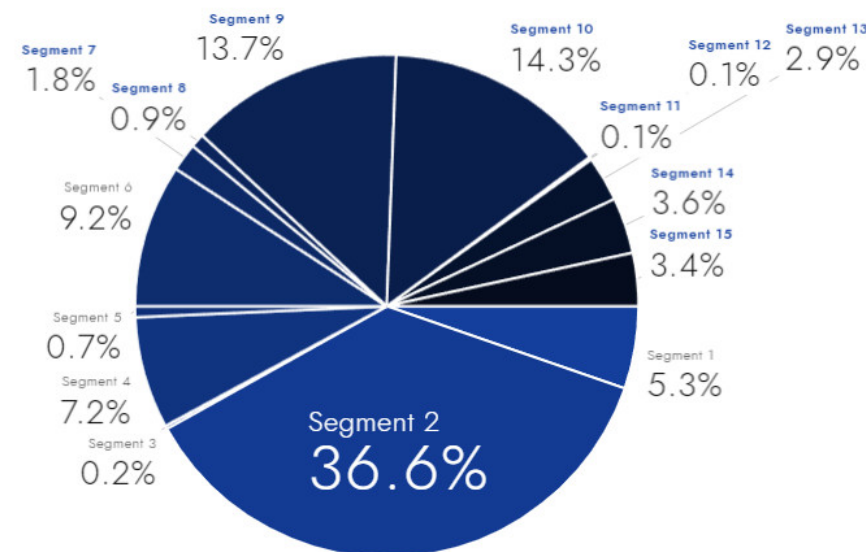
- Средний чек 4452р.
- Онлайн резервирование, безналичная оплата, при этом мелкие регулярные покупки, в среднем разные товары.

## 9 Молодое поколение - любители аксессуаров

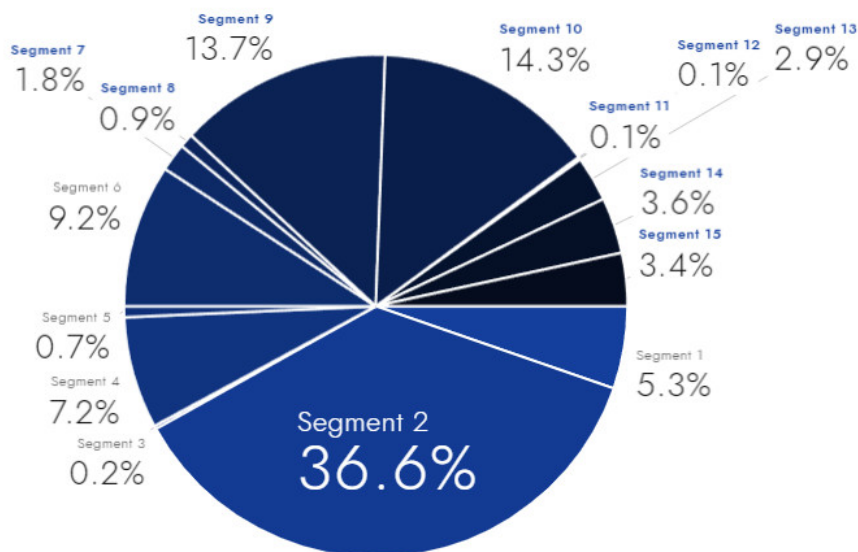
- Средний чек 3258р.
- Преобладает безналичная оплата, в основном покупают категорию аксессуары.

## 10 Любители крупной бытовой техники

- Средний чек 10000р.
- Обдуманные, серьезные покупки. Крупная бытовая техника, цифровая техника, при этом оплата в основном наличными средствами.



# ОПИШЕМ ПОТРЕБИТЕЛЕЙ



## 11 Оптовики повседневных покупок

- Средний чек 6500р.
- Заказы через кол-центр, оплата наличными, присутствует немного кредита, по категории покупают все подряд, но при этом что интересно - оптом.

## 12 Богатые покупатели цифровой техники

- Средний чек 130000р.
- Особенности: онлайн резервирование, есть кредит, но при этом их мало. Предпочитают покупать цифровую и аудио технику.

## 13 Должники с бытовой техникой

- Средний чек 36759р.
- Способ оплаты: онлайн резервирование, наличная в основном оплата и есть кредит в магазине. Категории товаров: крупная бытовая, аудиотехника, цифровая техника.

## 14 Покупатели, делающие ремонт





- Средний чек 17941р.
- Заказы через кол-центр, наличная оплата, преобладает категория крупная бытовая техника.

## 15 Кор сегмент без развлечений/серьезные покупатели





- Средний чек 7124р.
- Заказ через сайт, оплата наличными средствами, мелкая бытовая техника, покупают в основном все, кроме развлечений.

# СРАВНЕНИЕ МЕТОДОВ СЕГМЕНТАЦИИ

## K-means

-  Простота реализации
-  Разграничение не вручную на классы
-  Вариация кол-ва разделяемых классов
-  Не всегда равночисленное разделение на группы при большом наличии кластеров

## RFM

-  Наглядность, простота и интуитивность
-  Полученный результат легко интерпретируем
-  Мало рассматриваемых признаков
-  Анализ необходимо проводить несколько раз в квартал

# ЗАКЛЮЧЕНИЕ

