

# PRESENTATION

Бу ЭНД

Оспанов Әділхан и Қазбек Ілияс

Экономика и Наука о данных

Максут Нарикбаев универсиет

# OUTLINE

1. Введение
2. Методология
3. Описание предобработки
4. Обработка данных
5. Графовые нейронные сети
6. Результаты моделей
7. Описание кластеров от HDBSCAN
8. Описание кластеров от DBSCAN
9. Заключение

# ВВЕДЕНИЕ

Данная презентация это short-summary нашей работы. Со всем кодом, с комментариями и ходом наших мыслей (черновыми записями), можно ознакомиться в !runb файле (ЭНД\_team).

# МЕТОДОЛОГИЯ

Кейс решался через глубокую предобработку транзакционных данных, включающую извлечение временных признаков, агрегацию поведенческих метрик и категориальных трат по клиентам. Это позволило создать детальный профиль каждого пользователя.

Для выявления скрытых связей и создания более информативных представлений клиентов, был построен граф, а затем использована графовая нейронная сеть (GNN), в частности Graph Autoencoder с GATConv, для генерации высококачественных эмбедингов.

На основе этих эмбедингов проведена кластеризация с помощью DBSCAN и HDBSCAN, выявившая 4-5 четких сегментов. Каждый сегмент был детально проинтерпретирован по всем аспектам поведения, что дало ценные инсайты для персонализации продуктов и маркетинговых стратегий.

# ОПИСАНИЕ ПРЕДОБРАБОТКИ

## СОЗДАНИЕ ПОВЕДЕНЧЕСКИХ ФИЧЕЙ НА ОСНОВЕ ТРАНЗАКЦИОННОЙ АКТИВНОСТИ КЛИЕНТОВ

Извлечение временных признаков: День недели (day\_name),  
Месяц (month\_name), Час транзакции (hour)

1. Формирование поведенческих метрик: Среднее время между транзакциями (avg\_time\_between\_txn), Доля зарубежных операций (share\_foreign\_txn), Доля операций с конверсией валют (share\_converted\_txn)
2. Агрегация категориальных признаков: wallet\_type, pos\_entry\_mode, transaction\_type, merchant\_city, mcc\_category; временные: hour, day\_name, month\_name

# ОПИСАНИЕ ОБРАБОТКИ

## ЧИСТКА ДАННЫХ

1. Заполнение NA с помощью среднего
2. Лог трансформация все числовых переменных

### Почему выбрали среднее?

Количество пропусков (NaN) было незначительным и встречалось преимущественно в производных переменных одного типа. При большей доле пропусков мы бы применили более продвинутые методы, такие как MICE, а при доле  $>50\%$  — удалили бы переменную.

# ОПИСАНИЕ ОБРАБОТКИ

## ЧИСТКА ДАННЫХ

### Почему лог трансформация?

Мы протестировали несколько методов масштабирования признаков, включая StandardScaler, RobustScaler и комбинацию логарифмического преобразования с последующим стандартным масштабированием (Log + StandardScaler). Наилучшие результаты по всем ключевым метрикам показала именно логарифмическая трансформация. Кроме того, все переменные в датасете неотрицательные ( $\geq 0$ ), что позволяет применять лог-преобразование без потерь и необходимости дополнительных преобразований.

# ГРАФОВЫЕ НЕЙРОННЫЕ СЕТИ

## Почему мы используем графовые нейронные сети (GNN) для сегментации?

Мы рассматриваем поведение клиентов не только через индивидуальные признаки, но и через сходство в транзакционном поведении, что позволяет представить данные в виде графа: Узлы (nodes): клиенты; Рёбра (edges): наличие поведенческого сходства (например, через FAISS-контакты на основе kNN). Для расчета наиболее подходящих гиперпараметров, мы использовали циклы, и сравнивали каждую модель по метрикам. Также, мы фильтровали модели, у которых количество кластеров было меньше 3.

### GNN с GAE (Graph Autoencoder)

- GNN позволяет учитывать как собственные характеристики клиента, так и поведение его «похожих соседей»;
- Мы используем Graph Autoencoder (GAE) с GATConv, чтобы получить латентные векторные представления клиентов;
- Эти эмбединги отражают как личные, так и структурные поведенческие особенности.



# ГРАФОВЫЕ НЕЙРОННЫЕ СЕТИ

Почему мы применяем DBSCAN, HDBSCAN и Spectral Clustering?

## DBSCAN

- Не требует задавать число кластеров заранее;
- Устойчив к выбросам;
- Идеален для кластеров произвольной формы и разной плотности — а именно такие группы могут быть среди клиентов.

## HDBSCAN

- Улучшение DBSCAN:
- Строит иерархическую структуру кластеров;
- Лучше справляется с кластеризацией при сильно неравномерной плотности;
- Выделяет наиболее стабильные кластеры;
- Автоматически помечает шум (клиентов без устойчивых паттернов).
- Особенно полезен при UMAP/GAE эмбедингах, где плотность может быть неравномерной.

# ГРАФОВЫЕ НЕЙРОННЫЕ СЕТИ

Почему мы применяем DBSCAN, HDBSCAN и Spectral Clustering?

## Spectral Clustering

- Основан на графовой лапласианной матрице, что идеально дополняет использование GNN;
- Позволяет выявить сложные, нелинейные структуры кластеров, которые невозможно уловить простыми методами (например, KMeans);
- Отлично работает после снижения размерности (PCA/UMAP), т.к. опирается на глобальные структуры данных.

# РЕЗУЛЬТАТЫ МОДЕЛЕЙ

## Метрики

Silhouette Score (чем выше, тем лучше)

Calinski-Harabasz Index (чем выше, тем лучше)

Davies-Bouldin Index (чем ниже, тем лучше)

## HDBSCAN, Цель: Минимизировать Davies-Bouldin

🔧 embedding\_dim = 64

💧 dropout = 0.2

👤 min\_cluster\_size = 30

📈 Calinski-Harabasz = 17,143.60

📉 Davies-Bouldin = 0.259

📊 Silhouette Score = 0.553

# РЕЗУЛЬТАТЫ МОДЕЛЕЙ

## DBSCAN, Цель: Максимизировать Silhouette Score

🔧 embedding\_dim = 64

💧 dropout = 0.2

📏 eps = 0.5

👤 min\_samples = 15

📈 Calinski-Harabasz =  $6.42 \times 10^{15}$

📈 Davies-Bouldin =  $2.28 \times 10^{-7}$

📈 Silhouette Score = 1.000 ✅

## Spectral Clustering, Цель: Максимизировать Silhouette Score

🔧 embedding\_dim = 32

💧 dropout = 0.3

📊 n\_clusters = 10

📈 Calinski-Harabasz = 24.32

📈 Davies-Bouldin = 3.533

📈 Silhouette Score = 0.188

# РЕЗУЛЬТАТЫ МОДЕЛЕЙ

- **DBSCAN** показывает идеальную силу кластеров ( $Silhouette = 1.0$ ) и минимальную ошибку сегментации.
- **HDBSCAN** — более интерпретируем и устойчив к шуму, демонстрирует хороший баланс всех метрик.
- **Spectral Clustering** уступает по всем метрикам, но всё ещё может быть полезен для выявления нелинейных структур.

**По этой причине для дальнейшего анализа мы будем использовать кластеры, полученные с помощью HDBSCAN и DBSCAN.**

# ОПИСАНИЕ СЕГМЕНТОВ

Кластер со значением (-1) — это шум или выбросы в данных.

DBSCAN

Кластер	Количество	Доля
0	1795	0.8975
-1	165	0.08
1	22	0.011
2	18	0.009

HDBSCAN

Кластер	Количество	Доля
-1	1232	0.618
0	36	0.018
1	39	0.0195
2	643	0.3215
3	50	0.025

# ОПИСАНИЕ СЕГМЕНТОВ

DBSCAN продемонстрировал значительно меньше выбросов (значений с кластером -1), чем HDBSCAN.

DBSCAN

**0,08**

HDBSCAN

**0,618**

**<**

# ОПИСАНИЕ СЕГМЕНТОВ HDBSCAN

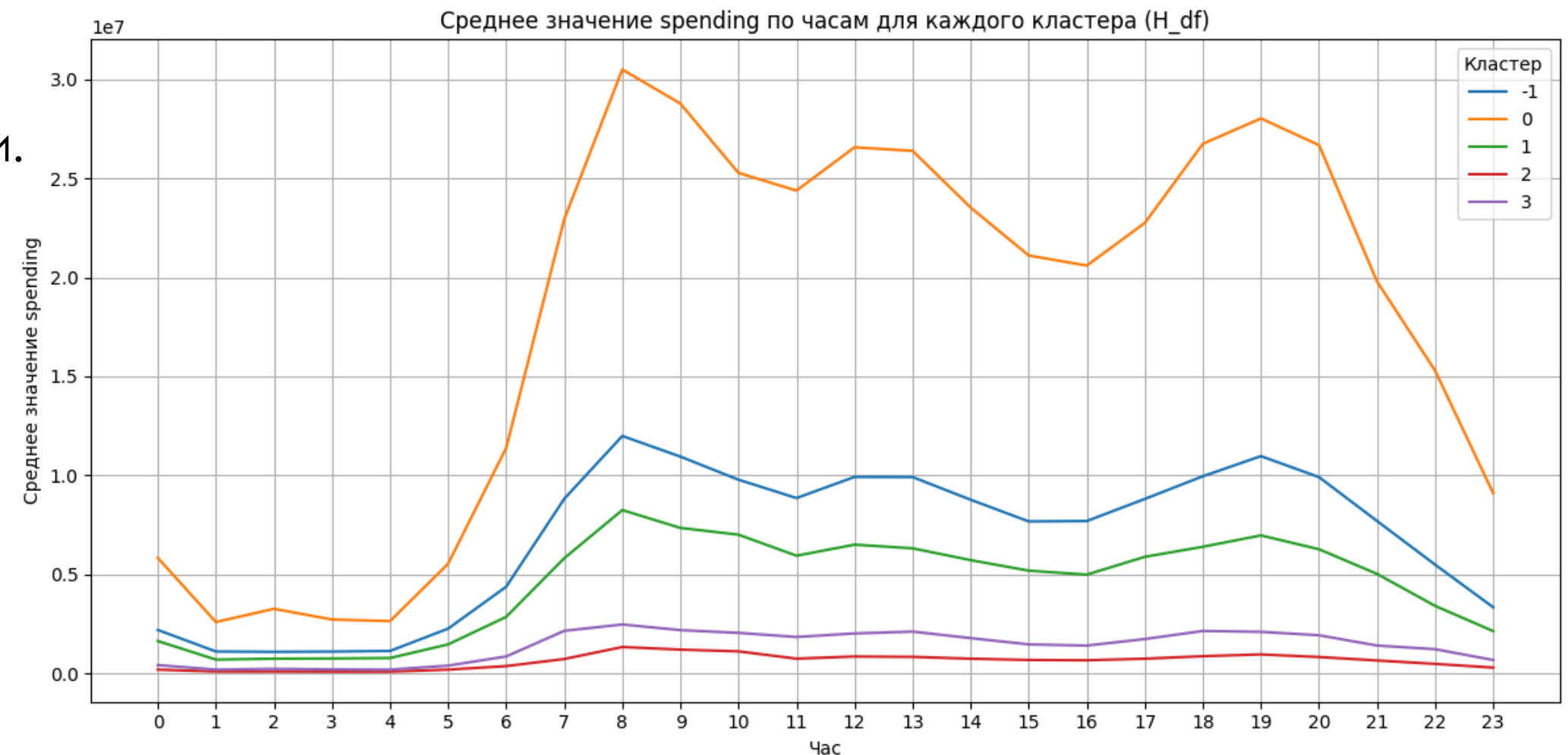
**Шум (кластер -1):** Атипичное поведение, равномерные расходы. Возможны технические записи или сбои.

**Кластер 0:** Клиенты с высокими расходами. Пик трат 8:00–15:00. Вероятно, премиум/корпоративные.

**Кластер 1:** Активные пользователи с умеренными расходами. Расходы распределены, пик 8:00–14:00. Розничные клиенты.

**Кластер 2:** Клиенты с низкими расходами. Основная активность 8:00–13:00. Возможно, молодые или с ограниченным использованием карты.

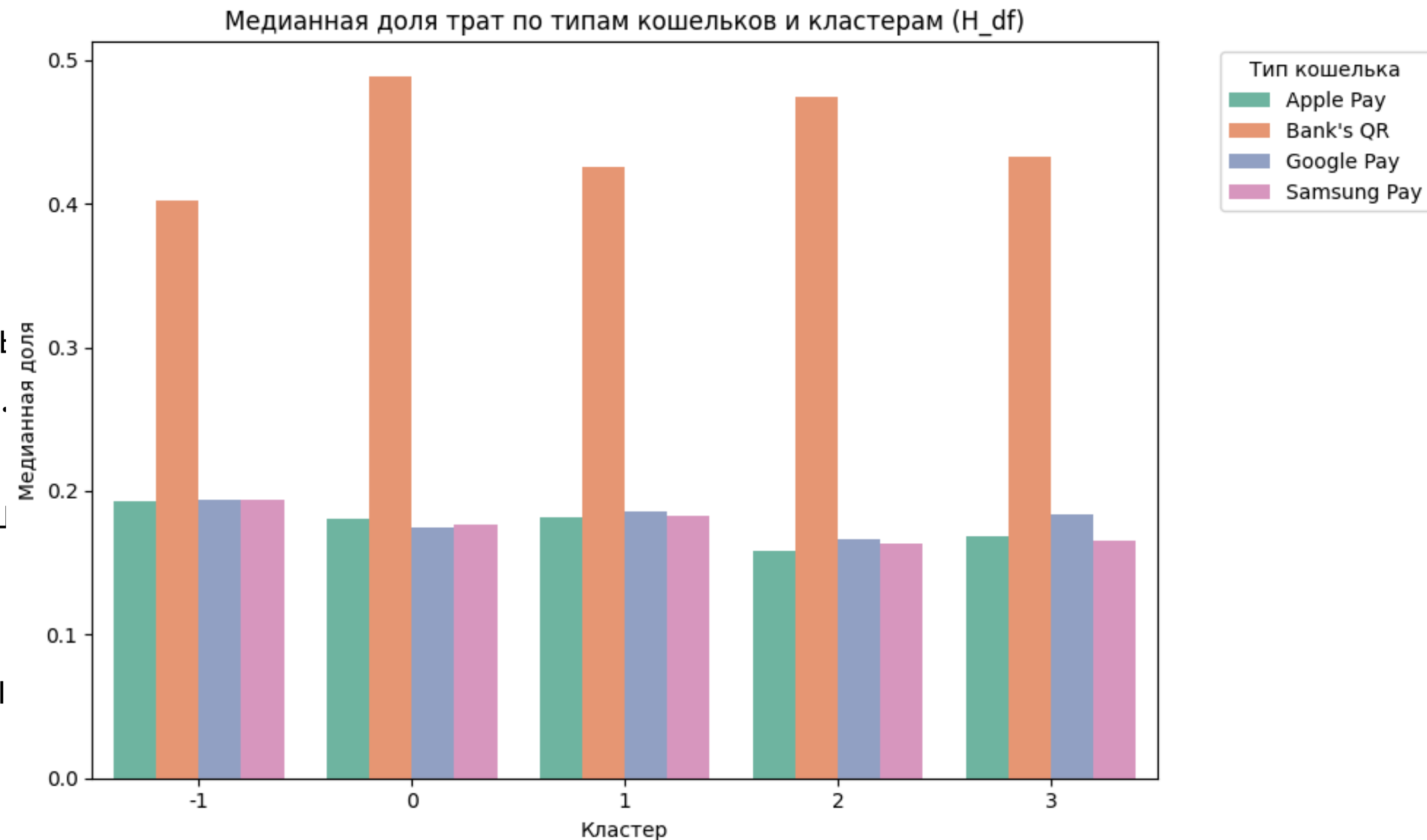
**Кластер 3:** Средний уровень трат, пик 8:00–18:00. Поведение указывает на рабочий график (офисные сотрудники, стабильные покупатели)."





# ОПИСАНИЕ СЕГМЕНТОВ HDBSCAN

- Bank's QR лидирует во всех кластерах: Значительно большая медианная доля трат по сравнению с другими кошельками.
- Bank's QR отражает поведение кластеров: Кластер 0 (Высокие расходы): Максимальная доля Bank's QR (около 0.48).
- Кластер 2 (Низкие расходы): Минимальная доля Bank's QR, но всё ещё доминирует.
- Apple/Google/Samsung Pay: Доли трат ниже, относительно схожи и стабильны по всем кластерам (0.15-0.2).



# ОПИСАНИЕ СЕГМЕНТОВ HDBSCAN

**Кластер -1 (Шум):** Атипичные, равномерные расходы.

- POS-режимы: Схожие доли Contactless, QR\_Code, ECOM (0.2-0.28).
- Кошельки: Bank's QR доминирует (0.4), другие ~0.2.

**Кластер 0:** (Высокие расходы / Премиум):

- Пик трат 8:00–15:00.
- POS-режимы: QR\_Code доминирует (0.35), Contactless высок (0.29).
- Кошельки: Bank's QR пик (~0.5), другие ~0.17.

**Кластер 1:** (Активные / Розничные):

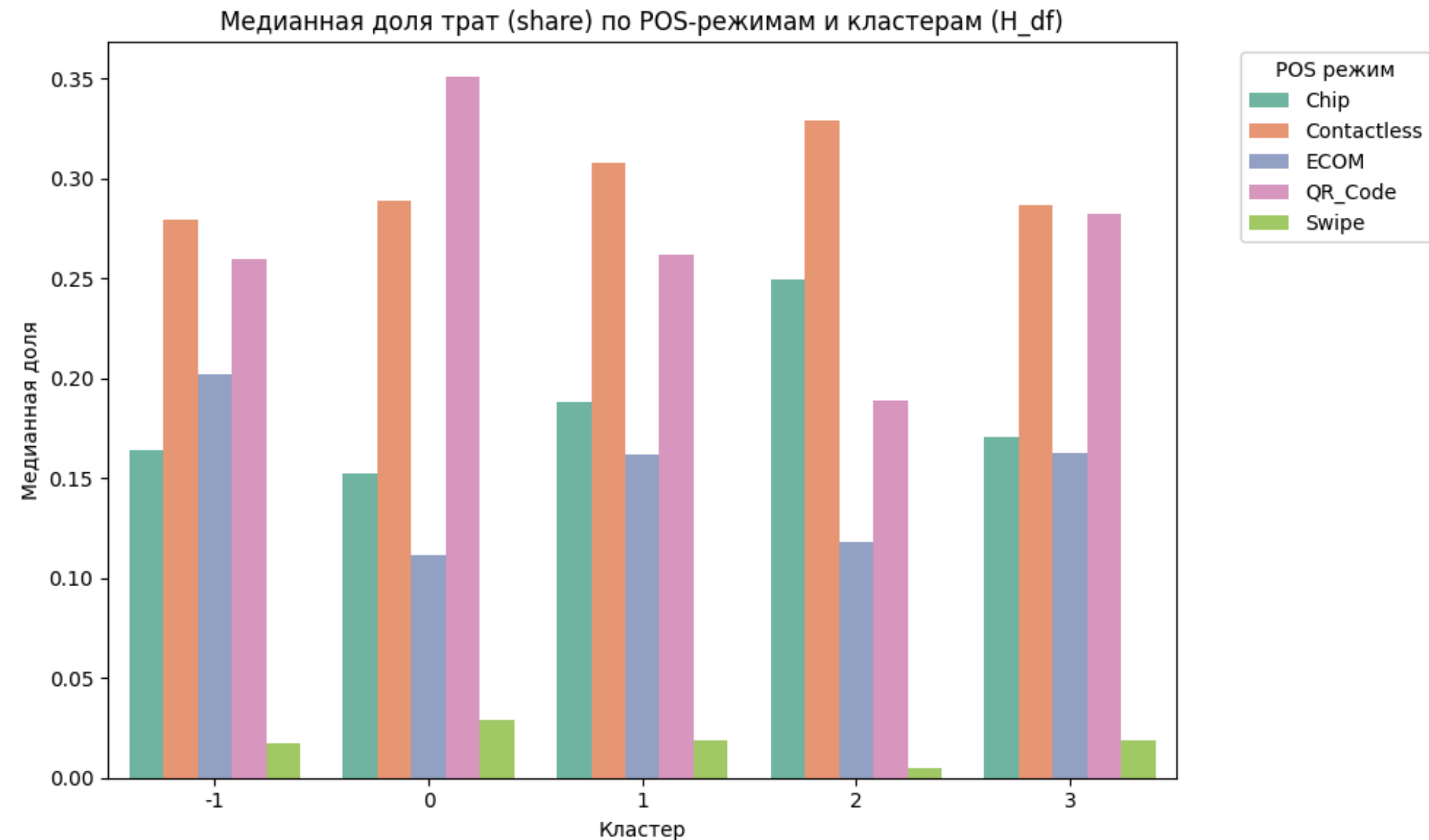
- Умеренные расходы, пик трат 8:00–14:00.
- POS-режимы: Contactless доминирует (0.31), QR\_Code высок (0.26).
- Кошельки: Bank's QR (0.42), другие ~0.18.

**Кластер 2:** (Низкие расходы / Молодые):

- Активность 8:00–13:00.
- POS-режимы: Contactless наиболее популярен (0.33), QR\_Code снижен (0.18).
- Кошельки: Bank's QR (0.4), другие наименьшие (0.15-0.16).

**Кластер 3:** (Средние траты / Офисные сотрудники):

- Пик трат 8:00–18:00 (рабочий график).
- POS-режимы: Сбалансированное использование Contactless (0.29) и QR\_Code (0.28).
- Кошельки: Bank's QR высок (0.43), другие ~0.17.



# ОПИСАНИЕ СЕГМЕНТОВ HDBSCAN

## Кластер -1 (Шум):

- Атипичные, равномерные траты. Время между транзакциями: ~1ч10м.
- POS: равные доли Contactless, QR, ECOM (~0.25).
- Кошельки: Bank's QR (0.4), другие ~0.2.

## Кластер 0 (Премиум):

- Высокие частые траты (пик 8:00–15:00), интервал ~53 мин.
- POS: QR (0.35), Contactless (0.29).
- Кошельки: Bank's QR (0.5), другие ~0.17.

## Кластер 1 (Розница):

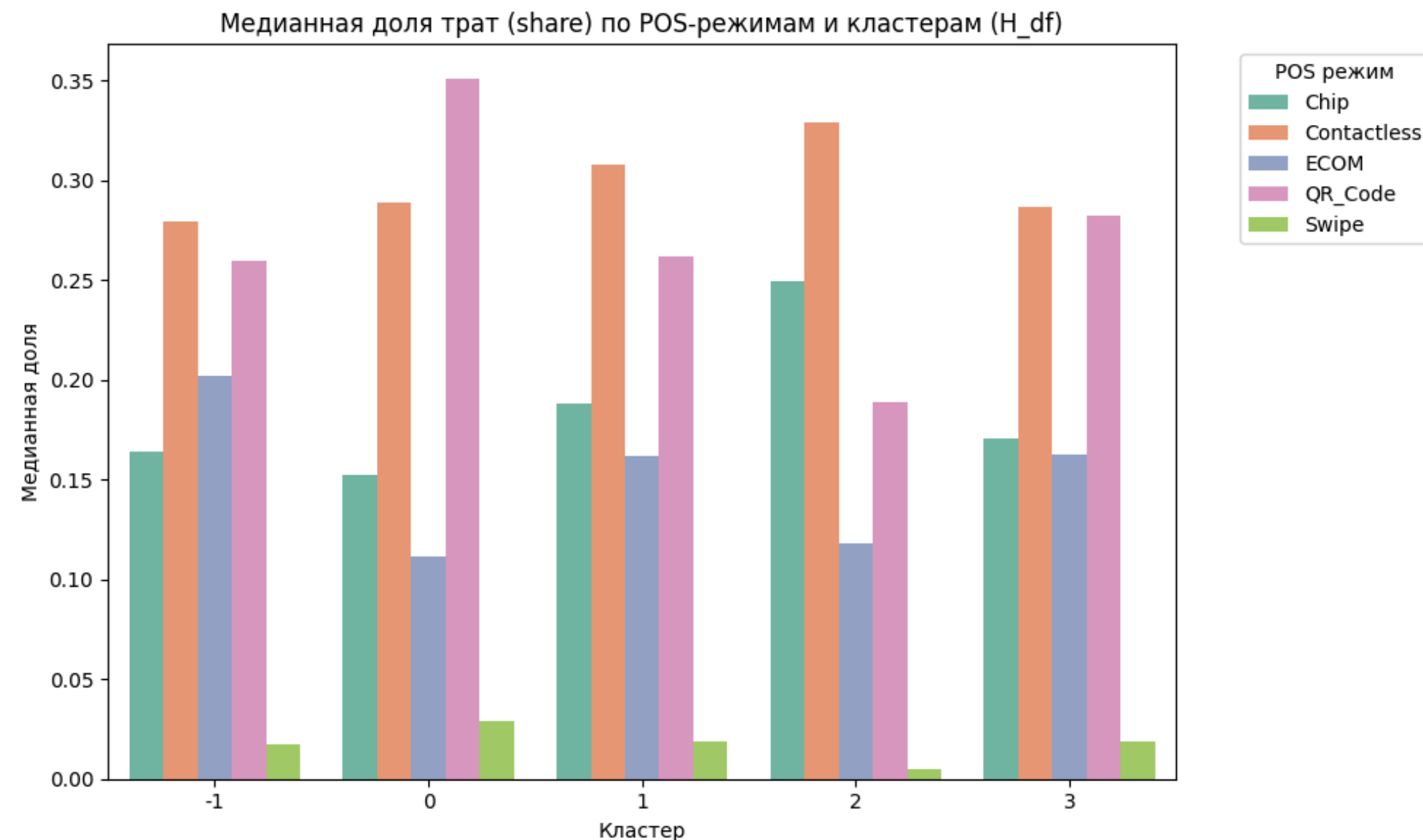
- Умеренные траты (пик 8:00–14:00), интервал ~1ч32м.
- POS: Contactless (0.31), QR (0.26).
- Кошельки: Bank's QR (0.42), другие ~0.18.

## Кластер 2 (Молодёжь):

- Редкие траты (пик 8:00–13:00), интервал ~3ч28м.
- POS: Contactless (0.33), QR (0.18).
- Кошельки: Bank's QR (0.4), другие ~0.15.

## Кластер 3 (Офисные):

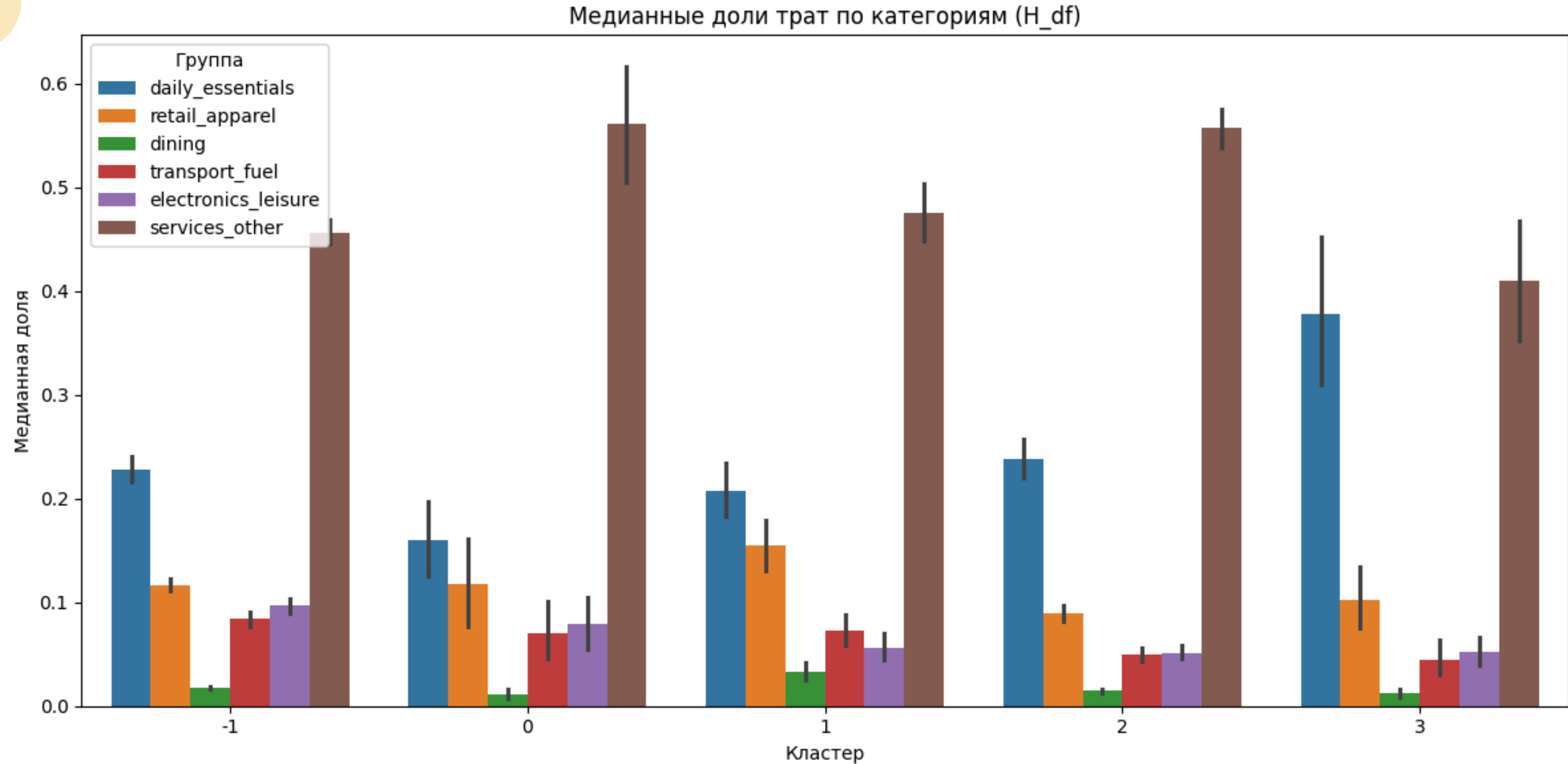
- Траты в рабочее время (8:00–18:00), интервал ~1ч22м.
- POS: Contactless (0.29), QR (0.28).
- Кошельки: Bank's QR (0.43), другие ~0.17.



# ОПИСАНИЕ СЕГМЕНТОВ HDBSCAN

- 🛒 **1. Повседневные покупки (Daily Essentials)**
  - Grocery & Food Markets
  - Pharmacies & Health
  - Utilities & Bill Payments
- 🛍️ **2. Розничная торговля и мода (Retail & Apparel)**
  - Clothing & Apparel
  - General Retail & Department
  - Home Furnishings & Supplies
- 🍴 **3. Еда и рестораны (Dining)**
  - Dining & Restaurants
- 🚛 **4. Транспорт и топливо (Transport & Fuel)**
  - Fuel & Service Stations
  - Travel & Transportation
- 💻 **5. Электроника и хобби (Electronics & Leisure)**
  - Electronics & Software
  - Hobby, Books, Sporting Goods
- 🛠️ **6. Услуги и прочее (Services & Other)**
  - Services (Other)
  - Unknown

# ОПИСАНИЕ СЕГМЕНТОВ HDBSCAN



# ВЫВОД ПО КЛАСТЕРАМ HDBSCAN

## 🧩 Кластер -1 — Шум / Атипичное поведение

- Равномерные расходы, возможны сбои или технические записи
- Траты: "Другие услуги" (45%), "Повседневные нужды" (23%)
- POS: Равномерно — Contactless, QR, ECOM
- Кошельки: Bank's QR (40%)
- Особенности: низкий объем иностранных транзакций, слабая сезонность

## 🇺🇸 Кластер 0 — Премиум / Корпоративные клиенты

- Высокие расходы, активность 8:00–15:00
- Траты: "Другие услуги" >55%
- POS: QR\_Code (35%), Contactless (29%)
- Кошельки: Bank's QR (почти 50%)
- Частые транзакции (~53 мин), локальный характер покупок

## 🛍️ Кластер 1 — Активные розничные клиенты

- Умеренные расходы, пик 8:00–14:00
- Траты: "Другие услуги" (47%), "Повседневные нужды", "Одежда"
- POS: Contactless (31%), QR\_Code (26%)
- Кошельки: Bank's QR (42%)
- Самая высокая доля иностранных транзакций (0.011%)

## 👛 Кластер 2 — Низкие траты / Молодые клиенты

Низкие расходы, активность 8:00–13:00

Траты: "Другие услуги" (55%), "Повседневные нужды" (24%)

POS: Contactless (33%)

Кошельки: Bank's QR (40%), прочие — минимум

Самая редкая частота трат (~3.5 ч)

## 🏢 Кластер 3 — Офисные / Стабильные клиенты

Средние траты, активность 8:00–18:00

Траты: "Другие услуги" (40%), "Повседневные нужды" (37%)

POS: Сбалансировано — Contactless и QR\_Code

Кошельки: Bank's QR (43%)

Сильная сезонность, пик в декабре

# ОПИСАНИЕ СЕГМЕНТОВ DBSCAN

## Кластер -1 (Шум):

- Атипичные, равномерные траты. Время между транзакциями: ~1ч10м.
- POS: Contactless, QR, ECOM — равные доли (~0.25).
- Кошельки: Bank's QR (0.42), Apple Pay (0.16), Google/Samsung Pay (~0.15).

## Кластер 0 (Премиум):

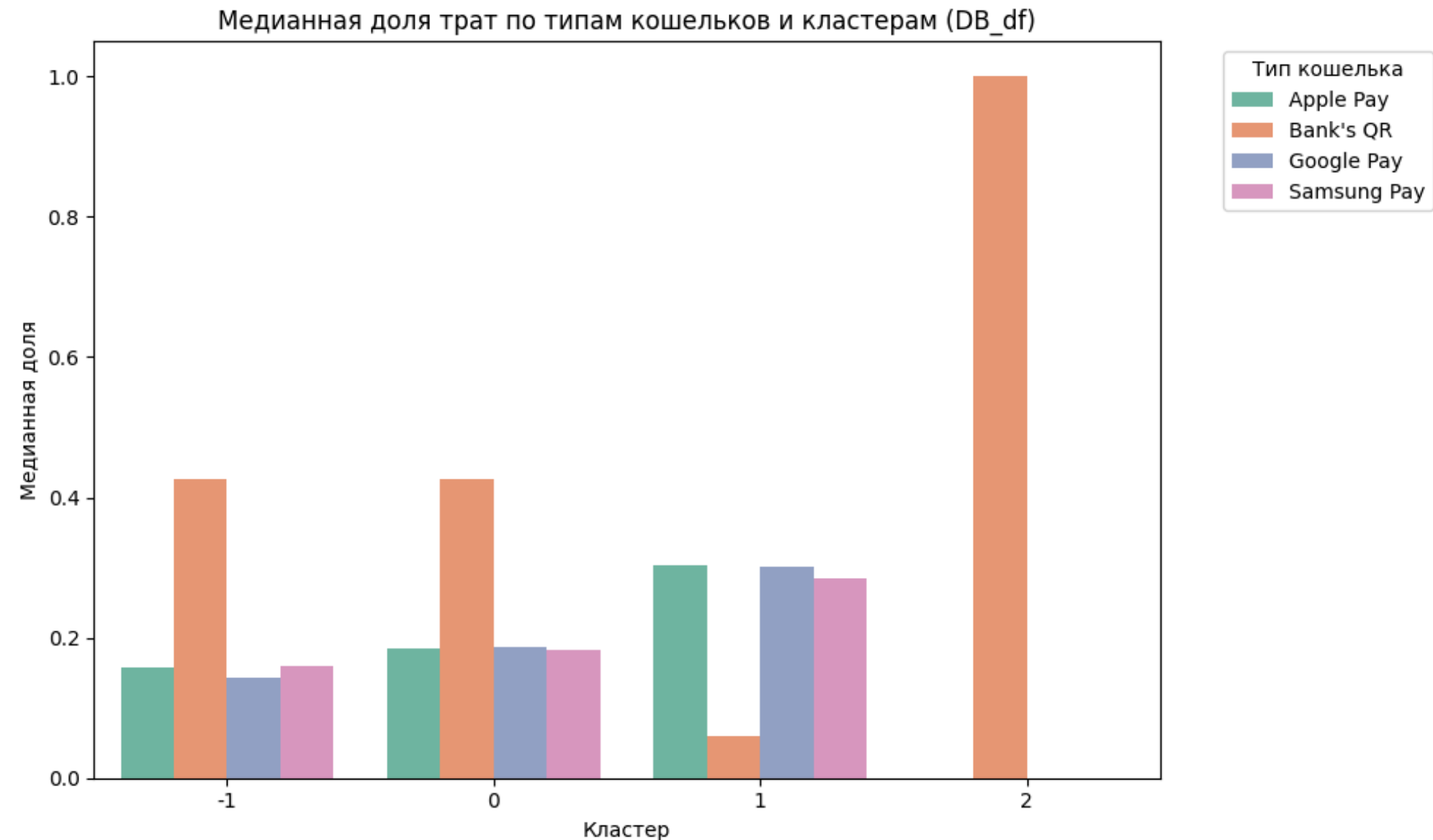
- Частые траты (пик 8:00–15:00), интервал ~53 мин.
- POS: QR (0.35), Contactless (0.29).
- Кошельки: Bank's QR (0.42), Apple Pay (0.18), Google/Samsung Pay (~0.18).

## Кластер 1 (Цифровые):

- Активные пользователи цифровых кошельков, интервал ~1ч20м.
- POS: Contactless (высокая доля), QR — второстепенный.
- Кошельки: Apple/Google/Samsung Pay (~29–30%), Bank's QR (0.05).

## Кластер 2 (Лоялисты QR):

- Почти полная зависимость от одного метода оплаты.
- POS: QR (доминирует).
- Кошельки: Bank's QR (~100%), остальные — ~0%.





# ОПИСАНИЕ СЕГМЕНТОВ DBSCAN

## Кластер -1 (Цифровой оффлайн):

Доминирует Bank's QR (~42%), другие кошельки ~15–20%.  
POS: QR (26%), Contactless (25%), ECOM (19%), Chip (17%).

Предпочитают банковские приложения и бесконтактные оффлайн-платежи. Онлайн и чип также заметны.

## Кластер 0 (Переходный):

Bank's QR ~42%, мобильные кошельки ~18–20%.  
POS: Contactless (29%), QR (25%), Chip/ECOM ~18%.

Схож с кластером -1, но выше открытость к мобильным кошелькам. На раннем этапе цифровизации.

## Кластер 1 (Онлайн-цифровой):

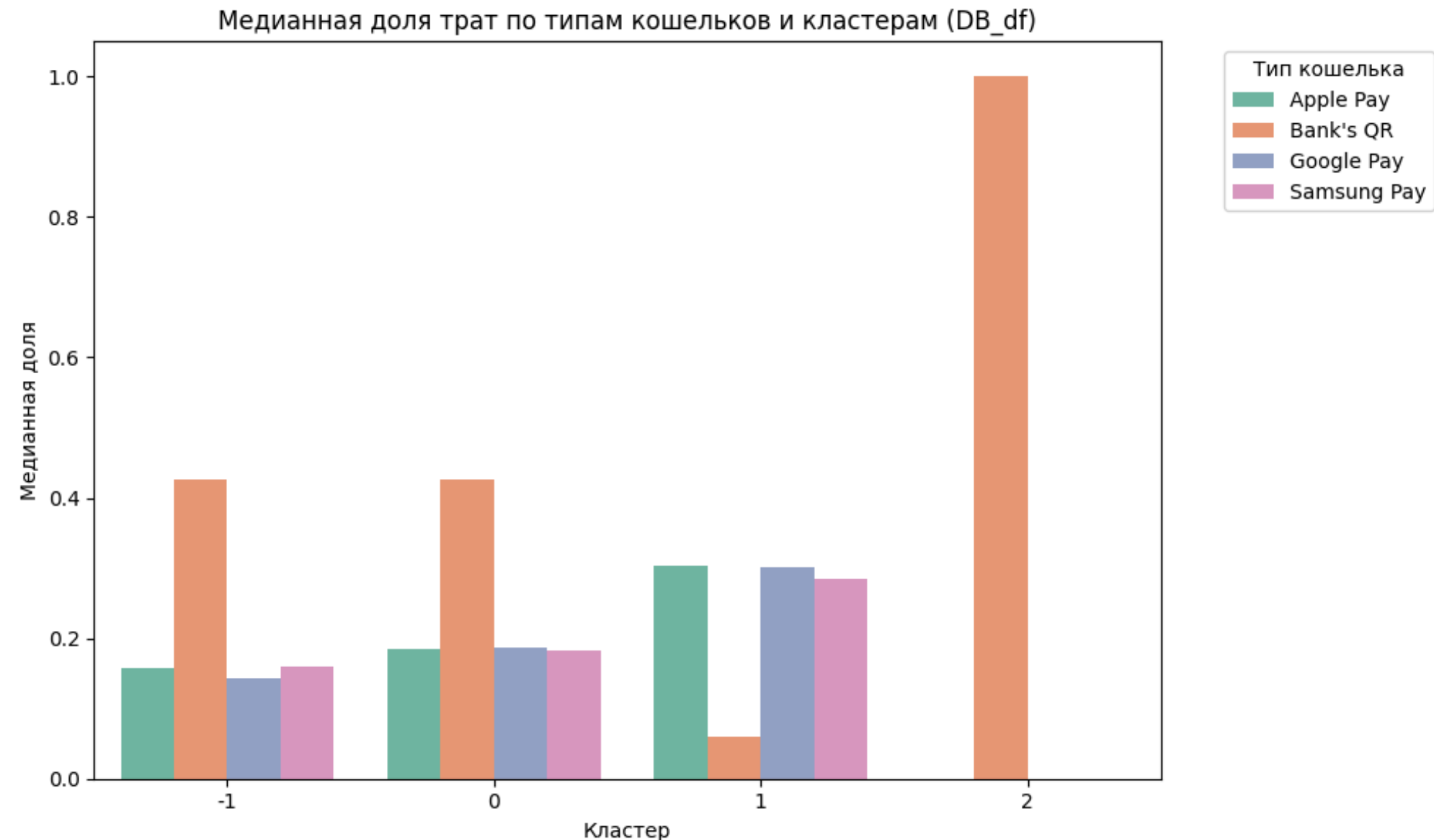
Apple/Google/Samsung Pay ~28–30%, Bank's QR <10%.  
POS: ECOM доминирует (~80%), остальные — <10%.

Пользователи цифровых кошельков, ориентированы на онлайн-покупки.

## Кластер 2 (Консервативный оффлайн):

Почти исключительно Bank's QR (~100%).  
POS: Chip (42%), Contactless (40%), QR (16%).

Не используют мобильные кошельки и онлайн-оплату.  
Ориентированы на физические карты и банковские POS-платежи.





# ОПИСАНИЕ СЕГМЕНТОВ DBSCAN

## Кластер -1 (Цифровой оффлайн):

- ◆ Кошельки: Bank's QR (>40%), Apple/Google/Samsung Pay ~15–20%.
- ◆ POS: Contactless (25%), QR (26%), ECOM (19%), Chip (17%).
- ◆ Время: ~12 500 — средняя частота.

💡 Вывод: Используют банковские QR, часто платят бесконтактно и онлайн. Среднеактивные пользователи.

## Кластер 0 (Активные платёжники):

- ◆ Кошельки: Bank's QR >40%, Apple/Google/Samsung Pay ~18–20%.
- ◆ POS: Contactless (29%), QR (25%), Chip/ECOM ~18%.
- ◆ Время: ~5 500 — самая высокая частота.

💡 Вывод: Часто совершают покупки. Чуть активнее используют мобильные кошельки. Очень активные покупатели.

## Кластер 1 (Онлайн-цифровой):

- ◆ Кошельки: Apple/Google/Samsung Pay ~28–30%, Bank's QR <10%.
- ◆ POS: ECOM ~80%, остальное <10%.
- ◆ Время: ~16 500 — редкие покупки.

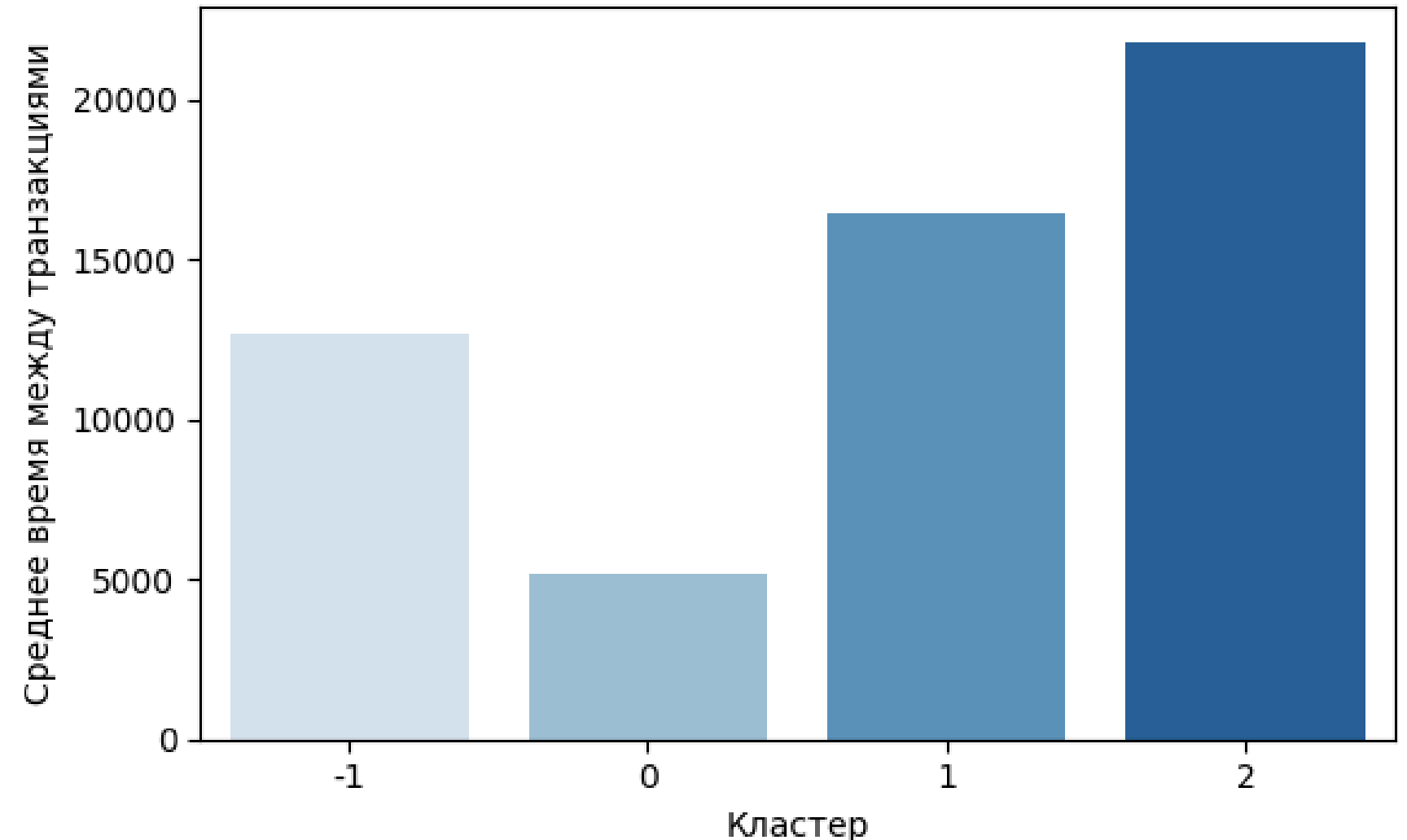
💡 Вывод: Ориентированы на онлайн-шопинг и мобильные кошельки. Практически не платят в оффлайне.

## Кластер 2 (Редкие оффлайн-покупатели):

- ◆ Кошельки: Bank's QR ~100%, другие — отсутствуют.
- ◆ POS: Chip (42%), Contactless (40%), QR (16%).
- ◆ Время: >21 000 — крайне редкие транзакции.

💡 Вывод: Не используют мобильные кошельки и онлайн. Консервативные пользователи с крайне низкой активностью.

Медианное время между транзакциями по кластерам (DB\_df)



# ОПИСАНИЕ СЕГМЕНТОВ DBSCAN



# ОПИСАНИЕ СЕГМЕНТОВ DBSCAN

- **Кластер -1 Активные горожане**

Кошелёк: Преобладает Bank's QR (~40%).

POS: Contactless и QR\_Code (~25% каждый), Chip и ECOM тоже заметны.

Частота: Средняя (медианное время ~12 500).

Траты по времени: Низкий уровень, небольшие пики в 7–8 и 19–20 ч.

Вывод: Повседневные оффлайн-платежи через Bank's QR, умеренная активность, ориентация на удобство.

- **Кластер 0 Повседневные покупатели**

Кошелёк: Bank's QR (>40%), мобильные кошельки ~20%.

POS: Contactless (29%), QR\_Code (25%), Chip и ECOM ~18%.

Частота: Очень высокая (медианное время ~5 500).

Траты по времени: Самые высокие, пик утром (7–9 ч) и вечером (19–20 ч).

Вывод: Самые активные пользователи с регулярными оффлайн-тратами, часто используют Bank's QR.

- **Кластер 1 Цифровые онлайн-покупатели**

Кошелёк: Apple Pay, Google Pay, Samsung Pay (~30% каждый), Bank's QR <10%.

POS: ECOM — почти 80%.

Частота: Низкая (медианное время ~16 500).

Траты по времени: Средний уровень, пик в 19–20 ч.

Вывод: Онлайн-покупатели, предпочитают мобильные кошельки, активны вечером.

- **Кластер 2 Консервативные потребители**

Кошелёк: Только Bank's QR (100%).

POS: Chip (42%) и Contactless (40%), QR\_Code ~16%.

Частота: Очень низкая (медианное время >21 000).

Траты по времени: Самые низкие, равномерные днём.

Вывод: Редкие оффлайн-платежи, нет использования мобильных или онлайн-методов.

# ОПИСАНИЕ СЕГМЕНТОВ DBSCAN

## Кластер 0 – Активные и разносторонние пользователи

- Часто тратят: наименьшие интервалы между транзакциями
- Основные каналы: Bank's QR, Contactless
- Время трат: пик утром и вечером
- Категории: услуги (40%), повседневное (12%)
- Активные городские жители, использующие платежи для широкого круга нужд, включая бизнес и сервисы

## Кластер -1 – Рутинные покупатели

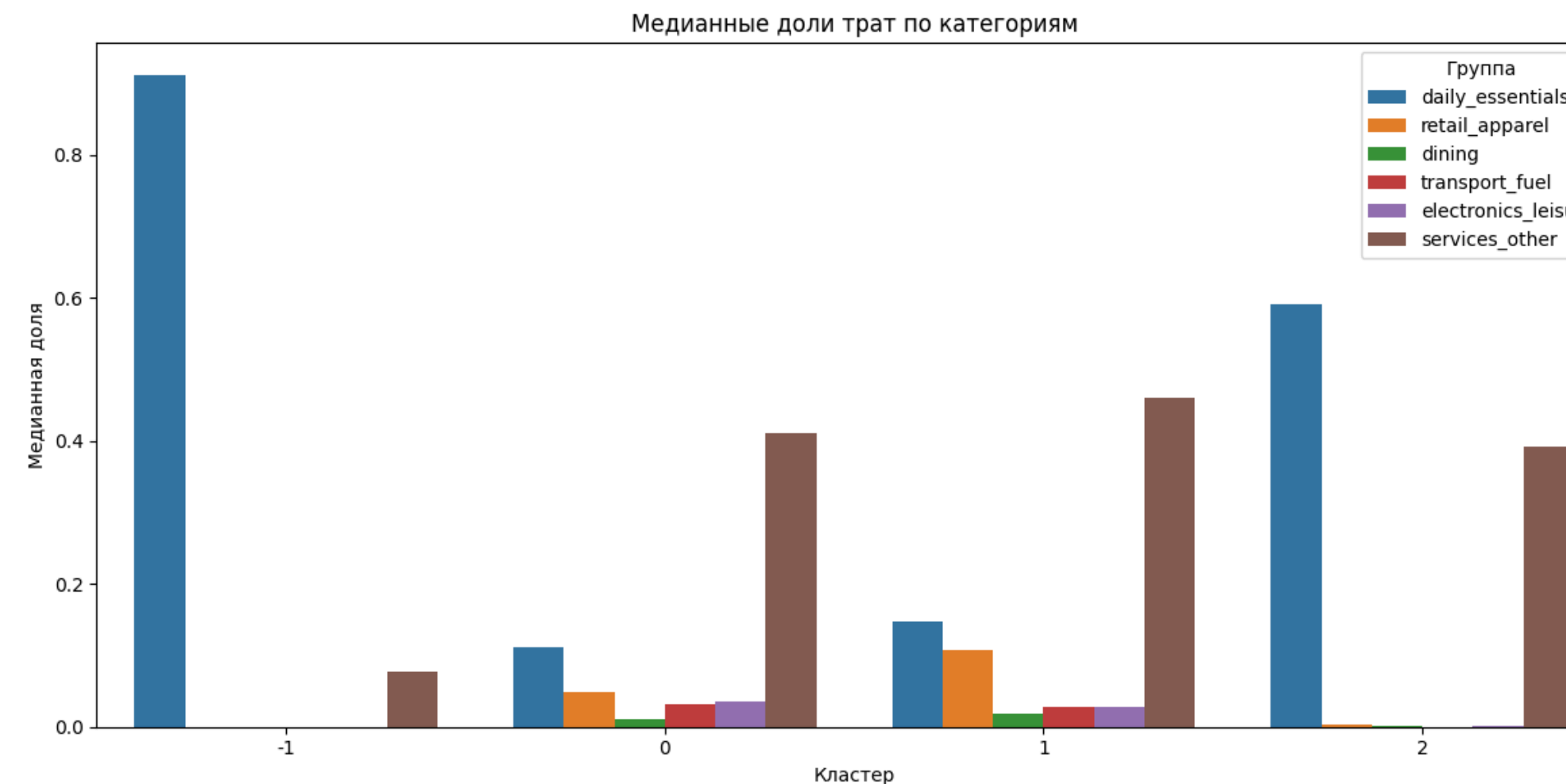
- Умеренная частота трат
- Основные каналы: Contactless, QR, также Chip и онлайн (ECOM)
- Время трат: небольшие пики утром и вечером
- Категории: повседневные товары (90%)
- Пользователи для повседневных покупок, ценят удобство бесконтактных и QR-платежей

## Кластер 1 – Онлайн-ориентированные цифровые покупатели

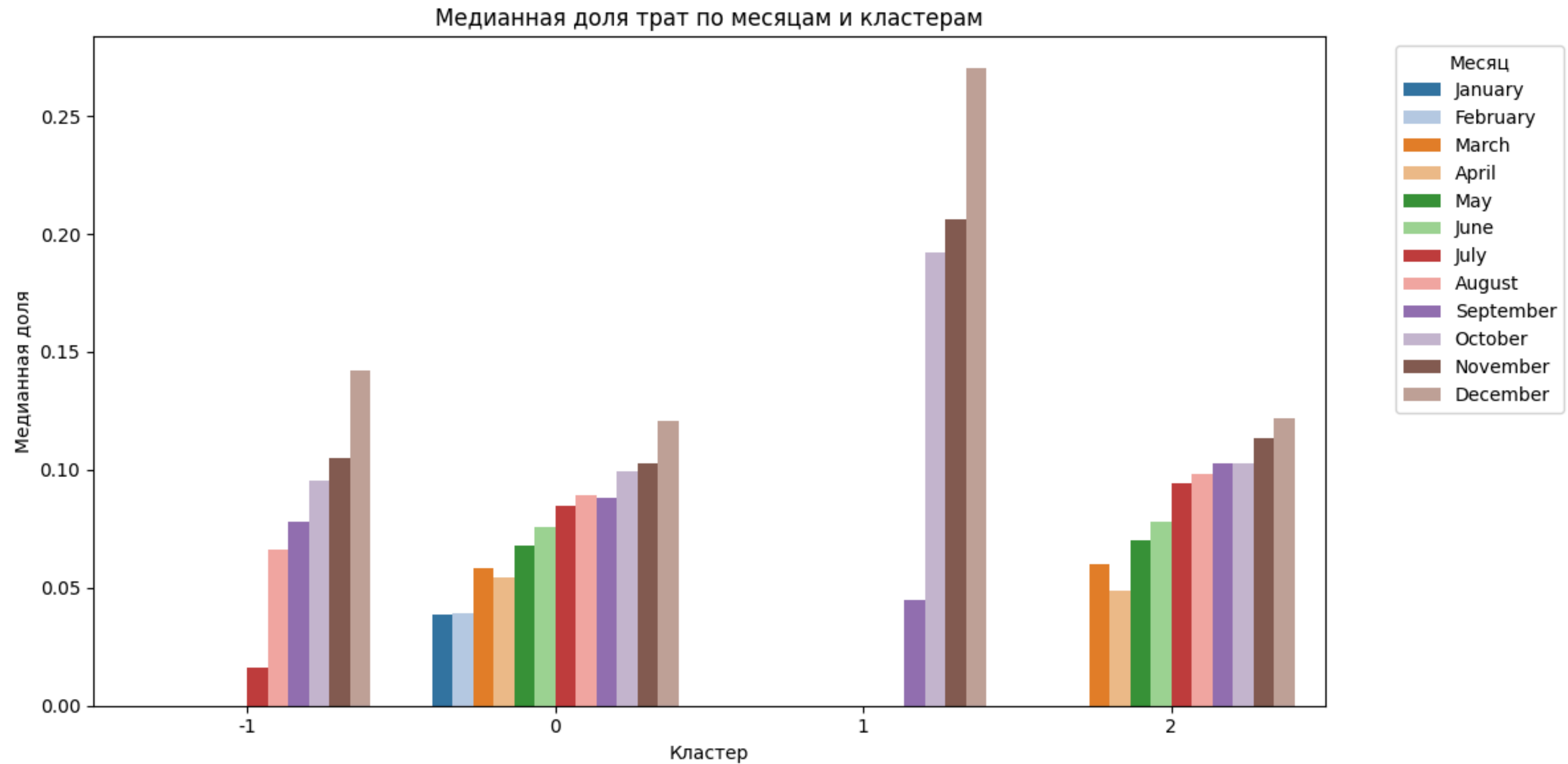
- Низкая частота, но стабильные вечерние траты
- Основные каналы: Apple Pay, Google Pay, ECOM (~80%)
- Категории: услуги (45%), повседневное, одежда
- Молодые и цифровые, предпочитают онлайн-шопинг и мобильные кошельки

## Кластер 2 – Редкоактивные, традиционные пользователи

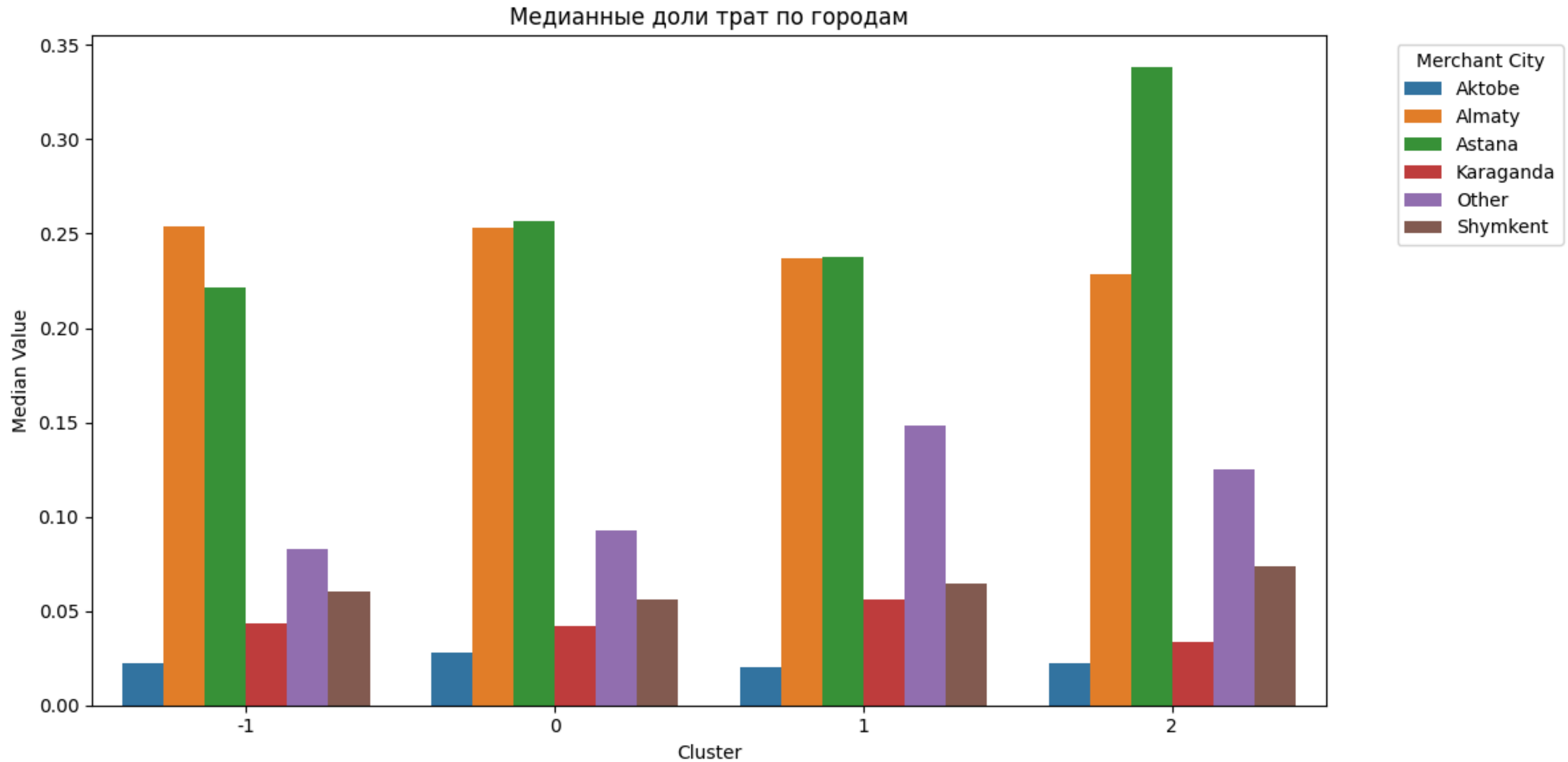
- Самая низкая частота и объёмы трат
- Основные каналы: Bank's QR (100%), Chip, Contactless
- Время трат: низкая активность днём
- Категории: услуги (40%), повседневное (30%)
- Консервативные и малоактивные, совершают редкие, целевые покупки оффлайн



# ОПИСАНИЕ СЕГМЕНТОВ DBSCAN



# ОПИСАНИЕ СЕГМЕНТОВ DBSCAN



# ВЫВОД ПО КЛАСТЕРАМ DBSCAN

 Кластер -1 — Консервативные повседневные покупатели

Средние расходы, пики 7:00–8:00 и 19:00–20:00

Траты: Повседневные нужды (до 90%)

POS: Contactless, QR, ECOM, Chip

Кошельки: Bank's QR (основной)

Особенности: рост активности в ноябре–декабре, средняя частота транзакций (~12 500)

 Кластер 0 — Активные потребители услуг

Самые частые транзакции, пик 7:00–9:00

Траты: Другие услуги (40%), повседневные нужды

POS: Bank's QR, Contactless

Кошельки: Bank's QR (высокая доля)

Особенности: стабильная активность круглый год, минимальное время между транзакциями (~5 500)

 Кластер 1 — Онлайн-энтузиасты и сезонные шоперы

Умеренные расходы, вечерний пик 19:00–20:00

Траты: Другие услуги (45%), повседневные нужды

POS: ECOM (почти 80%), Contactless

Кошельки: Apple/Google/Samsung Pay

Особенности: резкий рост в ноябре–декабре, редкие транзакции (~16 500)

 Кластер 2 — Редкие банковские пользователи

Низкие расходы, без ярко выраженных пиков

Траты: Другие услуги (40%), повседневные нужды (30%)

POS: Chip (42%), Contactless (40%)

Кошельки: Bank's QR (100%)

Особенности: очень редкие транзакции (~21 000), слабая сезонность



# ЗАКЛЮЧЕНИЕ

## Прогнозы

- Увеличение доли мобильных платежей: Сегмент "Онлайн-Энтузиаст и Сезонный Шопер" (Кластер 1 HDBSCAN) уже активно использует мобильные кошельки. Прогнозируется дальнейший рост их использования, особенно среди онлайн-покупателей, где удобство мобильных платежей играет ключевую роль.
- Сохранение значимости Bank's QR в оффлайне: Несмотря на рост популярности Apple Pay/Google Pay/Samsung Pay, сегменты "Консервативный Покупатель Повседневных Товаров" (Кластер -1 HDBSCAN) и "Активный Потребитель Услуг" (Кластер 0 HDBSCAN) продолжают активно использовать Bank's QR и бесконтактные платежи. Это указывает на глубоко укоренившиеся привычки и инфраструктурную поддержку.
- Выраженные сезонные пики потребления: "Онлайн-Энтузиаст и Сезонный Шопер" (Кластер 1 HDBSCAN) будет продолжать демонстрировать резкий рост трат в ноябре и декабре, что является предсказуемым для предпраздничного периода и "Черной пятницы". Подобные, но менее выраженные, пики будут наблюдаться и у "Консервативного Покупателя Повседневных Товаров" (Кластер -1 HDBSCAN).
- Стабильность потребления услуг: "Активный Потребитель Услуг" (Кластер 0 HDBSCAN) продолжит обеспечивать стабильный поток транзакций в течение года, поскольку услуги являются постоянной потребностью.
- Высокая предсказуемость поведения "Редких Банковских Пользователей" (Кластер 2 HDBSCAN): Из-за низкой частоты и объема трат, а также ограниченного набора используемых методов платежей, поведение этого сегмента будет наиболее предсказуемым, хотя и с наименьшей выручкой.



# ЗАКЛЮЧЕНИЕ

## Инсайты

Персонализация платежных решений:

- Для оффлайн-клиентов (Кластеры -1, 0 HDBSCAN): Фокус на улучшении банковских QR и NFC в приложении, особенно для утренних транзакций Кластера 0.
- Для онлайн-клиентов (Кластер 1 HDBSCAN): Обеспечить бесшовную интеграцию с Apple/Google/Samsung Pay, стимулировать активность в начале года.
- Для редких пользователей (Кластер 2 HDBSCAN): Целиться на увеличение среднего чека, а не частоты транзакций.

Таргетинг по времени и потребностям:

- Используйте утренние (7-9 утра) и вечерние (19-20 часов) пики активности для целенаправленных предложений, соответствующих потребностям сегментов (услуги для Кластера 0, онлайн-покупки для Кластера 1).
- Адаптируйте программы лояльности под основные категории трат: "повседневные нужды" для Кластеров -1 и 2, "услуги" для Кластеров 0, 1 и 2.

Ценность GNN: Применение графовых нейронных сетей позволило получить глубокие инсайты, выявив более тонкие и действенные сегменты поведения клиентов, что невозможно достичь традиционными методами.

# ЗАКЛЮЧЕНИЕ

## Основные выводы по сегментации (на основе объединенного анализа):

- Кластеры -1 и 0 (HDBSCAN), а также -1, 0, 1, 2, 3 (DBSCAN): Разделение на кластеры по обеим моделям кластеризации показало схожие паттерны поведения, хотя и с разным количеством кластеров и распределением шума. HDBSCAN оказался более интерпретируемым и устойчивым к шуму, в то время как DBSCAN продемонстрировал идеальную силу кластеров по Silhouette Score.
- Консервативные повседневные покупатели (Кластер -1 HDBSCAN): Этот сегмент характеризуется использованием Bank's QR и бесконтактных платежей для ежедневных нужд. Они проявляют умеренную активность и рост трат к концу года, что указывает на сезонные покупки.
- Активные потребители услуг (Кластер 0 HDBSCAN): Наиболее активная и высокодоходная группа, совершающая частые транзакции, в основном через Bank's QR и бесконтактные методы, с ярко выраженной утренней активностью. Их основные траты приходятся на различные услуги, а активность стабильна в течение года.
- Онлайн-энтузиасты и сезонные шоперы (Кластер 1 HDBSCAN): Этот сегмент активно использует мобильные платежные системы для онлайн-покупок. Они демонстрируют ярко выраженную сезонность трат с пиками в ноябре и декабре, что указывает на предпраздничный шопинг.
- Редкие банковские пользователи (Кластер 2 HDBSCAN): Самый консервативный сегмент, использующий исключительно Bank's QR через чиповые и бесконтактные транзакции. Их активность очень низка, траты минимальны и распределены равномерно в течение дня, с небольшим сезонным ростом к концу года.



**THANK YOU**