

# RFM Analysis

## Вступление

В рамках данного проекта мы осуществим RFM-анализ датасета с покупками.

---

## Источник данных

Датасет доступен по [ссылке](#) на Kaggle и также доступен в той же папке, что и данный файл, в формате .sql.

---

## Описание данных

**transaction\_id** - ID транзакции

**purchase\_date** - дата покупки

**customer\_id** - ID покупателя

**gender** - пол покупателя

**age** - возраст покупателя

**product\_category** - категория товара

**quantity** - количество приобретенного товара

**price\_per\_unit** - стоимость единицы товара

**total\_amount** - общая сумма транзакции

---

## Анализ

В результате нашего анализа мы ожидаем получить разделение базы клиентов на несколько сегментов, чтобы затем отталкиваться от сегмента в рамках email коммуникации.

Детализированные сегменты, а также рекомендации по дальнейшему взаимодействию основаны на [данной статье](#).

Приступим к анализу.

```
with customer_purchases as
--считаем по каждому клиенту кол-во покупок, общую сумму и последнюю дату покупки
(select distinct customer_id,
  count(transaction_id) as cnt_purchases,
  sum(total_amount) as total_sum,
  max(purchase_date) as last_purchased
 from retail
 group by customer_id),

rfm as
--добавляем давность последней покупки в днях
```

```

(select customer_id,
    current_date - last_purchased::date as recency,
    cnt_purchases as cnt_purchases,
    total_sum as total_sum
from customer_purchases),

customer_segment_base as
(select customer_id,
    --далее каждому клиенту присваиваем категорию (1-2-3) на основе этих трех показателей
    ntile(3) over (order by recency) as recency,
    ntile(3) over (order by cnt_purchases) as frequency,
    ntile(3) over (order by total_sum) as monetary
from rfm),

segments as
(select customer_id,
    recency,
    frequency,
    monetary,
    concat(recency, frequency, monetary) as segment
from customer_segment_base),

complete_rfm as
(select customer_id,
    recency,
    frequency,
    monetary,
    segment,
    case when recency = 1 then 'Уходящие'
        when recency = 2 then 'Спящие'
        else 'Постоянные' end as segment_type,
    case when segment = '111' then 'Потерянные экономные'
        when segment = '112' then 'Одноразовые'
        when segment = '113' then 'Одноразовые'
        when segment in ('121', '122', '123') then 'Уходящие редкие'
        when segment = '131' then 'Уходящие постоянные'
        when segment = '132' then 'Уходящие хорошие постоянные'
        when segment = '133' then 'Уходящие VIP'
        when segment in ('211', '212', '213') then 'Спящие разовые с маленьким чеком'
        when segment = '221' then 'Спящие редкие с маленьким чеком'
        when segment = '222' then 'Спящие редкие со средним чеком'
        when segment = '223' then 'Спящие редкие с высоким чеком'
        when segment = '231' then 'Спящие постоянные с маленьким чеком'
        when segment = '232' then 'Спящие постоянные со средним чеком'
        when segment = '233' then 'Спящие постоянные с высоким чеком'
        when segment = '311' then 'Новички с маленьким чеком'
        when segment = '312' then 'Новички со средним чеком'
        when segment = '313' then 'Новички с высоким чеком/Потенциальные VIP'

```

```

when segment = '321' then 'Постоянные с маленьким чеком'
when segment = '322' then 'Постоянные со средним чеком'
when segment = '323' then 'Постоянные с высоким чеком'
when segment = '331' then 'Постоянные с маленьким чеком'
when segment = '332' then 'Постоянные со средним чеком'
when segment = '333' then 'VIP'
else ''
end as segment_desc
from segments)

select *
from complete_rfm;

```

В итоге получаем следующую таблицу:

	customer_id character varying (512)	recency integer	frequency integer	monetary integer	segment text	segment_type text	segment_desc text
1	CUST650	1	1	1	111	Уходящие	Потерянные экономные
2	CUST211	1	3	3	133	Уходящие	Уходящие VIP
3	CUST857	1	1	1	111	Уходящие	Потерянные экономные
4	CUST805	1	3	3	133	Уходящие	Уходящие VIP
5	CUST908	1	3	3	133	Уходящие	Уходящие VIP
6	CUST233	1	3	3	133	Уходящие	Уходящие VIP
7	CUST520	1	2	2	122	Уходящие	Уходящие редкие
8	CUST989	1	1	1	111	Уходящие	Потерянные экономные
9	CUST664	1	3	3	133	Уходящие	Уходящие VIP
10	CUST429	1	1	1	111	Уходящие	Потерянные экономные
11	CUST386	1	3	3	133	Уходящие	Уходящие VIP
12	CUST062	1	2	2	122	Уходящие	Уходящие редкие
13	CUST879	1	1	1	111	Уходящие	Потерянные экономные
14	CUST082	1	2	2	122	Уходящие	Уходящие редкие

Теперь мы можем кастомизировать наши рассылки данным клиентам, а также исключить некоторых из них из перечня получателей, чтобы максимизировать эффективность рассылок.