

# Конспект по теме "Когортный анализ"

## Когортный анализ

**Когорта** — группа людей, объединённых одним и тем же событием, произошедшим в определённый период времени.

Для выделения когорты, надо определить:

- 1) **событие**, которое должно произойти с пользователем;
- 2) **период**, за который должно случиться событие.

**Событие** — зафиксированный факт совершения пользователем какого-либо действия.

**Период** — временной промежуток, в котором происходит событие.

Определившись с событием и периодом, **формируют когорту**: группируют пользователей, которые совершили одно и то же событие в одинаковый период времени.

Аналитики делят клиентскую базу на когорты, чтобы определить тренды в поведении пользователей. Берут когорту и изучают её на протяжении времени. Выводы из когортного анализа могут стать причиной изменений в продукте и процессах, он позволяет выдвинуть впечатляющие гипотезы. Но это метод ретроспективного анализа — он отвечает на вопрос, как пользователи вели себя *в прошлом*, поэтому его редко применяют для прогнозирования поведения пользователей, на которое в будущем повлияют внешние факторы.

## Оценка изменений абсолютных показателей по месяцам

Большинство количественных наблюдений фиксируют в абсолютных величинах. **Абсолютная величина** — это объём, величина или размер наблюдаемого явления или события. **Относительная величина** —

показатель, который даёт числовую меру соотношения двух сопоставимых величин.

Первая задача, которую можно решить путём когортного анализа: отследить, как изменяется количество покупающих пользователей внутри одной когорты. Это позволит понять, сколько людей продолжают покупать в следующих месяцах после совершения первой покупки.

Отслеживание абсолютных показателей по месяцам позволяет исследовать изменения с течением времени. А также заметить общие черты в поведении всех когорт — например, обнаружить сезонное снижение числа пользователей. Можно сделать и вывод о доле каждой когорты в общем количестве покупателей месяца.

## Оценка изменений относительных показателей по времени жизни

Ещё одна задача когортного анализа: понять, как с течением времени изменяется **средний чек** — сумма покупок, делённая на число покупателей. Для когорты покупателей рассчитывается показатель **lifetime** — метрика, которая для каждого события отображает, в каком по счёту месяце от месяца когорты оно произошло.

## Визуализация когортного анализа

**Тепловая карта** — визуализация таблицы, где ячейки со значениями меняют цвет в зависимости от их близости к максимальному или минимальному значению.

Для построения тепловых карт в Python используется метод `heatmap()` библиотеки `seaborn`. В качестве параметров он принимает:

- **таблицу с данными;**
- **`annot=True`** означает, что в ячейках тепловой карты будет отображаться значение самой величины;
- **`fmt='.1f'`** задаёт формат вывода значения — один знак после запятой;
- **`linewidths=1`** задаёт ширину линии, которая разграничит ячейки тепловой карты (1 пиксель);

- **linecolor='gray'** устанавливает цвет линии — серый.

```
import seaborn as sns
from matplotlib import pyplot as plt

plt.figure(figsize=(13, 9))
sns.heatmap(dataframe, annot=True, fmt='.1f', linewidths=1, linecolor='gray')
```

## Retention Rate и Churn Rate

Когортный анализ применяют для анализа поведения пользователей в цифровых продуктах. Особенно часто оценивают показатели **Retention Rate** (коэффициент удержания) и **Churn Rate** (коэффициент оттока).

**Retention Rate** показывает, сколько пользователей из когорты относительно их изначального числа, вернулось в последующие периоды.

Этот показатель рассчитывается так:  $\text{Retention Rate} = \frac{\text{количество активных пользователей на } n\text{-ый lifetime}}{\text{количество активных пользователей на } 0 \text{ lifetime}} * 100\%$ .

**Churn Rate** даёт понимание, какой процент пользователей прекращает использование продукта с течением времени. Если при вычислении *Retention Rate* за точку отсчёта берут общее количество клиентов когорты, то для *Churn Rate* — количество клиентов на определенный отрезок времени сравнивается с предыдущим:  $\text{Churn Rate} = \frac{\text{количество активных пользователей на } n\text{-ый lifetime}}{\text{количество активных пользователей на } (n-1)\text{-ый lifetime} - 1} * 100\%$ .

Для расчёта **Churn Rate** в Python используется метод **pct\_change()**. Он позволяет вычислить процентное изменение значения в столбце относительно предыдущей строки датафрейма. Если применить его с группировкой, метод сработает внутри группы:

```
cohorts['churn_rate'] = cohorts.groupby(['column1'])['column2'].pct_change()
```

## Поведенческие когорты

**Анализ поведенческих когорт** — метод анализа, в ходе которого выделяют когорту пользователей, совершивших событие или ряд событий в определённый промежуток времени. Затем изучают, как с течением времени изменяется целевой показатель поведенческой когорты, например, *Retention Rate*.