

СЕГМЕНТАЦИЯ КЛИЕНТОВ БАНКА ПО ПОТРЕБЛЕНИЮ

- Анализ клиентов регионального банка
- Сегментация пользователей по количеству потребляемых продуктов
- Анализ оттока клиентов

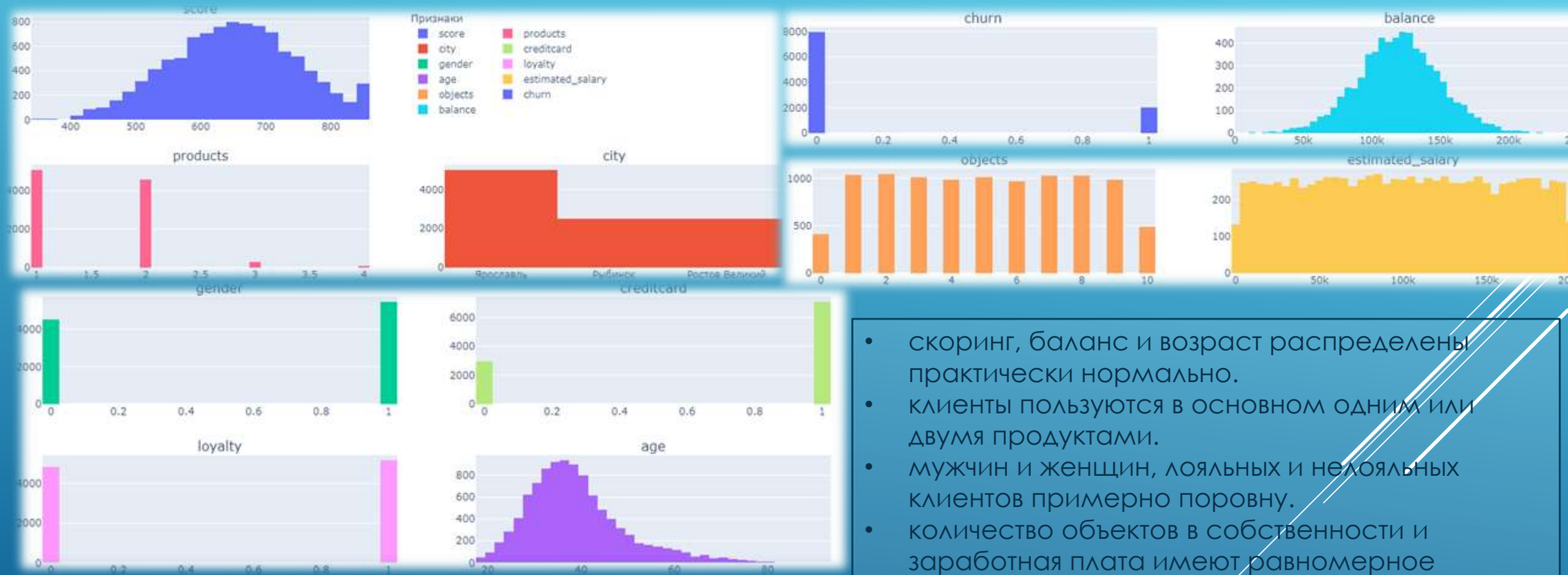
СЕГМЕНТАЦИЯ КЛИЕНТОВ БАНКА ПО ПОТРЕБЛЕНИЮ

- Анализ клиентов регионального банка
- Сегментация пользователей по количеству потребляемых продуктов
- Анализ оттока клиентов

ЦЕЛИ АНАЛИЗА

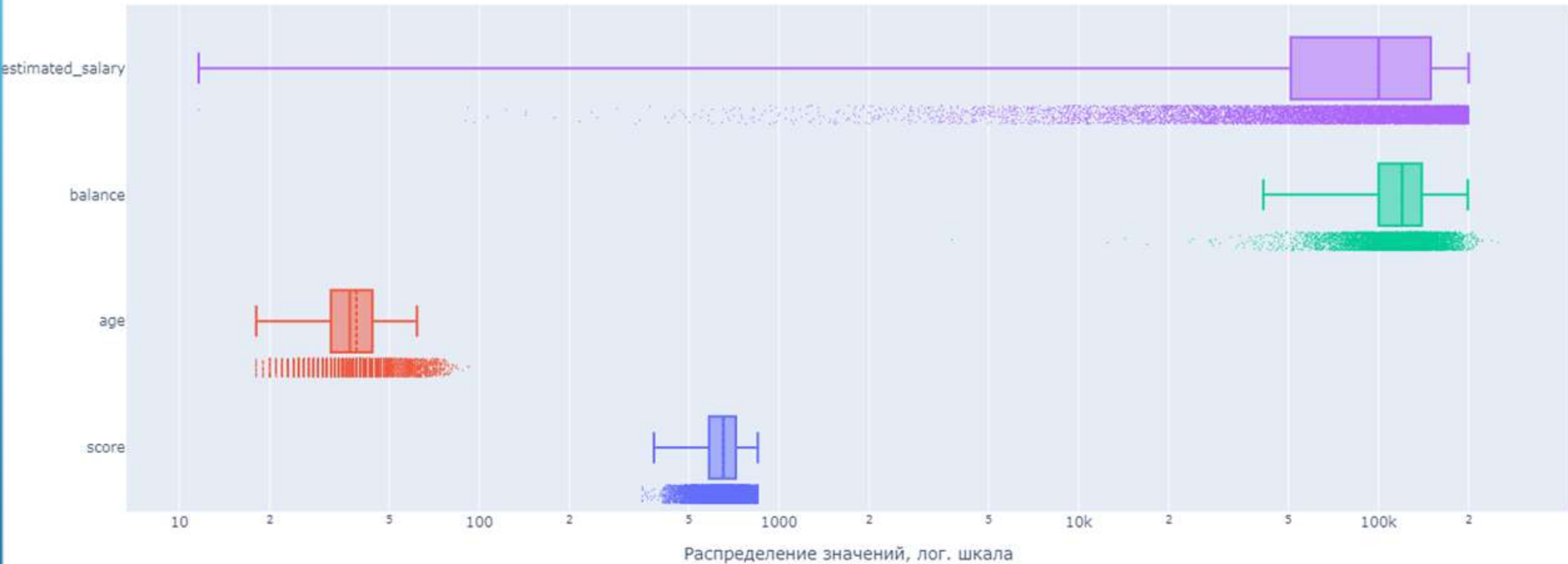
- ▶ Провести исследовательский анализ данных
- ▶ Сегментировать пользователей на основе данных о количестве потребляемых продуктов
- ▶ Сформулировать и проверить статистические гипотезы:
 - ▶ Гипотеза различия дохода между теми клиентами, которые пользуются двумя продуктами банка и теми, которые пользуются одним
 - ▶ Сформулировать и проверить статистическую гипотезу относительно представленных данных

РАСПРЕДЕЛЕНИЕ ПРИЗНАКОВ У КЛИЕНТОВ БАНКА

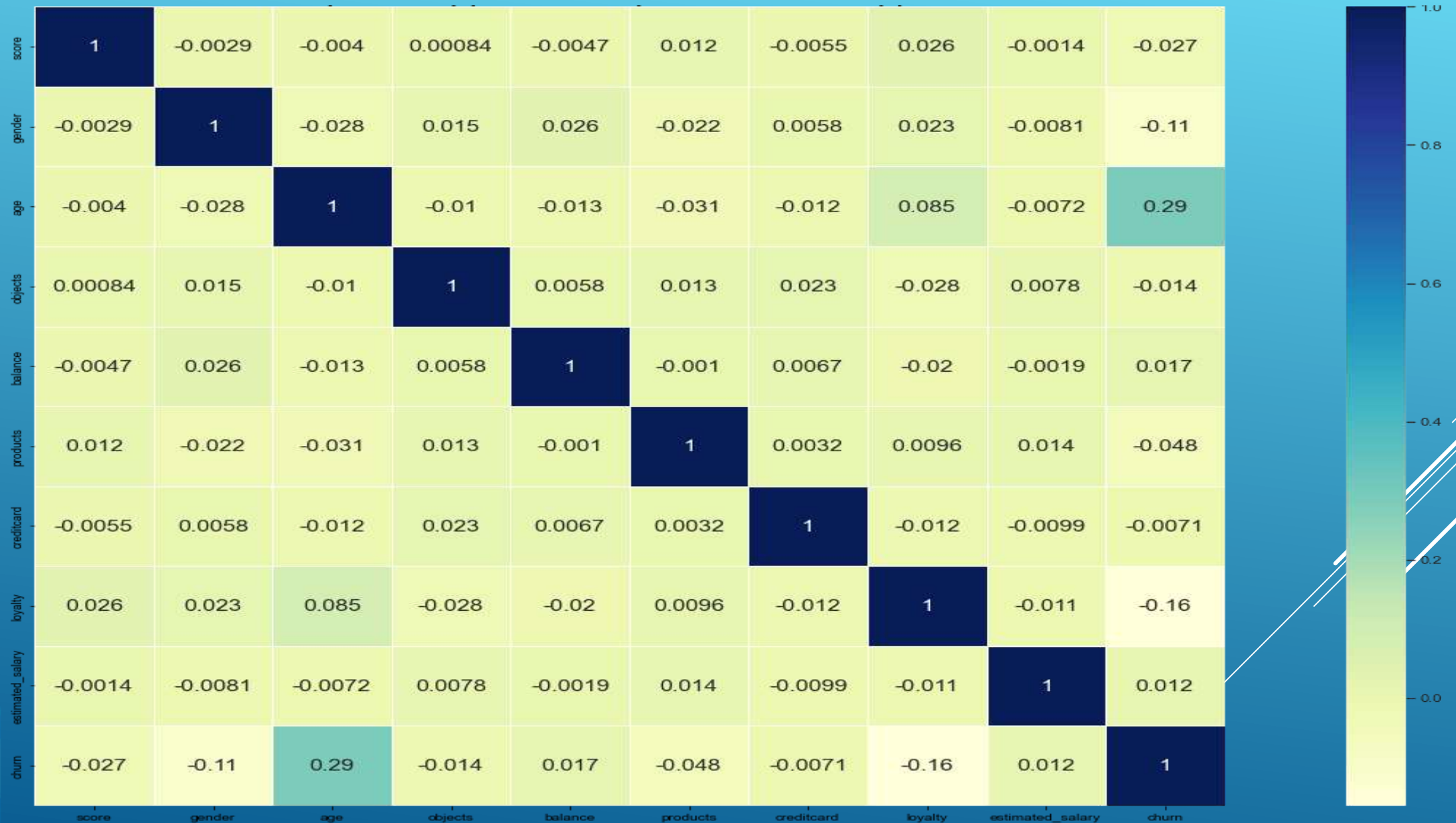


- скоринг, баланс и возраст распределены практически нормально.
- клиенты пользуются в основном одним или двумя продуктами.
- мужчин и женщин, лояльных и нелояльных клиентов примерно поровну.
- количество объектов в собственности и заработная плата имеют равномерное распределение.
- в оттоке находится примерно 20% клиентов.

РАЗМАХ РАСПРЕДЕЛЕНИЯ НЕБИНАРНЫХ ВЕЛИЧИН

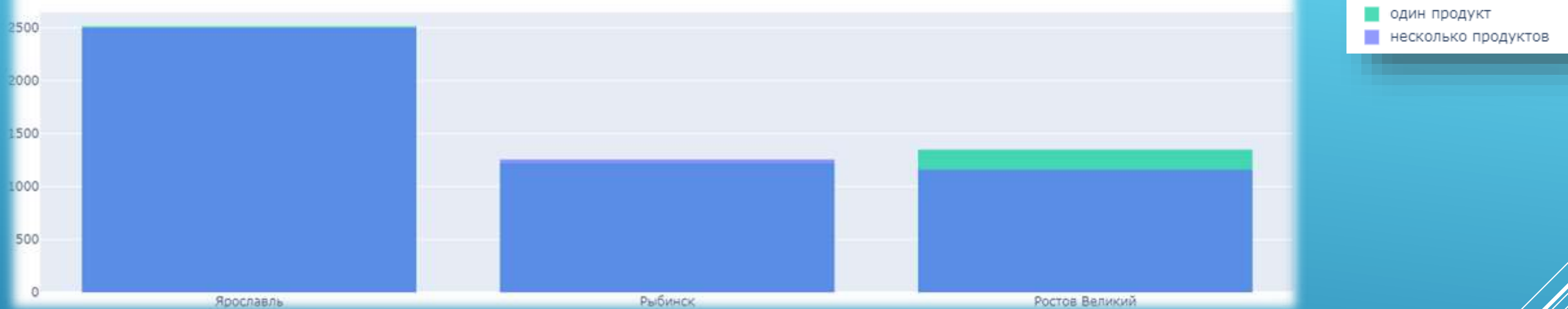


МАТРИЦА КОРРЕЛЯЦИЙ ПРИЗНАКОВ ДАТАФРЕЙМА

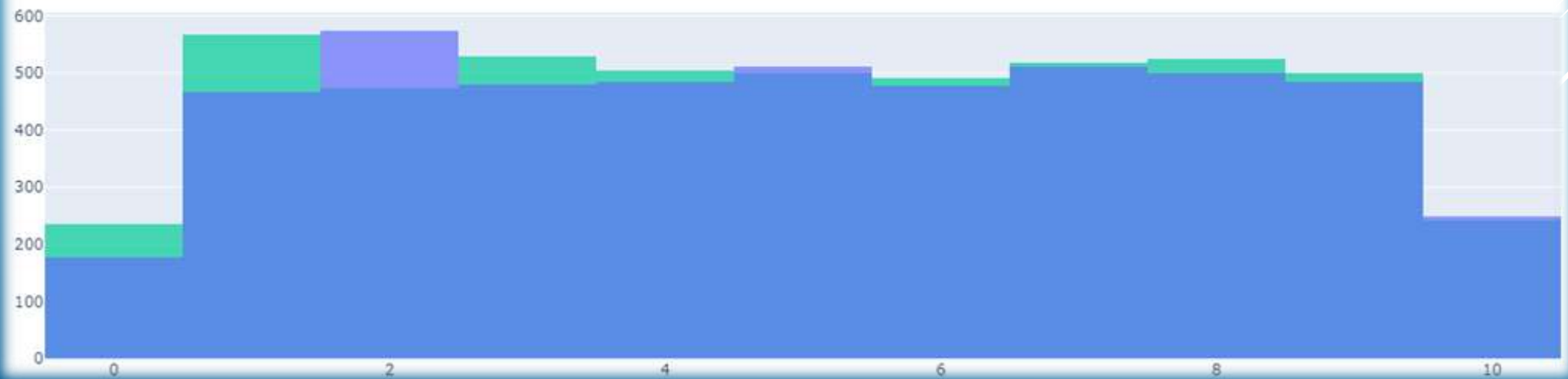


Сегментация клиентов по продуктам – распределение признаков

Распределение небинарных значений "city"

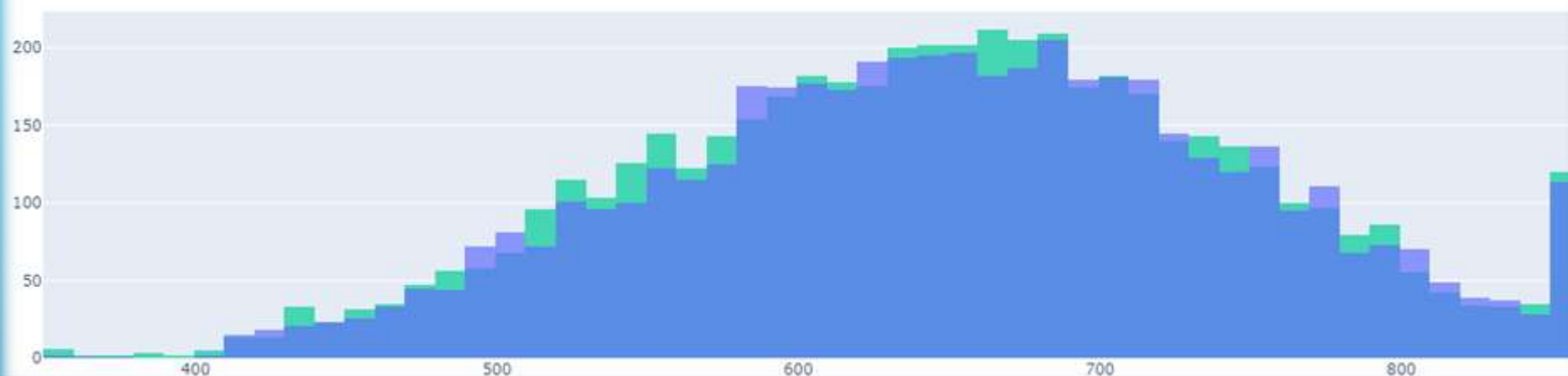


Распределение небинарных значений "objects"



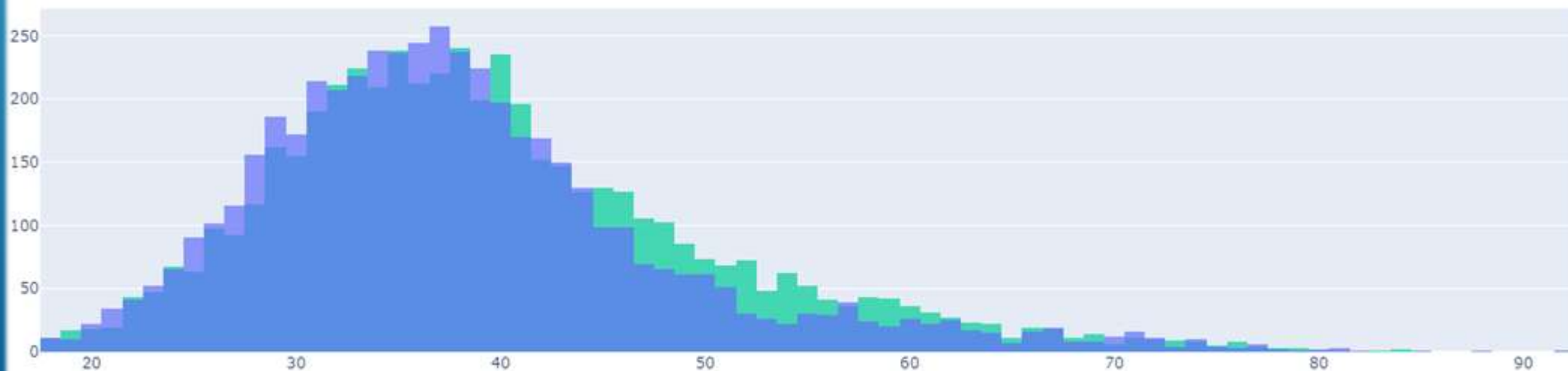
СЕГМЕНТАЦИЯ КЛИЕНТОВ ПО ПРОДУКТАМ – РАСПРЕДЕЛЕНИЕ ПРИЗНАКОВ

Распределение небинарных значений `score`



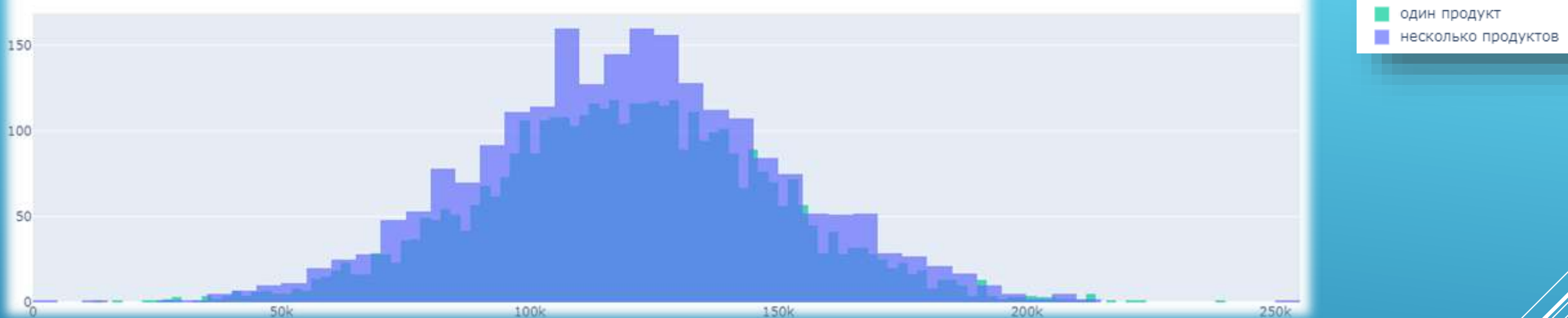
один продукт
несколько продуктов

Распределение небинарных значений `age`

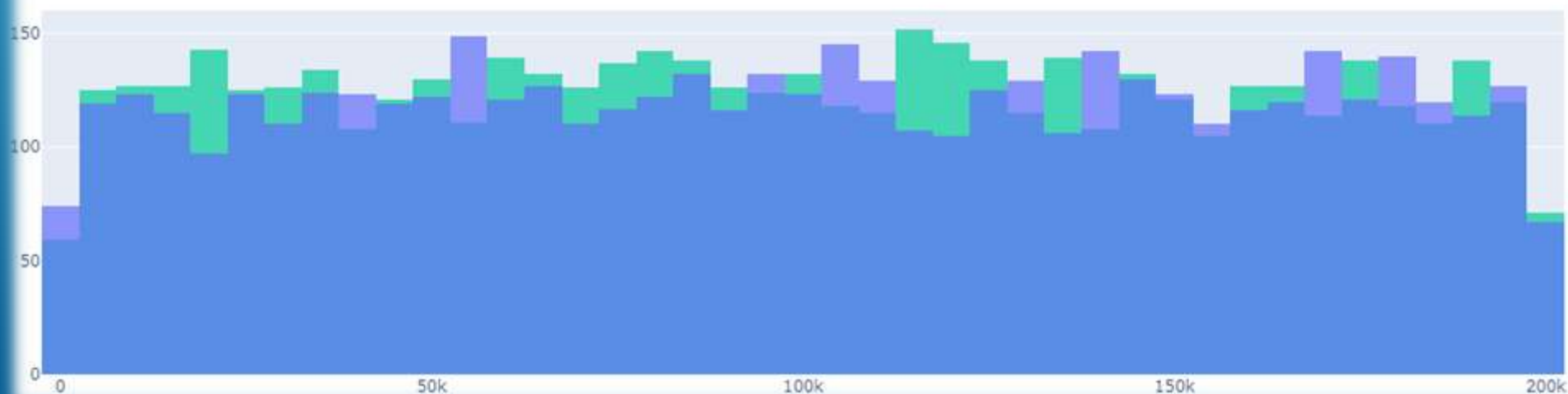


СЕГМЕНТАЦИЯ КЛИЕНТОВ ПО ПРОДУКТАМ – РАСПРЕДЕЛЕНИЕ ПРИЗНАКОВ

Распределение небинарных значений `balance`

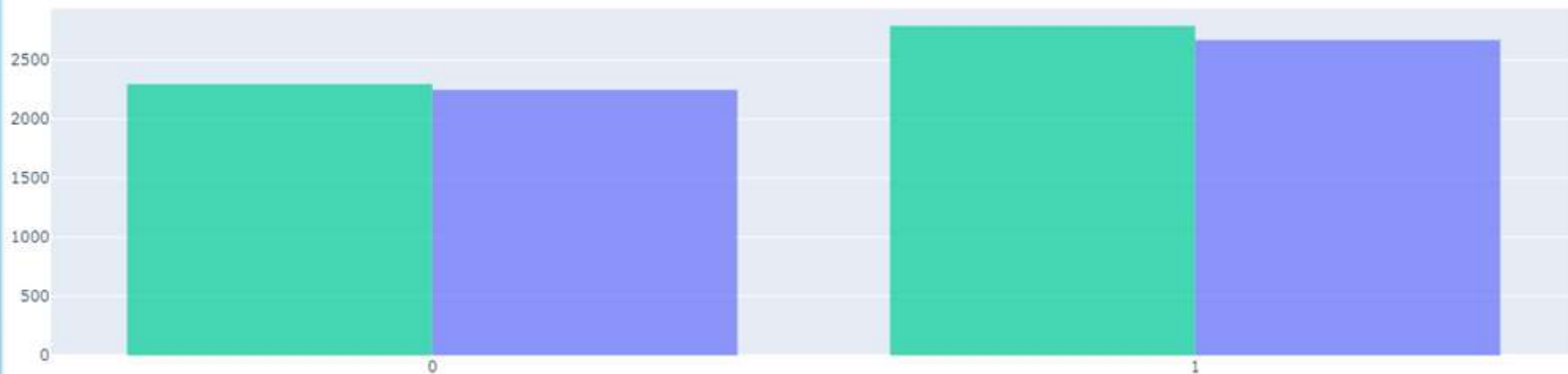


Распределение небинарных значений `estimated_salary`



СЕГМЕНТАЦИЯ КЛИЕНТОВ ПО ПРОДУКТАМ – РАСПРЕДЕЛЕНИЕ ПРИЗНАКОВ

Распределение бинарных значений `gender`



один продукт
несколько продуктов

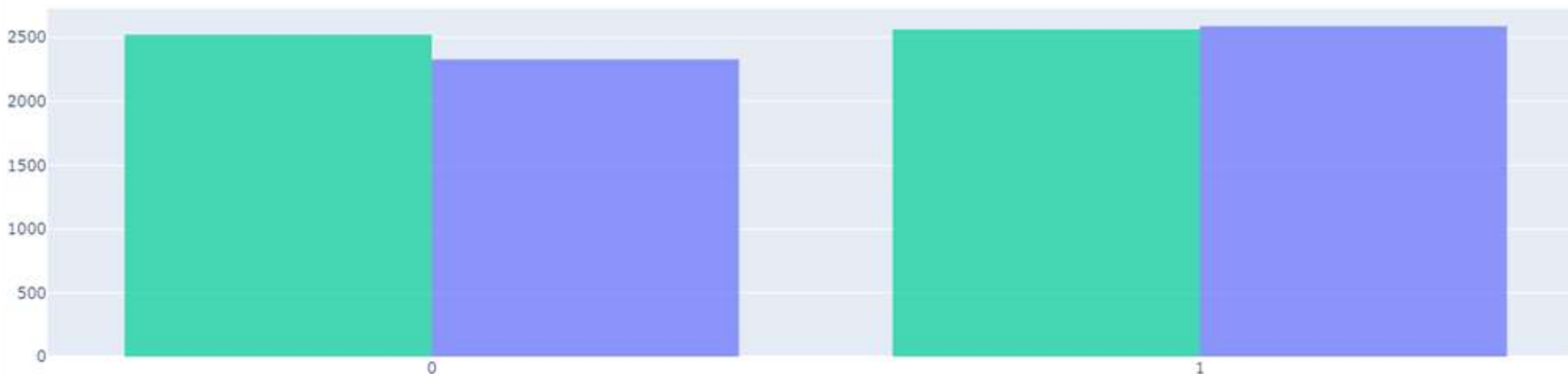
Распределение бинарных значений `creditcard`



АТМ

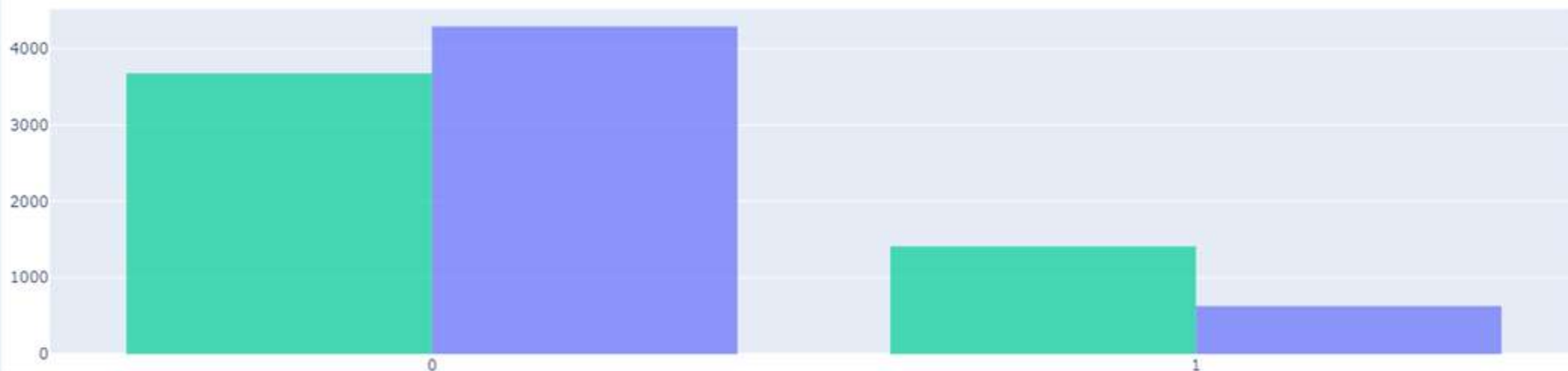
СЕГМЕНТАЦИЯ КЛИЕНТОВ ПО ПРОДУКТАМ – РАСПРЕДЕЛЕНИЕ ПРИЗНАКОВ

Распределение бинарных значений `loyalty`



один продукт
несколько продуктов

Распределение бинарных значений `churn`



СЕГМЕНТАЦИЯ КЛИЕНТОВ ПО ПРОДУКТАМ – РАСПРЕДЕЛЕНИЕ ПРИЗНАКОВ - ВЫВОДЫ

- *По небинарным признакам:*

- В Ростове Великом больше клиентов с одним продуктом.
- Среди клиентов с одним продуктом больше тех, к кого нет объектов недвижимости или такой объект всего один.
- Клиентов с одним продуктом с низким скорингом незначительно больше, чем клиентов с несколькими продуктами.
- Клиентов от 40 до 65 лет с одним продуктом больше, чем с несколькими.
- У людей с несколькими продуктами распределение значений баланса на счету сопоставим с клиентами, у которых один продукт.
- По заработной плате примерный паритет среди клиентов обеих групп.

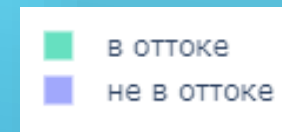
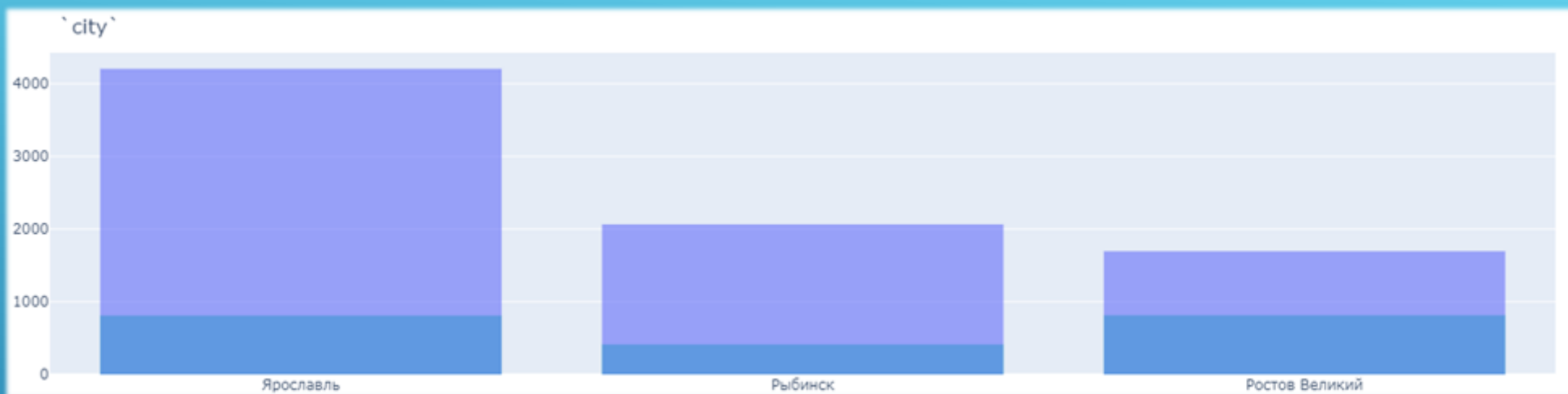
-

- *По бинарным признакам:*

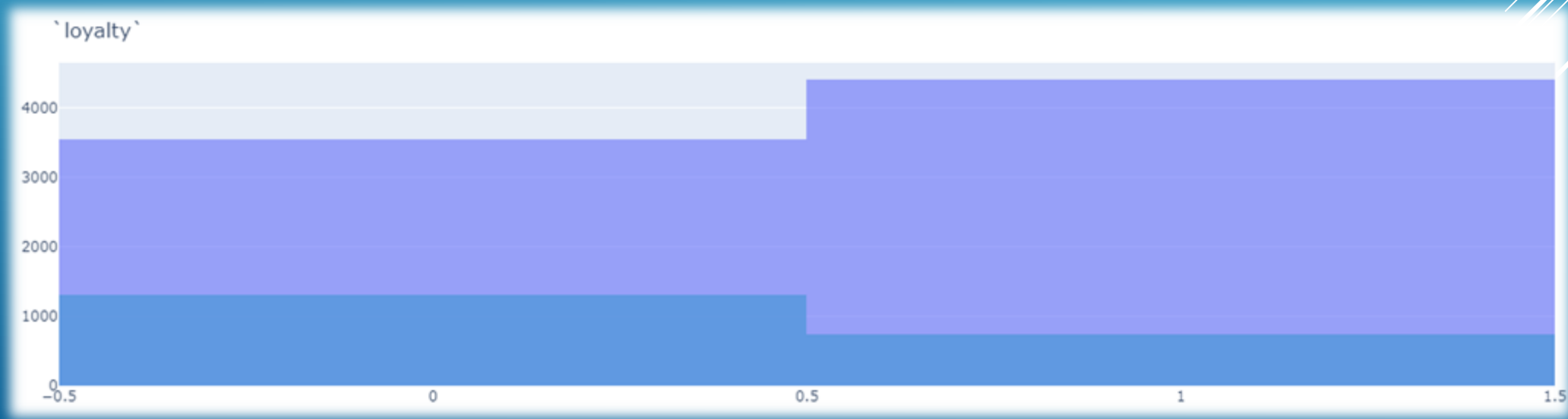
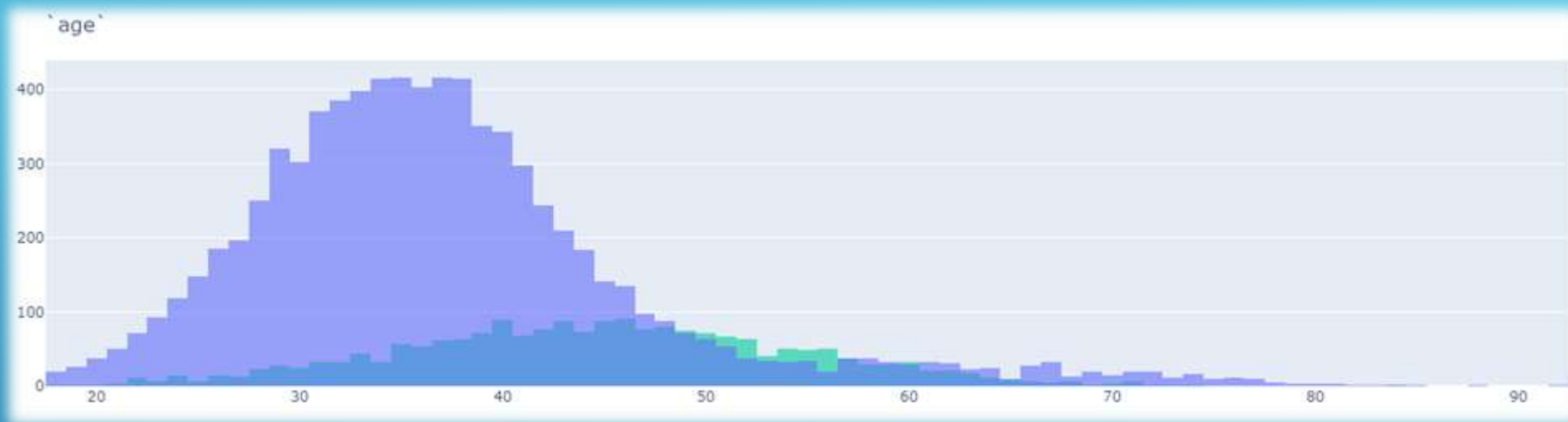
- Лояльных клиентов больше, у которых несколько продуктов.
- В отток попадают в основном клиенты с одним продуктом.



АНАЛИЗ РАСПРЕДЕЛЕНИЙ ДАННЫХ ДЛЯ КЛИЕНТОВ, ПОПАВШИХ И НЕ ПОПАВШИХ В ОТТОК – СУЩЕСТВЕННЫЕ **РАЗЛИЧИЯ**



АНАЛИЗ РАСПРЕДЕЛЕНИЙ ДАННЫХ ДЛЯ КЛИЕНТОВ, ПОПАВШИХ И НЕ ПОПАВШИХ В ОТТОК – СУЩЕСТВЕННЫЕ **РАЗЛИЧИЯ**



АНАЛИЗ РАСПРЕДЕЛЕНИЙ ДАННЫХ ДЛЯ КЛИЕНТОВ, ПОПАВШИХ И НЕ ПОПАВШИХ В ОТТОК – **ВЫВОДЫ**

- По городам - больше всего в отток попали клиенты из Ростова Великого.
- По гендерному признаку: в оттоке больше женщин, чем мужчин, не в оттоке - наоборот.
- По возрастам - распределение напоминает логнормальное со скошенностью справа. Средний возраст клиентов в оттоке смещен в большую сторону относительно не попавших в отток.
- По объектам владения и среднему доходу распределения похожи и напоминают равномерное.
- По балансу на счете, наличию кредитной карты распределения примерно похожи.
- Среди попавших в отток активных клиентов меньше, чем среди не попавших в отток.

РЕЗУЛЬТАТЫ ПРОВЕРКИ ГИПОТЕЗЫ И РАЗЛИЧИЯХ В ДОХОДЕ МЕЖДУ ГРУППАМИ

- Данные о доходе клиентов не имеют нормального распределения.
- На доступных данных не удалось отвергнуть нулевую гипотезу (`p-value` существенно выше критического уровня статистической значимости).
- **Скорее всего, различий в доходе между группами с одним и несколькими продуктами нет.**

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ – ГРУППЫ КЛИЕНТОВ

- Гипотеза 1:

«Между долями клиентов из разных групп (с одним и несколькими продуктами), попавших в отток по одному из критериев, нет статистически значимой разницы»

- Результат проверки ($\alpha = 0.01$):

Возможно, пол и лояльность являются характеристиками, немного влияющими на отток клиентов относительно двух групп (тех, у кого один продукт и тех, у кого их несколько).



ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

- Гипотеза 2:

«Между возрастными клиентами, попадающими в отток из разных групп, нет статистически значимой разницы»

- Результат проверки ($\alpha = 0.01$):

Между возрастными клиентами, попадающими в отток из разных групп, нет статистически значимой разницы

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ – ГРУППЫ КЛИЕНТОВ

- Гипотеза 3:

«Между долями клиентов из разных групп (с одним и несколькими продуктами), попавших в отток по одному из критериев, нет статистически значимой разницы»

- Результат проверки ($\alpha = 0.01$):

Возможно, пол и лояльность являются характеристиками, немного влияющими на отток клиентов относительно двух групп (тех, у кого один продукт и тех, у кого их несколько).

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ - ОТТОК

- Гипотеза 4:

«Между возрастaми клиентов, попадающих в отток из разных групп, нет статистически значимой разницы»

- Результат проверки ($\alpha = 0.01$):

Между возрастaми клиентов, попадающих в отток из разных групп, нет статистически значимой разницы.

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ - ОТТОК

- Гипотеза 5:

«Между возрастaми клиентов, попадающих в отток, нет статистически значимой разницы»

- Результат проверки ($\alpha = 0.01$):

Несмотря на схожие границы и стандартное отклонение возрастов для клиентов в оттоке и не в оттоке, средний возраст (как и медиана) имеют различия:

- Средний возраст клиента в оттоке - 45 лет.
- Средний возраст клиента НЕ в оттоке - 37 лет.

ВЫВОДЫ – ВОЗМОЖНЫЕ ПРИЧИНЫ ОТТОКА КЛИЕНТОВ

- 2 портрета клиентов:

- `Клиент, склонный к оттоку` - женщина в возрасте 45 лет, из любого из трех городов (но, скорее всего, из Ростова Великого), который имеет один продукт банка, является неактивным клиентом (клиент с низкой лояльностью).
- `Лояльный клиент` - мужчина в возрасте 37 лет, который имеет несколько продуктов банка и является активным клиентом.

- Возможные причины оттока:

- низкая экономическая активность в Ростове Великом,
- низкая активность клиентов,
- один банковский продукт,
- возраст клиента.

РЕКОМЕНДАЦИИ ПО УДЕРЖАНИЮ КЛИЕНТОВ

1. Внедрение CRM-системы для подготовки отчета по клиентам в разрезе активности их банковских счетов, с учетом возраста и пола клиента.
2. Включить дополнительную рассылку по электронной почте для таких клиентов с целью оповещения о новых продуктах банка.
3. Разработать возможность внедрения дополнительных продуктов банка на более льготных условиях, стимулирующих экономическую активность клиентов из небольших городов (Ростов Великий).
4. Для `**клиентов, склонных к оттоку**` разработать (и внедрить) программу лояльности, предлагающую дополнительные бонусы за пользование нескольких продуктов банка, а также за любую дополнительную активность.