

Анализ оттока клиентов

На наборе данных Bank Customer Churn Prediction



Содержание



01 Предметная область

Описание предметной области данных и описание набора данных

02 Инсайды

Полученные в ходе анализа данных и обучения модели

03 Интерактивный дашборд

Описание созданного интерактивного дашборда в Tableau

04 Бизнес интерпретация

Предположения о путях решения основных найденных проблем





1. Предметная область и данные

Банки стремятся удерживать клиентов, так как привлечение новых обходится дороже, чем сохранение существующих. **Отток клиентов** (churn) — это явление, когда клиенты закрывают счета или перестают пользоваться услугами банка. В данном датасете целевая переменная Exited (1 — клиент ушел, 0 — остался).

Кто считается "**ушедшим**" клиентом?

- Клиент закрыл все счета в банке.
- Клиент напрямую сообщил, что прекращает пользоваться услугами банка и вывел все средства.

Датасет: [Bank Customer Churn Prediction](#)

Основные характеристики клиентов:

- CreditScore – кредитный рейтинг.
- Geography – страна/регион.
- Age – возраст.
- Gender – пол.

Финансовые показатели:

- Balance – суммарный остаток на счетах.
- EstimatedSalary – предполагаемая зарплата.
- NumOfProducts – количество банковских продуктов.
- HasCrCard – наличие кредитной карты.

Поведенческие метрики:

- Tenure – срок сотрудничества с банком.
- IsActiveMember – активность.



2. Инсайды в данных



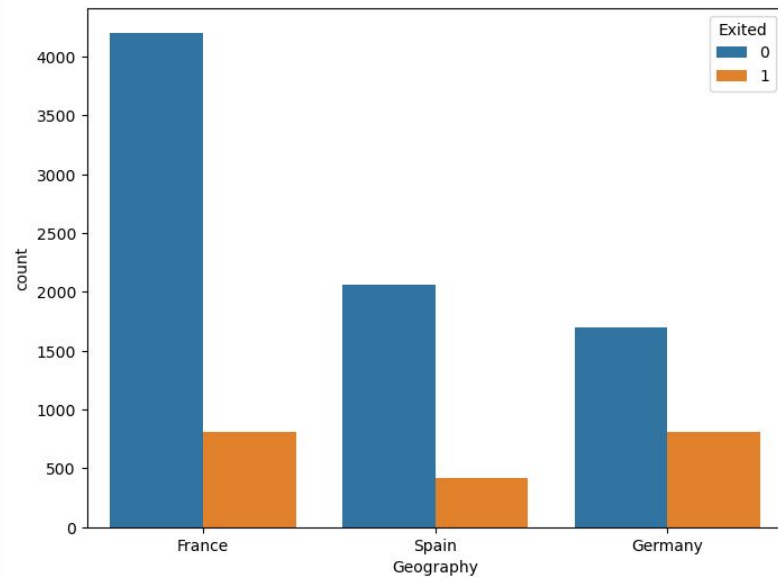
2.1 Отток по странам

В Германии доля ушедших клиентов составляет **32%**, в то время как во Франции и Испании доля ушедших не превышает 17%.

Разницу можно увидеть на графике. Синим отмечены не ушедшие клиенты, а оранжевым количество ушедших.

Для проверки влияет ли местоположение на отток был проведен статистический тест (подробнее в файле EDA.ipynb), показавший **статистически значимый результат**. Однако, связь является слабой.

Это означает, что страна влияет на отток, но это не главный фактор.



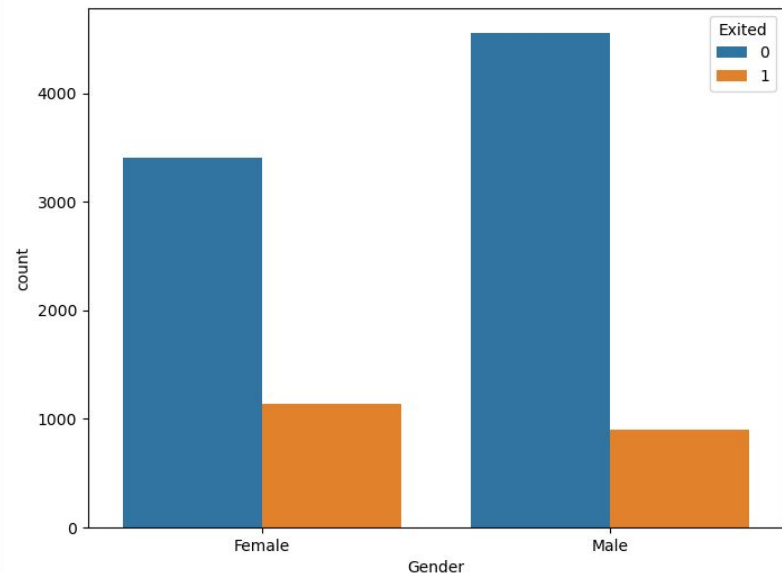
2.2 Отток в зависимости от пола

Доля ушедших клиентов среди женщин составляет **25%**, в то время как доля ушедших клиентов среди мужчин **16%**.

Разницу можно увидеть на графике. Синим отмечены не ушедшие клиенты, а оранжевым количество ушедших.

Для проверки влияет ли пол на отток был проведен статистический тест (подробнее в файле EDA.ipynb), показавший **статистически значимый результат**. Однако, связь также является слабой.

Это означает, что пол влияет на отток, но это не главный фактор.



2.3 Отток по возрастным группам

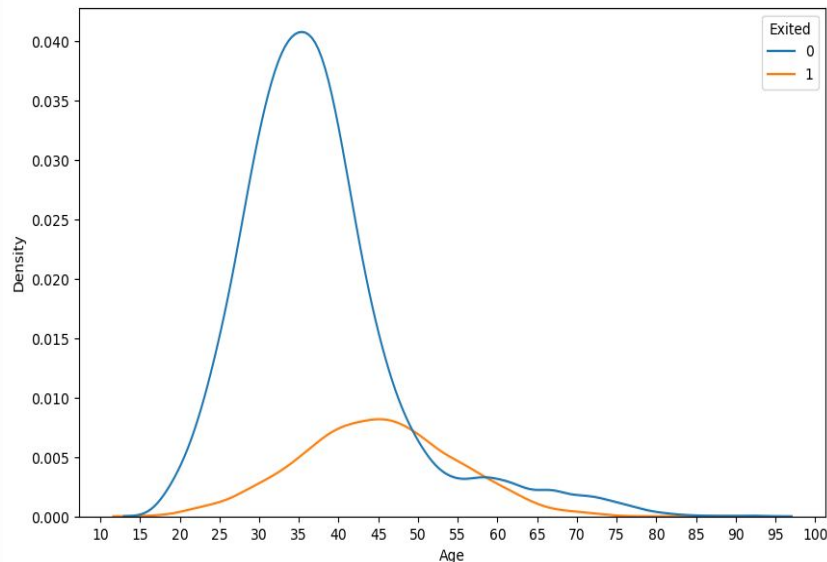
Доля ушедших клиентов в возрасте от 45 до 54 лет составляет **48%**, в возрасте более 55 лет **39%**.

Разницу можно увидеть на графике. Синим отмечена плотность не ушедших клиентов, а оранжевым плотность ушедших в зависимости от возраста.

На графике видим, что примерно с **49 и до 59 лет доля ушедших клиентов становится больше, чем доля оставшихся**.

Для проверки связи возраста с оттоком также был проведен статистический тест, **показавший статистически значимый результат и умеренную связь**.

Это означает, что возраст влияет на отток, при том довольно сильно.



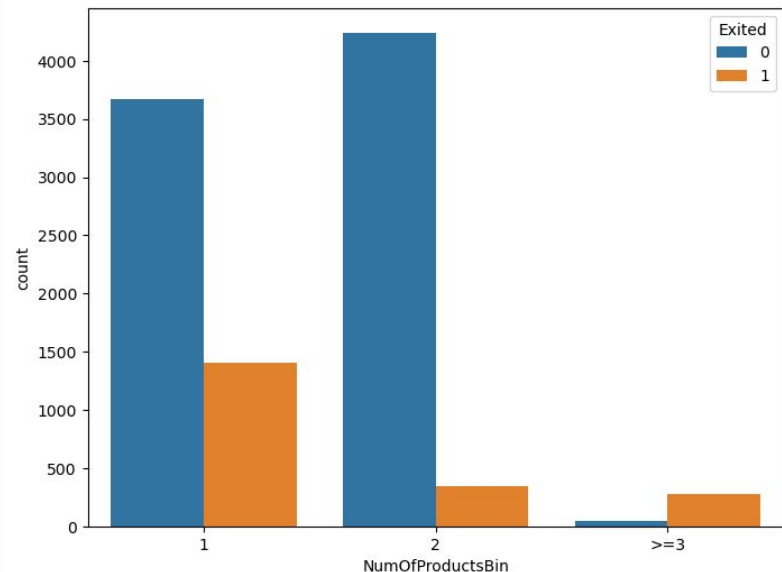
2.4 Отток по количеству продуктов

Доля ушедших клиентов с одним продуктом составляет **27%**, с тремя и более продуктами **85%**, в то время как с двумя всего **7%**.

Разницу можно увидеть на графике. Синим отмечены не ушедшие клиенты, а оранжевым количество ушедших.

Тестирование на взаимосвязь количества продуктов и оттока показало **статистически значимый результат и умеренную связь**.

Это означает, что количество продуктов влияет на отток достаточно сильно.



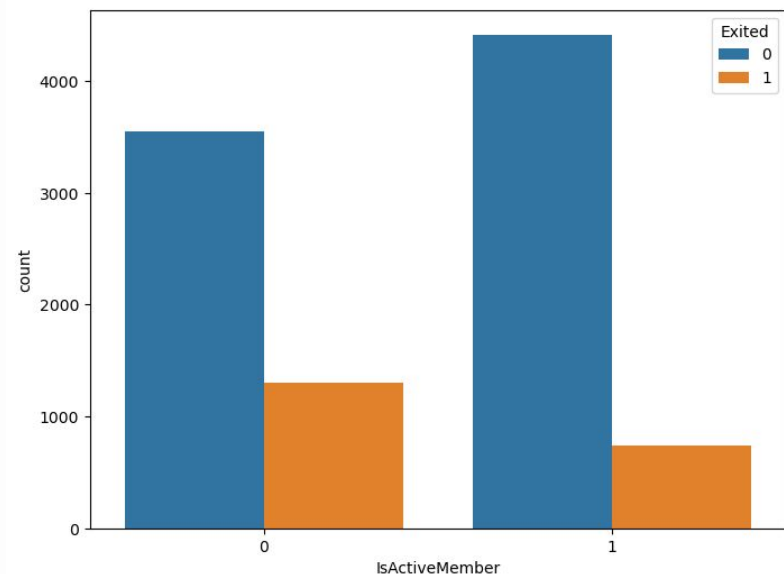
2.5 Отток в зависимости от активности

Доля ушедших неактивных клиентов составляет **27%**, в то время как активных **14%**.

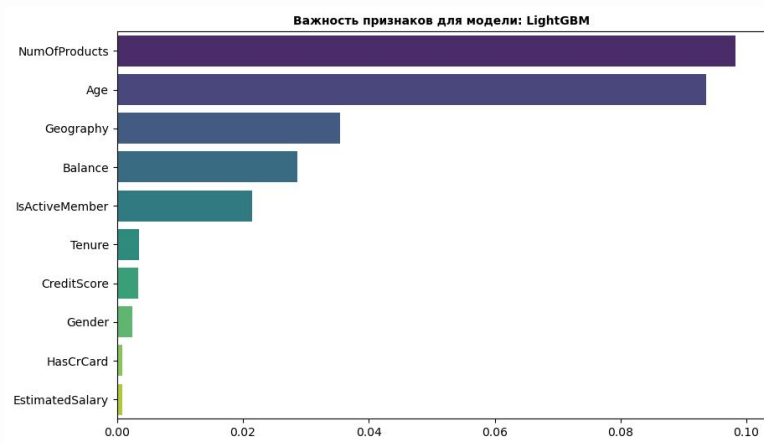
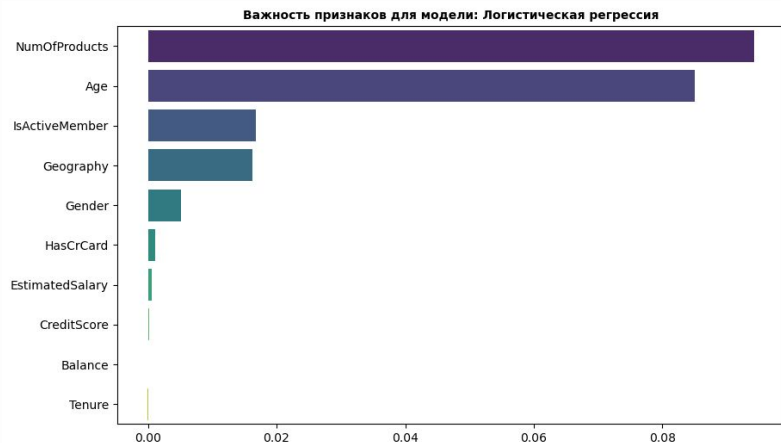
Разницу можно увидеть на графике. Синим отмечены не ушедшие клиенты, а оранжевым количество ушедших.

Тестирование показало наличие взаимосвязи, но слабую связь.

Это означает, что активность влияет на отток, но не является решающим фактором.



2.6 Важность признаков в обученных моделях



Были обучены две модели: логистическая регрессия и LightGBM.

По графикам видно, что обе модели выделили возраст и количество продуктов как самые важные признаки, также активность и местоположения пользователя тоже определились как значимые. Однако, пол обе модели не выделили как значимый признак.

3. Интерактивный дашборд





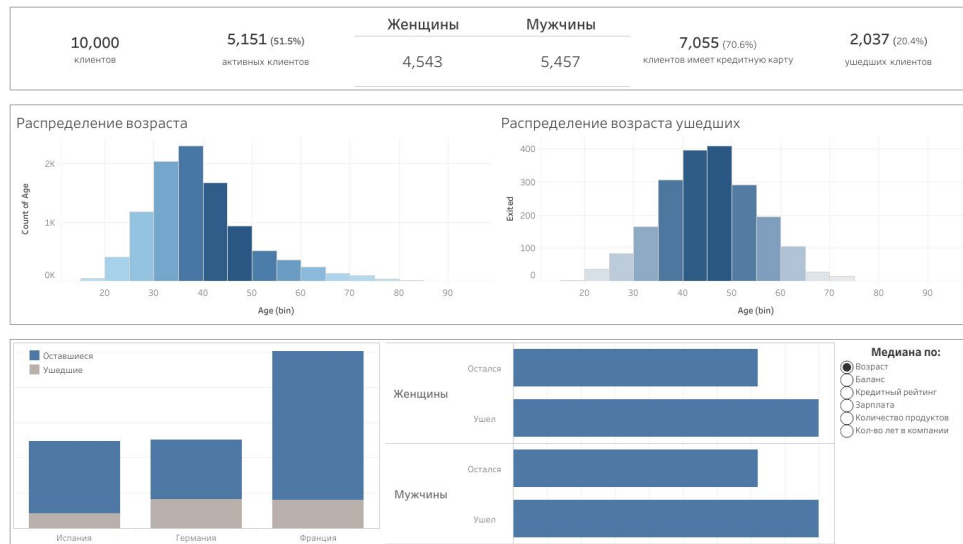
3. Интерактивный дашборд

На основе данных был создан интерактивный дашборд в Tableau. На нем можно увидеть все основные характеристики данных.

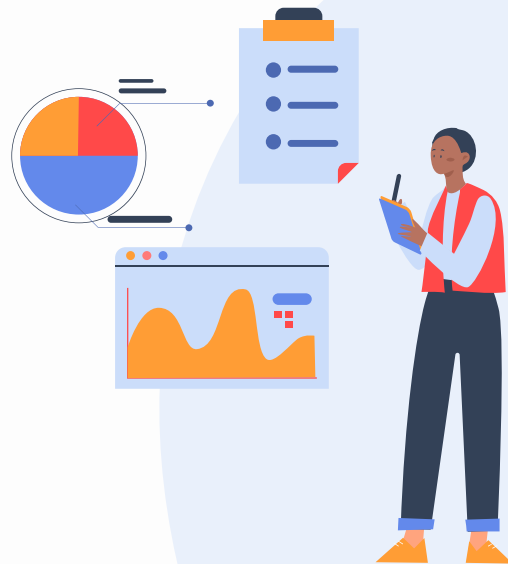
Столбцы всех графиков являются фильтрами, что позволяет удобно анализировать информацию о группах пользователей.

Повзаимодействовать с ним можно по ссылке: [Banking Churn Dashboard](#) (только с VPN).

Или открыв файл BankingChurnDashboard.twbx. В папке reports репозитория данного исследования.



4. Бизнес интерпретация





4. Бизнес интерпретация

Необходимо выяснить:

1. Кол-во продуктов

Почему клиенты с тремя и более продуктами уходят чаще. Список возможных причин:

- Дорогое обслуживание
- Проблемы с предыдущими продуктами, при создании нового

2. Возраст

Почему клиенты в возрасте с 45 лет начинают уходить чаще.

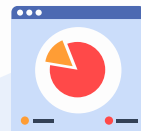
Возможные причины:

- Выгодные условия только для клиентов моложе
- Отсутствие оффлайн поддержки

3. Страна

Почему клиенты в Германии уходят чаще. Возможные причины:

- Наличие сильного конкурента
- Плохая локализация
- Отсутствие выгодных условий, которые есть в других странах



Спасибо за внимание!

Узнать технические подробности и ход анализа можно в репозитории данного проекта в ноутбуке EDA.ipynb и ModelTesting.ipynb

CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

