**Документация**

**Дипломная работа по теме:**

**«Анализ оттока клиентов банка (поиск инсайтов, составление рекомендаций стейкхолдерам)»**

Профессия “Аналитик данных”, DA-114

Бочериков Александр Сергеевич

г. Москва, 2025 года

Оглавление

Введение

Блок 1. Описание исходного датасета и типов данных

Блок 2. Проверка и изменение данных

Блок 3. Анализ данных для стейкхолдеров

3.1 гистограмма пользователей которые ушли из банка по возрасту

3.2 Соотношение лояльных и ушедший клиентов в долях

3.3 Графики по важным показателям

3.3.1 График возраста

3.3.2 График баланса на счете клиента

3.3.3 Статус активного клиента

3.3.4 Жалобы клиентов

3.3.5 Отток клиентов от числа приобретенных у банка услуг

3.3.6 Корреляционная диаграмма

Блок 4. Построение модель логистической регрессии

4.1 Цель логистической регрессии

4.2 Алгоритм модели

4.3 Оценка качества прогнозной модели

Итоги проекта и заключение

**Введение**

**Цели проекта:**

Произвести анализ клиентов склонных уходить из банка, найти ключевые данные которые взаимосвязаны с оттоком клиентов, построить модель линейной регрессии. Дать рекомендации различным бизнес-направлениям банка, подразделения сопровождения клиента.

**Бизнес-Задачи:**

- Выяснить основные причины оттока клиентов.

- Найти сегменты клиентов, которые в наибольшей степени склонны уходить из банка

- Определить, какие факторы чаще всего служат причиной потери клиентов

**Стейкхолдер**

**Руководство банка** - увеличение хороших клиентов и их сохранение, а также понимание с какими подразделениями работать в банке.

**Аналитики банка** - подготовка аналитических отчётов для руководства, мониторинг финансовых рынков и конкурентной среды, оценка клиентов.

**Менеджеры сопровождения** - анализ оттока клиентов на предмет жалоб, а также типа людей которые склонны уходить из банка.

**Блок 1. Описание исходного датасета и типов данных (18 столбцов)**

Для исследования был взят датасет “[**Bank Customer Churn**](https://www.kaggle.com/datasets/radheshyamkollipara/bank-customer-churn/data)” со статистикой клиентов банка

| № | Имя Столбца | описание | Тип данных |
| --- | --- | --- | --- |
| 1 | RowNumber | номер строки | int64 |
| 2 | CustomerId | ID клиента | int64 |
| 3 | Surname | фамилия клиента | object |
| 4 | CreditScore | кредитный рейтинг | int64 |
| 5 | Geography | местоположение | object |
| 6 | Gender | пол клиента | object |
| 7 | Age | возраст клиента | int64 |
| 8 | Tenure | сколько лет клиент пользуется услугами банка | int64 |
| 9 | Balance | баланс на счетах клиента в банке | float64 |
| 10 | NumOfProducts | количество услуг банка, которые приобрел клиент | int64 |
| 11 | HasCrCard | есть ли у клиента кредитная карта (1 — да, 0 — нет) | int64 |
| 12 | IsActiveMember | есть ли у клиента статус активного клиента банка (1 — да, 0 — нет) | int64 |
| 13 | EstimatedSalary | предполагаемая заработная плата клиента | float64 |
| 14 | Exited | статус лояльности (1 — ушедший клиент, 0 — лояльный клиент) | int64 |
| 15 | Complain | у клиента есть жалоба или нет (1 — есть, 0 — нет) | int64 |
| 16 | Satisfaction Score | оценка, предоставляемая клиентом за разрешение его жалобы | int64 |
| 17 | Card Type | тип карты, которой владеет клиент | object |
| 18 | Points Earned | баллы, заработанные клиентом за использование кредитной карты | int64 |

**Блок 2. Проверка и изменение данных**

В ходе исследования качества данных было сделано следующее:

**2.1 Проверка данных**

* название столбцов (без изменений)
* тип данных в столбцах (без изменений)
* данные на пропуски (без изменений)
* на наличие дубликатов (без изменений)
* столбцы с бинарными значениями (без изменений)
* проверка на выбросы (изменения)

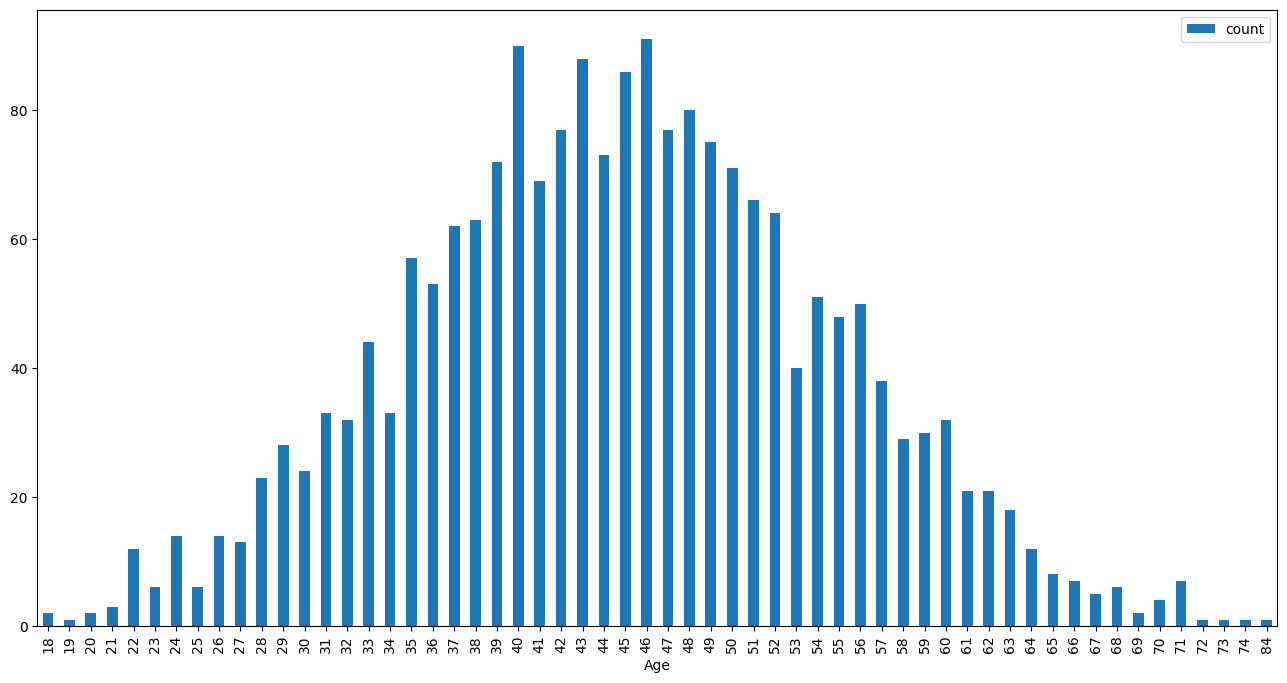
**2.2 Изменения данных**

Был удален один пользователь из за выброса в столбце (EstimatedSalary) предполагаемая заработная плата клиента

**Блок 3.** **Анализ данных для стейкхолдеров**

Целью блока является поиск тенденций и инсайтов для составления рекомендаций стейкхолдерам.

**3.1 Гистограмма пользователей которые ушли из банка по возрасту**

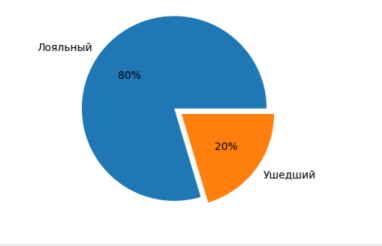


**вывод :** явных выбросов нет, нам это говорит что есть связь возраста и ушедшим клиентом

### По таблице видно, что наибольшая корреляция с показателем оттока клиентов :

* age связь между возрастом и оттоком клиентов
* balanse
* isActiveMember статус активного клиента
* Complain жалобы клиентов
* NumOfProducts кол-во услуг в банке

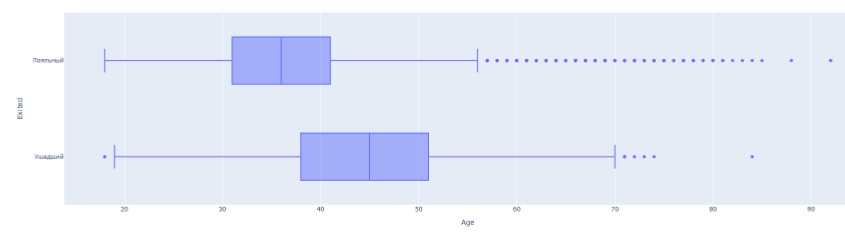
**3.2** **Соотношение лояльных и ушедший клиентов в долях**



Соотношение лояльных и ушедших клиентов банка 1к4

**3.3 Графики на основе корреляционной диаграммы**

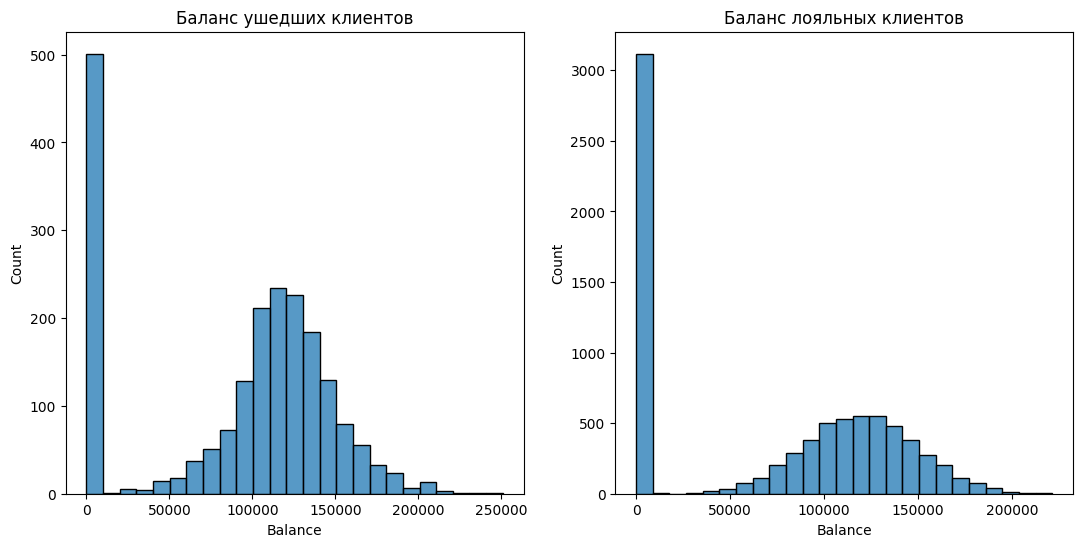
**3.3.1 График возраста**

****

Вывод:

На графики видно что у ушедших клиентов меньше выбросов это говорит о том что есть связь с возрастом

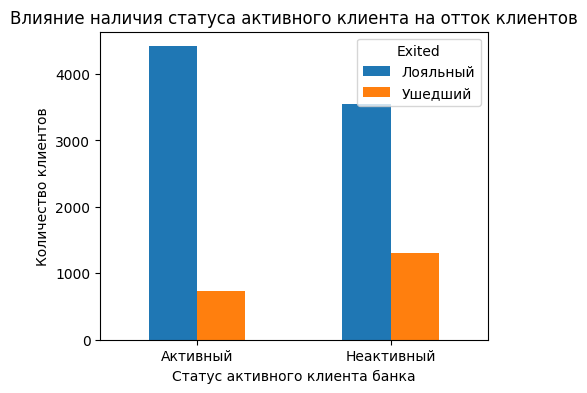
**3.3.2 График баланса на счете клиента**

****

Вывод:

На графики видно что большое кол-во людей с 0 счетом и баланс ушедших клиентов больше.

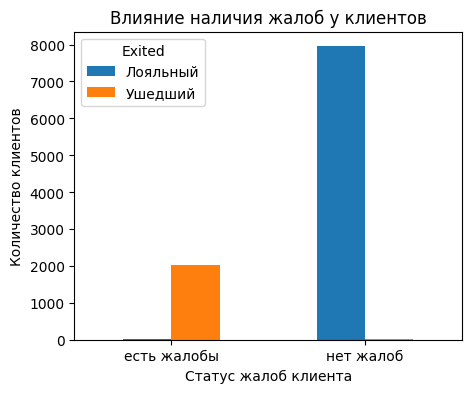
**3.3.3 Статус активного клиента**

****

Вывод:

На графики видно чем неактивнее пользователь тем отток клиентов увеличивается

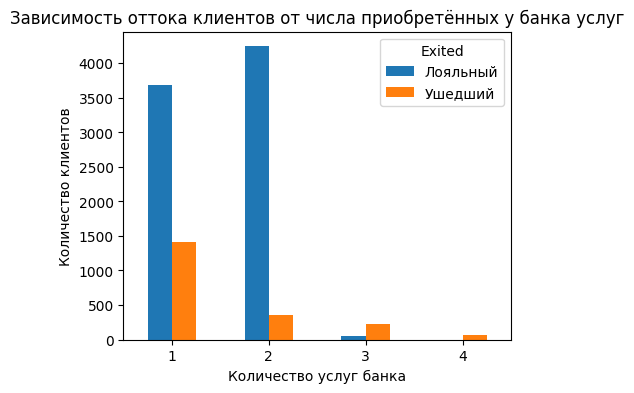
**3.3.4 Жалобы клиентов**

****

Вывод:

Кол-во клиентов которые ушли и имеют жалобы на момент ухода 2000 это 20% от общего числа клиентов.

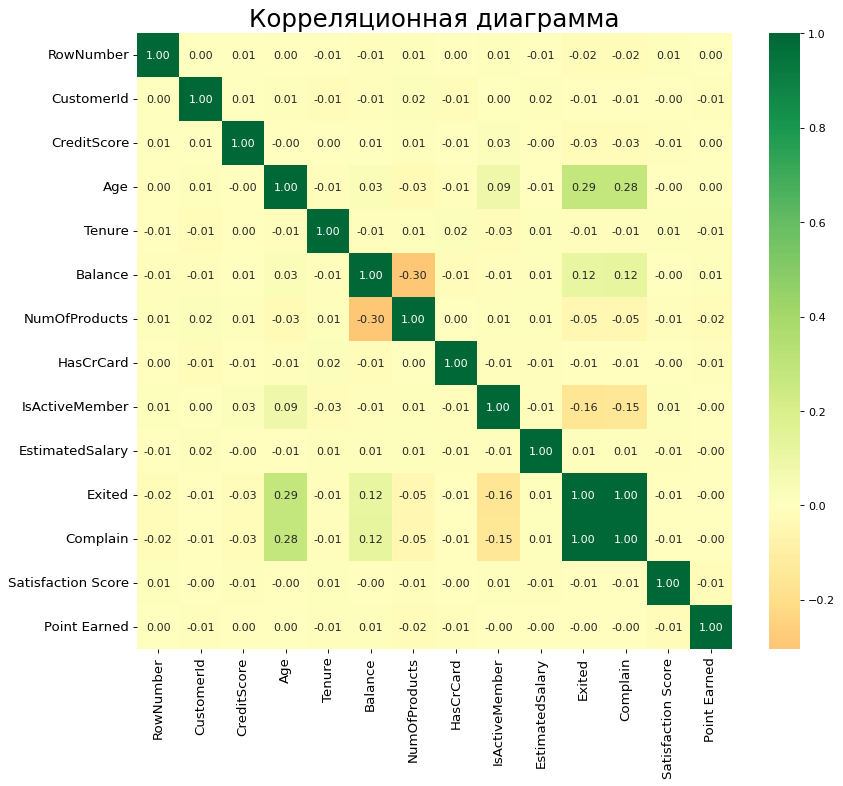
**3.3.5 Кол-во услуг в банке**

****

Вывод:

Чем больше приобретенных услуг у клиента тем меньше происходит отток.

**3.3.6** **Корреляционная диаграмма**



### По таблице видно, что наибольшая корреляция с показателем оттока клиентов :

* age связь между возрастом и оттоком клиентов
* balanse
* isActiveMember статус активного клиента
* Complain жалобы клиентов
* NumOfProducts кол-во услуг в банке

**Блок 4. Построение модели логистической регрессии**

**4.1 Цель логистической регрессии**

Цель регрессии заключается в предсказании значение одной переменной ‘Exited’ на основе переменных 'Age', 'Tenure', 'Balance', 'NumOfProducts', 'IsActiveMember', 'EstimatedSalary'.

**4.2 Алгоритм модели**

Создали переменную с признаками и переменную результирующую данные переменные разделили на тестовую выборку (20%) и главную выборку (80%).

Обучили модель логистической регрессии

**4.3 Оценка качества прогнозной модели**

Прогноз результатов тестового набора и вычисление точности показывают хорошую точность самой модели. Коэффициентом детерминации тестовой выборки и главной выборки схожи это нам говорит что выборка репрезентативна. коэффициентом детерминации 0,8 указывает на приемлемую точность модели.

**Вывод:** Логистическая модель показала себя хорошо поэтому на основе ее можно предсказывать переменную 'Exited'.

**Итоги проекта и заключение**

**Итог:**

* возраст ушедших клиентов сосредоточен 38-51 год
* баланс ушедших клиентов больше
* неактивные пользователи чаще уходят
* у ушедших клиентов большое кол-во жалоб
* чем меньше услуг использует клиент, тем больше увеличивается отток клиентов

**Вывод:**

* обратить на клиентов от 38 - 51 года, проработать индивидуальный подход, обрабатывать все жалобы.
* проработать качество обслуживания, условия обслуживания, а также программы лояльности
* мониторить рынок услуг, делать специальные предложения клиентам
* предоставление дополнительных услуг и сервисов
* организация мероприятий и акций
* клиенты с 0 счетом убедить их в надежности банка, предоставить хорошие условия по вкладам, получать отзывы от клиентов,обеспечить качественное обслуживание. Проработать ожидания клиента, на этом строить предложения клиенту.

**Заключение**:

Обращать внимание на клиентов от 38 лет до 51 года, проработать индивидуальные предложения, отслеживать жалобы клиентов и эффективность обработки таких жалоб. Увеличить специальные предложения банка, мониторить рынок предложений и информировать о новых предложениях и условиях банка. Клиентов с 0 счетом информировать о выгодных условиях и предоставить выгодные предложения по вкладам.