**Документация по проекту**

**Дипломная работа по теме:**

**"Исследование оттока клиентов банка (поиск инсайтов, составление рекомендаций стейкхолдерам, построение дашборда)"**

Профессия “Аналитик BI”, ABU-34

Круглова Виктория Дмитриевна

Новосибирск, 2023

Оглавление

[1. Постановка задачи 3](#_Toc93590297)

[1.1. Описание бизнес-задачи и ее актуальности 3](#_Toc93590298)

[1.2. Круг стейкхолдеров, сбор бизнес-требований по задаче, гипотезы для проверки 3](#_Toc93590299)

[1.3. Метрики для проверки гипотез, источники информации для сбора данных для расчета таких метрик 4](#_Toc93590300)

[2. Анализ данных 5](#_Toc93590301)

[2.1. Описание исходных данных 5](#_Toc93590302)

[2.2. Алгоритмы и техники, применяемые для решения задачи, их требования к исходным данным 5](#_Toc93590303)

[2.3. Форма представления итоговых результатов 5](#_Toc93590304)

[3. Методика решения 6](#_Toc93590305)

[3.1. Шаги преобразования и очистки данных 6](#_Toc93590306)

[3.2. Выбор методов работы с данными и метрик для решения бизнес-задачи, построена система метрик 6](#_Toc93590307)

[3.3. Сценарий проверки гипотез 7](#_Toc93590308)

[3.4. Модель данных 7](#_Toc93590309)

[4. Результаты 8](#_Toc93590310)

[5. Выводы 10](#_Toc93590311)

1. **Постановка задачи**
   1. **Описание бизнес-задачи и её актуальности**

Заказчик дашборда – банк.

Цель банковской деятельности – получение/максимизация финансовой прибыли, одним из способов достижения этой цели является максимизация количества клиентов, что в свою очередь декомпозируется на два направления: привлечение новых клиентов и удержание уже существующих. Удержание клиентов в целом обходится для банка дешевле, чем привлечение новых, т.к. новых ещё нужно изучить и привлечь, а об уже существующих клиентах есть информация, они уже были привлечены.

Кроме того, о существующих клиентах уже накоплены сведения, которые позволяют предлагать им новые банковские продукты с большей вероятностью того, что они примут предложения.

Поставлены задачи:

1. Выделить сегменты клиентов, склонных к оттоку, для которых впоследствии можно будет разработать планы маркетинговых мероприятий по их удержанию.
2. Подтвердить или опровергнуть ряд гипотез о зависимости клиентов к оттоку от различных факторов.
   1. **Круг стейкхолдеров, сбор бизнес-требований по задаче, гипотезы для проверки**

**Стейкхолдеры:**

**-** топ-менеджмент банка;

**-** отдел аналитики;

**-** отдел маркетинга.

**Бизнес-требования:**

Создать отчет в виде дашборда, на котором должны быть отражена зависимость клиентов, склонных к уходу из клиентов банка от различных факторов: Age (возраст), Gender (пол), Geography (место проживания), CreditScore (кредитный рейтинг), Balance (остаток на счете), EstimatedSalary (дохода), Tenure (стажа в качестве клиента банка), IsActiveMember (активности), NumOfProducts (количества приобретенных банковских продуктов), HasCrCard (наличия кредитной карты).

**Гипотезы:**

1. Клиент с низким кредитным рейтингом с большей вероятностью покинет банк.
2. Пожилые (клиенты старших возрастных групп) реже покидают банк, чем молодые.
3. Пол клиентов не оказывает значительного влияния на склонность покидать банк.
4. География клиентов оказывает влияние на склонность покидать банк.
5. Чем больше стаж клиента (время, в течение которого он является клиентом банка), тем реже он покидает банк.
6. Чем больше остаток на счете клиента, тем меньше вероятность, что он покинет банк.
7. Чем больше у клиента продуктов банка, тем меньше вероятность, что он покинет банк.
8. Клиенты, которые используются кредитную карту банка, менее склонны покидать банк.
9. Активные клиенты банка реже покидают банк.
10. Клиенты с более низким доходом чаще покидают банк.
    1. **Метрики для проверки гипотез, источники информации для сбора данных для расчета таких метрик**

Семейство метрик: зависимость процентного соотношения числа ушедших клиентов от общего числа клиентов от различных факторов (пол, возраст, география, наличие кредитной карты, кредитный рейтинг, баланс, доход, активность, стаж клиента).

1. **Анализ данных**
   1. **Описание исходных данных**

Источник данных: [Churn for Bank Customers](https://www.kaggle.com/datasets/mathchi/churn-for-bank-customers)

*Табл. Исходные данные*

|  |  |  |
| --- | --- | --- |
| **Показатель** | **Описание** | **Проблемы** |
| RowNumber | Порядковый номер | -- |
| CustomerId | Идентификационный номер клиента | -- |
| Surname | Фамилия клиента | -- |
| CreditScore | Кредитный рейтинг клиента | Текстовый тип данных |
| Geography | Местоположение клиента | -- |
| Gender | Пол клиента | -- |
| Age | Возраст клиента | Текстовый тип данных |
| Tenure | Стаж клиента в качестве клиента банка | Текстовый тип данных |
| Balance | Остаток денежных средств на счете клиента | Текстовый тип данных |
| NumOfProducts | Количество приобретенных клиентом банковских продуктов | Текстовый тип данных |
| HasCrCard | Наличие у клиента кредитной карты | Текстовый тип данных |
| IsActiveMember | Активность клиента | Текстовый тип данных |
| EstimatedSalary | Доход клиента | Текстовый тип данных |
| Exited | Покинул банк или нет | Текстовый тип данных |

**2.2. Алгоритмы и техники, применяемые для решения задачи, их требования к исходным данным**

Алгоритмы и техники: *дашборд*.

Требования к исходным данным:

- корректное указание типов данных;

- отсутствие ошибок в данных.

**2.3. Форма представления итоговых результатов**

Дашборд, выполненный в MS Power BI.

1. **Методика решения**
   1. **Шаги преобразования и очистки данных**

*Табл. Преобразование и очистка*

|  |  |
| --- | --- |
| **Показатель** | **Шаги** |
| CreditScore | Заменен тип данных на целочисленный |
| Age | Заменен тип данных на целочисленный |
| Tenure | Заменен тип данных на целочисленный |
| Balance | Знак “.” заменен на “,”,  тип данных заменен на целочисленный |
| NumOfProducts | Заменен тип данных на целочисленный |
| HasCrCard | Произведена замена значений: “0” на “нет”, “1” на “да” |
| IsActiveMember | Произведена замена значений: “0” на “нет”, “1” на “да” |
| EstimatedSalary | Знак “.” заменен на “,”,  тип данных заменен на целочисленный |

**3.2. Выбор методов работы с данными и метрик для решения бизнес-задачи**

**1) построена система метрик:**

* Количество клиентов – посчитано количество уникальных CustomerID в исходной таблице
* Количество ушедших клиентов – посчитано количество клиентов в контексте поля Exited
* Процент оттока клиентов – посчитан как частное от деления количества ушедших клиентов на общее количество клиентов.

**2) в Power Query созданы дополнительные столбцы:**

* Пользовательский столбец для разбивки по возрастным группам

*(4 возрастные группы определены следующим образом 18-25, 26-39, 40-59 и старше 60 лет – по ключевым задачам, стоящим перед группой: 18-25 – студенты, начинающие специалисты; 26-39 – молодые семьи; 40-59 – люди с большей финансовой свободой); старше 60 - пенсионеры и предпенсионеры)*

* *Условный столбец для разбивки по размеру баланса (4 группы: 1) от 0 до 50 тыс, 2) от 50 до 100 тыс, 3) от 100 до 200 тыс; 4) больше 200 тыс.)*

**3.3. Сценарий проверки гипотез**

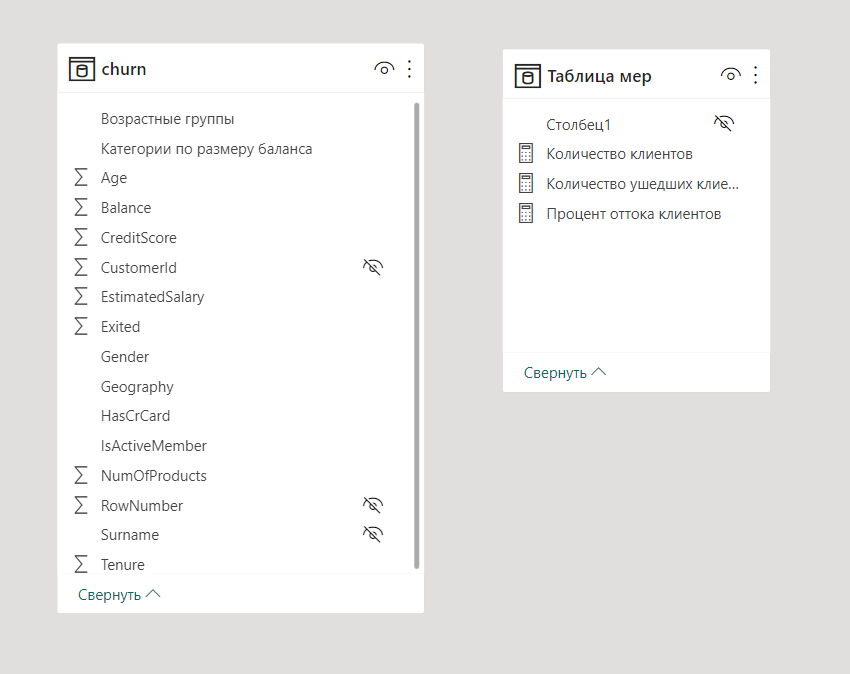
Гипотезы проверяются путем сравнения значений метрик при установке необходимых фильтров в дашборде, фильтры выведены в виде срезов на дашборд.

**3.4. Модель данных**

Модель данных включает 2 таблицы: исходную таблицу и таблицу мер.

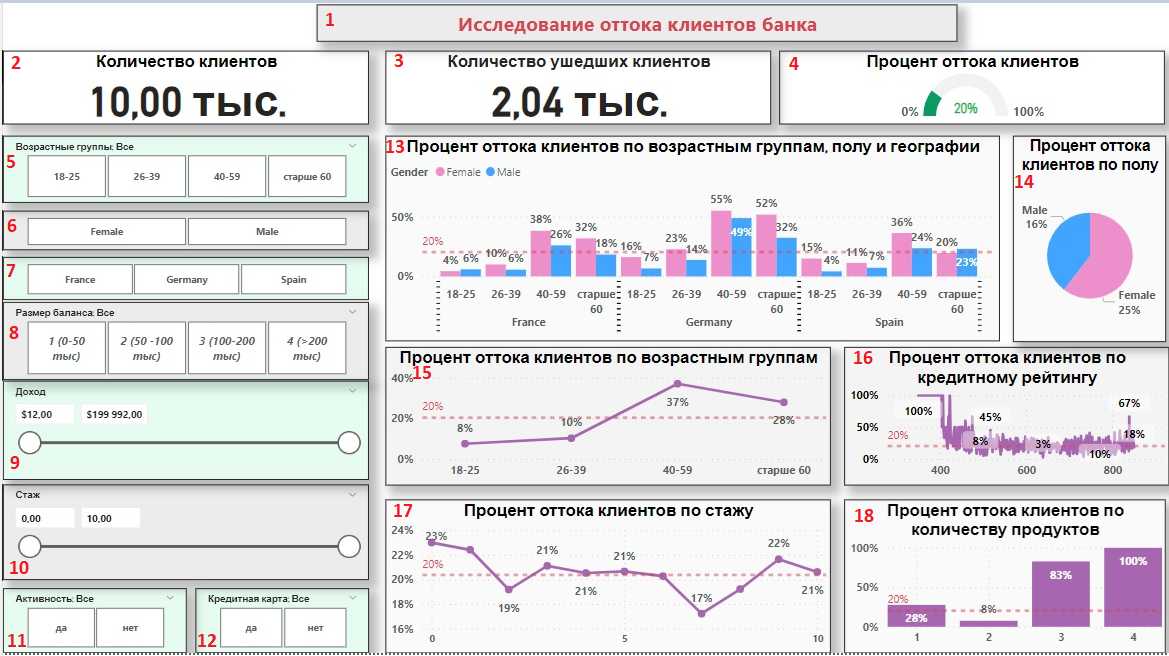
Связей между таблицами нет*.*

*Таблицы*



1. **Результаты**

Дашборд имеет 18 визуальных элементов, пронумерованных цифрами красного цвета. На Рисунке 1 представлен внешний вид дашборда.



*Рисунок 1. Внешний вид дашборда*

Далее описан каждый из 18-ти визуальных элементов:

1. Заголовок дашборда «Исследование оттока клиентов банка».
2. Показатель в виде карточки: «Количество клиентов», который представляет из себя общее количество клиентов с учетом выставленных пользователем фильтров.
3. Показатель в виде карточки: «Количество ушедших клиентов», который представляет из себя количество ушедших клиентов с учетом выставленных пользователем фильтров.
4. Показатель в виде датчика: «Процент оттока клиентов», который представляет из себя процентное соотношение ушедших клиентов от общего количества клиентов с учетом выставленных пользователем фильтров.
5. Срез данных в виде плитки по возрастным группам, где пользователь может выбрать нужный вариант.
6. Срез данных в виде плитки по полу, где пользователь может выбрать нужный вариант.
7. Срез данных в виде плитки по географии, где пользователь может выбрать нужный вариант.
8. Срез данных в виде плитки по размеру баланса, где пользователь может выбрать нужный вариант.
9. Срез данных по доходу, где пользователь может установить интересующий диапазон дохода. На данный момент минимальное значение 12 ($), максимальное значение 199 992 ($).
10. Срез данных по стажу, где пользователь может установить интересующий диапазон стажа. На данный момент минимальное значение 1 (год), максимальное значение 10 (лет).
11. Срез данных в виде плитки по активности клиентов, где пользователь может выбрать один из двух вариантов.
12. Срез данных в виде плитки по наличию кредитной карты, где пользователь может выбрать один из двух вариантов.
13. Гистограмма с группировкой, показывающая процент оттока клиентов с разбивкой по возрастным группам, по полу и по географии с учетом выставленных пользователем фильтров.
14. Круговая диаграмма, показывающая соотношение процента оттока клиентов по полу с учетом выставленных пользователем фильтров.
15. График изменения процента оттока клиентов по возрастным группам с учетом выставленных пользователем фильтров. К графику реализовано подключение подсказки с детализацией внутри возрастной группы.
16. График изменения процента оттока клиентов по кредитному рейтингу с учетом выставленных пользователем фильтров.
17. График изменения процента оттока клиентов по стажу с учетом выставленных пользователем фильтров.
18. Гистограмма с накоплением, демонстрирующая изменение процента оттока клиентов по количеству банковских продуктов с учетом выставленных пользователем фильтров.

Элементы имеют следующие взаимодействия:

- Визуализации 2,3,4 отражают все изменения, которые происходят при выборе данных по срезам (5-12) и при выборе элементов с визуализаций 13-18.

- Визуализации 13,14,15,16,17,18 могут фильтровать друг друга.

- Все визуализации в виде срезов (5-12) являются фильтрами для визуализаций 2-4, 13-18.

1. **Выводы**

Из 10 тысяч клиентов банка ушли 2 тысячи клиентов, процентный отток составил 20%. Предлагаю исходить из этой цифры как ориентира, который может показать, насколько больше/меньше отток клиентов по определенному фактору.

**Подтвержденные гипотезы:**

1. Клиент с низким кредитным рейтингом с большей вероятностью покинет банк. Да, это так, но эта вероятность работает для клиентов с самым низким кредитным рейтингом – от 350 до 500. После 500 рост кредитного рейтинга не оказывает значительного влияния на уход клиентов.
2. География клиентов оказывает влияние на склонность покидать банк. Действительно клиенты из Германии чаще покидают банк (32%), чем клиенты Франции (16%) и Испании (17%).
3. Активные клиенты банка реже покидают банк. Да, действительно, отток активных клиентов составляет 14%, в то время как неактивных – 27%.

**Неподтверждённые гипотезы:**

1. Пожилые (клиенты старших возрастных групп) реже покидают банк, чем молодые. Строго наоборот. Наибольший отток клиентов характерен для возрастной группы 40-59 лет – 37%
2. Пол клиентов не оказывает значительного влияния на склонность покидать банк. Напротив, женщины (25%) покидают банк чаще, чем мужчины (16%). Причём во всех возрастных категориях и географических регионах, при любых фильтрах.
3. Чем больше стаж клиента (время, в течение которого он является клиентом банка), тем реже он покидает банк. Не выявлено такой зависимости. Процент оттока клиентов изменяется от 17 до 23, что в среднем составляет 20%. Единственное, можно обратить внимание, что клиенты со стажем до 1 года покидают банк чаще (23%). А клиенты со стажем 3, 7 и 8 лет несколько реже.
4. Чем больше остаток на счете клиента, тем меньше вероятность, что он покинет банк. Напротив, клиента с остатком меньше 50 тыс наименее часто покидают банк (процент оттока 14%), клиенты же с остатком от 100 тыс до 200 тыс и более 200 тыс покидают банк чаще (25% и 56% соответственно).
5. Чем больше у клиента продуктов банка, тем меньше вероятность, что он покинет банк. Наоборот, клиенты с 4 и 3 продуктами покидают банк чаще (100% и 83%). Меньше всего склонны к оттоку клиенты с 2 продуктами (8%).
6. Клиенты, которые используются кредитную карту банка, менее склонны покидать банк. Не выявлено существенной разницы в оттоке между клиентами с кредитной картой (20%) и без кредитной карты (21%).
7. Клиенты с более низким доходом чаще покидают банк. Не выявлено зависимости оттока клиентов от величины дохода.

Можно составить портрет клиента, который чаще всего покидает банк. Это женщина возрастной группы 40-59 лет из Германии с балансом от 100 до 200 тыс, не активная обладательница кредитной карты. Для этой категории доля оттока 65%, но есть момент, что таких клиентов всего 173 человека.

В противовес портрет клиента, который не склонен покидать банк. Это мужчина 18-25 лет из Испании с балансом на счёте от 0 до 50 тыс, не активный.

**Рекомендации:**

1. Добавить данные по LTV (Lifetime Value), узнать сколько прибыли приносит клиент банку, т.к. важно понимать, что отток клиентов не равен количеству потерянных денег (упущенной выгоды) и проанализировать с этой точки зрения, какой бюджет на удержание той или иной категории клиентов стоит тратить.
2. Добавить данные по временному разрезу, чтобы посчитать коэффициент оттока клиентов и проанализировать, какая динамика оттока клиентов вообще от года к году и внутри года по сезонности, соответственно можно будет определить время, когда лучше проводить акции по удержанию клиентов.
3. Продумать маркетинговые активности для женской клиентуры, клиентов после 40 лет, обладателей баланса от 100 до 200 тыс.
4. Не перегружать клиентов банковскими продуктами, 2 банковских продукта – оптимально.
5. Стимулировать у клиентов активность.