

Задание 1 «Сбор требований»

Бабенко Ксения

1. Цель:

Сравнить каналы подключения (личный кабинет, офис продаж, площадки партнёров) по среднему количеству подключённых банковских продуктов на одного клиента за 2024 год и определить их относительный вклад в кросс-продажи.

Задача:

Подготовить аналитический отчёт с данными о среднем количестве подключённых продуктов на клиента по каждому каналу и долей каналов в общем числе подключений за 2024 год:

1. Определить источники данных, содержащие информацию о подключениях продуктов и каналах подключения.
2. Собрать и проверить данные о подключениях банковских продуктов за период с 01.01.2024 по 31.12.2024.
3. Сгруппировать данные по клиентам и каналам подключения.
4. Рассчитать среднее количество подключённых продуктов на одного клиента по каждому каналу.
5. Рассчитать долю каждого канала в общем числе подключений и в числе клиентов, подключивших 2+ продукта.
6. Подготовить отчёт с визуализациями, отражающими сравнение каналов и их вклад в кросс-продажи.
7. Представить результаты заинтересованным сторонам с интерпретацией и рекомендациями.

Ограничения:

1. Технологическое окружение. Проект реализуется в рамках существующего технологического стека: DWH на базе PostgreSQL, ETL с использованием Airflow, визуализация — Power BI. Доступ к данным возможен только через DBeaver и BI-платформу — прямые подключения к источникам запрещены политикой безопасности.
2. В анализ включаются только три канала подключения с чётко определёнными источниками данных:

- личный кабинет (**IB_SYSTEM**),
- CRM-система отделений — офис продаж (**CRM_BANK**),
- площадки партнёров — внешние системы: API внешних партнёров (**PARTNER_API_**), файловые выгрузки от партнёров (**PARTNER_FILE_**).

Другие каналы (например, колл-центр, мобильное приложение) не учитываются.

3. Анализ ограничен данными за период с 01.01.2024 по 31.12.2024. Более ранние данные не рассматриваются.

4. Не учитываются кросс-канальные сценарии (например, клиент узнал о продукте в офисе, но подключил в личном кабинете), так как такая информация не фиксируется в системах.

5. Отчёт не оценивает качество подключений (например, отток клиента в течение 30 дней после подключения) и уровень удовлетворённости клиентов.

6. Не анализируются затраты на каналы (например, зарплаты сотрудников, маркетинговые бюджеты), поэтому экономическая эффективность (ROI) не рассчитывается.

7. Проект не включает сбор первичных данных от клиентов (опросы, интервью), только анализ вторичных данных из систем.

8. Витрина проектируется только для агрегированного анализа по каналам, а не для детального поведенческого анализа (например, время сессии, количество отказов, путь клиента)

Зависимости:

1. Наличие в системах-источниках обязательных атрибутов для анализа:

- В **IB_SYSTEM**, **CRM_BANK**, **PARTNER_API_***, **PARTNER_FILE_*** должно быть поле **channel_cd** (канал подключения) со значением, соответствующим одному из трёх каналов: **IB**, **BRANCH**, **PARTNER**.
- Для каждого подключения должен быть указан **client_id** или, при его отсутствии, полный набор реквизитов для идентификации клиента: ФИО, серия и номер паспорта, дата рождения.

- Дата подключения должна фиксироваться в поле `activation_dttm` (дата и время активации продукта).

2. Регулярная и полная передача данных от партнёрских систем:

- Интерфейсы с партнёрами должны обеспечивать передачу данных ежедневно в согласованном формате (JSON, CSV или через API).
- Файловые выгрузки должны размещаться на SFTP-сервере в каталоге, доступном для ETL-процесса.
- API-интерфейсы должны возвращать данные по запросу с кодом ответа 200 и в согласованной структуре.

3. Доступ к BI-платформе с правами на создание и публикацию отчётов. Должен быть предоставлен доступ к BI-системе (Power BI) с возможностью подключения к витрине данных и публикации дашборда.

4. Утверждение бизнес-логики расчёта ключевых метрик:

- Момент подключения = дата активации продукта (`activation_dttm`)
- Канал подключения = значение из поля `channel_cd` в источнике
- Уникальный клиент = идентификация по `client_id`, при его отсутствии — по комбинации: ФИО + паспорт + дата рождения
- Подключённый продукт = продукт со статусом "активен"

Допущения:

1. Атрибут "канал подключения" (`channel_cd`) в источниках заполнен полностью и соответствует одному из трёх значений: `IB`, `BRANCH`, `PARTNER`. Отсутствуют пропуски, некорректные значения или смешанные форматы (например, `online`, `branch`, `partner_site`).

2. Уникальный идентификатор клиента (`client_id`) доступен в `IB_SYSTEM` и `CRM_BANK`, а в данных от партнёров — передаются достаточные реквизиты (ФИО, паспорт, дата рождения) для однозначного сопоставления с клиентом банка.

3. Все подключения продуктов зафиксированы в системах-источниках. Не существует "ручных" или неучтённых операций, влияющих на полноту данных.

4. Данные от партнёрских систем передаются регулярно (ежедневно) и в согласованном формате. Изменения структуры выгрузок или простои в передаче данных не ожидаются.
5. Дата активации продукта (**activation_dttm**) является достоверным и единым показателем момента подключения. Альтернативные даты (например, дата заявки) не используются для анализа.
6. Данные в DWH обновляются ежедневно, и задержка поступления данных не превышает 24 часа. На момент формирования отчёта информация актуальна.
7. Бизнес-логика расчёта метрик (среднее число продуктов на клиента, доля каналов) будет согласована до начала реализации витрины и не будет изменяться в ходе проекта.
8. Результаты анализа будут использоваться заинтересованными сторонами для принятия решений по развитию каналов подключения. Проект имеет ценность для бизнеса и не является разовым запросом.

2. Список заинтересованных лиц

Заинтересованное лицо	Роль / подразделение	Интерес / ожидания от проекта
Руководитель отдела аналитики	Заказчик проекта	Получить объективные данные для стратегических решений по каналам продаж
Руководитель цифровых каналов	Отвечает за онлайн-продукты	Оценить эффективность личного кабинета, доказать потребность в инвестициях
Руководитель розничной сети	Отвечает за офисы продаж	Защитить бюджет отделений, понять вклад офлайн-канала в кросс-продажи
Руководитель партнёрских программ	Работает с внешними партнёрами	Оценить вклад партнёрских площадок, выявить сильных/слабых партнёров

Маркетолог (по каналам продвижения)	Маркетинг	Понять, куда эффективнее вкладывать бюджет (онлайн vs офлайн vs партнёры)
Владелец данных "Клиент"	DWH / Мастер-данных	Обеспечить корректность идентификации клиентов, согласовать логику мэтчинг
Владелец данных "Продукт"	DWH / Продуктовая команда	Подтвердить состав и статусы продуктов, участвующих в анализе
Аналитик DWH	Команда DWH	Проектирование витрины и ЕТ
ETL-разработчик/инженер данных	Команда DWH	Получить чёткие требования к источникам, маппингу полей и логике загрузки
BI-аналитик / разработчик дашбордов	BI-команда	Получить готовую витрину для построения отчёта или дашборда
Владелец источника CRM_BANK	CRM-администратор	Подтвердить доступность данных и корректность поля channel_cd
Владелец источника IB_SYSTEM	Интернет-банк / Digital	Подтвердить фиксацию каналов в онлайн-среде
Представитель партнёрской интеграции	Интеграции с партнёрами	Обеспечить стабильность API/файловых выгрузок, согласовать формат данные
Менеджер по продуктам (кредиты, карты)	Продуктовая команда	Увидеть, через какие каналы чаще подключают их продукт

3. Матрица классификации заинтересованных лиц

	рук. аналитики	аналитик DWH	рук. цифровых каналов	рук. розницы	рук. партнёрских программ	маркетолог	владелец «клиент»	владелец «продукт»	ETL-разра-ботчик	BI-аналитик	CRM владелец	владелец IB	представитель партнёров	менеджер по продуктам
1. Формулировка цели и сбор требований	A	R	C	C	C	C								C
2. Проектирование витрины данных	C	R					C	C	C	C				
3. Согласование логики мэтчинга клиентов		R					A		C				C	
4. Согласование состава и статусов продуктов		R						A					C	A
5. Определение атрибута "канал подключения"	C	R	C	C	C				C		C	C	C	
6. Реализация ETL-процесса		C			C		C	C	R		C	C	C	
7. Разработка дашборда в BI		C	C	C	C	C				R				C
8. Согласование бизнес-логики (момент подключения, 2+ продукта)	A	R	C	C	C	C	C	C	C					C
9. Тестирование и валидация отчёта	A	R	C	C	C	C	C	C	C	C				C
10. Презентация результатов	A	R												
11. Поддержка и обновление отчёта	A	R			C				R	R			C	

5. Вопросы для интервью

Интервью с Руководителем отдела аналитики

1. Какова основная цель этого отчёта? Что вы будете делать с этими данными?
2. Есть ли у вас примеры похожих отчётов, которые можно взять за основу?
3. Какие метрики вы считаете наиболее важными (общее число подключений, среднее на клиента и т.д.)?
4. Нужна ли динамика по месяцам или достаточно среза за год?
5. Кто ещё из бизнеса будет использовать этот отчёт?

6. Какой уровень детализации вы ожидаете: по каналам, по продуктам, по регионам?
7. Готовы ли вы утвердить бизнес-логику (например, что считается подключением) до начала реализации?
8. После каких событий (акции, запуск продукта) растёт активность?
9. Планируются ли масштабирование каналов (например, больше партнёров)? Ожидается ли рост числа пользователей отчёта?

Интервью с ETL-разработчиком / инженером данных

1. Как организована загрузка данных из **IB_SYSTEM**, **CRM_BANK** и партнёрских систем в DWH?
2. Есть ли единый шлюз интеграции, или каждая система подключена отдельно?
3. Как часто обновляются данные из этих источников? Есть ли задержки?
4. Как передаются данные от партнёров: через API, файлы, SFTP? В каком формате?
5. Есть ли уже витрина или таблица с фактами подключений? Если да — как она называется?
6. Как реализован мэтчинг клиентов между источниками (особенно с партнёрами)?
7. Нужно ли нам интегрироваться со всеми партнёрскими системами, или есть единая платформа-агрегатор?
8. Какой объём данных по подключениям приходит ежемесячно из **IB_SYSTEM**, **CRM_BANK**, партнёрских систем? Какой общий прирост в DWH по этой теме?
9. Есть ли пики (например, в начале месяца, после маркетинговой кампании)? Какие нагрузки создают аналогичные отчёты на BI-сервере?

Интервью с Владелецem данных "Клиент"

1. Какой атрибут используется как уникальный идентификатор клиента (**client_id**) в DWH?
2. Как происходит сопоставление клиента из **IB_SYSTEM**, **CRM_BANK** и партнёрских систем?

3. Что делать, если в данных от партнёра нет `client_id`, только паспортные данные?
4. Есть ли справочник каналов подключения? Какие значения разрешены (например, `IB`, `BRANCH`, `PARTNER`)?
5. Как часто обновляются мастер-данные клиентов?
6. Кто отвечает за качество данных о клиентах в источниках?

Интервью с Владелецem источника `CRM_BANK`

1. В каком поле фиксируется канал подключения в `CRM_BANK`?
2. Какие значения принимает это поле? Есть ли справочник?
3. Как определяется момент подключения продукта (дата заявки, дата активации)?
4. Передаётся ли `client_id` в DWH? Или используется другой идентификатор?
5. Есть ли в системе признак, что клиент подключил несколько продуктов в рамках одного визита?
6. Есть ли ограничения на выгрузку данных (объём, частота, доступ)?

Интервью с Представителем партнёрской интеграции

1. Какие партнёры передают данные о подключениях? Есть ли единый список?
2. В каком формате передаются данные: API, файлы, SFTP?
3. Какие поля обязательных в выгрузке от партнёра (особенно: `client_id`, `product_id`, `channel`, `activation_dttm`)?
4. Есть ли единый шаблон или соглашение по формату данных от партнёров?
5. Как часто поступают выгрузки: ежедневно, раз в неделю?
6. Что происходит, если у партнёра нет `client_id`? Как мы можем идентифицировать клиента?

Интервью с BI-аналитиком / разработчиком дашбордов

1. Есть ли уже шаблоны дашбордов по каналам продаж? Можно ли их адаптировать?
2. Нужна ли возможность фильтрации по продуктам, регионам, периоду?
3. Как часто должен обновляться дашборд: ежедневно, еженедельно?
4. Кто и как будет получать доступ к отчёту: через веб, email, мобильное приложение?

5. Анкета сбора требований для отчёта

Какой основной показатель эффективности канала вы считаете наиболее релевантным? *(Выберите один ответ)*

- ☐ Общее количество подключений
- ☐ Среднее число продуктов на одного клиента
- ☐ Доля клиентов, подключивших 2+ продукта
- ☐ Рост подключений по месяцам
- ☐ Доля канала в общем объёме подключений
- ☐ Другое: _____

Нужна ли детализация по типам продуктов? *(Выберите один ответ)*

- ☐ Нет, достаточно общего числа подключений
- ☐ Да, по основным типам: Кредиты, Карты, Вклады, Страхование
- ☐ Да, с разбивкой по конкретным продуктам (например, "Премиум-карта", "Ипотека 2024")
- ☐ Не знаю / не важно

Какие сегменты клиентов важно анализировать отдельно? *(Выберите все подходящие ответы)*

- ☐ По возрасту (например, 18–30, 31–50, 50+)
- ☐ По региону
- ☐ По уровню дохода

☐ По частоте использования (активные / неактивные)

☐ Нужен только общий анализ

☐ Другое: _____

Как часто вы планируете использовать этот отчёт? *(Выберите один ответ)*

☐ Ежедневное

☐ Еженедельно

☐ Ежемесячно

☐ Ежеквартально

☐ По запросу

☐ Для разового анализа

☐ Не знаю

Должны ли в анализ включаться отменённые или расторгнутые продукты?
(Выберите один ответ)

☐ Нет, только активные на текущий момент

☐ Да, но с пометкой "отменён"

☐ Да, но анализироваться отдельно

☐ Не знаю / не важно

Какие данные вы используете для анализа каналов подключения? *(Выберите все подходящие ответы)*

☐ Данные из интернет-банка (IB_SYSTEM)

☐ Данные из CRM-системы (CRM_BANK)

☐ Данные от партнёрских площадок (API / файлы)

☐ Данные из колл-центра

☐ Данные из мобильного приложения

☐ Другое: _____

Как часто должны обновляться данные в отчёте? *(Выберите один ответ)*

- ☐ Ежедневно
- ☐ Еженедельно
- ☐ Ежемесячно
- ☐ По запросу
- ☐ В реальном времени (онлайн)

В какое время суток вы используете на отчёт / дашборд? *(Выберите один ответ)*

- ☐ Утро (8:00–10:00)
- ☐ День (10:00–15:00)
- ☐ Вечер (15:00–18:00)
- ☐ Нагрузка равномерная
- ☐ Не знаю

Какой уровень доступа Вам нужен? *(Выберите один ответ)*

- ☐ Просмотр
- ☐ Фильтрация
- ☐ Редактирование
- ☐ Экспорт

Нужна ли возможность экспорта данных из отчёта? *(Выберите один ответ)*

- ☐ Да, в Excel/CSV
- ☐ Да, в PDF
- ☐ Нет, достаточно визуализации
- ☐ Не знаю

Есть ли у вас требования к времени отклика отчёта? *(Выберите один ответ)*

- ☐ Мгновенно (до 3 сек)
- ☐ До 10 секунд
- ☐ До 30 секунд

☐ До 1 минуты

☐ Не критично

6. Функциональные требования

1. Отчёт должен отображать среднее количество подключённых продуктов на одного клиента по трём каналам: личный кабинет, офис продаж, площадки партнёров.
2. Отчёт должен показывать долю каждого канала в общем числе подключений за 2024 год.
3. Должна быть возможность сравнения каналов по количеству клиентов, подключивших 2+ продукта.
4. Должна быть детализация по типам продуктов (карты, кредиты, вклады, страхование).
5. Отчёт должен поддерживать фильтрацию по периоду(месяц, квартал, год).
6. Должна быть визуализация: столбчатые диаграммы, круговая диаграмма (доля каналов), линейный график (динамика).
7. Должна быть возможность экспорта данных в Excel/CSV.
8. Витрина данных (dm_channel_activation) должна содержать поля: client_id, product_id, channel_cd, activation_dttm, product_type.
9. Должна быть реализована агрегированная таблица с метриками: channel, client_count, total_products, avg_products_per_client, share_of_channel.
10. Отчёт должен быть построен на данных из DWH, загруженных из источников: IB_SYSTEM, CRM_BANK, PARTNER_API_*, PARTNER_FILE_*.

7. Нефункциональные требования

1. Частота обновления данных — ежедневно, с задержкой не более 24 часов.
2. Время загрузки отчёта — не более 10 секунд при стандартной нагрузке.
3. Пиковая нагрузка — *система должна поддерживать до 50 одновременных пользователей с временем отклика отчёта не более 10 секунд в пиковой нагрузке*.

4. Объём хранимых данных— витрина должна поддерживать хранение данных за 3 года (ожидаемый объём: до 5 ГБ).

5. Формат доступа — отчёт доступен через BI-платформу (Power BI) по защищённому внутреннему portalу.

6. Уровни доступа:

- Руководители и аналитики — просмотр, фильтрация, экспорт
- ETL/BI-разработчики — полный доступ для настройки
- Остальные — только просмотр (по необходимости)

7. Резервное копирование — данные витрины резервируются ежедневно.

8. Аудит доступа — фиксация входа пользователей в BI-систему.

9. Источники данных:

- Данные из **IB_SYSTEM** и **CRM_BANK** загружаются через ETL-процессы
- Данные от партнёров — через API или SFTP в формате JSON/CSV
- Все данные должны быть доступны в DWH до 06:00 следующего дня

10. Документирование:

- Должна быть документация по витрине: маппинг полей, бизнес-логика, источники
- Должен быть описан ETL-процесс загрузки