

ДЕПАРТАМЕНТ ОБРАЗОВАНИЯ И НАУКИ ГОРОДА МОСКВЫ

Государственное автономное образовательное учреждение

высшего образования города Москвы

«Московский городской педагогический университет»

(ГАОУ ВО МГПУ)

Институт цифрового образования

Департамент информатики, управления и технологий

Практическая(лабораторная) работа № 3.2

по дисциплине «Платформы Data Engineering»

Выполнил:

студент группы БД-251м

Направление подготовки/Специальность

38.04.05 - Бизнес-информатика

Бобылева Варвара Владимировна

(Ф.И.О.)

Проверил:

Кандидат технических наук

(ученая степень, звание)

Босенко Тимур Муртазович

(Ф.И.О.)

Москва 2025

---

-- Автор: Бобылева Варвара  
-- Группа: БД-251м  
-- Проект: Интеграция данных из разнородных источников

---

## Постановка цели

Основная цель дашборда — дать обзор текущей инфраструктуры компании, проанализировать проблемы и возможности, связанные с интеграцией данных, а также оценить удовлетворенность сотрудников текущими инструментами.

## Тема

Мы хотим обсудить наш текущий опыт в интеграции данных из различных источников. Речь идет о том, как мы собираем, обрабатываем и используем информацию, поступающую из разных баз, сервисов и файлов.

## Целевая аудитория

Опрос предназначен для всех коллег, которые работают с данными: от аналитиков и разработчиков до менеджеров, принимающих решения на основе этой информации. Ваш опыт и мнение очень важны для нас.

## Процесс разработки

Процесс подготовки и создания дашборда в DataLens, использующего данные из Google Forms, можно разделить на два больших этапа: подготовка данных и разработка в DataLens.

### Этап 1: Подготовка данных (Google Forms и Google Sheets)

#### 1. Создание опроса в Google Forms.

- Для выбранной темы был сформулирован перечень вопросов и вариантов ответа к ним для опроса.

- Опрос был создан в Google Forms.

Пример полученного опроса представлен на рисунке 1.

### Интеграция данных из разнородных источников

Анализ процесса объединения информации из множества систем источников

\*Обязательный вопрос

1. Из каких систем вы собираете данные для анализа? \*

Отметьте все подходящие варианты.

- ☐ CRM
- ☐ ERP
- ☐ SQL-базы
- ☐ Облачные сервисы
- ☐ Excel
- ☐ Другое: \_\_\_\_\_

2. Насколько критична для ваших бизнес-процессов синхронизация данных в реальном времени по сравнению с периодическим обновлением?

Отметьте только один овал.

- ☐ Критично
- ☐ Важно
- ☐ Не очень важно
- ☐ Не важно

Рисунок 1 – Пример опроса

## 2. Настройка и очистка данных опроса.

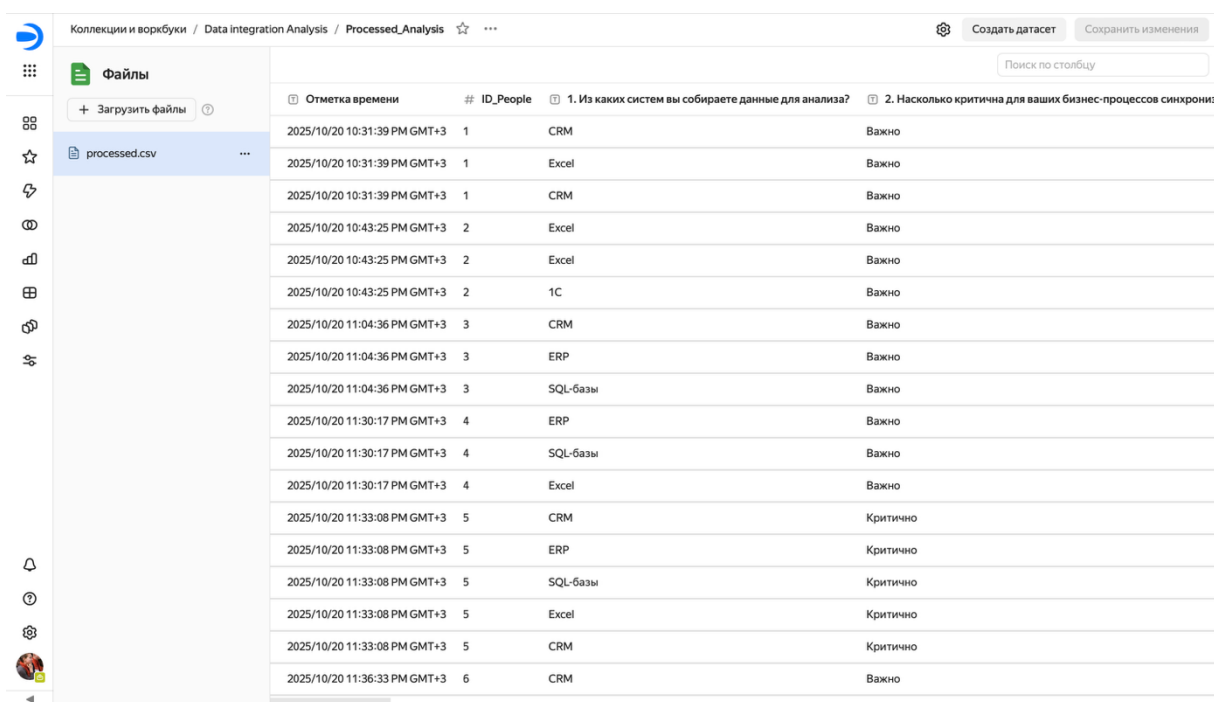
- По завершению опроса итоговые ответы были выгружены в файл формата .csv
- После была произведена проверка и при необходимости корректировка заголовков столбцов.
- Также были проверены и очищены пустые строки.
- Данные были приведены к единому формату.

## Шаг 2: Разработка дашборда в DataLens

### 1. Создание подключения в DataLens:

В интерфейсе Yandex DataLens создается новое подключение к выбранному источнику данных, указываются необходимые параметры доступа.

Для проекта было создано подключение с загрузкой данных из полученного и обработанного файлов-источников в шаге 1. Результат создания подключения и загрузки данных представлен на рисунке 2.



Отметка времени	# ID_People	1. Из каких систем вы собираете данные для анализа?	2. Насколько критична для ваших бизнес-процессов синхрониз
2025/10/20 10:31:39 PM GMT+3	1	CRM	Важно
2025/10/20 10:31:39 PM GMT+3	1	Excel	Важно
2025/10/20 10:31:39 PM GMT+3	1	CRM	Важно
2025/10/20 10:43:25 PM GMT+3	2	Excel	Важно
2025/10/20 10:43:25 PM GMT+3	2	Excel	Важно
2025/10/20 10:43:25 PM GMT+3	2	1C	Важно
2025/10/20 11:04:36 PM GMT+3	3	CRM	Важно
2025/10/20 11:04:36 PM GMT+3	3	ERP	Важно
2025/10/20 11:04:36 PM GMT+3	3	SQL-базы	Важно
2025/10/20 11:30:17 PM GMT+3	4	ERP	Важно
2025/10/20 11:30:17 PM GMT+3	4	SQL-базы	Важно
2025/10/20 11:30:17 PM GMT+3	4	Excel	Важно
2025/10/20 11:33:08 PM GMT+3	5	CRM	Критично
2025/10/20 11:33:08 PM GMT+3	5	ERP	Критично
2025/10/20 11:33:08 PM GMT+3	5	SQL-базы	Критично
2025/10/20 11:33:08 PM GMT+3	5	Excel	Критично
2025/10/20 11:33:08 PM GMT+3	5	CRM	Критично
2025/10/20 11:36:33 PM GMT+3	6	CRM	Важно

Рисунок 2 – Создание подключения и загрузка исходных данных по опросу

## 2. Формирование датасета:

Датасет — это центральный элемент, который связывает источник данных с чартами. Он позволяет структурировать данные, задать типы полей и создать вычисляемые поля, необходимые для расчётов.

Датасет «Processed\_analisys» основан на данных из таблицы «processed» по ответам опрашиваемой аудитории.

Результат создания датасета и преобразования полей представлен на рисунках 3 и 4.

#	Уникальный пользователь	Отметка времени	ID_People	Источник данных	Критичность синхронизации	Инструменты и технологии	Удовлетворение технологиями и инструментами
1	1	2025/10/21 9:32:19 AM GMT+3	9	ERP	Важно	Интеграционная шина (EBS)	Полностью удовлетворяют
2	1	2025/10/25 8:40:07 AM GMT+3	28	1C	Не важно	Ручное выполнение	Скорее не удовлетворяют
3	1	2025/10/26 11:04:36 PM GMT+3	33	SQL-базы	Важно	API-интеграция	Скорее удовлетворяют
4	1	2025/10/25 8:39:42 AM GMT+3	27	Excel	Не важно	Интеграционная шина (EBS)	Скорее удовлетворяют
5	1	2025/10/21 9:32:19 AM GMT+3	9	ERP	Важно	Файловый обмен	Полностью удовлетворяют
6	1	2025/10/21 11:20:02 AM GMT+3	20	CRM	Не очень важно	Интеграционная шина (EBS)	Скорее не удовлетворяют
7	1	2025/10/25 8:37:50 AM GMT+3	24	CRM	Не важно	ETL	Ограничивают возможности
8	1	2025/10/22 12:33:33 PM GMT+3	22	SQL-базы	Важно	Файловый обмен	Скорее удовлетворяют
9	1	2025/10/25 8:38:21 AM GMT+3	25	1C	Критично	Микросервисы	Скорее не удовлетворяют
10	1	2025/10/26 8:40:07 AM GMT+3	39	Облачные сервисы	Не важно	Микросервисы	Скорее не удовлетворяют

Рисунок 3 – Результат создания датасета, выявление источника данных и предпросмотр данных

Имя	Источник поля	Тип	Агрегация
1	Имя	Целое число	Авто
2	Отметка времени	Строка	Нет
3	ID_People	Целое число	Нет
4	Источник данных	Строка	Нет
5	Критичность синхронизации	Строка	Нет
6	Инструменты и технологии	Строка	Нет
7	Удовлетворение технологиями и инструментами	Строка	Нет
8	Проблемы масштабирования	Строка	Нет
9	Проблемы качества данных	Строка	Нет
10	Правила управления качеством данных	Строка	Нет
11	Отслеживание конфликтов данных	Строка	Нет
12	Методы разрешения конфликтов данных	Строка	Нет
13	Обеспечение безопасности данных	Строка	Нет
14	Политики доступа к данным	Строка	Нет
15	Проблемы с разрозненными данными	Строка	Нет
16	Варианты реализации при едином хранилище	Строка	Нет

Рисунок 4 – Настройка полей датасета

## Шаг 2: Создание вычисляемых полей

В связи с тем, что опрос включал в себя множественный выбор вариантов ответа, данные для анализа имеют дублирующие значения. Для их исключения была создана метрика подсчета уникальности пользователя, чтобы исключить повторения ответов по одному человеку.

Формула расчета – `COUNTD([ID_People])`

## Шаг 3: Выбор чартов и их построение

Для визуализации метрик будут использованы следующие типы чартов:

1. Показатели удовлетворенности и проблем (Круговая диаграмма).

Что показывает: Фокусировку на оценку удовлетворенности сотрудников текущими инструментами и технологиями для работы с данными.

Обоснование: Визуализация помогает быстро понять, в каких сегментах сконцентрированы проблемные и слабые места, требующие особого внимания.

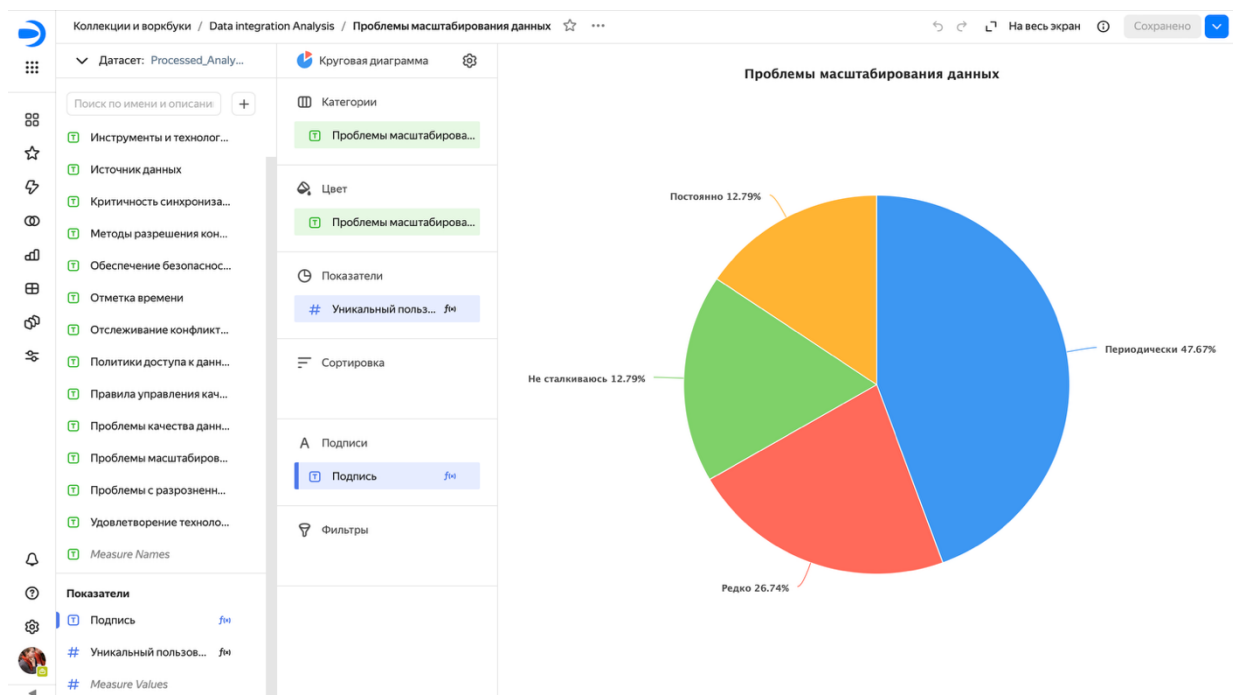


Рисунок 5 – Пример построение чарта с типом круговая диаграмма

## 2. Инфраструктура компании (Гистограмма):

○ Что показывает: Обзор текущей инфраструктуры данных и инструментов, используемых в компании.

○ Обоснование: Позволяет понять, какие источники данных наиболее распространены и какие технологии применяются для их интеграции.

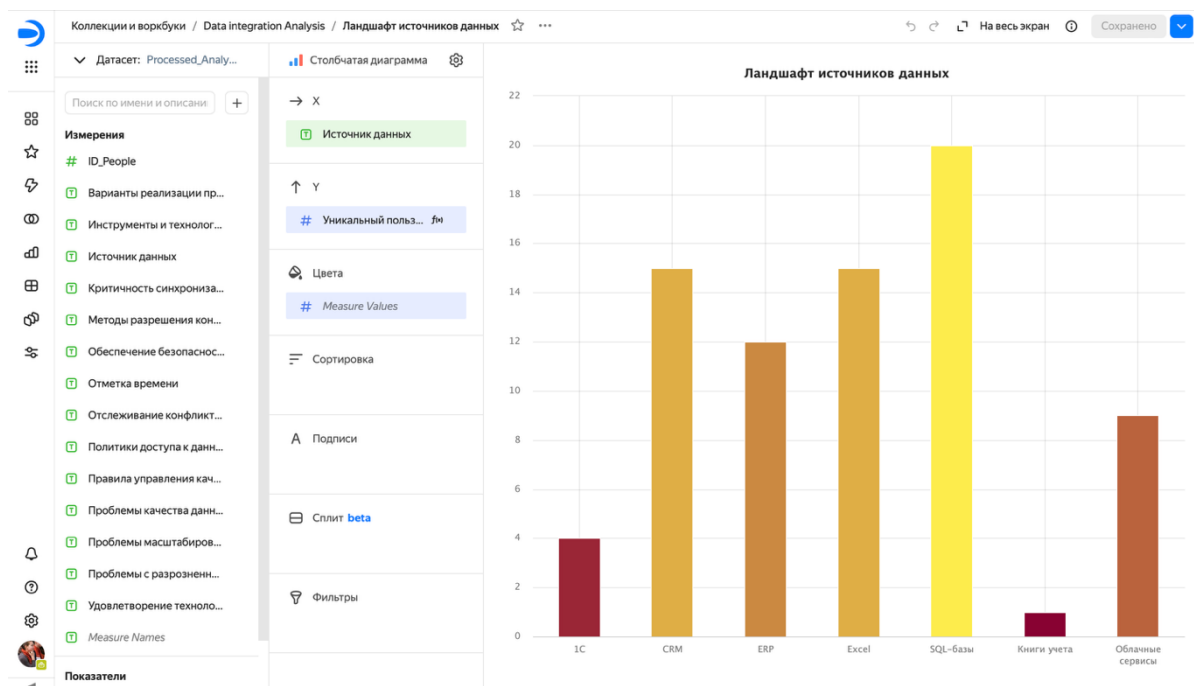


Рисунок 6 – Пример построение чарта гистограммы

## Шаг 4: Сборка и настройка дашборда

1. Создание нового дашборда: В интерфейсе Yandex DataLens создается новый дашборд.
2. Добавление чартов: Поочерёдно добавляются созданные чарты, задаётся их расположение на холсте.
3. Настройка виджетов с текстом:
  - Наполнение: Добавление заголовков, описаний и пояснений к чартам.
  - Обоснование: Текстовые блоки делают дашборд более понятным для конечного пользователя, объясняя, что показывает тот или иной чарт.

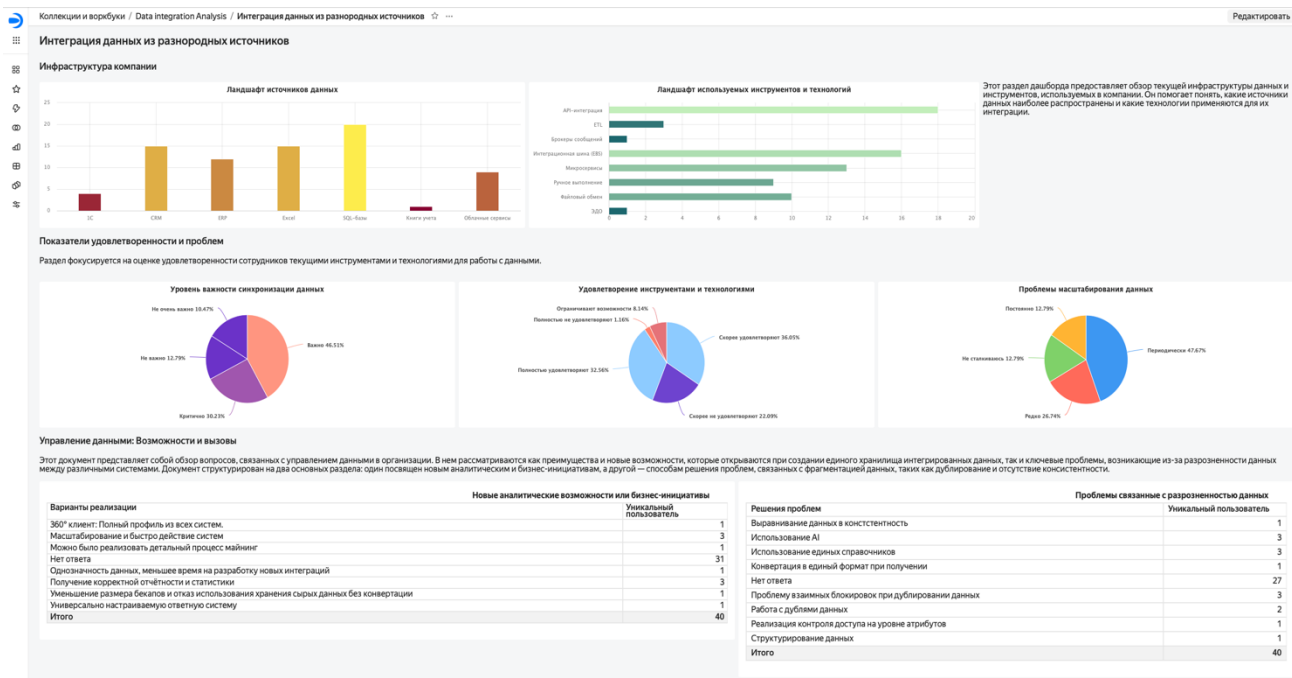


Рисунок 7 – Итоговый дашборд по проекту

## Шаг 5: Тестирование и обратная связь

1. Проверка корректности расчётов: Убедиться, что вычисляемые поля и отображаемые данные соответствуют действительности.
2. Демонстрация дашборда: Показать дашборд заинтересованным лицам для сбора обратной связи.
3. Доработка: На основе обратной связи внести корректировки в дизайн, чарты или вычисляемые поля.
4. Публикация: Предоставить доступ к готовому дашборду для регулярного мониторинга.

## Выводы

Основываясь на проделанной работе по созданию дашборда в DataLens с использованием данных из Google Forms и Google Sheets, можно выделить следующие детальные выводы:

### Визуализация данных как мощный инструмент коммуникации и анализа

- **Трансформация данных в инсайты:** Сами по себе данные, представленные в файле формата .csv, являются массивом информации, который сложно анализировать без специальной обработки. Визуализация в DataLens позволила превратить эти данные в

наглядные и легко интерпретируемые графики, диаграммы и таблицы. Это сократило время, необходимое для поиска закономерностей и выводов, поскольку коллеги могут моментально оценить результаты опроса.

- **Повышение доступности информации:** С помощью дашборда мы сделали результаты анализа доступными для всех заинтересованных лиц в удобной и интерактивной форме. Теперь не нужно разбираться в сложных таблицах или составлять отчёты вручную — вся ключевая информация всегда под рукой, обновляется автоматически и доступна по ссылке.

#### **Качество исходных данных определяет качество результата**

- **Критическая важность предобработки:** Создание дашборда показало, что качество конечной визуализации и достоверность выводов напрямую зависят от качества исходных данных. На этапе работы с файлом формата .csv была проведена критически важная работа по очистке и форматированию. Некорректные типы данных или несогласованные ответы могли бы привести к ошибочным результатам, делая дашборд бесполезным.

- **Опыт в подготовке данных:** Этот опыт подчёркивает, что инвестиции времени в структурирование и проверку данных на начальном этапе окупаются. Навыки по очистке, нормализации и приведению данных к единому формату являются ключевыми для любого BI-проекта.

#### **Интерактивные дашборды расширяют возможности для самостоятельного анализа**

- **Децентрализация аналитики:** Создание интерактивных элементов, таких как фильтры и селекторы, на дашборде позволило коллегам самостоятельно проводить более глубокий анализ. Например, они могут отфильтровать ответы по конкретному отделу, дате или другому параметру, не обращая за помощью к аналитику. Это расширяет доступ к данным и способствует развитию культуры принятия решений, основанных на данных, во всей команде.

- **Фокус на конкретных потребностях:** Каждый пользователь может настроить дашборд под свои нужды, получая ответы на специфические вопросы, что повышает ценность инструмента для разных подразделений и ролей в компании.

#### **Полученный опыт — основа для дальнейших BI-проектов**

- **Демонстрация возможностей:** Проект по созданию дашборда на основе Google Forms и DataLens стал отличной демонстрацией возможностей BI-инструментов в целом. Он показал, как можно быстро и эффективно создавать аналитические продукты.

- **База для масштабирования:** Освоенные шаги и полученный опыт могут быть легко масштабированы для решения более сложных задач, с более крупными и разнородными источниками данных. Команда теперь обладает необходимой компетенцией для интеграции данных из других систем, таких как CRM, ERP или внутренних баз данных, что открывает новые перспективы для аналитики в компании.

## **Заключение**

Проделанная работа представляет собой полный цикл аналитического проекта — от постановки целей до детальной разработки и анализа инструмента. Основные выводы и полученный опыт можно сформулировать следующим образом:

1. **Умение работать с данными и структурировать их.** Появилось понимание важности этапа подготовки данных и умение применять базовые методы очистки и форматирования их для последующего анализа.

2. **Навыки работы с BI-инструментами:** На примере Yandex DataLens был детально разобран процесс создания дашборда — от подготовки данных и создания вычисляемых полей до выбора оптимальных типов визуализации. Особое внимание было уделено обоснованию каждого шага, что является ключевым навыком в аналитике.

3. **Понимание ценности автоматизации.** Стало очевидно, что даже небольшие проекты по автоматизации сбора и анализа данных могут принести значительную экономию времени и повысить общую продуктивность.

4. **Оценка возможностей облачных решений.** Проект продемонстрировал, как облачные сервисы (Google Forms, Google Sheets) и BI-инструменты (DataLens) могут бесшовно взаимодействовать, создавая мощные и доступные решения для бизнеса.

5. **Систематизация знаний:** Работа позволила систематизировать знания о том, как превратить необработанные данные в наглядные и полезные для бизнеса инсайты. Был закреплён опыт последовательного перехода от общей концепции и стратегии к конкретным шагам по реализации.

## **Ссылки**

Итоговый дашборд по проекту доступен по ссылке – <https://datalens.yandex/itkerzarz2fi3>

Ссылка на GitHub проекта - <https://github.com/VarvaraBobyleva/DEP-MGPU/tree/main/Module6>