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

Варданян Роберт Барсегович

(Ф.И.О.)

Проверил:

Доцент департамента информатики, управления и технологий, доктор экономических наук

(ученая степень, звание)

Босенко Т.М.

(Ф.И.О.)

Москва 2025

Оглавление

[1. Введение 3](#_Toc212419828)

[1.1. Цель исследования 3](#_Toc212419829)

[1.2. Методология 3](#_Toc212419830)

[1.3. Используемые технологии 3](#_Toc212419831)

[2. Проектирование и проведение опроса 3](#_Toc212419832)

[2.1. Структура опроса 3](#_Toc212419833)

[2.2. Подготовка данных 4](#_Toc212419834)

[3. Обогащение данных в Yandex DataLens 4](#_Toc212419835)

[3.1. Созданные вычисляемые поля 4](#_Toc212419836)

[4. Визуализация и аналитика 5](#_Toc212419837)

[4.1. Созданные чарты 5](#_Toc212419838)

[4.2. Интерактивный дашборд 5](#_Toc212419839)

[5. Ключевые выводы и инсайты 6](#_Toc212419840)

[5.1. Основные выводы 6](#_Toc212419841)

[5.2. Рекомендации 7](#_Toc212419842)

[6. Заключение 8](#_Toc212419843)

[7. Ссылки 8](#_Toc212419844)

# **1. Введение**

## **1.1. Цель исследования**

Провести комплексный анализ текущего состояния безопасности данных в области Data Engineering, выявить ключевые проблемы, инструменты и практики, используемые специалистами индустрии.

## **1.2. Методология**

* **Выборка:** 10 специалистов из различных отраслей
* **Метод сбора данных:** Online-опрос
* **Целевая аудитория:** Data Engineers, Security Engineers, DevOps Engineers, Data Scientists, Team Leads

## **1.3. Используемые технологии**

* **Сбор данных:** Google Forms/Typeform
* **Хранение данных:** CSV-файл
* **Аналитика:** Yandex DataLens
* **Визуализация:** Интерактивный дашборд

# **2. Проектирование и проведение опроса**

## **2.1. Структура опроса**

Опрос состоял из 16 вопросов, сгруппированных в 5 тематических блоков:

**Блок A: Демография и контекст** (4 вопроса)

* Роль, опыт работы, размер компании, отрасль

**Блок B: Текущее состояние безопасности** (3 вопроса)

* Общий уровень безопасности, типы данных, compliance требования

**Блок C: Инструменты и практики** (3 вопроса)

* Используемые инструменты, управление доступом, частота аудитов

**Блок D: Риски и инциденты** (3 вопроса)

* История инцидентов, основные риски, время обнаружения

**Блок E: Культура и будущее** (3 вопроса)

* Обучение безопасности, бюджет, планы по улучшению

## 

## **2.2. Подготовка данных**

* Сбор ответов от 10 респондентов
* Экспорт в CSV-формат
* Очистка и валидация данных

# **3. Обогащение данных в Yandex DataLens**

## **3.1. Созданные вычисляемые поля**

| Поле |  | Назначение |
| --- | --- | --- |
| **Общий индекс зрелости безопасности** |  | Количественная оценка уровня безопасности |
| **Наличие высокорисковых данных** |  | Флаг работы с чувствительными данными |
| **Оценка частоты аудитов** |  | Числовая оценка процессов аудита |
| **Уровень серьезности инцидентов** |  | Группировка по серьезности инцидентов |
| **Проактивная безопасность** |  | Флаг проактивного подхода к безопасности |

**3.2. Преобразования данных**

* Приведение категориальных переменных к числовым шкалам
* Создание бинарных флагов для анализа
* Нормализация данных для сравнения

# **4. Визуализация и аналитика**

## **4.1. Созданные чарты**

**Ключевые метрики (индикаторы):**

* Средняя зрелость безопасности: 3.4/5
* Работают с рисковыми данными: 80%
* Проактивная безопасность: 40%
* Высокие инциденты: 20%

**Основные визуализации:**

1. **Распределение по уровню безопасности** (круговая диаграмма)
2. **Зрелость безопасности по отраслям** (столбчатая диаграмма)
3. **Compliance требования по отраслям** (нормированная столбчатая)
4. **Аудиты vs Инциденты по опыту** (комбинированная диаграмма)
5. **Инструменты vs Зрелость** (точечная диаграмма)
6. **Детали по респондентам** (сводная таблица)

## **4.2. Интерактивный дашборд**

**Структура дашборда:**

* Верхняя часть: заголовок и 4 KPI индикатора
* Центральная часть: основные инсайты и распределения
* Нижняя часть: детальная аналитика и таблицы
* Боковая панель: фильтры по роли, отрасли, уровню безопасности

**Функциональность:**

* Интерактивные фильтры для сегментации данных
* Возможность детализации по клику на элементы чартов
* Связь между всеми визуализациями

# **5. Ключевые выводы и инсайты**

## **5.1. Основные выводы**

1. **Отраслевые различия**
   * FinTech и Healthcare демонстрируют самый высокий уровень безопасности (4.5/5)
   * Образовательные учреждения имеют наименьшие показатели (1.5/5)
   * 100% компаний в Healthcare соответствуют HIPAA требованиям
2. **Эффективность практик**
   * Компании с ежеквартальными аудитами в 3 раза реже сталкиваются с инцидентами
   * Проактивный подход к безопасности сокращает время обнаружения инцидентов до 1 часа
   * Использование RBAC систем повышает общий индекс зрелости на 1.2 пункта
3. **Инструментальный стек**
   * Шифрование данных (at-rest и in-transit) - наиболее распространенная практика (90%)
   * DLP-системы используются только в 30% компаний, преимущественно в FinTech
   * Количество инструментов положительно коррелирует с зрелостью безопасности
4. **Культура безопасности**
   * Только 40% компаний проводят регулярное обучение безопасности
   * 60% респондентов отметили увеличение бюджета на безопасность
   * Основные барьеры: нехватка экспертизы (50%) и бюджетные ограничения (30%)

## **5.2. Рекомендации**

1. **Для стартапов и малых компаний:**
   * Внедрить базовые практики шифрования и управления доступом
   * Начать с compliance требований, релевантных отрасли
   * Регулярно проводить security training для команды
2. **Для средних и крупных компаний:**
   * Внедрить проактивный мониторинг безопасности
   * Автоматизировать процессы compliance
   * Инвестировать в DLP и SIEM системы
3. **Общие рекомендации:**
   * Внедрить регулярные security аудиты (не реже чем раз в полгода)
   * Разработать инцидент-response планы
   * Создать культуру security-first в командах

# **6. Заключение**

Проведенное исследование демонстрирует значительные различия в подходах к безопасности данных между компаниями разных отраслей и размеров. Наблюдается прямая корреляция между зрелостью процессов безопасности и частотой инцидентов.

**Основные достижения проекта:**

* ✅ Разработан методологически корректный опрос
* ✅ Собраны и подготовлены данные
* ✅ Создана обогащенная витрина данных
* ✅ Разработан интерактивный аналитический дашборд
* ✅ Сформулированы практические рекомендации

# **7. Ссылки**

Ссылка на репозиторий GitHub: <https://github.com/vardanyan4ik/DEP-MGPU/tree/5d3bfbe4b7e86d953f911053c08528762929f4f0/Module03>

Ссылка на Дашборд Yandex DataLens: <https://datalens.ru/kva0tx3mwz4u5>