

Департамент образования и науки города Москвы
Государственное автономное образовательное учреждение
высшего образования города Москвы
«Московский городской педагогический университет»
Институт цифрового образования
Департамент информатики управления и технологий

Кузьмина Дарья Юрьевна БД-241м

Практическая работа 3.2 Исследование актуальных трендов в Data Engineering
Вариант 11

Направление подготовки/специальность
38.04.05 - Бизнес-информатика
Бизнес-аналитика и большие данные
(очная форма обучения)

Руководитель дисциплины:
Босенко Т.М., доцент департамента
информатики, управления и технологий,
доктор экономических наук

Москва
2024

Содержание

| | |
|----------------------|---|
| Введение | 2 |
| Основная часть | 3 |
| Основные выводы..... | 3 |
| Заключение..... | 3 |
| Приложения | 3 |

Введение

Цель:

Разработать, провести и проанализировать опрос по актуальным темам **Data Engineering**, применяя полученные знания на практике.

Задачи:

1. Подготовка опроса:

- Создайте опрос, используя Google Forms (<https://forms.google.com/>) или Survey Monkey (<https://www.surveymonkey.com/>).
- Тематика опроса должна быть связана с одной из предложенных тем (см. таблицу ниже).

2. Материалы для подготовки вопросов:

- [The Rise of the Data Engineer](#) от Maxime Beauchemin
- [Gartner Data Management Blog](#)
- [Fundamentals of Data Engineering \(O'Reilly\)](#)
- [Data Engineering Podcast](#)

3. Требования к опросу:

- 10-15 вопросов различных типов (множественный выбор, шкала, открытые вопросы).
- Минимум 5 респондентов (коллеги по работе, сокурсники или профессиональное сообщество).
- Вопросы должны быть сформулированы четко и профессионально.

4. Анализ результатов:

- Соберите ответы и проведите базовый анализ данных.
- Создайте минимум 3 визуализации полученных результатов (графики, диаграммы).
- Сформулируйте 3-5 основных выводов на основе полученных данных.

5. Документация и сохранение:

- Загрузите в Git-репозиторий:
 1. PDF-версию созданного опроса или скриншоты всех страниц.
 2. Экспортированные данные результатов (CSV/Excel).
 3. Jupyter Notebook или R Markdown с анализом и визуализацией результатов.

4. README.md с описанием проекта, процесса и основных выводов.

Основная часть

Тема опроса: 11Data Mesh: принципы и применение

Все необходимые файлы находятся в репозитории.

Основные выводы

1. Опыт респондентов:

Большинство респондентов имеют начальный или средний уровень опыта в Data Engineering. Это указывает на то, что тема Data Mesh пока что больше интересует начинающих и специалистов среднего уровня.

2. Знакомство с Data Mesh:

Уровень знакомства с концепцией Data Mesh варьируется, но большинство оценивают его на **3-4 балла из 5**. Это говорит о том, что концепция пока что не является широко известной, но уже вызывает интерес.

3. Важность аспектов:

Наиболее важным аспектом для внедрения Data Mesh считается **технологическая инфраструктура**. Также важными являются организационная культура, компетенции специалистов и финансовые ресурсы.

4. Ключевые принципы:

Наиболее важными принципами Data Mesh считаются **децентрализация владения данными и доменно-ориентированная архитектура**.

5. Проблематичные аспекты:

Наиболее проблематичным аспектом внедрения Data Mesh считаются **организационные барьеры**.

6. Используемые технологии:

Наиболее популярными технологиями для поддержки Data Mesh являются **Apache Kafka, Apache Airflow и Databricks**.

Заключение

Проведенный анализ показал, что концепция Data Mesh вызывает интерес у специалистов с разным уровнем опыта. Однако для успешного внедрения необходимо уделить внимание технологической инфраструктуре, организационным изменениям и обучению специалистов. Наиболее проблематичными аспектами являются организационные барьеры, что указывает на необходимость более глубокого изучения этой темы.

Приложения

1. **Исходные данные:** [Lab3.2 - Ответы на форму (1)-2.csv]
2. **Код анализа:** [lab3_2.ipynb]

3. **Графики:** Все графики сохранены в формате PDF.