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 № 3.1

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**

[Введение 4](#_Toc212397657)

[Задачи: 4](#_Toc212397658)

[Описание выбранного варианта 4](#_Toc212397659)

[Основные метрики: 4](#_Toc212397660)

[Рекомендуемые визуализации: 4](#_Toc212397661)

[Процесс разработки 5](#_Toc212397662)

[Заключение 9](#_Toc212397663)

[Сравнительный анализ Yandex DataLens и Tableau: 10](#_Toc212397664)

[Рекомендации по выбору инструмента: 10](#_Toc212397665)

# Введение

**Цель работы:** Разработать интерактивный аналитический дашборд для мониторинга и анализа производственной цепочки поставок с использованием Yandex DataLens.

## Задачи:

- Создать дашборд в Yandex DataLens на основе индивидуального варианта

- Реализовать ключевые метрики анализа производственных процессов

- Применить принципы эффективного дизайна дашбордов

- Обеспечить интерактивность для глубокого анализа данных

- Сравнить функциональность Yandex DataLens с Tableau

## Описание выбранного варианта

**Тема:** Анализ цепочки поставок

**Набор данных:** Производственные данные оборудования (manafcturingdataset.csv)

## Основные метрики:

- Надежность оборудования

- Уровень производственных рисков

- Качество продукции (уровень сервиса)

- Эффективность использования ресурсов

## Рекомендуемые визуализации:

- Матрица рисков (реализована через точечную диаграмму)

- Диаграмма Санкея движения товаров (адаптирована в круговую и линейчатую диаграммы)

- Карта поставщиков (заменена на рейтинг проблемных зон)

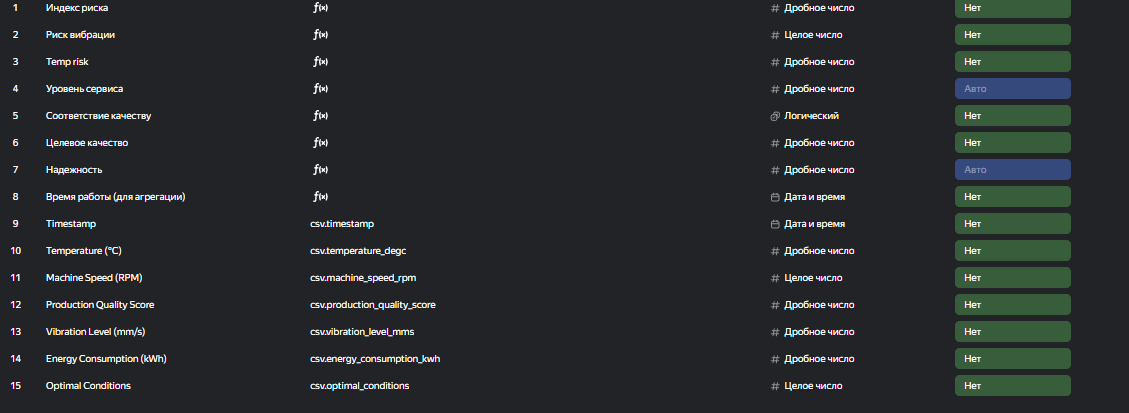
## Процесс разработки

**1. Подключение к источнику данных**

**В качестве источника данных использовался CSV-**файл (manafcturingdataset.csv), содержащий информацию о работе производственного оборудования с следующими полями:

**Таблица 1 –** Поля источника данных

|  |  |  |
| --- | --- | --- |
| **Поле** | **Описание** | **Тип данных** |
| **Timestamp** | **Временная метка измерения** | **Дата и время** |
| **Machine Speed** | **Скорость работы оборудования** | **Дробное число** |
| **Production Quality Score** | **Оценка качества продукции** | **Дробное число** |
| **Vibration Level** | **Уровень вибрации** | **Дробное число** |
| **Energy Consumption** | **Потребление энергии** | **Дробное число** |
| **Optimal Conditions** | **Флаг оптимальных условий** | **Целое число** |

****

**2. Создание и подготовка датасета**

Все поля были проверены и приведены к корректным типам данных. Далее были созданы следующие вычисляемые поля для бизнес-анализа:

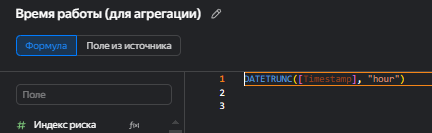
Ключевые вычисляемые поля:

**1. Рабочий час (для агрегации):**

**```sql**

**DATETRUNC([Timestamp], 'hour')**

**```**

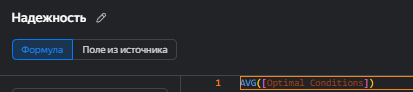
****

**2. Надежность оборудования:**

**```sql**

**AVG([Optimal Conditions])**

**```**

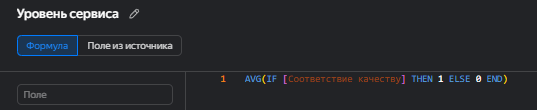
****

**3. Уровень сервиса (качество продукции):**

**```sql**

**AVG(IF [Production Quality Score] >= 8.5 THEN 1 ELSE 0 END)**

**```**

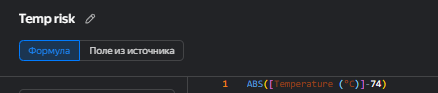
****

**4. Температурный риск:**

**```sql**

**ABS([Temperature] - 74)**

**```**

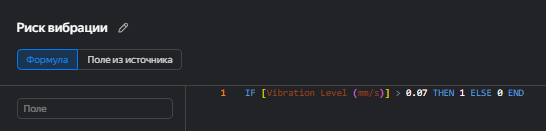
****

**5. Риск вибрации:**

**```sql**

**IF [Vibration Level] > 0.07 THEN 1 ELSE 0 END**

**```**

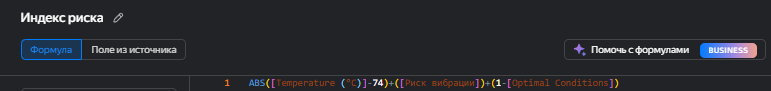
****

**6. Индекс риска:**

**```sql**

**[Температурный риск] + [Риск вибрации] + (1 - [Optimal Conditions])**

**```**

****

**7. Категория риска:**

**```sql**

**IF [Индекс риска] >= 2.5 THEN 'Критический'**

**ELSEIF [Индекс риска] >= 1.5 THEN 'Высокий'**

**ELSEIF [Индекс риска] >= 0.5 THEN 'Средний'**

**ELSE 'Низкий'**

**END**

**```**

**3. Разработка чартов**

**Индикаторы ключевых метрик:**

- Средний уровень надежности: AVG([Надежность])

- Средний уровень риска: AVG([Индекс риска])

- Средний уровень сервиса: AVG([Уровень сервиса])

**Матрица рисков (Точечная диаграмма):**

- Обоснование выбора: Классическая матрица рисков позволяет оценить корреляцию между температурными рисками и вибрацией оборудования

- Настройка:

- Ось X: Температурный риск

- Ось Y: Vibration Level

- Цвет: Категория риска

- Размер точек: Energy Consumption

- Подписи: Рабочий час

**Доля измерений по категориям риска (Круговая диаграмма):**

- Обоснование выбора: Наглядно показывает распределение рабочего времени по уровням производственного риска

- Настройка:

- Цвет: Категория риска

- Показатели: COUNT()

**Проблемные часы (Столбчатая диаграмма):**

- Обоснование выбора: Выявляет наиболее проблемные периоды работы для фокусировки улучшений

- Настройка:

- Ось X: Рабочий час

- Ось Y: COUNT\_IF([Категория риска] = 'Критический')

- Сортировка: по оси Y (по убыванию)

**Детальная статистика по часам (Сводная таблица):**

- Обоснование выбора: Предоставляет детализированную информацию по всем показателям в разрезе часов работы

- Настройка:

- Строки: FORMAT\_DATETIME([Рабочий час], 'HH:mm')

- Показатели: AVG([Production Quality Score]), AVG([Machine Speed]), COUNT\_IF([Категория риска] = 'Критический'), AVG([Energy Consumption])

**4. Сборка и настройка дашборда**

**Композиция дашборда:**

- Верхний ряд: Индикаторы KPI (самые важные метрики)

- Центральная часть: Матрица рисков и круговая диаграмма (ключевые визуализации)

- Нижняя часть: Детальная аналитика (таблицы и рейтинги)

**Настройка селекторов:**

- Период времени (Рабочий час)

- Категория риска

- Уровень качества продукции

**Интерактивность:**

- Настроено связывание всех чартов

- Реализована перекрестная фильтрация при клике на элементы

- Добавлены всплывающие подсказки и описания

# Заключение

В ходе работы был успешно разработан интерактивный дашборд для анализа производственной цепочки поставок в Yandex DataLens. **Основные достижения:**

- Реализованы ключевые метрики анализа производственных процессов через вычисляемые поля

- Создана комплексная система визуализации, включающая матрицу рисков, анализ распределения и детализацию по временным периодам

- Обеспечена высокая интерактивность через систему связанных селекторов и перекрестной фильтрации

- Применены принципы эффективного дизайна для создания понятного и информативного интерфейса

**Дашборд позволяет бизнес-пользователям:**

- Выявлять проблемные периоды работы через матрицу рисков

- Анализировать распределение рабочего времени по уровням риска

- Детализировать анализ по конкретным часам и категориям

- Принимать обоснованные решения по оптимизации производственных процессов

- Мониторить ключевые показатели эффективности в реальном времени

# Сравнительный анализ Yandex DataLens и Tableau:

|  |  |  |
| --- | --- | --- |
| **Критерий** | **Yandex DataLens** | **Tableau** |
| **Простота подключения** | **5** | **4** |
| **Подготовка данных** | **3** | **5** |
| **Язык вычисляемых полей** | **3** | **5** |
| **Визуализации** | **3** | **5** |
| **Интерактивность дашбордов** | **4** | **5** |
| **UI/UX и порог вхождения** | **5** | **3** |
| **Стоимость** | **5** | **3** |

# Рекомендации по выбору инструмента:

- Yandex DataLens предпочтительнее для быстрого старта, ограниченного бюджета и стандартных задач анализа

- Tableau рекомендуется для сложных аналитических задач, требующих глубокой кастомизации и мощных вычислительных возможностей (отдаю предпочтение ему т.к работал с ним, мб поэтому кажется удобнее)

Работа подтвердила, что Yandex DataLens является эффективным инструментом для оперативного создания дашбордов мониторинга производственных процессов.