

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ИТМО»

Отчет

по лабораторной работе №1

«Сбор, подготовка и анализ данных»

по дисциплине «**Креативные технологии в представлении данных**»

Автор: Малаев С.Г.

Факультет: ИКТ

Группа: К34422

Преподаватель: Валитова Ю. О.



Санкт-Петербург

2025

Цель работы.

Провести сбор, подготовку и анализ данных.

Задание на лабораторную работу.

1. Установите программное обеспечение, в котором будут выполняться работы, на свой компьютер. (Это задание выполняется в случае необходимости)
2. Запустите программу и освоите интерфейс. Изучите закладки и панели, посмотрите какие источники данных можно подключить.
3. Найдите несколько датасетов содержащих экономические данные, с которыми вы в дальнейшем будете работать. Эти данные могут содержать сведения о результатах работы конкретной организации, региона и т.п.
4. Подберите другие источники данных (ресурсы интернет, текстовые файлы и т.д.).
5. Опишите подобранные наборы данных. Какие данные там хранятся, в каком виде и т.п.
6. Загрузите наборы данных в выбранное программное средство. Если они нуждаются в преобразовании, то сделайте это.

Ход работы.

1. Установка программного обеспечения.

Для использования сервисов Yandex Datalens и Power BI не требуется осуществление локальной установки программного обеспечения. Данные решения предоставляются по модели облачных вычислений SaaS, что позволяет пользователю получить доступ к их функциональности посредством веб-интерфейса.

Ниже приведены ссылки на соответствующие ресурсы:

- [Yandex Datalens](#)
- [Power BI](#)

2. Исследование интерфейса.

Интерфейс Yandex DataLens устроен максимально интуитивно и понятно.

Боковое меню содержит вкладки:

- Подключения
- Датасеты
- Чарты
- Дашборды
- Избранное
- Личная папка

Панель управления помогает быстро находить нужные объекты, а система закладок обеспечивает сохранение выбранных представлений.

Сервис поддерживает широкий спектр подключений, включая файловые форматы и базы данных, таких как ClickHouse, PostgreSQL, а также специализированные сервисы, по типу Яндекс.Метрика.

Yandex DataLens

Тариф Business открывает дополнительные возможности
Статистика использования, стилизация UI, безопасное встраивание, SLA, приоритетная поддержка и другие преимущества.

[Попробовать](#) [Узнать больше](#)

Подключения
Подключайте свои источники данных
[Создать подключение](#)

Датасеты
Формируйте наборы данных с вычисляемыми полями и агрегациями
[Создать датасет](#)

Чарты
Визуализируйте данные в виде диаграмм и таблиц
[Создать чарт](#)

Дашборды
Создавайте страницы с наборами диаграмм, таблиц и фильтров
[Создать дашборд](#)

Примеры дашбордов

- Демонстрационный дашборд DataLens
- Статистика по COVID-19
- Исторические данные Яндекс.Погоды
- Аналитика по счетчику Яндекс.Метрики
- Геослой по спросу и предложению
- [Перейти в Marketplace →](#)

Рисунок 1 – Интерфейс Yandex Datalens сервиса

Интерфейс Power BI имеет более комплексным решения для бизнес-аналитики, который сочетает возможности создания интерактивных отчетов, дашбордов и панелей мониторинга.

Большой набор функций, рабочая область разделена на панели, такие как “Наборы данных”, “Отчеты” и “Дашборды”. Боковые панели позволяют переключаться между режимами создания визуализаций, работы с данными и настройками фильтров.

Интерфейс выглядит более продвинутым, но требует время для освоения.

Power BI также поддерживает подключение к множеству источников данных, от локальных файлов до корпоративных баз данных и облачных сервисов.

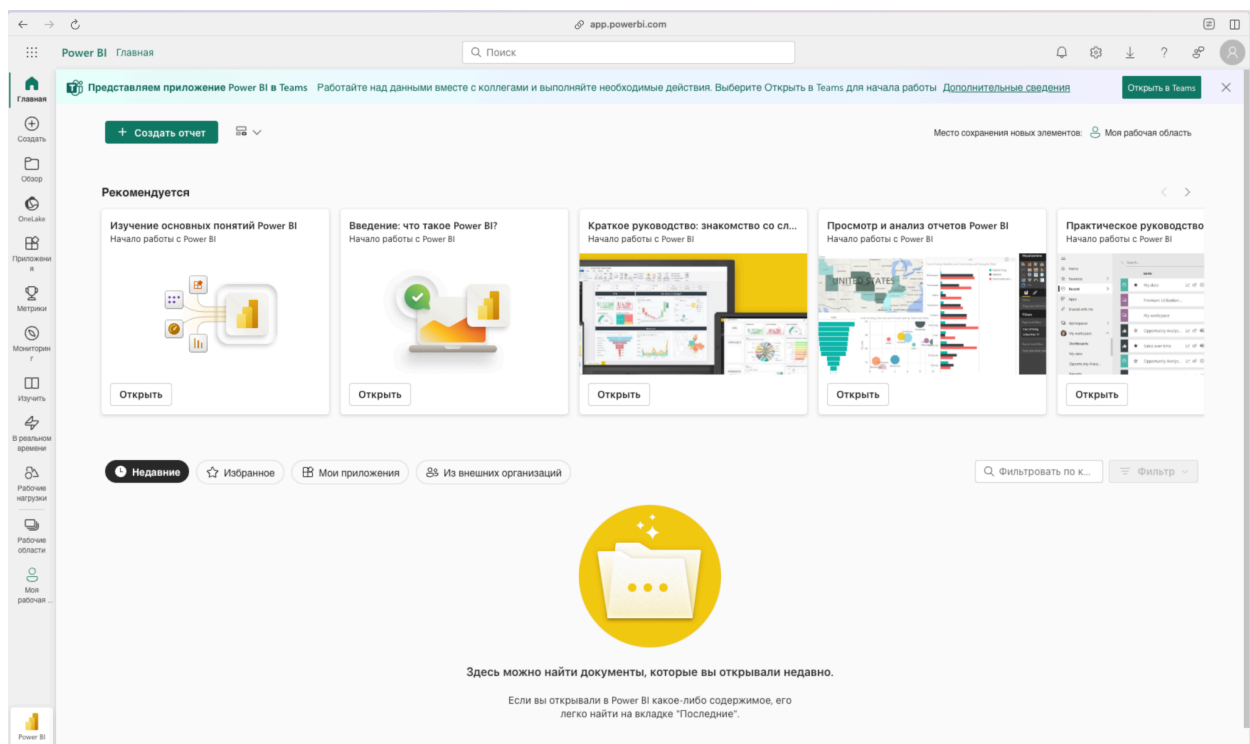


Рисунок 2 – Интерфейс Power BI сервиса

3. Поиск датасетов

Были выбраны два датасета: [Walmart Stocks Data 2025](#) и [Chocolate Sales Data](#) на платформе Kaggle

Walmart Stocks Data 2025 содержит исторические данные акций Walmart, собранные через yfinance с Yahoo Finance. Он включает ежедневные значения таких показателей, как цена открытия, максимум, минимум, закрытие и скорректированная цена, объём торгов, дивиденды и события сплитов. Датасет может использоваться для проведения разведывательного анализа EDA, расчета статистических характеристик, построения различных графиков и изучения временных рядов.

Chocolate Sales Data содержит данные о продажах шоколадных изделий. В нём представлены сведения о продукте, объемах продаж, выручке и сегментации покупателей. Датасет может быть использован для решения задач прогнозирования продаж, анализа тенденций и изучения покупательского поведения.

4. Подбор различных источников данных

Можно проверить возможность интеграции альтернативных источников данных. Например, можно подключиться к публичным API, таким как BBC API, для получения актуальных новостных в формате JSON или XML.

5. Описание подобранных наборов данных

Walmart Stocks Data 2025:

- Структура: Табличный CSV-файл, где каждая строка соответствует одному торговому дню.
- Данные:
 - Цена открытия
 - Максимальная цена
 - Минимальная цена
 - Цена закрытия
 - Скорректированная цена
 - Объем торгов
 - Дивиденды
 - Информация о сплитах
- Назначение: Используется для анализа динамики и волатильности акций Walmart, а также для расчёта статистических характеристик временных рядов.

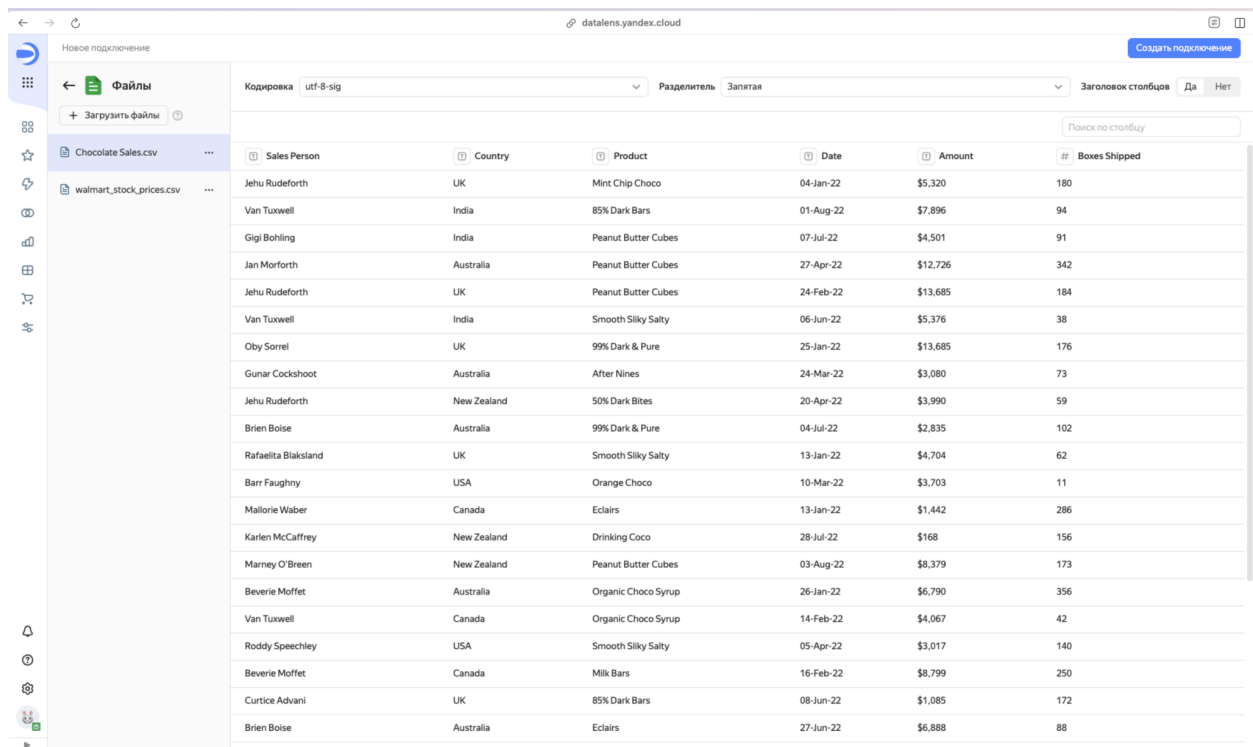
Chocolate Sales Data:

- Структура: Табличный CSV-файл, представляющий данные по продажам шоколада, где каждая запись отражает отдельную транзакцию.

- Данные:
 - Дата транзакции
 - Название продукта
 - Категория
 - Количество проданных единиц
 - Общая выручка
 - Сегментация клиентов
 - Местоположение продажи
- Назначение: Предназначен для анализа тенденций продаж, прогнозирования и исследования покупательского поведения в сегменте шоколадной продукции.

6. Загрузка данных в облачные сервисы

Данные были успешно загружены в оба сервиса



Sales Person	Country	Product	Date	Amount	Boxes Shipped
Jehu Rudeforth	UK	Mint Chip Choco	04-Jan-22	\$5,320	180
Van Tuxwell	India	85% Dark Bars	01-Aug-22	\$7,896	94
Gigi Bohling	India	Peanut Butter Cubes	07-Jul-22	\$4,501	91
Jan Morforth	Australia	Peanut Butter Cubes	27-Apr-22	\$12,726	342
Jehu Rudeforth	UK	Peanut Butter Cubes	24-Feb-22	\$13,685	184
Van Tuxwell	India	Smooth Silky Salty	06-Jun-22	\$5,376	38
Oby Sorrel	UK	99% Dark & Pure	25-Jan-22	\$13,685	176
Gunar Cockshoot	Australia	After Nines	24-Mar-22	\$3,080	73
Jehu Rudeforth	New Zealand	50% Dark Bites	20-Apr-22	\$3,990	59
Brien Boise	Australia	99% Dark & Pure	04-Jul-22	\$2,835	102
Rafaellita Blaksland	UK	Smooth Silky Salty	13-Jan-22	\$4,704	62
Barr Faughny	USA	Orange Choco	10-Mar-22	\$3,703	11
Mallorie Weber	Canada	Eclairs	13-Jan-22	\$1,442	286
Karlen McCaffrey	New Zealand	Drinking Coco	28-Jul-22	\$168	156
Marney O'Brien	New Zealand	Peanut Butter Cubes	03-Aug-22	\$8,379	173
Beverie Moffet	Australia	Organic Choco Syrup	26-Jan-22	\$6,790	356
Van Tuxwell	Canada	Organic Choco Syrup	14-Feb-22	\$4,067	42
Roddy Speechley	USA	Smooth Silky Salty	05-Apr-22	\$3,017	140
Beverie Moffet	Canada	Milk Bars	16-Feb-22	\$8,799	250
Curtice Advani	UK	85% Dark Bars	08-Jun-22	\$1,085	172
Brien Boise	Australia	Eclairs	27-Jun-22	\$6,888	88

Рисунок 3 – Загрузка датасетов в Yandex DataLens

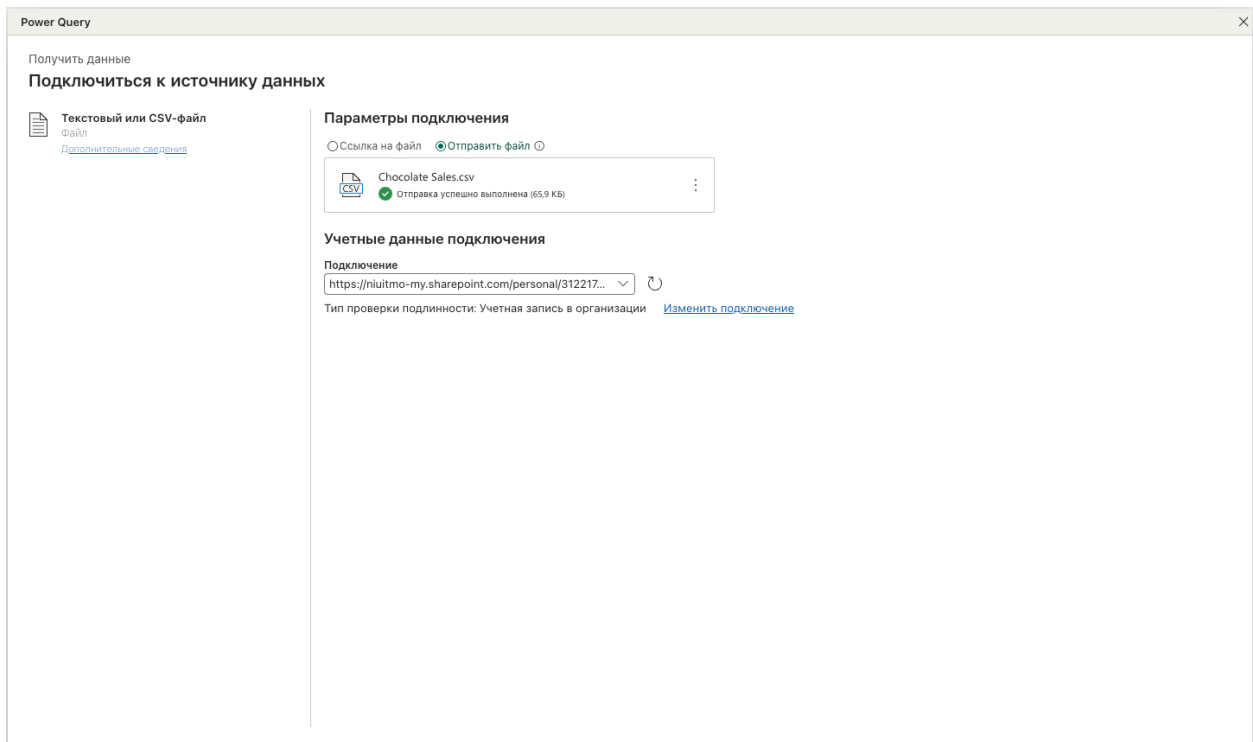


Рисунок 4 – Загрузка датасета в Power BI

Вывод.

В рамках лабораторной работы были успешно выполнены все этапы: освоение интерфейсов облачных сервисов Yandex DataLens и Power BI, выбор и анализ подходящих наборов данных: Walmart Stocks Data 2025 и Chocolate Sales Data, а также подбор альтернативных источников данных. Эти действия позволили создать базовую модель для дальнейшего анализа и визуализации, продемонстрировав возможности интеграции разнородных данных и подготовки аналитических отчетов.