

Лабораторная работа № 5

Тема: «Основы использования PowerBi для создания дэшбордов»

Цель работы: изучить основы работы с PowerBi.

Теоретическая справка

BI-системы – это набор инструментов и технологий для сбора, анализа и обработки данных, представляющий собой цифровое решение для поддержки принятия решений в бизнесе. Впервые термин «intelligence system» появился в статье «A Business Intelligence System», изданной в IBM System Journal (1958 г.), автор которой Ханс Петер Лун (Hans Peter Luhn) представил бизнес как набор различных видов деятельности в науке, технологиях, коммерции, индустрии и даже в законодательной сфере, а обеспечивающие его системы определил как системы, поддерживающие разумную деятельность.

BI-систему можно настроить под потребности конкретного бизнеса: выбрать источники данных, задать принцип обработки информации, определить формат отчетности. BI-системы позволяют объединять данные из внешних источников (открытые правительственные данные, рыночные показатели, данные из отчетности мониторинга и контроля государственных и административных структур, приобретенные базы данных) и из внутренних источников (CRM-систем (Customer Relationship Management, управление взаимоотношениями с клиентами), ERP-систем (Enterprise Resource Planning, управление ресурсами предприятия), SCM-систем (Supply Chain Management, управление цепочками поставок (снабжения)), workflow-систем (управление потоком работ), BPM-систем (business process management, управление бизнеспроцессами), иных учетных систем в бизнесе).

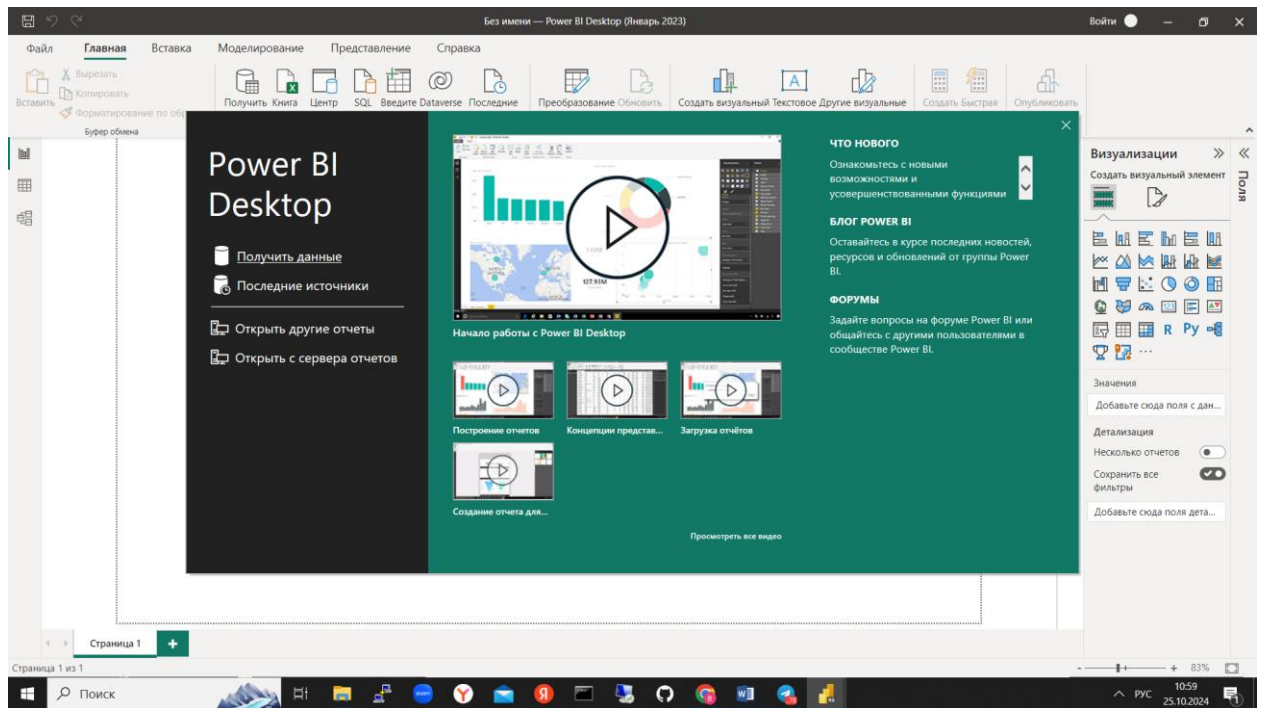
BI-системы решают следующие задачи:

- импорт/экспорт данных, подключение различных источников данных, очистка и преобразование данных;
- анализ данных (структурированных, неструктурированных),
- проверка гипотез для разработки управленческих решений на основе аналитики;
- формирование отчетности в виде панелей мониторинга (англ. Dashboards), данные на которых представлены в виде различных визуализаций: диаграмм, инфографики, контрольных показателей и т.п., предоставляя краткую обзорную информацию, которую можно уточнять, детализировать и обобщать по желанию пользователей;
- систематизация знаний о бизнесе, которая позволяет получить ретроспективу бизнес-деятельности, а также передать контексты операционных процессов новым сотрудникам.

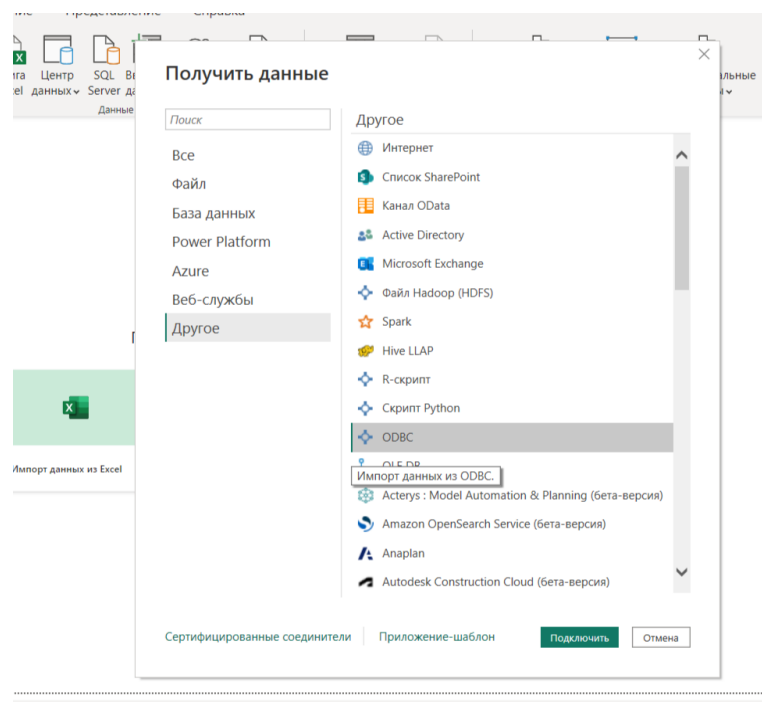
Power BI – это набор продуктов от Microsoft для решения задач бизнесаналитики. Основное назначение Power BI – это создание визуализаций разнообразных интерактивных данных, полученных из самых разных источников. Первый публичный показ Power BI состоялся в 2014.

Главная страница Power BI service

«Визуальная аналитика», Киреев В.С.

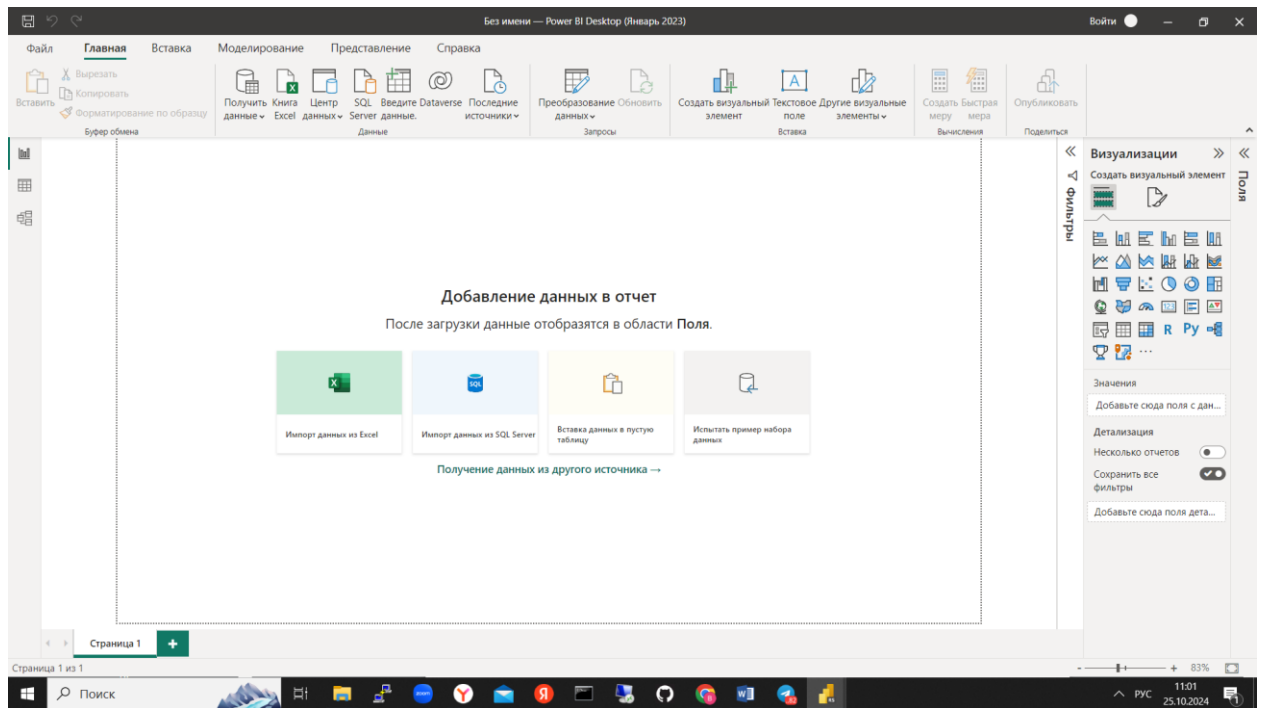


После выбора ссылки **Получить данные** происходит переход к диалоговому окну поиска и создания содержимого



В рабочей панели слева находятся пиктограммы для выбора одного из трех представлений: **Отчет**, **Данные** и **Модель**. Например, если выбрано представление **Отчет**, то в Power BI Desktop активируется набор визуализаций. Однако пока не проведено подключение данных, его просто не к чему применять. При наличии регистрации в Power BI service и авторизации в Power BI Desktop возможна передача подготовленных отчетов из Power BI Desktop в рабочую область Power BI service

«Визуальная аналитика», Киреев В.С.



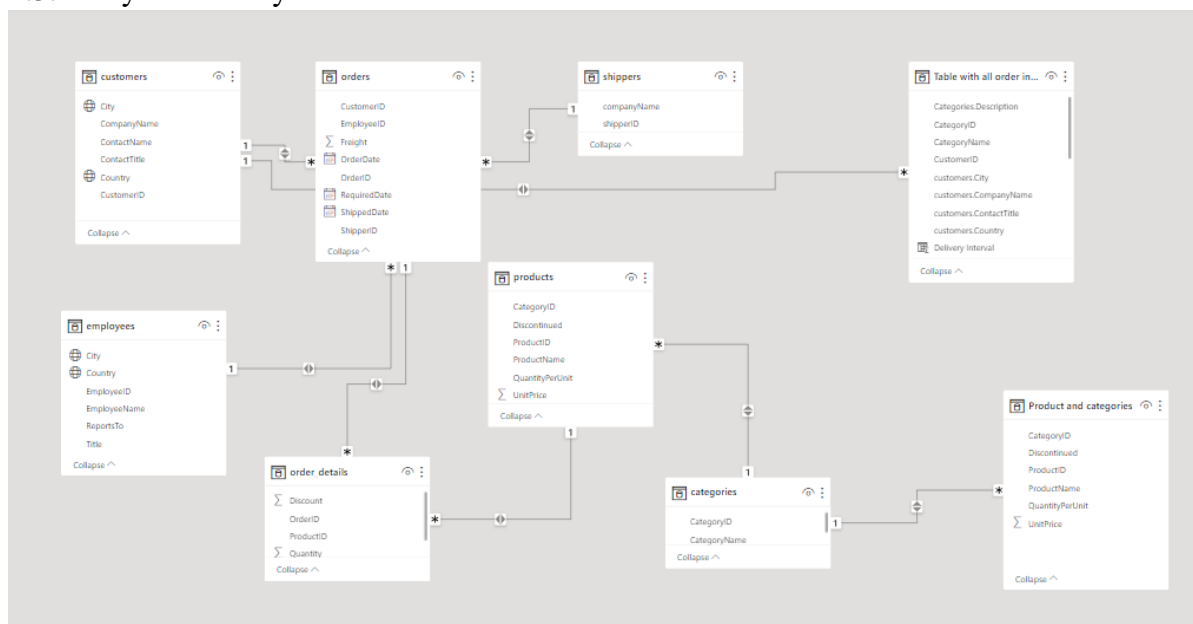
Power BI service Power BI Desktop включает в себя три компонента:

- Power Query – надстройку, которая позволяет подключиться к данным, провести их очистку и преобразование.
- Power Pivot – надстройку, позволяющую создавать сводные таблицы, проводить необходимые вычисления.
- Power View – компонент, отвечающий за визуализации данных.

«Визуальная аналитика», Киреев В.С.

1. Самостоятельное задание

1. Установить и настроить источник данных ODBC (создать новое подключение Northwind и указать путь к файлу бд, при необходимости скачать ODBC по [ссылке](#)). Открыть PowerBi и создать подключение к настроенному источнику данных.
2. Загрузить таблицы:
 - Заказы: содержат информацию о orderID, customerID, employeeID, orderDate, requiredDate, shippedDate, shipperID, freight.
 - Детали заказа: содержит информацию о orderID, productID, unitPrice, quantity, discount.
 - Сотрудники: содержит информацию о employeeID, employeeName, title, city, country, reportsTo.
 - Клиенты: содержит информацию о customerID, companyName, contactName, contactTitle, city, country.
 - Категории: содержит информацию о categoryID, categoryName, description.
 - Грузоотправители: содержит информацию о shipperID, companyName.
 - Продукты: содержат информацию о productID, productName, quantityPerUnit, unitPrice, прекращенном categoryID.
3. Провести разведочный анализ – пропущенные значения, число строк, столбцов и т.п.
4. Обработать данные:
 - 4.1. Объединить набор данных Product с Category, используя CategoryID -: «Продукты и категории».
 - 4.2. Создать новую таблицу «Таблица со всей информацией о заказах»
 - 4.3. Получить схему:



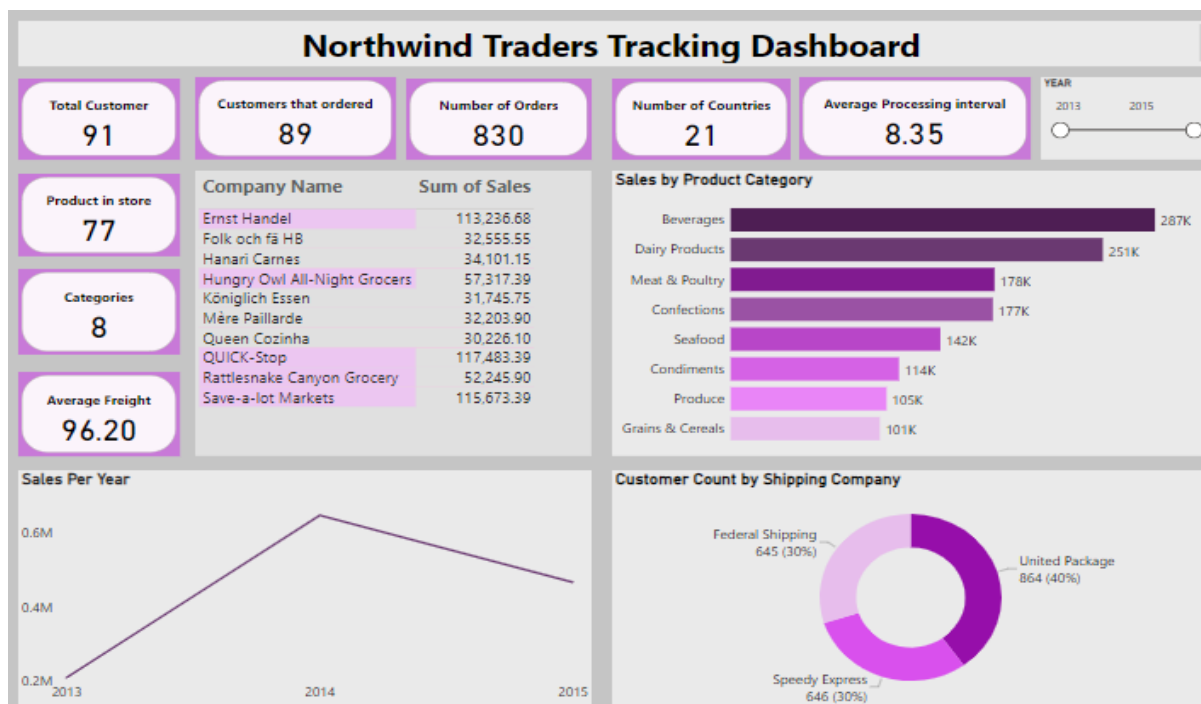
5. Создать меры:
 - 5.1. **Общее количество клиентов.** Вычисляет общее количество уникальных клиентов в наборе данных, используя агрегатную функцию COUNT(DISTINCT) из таблицы Customer Table.
 - 5.2. **Число уникальных клиентов, совершивших покупки в разные годы.** Здесь вычисляется количество клиентов, совершивших покупки в разные годы, с помощью агрегатной функции COUNT(DISTINCT) из таблицы Orders.

«Визуальная аналитика», Киреев В.С.

- 5.3. **Общее количество заказов за годы.** Рассчитывается общее количество отдельных заказов, сделанных клиентами за три года, то есть с 2013 по 2015 год, с помощью агрегатной функции COUNT(DISTINCT)
- 5.4. **Охваченные страны.** Здесь подсчитывается количество уникальных стран, включенных в набор данных, с помощью агрегатной функции COUNT(DISTINCT).
- 5.5. **Средний интервал обработки отправления.** Здесь вычисляется средний интервал обработки (в днях) для отправок, с помощью агрегатной функции Average. Для этого нужно создать новый столбец. Используя функцию DATEDIFF, вычислить разницу между датой заказа и датой отгрузки. В таком случае формула будет: Интервал обработки = DATEDIFF('Таблица со всей информацией о заказе'[OrderDate], 'Таблица со всей информацией о заказе'[ShippedDate], DAY)
- 5.6. **Продукты в магазине.** Здесь подсчитывается количество уникальных товаров, доступных в магазине, с помощью агрегатной функции COUNT(DISTINCT).
- 5.7. **Категории товаров.** Здесь подсчитывается количество уникальных категорий товаров, доступных в магазине, с помощью агрегатной функции COUNT(DISTINCT).
- 5.8. **Средний фрахт.** Рассчитывается средняя стоимость доставки для всех заказов с помощью агрегатной функции Average.
6. Создать визуализации:
 - 6.1. **Десять лучших клиентов по объему продаж.** Таблица отображает десять лучших клиентов по SUM продаж, представляя их названиями компаний. Для этого необходимо создать новый столбец.

Перед созданием визуального образа написать выражение для генерации продаж: Продажи = 'Таблица со всей информацией о заказах'[UnitPrice] * 'Таблица со всей информацией о заказах'[Quantity]

Использовать панель фильтрации. Отредактировать для Top N, используя столбец «Название компании».
 - 6.2. **Продажи из каждой категории товаров.** Гистограмма показывает СУММУ продаж, полученных от каждой категории товаров, помогая определить наиболее прибыльные категории.
 - 6.3. **Продажи за год.** Линейная диаграмма представляет собой разбивку SUM продаж по годам, позволяя анализировать тенденции продаж с течением времени.
 - 6.4. **Предпочтения клиентов по доставке.** Пончиковая диаграмма дает представление о предпочтениях клиентов по доставке, т. е. о распределении клиентов по транспортным компаниям.
7. Создать конечный дэшборд вида:



8. Проанализировать полученный дэшборд и попробовать получить инсайты о ситуации в компании.