

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

Платформы Data Engineering

**Лабораторная работа 1 Основы Data Engineering: работа с Git,
архитектурой и дашбордами**

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
2025

Содержание

Введение	2
Основная часть.....	2
Заключение	9

Введение

Цель

Познакомиться с основами Data Engineering: работа с Git, архитектурой и дашбордами

Задачи

1. Настроить Git.
2. Проиллюстрировать архитектуру аналитического решения, загрузить в репозиторий.
3. Выполнить подготовку дашборда по примеру, выгрузить в git.
4. Подготовить отчет по лабораторной работе.

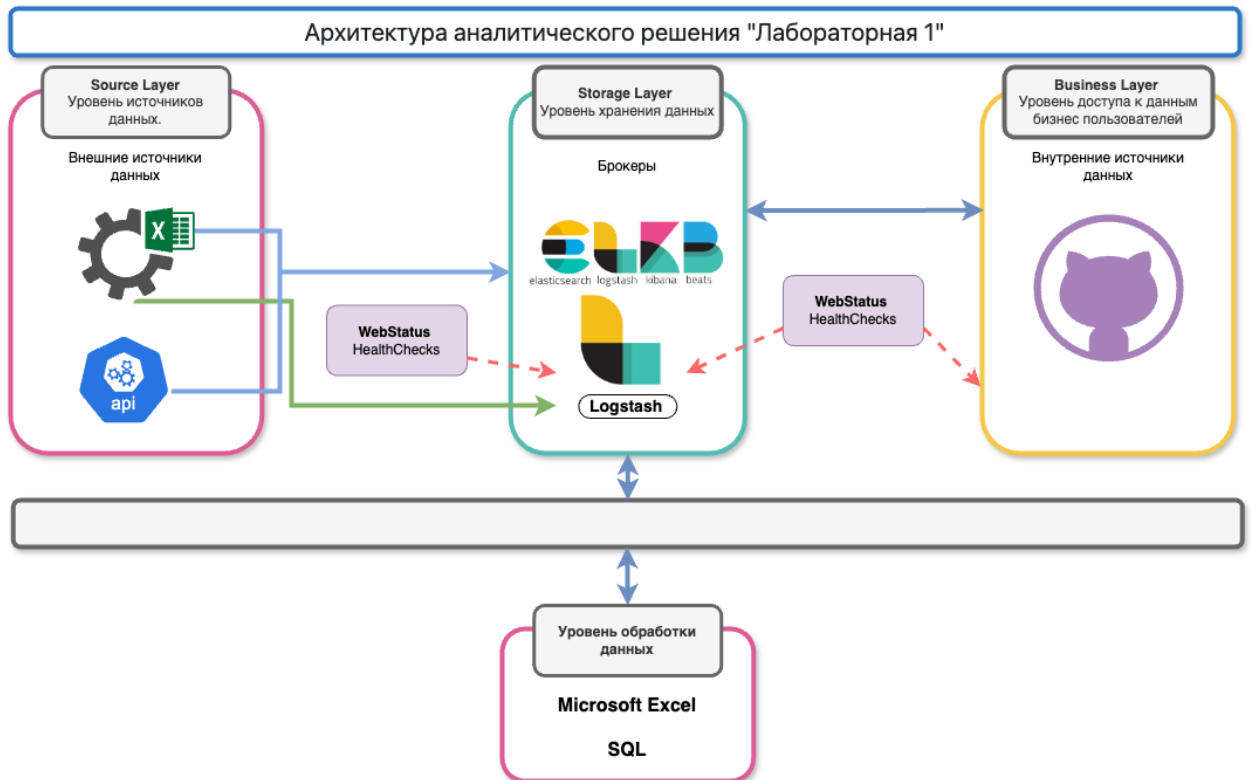
Основная часть

Задача 1. Подготовила Git-репозиторий для выполнения заданий дисциплины.

The screenshot shows the GitHub interface for the repository 'lezekiss / Data_Engineering'. The top navigation bar includes links for Code, Issues, Pull requests, Actions, Projects, and Wiki. Below this, the repository path 'Data_Engineering / Lab1' is shown. A commit history table is visible, showing a recent commit by 'lezekiss' with the message 'Добавила папку lab1 и нарисованную схему' (Added folder lab1 and a diagram) at 1b474ce, 29 minutes ago. The commit details table below shows the file 'Архитектура аналитического...' (Analytical architecture...) was added in the same commit.

Name	Last commit message	Last commit date
..		
Архитектура аналитического...	Добавила папку lab1 и нарисованную...	29 minutes ago

Задача 2. Построила архитектуру аналитического решения, добавила на Git



Задача 3. Создала таблицу из предоставленных данных

Задаем имя таблицы

Главная Вставка Рисование Разметка

Имя таблицы:

Data

Сводная таблица
Удалить дубликаты
Преобразовать в диапазон

C15 05.12.2018

	A	B	C	D	E	F	G
1	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
2	1	CA-2018-152156	08.11.2018	11.11.2018	Second Class	CG-12520	Claire Gute
3	2	CA-2018-152156	08.11.2018	11.11.2018	Second Class	CG-12520	Claire Gute
4	3	CA-2018-138688	12.06.2018	16.06.2018	Second Class	DV-13045	Darrin Van H
5	4	US-2017-108966	11.10.2017	18.10.2017	Standard Class	SO-20335	Sean O'Donne
6	5	US-2017-108966	11.10.2017	18.10.2017	Standard Class	SO-20335	Sean O'Donne
7	6	CA-2016-115812	09.06.2016	14.06.2016	Standard Class	BH-11710	Brosina Hoffin
8	7	CA-2016-115812	09.06.2016	14.06.2016	Standard Class	BH-11710	Brosina Hoffin
9	8	CA-2016-115812	09.06.2016	14.06.2016	Standard Class	BH-11710	Brosina Hoffin
10	9	CA-2016-115812	09.06.2016	14.06.2016	Standard Class	BH-11710	Brosina Hoffin

Добавила недостающие столбцы

	U	V	W	X
	Profit	Person	Return	
0	41,9136			
0	219,582			
0	6,8714			
15	-383,031			
,2	2,5164			
0	14,1694			
0	1,9656			
,2	90,7152			
,2	5,7825			

Подтянули недостающие данные при помощи формул

=ЕСЛИ(M2=People!\$B\$5;People!\$A\$5;ЕСЛИ(M2=People!\$B\$4;People!\$A\$4;ЕСЛИ(M2=People!\$B\$3;People!\$A\$3;ЕСЛИ(M2=People!\$B\$2;People!\$A\$2;0))))

и

=ЕСЛИОШИБКА(ИНДЕКС>Returns!A:A;ПОИСКПОЗ(B2>Returns!B:B;0));"No")

T	U	V	W	X
count	Profit	Person	Return	
0	41,9136	Cassandra Brandow	No	
0	219,582	Cassandra Brandow	No	
0	6,8714	Anna Andreadi	No	
0,45	-383,031	Cassandra Brandow	No	
0,2	2,5164	Cassandra Brandow	No	
0	14,1694	Anna Andreadi	No	
0	1,9656	Anna Andreadi	No	
0,2	90,7152	Anna Andreadi	No	
0,2	5,7825	Anna Andreadi	No	
0	34,47	Anna Andreadi	No	
0,2	85,3092	Anna Andreadi	No	
0,2	68,3568	Anna Andreadi	No	
0,2	5,4432	Cassandra Brandow	No	
0,2	132,5922	Anna Andreadi	No	
0,8	-123,858	Kelly Williams	No	
0,8	-3,816	Kelly Williams	No	

Определились с отчетах:

Параметры и показатели	Визуализация
Динамика дохода и прибыли	Диаграмма-график
Категории товаров(сравнение)	Линейчатая диаграмма
Региональные менеджеры(сравнение)	Гистограмма
Сегменты(сравнение)	Гистограмма
Динамика по сегментам	Диаграмма-график
Основные показатели	Спарклайны
По штатам	Картограмма
По регионам(сравнение)	Круговая
По возвратам в %	Круговая

Создаем первую сводную таблицу:

The screenshot shows the Excel interface for creating a PivotTable. The 'Вставка' (Insert) ribbon is active, and the 'Сводная таблица' (PivotTable) button is selected. The 'Поля сводной таблицы' (PivotTable Fields) task pane is open, displaying a list of fields: Row ID, Order ID, Order Date, and Ship Date. The main worksheet area shows a blank grid with a message 'Перетащите сюда поля значений' (Drag fields here).

Отобразила динамику дохода и прибыли.

Т.к это динамика, нам нужна дата. Перетаскиваем параметр даты (у нас это Order Date) в область строк.

Скриншот Excel с PivotTable. Таблица содержит данные по годам (2016, 2017, 2018) и месяцам (январь, февраль, март, апрель, май, июль, август, сентябрь, октябрь, ноябрь, декабрь). Показатели: Сумма по полю Sales, Сумма по полю Profit. В правой панели «Поля сводной таблицы» добавлены поля Profit и Order Date в область Значения.

Добавила в таблицу данные, перетащила нужные показатели в область «Значения».

Скриншот Excel с PivotTable. Таблица содержит данные по годам (2016, 2017, 2018, 2019) и кварталам (Кв-1, Кв-2, Кв-3, Кв-4). Показатели: Order Date, Ship Date. В правой панели «Поля сводной таблицы» добавлены поля Order Date и Ship Date в область Значения.

Преобразуем числа в таблице в денежный формат и округлим до целого

	A	B	C	D	E	F
1	Перетащите сюда поля фильтра отчета					
2						
3		Данные				
4	Названия строк	Сумма по полю Sales	Сумма по полю Profit			
5	2016	484247,4981	\$49 544			
6	январь	14236,895	\$2 450			
7	февраль	4519,892	\$862			
8	март	55691,009	\$499			
9	апрель	28295,345	\$3 489			
10	май	23648,287	\$2 739			
11	июнь	34595,1276	\$4 977			
12	июль	33046,303	\$841			

Поля сводной таблицы

ИМЯ ПОЛЯ

☒ Profit

☐ Person

☐ Returned

☒ Годы

Форматирование ячеек

Число

Числовые форматы: Образец

Общий

Числовой

Денежный

Финансовый

Дата

Время

Процентный

Дробный

Экспоненциальный

Текстовый

Дополнительный

(все форматы)

\$484 247

Число десятичных знаков: 0

Символ: \$ английский (США)

Отрицательные числа:

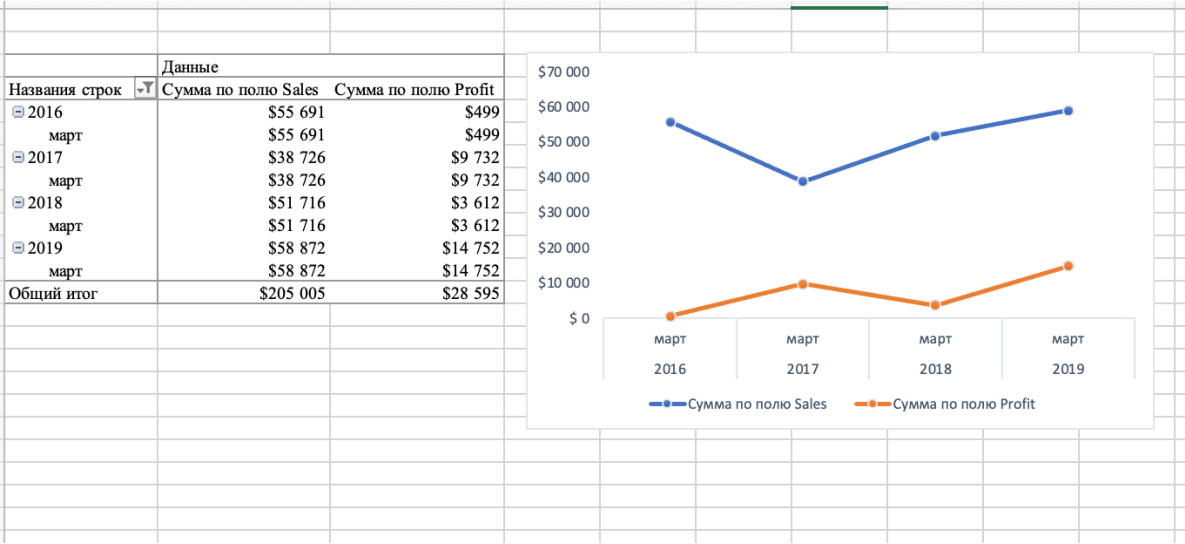
-\$1 234

\$1 234

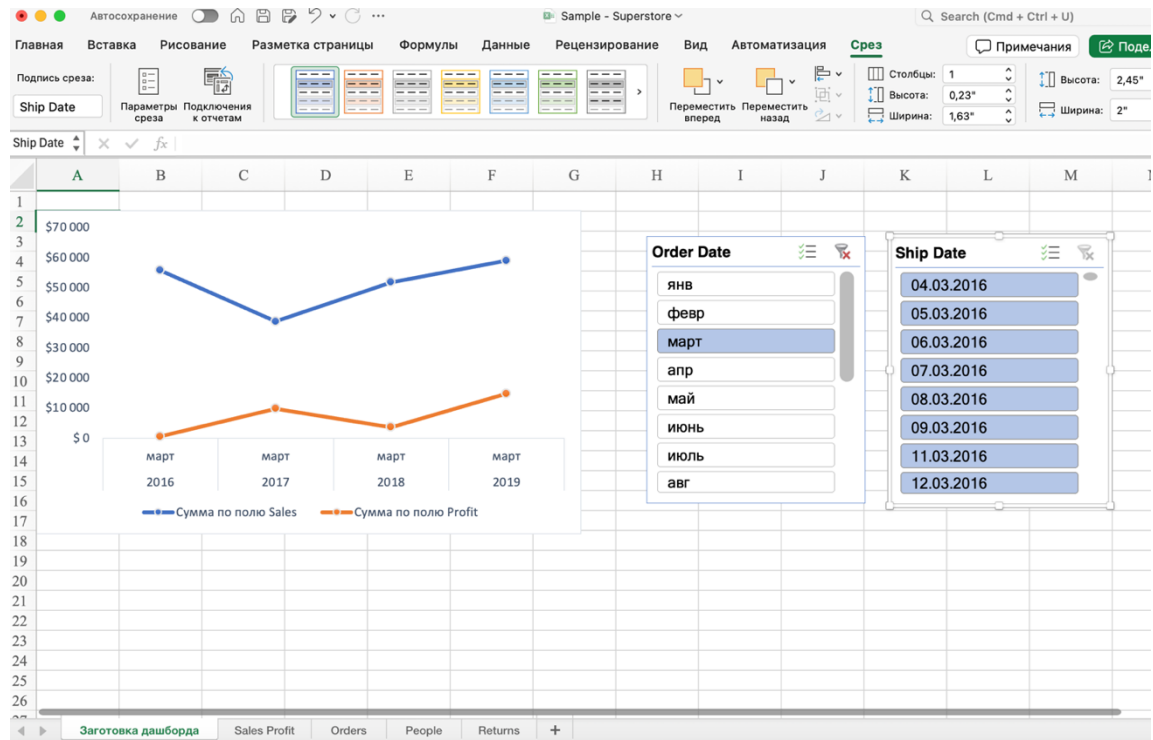
-\$1 234

-\$1 234

Создала диаграмму



Добавила ее на лист Дашборда и подготовила срез



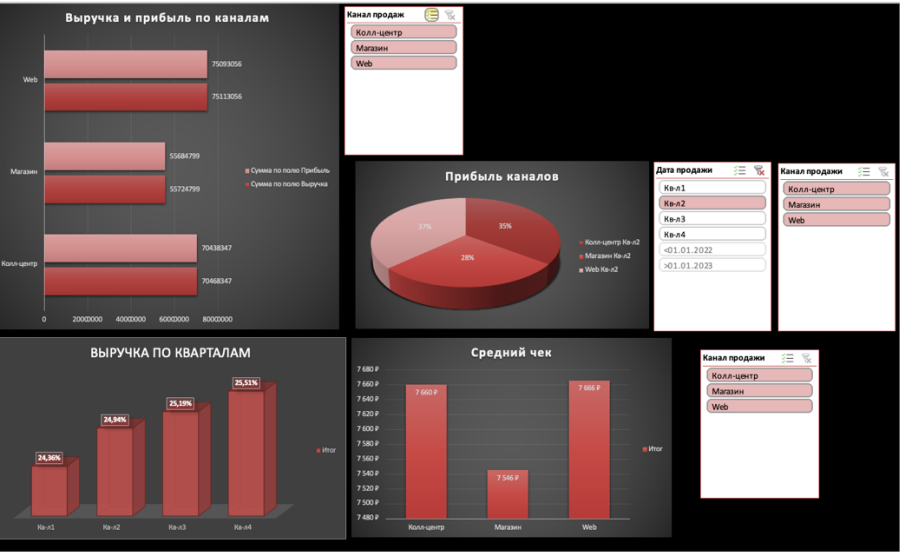
Перехожу к выполнению индивидуального задания

Перенесла затраты в общую таблицу при помощи формулы

=ЕСЛИИ(F4=returns!\$B\$4;returns!\$C\$4;ЕСЛИИ(F4=returns!\$B\$5;returns!\$C\$5;ЕСЛИИ(F4=returns!\$B\$6;returns!\$C\$6;0)))

Факты продаж, 2022 г.						
Артикул	Название товара	Цена, руб.	Дата продажи	Канал продажи	Затраты	Кол-во
6468	Миксер Philips HR 1560	2 453	25.01.2022	Магазин	40 000,00 Р	
7529	Электробритва Panasonic E5-3042	1 847	11.04.2022	Web	20 000,00 Р	
7591	Кухонный комбайн Bosch MUM 4655	9 240	08.12.2022	Web	20 000,00 Р	
7591	Кухонный комбайн Bosch MUM 4655	9 240	19.09.2022	Web	20 000,00 Р	
7604	Электронайчик Bosch TWK 1201	2 197	10.04.2022	Магазин	40 000,00 Р	
7604	Электронайчик Bosch TWK 1201	2 197	22.11.2022	Web	20 000,00 Р	
8302	Соковыжималка Philips HR 2744	1 716	01.05.2022	Колл-центр	30 000,00 Р	
31692	Миксер Philips HR 1565	3 740	19.07.2022	Колл-центр	30 000,00 Р	
31838	Электронайчик Braun WK 300	2 011	28.10.2022	Магазин	40 000,00 Р	
31838	Электронайчик Braun WK 300	2 011	04.10.2022	Web	20 000,00 Р	
31838	Электронайчик Braun WK 300	2 011	03.12.2022	Web	20 000,00 Р	
31838	Электронайчик Braun WK 300	2 011	28.05.2022	Web	20 000,00 Р	
31838	Электронайчик Braun WK 300	2 011	27.05.2022	Web	20 000,00 Р	
31901	Пылесос Thomas Bravo 20S Aquafilter	10 366	09.05.2022	Web	20 000,00 Р	
31901	Пылесос Thomas Bravo 20S Aquafilter	10 366	01.01.2022	Web	20 000,00 Р	

Результат выполнения индивидуального задания вы можете посмотреть в прикрепленном файле.



Вывод: Квартальная выручка демонстрирует устойчивый рост, что свидетельствует о положительной динамике развития бизнеса. Наиболее прибыльным каналом продаж является Web, за ним следует Колл-центр, а физический магазин занимает последнее место. Средние чеки по каналам Web и Колл-центр находятся на сопоставимом уровне, в то время как средний чек физического магазина значительно ниже. Для дальнейшего роста рекомендуется сосредоточить усилия на развитии продаж через Web и Колл-центр, так как они показывают наибольшую эффективность.



Заключение

Вывод: Квартальная выручка демонстрирует устойчивый рост, что свидетельствует о положительной динамике развития бизнеса.

Наиболее прибыльным каналом продаж является Web, за ним следует Колл-центр, а физический магазин занимает последнее место.

Средние чеки по каналам Web и Колл-центр находятся на сопоставимом уровне, в то время как средний чек физического магазина значительно ниже.

Для дальнейшего роста рекомендуется сосредоточить усилия на развитии продаж через Web и Колл-центр, так как они показывают наибольшую эффективность.