

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	Ошибка! Закладка не определена.

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. The main content area shows the 'Files' view for the 'main' branch, specifically the 'Data_Engineering / Lab1' directory. 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. Below the commit history, a table lists the files in the directory, including a folder named '..' and a file named 'Архитектура аналитического...' (Analytical architecture...).

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

[illegible]

Главная Вставка Рисование Разметка страницы Формулы Данные Рецензирование Вид Примечания Поделитесь

Имя таблицы: Таблица1

Сводная таблица
Удалить дубликаты
Преобразовать в диапазон

Вставить срез
Обновить

☒ Строка заголовка
☐ Строка итогов
☐ Чередующиеся строки

☐ Первый столбец
☐ Последний столбец
☐ Чередующиеся столбцы

☒ Кнопка фильтра

Экспресс-стили

C15 X ✓ fx 05.12.2018

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category
1	1	CA-2018-152156	08.11.2018	11.11.2018	Second Class	CG-12520	Claire Garte	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BD-100017 Furniture	I
2	2	CA-2018-152156	08.11.2018	11.11.2018	Second Class	CG-12520	Claire Garte	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-100004 Furniture	(
3	3	CA-2018-136868	12.06.2018	16.06.2018	Second Class	DV-13450	Darin Van Huff	Corporate	United States	Los Angel	California	90036	West	OFF-LA-100002 Office Supplies	I
4	4	US-2017-108966	11.10.2017	18.10.2017	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Laud	Florida	33311	South	FUR-TA-100005 Furniture)
5	5	US-2017-108966	11.10.2017	18.10.2017	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Laud	Florida	33311	South	OFF-ST-100007 Office Supplies	I
6	6	CA-2016-115812	09.06.2016	14.06.2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angel	California	90032	West	FUR-FU-100014 Furniture	I
7	7	CA-2016-115812	09.06.2016	14.06.2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angel	California	90032	West	OFF-AR-100028 Office Supplies	I
8	8	CA-2016-115812	09.06.2016	14.06.2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angel	California	90032	West	TEC-PH-100022 Technology	I
9	9	CA-2016-115812	09.06.2016	14.06.2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angel	California	90032	West	OFF-BI-1000391 Office Supplies	I
10	10	CA-2016-115812	09.06.2016	14.06.2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angel	California	90032	West	OFF-AP-100028 Office Supplies	I
11	11	CA-2016-115812	09.06.2016	14.06.2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angel	California	90032	West	FUR-TA-100015 Furniture)
12	12	CA-2016-115812	09.06.2016	14.06.2016	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angel	California	90032	West	TEC-PH-100020 Technology	I
13	13	CA-2019-14412	15.04.2019	20.04.2019	Standard Class	AA-10480	Andrew Allen	Consumer	United States	Concord	North Car	28027	South	OFF-PA-100023 Office Supplies	I
14	14	CA-2018-161389	05.12.2018	10.12.2018	Standard Class	IM-15070	Irene Maddox	Consumer	United States	Seattle	Washington	98103	West	OFF-BI-1000365 Office Supplies	I
15	15	US-2017-118983	22.11.2017	26.11.2017	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-AB-100023 Office Supplies	I
16	16	US-2017-118983	22.11.2017	26.11.2017	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-1000365 Office Supplies	I
17	17	CA-2016-105893	11.11.2016	18.11.2016	Standard Class	PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-100041 Office Supplies	\$
18	18	CA-2016-167164	13.05.2016	15.05.2016	Second Class	AG-10270	Alejandro Grove	Consumer	United States	West Jor	Utah	84084	West	OFF-ST-100001 Office Supplies	I
19	19	CA-2016-143336	27.08.2016	01.09.2016	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Franci	California	94109	West	OFF-AR-100030 Office Supplies	I
20	20	CA-2016-143336	27.08.2016	01.09.2016	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Franci	California	94109	West	TEC-PH-100119 Technology	I
21	21	CA-2016-143336	27.08.2016	01.09.2016	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Franci	California	94109	West	OFF-BI-1000221 Office Supplies	I
22	22	CA-2018-173730	09.12.2018	13.12.2018	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AR-100002 Office Supplies	I
23	23	CA-2018-173730	09.12.2018	13.12.2018	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AP-100014 Office Supplies	I

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(М2=People!\$B\$5;People!\$A\$5;ЕСЛИ(М2=People!\$B\$4;People!\$A\$4;ЕСЛИ(М2=People!\$B\$3;People!\$A\$3;ЕСЛИ(М2=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:

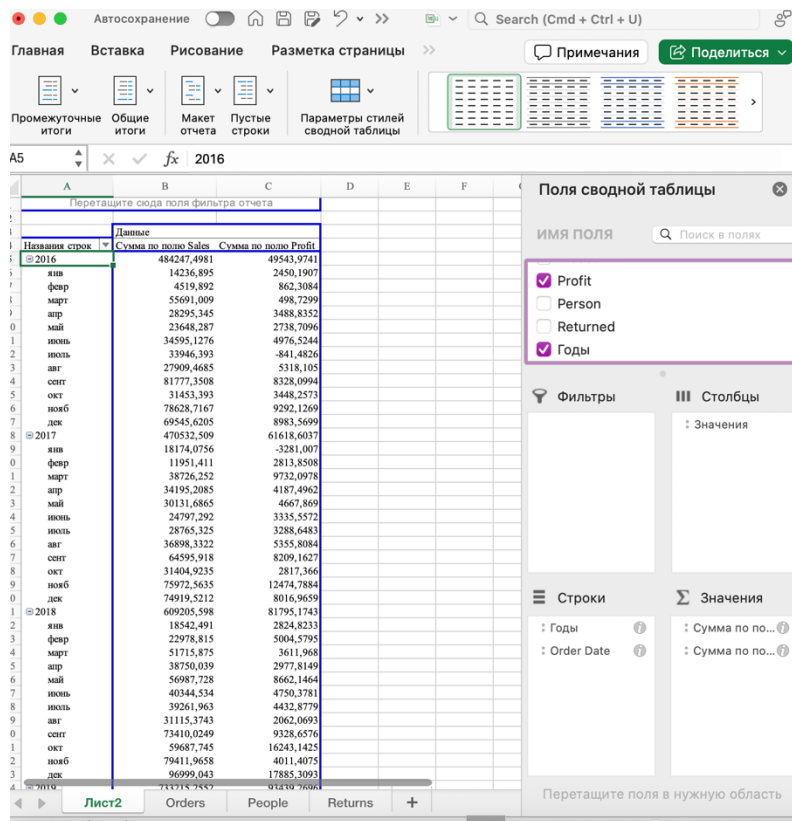
Параметры и показатели	Визуализация
Динамика дохода и прибыли	Диаграмма-график
Категории товаров(сравнение)	Линейчатая диаграмма
Региональные менеджеры(сравнение)	Гистограмма
Сегменты(сравнение)	Гистограмма
Динамика по сегментам	Диаграмма-график
Основные показатели	Спарклайны
По штатам	Картограмма
По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

Создаем первую сводную таблицу:

Скриншот Excel, показывающий процесс создания сводной таблицы. Вкладка 'Вставка' (Insert) активна, и выбран пункт 'Сводная таблица' (PivotTable). В панели задач 'Поля сводной таблицы' (PivotTable Fields) перечислены доступные поля: Row ID, Order ID, Order Date, Ship Date. Основная рабочая область Excel содержит таблицу с заголовками и данными, готовую для анализа.

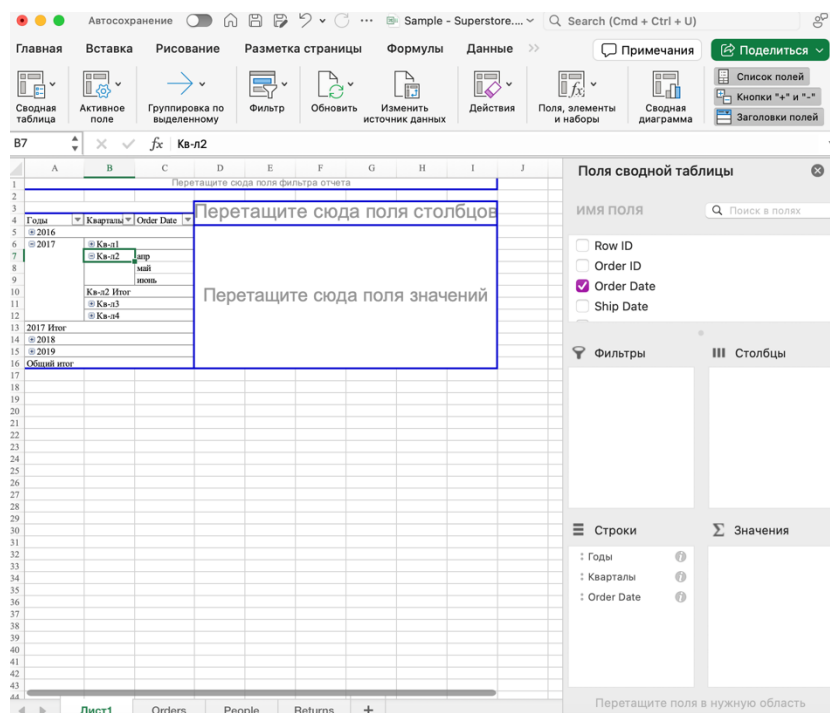
Отобразила динамику дохода и прибыли.

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



Скриншот Excel, демонстрирующий PivotTable с данными по годам и датам заказов. В правой панели «Поля сводной таблицы» (PivotTable Fields) в разделе «Фильтры» (Filters) выбраны «Годы» (Years) и «Order Date». В разделе «Значения» (Values) добавлены «Сумма по полю Profit» (Sum of Profit) и «Сумма по полю Order Date» (Sum of Order Date). Таблица данных отображает суммы по годам (2016, 2017, 2018) и по месяцам (январь, февраль, март, апрель, май, июль, август, сентябрь, октябрь, ноябрь, декабрь).

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



Скриншот Excel, демонстрирующий PivotTable с данными по годам, кварталам и датам заказов. В правой панели «Поля сводной таблицы» (PivotTable Fields) в разделе «Фильтры» (Filters) выбраны «Годы» (Years), «Кварталы» (Quarters) и «Order Date». В разделе «Значения» (Values) добавлены «Сумма по полю Order Date» (Sum of Order Date) и «Сумма по полю Ship Date» (Sum of Ship Date). Таблица данных отображает суммы по годам (2016, 2017, 2018) и по кварталам (Кв-1, Кв-2, Кв-3, Кв-4).

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:

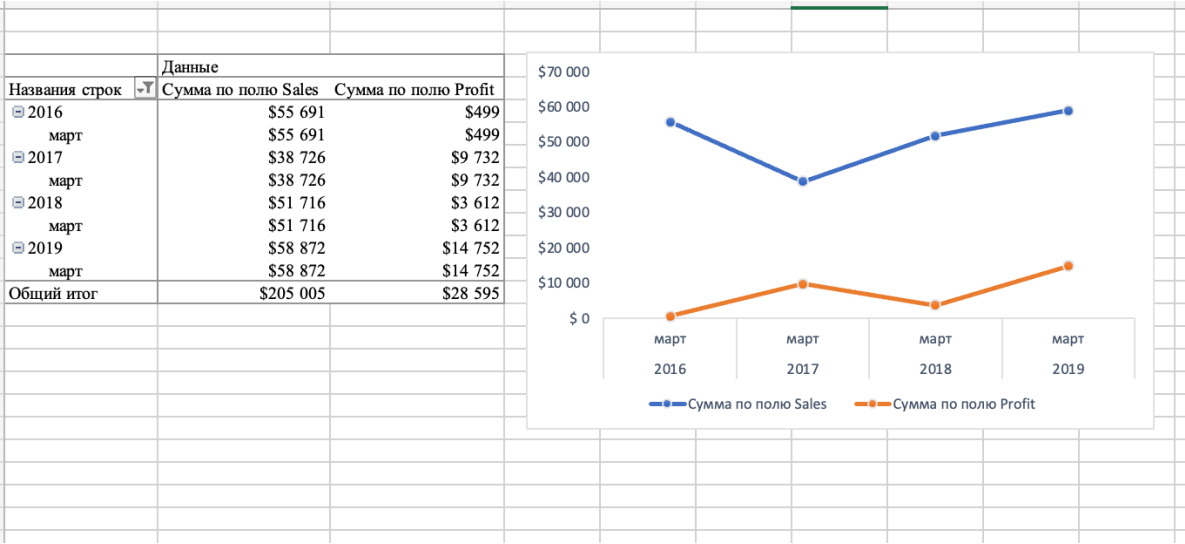
-\$1 234

\$1 234

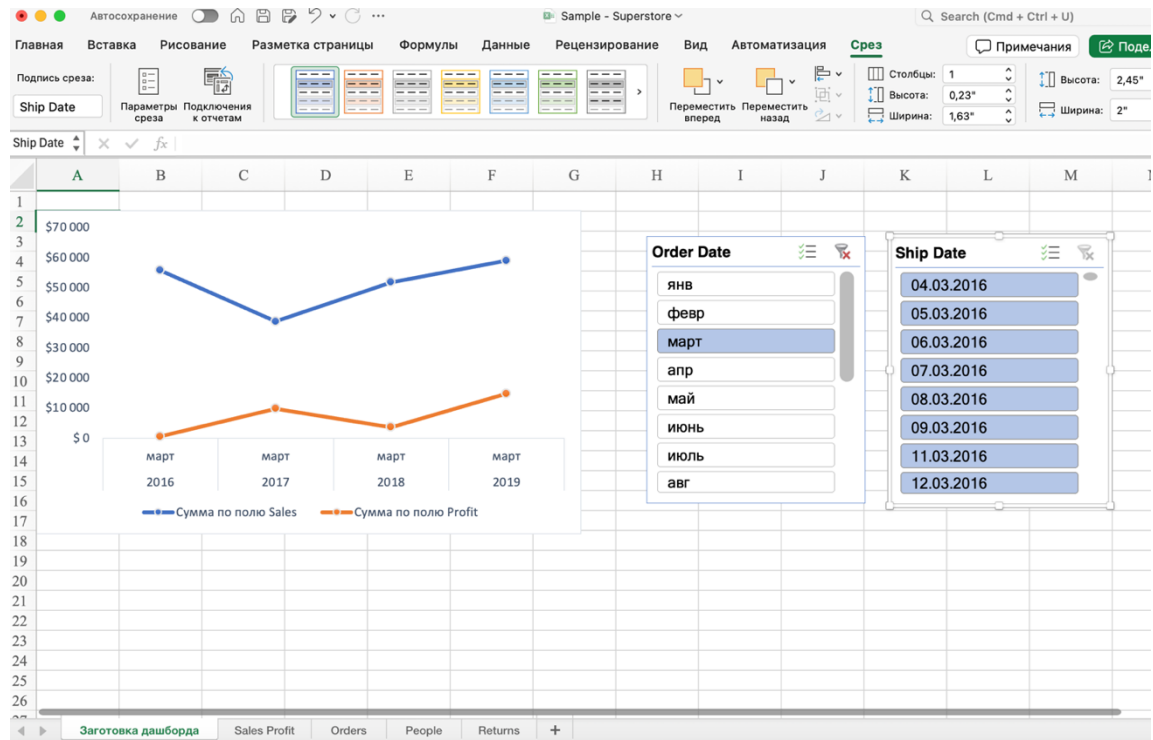
-\$1 234

-\$1 234

Создала диаграмму



Добавила ее на лист Дашборда и подготовила срез



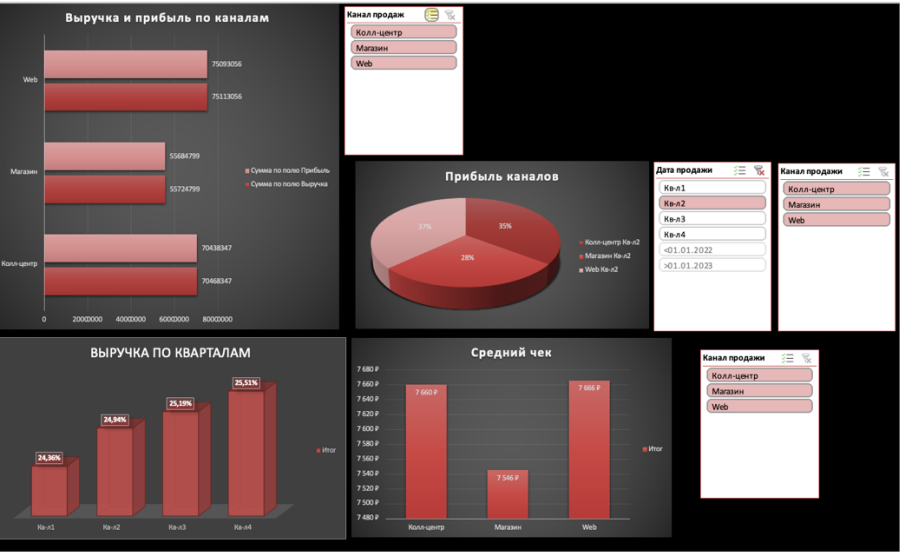
Перехожу к выполнению индивидуального задания

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.



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.