

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

st_79

Тема практической работы

«Основы Data Engineering: работа с архитектурой и дашбордами»

Практическая работа 1.1

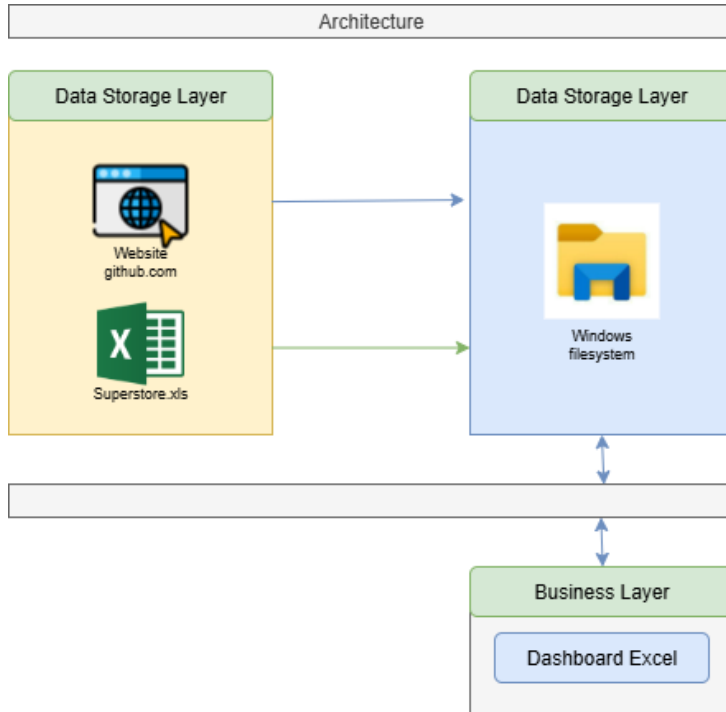
Москва

2025

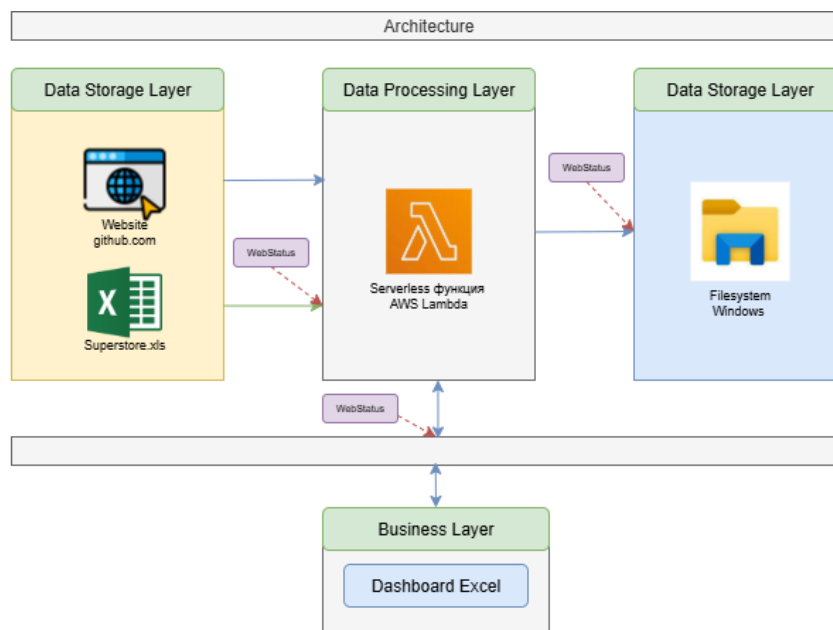
Вариант 20

Анализ потребительского сегмента (Consumer). Что покупают и насколько прибыльны частные клиенты ?

Фактическая архитектура решения:



1 Добавить в Processing Layer иконку "Serverless функции" (AWS Lambda, Google Functions).



Комплексный анализ данных в Excel

Источник данных – лист эксель, для ответа на вопрос «Что покупают и насколько прибыльны частные клиенты?» был использован лист «Orders», в категориальных данных не было выявлено пропущенных или неверно форматированных данных. При этом в столбце «Sales» стандартное отклонение выше, чем среднее значение, что изображено в таблице 1, что говорит о выбросах в данных, на рисунке 1 можно заметить, что в данных заметны несколько значений-выбросов.

Таблица 1 – статистические характеристики столбца «Sales»

Среднее по полю Sales	Минимум по полю Sales	Максимум по полю Sales	Стандартное отклонение по полю Sales
229,8580008	0,444	22638,48	623,2451005



Рисунок 1- box plot по полю «Sales».

При этом в разбивке по категориям и суб-категориям распределение «Sales» становится более «нормальным», что изображено в таблице 2. Поэтому анализ прибыли необходимо проводить учитывая категории.

Таблица 2 – статистические характеристики столбца «Sales» в разбивке по категориям и суб-категориям

Category	Sub-Category	Среднее по полю Sales	Минимум по полю Sales	Максимум по полю Sales	Стандартное отклонение по полю Sales
Furniture	Bookcases	503,86	35,49	4404,90	638,75
	Chairs	532,33	26,64	4416,17	550,15
	Furnishings	95,83	1,89	1336,44	147,89
	Tables	648,79	24,37	4297,64	615,77
Office Supplies	Appliances	230,76	0,44	2625,12	388,95
	Art	34,07	1,34	1113,02	60,12
	Binders	133,56	0,56	9892,74	563,25
	Envelopes	64,87	1,63	604,66	84,35
	Fasteners	13,94	1,24	93,36	12,42
	Labels	34,30	2,09	786,48	74,12
	Paper	57,28	3,38	733,95	78,17
	Storage	264,59	4,46	2934,33	355,22
	Supplies	245,65	1,74	8187,65	923,83
	Technology	215,97	0,99	3347,37	334,97
Technology	Copiers	2198,94	299,99	17499,95	3175,67
	Machines	1645,55	11,56	22638,48	2765,10
	Phones	371,21	2,97	4548,81	491,46

Похожая ситуация была обнаружена по полю «Profit», при этом в данном столбце присутствуют отрицательные значения.

На этапе трансформации данных было выполнено объединение таблиц «Orders» и «Returns» и добавление столбца Return со значениями Yes/No, а также расчёт метрик.

2 Детальный дашборд по сегменту Consumer: топ категории, топ штаты, динамика продаж.

В среднем и по сумме категория «Technology» приносит наибольший доход, что изображено на рисунках 2 и 3.

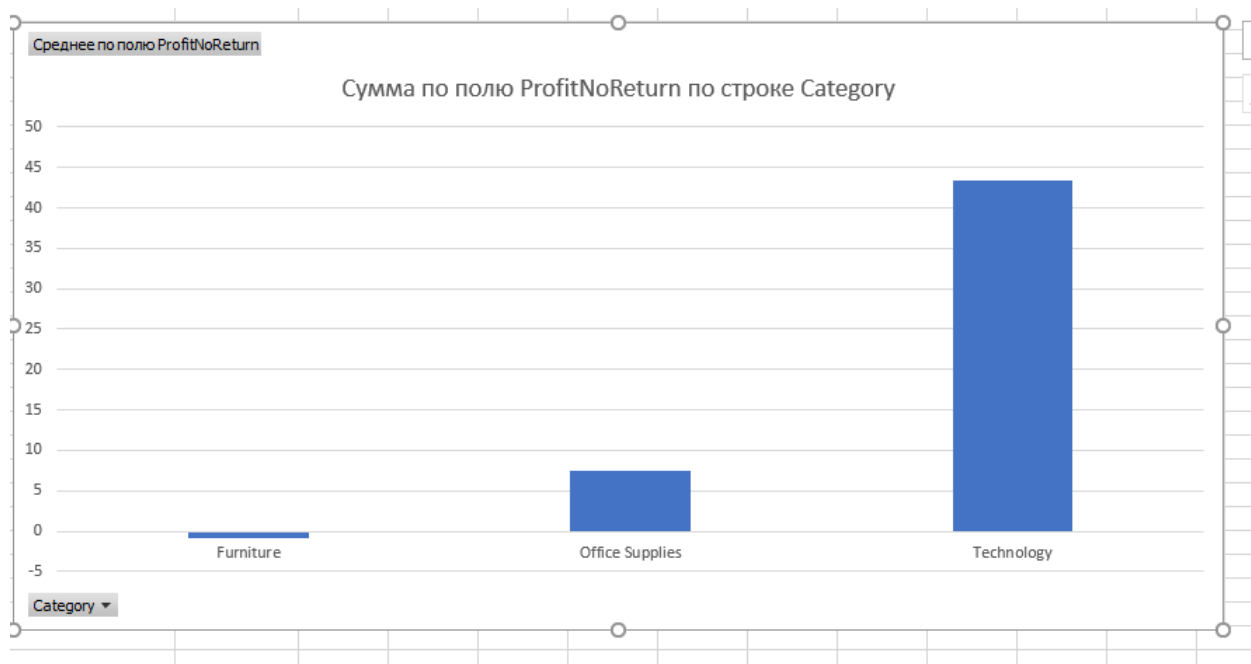


Рисунок 2 - прибыль в среднем по категориям.

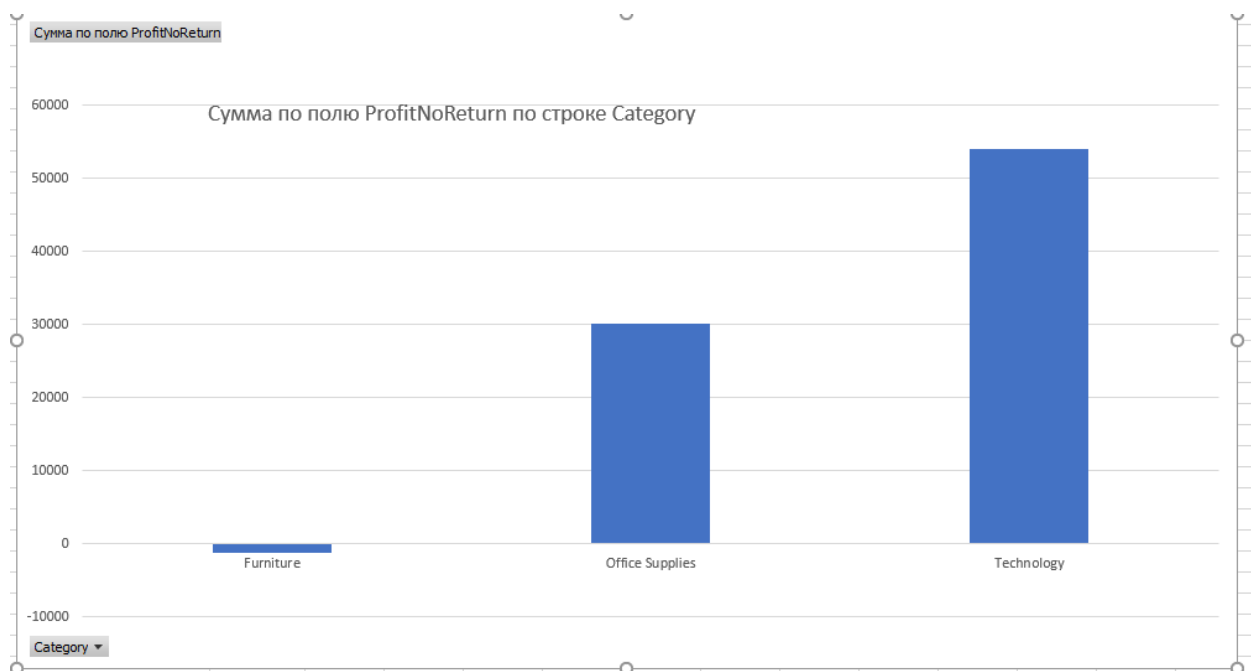


Рисунок 3 – сумма прибыли по категориям.

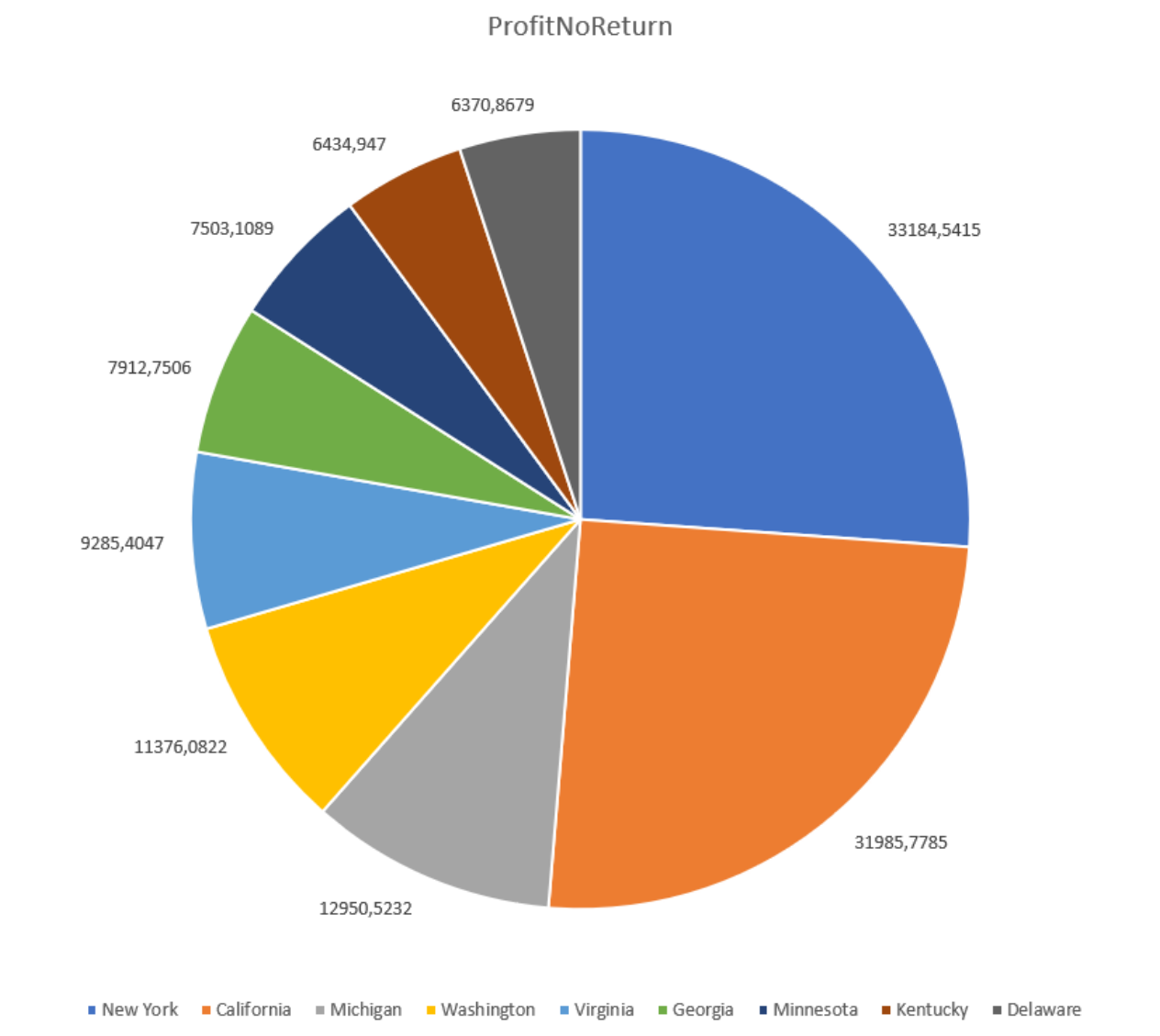


Рисунок 4 – топ 10 штатов по прибыли.

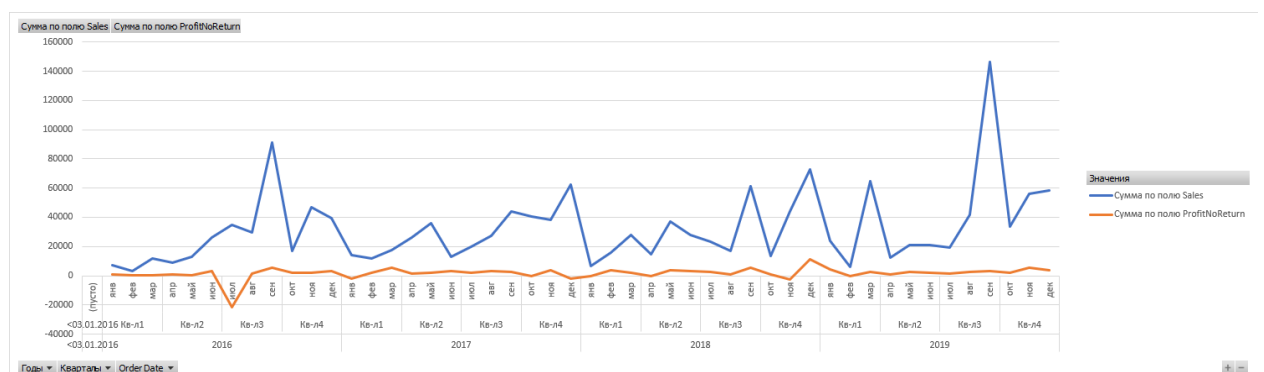


Рисунок 5 – динамика продаж и прибыли с 2016 по 2019 годы

3 Рассчитать средний чек (Sales / кол - во Order ID) для сегмента Consumer.

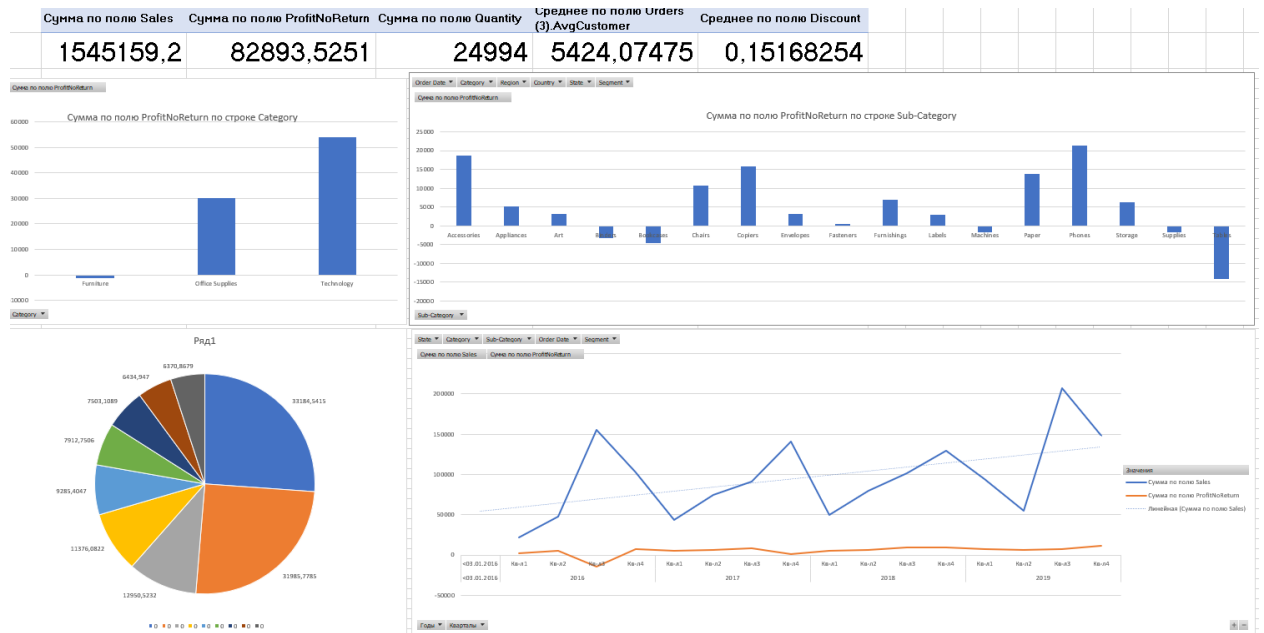
Расчёт был выполнен по следующим этапам, результат на рисунке :

1. Загрузка таблицы «Orders» в Power Query
2. Создание столбца SumSale умножением Sales на Quantity
3. Создание SumOrder суммированием SumSale по Order ID
4. Создание AvgCustomer, среднее SumOrder по Customer ID

	A	B	C
1	Customer ID ▾	AvgCustomer ▾	
2	SM-20320	66833,23267	
3	TC-20980	37768,83867	
4	RB-19360	10554,21617	
5	TA-21385	26609,6672	
6	HL-15040	7894,386	
7	AB-10105	8316,80035	
8	SC-20095	10736,28036	
9	BS-11365	22197,12678	
10	SE-20110	8133,423263	
11	CC-12370	16727,26218	
12	KL-16645	6896,773276	
13	BM-11140	37280,80713	
14	GT-14635	12981,34467	
15	TB-21400	4196,215882	
16	CM-12385	16251,6958	
17	AP-10540	8322,264889	
18	DP-12940	14254,75262	
19	TS-21370	8687,5734	
20	JW-15220	9369,171857	
21	EH-13765	3429,193875	
22	KD-16270	14185,4915	
23	HM-14860	5322,36468	
24	NC-18535	4896,5152	
25	JA-15970	5274,748625	
26	KD-16495	4417,894286	
27	JM-15865	6599,238	
28	DP-13390	3339,784429	
29	CS-12505	8227,498778	
30	MJ-17740	4412,2132	
31	TP-21415	4735,67848	
32	QJ-19255	3073,500154	
33	PO-18850	14674,88266	
34	AB-10060	7830,908444	
35	PK-19075	4370,65424	
36	JH-15985	7372,129571	
37	PF-19120	21864,94168	
38	GM-14695	10647,93145	
39	CP-12085	6868,229714	
40	KF-16285	8329,182889	
41	GT-14710	6892,833931	
42	CJ-12010	11563,3132	
43	NM-18445	5022,438143	
44	AA-10315	5350,340364	
45	SD-20485	4046,689619	

Рисунок т - Средний чек для Consumer.

4 Слайд: "Портрет потребителя: что, где и когда покупает наш ключевой сегмент."



Анализ данных по частным клиентам выявил выраженную географическую концентрацию прибыли: Нью-Йорк (33,2 тыс.) и Калифорния (32,0 тыс.) являются ключевыми драйверами, значительно опережая другие штаты. Наблюдается резкий дисбаланс в прибыльности товарных категорий: Technology доминирует с высокой маржой и абсолютной прибылью, благодаря успеху Phones и Copiers; Office Supplies остается стабильно прибыльным, в то время как Furniture — убыточна из-за катастрофических показателей Tables и Bookcases. Динамика продаж показывает устойчивый рост с 2016 по 2019 год с выраженной сезонностью в 3-4 кварталах, однако общая маржа остается под давлением высоких средних скидок, что требует оптимизации ассортимента в пользу высокомаржинальных категорий и пересмотра ценовой политики в проблемных сегментах.