

# «Анализ продаж e-commerce магазина (поиск инсайтов, составление рекомендаций стейкхолдерам)»

Дипломная работа по программе «Аналитик данных»

Букшенко А. М.  
Группа: DA-114

2025г.



# Описание бизнес-задачи, стейкхолдеров отчёта



1

## Описание бизнес-задач

- Повышения эффективности товарного ассортимента
- Оптимизации маркетинговых активностей
- Улучшения клиентского опыта
- Максимизации прибыли

## Стейкхолдеры

- Генеральный директор
- Директор по маркетингу

## Основные задачи анализа

- Оценка динамики продаж
- ABC-анализ товарного ассортимента
- RFM-сегментация клиентской базы
- Выявление сезонных трендов



# Описание данных и их предобработки



2

# Описание данных и их предобработки

Для исследования был взят датасет "[TATA: Online Retail Dataset](#)" со статистикой продаж интернет-магазина.

На этапе предобработки данных выявлено и удалено 5,268 дубликатов по всем столбцам (0.97% от общего объема записей)

Данные до преобразований

№	Имя Столбца	Описание	Тип данных
1	InvoiceNo	Номер счета - фактуры (Уникальный номер счета - фактуры который идентифицирует каждую транзакцию)	object
2	StockCode	Код акций (Уникальный код товара, который используется для идентификации конкретного продукта в системе)	object
3	Description	Описание (Описание товара )	object
4	Quantity	Количество (Количество единиц товара, купленных в рамках данной транзакции)	int64
5	InvoiceDate	Дата счета - фактуры (Дата и время, когда была совершена покупка)	object
6	UnitPrice	Цена (Цена за единицу товара)	float64
7	CustomerID	Идентификатор клиента (Уникальный идентификатор клиента, который позволяет отслеживать покупки конкретного клиента)	float64
8	Country	Страна (Страна, из которой совершен заказ)	object

Данные после очистки и преобразований

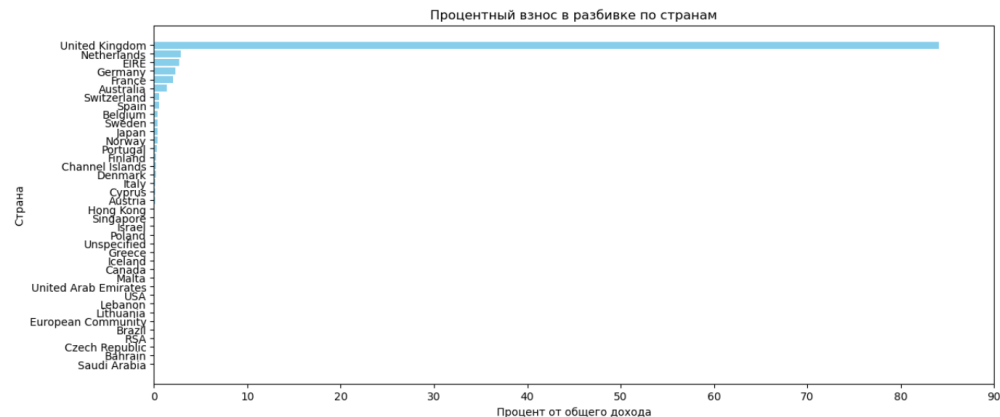
№	Имя столбца	Преобразование данных	% NaN	Очистка данных
1	InvoiceNo	object	0.00	Без изменений. Нет пустот. Обнаружены возвраты с отрицательным значением Quantity, где номер транзакции (InvoiceNo) начинается на букву "C". Данные транзакции будут удаляться только для проведения RFM-анализа. Для других видов анализа они сохранены, так как при расчете общих сумм положительные и отрицательные значения компенсируют друг друга. Удалена 1 строка с большой суммой и отрицательным количеством для того чтобы не портить общую картину анализов.
2	StockCode	object	0.00	Без изменений. Нет пустот
3	Description	object	0.27	Пустоты заменены на 'unknown', чтобы не исказить общие показатели, основной ориентир будет на StockCode(столбец в котором не было обнаружено пропусков), при необходимости позицию можно определить по StockCode).
4	Quantity	int64	0.00	Удалены 1336 строк с отрицательным значением и 0 ценной, которые сложно отнести к какой либо категории товара
5	InvoiceDate	datetime64[ns]	0.00	Без изменений. Нет пустот. 2011.12 не полный месяц, данные только до 2011-12-09
6	UnitPrice	float64	0.00	Удалены 3 строки с задолженностями 'Adjust bad debt'. 2 из них с отрицательным значением, чтобы не исказить общие показатели
7	CustomerID	float64	24.93	Пустоты заменены на '0', чтобы не исказить общие показатели. Для RFM анализа будет сделана дополнительная очистка по удалению строк с 0 значением
8	Country	object	0.00	Без изменений. Нет пустот

# Описание проведенного исследования

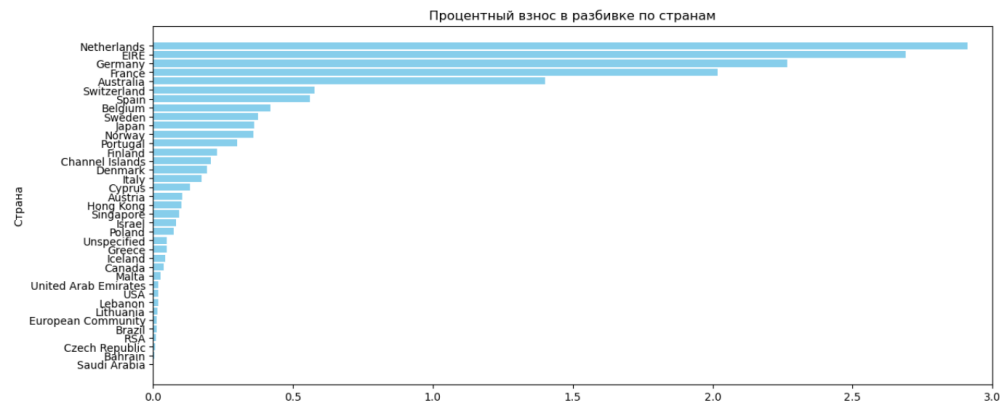


3

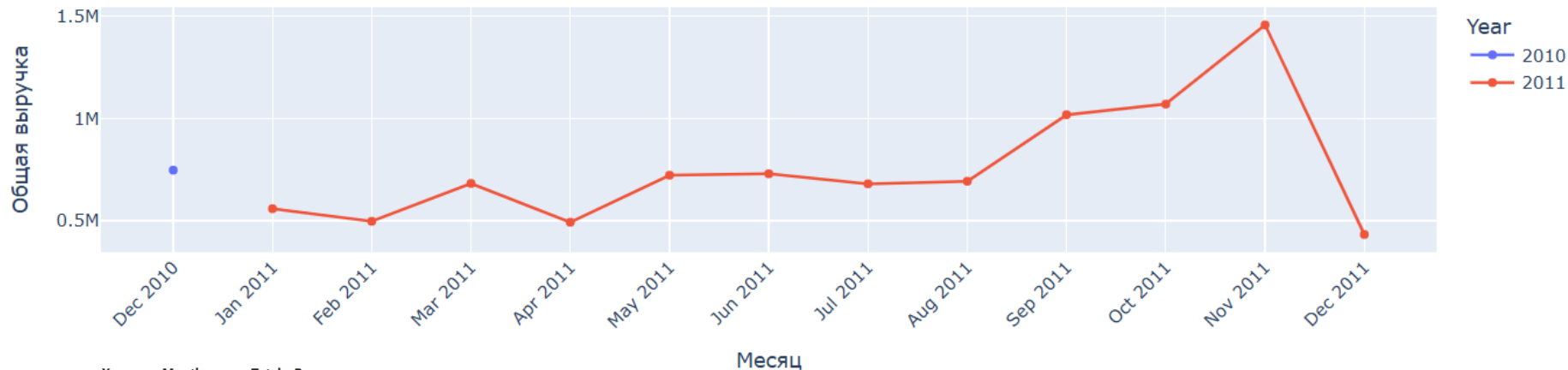
# Динамика продаж



Выделяются 5 топ стран:  
Великобритания, Нидерланды,  
Ирландия, Германия, Франция,  
обеспечивают 93.9% выручки. 35  
остальных рынков в сумме дают  
лишь 6.1% дохода



## Динамика продаж по месяцам



Year	Month	Total	Revenue
2010	December	746723.61	NaN
2011	January	558448.56	-0.25
2011	February	497026.41	-0.11
2011	March	682013.98	0.37
2011	April	492367.84	-0.28
2011	May	722094.10	0.47
2011	June	728947.23	0.01
2011	July	680156.99	-0.07
2011	August	692448.52	0.02
2011	September	1017596.68	0.47
2011	October	1069368.23	0.05
2011	November	1456145.80	0.36
2011	December	432701.06	-0.70

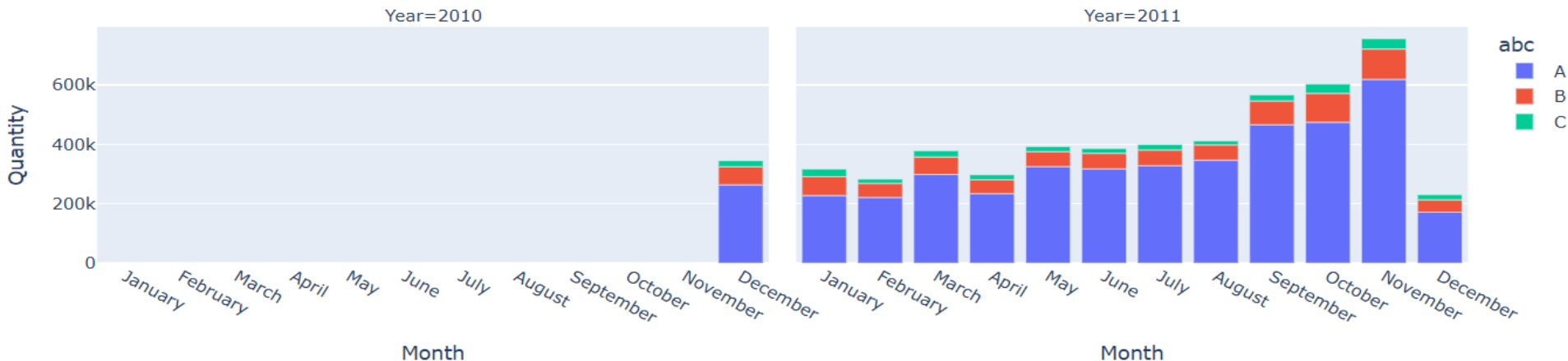
Данные показывают значительные колебания в динамике продаж. Например, в марте 2011 года наблюдается внезапный рост на 37.2%, за которым следует падение на 27.8% в апреле.





# ABC анализ по товарам

## Продажи по ABC-категориям



StockCode	Quantity	Share	Cumsum	abc
22197	56427	0.01	0.01	A
84077	53751	0.01	0.02	A
85099B	47260	0.01	0.03	A
85123A	39067	0.01	0.04	A
84879	36282	0.01	0.04	A
...	...	...	...	...
85044	1	0.00	1.00	C
20950	1	0.00	1.00	C
23843	0	0.00	1.00	C
85047	0	0.00	1.00	C
79323B	0	0.00	1.00	C

3932 rows × 4 columns

Распределил товар согласно правилу Парето, по категориям:  
Категория А (первые 20% товаров): Вклад в продажи: ~80% от общего объема.

Категория В (следующие 15% товаров): Вклад в продажи: ~15%.

Категория С (оставшиеся 65% товаров): Вклад в продажи: ~5%.

Посмотрев на график можно выделить сезонность для товаров категории А в сентябре, октябре и ноябре

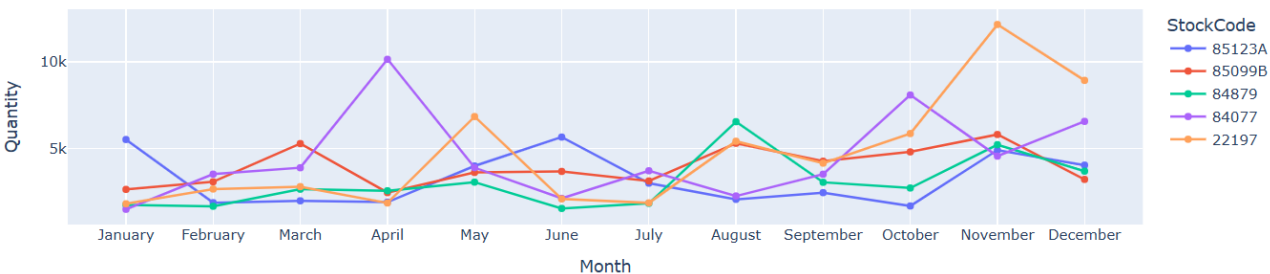
## Топ 5 товаров категории A

	Quantity	Share	Cumsum	abc
StockCode				
22197	56427	0.01	0.01	A
84077	53751	0.01	0.02	A
85099B	47260	0.01	0.03	A
85123A	39067	0.01	0.04	A
84879	36282	0.01	0.04	A

Получившийся топ показывает абсолютных лидеров продаж.

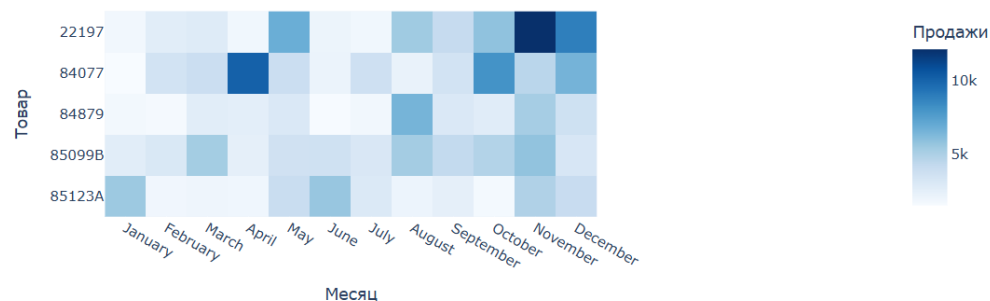


Сезонность топ-5 товаров категории А



Ярко выраженная сезонность.  
Ноябрь (макс. продажи 22197),  
апрель(84077), декабрь (22197),  
май (22197).

Тепловая карта сезонности топ-5 (категория А)



# RFM анализ

	CustomerID	Recency	Frequency	Monetary	R	F	M	rfm_score	RFM_Label	rfm_segment
0	12346.00	325	1	77183.60	1	1	5	7	1 1 5	Hibernating
1	12347.00	1	182	4310.00	5	5	5	15	5 5 5	Champions
2	12348.00	74	31	1797.24	2	3	4	9	2 3 4	At risk
3	12349.00	18	73	1757.55	4	4	4	12	4 4 4	Loyal customers
4	12350.00	309	17	334.40	1	2	2	5	1 2 2	Hibernating
...	...	...	...	...	...	...	...	...	...	...
4334	18280.00	277	10	180.60	1	1	1	3	1 1 1	Hibernating
4335	18281.00	180	7	80.82	1	1	1	3	1 1 1	Hibernating
4336	18282.00	7	12	178.05	5	1	1	7	5 1 1	New customers
4337	18283.00	3	721	2045.53	5	5	4	14	5 5 4	Champions
4338	18287.00	42	70	1837.28	3	4	4	11	3 4 4	Loyal customers

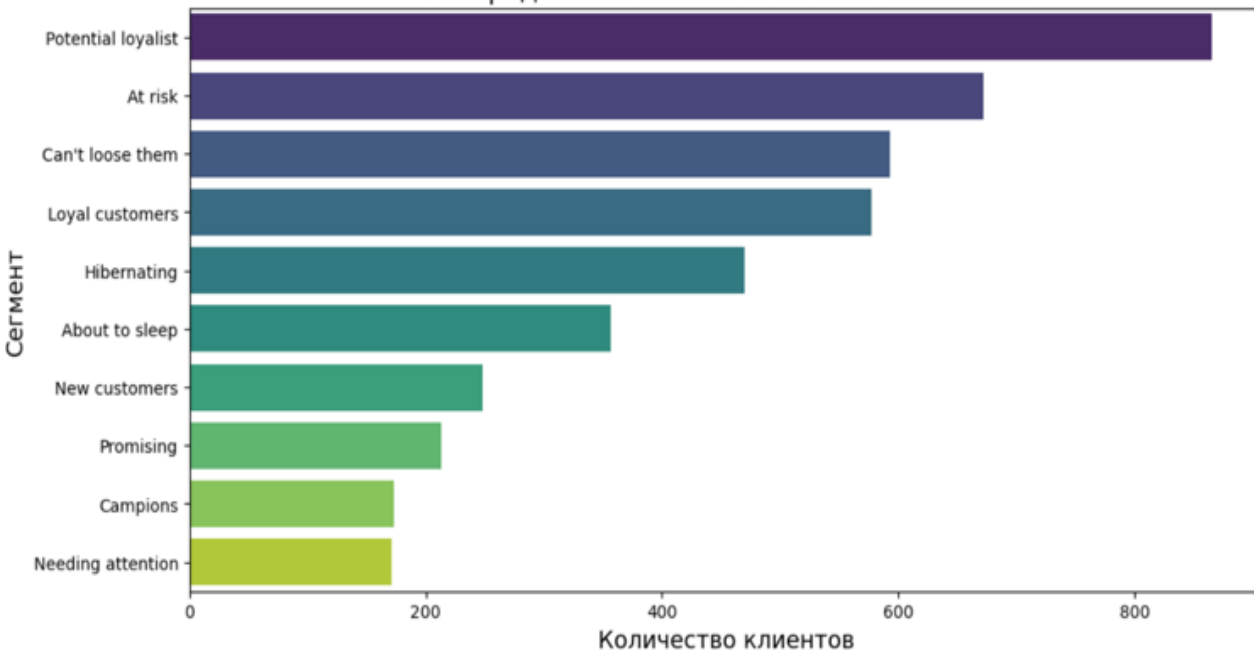
4339 rows × 10 columns

Разработана система меток которая позволила разбить клиентов на 10 ключевых групп:

- Champions
- Loyal Customers
- Potential Loyalist
- New Customers
- Promising
- Needing Attention
- About To Sleep
- At Risk
- Can't Loose Them
- Hibernating



Распределение клиентов по RFM-сегментам



Можно выделить:

**Доминирующие группы**

- Potential loyalist
- At risk
- Can't loose them

**Наиболее ценные сегменты**

- Champions
- Loyal customers

**Проблемные зоны**

- Hibernating
- About to sleep
- At risk



# Выводы и рекомендации



4

# Выводы и рекомендации

Проведенный комплексный анализ данных позволил выявить ключевые закономерности в продажах, клиентском поведении и эффективности товарного ассортимента.

## Географический анализ

- 84% выручки генерирует Великобритания
- 25 стран обеспечивают менее 0.1% дохода каждая

## Сезонность

- Пики продаж в ноябре (+36%), сентябре (+47%) и мае (+47%)
- Спад в январе (-25%) и апреле (-28%)

## ABC-анализ

- 20% товаров (категория A) дают 80% выручки

## RFM-анализ

- Выделено 10 клиентских сегментов
- 593 VIP-клиента ("Can't loose them") требуют особого внимания
- 865 перспективных клиентов ("Potential loyalist")

Данное исследование предоставляет аналитическую основу для принятия управленческих решений. Следующий шаг — тестирование рекомендаций на практике с последующим контролем метрик. Реализация предложенных мер позволит компании увеличить прибыльность и укрепить позиции на рынке.



**Спасибо за  
внимание!**

