

Дипломная работа

на тему: “Анализ продаж британского e-commerce (поиск инсайтов, составление рекомендаций стейкхолдерам, построение предиктивной модели объёмов продаж).”

Студент группы DAU -9:
Гетьман Александр Александрович

Описание задачи

- Посмотреть на качество данных
- Проанализировать имеющиеся данные по продажам в период с 12.2010 по 12.2011.
- Построить модель финансового состояния магазина
- Сделать выводы и написать рекомендации.

Метрики

- При построении модели финансового состояния магазина мы будем использовать модель линейно регрессии
- Для проверки данной модели будем использовать RMSE

Описание данных

Это транснациональный набор данных, который содержит все транзакции, произошедшие в период с 01.12.2010 по 09.12.2011 для британского и зарегистрированного интернет-магазина.

Компания в основном продает уникальные подарки на все случаи жизни. Многие клиенты компании являются оптовиками.

InvoiceNo	object
StockCode	object
Description	object
Quantity	float64
InvoiceDate	object
UnitPrice	float64
CustomerID	float64
Country	object

Описание обработки данных

В ходе проведения работ по обработке данных было проделано большое количество работ по приведению их в хорошее состояние для дальнейшего анализа.

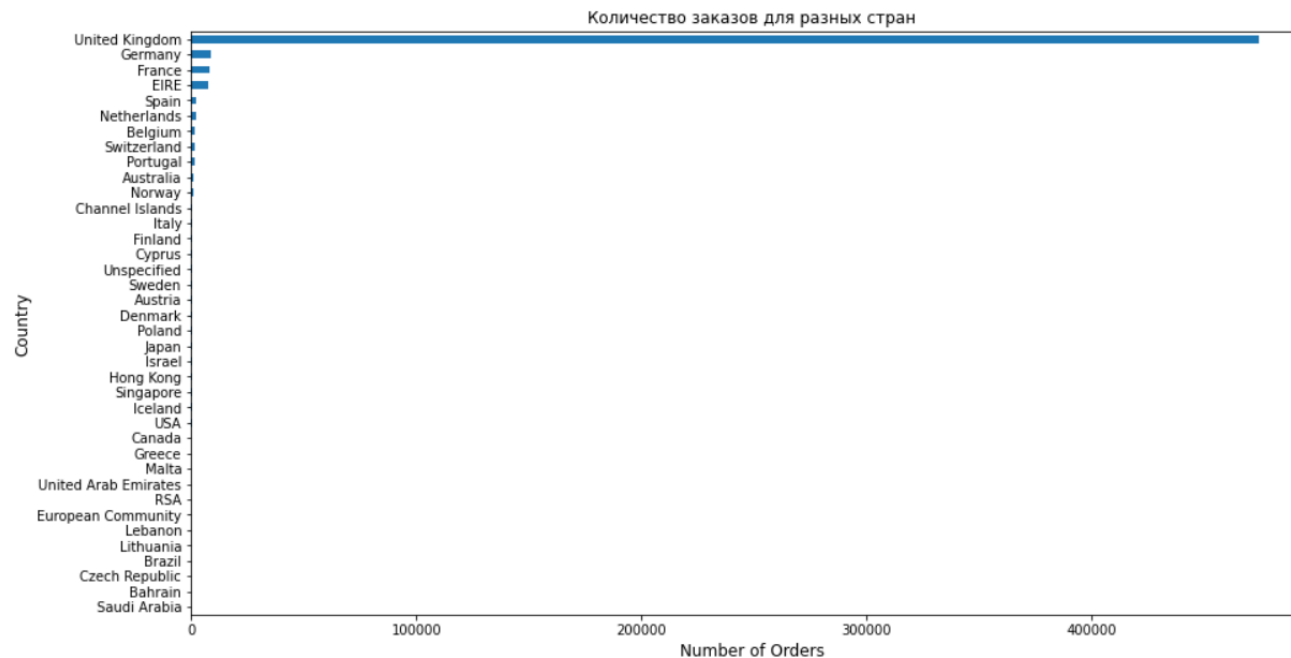
Так же были добавлены дополнительные ключ-значения.

#	Column	Non-Null Count		Dtype
--	-----	-----		-----
0	InvoiceNo	517908	non-null	object
1	StockCode	517908	non-null	object
2	Description	517908	non-null	object
3	Quantity	517908	non-null	float64
4	InvoiceDate	517908	non-null	datetime64[ns]
5	UnitPrice	517908	non-null	float64
6	CustomerID	517908	non-null	int64
7	Country	517908	non-null	object
8	Date	517908	non-null	datetime64[ns]
9	Month	517908	non-null	int64
10	Day	517908	non-null	int64
11	Year	517908	non-null	int64
12	Time	517908	non-null	int64
13	Amount_spent	517908	non-null	float64
14	Daymonth	517908	non-null	object
15	Monthyear	517908	non-null	object

Анализ данных

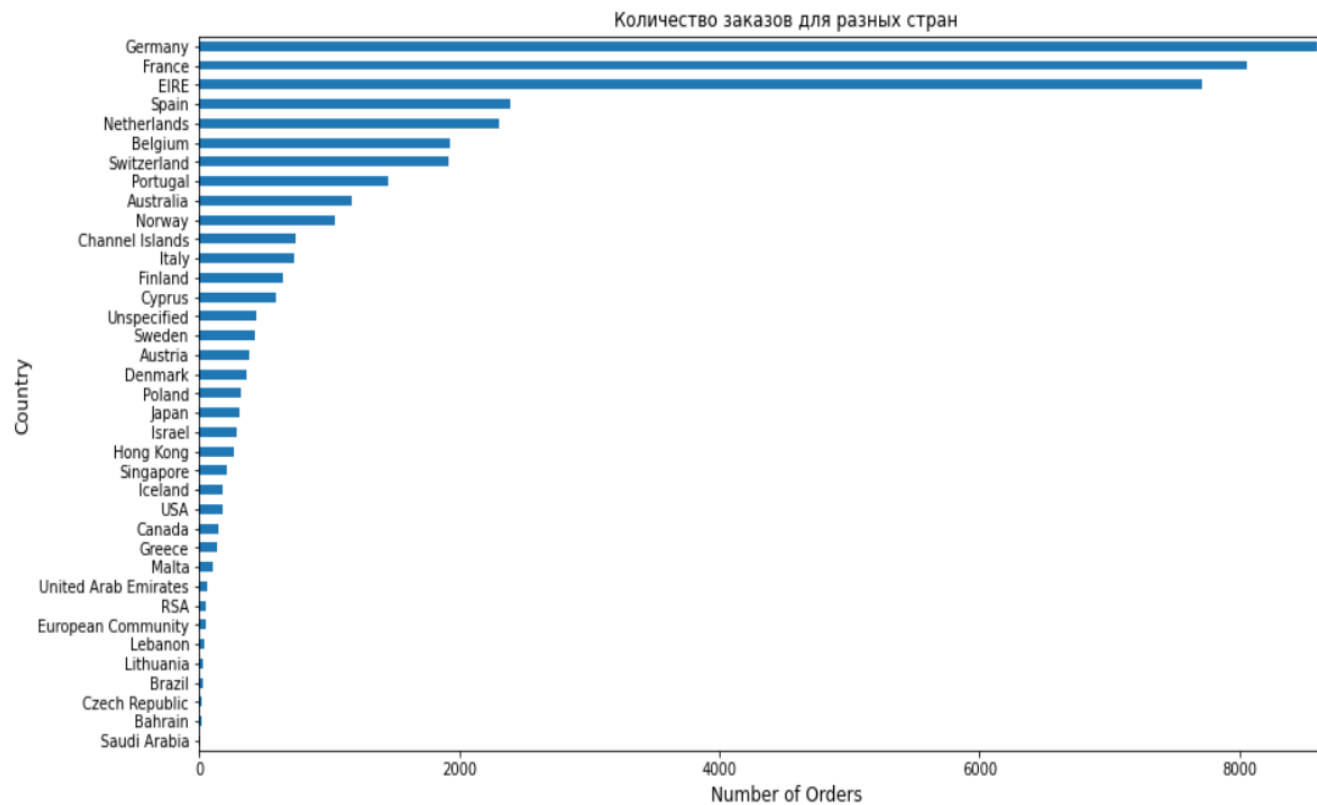
При построении графика показывающего: «Количество заказов по странам» мы видим что пик продаж приходится на Великобританию

Это не удивительно так как компания базируется в Британии



Анализ данных

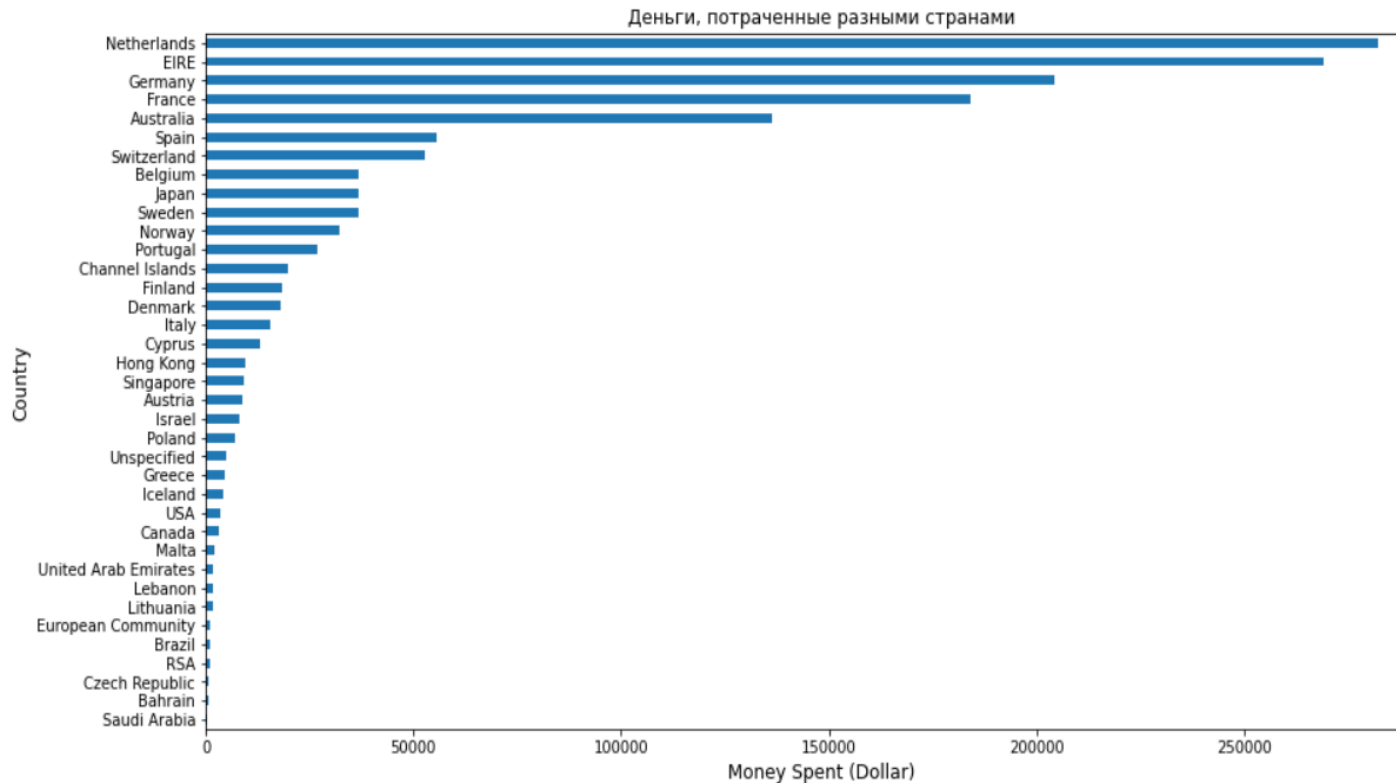
За исключением Соединённого Королевства, Германия, Франция и Ирландия являются тремя странами, где клиенты потратили больше всего денег



Анализ данных

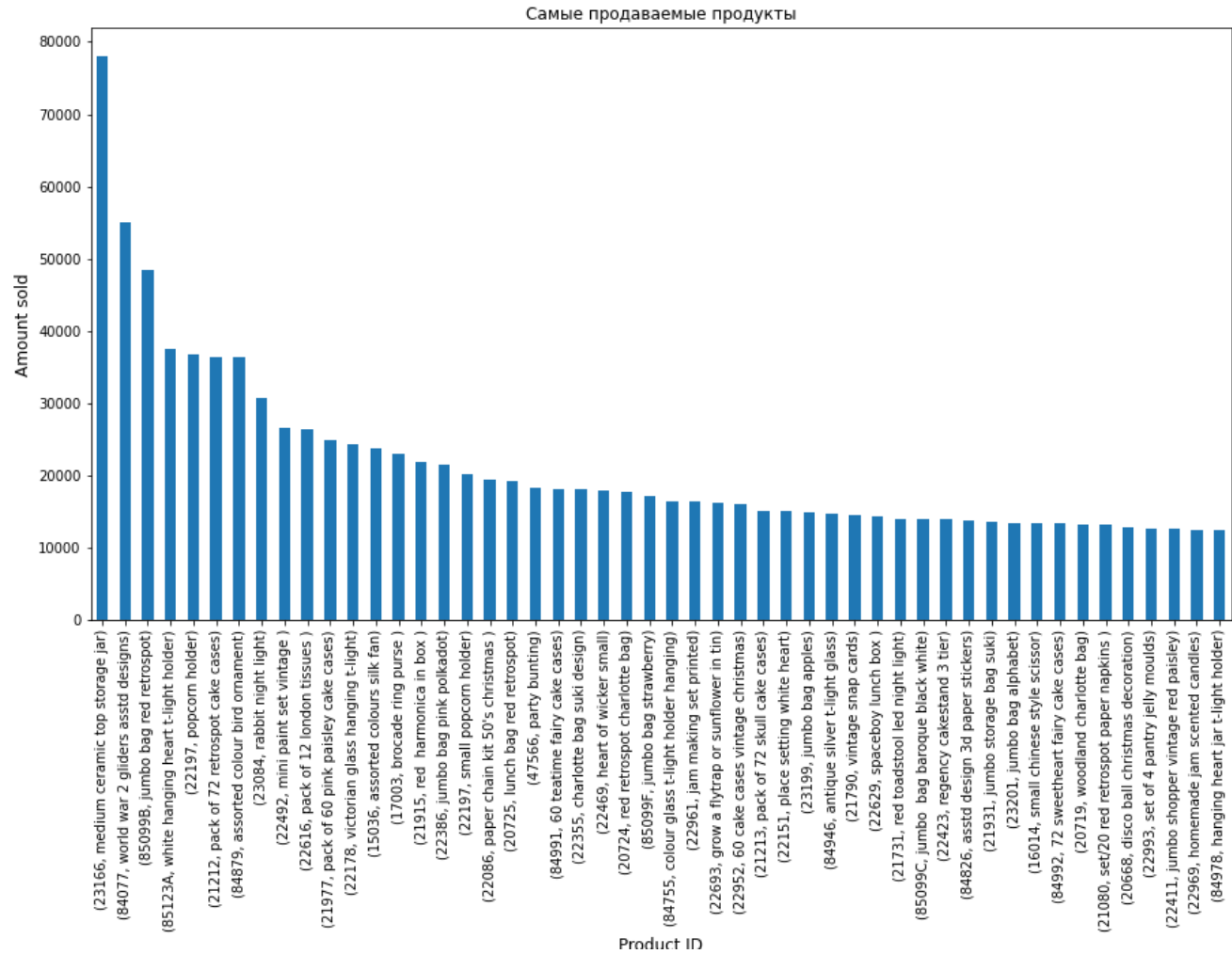
Так как компания базируется в Британии логично что при построении графика показывающего «Деньги, потраченные разными странами» она покажет самый большой показатель можно её сразу исключить из списка

За исключением Великобритании больше всего денег потратили клиенты из Нидерландов, Ирландии, Германии, Франции и Австралии.



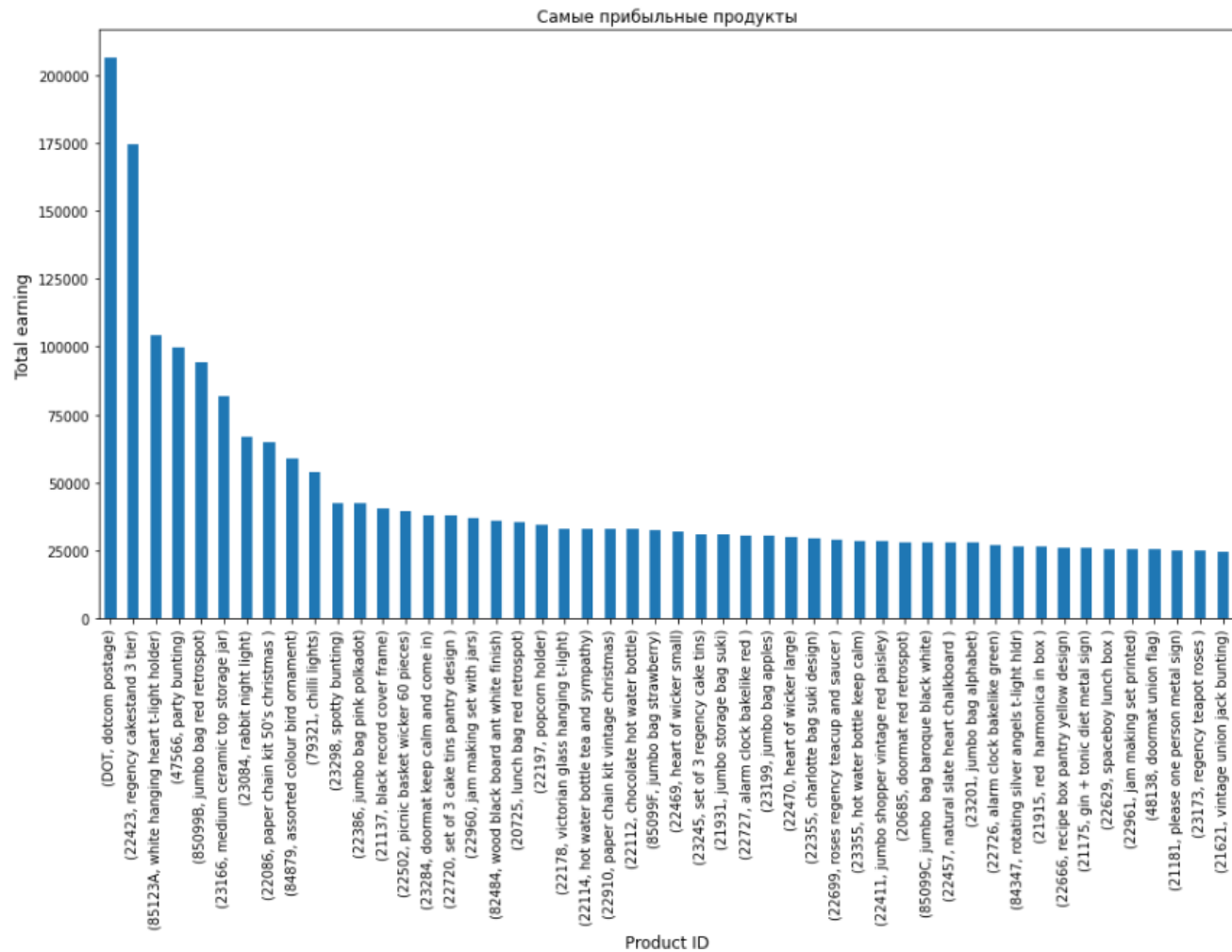
Анализ данных

Далее посмотрим какие продукты продавались чаще всего. Видим что самыми продаваемыми продуктами являются товары с индексами 23166, 84077, 85099B, 85123A и 22197.



Анализ данных

Самыми прибыльными для компании являются товары с индексами DOT, 22423, 85123A, 47566 и 85099B.



Построение модели финансового состояния магазина

После получения прогноза по количеству продаж будущего периода и склейки данных для сравнения показателей мы получаем следующие данные.

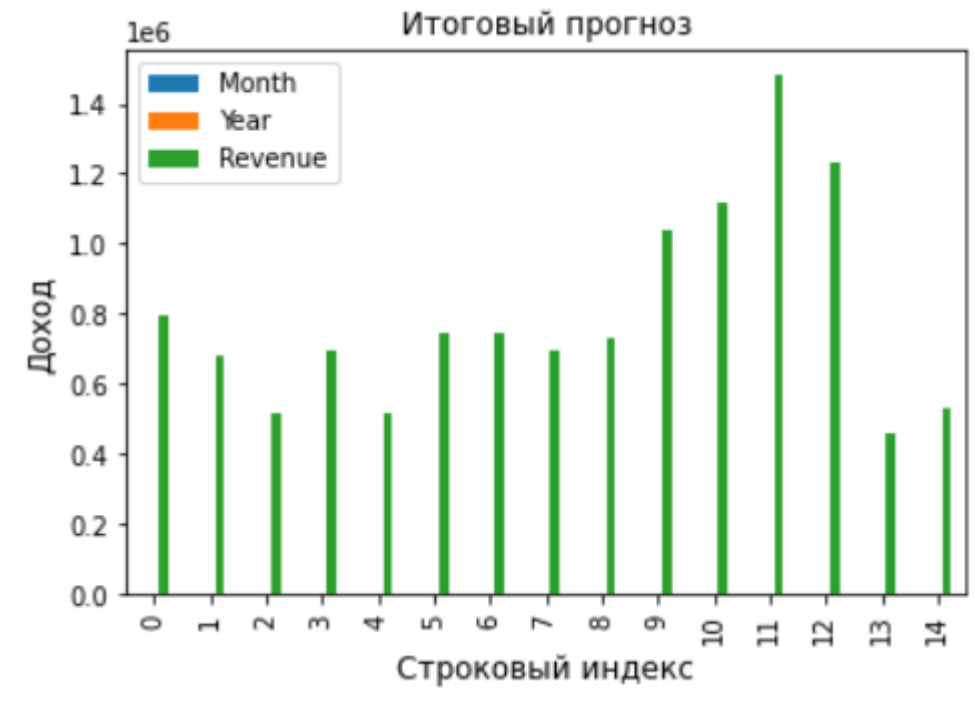
Построим простую визуализацию для упрощения восприятия.

	Month	Year	Revenue
0	12	2010	794182.6
1	1	2011	677874.5
2	2	2011	512461.3
3	3	2011	696294.2
4	4	2011	518098.9
5	5	2011	744369.1
6	6	2011	744467.4
7	7	2011	696335.0
8	8	2011	733891.3
9	9	2011	1037649.2
10	10	2011	1116611.7
11	11	2011	1482838.1
12	12	2011	1230293.4
13	1	2012	461012.8
14	2	2012	530947.4

Построение модели финансового состояния магазина

После построения визуализации мы наблюдаем резкий спад продаж в период с 01.2012 по 02.2012

Пик продаж приходится в период с 10.2011 по 12.2011. Ожидаемо что в преддверии Нового Года данные сервисы пользуются большой популярностью для поиска и покупок подарков. Так как есть трафик из-за рубежа предполагаем что клиенты закупаются подарки заранее чтоб успеть к празднику.



Выводы и рекомендации

- Делая вывод о ведении БД в данной компании. Качество данных не сильно хорошее.
- Данный бизнес неплохо масштабирован по странам но из-за того что там нет указанного города клиента то мы не можем посмотреть регионы где приобретен товар. Если бы было указано местное время клиента то можно было бы посмотреть на какие часы приходится пик продаж и стимулировать клиента на покупку в не пиковое время различными акциями.
- Для данной компании неплохо подошла бы воронка AARRR.
- Можно рассчитать время жизни клиента, сколько стоило его завлечь (при имеющихся данных о бюджете компании). Определить средний чек и в зависимости от того сколько он тратит посмотреть стоит ли нам увеличить стоимость товара или дальше работать на кол-во.
- По итогам получившегося прогноза рекомендую сделать клиентские акции для увеличения конверсии по продажам и предотвращения возможного спада в продажах.