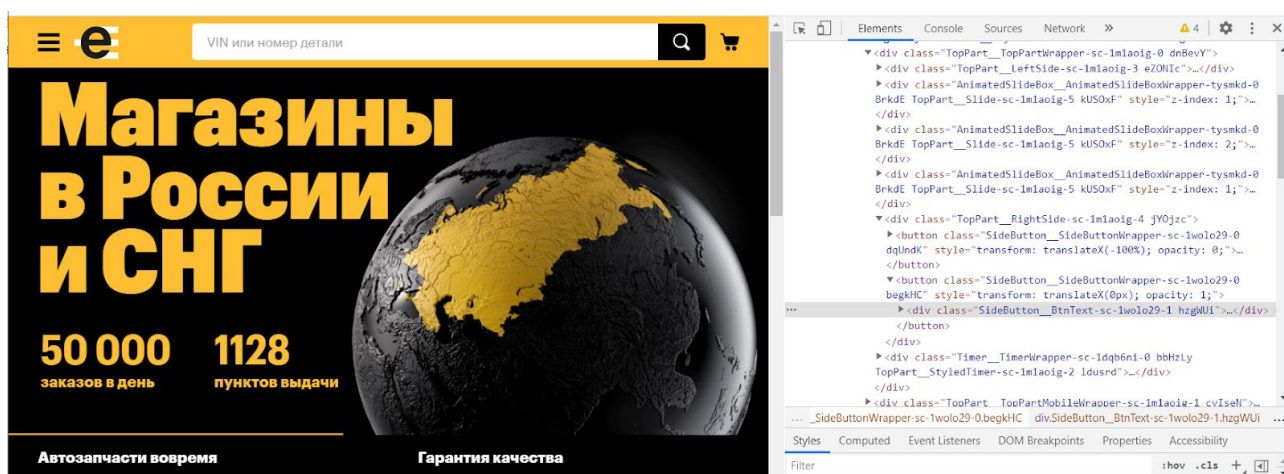


## Маркетинг: парсим скриптовые сайты - решаем задачи бизнеса.

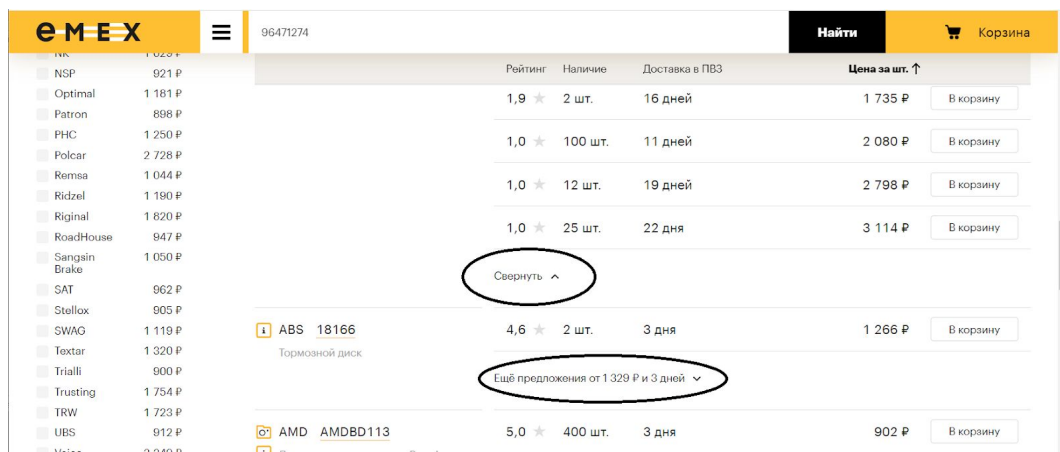
Например, я торгую автозапчастями. Товар не эксклюзивный, представлен многими продавцами. Если я знаю ситуацию на рынке, то имею все основания повысить свою маржинальность, а также решать важные задачи по тактике продаж. Если не знаю, то всегда проигрываю: занижаю цену — снижаю маржинальность, завышаю цену — снижаю объем продаж.

Нужную мне информацию я планирую взять у агрегаторов запчастей, которые продвигают товар независимых продавцов. Сайтом emex.ru, например, пользуются более 700 поставщиков. Для агрегатора выгодно, если продавцы и покупатели не видят картину в целом и в конкурентной борьбе между собой увеличивают ценность агрегатора. Сайты emex.ru и exist.ru — скриптовые. Данные в них подгружаются по мере необходимости из внутренних баз данных. Если парсить традиционный сайт, то данные отображаемые на web-странице можно увидеть и в исходном html коде. В скриптовых сайтах это не так — в коде web-страницы нет никаких явных данных. И дело не в том, что они отсутствуют, а в том, что они еще не подгрузились.



Для того чтобы получить данные в коде, надо иметь инструментарий, позволяющий открыть виртуальную сессию и делать паузы для подгрузки данных.

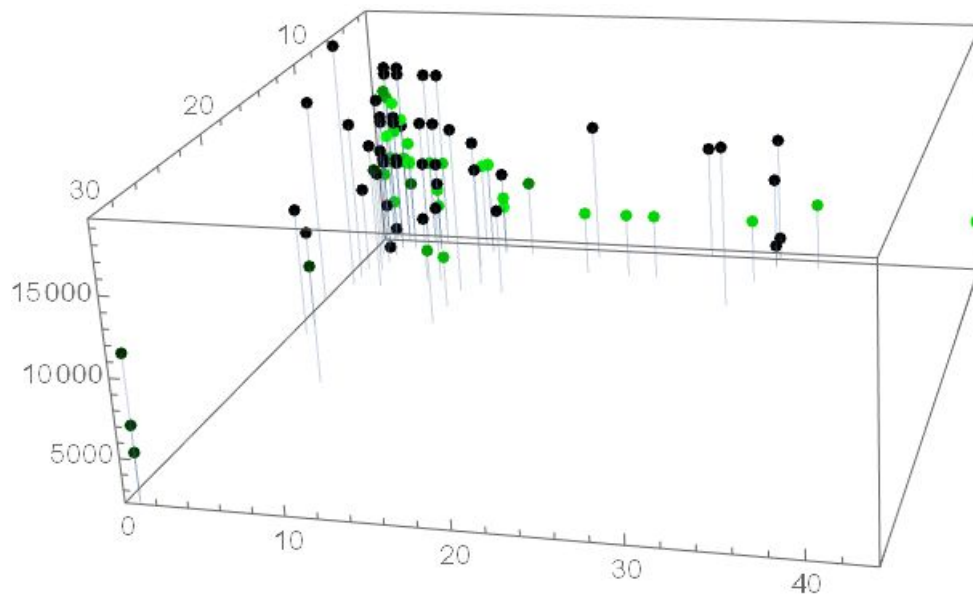
Это не исчерпывает перечень проблем. Например надо уметь раскрывать вложенные списки. В противном случае много данных будет потеряно (не будут отображены в коде).



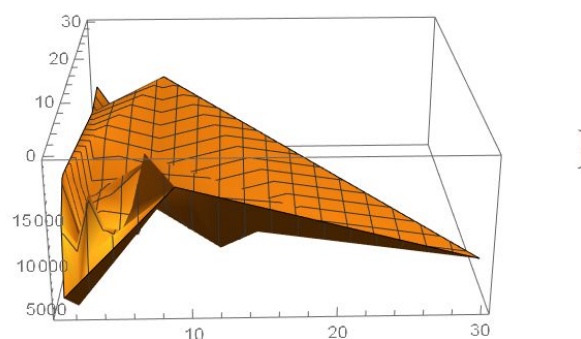
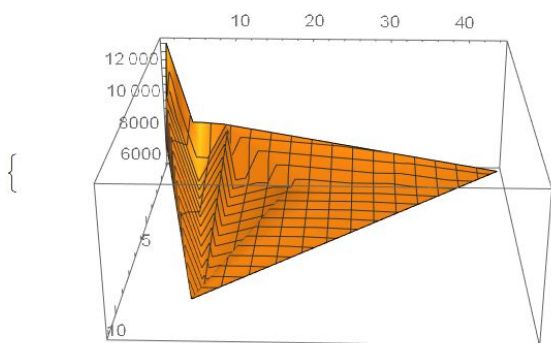
Хорошо, пусть я преодолел перечисленные трудности и сумел получить желаемые данные. Что это даст для бизнеса?

Первая проблема при работе с полученными данными связана с непонятностью в применении статистических методов к выставленным продавцами ценам.

Пример: на оси z (значения от 0 до 20000 рублей) — цена, на оси x (значения от 0 до 40 дней) — время ожидания доставки, на оси y (значения от 0 до 30 штук) — количество товара на складе.



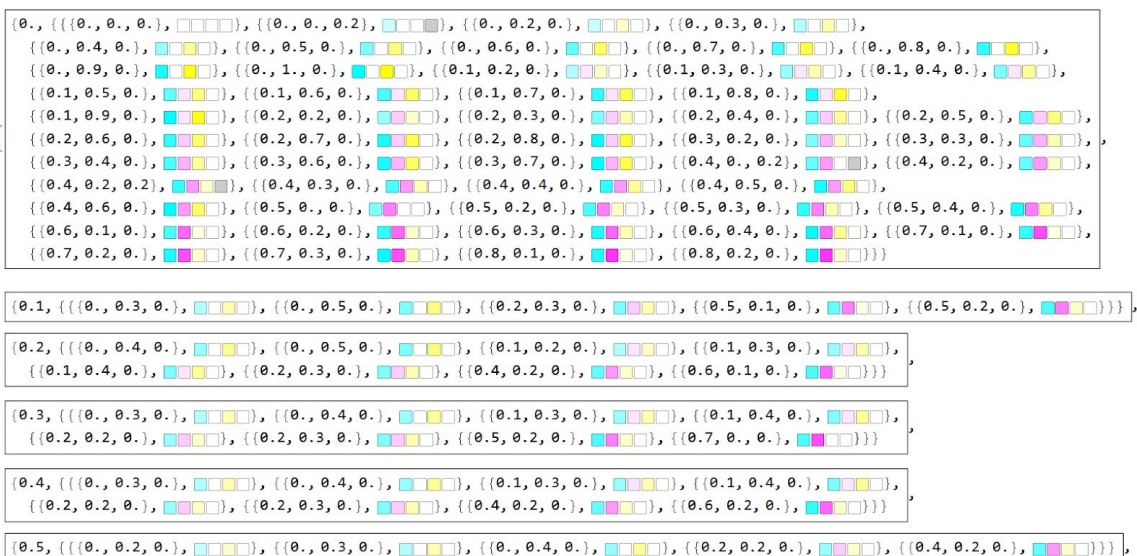
Тот же пример, но на первом рисунке присутствуют все предложения, а на втором — только предложения продавцов с высоким рейтингом. На втором рисунке уже прослеживаются разумные тренды — чем дольше ждешь — тем цена ниже.



Вывод простой. Некоторые продавцы «вводят в заблуждение» и если не понять как, то всегда будет иметь место занижение вами цены.

Для этого можно использовать данные по рейтингам продавцов за 90 дней.

Данные по рейтингам продавцов можно обобщить и визуализировать. Таблица рейтинга для диапазона от 0 до 0.5 (весь диапазон от 0 до 1).



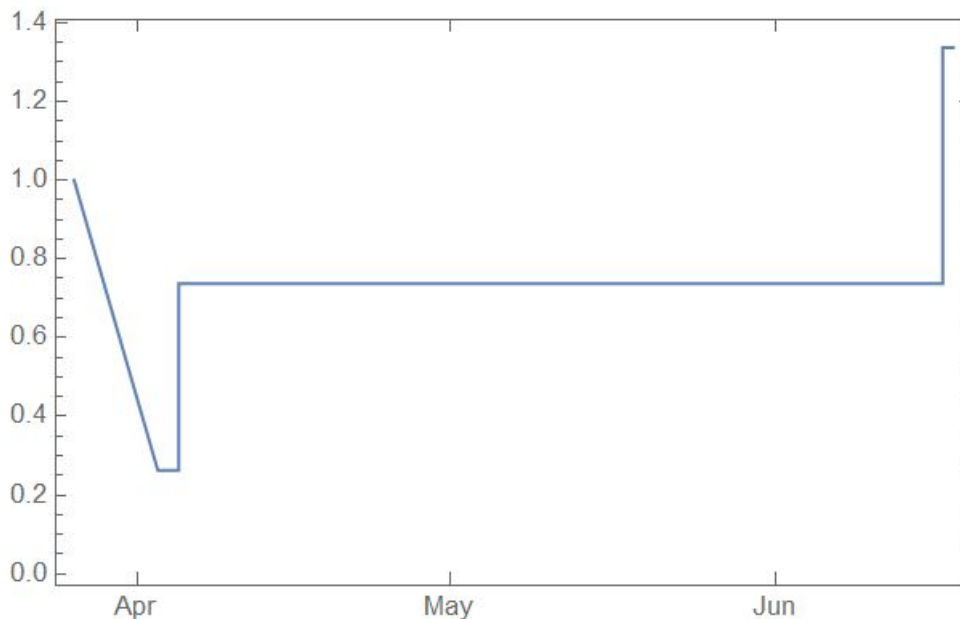
Первое число в группе — суммарный рейтинг. При этом он складывается из значений триплексов. Первое число в триплексе — доля отказов от продажи, второе число — доля опозданий и третье — доля предложений о перемене одной запчастей на другую. Из рисунка видно, что часть продавцов играет на более дешевых предложениях (запчастей по такой цене у них нет), а часть на занижении сроков доставки (запчасть имеется, но логистически очень далеко).

Хорошо, поняли на какие данные надо ориентироваться. Тогда можно начать решать задачи ценообразования.

5E0821105A			5E0821105A		
Day: Sat 4 Apr 2020			Day: Thu 18 Jun 2020		
(Крыло перед. левое 5E0821105A Octavia (13-))			(Крыло перед. левое 5E0821105A Octavia (13-))		
<b>AP</b>					
3	SD26101600L00 analog 1	3793 (4.5, OYJA)	2	SD26101600L00 analog 5	3647 (4.5, ATOS)
4	SD26101600L00 analog 7	3923 (4.4, ATOS)	4	SD26101600L00 analog 15	3839 (4.4, KOLC)
<b>O.E.M.</b>					
4	001282241002052017 analog 35	2848 (4.4, PNGI)	6	SD26101600L00 analog 20	3282 (3.9, VKEL)
<b>Polcar</b>					
3	69C101 analog 1	5428 (3.6, POLC)	20	SD26101600L00 analog 20	3971 (4.2, AVUG)
<b>SAT</b>					
2	STS0270162 analog 3	2994 (4.2, ASAA)	<b>Gordon</b>		
<b>TYG</b>					
2	SD100144L analog 10	3879 (4.8, MASJ)	4	GD5223CL analog 10	3839 (4.4, KOLC)
	SD100144L analog 10	3752 (4.7, MAUJ)	6	GD5223CL analog 20	3282 (3.9, VKEL)
3	SD100144L analog 1	4420 (3.8, GORR)	<b>NTP</b>		
	SD100144L analog 1	4297 (3.9, GORC)	4	SK124021FL analog 19	3839 (4.4, KOLC)
4	SD100144L analog 3	5833 (4.6, UFEW)	<b>O.E.M.</b>		
<b>SAT</b>					
			5	001282241002052017 analog 54	3022 (4.9, NRAS)
2	STS0270162CN analog 9	3295 (4.2, ROAP)	<b>TYG</b>		
	STS0270162CN analog 9	2714 (4.2, ASAA)	2	SD100144L analog 6	3333 (4.4, AMTL)
	STS0270162 analog 6	3755 (4.2, ROAP)		SD100144L analog 10	3215 (4.8, MASJ)
	STS0270162 analog 6	3092 (4.2, ASAA)		SD100144L analog 10	3110 (4.7, MAUJ)
12	STS0270162 analog 2	15076 (5., TLEE)	4	SD100144L analog 10	3839 (4.4, KOLC)
	STS0270162CN analog 4	6015 (4.4, DVZV)	<b>Yih Sheng</b>		
16	STS0270162 analog 2	5660 (4.4, DVZV)	4	SD2402AL analog 20	3839 (4.4, KOLC)
<b>TYG</b>					
			6	SD2402AL analog 20	3282 (3.9, VKEL)

Первое число в строке в рамке — декларируемое количество запчастей на складе продавца. Красное число — цена. Последний дуплекс — рейтинг и код продавца.

В тактике продаж важно понимать динамику рынка. По имеющимся данным расчетным методом можно посчитать только относительную динамику в дамках скользящей 90-дневной статистики единиц продаж.



Логика простая — ваши продажи не должны сильно отличаться от обобщенных показателей (в штуках) рынка.

Анализ ценообразования продавцов показывает, что пока в тактике продаж они слепо следуют за курсом доллара. При этом важно в тактику продаж ввести режим «продавать/не продавать», если себестоимость запчастей не соответствует курсу доллара. Массовое запаздывание по срокам доставки, кроме случаев специального ввода в заблуждение, может быть связано с недостаточной проработкой логистической составляющей: где хранить, в каком количестве и какой вес/объем.

Если встать на путь систематического улучшения бизнеса, то можно к парсингу основных агрегаторов добавить статистику поисковых запросов, собираемую браузерами.