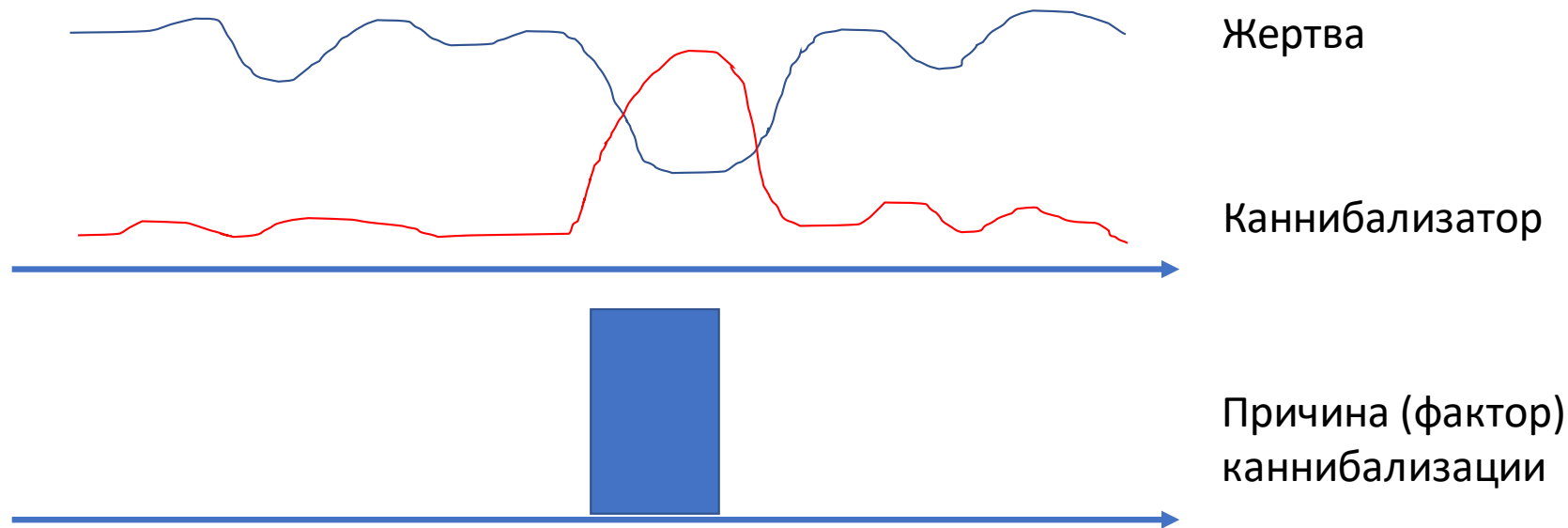


# План дальнейших работ

Д. Звежинский, 16.01.2019

[Dmitry.Zvezhinsky@sas.com](mailto:Dmitry.Zvezhinsky@sas.com)

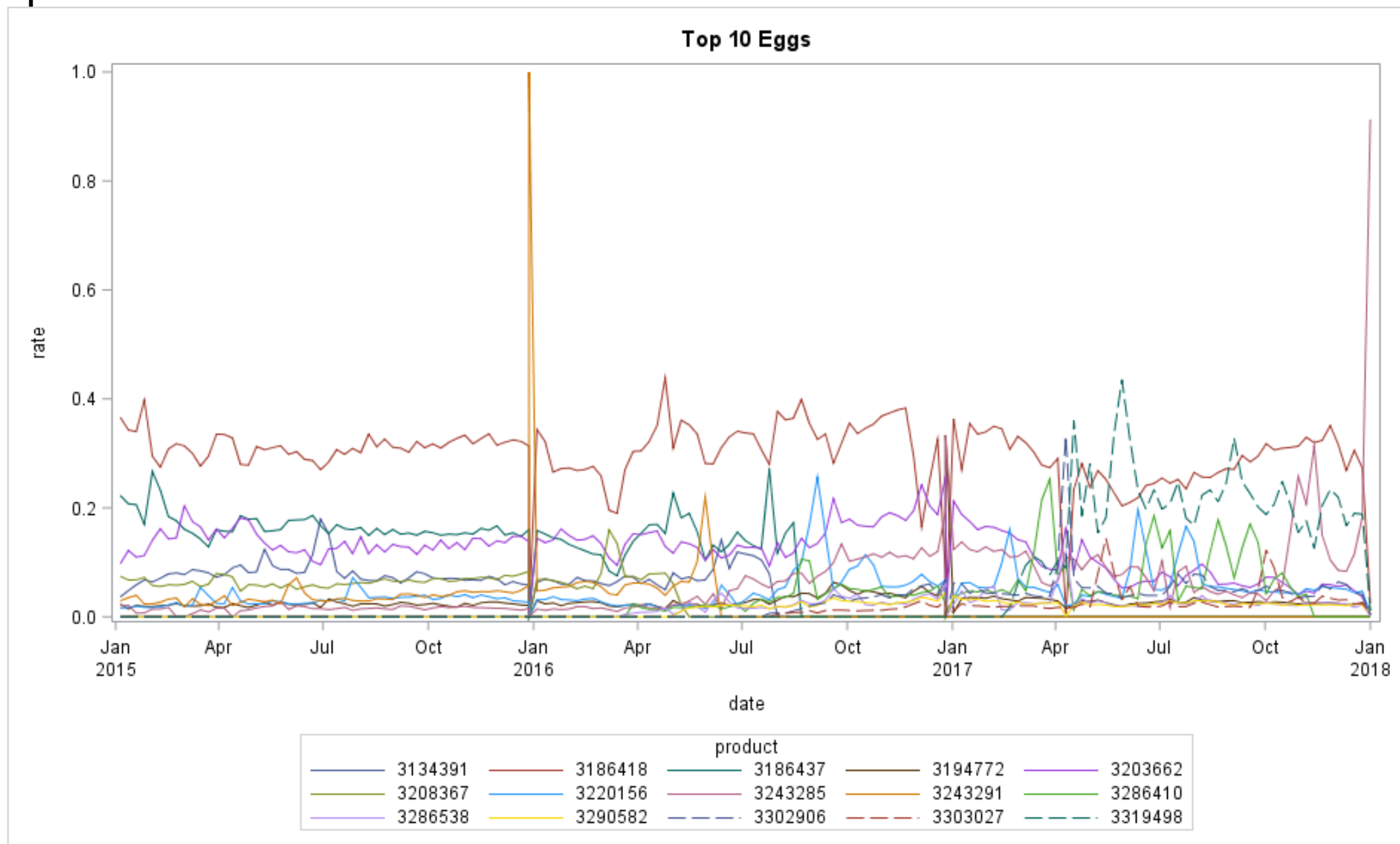
# Общий план решения задачи



Для корректного учета эффекта каннибализации нужно:

- 1) Выделить группу рядов (признак товарной группы), внутри которой ожидается эффект.
- 2) Определить факторы и правила, позволяющие обнаружить момент времени, где возникает эффект (в том числе, в будущем). Для каннибализации должны быть причины!
- 3) Вычислить размер эффекта (относительный или абсолютный). На сколько единиц изменились продажи жертвы и каннибализатора?
- 4) Вычислить базовый прогноз (где каждый временной ряд знает только про себя). Основная масса временных рядов.
- 5) На основании 2,3,4 сделать прогноз с учетом взаимного влияния товаров.
- 6) Оценить значимость результатов.

# Данные «Утконос»



# Обратите внимание на особенности данных

- Главн. таблица: Items, доступность: Storehouse\_available
- Нет цен. Можно поделить обороты на продажи.
- Нет остатков, но есть доступность. Доступность < 100 может означать дефицит.
- Есть два флага промо, редко бьющихся друг с другом:

Frequency Percent Row Pct Col Pct	Table of promo_sap by promo_site			
	promo_sap	promo_site		
			1	Total
0		17927	4966	22893
		73.28	20.30	93.58
		78.31	21.69	
		92.15	99.12	
1		1527	44	1571
		6.24	0.18	6.42
		97.20	2.80	
		7.85	0.88	
Total		19454	5010	24464
		79.52	20.48	100.00

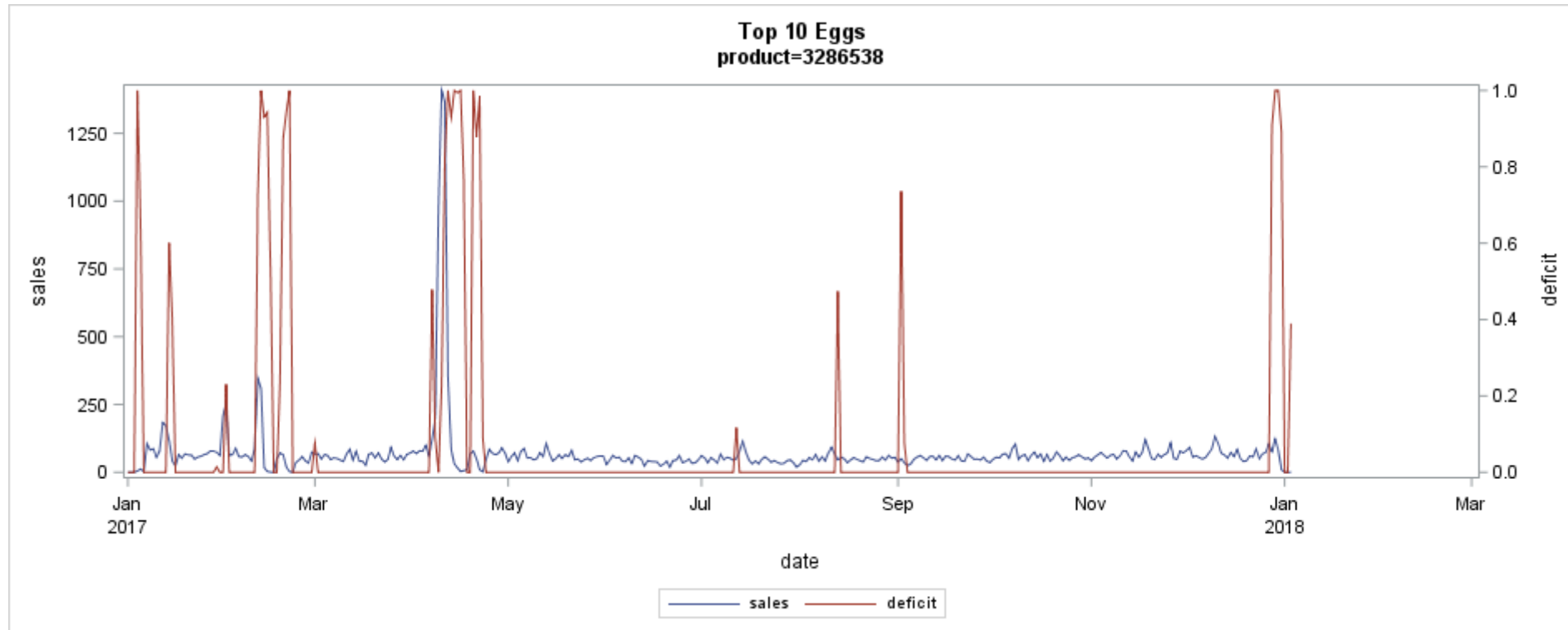
# Обратите внимание на особенности данных

- Есть пропущенные интервалы по датам – заполнить!
- Есть дубликаты по датам – проагрегировать!

3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	0				0	2015-07-07	12	1142.4	100
3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	1	Регулярное промо			0	2015-07-08	2	170	100
3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	0				0	2015-07-08	5	476	100
3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	1	Регулярное промо			0	2015-07-09	25	2125	100
3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	0				0	2015-07-09	2	190.4	100
3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	1	Регулярное промо			0	2015-07-10	27	2295	100
3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	1	Регулярное промо			0	2015-07-11	38	3230	100
3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	1	Регулярное промо			0	2015-07-12	24	2040	100
3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	1	Регулярное промо			0	2015-07-13	20	1700	100
3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	1	Регулярное промо			0	2015-07-14	28	2380	100
3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	1	Регулярное промо			0	2015-07-15	8	680	100
3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	0				0	2015-07-15	4	356.74	100
3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	1	Регулярное промо			0	2015-07-16	2	170	100
3134366	Яйцо Экстра ку р. С1 кор. дес	Яйцо	001009	Яйцо	001009005	Яйцо первой категории	0				0	2015-07-16	22	2084.88	100

# Обратите внимание на особенности данных

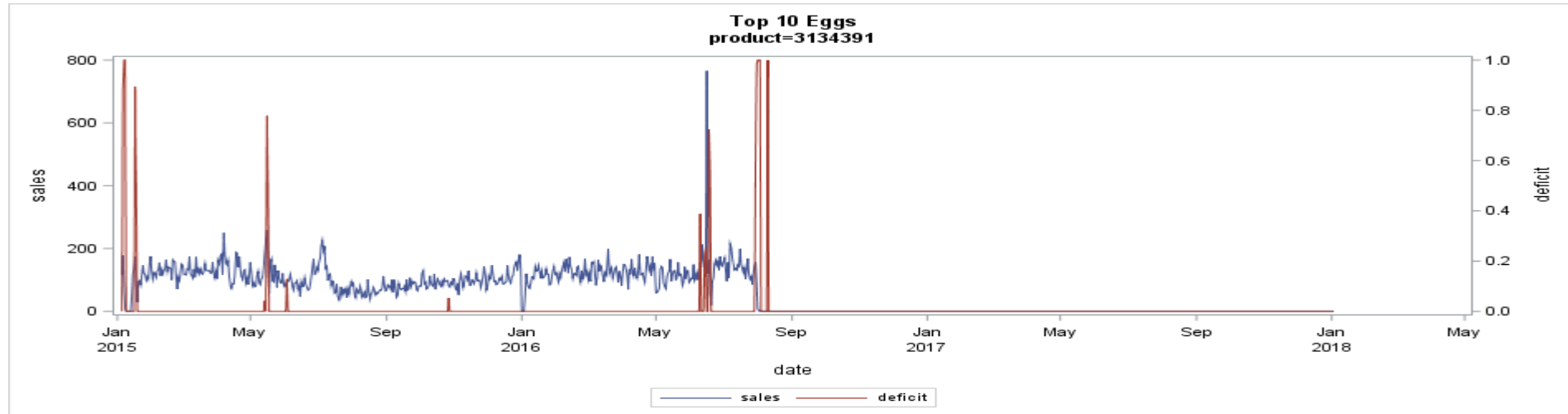
Восстановление спроса, очевидно, понадобится:



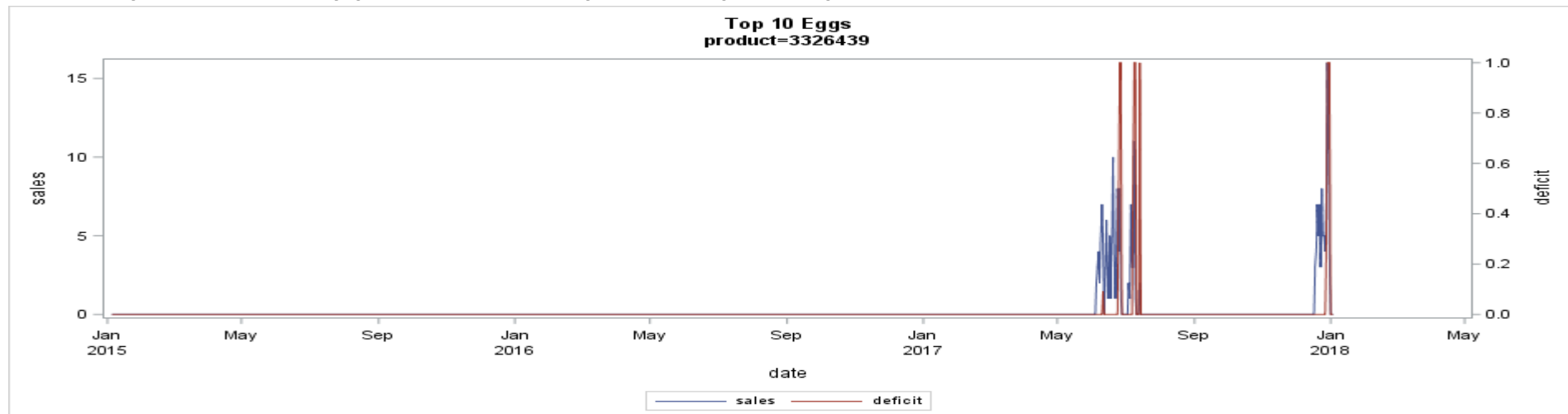
Примеры в категории яйца - см. пасха

# Обратите внимание на особенности данных

Есть много выведенных товаров:



При этом нужно аккуратно выбрать критерий:



# Основные этапы решения задачи

	Оценка
• Анализ, подготовка исходных данных	
• Исследование данных – какой ключ у исходных данных? Есть ли пропущенные интервалы? Какие независимые переменные есть в данных?	5/10
• Подготовка временных рядов – заполнение пропущенных интервалов, для каждого разреза «товар-магазин-день или неделя» должна быть только одна строка (значение продаж)	6/10
• Восстановление спроса – выделение дефицита, заполнение продаж в дефицитные дни на основании недефицитных.	7/10
• Визуализация случаев каннибализации – нужно придумать правила, которые помогают автоматически выделить случаи каннибализации для пары рядов внутри группы товаров или рядов. Правила опираются на доступные независимые переменные (цены, промо, погода итд)	8-9/10
• Реализация правил разметки случаев каннибализации в SAS	10/10

Илья Паузнер



# Основные этапы решения задачи

	Оценка
• Моделирование эффекта каннибализации	
• Собрать обучающую выборку для случаев каннибализации.	5/10
• Feature engineering – анализ и создание признаков. Какие признаки могут описать величину эффекта каннибализации (у каннибализатора и у жертвы)? Какие признаки являются значимыми?	6-7/10
• Построение моделей на hold-out и выбор лучшей.	8/10
• Реализация stacking (на вход приходит базовый прогноз временного ряда, разметка будущих случаев каннибализации и другие признаки).	9/10
• Реализация поправки для случаев каннибализации в SAS	10/10

Наталья Бернштейн

# Основные этапы решения задачи

	Оценка
• Базовая модель прогнозирования спроса	
• Обзор существующих инструментов автоматического прогнозирования иерархических ВР.	5/10
• Получение наивного (прогноз=предыдущему значению) и базового (без учета каннибализации, только по своей истории) прогнозов.	6/10
• Реализация расчета поправки на каннибализацию для всех товаров (нужны правила разметки случаев каннибализации)	7/10
• Расчет прогноза с учетом каннибализации как поправки к базовому прогнозу	8/10
• Реализация расчета базового прогноза на SAS и учёт в нём каннибализации	9-10/10

Мария Ушерович

# Основные этапы решения задачи

- |  |        |
|--|--------|
| • Оценка значимости эффекта каннибализации для данных «Утконоса». Измерение качества прогноза, оценка статистической значимости результатов.                     | Оценка |
| • Получение оценок точности для базового и наивного прогнозов спроса используя метрику MAE (горизонты 1,2,3 недели – категория товаров УИ2). Уточнение методики. | 5/10   |
| • Анализ ошибки. На каких разрезах базовый прогноз (без учёта каннибализации) лучше всего / хуже всего? Почему?  | 6/10   |
| • Получение оценок точности для прогноза с учетом каннибализации.  | 7/10   |
| • Статистические тесты, сравнивающие качество прогнозов: Диболд-Мариано, Harvey-Leybourne-Newbold  | 8-9/10 |
| • Реализация расчета точности и стат. тестов в SAS   | 10/10  |

Линара Кармысова

# Основные вводные:

- Данные агрегируются по неделям.
- Строим прогноз еженедельного спроса, а не продаж.
- Каннибализацию ищем внутри группы УИ2 (Товарная иерархия (уровень 2))
- Проверка точности прогноза – на четырёх неделях (20.11.2017 – 17.12.2017), метрика = MAE.