

Эффект каннибализации

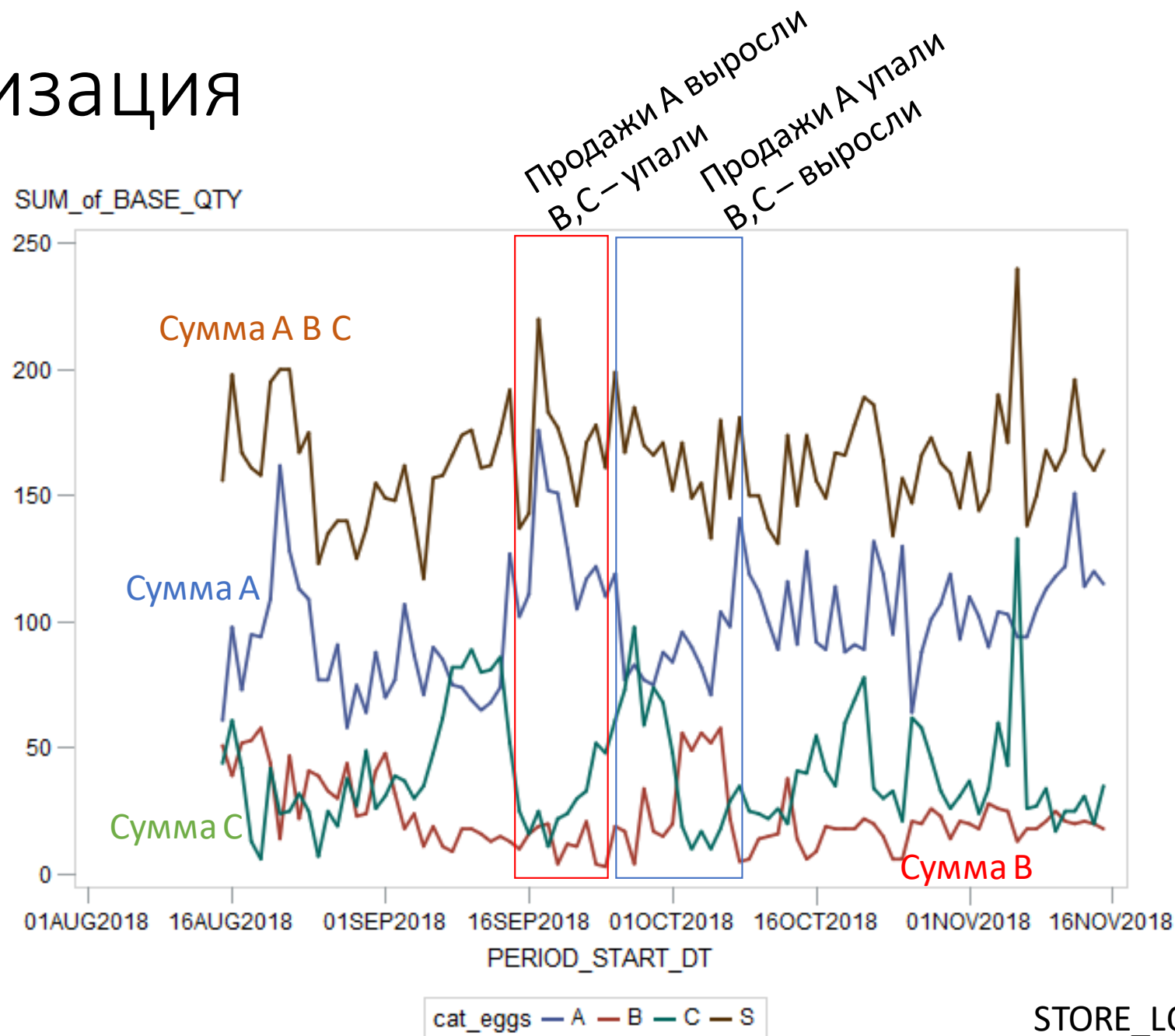
Д. Звежинский, 28.11.2018

Dmitry.Zvezhinsky@sas.com

Пример эффекта

- В food-ритейле пример эффекта каннибализации можно обнаружить в категории «яйца».
- Почему это именно «каннибализация»? Потому что он вызван нестационарностью ассортимента.
- Драйвером эффекта может служить промо-акция на отдельные позиции.
- Или – дефицит на популярные позиции (на первый взгляд, неочевидно).
- Для визуализации просуммировали продажи по дням яиц разных категорий (А, В, С)
- Яйца категории А – самые популярные (самые высокие обороты в штуках), С – наименее популярные (самые низкие обороты).

Визуализация

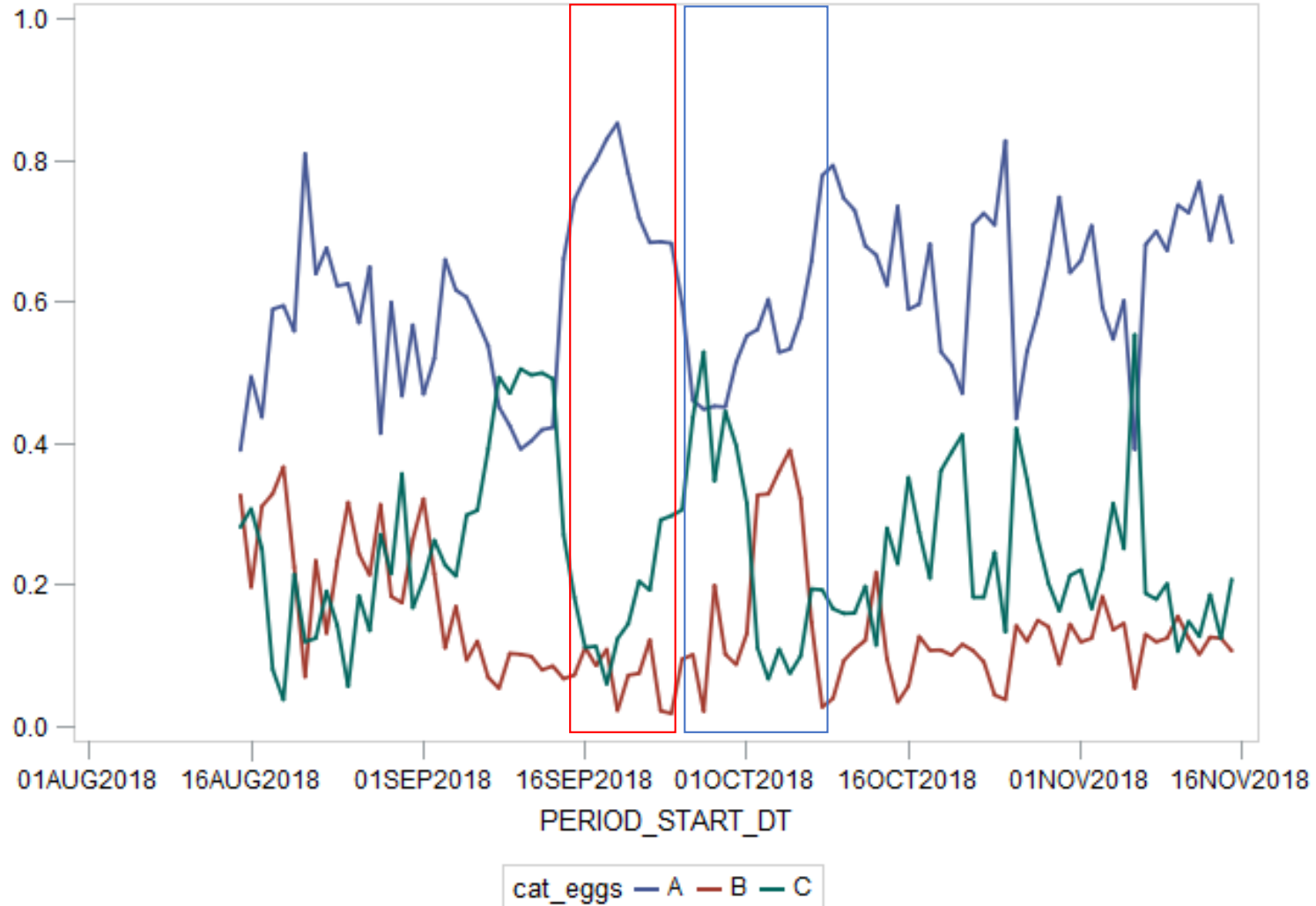


STORE_LOCATION_ID = 836426

Визуализация

Доля продаж

base_rate



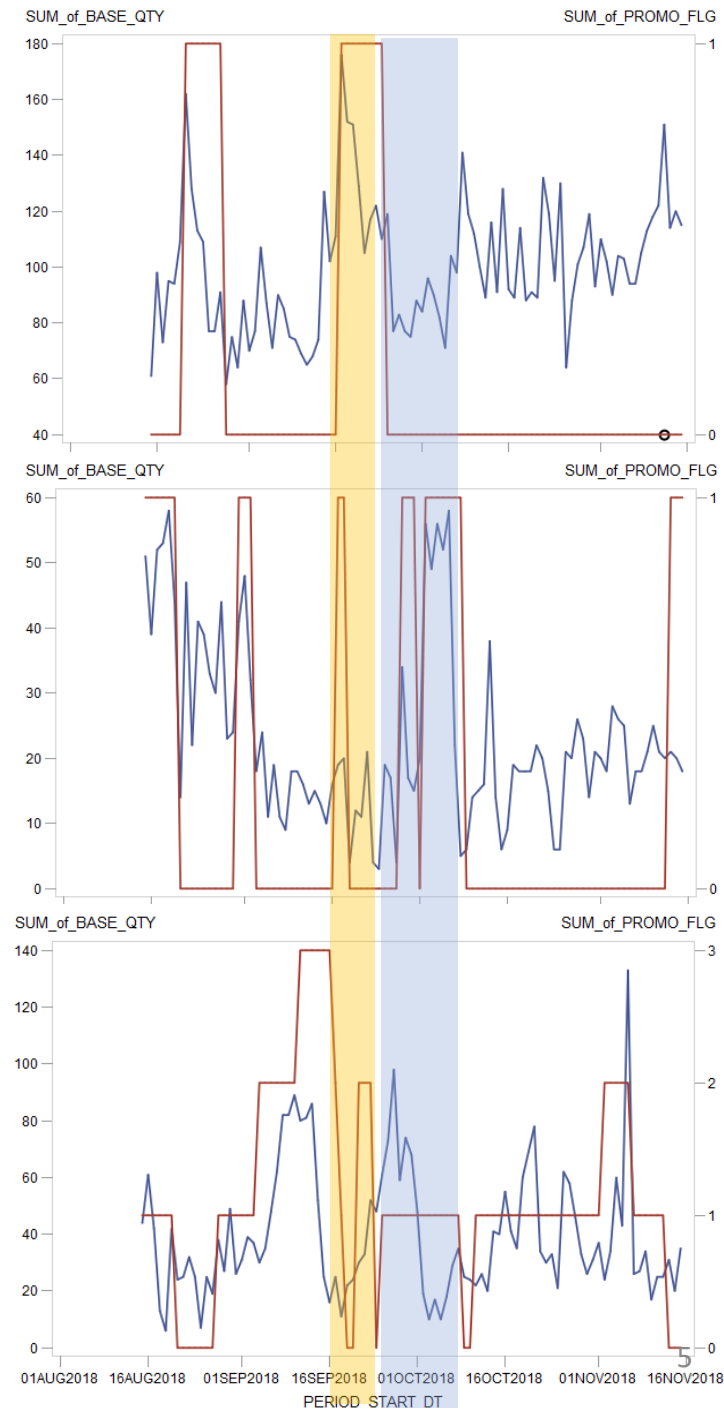
Причины

- 1) Промо на кат. А (самые популярные позиции) в первом диапазоне.
- 2) Промо на кат В, С во втором
- 3) Мы характеризуем «промо» на категории самым простым способом – агрегатом флагов наличия ценового промо, но не видим, собственно, цену или скидку.

Сумма А

Сумма В

Сумма С



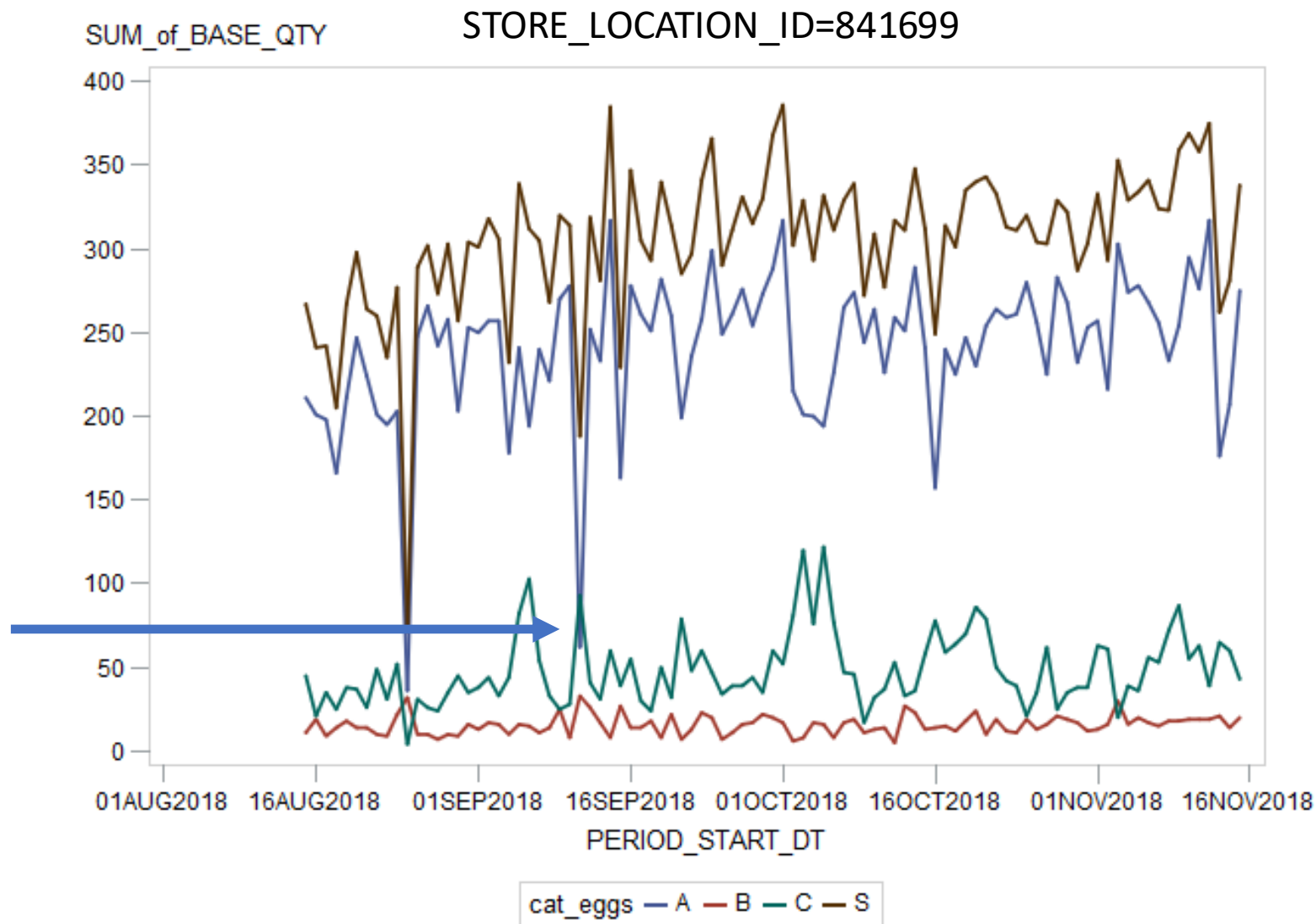
Сумма флагов наличия промо (по данной категории)

Дефицит как драйвер каннибализации

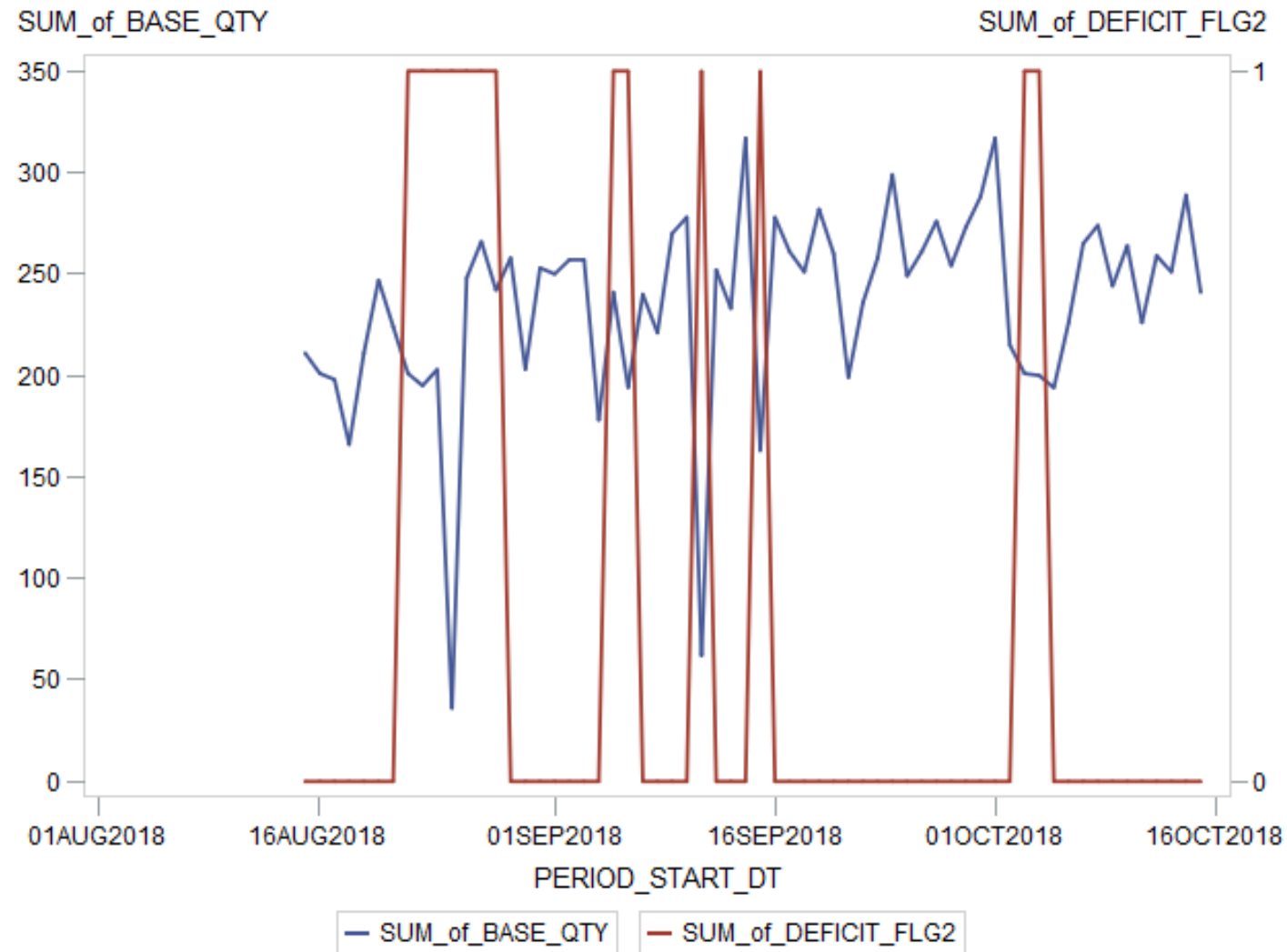
Очень сильное падение продаж в категории А приводит к пику в продажах С и В.

На следующем слайде посмотрим на агрегат флага дефицита, который считается в разрезе товар-магазин-день.

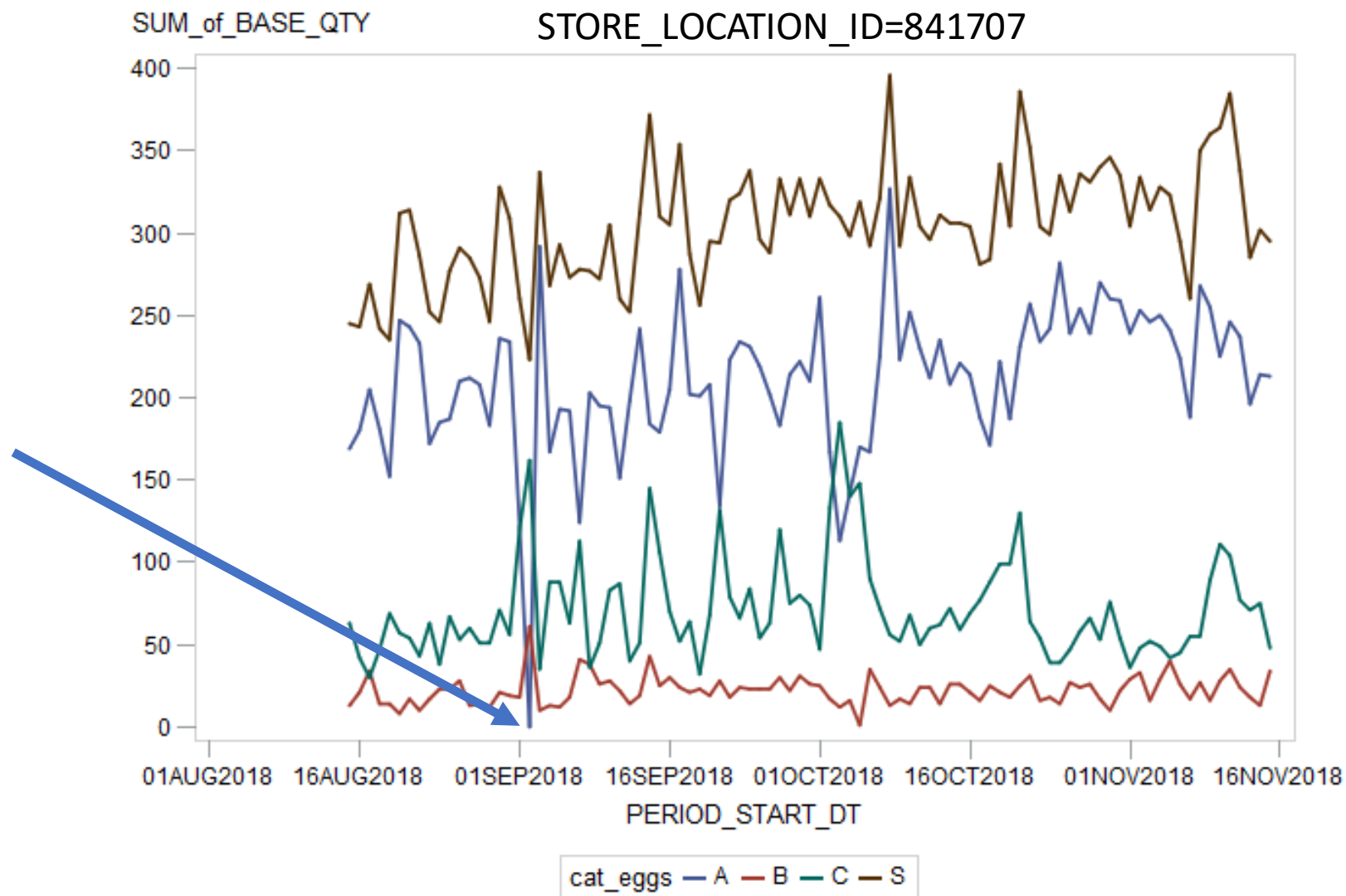
Флаг дефицита определяется по ненулевым остаткам, а также сильным отклонениям продаж в конкретный день от скользящего среднего продаж по дням, где были ненулевые остатки.



Дефицит как драйвер каннибализации



Дефицит как драйвер каннибализации



План решения задачи

- Выделение групп товаров, которые взаимно каннибализируют друг друга. Здесь может помочь знание бизнеса.
 - Пример - товары внутри одной товарной группы, которые НЕ каннибализируют друг друга.
- Понимание главных причин каннибализации - драйверов.
 - Они должны быть наблюдаемыми (присутствовать в данных) и
 - Осмысленными (т.е. причины их влияния на спрос интуитивно понятны, как и «направление» связанного эффекта – уменьшает он спрос или увеличивает).
- Факторы (фичи), построенные на драйверах (см выше) войдут в итоговую модель.

Домашнее задание

- Обзор по методу иерархического прогнозирования: <https://otexts.org/fpp2/hts.html> Интересен как подход, пытающийся учитывать каннибализацию.
- Визуализация случаев каннибализации в данных при помощи Python
- Способы построения базового прогноза спроса без учета каннибализации.
 - Восстановление спроса. <https://fnow.ru/articles/pochemu-ne-stoit-prognozirovat-prodazhi-pri-upravlenii-zapasami-chast-1>
 - Auto Arima в Python.
 - Более сложные модели для прогнозирования временных рядов. Библиотека TSFresh для подготовки фичей. <https://tsfresh.readthedocs.io/en/latest/>
- Методы нахождения групп взаимно каннибализирующих товаров:
 - Тест Грэйнджера на наличие причинности во временных рядах. (https://en.wikipedia.org/wiki/Granger_causality) Подходит ли он нам? (кажется, нет)
 - DTW (<https://www.cs.unm.edu/~mueen/DTW.pdf>)

Домашнее задание - данные

- Kaggle: <https://www.kaggle.com/c/competitive-data-science-predict-future-sales>
- Ссылка на постановку задачи: https://raif.jet.su/hackathon/?utm_source=habr&utm_medium=post&utm_campaign=raif_hack_2018
 - Основная задача – поиск влияния цен, промо и наличия одних товаров на спрос по другим товарам.
 - Ссылка на данные Утконоса: <https://drive.google.com/open?id=1ER2ucshnLPCblHaHiq2vy3jNeL1wl6go>
 - Описание данных:
 - Orders – данные о заказах
 - Сырье/Товар ID – код товара
 - Заказчик – код покупателя
 - Дата создания
 - Продажи, БЕИ – Продажи в базовых единицах изменения (где-то кг, где-то штуки итд)
 - Items – данные о продажах
 - Товарная иерархия
 - Индикатор промо в SAP (ведется вручную)
 - Тип промо в SAP (ведется вручную)
 - Признак участия товаров в акции сайта (начался вести позднее, но данные более качественные, чем в SAP)
 - Тип промо сайта
 - Дата
 - Продажи в шт.
 - Продажи в руб.
 - Storehouse_available
 - Товар
 - День
 - Наличие (Каждый день на складе каждые 2 часа снимается данные о присутствии товара на остатках склада. По внутридневному 2часовому профилю заказов Утконос оценивает % доступности товара к заказу за день. 100 – доступен весь день, 0 – весь день не доступен)
- 4й файл можно не использовать