Дисклеймер: в данном кейсе используются замаскированные «живые» данные настоящей задачи, при этом все названия в кейсе подобраны случайно сохраняя смысл данных, все совпадения с реально существующими – случайны.

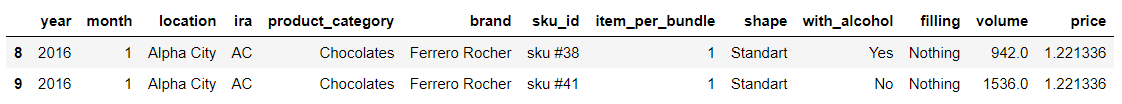
# Контекст

В некой стране **Компания** производит различные товары-сладости (SKU) и осуществляет продажу в 4х туристических городах (локациях), где количество и тип потребителей меняется от месяца к месяцу. Основную часть ассортимента **Компании** составляет шоколад, а основной объем продаж приходится на город Alpha City.

# Исходные данные

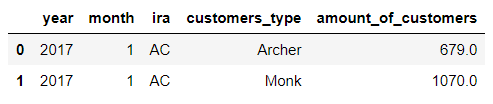
**Компании** доступно 2 источника информации, оба помесячные: внутренние данные по отгрузкам и розничным ценам в каждом из 4х городов, а также внешние данные, которые компания закупает у стороннего подрядчика, по количеству и типам клиентов в городах.   
  
Формат обоих источников - .tsv (tab-separated values).

# Данные по продажам



* **sku\_id** = уникальный идентификатор одной позиции товара (SKU = Stock Keeping Unit).
* **ira** = International Retail Association = уникальный код локации согласно классификации международной ассоциации ретейлеров.
* **product\_category** – категория продукта.
* **item\_per\_bundle** – количество единиц товара на 1 упаковку (например, бывают наборы по 3 шоколадки сразу).
* **shape** – форма упаковки.
* **with\_alcohol** – содержит ли алкоголь в начинке (yes/no).
* **filling** – начинка.
* **volume** – объем продаж в штуках
* **price** – цена 1 штуку товара в USD.

# Данные по потоку потребителей в городах

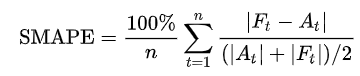


Сторонняя компания, предоставляющая данные по типу и количеству клиентов (потоку клиентов), имеет глубокую экспертизу и дополнительные внешние источники (например, забронированные путевки), на основании которых предоставляет как фактические данные (до июня 2019 года включительно), так и достаточно точной прогноз на 3 месяца вперед (июль-сентября 2019).

# Задача

Сейчас 01 июля 2019 года, **Компания** только что получила свежие данные по продажам до июня включительно. Вас, как специалиста по продвинутой аналитике, наняли для того, чтобы найти алгоритм, максимально точно описывающий поведение рынка, то есть зависимость продаж каждой позиции товара в каждой локации от различных факторов.

Задача состоит из 3х последовательных этапов:

1. Спрогнозировать продажи по каждому SKU в ассортименте для каждой локации на 3 месяца вперед помесячно (июль, август, сентябрь) с предположением, что цены останутся неизменными по сравнению с последним месяцем.  
   Точность оценивается по метрике SMAPE:  
   ,   
   где At – фактический объем, Ft – спрогнозированный для каждого месяца.  
     
   Если для какого-то SKU в какой-то локации прогноз сделать невозможно, необходимо указать причину.
2. Найти на тот же период цену (в пределах варьирования от -15% до +15%), при которой достигался бы максимум выручки (выручка = продажи \* цена). Гранулярность анализа остается на ваше усмотрение (SKU/бренд или весь портфель), **Компания** заинтересована в вашем экспертном мнении.
3. Обосновать, насколько точен ваш подход к определению «наилучшей» цены.

Формат предоставления результатов:

* Оформленный репозиторий с кодом проекта (в виде .zip файла или ссылки на github/аналог), удовлетворяющий следующим критериям:
  1. Воспроизводимый.
  2. Содержащий Jupyter Notebook (или несколько) c комментариями к проделанным шагам.
* tsv файл с результатами прогноза в формате:



* Краткий текстовый отчет для **Компании** в любом форматео качестве прогноза ипредложениипонаилучшей цене. Также приветствуются дополнительные комментарии: a) по имеющимся данным, b) чтобы вы сделали, имея больше ресурсов/времени/данных (каких?).