## Ситуация

1. Происходит переезд на новую ERP систему.
   1. В систему нужно перенести информацию о ранее проданных товарах.
   2. Все товары имеют уникальный серийный номер.
   3. Все товары разделены на категории (направления, activities) и подкатегории (номиналы, activity\_values).
2. Существуют следующие базы данных:
   1. База АМО - df\_amocrm\_export\_samp\_DD\_MM\_YY\_DD\_MM\_YY.csv - архив, выгружаемый из АМО СРМ.
   2. Реестр подкатегорий - df\_list\_of\_values.csv - Реестр подкатегорий (номиналов).
   3. Список товаров - база, которую нужно пополнить данными из предыдущих баз.
   4. df\_certificates\_list\_sample - тестовая выгрузка из Списка товаров (чтобы знать, как он выглядит в натуре)
   5. df\_codes\_archieve - архив префиксов.

## Задача

1. Написать скрипт, который перенесёт данные о проданных товарах из Базы АМО в новый файл r\_certs\_to\_import.csv (Список товаров), чтобы потом это дело импортировать в систему.

## Комментарии

1. Поля Списка товаров и источник данных:

|  | Источник | Примечание |
| --- | --- | --- |
| id | - | Не заполнять |
| amo\_id |  | Не заполнять |
| amo\_lead\_id | База АМО | Короче, парни поменяли название поля за каким-то, а прежнее поле amo\_id теперь для другой цели. |
| activity\_value\_id | Реестр подкатегорий |  |
| prefix | Реестр подкатегорий | Префиксов с буквой V в начале нет в реестре подкатегорий, потому что V добавляется к префиксу автоматически при генерации электронного сертификата.  То есть для целей проверки наличия префиксов в реестре подкатегорий, префикс VAT4 = AT4.  Товар с таким префиксом ничем не отличается от товара с обычным префиксом, кроме значения поля card\_type (см. ниже).  Товары с буквой V в начале префикса переносятся и обрабатываются в обычном порядке. |
| code | База АМО | prefix + number  Поля:  'Номер сертификата',  'Номер 2-го сертификата',  'Номер 3-го сертификата',  'Номер 4-го сертификата',  'Номер 5-го сертификата', |
| card\_type | База АМО | Значение ‘electronic’ если в начале prefix стоит буква V, значение ‘paper’ в остальных случаях. |
| design\_type | - | Везде указать значение ‘design1’ |
| status | База АМО | Значение 'sold\_used' если ‘Статус погашения’ в Базе АМО = ‘Погашен’ или если поля ‘Дата оплаты подрядчику’ в Базе АМО заполнено  В остальных случаях оставить пустым |
|  |  |  |
| supplier\_payment\_date | База АМО | Значение колонки “дата оплаты подрядчику” |
| sale\_price | База АМО | Если в сделке 1 товар и сделка не имеет производных сделок, то использовать значение поля SUBTOTAL  Если в сделке больше одного товара или сделка имеет производную сделку (то есть упоминается в поле КОРНЕВАЯ СДЕЛКА любой другой сделки), то sale\_price можно определить по ближайшей для этого префикса. |
| sale\_date | База АМО | Значение поля ‘Дата закрытия’ в Базе АМО |
| expiry\_date | - | Не заполнять |
| card\_used\_date | База АМО | Значение поля  ‘Дата оплаты подрядчику’ или ‘Дата получения услуги’ в Базе АМО |
| created\_at | - | Не заполнять |
| updated\_at | - | Не заполнять |
| deleted\_at | - | Не заполнять |
| number | База АМО | number без prefix |

1. Комментарий к работе с префиксами:
   1. В codes\_archieve должны быть все префиксы, как действующие, так и архивные, кроме, возможно, нескольких самых новых, которые есть только в list\_of\_values.
   2. Пометку OLD в ячейках с некоторыми префиксами в codes\_archieve нужно удалить на предобработке, так как если такие префиксы встретятся в АМО, они пометку OLD иметь, соответственно, не будут.
   3. Из базы codes\_archieve удалить префиксы, которые есть в list\_of\_values (чтобы префиксы не дублировались и присутствовали либо в одной базе, либо в другой).
   4. В Базе АМО идентифицировать товары с префиксами из list\_of\_values
   5. В Базе АМО идентифицировать товары с префиксами из codes\_archieve
   6. В оставшихся товарах не идентифицированные префиксы, вероятно, будут иметь следующую природу:
      1. Эти префиксы были неправильно выделены при обработке столбца с номерами сертификатов в базе АМО (то есть при выделении префикс забрал себе часть номера или ещё что).
      2. При внесении номера сертификата в АМО, менеджером была допущена ошибка.
   7. В моём представлении, товаров с такими не идентифицированными префиксами должно быть не более нескольких процентов. Их нужно выделить в отдельный df, который отправится в руки менеджеров для корректировки данных в АМО.
   8. Если на выходе товаров с не идентифицированными префиксами будет дохрена, то нужно будет искать причину.

1. Товары с префиксами, которые (а) отсутствуют в Списке подкатегорий и codes\_archieve, (б) товары, где в номере сертификата префикс вообще не указан (есть только цифровая часть), нужно загрузить в отдельный файл cert\_w\_missing\_prefix.csv.
2. В Базе АМО не должно быть сделок, где не указан номер сертификата. Если такие сделки есть, их нужно выделить в отдельный df.
3. Корневая и производная сделки, subtotal:
   1. С определённого момента, если в сделке более одного сертификата, было принято решение создавать новую сделку под каждый последующий сертификат, купленный в той же сделке. Это было сделано для удобства погашения сертификатов. В такой специально созданной производной сделке указывался номер основной, корневой, сделки, чтобы можно было в случае чего найти эту корневую сделку, где указана вся инфа по покупке.
   2. В таких случаях поле subtotal в корневой сделке содержит сумму цен двух и более товаров и использовать это значение для определения sale\_price уже нельзя.
4. Сделки, имеющие номер в поле “номер 2-го сертификата” (3-го, 4-го, 5-го):
   1. Тут аналогичная история, поле subtotal нельзя использовать для определения sale\_price, так как subtotal содержит сумму цен двух товаров.
5. Товары, указанные в поле “номер 2-го сертификата” - это всё отдельные товары, подлежащие включению в Список товаров в качестве отдельных строк с указанием номера той же сделки AMO, из которой они были извлечены.
6. Товары, указанные в полях “сертификат на тренера” и “сертификат на фотографа” нужно также вынести в отдельные строки, но сохранить в отдельный df.

## Результат

В конечном итоге предполагается получить следующие данные:

1. r\_certs\_to\_import\_DD\_MM\_YY\_DD\_MM\_YY - Список товаров, все необходимые поля которого заполнены. В названии зафиксировать период экспорта как в исходной выгрузке АМО.
2. r\_certs\_to\_import\_archieve\_prefix\_DD\_MM\_YY\_DD\_MM\_YY - Список товаров, чьи префиксы не обнаружены в list\_of\_values.csv, но есть в codes\_archieve
3. r\_cert\_w\_missing\_prefix\_DD\_MM\_YY\_DD\_MM\_YY - сделки без номера серта, номера сертов без префикса, номера с не идентифицированными префиксами
4. r\_missing\_prefix - список не идентифицированных префиксов