Необходимо сделать парсер с сайта meshok.net с выгрузкой данных в гугл.докс.

1. Сделать парсинг по лотам исходя из категории, продавца, и цены.

https://meshok.net/?f_p=100&good=252&pp=100&opt=3&pN=10900&sort=end_date&user=352195
|мин. цена||категория||тов. на стр.||фикс| |стр. 110| |продавец|

2. В гугл.докс выгружаются следующие данные
- название лота, как есть
- ник продавца
- id продавца
- фикс.цена товара на мешке (цифрой)
- цена товара для нашего каталога (высчитывается автоматически по формуле, цифрой, если на мешке цена в $ то умножать на актуальный курс доллара а потом по формуле)
- ссылка на товар на мешке
- количество, доступное на мешке
- основная категория товаров (Почтовые марки///по темам///[тематика, как на мешке])
- вторая категория товара (Почтовые марки///по странам///[страна, из названия, первая, которая найдется])
- теги через запятую (формируются из названия товара см. ниже)
- характеристики товара (формируются из названия товара см. ниже)
- название для магазина (формируются из названия товара см. ниже)
- артикул товара - формируется по принципу [m]+[номер лота на мешке] Например, m25863270

Пример разбора марки: https://goo.gl/T4DvWp

Название лота на мешке - Сан Томе и Принсипи 1984 M# 158 ** Корабль Пароход
ник продавца - Sammler-M
id продавца - 430787
цена товара на мешке – 105
цена товара на сайте – 210 (по формуле “*2”)
доступное количество – 12
основная категория - Почтовые марки///по
темам///Техника (вверху в хлебных крошках указана)
теги – корабль, пароход
Формат – не указано
Состояние MNH OG **
Номер по каталогу Michel – 158
Тираж – не указано
Страна - Сан Томе и Принсипи
Год - 1984

Пример разбора монеты: https://vk.cc/7Zub2l

Название лота на мешке - 25 центов монета США без букв (quarter Washington) 1951 год

ник продавца - oldcity49
id продавца - 303055
цена товара на мешке – 480
цена товара на сайте – 960 (по формуле “*2”)
доступное количество – 1
основная категория - Америка///США (вверху в хлебных крошках указана, но номинал в категории не берем)
теги – Washington, без букв
Состояние - не указано
Страна - США
Год - 1984

Пример парсинга.

4. Front End

Входные данные

- id профиля продавца, от которого берем товары
- id категории, из которой парсить товары данного продавца
- массивы данных для анализа названия товара: “Год” “Страна” “Номинал” “Состояние” “Теги” “Характеристики” “Доп информация” “Упаковка”.
- массив “Лишние слова”, которые не нужны в “Наименовании товара”
- минимальная цена товара ниже которой товары обрабатывать не нужно (по умолчанию 100 р.)
- формула подсчета цены (по умолчанию *2)

Схема работы скрипта:

1. берется первый ID клиента из списка, номера категорий и список прокси, запускается скрипт.
2. Скрипт проверяет наличие файла таблицы, с названием [id клиента].xls
3. Если файла нет, то файл создается, и начинает выполняться скрипт наполнения файла.
   3.1. Открывается страница https://meshok.net/ с параметрами:
      - f_p=100 – цена более 100 р.
      - good=252 – id категории (берется из списка категорий)
      - pp=100 – товаров на странице
      - opt=3 – только фикс цена
      - sort=end_date – сортировка по дате
      - user=352195 – id поставщика

   3.2. Парсится количество страниц(внизу есть), и далее последовательно открываются все страницы и собираются ссылки на товары и сохраняются в текстовом виде в спец. файле.
   3.3. После обработки последней страницы открывается файл и последовательно парсятся все товары, и в файле проставляются отметки напротив обработанных ссылок.
   3.4. Картинки товаров скачиваются и кладутся в папку img/[поставщик]/[категория]/[подкатегория]

3.5. Ссылка на картинку сохраняется в таблицу в виде /img/[поставщик]/[категория]/[подкатегория]/[имя картинки].jpg
3.6. В случае обрыва соединения парсинг товаров можно перезапустить с последнего не завершенного товара.

4. Если файл есть, то запускается скрипт обновления. (Вторая часть разработки.)
4.1. Открывается страница https://meshok.net/ с параметрами:
- f_p=100 – цена более 100 р.
- good=252 – id категории
- pp=100 – товаров на странице
- opt=3 – только фикс цена
- sort=end_date – сортировка по дате
- user=352195 – id поставщика

4.2. Парсится количество страниц(внизу есть цифра), и далее последовательно открываются все страницы и собираются ссылки на товары И ЦЕНЫ и сохраняются в текстовом виде в спец. файле.
4.3. Открывается таблица [id поставщика].xls и по ссылкам на товар сравниваются цены из таблицы и из спец. файла. Если цена изменилась, то данные сохраняются в таблицу.
4.4. Если товар присутствует в таблице но отсутствует в файле, то из таблицы товар удаляется.
4.5. Если товар отсутствует в таблице, но присутствует в файле, то парсится страница товара и данные добавляются в таблицу.
4.6. Картинки товаров скачиваются и кладутся в папку img/[поставщик]/[категория]/[подкатегория]
4.7. Ссылка на картинку сохраняется в таблицу в виде /img/[поставщик]/[категория]/[подкатегория]/[имя картинки].jpg

5. Таблица сохраняется в гугл диск.

Парсинг свойств и наименования из названия товара.

1. В скрипт загружаются массивы данных “Год” “Страна” “Номинал” “Состояние” и т.д. А также массив “Лишние слова”
2. Из названия товара удаляются все символы из массива “Символы”
3. Каждое слово из названия товара прогоняется по всем массивам и определяется в какой массив оно входит и к какому значению присваивается.
4. В соответствии с этим заполняются ячейки в таблице “Год” “Страна” “Номинал” и т.д.
5. Все слова которые не найдены ни в одном массиве, уходят в переменную “Наименование товара”
6. Все слова которые найдены в массиве “Лишние слова” удаляются из наименования товара.
7. “Наименование товара” приводится в правильный регистр, первая буква названия заглавная, и после точки с заглавной буквы.
8. “Наименование товара” формируется по шаблону [Год]. [Страна]. [Наименование], [Формат, или Номинал]. [Состояние] Например: 1959. Бурунди. Собаки, Почтовый блок. MNH OG **