**HD竞品信息爬取和Carro前台信息爬取Q&A**

1. **推荐页面每次登入都会刷新，记录结果并不固定，需要如何确定提取的时间。**
2. 能否在美国消费者购物集中的时间段自动做2次爬取前台资料（能否不需要人工操作）：

西雅图时间：12：00-14：00和17：00-20：00这两个时间段；

1. 提取后的数据能否自动保存并上传至google云端（能否不需要人工操作），然后我们可以不定时人工下载；
2. **HD的网页解析有问题，提取产品详细数据时会有报错的情况，是否需要人工操作。**

（1）请列一下具体哪些问题？

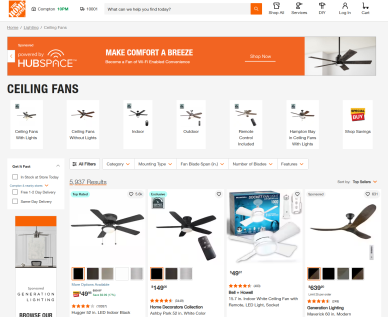
（2）另外我看到Ben导出的csv文档只有250个产品，我们手动数了一下，HD前5页应该一共有280-283个产品的。

（3）我已经在csv表格上新增了G,H,I列内容需求，还有将表头的顺序调整了一下。如附件一。

1. **筛选和排序，是筛选先还是排序先，是否过滤’sponsored‘和HD自家产品。**

（1）以下为从homedepot.com前台自然搜索类目CEILING FAN后的结果(不需要筛选和排序，同时不需要过滤Sponsored和HD自家产品)，需要每周操作一次：

https://www.homedepot.com/b/Lighting-Ceiling-Fans/N-5yc1vZbvlq?catStyle=ShowProducts



（2）以下为从homedepot.com前台自然搜索类目CEILING FAN WITH LIGHTS后的结果(不需要筛选和排序，同时不需要过滤Sponsored和HD自家产品) ，需要每周操作一次：

<https://www.homedepot.com/b/Lighting-Ceiling-Fans-Ceiling-Fans-With-Lights/N-5yc1vZcjnu>



1. **新增爬取需求一：**将指定品牌（Carro,Parrot Uncle,Matrix Decro,Breezary）占据CEILING FAN（https://www.homedepot.com/b/Lighting-Ceiling-Fans/N-5yc1vZbvlq?catStyle=ShowProducts）和CEILING FAN WITH LIGHTS（<https://www.homedepot.com/b/Lighting-Ceiling-Fans-Ceiling-Fans-With-Lights/N-5yc1vZcjnu>）类目前5页排名的产品提取并生成一个表格，如附件二，看能否尽可能多地提取到跟表格一样的信息。需要每天都操作一次；
2. **新增爬取需求二：**使用Carro在HD的后台的SKU在前台自动搜索出该SKU所有的页面信息并保存下载。或者说能做到的自动化程度如何？

例如：

登陆 [www.homedepot.com](http://www.homedepot.com)

在搜索框输入HD SKU：319929479

