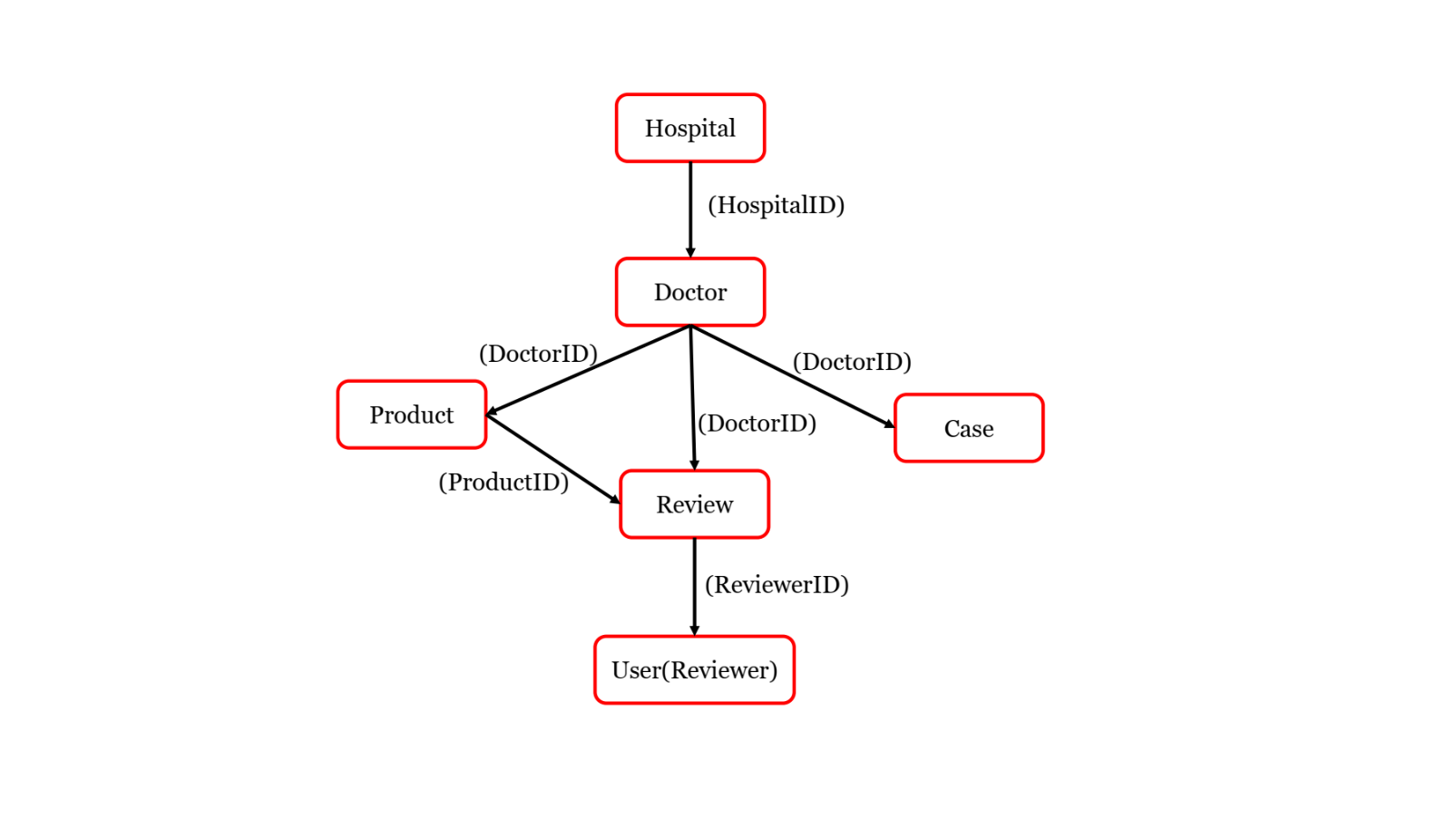
总需求：全网数据；每周/两周/一个月收一次；标黄部分可能需要计算或无法爬取，若无法爬取可忽略。

web主页：<https://www.soyoung.com/>

数据结构（括号内为匹配字段）：

1. **产品层面**

产品list：<https://www.soyoung.com/search/product>

产品ID(ProductID)

产品爬取时间(PCrawlDate)

产品名称(ProductName)

机构ID(HospitalID)

机构名称(HospitalName)

机构评分（HospitalRating）

机构医生数量（DoctorNum）

机构地址(HospitalAddress)

产品原价（ProductOPrice）

产品券后价（ProductPrice，实际价格）

产品销售量(ProductSale)

产品退单率(ReturnRatio,)

产品额外费用（ProductAPrice）

产品评价数量(PReviewNum)

产品消费后日记数量(PPostReviewNum)

产品差评日记数量(PNegReviewNum)

产品追评日记数量（PAddReviewNum）

产品有图日记数量(PImageReviewNum)

产品有视频日记数量(PVideoReviewNum)

产品第一条评论出现的时间（根据日记板块时间计算）

产品平均评分（根据所有日记评分计算）



1. **日记评论层面**

日记ID (ReviewID)

爬取时间（RCrawlDate）

是否通过新氧消费（IsCustom1Review, 是=1，否=0）

是否上传消费凭证（IsCustom2Review, 是=1，否=0）

是否为优质评价（IsHQReview, 是=1，否=0）

是否为体验官日记（IsEOReview, 是=1，否=0）

是否为图片日记（IsImageReview, 是=1，否=0）

是否为视频日记（IsVideoReview, 是=1，否=0）

日记发布时间(ReviewDate,如2021-12-21)

日记浏览量(ReviewViews)

日记评分(ReviewRating)

机构环境评分(ReviewERating)

服务态度评分(ReviewSRating)

项目效果评分(ReviewTRating)

医生专业评分(ReviewPRating)

日记图片数量(ReviewImageNum，纯文本评论和视频评论记为0)

日记文本内容(ReviewContent)

日记文本长度(ReviewTextLen，可后续根据文本内容计算)

机构回复内容（HospitalResponseText，无回复则为空）

**图文日记的图片内容（ReviewImage，图文日记需保存所有的图片，以ReviewID与所在日记进行匹配）**

**视频日记的视频内容（ReviewVideo，带视频的日记需保存视频，以ReviewID与所在日记进行匹配）**

日记获赞数量（ReviewLikeNum）

日记收藏数量（ReviewFollowNum）

日记回复数量（ReviewReplyNum）

日记追评内容（ReviewAddText，没有追评则为空）

合集日记数量（FollowReviewNum，单个日记则为0）

合集日记内容（对于有合集的患者日记，将合集日记中的每一篇追加内容包括文本和图片等单独保存，后续以日记ID与主日记合并）



日记发布者ID（ReviewerID)

日记发布者名称（ReviewerName）

日记发布者年龄（ReviewerGender）

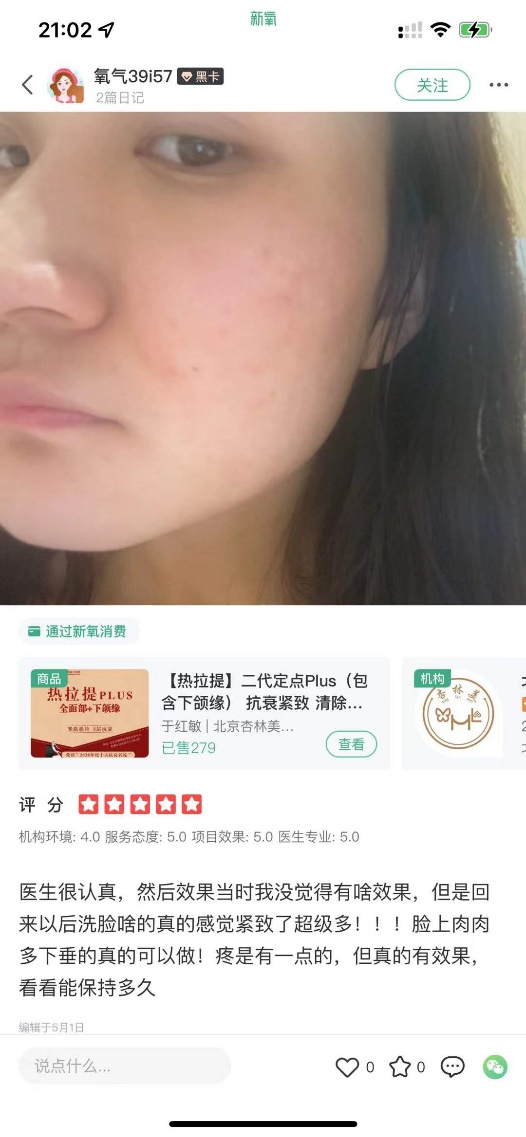
日记发布者性别（ReviewerGender）

日记发布者所在城市（ReviewerCity）

日记发布者日记数量（ReviewerPosts）

日记发布者粉丝数量（ReviewerFollwer）

日记发布者点赞与收藏数量(ReviewerLikes)



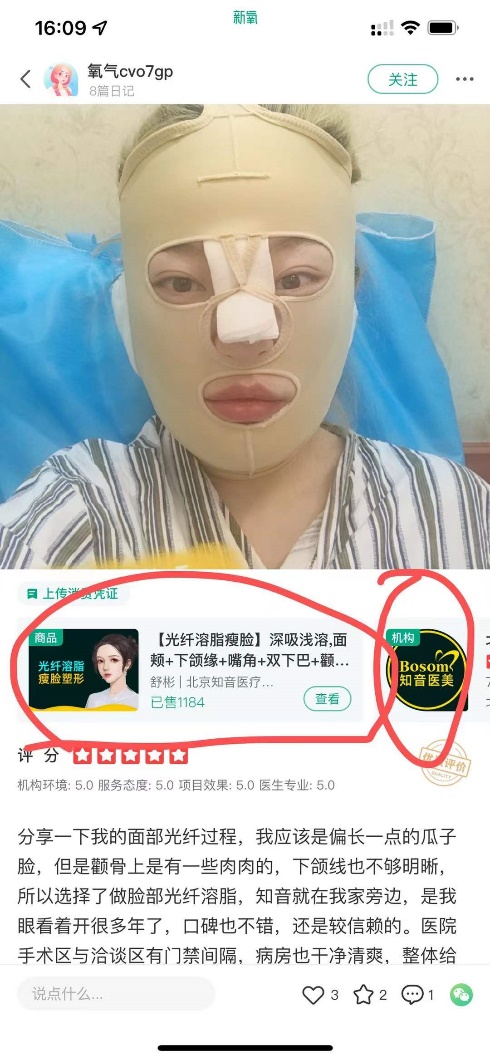
日记关联产品ID(ProductID

日记关联产品名称（ProductName）

日记关联医生ID（DoctorID）

日记关联医生姓名（DoctorName）

日记关联医疗结构ID(HospitalID)

日记关联医疗结构名称(HospitalName)

1. **医疗机构层面**

机构list: <https://www.soyoung.com/search/hospital>

机构ID(HospitalID)

爬取时间（CrawlTime）

机构名称（HospitalName）

机构评分（HospitalRating）

服务人次（ServiceNum）

机构地址(HospitalAddress)

机构分类（HospitalType，医疗美容医院、公立医院等）

评价数量(ReviewNum)

医生数量（DoctorNum）

机构城市(HospitalCity)

机构荣誉榜单数量（ReputNum）

机构荣誉榜单名称(Reputation, 以#间隔,)

收录年份(RecodeYear)

专利数量(APatentNum)

机构总体评分(HospitalRating)

机构效果评分(HospitalURating)

机构环境评分(HospitalTRating)

机构服务评分(HospitalSRating)

机构粉丝数量（HospitalFollower）

项目数量(ProjectNum)

机构动态数量(MGCNum)

日记数量（HReviewNum）

差评日记数量（HNegReviewNum）

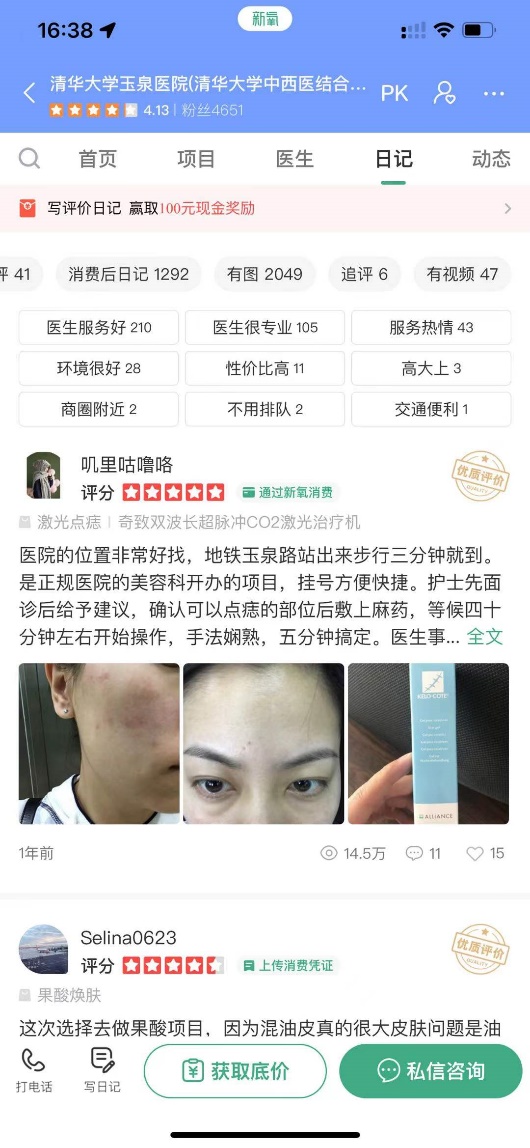
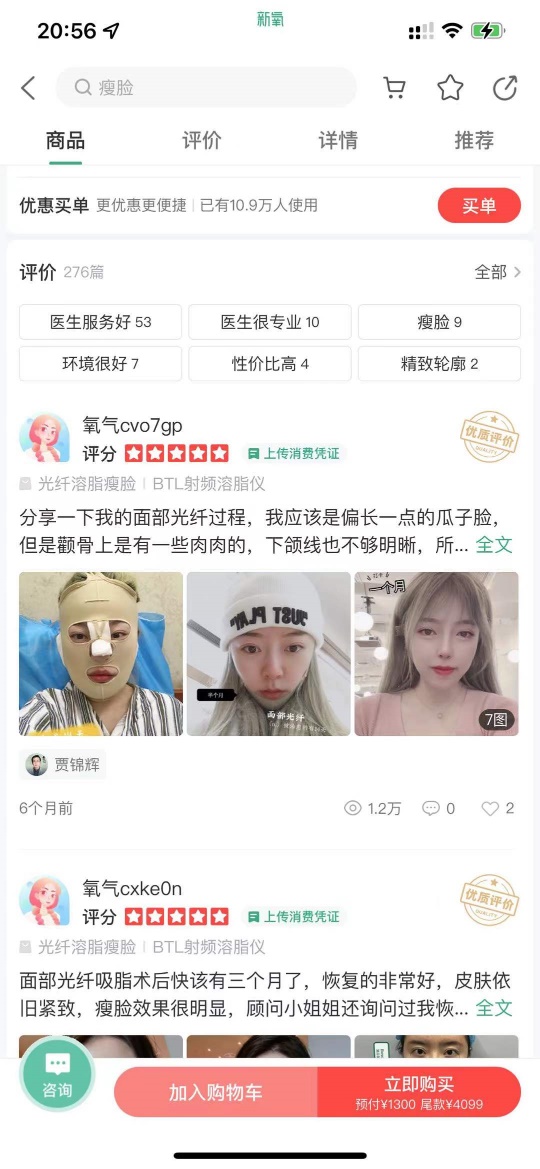
追评日记数量（HAddReviewNum）

优质日记数量（HHQReviewNum）

图文日记数量（HImageReviewNum）

视频日记数量（HVideoReviewNum）

消费后日记数量（HPostReviewNum）



1. **医生层面**

医生ID（DoctorID）

爬取时间(DCrawlDate)

医生姓名（DoctorName）

专业职称(ProfessionalTitle)

医生所在城市（DoctorCity）

从业年限(WorkYear)

医生所在医疗机构ID（HospitalID）

医生所在医疗机构名称（HospitalName）

性别(DoctorGender)

患者评分(DoctorRating)

服务人数（DoctorConsultation，面诊人数）

案例数量(DGCNum)

擅长领域（ExpertArea, 以#分隔）

医生履历文本（DoctorBio）

医生服务评分（DoctorSRating）

医生专业评分（DoctorPRating）

医生效果评分（（DoctorTRating）

日记数量（DoctorReviewNum）

粉丝数量(DoctorFollower)

好评率（DoctorPosrate）

是否开通视频面诊（VideoService）

视频面诊价格（VideoPrice）

是否开通语音面诊（VoiceService）

语音面诊价格（VoicePrice）

是否开通图文面诊（TextService）

图文面诊价格（TextPrice）

项目数量（DoctorProjectNum）

销售数量（DoctorSales）

日记数量（DReviewNum）

追评日记数量（DAddNum）

差评日记数量（DNegReviewNum）

图文日记数量（DImageReviewNum）

视频日记数量（DVideoReviewNum）

消费后日记数量（DPostReviewNum）



1. **医生案例层面**

案例爬取时间（CCrawlDate）

案例ID（CaseID）

医生ID（DoctorID）

医生姓名（DoctorName）

案例发布时间（CaseDate）

是否为图文案例（IsImageCase）

是否为视频案例（IsVedioCase）

案例图片数量（CaseImage）

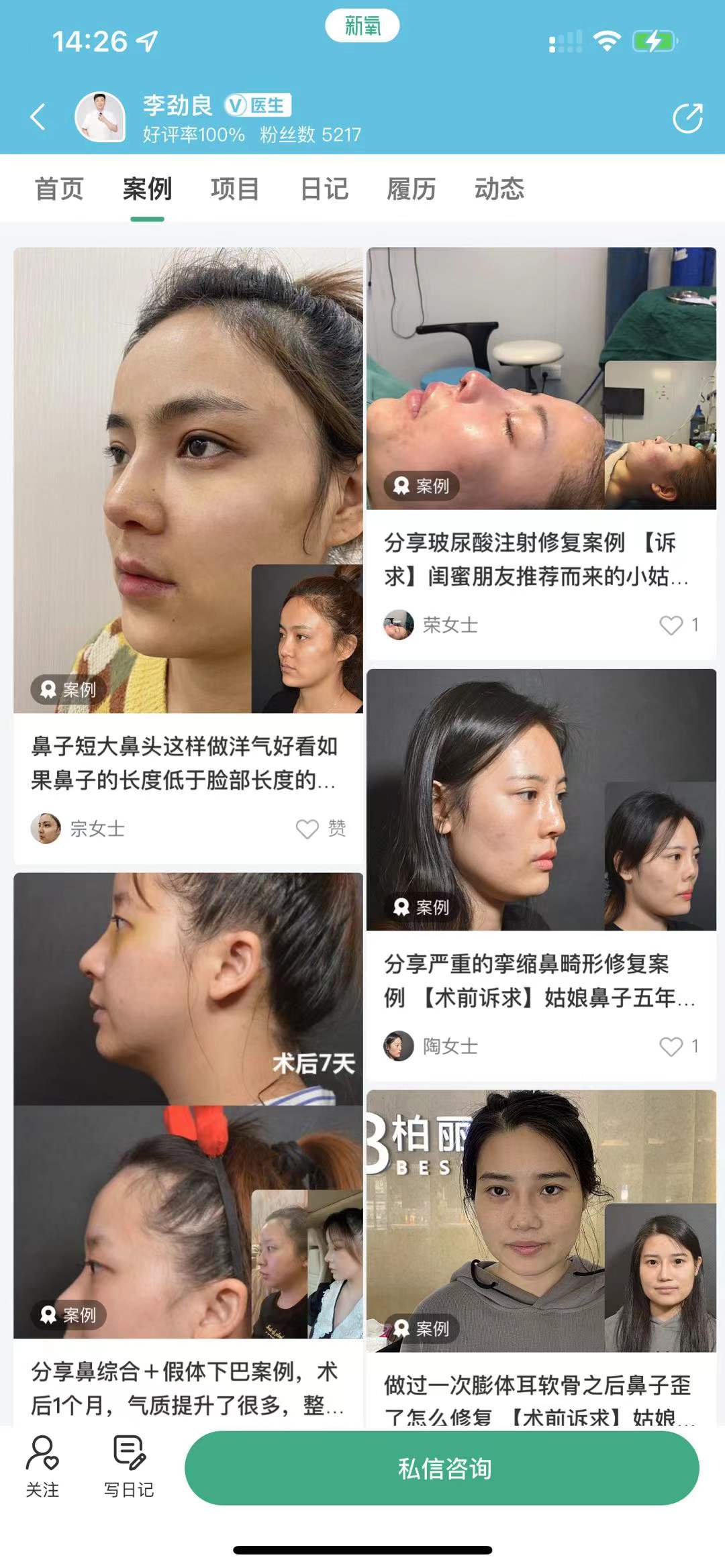
案例获赞数量（CaseLikes）

案例收藏数量（CaseFollow）

案例评论数量（CaseReply）

案例文本内容（CaseText）

**案例图片内容（CaseImage，保存每个案例的所有图片，以CaseID进行匹配）**

**案例视频内容（CaseVideo，保存每个案例的所有视频，以CaseID进行匹配）**

1. **用户层面（日记发布者）**

用户ID（ReviewerID)

用户名称（ReviewerName）

用户年龄（ReviewerAge）

用户性别（ReviewerGender）

用户动态数量（ReviewerPosts）

用户关注数量（ReviewerFollowee）

用户粉丝数量（ReviewerFollwer）

用户点赞与收藏数量(ReviewerLikes)

用户经验值(ReviewerExp, Web端)

用户氧分(ReviewerLevel，Web端)