



تمرین کامپیوتری سوم

سیستم‌های عامل - بهار ۱۳۹۹

مجموعه‌های داده

دانشکده مهندسی برق و کامپیوتر

مسئول تمرین: محمد مریدی

استاد: مهدی کارگهی

مقدمه



در این تمرین شما به تحلیل داده‌هایی که از وبگاه¹ Goodreads استخراج شده است می‌پردازید. در این پرونده² به بیان ویژگی‌های برخی از موجودیت³‌هایی که در این تمرین به آن‌ها احتیاج است پرداخته می‌شود. Goodreads یک شبکه اجتماعی کتاب است، که کاربران در آن کتاب‌هایی را که خوانده‌اند پیدا می‌کنند، به کتاب‌ها امتیاز می‌دهند و کتاب‌های محبوب خود را به دوستان خود توصیه می‌کنند.

مجموعه‌های داده⁴



مجموعه‌های داده‌ای که در این تمرین به شما داده شده است در قالب CSV⁵ هستند. CSV نام یک قالب برای پرونده‌های متنی است که در آن مقادیر با استفاده از نماد کاما (,) از یکدیگر جدا می‌شوند. این قالب یکی از روش‌های پرطرفدار برای تبادل اطلاعات است. در زیر درباره هر ویژگی و نوع داده⁶ مربوط به مجموعه‌های داده‌ای که در اختیار شما قرار داده شده است، توضیح مختصری آمده است.

¹ Website

² File

³ Entity

⁴ Datasets

⁵ Comma-Separated Values

⁶ Data Type

کتاب‌ها



اطلاعات مربوط به کتاب‌ها در پرونده‌ای به نام books.csv آورده شده است.

نام ویژگی	توضیح	نوع داده
book_id	شناسه یکتای کتاب	عدد صحیح
book_title	عنوان کتاب	رشته‌ای از کاراکترها
genre_1	اولین ژانر ادبی کتاب	رشته‌ای از کاراکترها
genre_2	دومین ژانر ادبی کتاب	رشته‌ای از کاراکترها
pages	تعداد صفحات کتاب	عدد صحیح
author_name	نام نویسنده	رشته‌ای از کاراکترها
author_average_rating	امتیاز متوسط نویسنده	عدد اعشاری

نظرات کاربران



اطلاعات مربوط به نظراتی که کاربران درباره کتاب‌ها داده‌اند در پرونده‌ای به نام reviews.csv آورده شده است.



نام ویژگی	توضیح	نوع داده
book_id	شناسه کتابی که بر روی آن نظر داده شده است	عدد صحیح
rating	امتیازی که به کتاب داده است	عدد صحیح
number_of_likes	تعداد کاربرانی که این نظر مورد علاقه‌شان بوده است	عدد صحیح

نکات تکمیلی

- ستون‌هایی که تحت عنوان id مشخص شده‌اند در آن مجموعه داده یکتا هستند.
- ژانرهای کتاب که چندبخشی هستند با استفاده از جداکننده '-' از یکدیگر جدا شده‌اند و جداکننده فاصله⁷ در این ستون‌ها وجود ندارد.
- تضمین می‌شود که در مجموعه‌های داده‌ای که در اختیار شما قرار گرفته است، از کاراکتر ',' تنها برای جداسازی اجزا استفاده شده است.

منابع

<https://www.goodreads.com>

<https://www.iconfinder.com>

⁷ Space : ' '