

تمرین مربوط به کتابخانه Pandas: تحلیل تعامل کاربران سرویس استریمینگ

شما به عنوان تحلیل‌گر داده در یک سرویس استریمینگ جدید کار می‌کنید. وظیفه شما این است که الگوهای تعامل کاربران را تحلیل کنید تا مشخص شود کدام پلتفرم‌ها بیشترین تعامل را ایجاد می‌کنند و روند تعامل در طول زمان چگونه است. خروجی نهایی باید یک **heatmap** باشد که امتیاز تعامل روزانه را در پلتفرم‌های مختلف وب، iOS، اندروید نمایش دهد. این تحلیل به تیم محصول کمک می‌کند تا تصمیم بگیرند که تلاش‌های توسعه‌ای خود را روی کدام پلتفرم متمرکز کنند.

مجموعه داده

ابتدا باید یک مجموعه داده مصنوعی ایجاد کنید که رویدادهای تعامل کاربران را نشان دهد. این مجموعه داده باید شامل ستون‌های زیر باشد:

- `event_id`: شناسه یکتا برای هر رویداد.
- `user_id`: شناسه یکتا برای هر کاربر.
- `timestamp`: تاریخ و زمانی که رویداد رخ داده است (ابتدا به صورت رشته).
- `platform`: پلتفرمی که رویداد در آن رخ داده است. ('web', 'ios', 'android')
- `event_type`: نوع تعامل کاربر. ('login', 'play_video', 'like', 'comment', 'logout')
- `video_id`: شناسه ویدیو مرتبط با رویداد (می‌تواند خالی باشد اگر رویداد مربوط به ویدیو نباشد).
- `watch_time_sec`: مدت زمانی که کاربر ویدیو را تماشا کرده است فقط برای رویدادهای 'play_video'
- `video_duration_sec`: مدت کل ویدیو.

مراحل پیاده‌سازی (راهنما)

ایجاد داده‌ها:

- یک تابع ایجاد کنید که یک `DataFrame` با ساختار بالا تولید کند. از بازه زمانی حدود ۲ تا ۳ ماه برای `timestamp` استفاده کنید.
- اطمینان حاصل کنید که داده‌ها دارای ویژگی‌های تصادفی و واقع‌گرایانه هستند. مثلاً `watch_time_sec` نباید از `video_duration_sec` بیشتر باشد.

پاک سازی داده ها و مهندسی ویژگی ها:

- ستون timestamp را به یک شیء datetime تبدیل کنید.
- یک امتیاز تعامل (engagement_score) برای هر رویداد تعریف و محاسبه کنید. همه رویدادها به یک اندازه مهم نیستند. به عنوان مثال، یک 'comment' تعامل بیشتری نسبت به 'play_video' دارد. به هر event_type امتیاز زیر را اختصاص دهید:

- comment: 5 points
- like: 3 points
- play_video: 1 point
- login / logout: 0 point

- یک ستون جدید به نام engagement_score بر اساس این وزن ها ایجاد کنید.

تجمیع سری های زمانی:

- هدف این است که یک خلاصه روزانه از تعاملات به دست آورید.
- داده ها را بر اساس بخش تاریخ timestamp و platform گروه بندی کنید.
- مجموع engagement_score را برای هر گروه محاسبه کنید. این کار امتیاز تعامل روزانه برای هر پلتفرم را به شما می دهد.

تغییر شکل داده ها برای مصور سازی:

- داده های تجمیع شده احتمالاً در قالب "طولانی" هستند (با ستون هایی برای تاریخ، پلتفرم و امتیاز).
- این داده ها را به قالب "عریض" تغییر دهید که در آن:
 - شاخص تاریخ باشد.
 - ستون ها پلتفرم ها ('web', 'ios', 'android') باشند.
 - مقادیر امتیاز تعامل روزانه باشند.
- هر مقدار گم شده ای که پس از تغییر شکل ظاهر می شود را مدیریت کنید (مثلاً روزی که هیچ فعالیتی در یک پلتفرم خاص نداشته باشد). مقدار * مناسب خواهد بود.

مصور سازی:

- از یک کتابخانه مانند matplotlib یا seaborn برای ایجاد یک heatmap از DataFrame نهایی استفاده کنید.

- heatmap باید به وضوح امتیازات تعامل روزانه را نشان دهد، با تاریخ‌ها در یک محور و پلتفرم‌ها در محور دیگر. از یک نقشه رنگی استفاده کنید که نقاط تعامل بالا ("hot spots") را به راحتی نشان دهد.
 - اطمینان حاصل کنید که نمودار شما دارای عنوان و برچسب‌های محورها باشد.
- این مصورسازی نهایی، راه‌حل مسئله است. می‌توان آن را بر اساس درستی و وضوح در ارائه روندهای تعامل ارزیابی کرد.