

Stage 1:

با توجه به فایل `stage1_qc_prep.py_01`، فایل `Q1_clean.csv` نسبت به فایل `Q1.csv` چه تغییراتی داشته؟

فایل `01_stage1_qc_prep.py` یک اسکریپت پایتون است که برای انجام QC (کنترل کیفیت) بر روی داده‌های ردیابی چشم، خلاصه‌سازی داده‌ها و مرتب‌سازی آن‌ها بر اساس شماره شرکت‌کننده و شماره سؤال طراحی شده است. این اسکریپت فایل‌های CSV ورودی را پردازش می‌کند، نمونه‌های نامعتبر (به عنوان مثال، داده‌های مربوط به پلک زدن) را حذف می‌کند، و یک فایل CSV خلاصه ایجاد می‌کند.

با توجه به عملکرد این اسکریپت، می‌توان انتظار داشت که فایل `Q1_clean.csv` نسبت به فایل `Q1.csv` تغییرات زیر را داشته باشد:

1. حذف ردیف‌های نامعتبر: ردیف‌هایی که در آن‌ها داده‌های ردیابی چشم نامعتبر هستند (به عنوان مثال، ردیف‌هایی که در آن‌ها شرکت‌کننده در حال پلک زدن بوده یا داده‌ها به دلایل دیگر نامعتبر تشخیص داده شده‌اند) حذف شده‌اند. تابع `valid_mask` برای تعیین ردیف‌های معتبر و تابع `apply_blink_padding` برای حذف ردیف‌های اطراف پلک زدن‌ها استفاده می‌شوند.

2. مرتب‌سازی: داده‌ها بر اساس شماره شرکت‌کننده و شماره سؤال مرتب شده‌اند. اگرچه در این مورد خاص هر دو فایل مربوط به یک شرکت‌کننده و یک سؤال هستند و این مورد تأثیری ندارد.

3. احتمالاً تغییر در ستون‌ها: اگر اسکریپت طوری طراحی شده باشد که ستون‌های خاصی را حذف یا اضافه کند، این تغییرات نیز در فایل `Q1_clean.csv` اعمال می‌شوند. با این حال، با توجه به کد ارائه شده، به نظر نمی‌رسد که ستونی حذف یا اضافه شود، بلکه بیشتر تمرکز بر روی فیلتر کردن ردیف‌ها است.

به طور خلاصه، `Q1_clean.csv` نسخه‌ای از `Q1.csv` است که داده‌های نامعتبر آن حذف شده و احتمالاً مرتب شده است تا برای تجزیه و تحلیل بیشتر آماده شود.

Stage 2:

با توجه به فایل `stage2_aoi_and_phases.py_02` فایل `Q1_stage2.csv` نسبت به فایل `Q1_clean.csv` چه تغییراتی داشته؟

اضافه شدن ستون‌های جدید: ستون‌های متعددی به فایل `Q1_stage2.csv` اضافه شده‌اند که اطلاعات مربوط به حرکات چشم، مناطق مورد توجه (AOI)، فازهای آزمایش و محاسبات مرتبط با سرعت و مسافت حرکات چشم را در بر می‌گیرند. برخی از این ستون‌ها عبارتند از: BKDUR, BKID, BKPMIN, BPOGV, BPOGX, BPOGY, CS, CX, CY, FPOGD, FPOGID, FPOGS, FPOGV, FPOGX, FPOGY, LEYEX, LEYEV, LEYEZ, LPCX, LPCY, LPD, LPOGV, LPOGX, LPOGY, LPS, LPUPILD, LPUPILV, LPV, REYEX, REYEV, REYEZ, RPCX, RPCY, RPD, RPOGV, RPOGX, RPOGY, RPS, RPUPILD, RPUPILV, RPV, FPOG_X, FPOG_Y, BPOG_X, BPOG_Y, x, y, sample_valid, in_left, in_right, in_question, in_answers_rect, in_submit, in_right_timer, in_right_bg, area_id, opt_id_at_sample, aoi_correct, aoi_wrong, area_code, area_is_question, area_is_wrong, area_is_correct, area_is_timer, area_is_submit, phase, phase_code, dx_f, dy_f, speed_f, dx_b, dy_b, speed_b, d_lpd, d_rpd, is_blink, participant, section, question_id, question_number, has_correct_answer.

محاسبه و اضافه شدن اطلاعات AOI: ستون‌هایی مانند `aoi_wrong` و `aoi_correct` نشان می‌دهند که آیا نقطه نگاه در هر نمونه‌برداری در یک ناحیه مورد توجه درست یا غلط بوده است. ستون `area_code` نیز کد مربوط به ناحیه مورد توجه را مشخص می‌کند.

محاسبه و اضافه شدن اطلاعات مربوط به فازهای آزمایش: ستون‌هایی مانند `phase_code` و `phase` فازهای مختلف آزمایش را مشخص می‌کنند.

محاسبه و اضافه شدن اطلاعات مربوط به حرکات چشم: ستون‌هایی مانند `dx_f`, `dy_f`, `speed_f`, `dx_b`, `dy_b`, `speed_b`, `d_lpd`, `d_rpd` اطلاعات مربوط به سرعت و مسافت حرکات چشم را نشان می‌دهند. ستون `is_blink` نیز مشخص می‌کند که آیا در آن لحظه پلک زدن رخ داده است یا خیر.

اضافه شدن اطلاعات مربوط به شرکت‌کننده و سوال: ستون‌هایی مانند `participant`, `section`, `question_id`, `question_number`, `has_correct_answer` اطلاعات مربوط به شرکت‌کننده، بخش آزمایش، شناسه سوال و درست بودن پاسخ را نشان می‌دهند.

به طور خلاصه، فایل `Q1_stage2.csv` حاوی اطلاعات بسیار غنی‌تری در مورد حرکات چشم، مناطق مورد توجه و فازهای آزمایش نسبت به فایل `Q1_clean.csv` است. این اطلاعات برای تحلیل دقیق‌تر رفتار شرکت‌کنندگان در طول آزمایش مورد استفاده قرار می‌گیرند.

توضیحات ستون‌های اضافه شده:

- FPOG_X, FPOG_Y:** موقعیت نقطه نگاه (foveal point of gaze) در محور X و Y، احتمالاً نرمالایز شده به محدوده 0 تا 1. این ستون‌ها مستقیماً از داده‌های ردیابی چشم بدست می‌آیند و نشان‌دهنده محل تمرکز نگاه شرکت‌کننده روی صفحه نمایش هستند.
- BPOG_X, BPOG_Y:** موقعیت نقطه نگاه (binocular point of gaze) در محور X و Y، احتمالاً نرمالایز شده به محدوده 0 تا 1. این ستون‌ها مستقیماً از داده‌های ردیابی چشم بدست می‌آیند و نشان‌دهنده محل تمرکز نگاه دوچشمی شرکت‌کننده روی صفحه نمایش هستند.

- **x, y:** این ستون‌ها احتمالاً موقعیت نقطه نگاه را در مختصات پیکسل صفحه نمایش نشان می‌دهند. این مقادیر با تبدیل مختصات نرمالایز شده (FPOG_X, FPOG_Y) یا (BPOG_X, BPOG_Y) به مختصات پیکسل بدست می‌آیند.
- **sample_valid:** یک مقدار بولی که نشان می‌دهد آیا نمونه داده در یک زمان خاص معتبر است یا خیر. این ستون بر اساس معیارهای مختلفی مانند کیفیت سیگنال ردیابی چشم و وجود پلک زدن تعیین می‌شود.
- **in_left, in_right:** مقادیر بولی که نشان می‌دهند آیا نقطه نگاه در پنل سمت چپ یا راست صفحه نمایش قرار دارد. این ستون‌ها بر اساس موقعیت X نقطه نگاه و نسبت پنل‌ها در تنظیمات (config) محاسبه می‌شوند.
- **in_question, in_answers_rect, in_submit, in_right_timer, in_right_bg:** مقادیر بولی که نشان می‌دهند آیا نقطه نگاه در ناحیه سوال، ناحیه پاسخ‌ها، دکمه ثبت، تایمر سمت راست یا پس‌زمینه سمت راست قرار دارد. این ستون‌ها با استفاده از تابع rect_contains و بر اساس موقعیت X و Y نقطه نگاه و مختصات مستطیل‌های مربوط به هر ناحیه محاسبه می‌شوند.
- **area_id:** یک شناسه عددی که ناحیه مورد توجه (AOI) فعلی را مشخص می‌کند. این ستون بر اساس ستون‌های **in_*** تعیین می‌شود و نشان می‌دهد که نقطه نگاه در کدام ناحیه قرار دارد.
- **opt_id_at_sample:** شناسه گزینه انتخاب شده در هر نمونه. اگر نقطه نگاه در ناحیه پاسخ‌ها باشد، این ستون با استفاده از تابع option_grid_rects و بر اساس موقعیت X و Y نقطه نگاه و مختصات گزینه‌ها تعیین می‌شود.
- **aoi_correct, aoi_wrong:** مقادیر بولی که نشان می‌دهند آیا نقطه نگاه در یک ناحیه مورد توجه درست یا غلط بوده است. این ستون‌ها بر اساس area_id و اطلاعات مربوط به پاسخ صحیح محاسبه می‌شوند.
- **area_code:** کد مربوط به ناحیه مورد توجه. این ستون بر اساس area_id تعیین می‌شود و نشان می‌دهد که نقطه نگاه در کدام ناحیه قرار دارد (به عنوان مثال، سوال، پاسخ، دکمه ثبت).
- **area_is_question, area_is_wrong, area_is_correct, area_is_timer, area_is_submit:** مقادیر بولی که نشان می‌دهند آیا ناحیه مورد توجه فعلی مربوط به سوال، پاسخ غلط، پاسخ درست، تایمر یا دکمه ثبت است. این ستون‌ها بر اساس area_code تعیین می‌شوند.
- **phase:** فاز آزمایش (به عنوان مثال، فاز مشاهده سوال، فاز پاسخ دادن). این ستون بر اساس تعداد نمونه‌های پیاپی که در ناحیه پاسخ‌ها قرار دارند، تعیین می‌شود.
- **phase_code:** یک کد عددی که فاز آزمایش را نشان می‌دهد.
- **dx_f, dy_f:** تغییرات در موقعیت نقطه نگاه (foveal) در محور X و Y نسبت به نمونه قبلی. این ستون‌ها با محاسبه تفاضل موقعیت فعلی و قبلی نقطه نگاه بدست می‌آیند.
- **speed_f:** سرعت حرکت نقطه نگاه (foveal). این ستون با استفاده از dx_f, dy_f و نرخ نمونه‌برداری محاسبه می‌شود.
- **dx_b, dy_b:** تغییرات در موقعیت نقطه نگاه (binocular) در محور X و Y نسبت به نمونه قبلی. این ستون‌ها با محاسبه تفاضل موقعیت فعلی و قبلی نقطه نگاه بدست می‌آیند.
- **speed_b:** سرعت حرکت نقطه نگاه (binocular). این ستون با استفاده از dx_b, dy_b و نرخ نمونه‌برداری محاسبه می‌شود.

- **d_lpd, d_rpd:** تغییرات در قطر مردمک چشم چپ و راست نسبت به نمونه قبلی. این ستون ها با محاسبه تفاضل قطر مردمک فعلی و قبلی بدست می آیند.
 - **is_blink:** یک مقدار بولی که نشان می دهد آیا در این نمونه پلک زدن رخ داده است یا خیر. این ستون بر اساس ستون BKID تعیین می شود.
 - **participant:** شناسه شرکت کننده در آزمایش.
 - **section:** بخش آزمایش.
 - **question_id:** شناسه سؤال.
 - **question_number:** شماره سؤال.
 - **has_correct_answer:** یک مقدار بولی که نشان می دهد آیا این سؤال پاسخ صحیح دارد یا خیر.
- این توضیحات به شما کمک می کنند تا درک بهتری از ستون های اضافه شده در فایل Q1_stage2.csv داشته باشید و نحوه محاسبه و ماهیت هر یک را بدانید.

Stage 3:

با توجه به فایل `stage3_join_and_metrics.py_03`، فایل‌های `*_stage3` چی هستند و شامل چه فیلدهایی هستند و چه تغییراتی داشته اند؟

فایل `03_stage3_join_and_metrics.py` یک اسکریپت پایتون است که برای ادغام داده‌های مرحله دوم (خلاصه‌سازی AOI و فازها (با اطلاعات پاسخ‌های شرکت‌کنندگان) از فایل‌های (JSON و تولید خلاصه‌های مختلف استفاده می‌شود. این اسکریپت سه فایل CSV به عنوان خروجی تولید می‌کند:

1. `stage3_per_question.csv`:

○ **فیلدها:** این فایل شامل اطلاعات مربوط به هر پرسش برای هر شرکت‌کننده است. فیلدهای آن شامل موارد زیر است:

- `participant`: شناسه شرکت‌کننده
- `section`: بخش آزمایش
- `question_number`: شماره سؤال
- `question_id`: شناسه سؤال
- `chosen_option`: گزینه انتخاب‌شده توسط شرکت‌کننده
- `is_correct`: نشان‌دهنده صحیح یا غلط بودن پاسخ (NaN) اگر پاسخ صحیح وجود نداشته باشد)
- `time_spent`: زمان صرف‌شده برای پاسخ دادن به سؤال
- `rows_total`: تعداد کل ردیف‌ها در فایل داده‌های ردیابی چشم
- `rows_valid`: تعداد ردیف‌های معتبر در فایل داده‌های ردیابی چشم
- `duration_total_s`: مدت زمان کل پاسخگویی به سؤال (ثانیه)
- `duration_Q_s`: مدت زمان فاز سؤال (ثانیه)
- `duration_A_s`: مدت زمان فاز پاسخ (ثانیه)
- `phase_onset_s`: زمان شروع فاز پاسخ (ثانیه)
- `hits_timer`: تعداد دفعاتی که شرکت‌کننده به تایمر نگاه کرده است
- `hits_submit`: تعداد دفعاتی که شرکت‌کننده به دکمه ثبت نگاه کرده است
- `hits_answers`: تعداد دفعاتی که شرکت‌کننده به ناحیه پاسخ‌ها نگاه کرده است
- `hits_question`: تعداد دفعاتی که شرکت‌کننده به ناحیه سؤال نگاه کرده است
- `correct_samples`: تعداد نمونه‌های داده که در ناحیه پاسخ صحیح قرار دارند
- `wrong_samples`: تعداد نمونه‌های داده که در ناحیه پاسخ غلط قرار دارند

▪ has_correct_answer: نشان‌دهنده این که آیا سؤال پاسخ صحیح دارد یا خیر

▪ correct_option_id: شناسه گزینه صحیح

○ **تغییرات:** این فایل از ادغام فایل stage2_summary.csv که خود شامل داده‌های پردازش شده ردیابی چشم و AOI ها است (با فایل‌های answers.json (که شامل پاسخ‌های شرکت‌کنندگان است) ایجاد می‌شود. فیلد is_correct بر اساس مقایسه پاسخ انتخاب‌شده با پاسخ صحیح محاسبه می‌شود.

2. stage3_by_participant.csv:

○ **فیلدها:** این فایل شامل خلاصه‌ای از عملکرد هر شرکت‌کننده در کل آزمایش است. فیلدهای آن شامل موارد زیر است:

▪ participant: شناسه شرکت‌کننده

▪ n_questions: تعداد سؤالات پاسخ داده شده توسط شرکت‌کننده

▪ mean_phase_onset_s: میانگین زمان شروع فاز پاسخ (ثانیه)

▪ mean_duration_Q_s: میانگین مدت زمان فاز سؤال (ثانیه)

▪ mean_duration_A_s: میانگین مدت زمان فاز پاسخ (ثانیه)

▪ sum_hits_timer: مجموع تعداد دفعاتی که شرکت‌کننده به تایمر نگاه کرده است

▪ sum_hits_submit: مجموع تعداد دفعاتی که شرکت‌کننده به دکمه ثبت نگاه کرده است

▪ sum_hits_answers: مجموع تعداد دفعاتی که شرکت‌کننده به ناحیه پاسخ‌ها نگاه کرده است

▪ sum_hits_question: مجموع تعداد دفعاتی که شرکت‌کننده به ناحیه سؤال نگاه کرده است

▪ accuracy_overall: دقت کلی شرکت‌کننده (فقط برای سؤالات دارای پاسخ صحیح)

▪ accuracy_part1: دقت شرکت‌کننده در بخش اول آزمایش (فقط برای سؤالات دارای پاسخ صحیح)

▪ accuracy_part2: دقت شرکت‌کننده در بخش دوم آزمایش (فقط برای سؤالات دارای پاسخ صحیح)

▪ participant_num: شناسه عددی شرکت‌کننده

○ **تغییرات:** این فایل با گروه‌بندی داده‌ها بر اساس شرکت‌کننده و محاسبه میانگین‌ها و مجموع‌های مختلف ایجاد می‌شود. دقت کلی و دقت در هر بخش نیز محاسبه و اضافه می‌شوند.

3. stage3_by_question.csv:

○ **فیلدها:** این فایل شامل خلاصه‌ای از عملکرد شرکت‌کنندگان در هر سؤال است. فیلدهای آن شامل موارد زیر است:

▪ question_id: شناسه سؤال

- `n_participants`: تعداد شرکت‌کنندگانی که به سؤال پاسخ داده‌اند
- `accuracy_mean`: میانگین دقت در پاسخگویی به سؤال
- `mean_phase_onset_s`: میانگین زمان شروع فاز پاسخ برای سؤال (ثانیه)
- `mean_duration_Q_s`: میانگین مدت زمان فاز سؤال برای سؤال (ثانیه)
- `mean_duration_A_s`: میانگین مدت زمان فاز پاسخ برای سؤال (ثانیه)

○ **تغییرات**: این فایل با گروه‌بندی داده‌ها بر اساس سؤال و محاسبه میانگین‌های مختلف ایجاد می‌شود.

به طور خلاصه، اسکریپت `03_stage3_join_and_metrics.py` داده‌های ردیابی چشم و پاسخ‌های شرکت‌کنندگان را ترکیب کرده و خلاصه‌هایی در سطوح مختلف (پرسش، شرکت‌کننده و غیره) ایجاد می‌کند تا برای تحلیل‌های بعدی آماده شوند. تغییرات اصلی شامل اضافه کردن اطلاعات پاسخ‌ها (گزینه انتخاب‌شده، زمان صرف‌شده، صحت پاسخ) و محاسبه خلاصه‌های آماری مختلف است.

Stage 4:

کد 04_stage4_labeling.py دقیقاً چه کاری باید انجام بدهد؟

کد 04_stage4_labeling.py برای برچسب‌گذاری داده‌ها بر اساس زمان خواندن سؤال و با استفاده از قواعد مشخص‌شده، طراحی شده است. به طور خلاصه، این کد مراحل زیر را انجام می‌دهد:

1. بارگیری داده‌ها:

- فایل stage3_per_question.csv را که شامل اطلاعات مربوط به هر پرسش برای هر شرکت‌کننده است، بارگیری می‌کند.
- به صورت اختیاری، یک فایل time_caps_csv را بارگیری می‌کند که شامل محدودیت‌های زمانی برای هر سؤال است.

2. آماده‌سازی داده‌ها:

- ستون‌های مورد نیاز (مانند زمان خواندن سؤال، صحت پاسخ، شناسه سؤال و غیره) را شناسایی و نرمال‌سازی می‌کند.
- یک ستون a_ij ایجاد می‌کند که نشان‌دهنده صحت پاسخ است (1 برای درست، 0 برای غلط).
- یک ستون valid_data ایجاد می‌کند که نشان‌دهنده معتبر بودن داده‌ها است (بر اساس معیارهایی مانند وجود زمان خواندن معتبر، تعداد ردیف‌های معتبر و غیره).

3. اعمال محدودیت‌های زمانی (Time Caps):

- اگر فایل time_caps_csv ارائه شده باشد، محدودیت‌های زمانی را برای هر سؤال اعمال می‌کند.
- یک ستون is_censored ایجاد می‌کند که نشان می‌دهد آیا زمان خواندن سؤال از محدودیت زمانی تجاوز کرده است یا خیر.

4. محاسبه آستانه‌ها (Thresholds):

- برای هر سؤال، آستانه‌های پایین (LB) را بر اساس زمان خواندن سؤال برای همه شرکت‌کنندگان معتبر محاسبه می‌کند. این آستانه‌ها با استفاده از چارک اول (Q1) و دامنه بین چارکی (IQR) محاسبه می‌شوند.
- برای هر سؤال، آستانه‌های بالا (UF_C) را بر اساس زمان خواندن سؤال برای شرکت‌کنندگانی که پاسخ صحیح داده‌اند، محاسبه می‌کند. این آستانه‌ها نیز با استفاده از چارک‌ها و IQR محاسبه می‌شوند.
- در صورت تمایل، ردیف‌هایی که زمان پاسخگویی آن‌ها محدود شده است (is_censored=1) را از محاسبات آماری حذف می‌کند.

5. برچسب‌گذاری داده‌ها:

- بر اساس قواعد زیر، به هر ردیف یک برچسب اختصاص می‌دهد:
 - اگر هیچ پاسخ صحیح معتبری برای یک سؤال وجود نداشته باشد، برچسب NA_no_correct اختصاص داده می‌شود.

- اگر پاسخ صحیح باشد و زمان خواندن سؤال از آستانه بالا (UF_C) بیشتر باشد، یا اگر پاسخ غلط باشد، برچسب UP (مشکوک) اختصاص داده می‌شود.
- در غیر این صورت، برچسب NP (عادی) اختصاص داده می‌شود.

6. ذخیره نتایج:

- دو فایل CSV ایجاد می‌کند:
- stage4_labels_per_question.csv: شامل داده‌های برچسب‌گذاری شده به همراه ستون‌های مربوطه.
- stage4_thresholds_by_question.csv: شامل آستانه‌های محاسبه شده برای هر سؤال.

به طور خلاصه، هدف این اسکریپت شناسایی الگوهای غیرعادی در زمان خواندن سؤالات است. شرکت‌کنندگانی که زمان زیادی را صرف خواندن سؤالات می‌کنند (به خصوص اگر پاسخ آن‌ها صحیح باشد) یا کسانی که پاسخ غلط می‌دهند، ممکن است رفتارهای مشکوکی داشته باشند که با برچسب UP مشخص می‌شوند. این برچسب‌ها می‌توانند در تحلیل‌های بعدی برای شناسایی تقلب یا رفتارهای غیرمعمول مورد استفاده قرار گیرند.

با توجه به کد `stage4_labeling.py_04` فیلدهای دوفایل `csv` خروجی رو به طور دقیق توضیح بده

با توجه به کد `04_stage4_labeling.py`، در اینجا توضیح دقیقی از فیلدهای دو فایل `CSV` خروجی `stage4_labels_per_question.csv` و `stage4_thresholds_by_question.csv` ارائه شده است:

1. stage4_labels_per_question.csv:

این فایل شامل اطلاعات مربوط به هر پرسش برای هر شرکت‌کننده است، به همراه برچسبی که نشان‌دهنده وضعیت آن پرسش است (مشکوک، عادی یا غیرقابل ارزیابی).

- **participant:** شناسه شرکت‌کننده.
- **section:** بخش آزمایش (به عنوان مثال، Part1، Part2).
- **question_number:** شماره سؤال در بخش مربوطه.
- **question_id:** شناسه یکتای سؤال.
- **chosen_option:** گزینه انتخاب‌شده توسط شرکت‌کننده (اگر در داده‌ها موجود باشد).
- **is_correct:** نشان‌دهنده صحیح یا غلط بودن پاسخ (اگر در داده‌ها موجود باشد).
- **a_ij:** مقدار 0 یا 1 که نشان‌دهنده غلط یا درست بودن پاسخ است.
- **t_ij:** زمان صرف‌شده برای خواندن سؤال (مدت زمان فاز سؤال).
- **valid_data:** مقدار بولی که نشان می‌دهد آیا داده‌های مربوط به این ردیف معتبر هستند یا خیر (بر اساس معیارهایی مانند وجود زمان خواندن معتبر، تعداد ردیف‌های معتبر و غیره).
- **valid_time:** مقدار بولی که نشان می‌دهد آیا زمان خواندن سؤال معتبر است یا خیر (بر اساس آستانه پایین LB).
- **is_censored:** مقدار 0 یا 1 که نشان می‌دهد آیا زمان خواندن سؤال بیشتر از یک محدودیت زمانی (time cap) بوده است یا خیر.
- **time_cap_s:** محدودیت زمانی برای سؤال (اگر اعمال شده باشد).
- **rows_total:** تعداد کل ردیف‌ها در فایل داده‌های ردیابی چشم مربوط به این سؤال.
- **rows_valid:** تعداد ردیف‌های معتبر در فایل داده‌های ردیابی چشم مربوط به این سؤال.
- **duration_total_s:** مدت زمان کل پاسخگویی به سؤال (ثانیه).
- **LB:** آستانه پایین (Lower Bound) برای زمان خواندن سؤال (بر اساس چارک اول و IQR).
- **UF_C:** آستانه بالا (Upper Fence for Correct answers) برای زمان خواندن سؤال (بر اساس چارک سوم و IQR، فقط برای پاسخ‌های صحیح).
- **label:** برچسب اختصاص داده شده به این ردیف، که می‌تواند یکی از مقادیر زیر باشد:

- UP: مشکوک - (Unusual Pattern) زمانی که پاسخ صحیح است و زمان خواندن از UF_C بیشتر است، یا زمانی که پاسخ غلط است.
- NP: عادی - (Normal Pattern) زمانی که پاسخ صحیح است و زمان خواندن از UF_C بیشتر نیست.
- NA_no_correct: غیرقابل ارزیابی - زمانی که هیچ پاسخ صحیح معتبری برای این سؤال وجود ندارد.

2. stage4_thresholds_by_question.csv:

این فایل شامل آستانه‌های محاسبه شده برای هر سؤال است.

- **question_id:** شناسه یکتای سؤال.
 - **n_all:** تعداد کل شرکت‌کنندگانی که به این سؤال پاسخ داده‌اند و داده‌های معتبر دارند.
 - **Q1:** چارک اول (Q1) زمان خواندن سؤال برای همه شرکت‌کنندگان معتبر.
 - **median:** میانه (Median) زمان خواندن سؤال برای همه شرکت‌کنندگان معتبر.
 - **Q3:** چارک سوم (Q3) زمان خواندن سؤال برای همه شرکت‌کنندگان معتبر.
 - **IQR:** دامنه بین چارکی (Interquartile Range) زمان خواندن سؤال برای همه شرکت‌کنندگان معتبر (Q3 - Q1).
 - **LB:** آستانه پایین (Lower Bound) برای زمان خواندن سؤال. $(Q1 - 1.5 * IQR)$
 - **n_correct_valid:** تعداد شرکت‌کنندگانی که پاسخ صحیح داده‌اند و داده‌های معتبر دارند (و در صورت فعال بودن، سانسور نشده‌اند).
 - **Q1_C:** چارک اول (Q1) زمان خواندن سؤال برای شرکت‌کنندگانی که پاسخ صحیح داده‌اند.
 - **median_C:** میانه (Median) زمان خواندن سؤال برای شرکت‌کنندگانی که پاسخ صحیح داده‌اند.
 - **Q3_C:** چارک سوم (Q3) زمان خواندن سؤال برای شرکت‌کنندگانی که پاسخ صحیح داده‌اند.
 - **IQR_C:** دامنه بین چارکی (Interquartile Range) زمان خواندن سؤال برای شرکت‌کنندگانی که پاسخ صحیح داده‌اند. $(Q3_C - Q1_C)$
 - **UF_C:** آستانه بالا (Upper Fence) برای زمان خواندن سؤال برای شرکت‌کنندگانی که پاسخ صحیح داده‌اند $(Q3_C + 1.5 * IQR_C)$
- این توضیحات به شما کمک می‌کنند تا درک کاملی از فیلدهای موجود در فایل‌های خروجی stage4_labels_per_question.csv و stage4_thresholds_by_question.csv داشته باشید.