

Case Study

December 22, 2021

1 資料說明

以下有五份資料的說明，分別是：subscriptions.csv、state_change.csv、plans.csv、user_missions.csv、answers.csv。

1.1 subscriptions.csv

檔案中有 5936 列，每列有 6 個欄位。每一列代表一張訂單的相關資訊，各欄位說明如下：

- id：每一張訂單的 id，為一個字串。
- subscribe_plan_id：訂單所使用的方案 id，為一個字串。
- payer_id：此訂單付款人的 id，為一個字串。
- subscriber_id：此訂單對應遊戲使用者的 id，為一個字串。
- created_at：一個 yyyy-mm-dd hh:mm:ss 的字串，代表該訂單成立的時間點，包含日期與時間。
- ended_at：一個 yyyy-mm-dd hh:mm:ss 的字串，代表該訂單到期的時間點，包含日期與時間。

1.2 state_change.csv

檔案中有 10215 列，每列有 4 個欄位。每一列代表一張訂單在某個時間點的狀態變化，各欄位說明如下：

- id：每一張訂單的 id，為一個字串。
- from_state：變化前的狀態，有 wait_purchase、paid、canceled 三種可能。
- to_state：變化後的狀態，有 paid、canceled 兩種可能。
- created_at：一個 yyyy-mm-dd hh:mm:ss 的字串，代表訂單狀態變化的時間點，包含日期與時間。

1.3 plans.csv

檔案中有 33 列，每列有 2 個欄位。每一列代表一個方案的 id 與名稱，各欄位說明如下：

- id：每一個方案的 id，為一個字串。
- name：每一個方案的名稱，為一個字串。名稱中會恰好包含「國小」、「國中」、「高中」這三個詞彙之一。

1.4 user_missions.csv

檔案中有 441562 列，每列有 10 個欄位。每一列代表一位玩家對一個任務的作答記錄，各欄位說明如下：

- id：每一筆作答記錄的 id，為一個字串。
- user_id：每一位玩家的 id，為一個字串。
- mission_id：每一個任務的 id，為一個字串。
- status：玩家對任務的狀態。「已派發」代表系統已將該任務派給玩家，但玩家尚未接取該任務；「進行中」代表玩家已接取該任務但尚未完成；「已暫停」代表玩家接取該任務後，在任務尚未完成前主動暫停任務；「已完成」代表玩家已接取該任務，並完成該任務。
- correct_count：玩家作答該任務時的答對題數。
- answered_count：玩家作答任務時的作答題數，此數值必大於或等於 1，但可能不等於該任務中的題目總數量。
- answer_duration：玩家作答該任務的花費時間，單位為秒。
- started_at：一個 yyyy-mm-dd hh:mm:ss 的字串，代表玩家接取該任務的時間點，包含日期與時間。
- finished_at：一個 yyyy-mm-dd hh:mm:ss 的字串，代表玩家完成該任務的時間點，包含日期與時間。
- created_at：一個 yyyy-mm-dd hh:mm:ss 的字串，代表該任務派給玩家的時間點，包含日期與時間。

資料中的一些欄位可能出現空值的情形，取決於玩家對任務的狀態（status），對應關係如表 1 所示。

status	correct_count	answered_count	answer_duration	started_at	finished_at	created_at
已派發	空值	空值	空值	空值	空值	非空值
進行中	空值	空值	空值	非空值	空值	非空值
已暫停	空值	空值	空值	非空值	空值	非空值
已完成	非空值	非空值	非空值	非空值	非空值	非空值

表 1: 空值說明

1.5 answer.csv

檔案中有 3515994 列，每列有 14 個欄位，每一列代表玩家在一個題目中的作答資訊。各欄位說明如下：

- id：每一筆作答記錄的 id，為一個字串。
- user_id：每一位玩家各自的 id，為一個字串。
- question_id：每一道題目的 id，為一個字串。
- mission_id：若玩家正在作答某個任務中的題目，則此欄位記錄該任務的 id，為一個字串；若玩家並非作答任務中的題目，則此欄位為空值。
- section_name：該題目所屬的章節名稱。
- book_volume_name：該題目所屬的冊次名稱。
- subject_name：該題目所屬的科目名稱。
- publisher_name：該題目所屬的出版商名稱。
- course_name：玩家在作答時所在的世界名稱。
- is_correct：該題目是否作答正確，正確為 1，錯誤為 0。
- cost_time：玩家花費在該題目的時間總數（秒）。
- is_gc_attacked：是否有他人因為該玩家作答該題目而受到攻擊。
- battle_type：玩家作答該題目前選擇的遊戲效果。
- created_at：一個 yyyy-mm-dd hh:mm:ss 的字串，代表該作答記錄創造的時間，包含日期與時間。

2 分析工作

1. 針對每個任務（mission），計算完成比例（「已完成」佔相關資料筆數的比例）。
2. 針對每個用戶，計算其任務總完成率（「已完成」佔相關資料筆數的比例）。
3. 找出任務（mission）完成率（「已完成」的比例）最高、最低的那些用戶。
4. 看看每個用戶過往逐月的題目（question）回答數、答對率、所花的時間，來找出活躍用戶和潛在流失用戶。
5. 有些題目（question）說不定出得太難了，大部分同學第一次答都答錯。請看看有沒有這種題目的存在。