# PWS CUP 2016 トランザクションデータの書式 Ver. 0.1

匿名加工コンテストで提出する匿名加工トランザクションデータは、以下の書式に従う必要があります。従っていないデータは受け付けられません。

元（加工前）のデータは本書式に従っているので、参考にしてください。

## 全体的な書式

全体的な書式は次の通りです。

|  |  |
| --- | --- |
| 項目 | 書式 |
| ファイル形式 | CSV (ASCII)、二重引用符なし |
| 改行文字 | LF (\n) または CRLF (\r\n) |
| 行数 | 0行以上1,000,000 （百万）行以下 |
| 各行の文字数 | 80文字以下（改行文字は含まない） |

各行は6列（セル）で構成されます。つまり、各行にちょうど5つのカンマが含まれる必要があります。空行があってはいけません。最終行の改行はあってもなくても良いです。たとえば、最終行は

|  |
| --- |
| 12680,581587,2011/12/9 12:50,22138,4.95,3<LF>[EOF] |

でも

|  |
| --- |
| 12680,581587,2011/12/9 12:50,22138,4.95,3[EOF] |

でも良いです（<LF>は改行文字、[EOF]はファイルの終わりを示します）。

## 各列の書式

各列の書式は次の通りです。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 番号 | 名前 | 通称 | 書式パターン（正規表現） | 書式の概要 |
| 1 | CustomerID | 顧客ID | [1-9][0-9]{4} | 5桁の自然数 |
| 2 | InvoiceNo | 伝票ID | [1-9][0-9]{5} | 6桁の自然数 |
| 3 | InvoiceDate | 日時 | <後に掲載> | 2010年と2011年の YYYY/M/D h:mm 形式 |
| 4 | StockCode | 製品ID | BANK CHARGES|[a-zA-Z0-9]+ | BANK CHARGES または任意の英数文字列 |
| 5 | UnitPrice | 単価 | <後に掲載> | 整数部5桁以下、小数部2桁以下の正小数 |
| 6 | Quantity | 数量 | [1-9][0-9]{0,5} | 6桁以下の自然数 |

上記列番号3 (InvoiceDate) の書式パターン（正規表現）は次の通りです。

|  |
| --- |
| (?x) #コメントモード開始  201[01]/ #年  ( #月日開始  (0?[13578]|1[02])/(0?[1-9]|[1-2][0-9]|3[01])| #大の月  (0?[469]|11)/(0?[1-9]|[1-2][0-9]|30)| #小の月（2月を除く）  (0?2)/(0?[1-9]|1[0-9]|2[0-8]) #2月  )\x20 #月日終了、空白1つ  ([01]?[0-9]|2[0-3]):[0-5][0-9] #時分 |

上記列番号5 (UnitPrice) の書式パターン（正規表現）は次の通りです。

|  |
| --- |
| (?x) #コメントモード開始  0\.0[1-9]|0\.[1-9][0-9]?| #範囲 (0, 1)  [1-9][0-9]{0,4}(\.[0-9]{1,2})? #1以上 |

各列とも、値の前後に不要な空白があるだけでも不正になることに注意してください。

## 各列の例

1列目（CustomerID）の例を挙げます。

* ○ 99999
* × 01234 ∵ 0が先行してはいけません。

2列目（InvoiceNo）の例を挙げます。

* ○ 100000
* × 012345 ∵ 0が先行してはいけません。

3列目（InvoiceDate）の例を挙げます。

* ○ 2010/1/1 0:00
* ○ 2011/01/01 00:00
* × 2011/1/1 0:0 ∵ 分は2桁必要です。
* × 2010/2/29 0:00 ∵ 2/29は（2010年には）ありません。

4列目（StockCode）の例を挙げます。

* ○ 0
* ○ M
* ○ BANK CHARGES
* × S P ∵ BANK CHARGES以外は空白を含めてはいけません。

5列目（UnitPrice）の例を挙げます。

* ○ 99999.99
* ○ 1
* ○ 1.0
* ○ 1.00
* ○ 0.1
* × 01234 ∵ 整数部で0が先行してはいけません。
* × 0 ∵ 正数でなければいけません。

6列目（Quantity）の例を挙げます。

* ○ 999999
* × 012345 ∵ 0が先行してはいけません。

2016年7月1日 Ver. 0.1

PWS CUP実行委員会