データハンドリング実技試験

ver2021-10-14

# はじめに

表形式のデータを操作するスキルを測る実技試験です。

### 回答方法

* 後述する問題を解いて、答えを解答欄に書き込んでください。
* 集計に使ったExcelファイルまたはソースコードを提出してください。

### 使用できるツール

* 以下のいずれかのツールを使ってください。
  + Excel
  + プログラミング言語 (SQL, Python, Rのいずれか)

### 注意点

* この課題を他人に公開する事は禁止です。
* 回答を一部でも他人に手伝ってもらう事も禁止です。
* ただし、ツールの一般的な使用方法をWebや書籍などで調べる事はOKです。(積極的に調べてください。)

# データの説明

これから、架空の小売店チェーンのデータを集計してもらいます。

以下、4つのファイルに入ったデータを使ってください。  
各データの意味は以下の通りです。

### transactions.csv

商品が購入されたログです。  
商品が購入された数だけ行があります。

|  |  |
| --- | --- |
| 列名 | 意味 |
| basket\_id | 1回の購入ごとに割り振られるuniqueなID |
| shop\_id | 購入が行われた店舗を示すID |
| user\_id | 購入したユーザーを示すID |
| item\_id | 購入が行われた商品を示すID |
| buy\_date | 購入日 |

### items.csv

商品のマスターデータです。  
1商品につき1行のデータがあります。

|  |  |
| --- | --- |
| 列名 | 意味 |
| item\_id | 1つの商品ごとに割り振られるuniqueなID |
| item\_name | 商品名 |
| item\_price | 商品価格 |
| category\_lv1 | 商品大カテゴリ |
| category\_lv2 | 商品カテゴリ |

### shops.csv

店舗のマスターデータです。  
1店舗につき1行のデータがあります。

|  |  |
| --- | --- |
| 列名 | 意味 |
| shop\_id | 1つの店舗ごとに割り振られるuniqueなID |
| reginon | 店舗が出店している地域 |
| name | 店舗名 |

### users.csv

お客さまのマスターデータです。  
1顧客につき1行のデータがあります。

|  |  |
| --- | --- |
| 列名 | 意味 |
| user\_id | 1顧客ごとに割り振られるuniqueなID |
| name | 顧客の名前 |
| sex | 顧客の性別 |
| age | 顧客の年齢 |

# 問題1

2020年1月1日から12月31日までの1年間について、商品カテゴリ別に売上額と購入者数を計算してください。

その上で、売上額Top3の商品カテゴリと、その売上額、購入者数を抽出してください。

* 売上額は、商品価格を合算した物です。税などを考慮する必要はありません。
* 購入者数は、ある商品を購入した人を重複なく数えたユニークユーザー数の事です。
* 商品カテゴリは、category\_lv2を使ってください。

### 解答欄

|  |  |  |  |
| --- | --- | --- | --- |
| 順位 | カテゴリ名 | 売上額 | 購入者数 |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |

# 問題2

2020年1月1日から12月31日までの1年間について「バスボディケア用品」の売上額が最も大きかった店舗を1つ抽出してください。

さらに、その店舗の客単価と購入者数を月毎に計算してください。  
客単価と購入者数は、「全商品カテゴリ」と「バスボディケア用品のみ」の2パターンを計算してください。

* 客単価とは、購入者1人あたりの平均売上の事です。
* 商品カテゴリは、category\_lv2を使ってください。

### 解答欄

#### バスボディケア用品の売上が最も大きかった店舗

|  |
| --- |
| 店舗名 |
|  |

#### 客人数・客単価

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 月 | 購入者数 (全カテ) | 客単価 (全カテ) | 購入者数 (バスボディケア用品) | 客単価  (バスボディケア用品) |
| 2020年1月 |  |  |  |  |
| 2020年2月 |  |  |  |  |
| 2020年3月 |  |  |  |  |
| 2020年4月 |  |  |  |  |
| 2020年5月 |  |  |  |  |
| 2020年6月 |  |  |  |  |
| 2020年7月 |  |  |  |  |
| 2020年8月 |  |  |  |  |
| 2020年9月 |  |  |  |  |
| 2020年10月 |  |  |  |  |
| 2020年11月 |  |  |  |  |
| 2020年12月 |  |  |  |  |