



顧客の声

弊社は長年、ECサイトを運営しています。弊社のお客様情報は、すべてECサイトで管理しているので、多くのデータが集まっています。今、多くの企業はデータ分析で結果を出していると聞いているので、その流行りのデータ分析に、弊社も取り組んでいきたいと思っています。そうは言っても、多くの社員は文系畑出身で、データの分析というものに疎く、何から手をつけていけば良いのかわからないのです。現状では、当月の売上の数字を把握するので精一杯です。弊社のECサイトのデータ分析、お願いできないでしょうか。

前提条件

本章の10本ノックでは、ECサイトのデータを扱っていきます。

ここで扱うECサイトは、PCを商品として扱っています。それぞれ価格帯別に5商品が存在します。データは表に示した4種類6個のデータとなります。

customer_master.csvは、このECサイトを利用した会員の顧客情報となります。みなさんもECサイトなどを利用する際に、初期登録をされるかと思いますが、それらの情報が格納されています。item_master.csvは、取り扱い商品のデータとなっており、商品名や価格などが存在しています。

transaction_1.csv、transaction_2.csvは、購入明細のデータとなります。どの顧客が、いつ、どのくらい買ったのかなどの情報を持っています。transaction_detail_1.csv、transaction_detail_2.csvは、購入明細の詳細となっており、具体的になんの商品をいくつ買ったかの情報が入っています。データベースに入っているデータは、システムから落としてくる際に、毎月や1000件単位に分割されてしまうことがあります。その場合、結合してデータを使用する必要があります。今回の明細関連のデータもそれに該当し、1、2の2つが存在します。

■表1-1：データ一覧

No.	ファイル名	概要
1	customer_master.csv	顧客データ。名前、性別等

2	item_master.csv	取り扱っている商品データ。商品名、価格等
3-1	transaction_1.csv	購入明細データ
3-2	transaction_2.csv	3-1の続き。システムの都合上分割して出力
4-1	transaction_detail_1.csv	購入明細の詳細データ
4-2	transaction_detail_2.csv	4-1の続き。システムの都合上分割して出力

ノック1： データを読み込んでみよう

まずは、4種類のデータをそれぞれ読み込んで、先頭5行のみ表示し、データを眺めてみましょう。4種類のデータは、データベースから抽出したCSV形式のファイルとなります。

最初に、customer_master.csvをJupyter-Notebookで読み込んでみましょう。

```
import pandas as pd
customer_master = pd.read_csv('customer_master.csv')
customer_master.head()
```

■図1-1：customer_masterの読み込み結果

ノック1：データを読み込んでみよう										
In [1]: import pandas as pd customer_master = pd.read_csv('customer_master.csv') customer_master.head()										
Out[1]:										
	customer_id	customer_name	registration_date	customer_name_kana	email	gender	age	birth	pref	
0	H1606473	高木 ちえみ	2019-01-01 00:44:12	たかぎ ちえみ	takagi_chiemi@example.com	F	58	1960/10/13	神奈川県	
1	P1480849	安田 誠治	2019-01-01 03:46:10	やすだ せいじ	yasuda_seiji@example.com	M	64	1955/5/24	兵庫県	
2	AS483225	栗田 公顕	2019-01-01 04:04:17	しばた きみあき	shibata_kimio@example.com	M	63	1955/11/23	福岡県	
3	PL372642	宮脇 ひとみ	2019-01-01 07:04:59	みやわき ひとみ	miyawaki_hitomi@example.com	F	48	1970/9/11	東京都	
4	OA681421	徳田 晋也	2019-01-01 07:22:10	とくだ しんや	tokuta_shinya@example.com	M	34	1985/4/20	東京都	

1行目では、Pythonライブラリのpandasの読み込み、2行目では、pandasのread_csvを使用し、pandasのデータフレーム型として、変数customer_masterに格納しています。3行目で読み込んだcustomer_masterの先頭5行