第2章 小売店のデータでデータ加工を行う10本ノック

これでクレンジングと整形まで終わったデータが無事ダンプされました。

元のデータに変更があった場合は、再度データ加工処理を行い直す必要がありますが、元のデータが変わらない限りは、このファイルを読み込んで分析する事ができます。

さっそくダンプファイルを読み込んで、集計を行ってみましょう。

ノック20:データを集計しよう

まずは、ダンプファイルを読み込みます。

import_data = pd.read_csv("dump_data.csv")
import_data

■図2-26: インポート結果

(27)										
		purchase_date	purchase_month	item_name	item_price	顧客名	かな	地域	メールアドレス	登録日
	0	2019-06-13 18:02:34	201906	商品A	100.0	深井菜で養	ふかい ななみ	C市	fukai_nanami@exampte.com	2017-01-26 00 00 00
	1	2019-07-13 13:05:29	201907	意品の	1900.0	浅田賀二	あさだ けんじ	C市	asada_kenji@example.com	2018-04-07 00:00:00
	2	2019-05-11 19:42:07	201905	商品A	100.0	南部庚二	なんぶ けいじ	ΑĦ	nannbu_keiji@exampte.com	2018-06-19 00:00:00
	3	2019-02-12 23:40:45	201902	商品Z	2600 0	聚生斯维	あそう りお	市	asou_rio@example.com	2018-07-22 00:00:00
	4	2019-04-22 03.09:35	201904	度品A	100 0	平田鉄二	ひらた てつじ	D市	hirata_tetsuji@exampte.com	2017-06-07 00:00:00
	5	2019-03-20 19:15:01	201903	惹品S	1900.0	海江街	ほりえたすく	H市	horie_tasuku@example.com	2018-95-14 00:00:00
	6	2019-05-18	201905	商品A	100.0	深井勝生	ふかいてるお	ΑĦ	fukal_teruo@exampte.com	2018-02-21

もはや説明するまでもありませんが、read_csvにてダンプデータを読み込んでいます。

続いて、purchase_monthを縦軸に、商品毎の集計を行いましょう。

byItem = import_data.pivot_table(index="purchase_month", columns="item_nam
e", aggfunc="size", fill_value=0)
byItem

■図2-27:購入年月、商品の集計結果

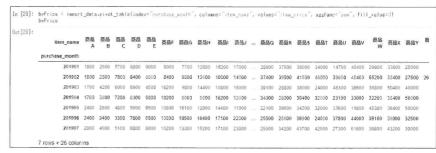
սէ[28]։	item_name	商品	商品	商品 C	商品	商品	商品 F	商品 G	商品	商品	商品		商品 Q	商品	商品	商品	商品	商品V	商品W	商品 X	商品	商品
	purchase_month	~	b		D		, F	G	n		J		u	K	3		U	٧		^		2
	201901	18	13	19	17	18	15	11	16	18	17	111	17	21	20	17	7	22	13	14	10	6
	201902	19	14	26	21	16	14	14	17	12	14	7.14	22	22	22	23	19	22	24	16	11	1
	201903	17	21	20	17	9	27	14	18	12	16		23	16	20	12	23	18	16	21	16	0
	201904	17	19	24	20	18	17	14	15	18	13	100	20	20	16	16	11	15	14	16	20	0
	201905	24	14	16	14	19	18	23	15	16	11		13	22	18	16	16	9	21	16	20	0
	201906	24	12	11	19	13	18	15	13	19	22	911	15	16	21	12	18	20	17	15	13	0
	201907	20	20	17	17	12	17	19	19	19	23		15	19	23	21	13	28	16	18	12	0

下部に「26 columns」とあるように、正しく商品A~Zの26商品毎の購入年月の集計が行えました。

続いて、purchase_monthを縦軸に、売上金額、顧客、地域の集計を行いましょう。

byPrice = import_data.pivot_table(index="purchase_month", columns="item_na
me", values="item_price", aggfunc="sum", fill_value=0)
byPrice

■図2-28:購入年月、売上金額の集計結果



byCustomer = import_data.pivot_table(index="purchase_month", columns="顧客名", aggfunc="size", fill_value=0)
byCustomer