■図2-29:購入年月、各顧客の購入数の集計結果

l:	顧客名	さだチ	中仁品	中田美智子	丸山 光臣	久保田 倫子	亀井 一徳	五十萬 春樹	井上 桃子	井口	井川真 悠子	***	善推 盛一	高原充則	高梨結衣	高沢美咲	高田さ んま	島居広司	鶴岡薫	麻生 料報	英川田 博之	黒谷長利
purchase	_month																					
	201901	3.	1	4	2	2	0	5	3	3	1	17	D	d	11	1	5	2	0	2	2	5
	201902	9	1	2	2	2	4	2	1	0	4		4	0	3	2	0	1	2	4	0	1
	201903	1	2	1	6	1	4	3	3	2	2		3	4	6	2	4	2	4	2	2	1
	201904	0	3	1	2	0	2	2	0	3	2		2	4	2	3	4	3	2	1	2	0
	201905	3	2	5	2	4	1	2	Ť	3	3		1	1	1	0	2	2	3	4	4	,
	201906	1	3	0	4	1	1	1	2	2	3		7	3	0	2	1	0	2	1	2	4
	201907	3	0	3	2	5	3	5	- 2	5	5		2	4	4	2	0	2	4	3	-4	-1

byRegion = import_data.pivot_table(index="purchase_month", columns="地域", aggfunc="size", fill_value=0)
byRegion

■図2-30:購入年月、地域における販売数の集計結果

U	Region									
[31]:										
	地域	Аm	日市	C市	助	E市	Frib	Giff	Hт	
	purchase_month									
	201901	59	55	72	34	49	57	49	42	
	201902	71	46	65	48	61	52	43	63	
	201903	64	52	57	43	52	59	51	59	
	201904	64	48	54	45	48	58	40	52	
	201905	57	52	68	48	59	65	35	43	
	201906	53	47	61	30	51	51	58	58	
	201907	76	53	61	42	54	64	47	54	

最後に、集計期間で購入していないユーザーがいないかチェックしてみましょう。

away_data = pd.merge(uriage_data, kokyaku_data, left_on="customer_name", r ight_on="顧客名", how="right")

away_data[away_data["purchase_date"].isnull()][["顧客名", "メールアドレス", "登録日"]]

■図2-31:集計期間内での離脱顧客

売上履歴と顧客台帳をライトジョインし、顧客台帳を主体として結合してみま しょう。

集計期間内に購買を行っていない顧客は「購買日」等がNaNで結合されます。 それを条件にデータを抽出する事で、集計期間に購入を行っていない顧客を調 べる事が可能です。



これで、第2章の10本ノックは終了です。

データ分析を行うにあたり、とても大事なデータ加工の基礎について学んできました。冒頭にもあります通り、データの状態を見極め、どのように加工するかを現場の方と密に相談しながらすすめ、的確にデータ加工を行う必要があります。ここを疎かにすると後でとても痛い目にあうので、今回のような基礎的なクレンジングだけではなく、複雑なデータのクレンジングに挑戦してみてください。本章がそのきっかけとなれば幸いです。

次章からは、より高度で具体的な分析に踏み込んでいきます。