

Exercise 5.4

Dan Franklin

Cleaned dataset

	A	B	C	D	E	F	G	H	I	J	K	L
1	Customer_ID	Credit Score	Country	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard?	IsActiveMember	Estimated Salary	ExitedFromBank?
2	15634602	619	France	Female	42	2	\$0.00	1	1	1	\$101,348.88	1
3	15647311	608	Spain	Female	41	1	\$83,807.8	1	0	1	\$112,542.58	0
4	15619304	502	France	Female	42	8	\$159,660	3	1	0	\$113,931.57	1
5	15701354	699	France	Female	39	1	\$0.00	2	0	0	\$93,826.63	0
6	15737888	850	Spain	Female	43	2	\$125,510	1	1	1	\$79,084.10	0
7	15574012	645	Spain	Male	44	8	\$113,755	2	1	0	\$149,756.71	1
8	15592531	822	France	Male	50	7	\$0.00	2	1	1	\$10,062.80	0
9	15656148	376	Germany	Female	29	4	\$115,046	4	1	0	\$119,346.88	1
10	15792365	501	France	Male	44	4	\$142,051	2	0	1	\$74,940.50	0
11	15592389	684	France	Male	27	2	\$134,603	1	1	1	\$71,725.73	0
12	15767821	528	France	Male	31	6	\$102,016	2	0	0	\$80,181.12	0
13	15737173	497	Spain	N/A	24	3	\$0.00	2	1	0	\$76,390.01	0
14	15632264	476	France	Female	34	10	\$0.00	2	1	0	\$26,260.98	0
15	15691483	549	France	Female	25	5	\$0.00	2	0	0	\$190,857.79	0
16	15600882	635	Spain	Female	35	7	\$0.00	2	1	1	\$65,951.65	0
17	15643966	616	Germany	Male	45	3	\$143,129	2	0	1	\$64,327.26	0
18	15737452	653	Germany	Male	58	1	\$132,602	1	1	0	\$5,097.67	1
19	15788218	549	Spain	Female	24	9	\$0.00	2	1	1	\$14,406.41	0
20	15661507	587	Spain	Male	45	6	\$0.00	1	0	0	\$158,684.81	0
21	15568982	726	France	Female	24	6	\$0.00	2	1	1	\$54,724.03	0
22	15577657	732	France	Male	41	8	\$0.00	2	1	1	\$170,886.17	0
23	15597945	636	Spain	Female	32	8	\$0.00	2	1	0	N/A	0
24	15699309	510	Spain	Female	N/A	4	\$0.00	1	1	0	\$118,913.53	1
25	15725737	669	France	Male	46	3	\$0.00	2	0	1	\$8,487.75	0
26	15625047	846	France	Female	38	5	\$0.00	1	1	1	\$187,616.16	0
27	15738191	577	France	Male	25	3	\$0.00	2	0	1	\$124,508.29	0
28	15736816	756	Germany	Male	36	2	\$136,815	1	1	1	\$170,041.95	0
29	15700772	571	France	Male	44	9	\$0.00	2	0	0	\$38,433.35	0
30	15732603	534	Germany	Female	43	3	\$141,340	1	1	1	\$189,187.43	0

Original dataset
Ex-customers
Remaining customers
clean dataset
Changes

Changes Table

	A	B	C	D
1				
2	Column Name	Issues	Actions taken	Notes
3	Row_Number	Not necessary	Deleted	not needed for analysis
4	Customer_ID			
5	Last_Name	Contains PII	Deleted	not needed for analysis
6	Credit Score	contains blanks	replaced blanks with NULL	
7	Country	contained countries in 2 formats eg FR and France	replaced with text rather than letters	
8	Gender	Genders listed in mixed formats	replaced with text rather than letters	
9	Age	1 NULL, and 11 entries under 2	replaced all with N/A	
10	Tenure			
11	Balance			
12	NumOfProducts			
13	HasCrCard?			
14	IsActiveMember			
15	Estimated Salary	1 NULL and 1 blank entry	changed both to N/A	
16	ExitedFromBank?			
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				

Original dataset
Ex-customers
Remaining customers
clean dataset
Changes

Ex-customers information

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Customer ID	Credit Score	Country	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	Estimated Salary	ExitedFromBank?						
15634602	619	France	Female	42	2	USD 0.00	1	1	1	101,348.88	1		Row Labels	Count of Gender		Average of Age	
15619304	502	France	Female	42	8	USD 159,660.80	3	1	0	113,931.57	1		Female			45.27722772	
15574012	645	Spain	Male	44	8	USD 113,755.78	2	1	0	149,756.71	1		Male				
15656148	376	Germany	Female	29	4	USD 115,046.74	4	1	0	119,346.88	1		Grand Total	204		Average of Credit Score	
15737452	653	Germany	Male	58	1	USD 132,602.88	1	1	0	5,097.67	1					636.5123153	
15699309	510	Spain	Female	N/A	4	USD 0.00	1	1	0	118,913.53	1		Row Labels	Count of Country		Average of NumOfProducts	
15589475	591	Spain	Female	39	3	USD 0.00	3	1	0	140,469.38	1		France				
15794171	475	France	Female	45	0	USD 134,264.04	1	1	0	27,822.99	1		Germany			1.460784314	
15738148	465	France	Female	51	8	USD 122,522.32	1	0	0	181,297.65	1		Spain				
15755196	834	France	Female	49	2	USD 131,394.56	1	0	0	194,365.76	1		Grand Total	204		Row Labels	Count of HasCrCard?
15602280	829	Germany	Female	27	9	USD 112,045.67	1	1	1	119,708.21	1					0	60
15771573	637	Germany	Female	39	9	USD 137,843.80	1	1	1	117,622.80	1		Row Labels	Count of IsActiveMember			144
15702298	655	Germany	Male	41	8	USD 125,561.97	1	0	0	164,040.94	1		0			Grand Total	204
15659590	601	Germany	Male	42	1	USD 98,495.72	1	1	0	40,014.76	1		1				
15623944	511	Spain	Female	66	4	USD 0.00	1	1	0	1,643.11	1		Grand Total	204		Average of Tenure	
15703793	738	Germany	Male	58	2	USD 133,745.44	4	1	0	28,373.86	1					4.71568275	
15663706	777	France	Female	32	2	USD 0.00	1	1	0	136,458.19	1		Average of Balance				
15762418	750	Spain	Male	22	3	USD 121,681.82	1	1	0	128,643.35	1		90239.21887				
15622897	646	France	Female	46	4	USD 0.00	3	1	0	93,251.42	1						
15757535	647	Spain	Female	44	5	USD 0.00	3	1	1	174,205.22	1		Average of Est				
15804919	670	Spain	Female	65	1	USD 0.00	1	1	1	177,655.68	1		97155.20461				
15613854	622	Spain	Female	46	4	USD 107,073.27	2	1	1	30,984.59	1						
15744689	479	Germany	Male	35	9	USD 92,833.89	1	1	0	99,449.86	1						
15609618	721	Germany	Male	28	9	USD 154,475.54	2	0	1	101,300.94	1						
15661670	524	Germany	Female	31	8	USD 107,818.63	1	1	0	199,725.39	1						
15627360	432	France	Male	42	9	USD 152,603.45	1	1	0	110,265.24	1						
15671137	549	France	Female	52	1	USD 0.00	1	0	1	8,636.05	1						
15782688	625	Germany	Male	56	0	USD 148,507.24	1	1	0	46,824.08	1						
15594408	584	Spain	Female	48	2	USD 213,146.20	1	1	0	75,161.25	1						
15640905	579	Spain	Female	35	1	USD 129,490.36	2	0	1	8,590.83	1						

Remaining customers information

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	Customer_ID	Credit Score	Country	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard?	IsActiveMember	Estimated Salary	ExitedFromBank?							
1	15647311	608	Spain	Female	41	1	83,807.86	1	0	1	112,542.58	0		Row Labels	Count of Gender		Average of Age		
2	15701354	699	France	Female	39	1	0	2	0	0	93,826.63	0		Female			341	37.49034749	
3	15737888	850	Spain	Female	43	2	125,510.82	1	1	1	79,084.10	0		Male			445		
4	15592531	822	France	Male	50	7	0	2	1	1	10,062.80	0		N/A			1	Average of Credit Score	
5	15792365	501	France	Male	44	4	142,051.07	2	0	1	74,940.50	0		Grand Total	787		651.6152866		
6	15592389	684	France	Male	27	2	134,603.88	1	1	1	71,725.73	0							
7	15767821	528	France	Male	31	6	102,016.72	2	0	0	80,181.12	0		Row Labels	Count of Country		Average of NumOfProducts		
8	15737173	497	Spain	N/A	24	3	0	2	1	0	76,390.01	0		France			403	1.538754765	
9	15632264	476	France	Female	34	10	0	2	1	0	26,280.98	0		Germany			182		
10	15691483	549	France	Female	25	5	0	2	0	0	190,857.79	0		Spain			202	Row Labels	Count of HasCrCard?
11	15600882	635	Spain	Female	35	7	0	2	1	1	65,551.65	0		Grand Total	787		0	231	
12	15643966	616	Germany	Male	45	3	143,129.41	2	0	1	64,327.26	0					1	556	
13	15788218	549	Spain	Female	24	9	0	2	1	1	14,406.41	0		Row Labels	Count of IsActiveMember		Grand Total	787	
14	15661507	587	Spain	Male	45	6	0	1	0	0	158,684.81	0		0			345		
15	15588882	726	France	Female	24	6	0	2	1	1	54,724.03	0		1			442		
16	15577657	732	France	Male	41	8	0	2	1	1	170,886.17	0		Grand Total	787				
17	15597945	636	Spain	Female	32	8	0	2	1	0	N/A	0						Average of Tenure	
18	15725737	669	France	Male	46	3	0	2	0	1	8,487.75	0		Average of Balance			5.157560356		
19	15625047	846	France	Female	38	5	0	1	1	1	187,616.16	0		74830.86779					
20	15738191	577	France	Male	25	3	0	2	0	1	124,508.29	0							
21	15738816	756	Germany	Male	36	2	136,815.64	1	1	1	170,041.35	0		Average of Estimated					
22	15700772	571	France	Male	44	9	0	2	0	0	38,433.35	0		98943.39019					
23	15728693	574	Germany	Female	43	3	141,349.43	1	1	1	100,187.43	0							
24	15656300	411	France	Male	29	0	59,697.17	2	1	1	53,483.21	0							
25	15706552	533	France	Male	36	7	85,311.70	1	0	1	156,731.91	0							
26	15750181	553	Germany	Male	41	9	110,112.54	2	0	0	81,898.81	0							
27	15659428	520	Spain	Female	42	6	0	2	1	1	34,410.55	0							
28	15732963	722	Spain	Female	29	9	0	2	1	1	142,033.07	0							
29	15788448	490	Spain	Male	31	3	145,260.23	1	0	1	114,066.77	0							
30	15725999	804	Spain	Male	33	7	76,548.60	1	0	1	98,453.45	0							
31	15717426	850	France	Male	36	7	0	1	1	1	40,812.90	0							
32	15585768	582	Germany	Male	41	6	70,349.48	2	0	1	178,074.04	0							
33	15619360	472	Spain	Male	40	4	0	1	1	0	70,154.22	0							
34	Original dataset		Ex-customers		Remaining customers			clean dataset		Changes		⊕							

Below is the decision tree based on my findings whilst cleaning the data and mining it for info. I saw patterns that suggested active customers are more likely to stay, females more like to leave, as well as people aged under 40. Interestingly, a much smaller percentage of French customers leave than the Germans or Spanish do.

I also looked at salaries, balances, and tenure (which I assumed to mean how long they are with the bank). What varied most was balances, with those remaining having on average a lower balance than those that leave – perhaps they take their money elsewhere to try and make more from it suggesting interest rates at Pig E. Bank might be lower than competitors. Average tenure suggests that those who have banked with them for over 5 years are more likely to stay, but the average tenure of those that left was 4.7 years, so not much different.



