

Data Cleaning

1. Import data.
2. Clean Format. Align data. "Home"-> "Format"-> "AutoFill Column Width"
3. Used function "PROPER" to make all the first letters in the data frame upper case for consistency and good look for columns B,C,D,E,G,H,I,K,P,Q (cells' number 2:4526). Only values are pasted with a special formula "Paste"->" Values".
4. Remove blank lines. Copied column D in a new column. Column D "TRIM(PROPER())" function applied. Only values are pasted with a special formula "Paste"->" Values".
5. Remove redundancy. "Data" -> "Remove Duplicates". Four duplicates removed.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Age Job	Marital	Education	Default Balance	Housing	Loan	Contact	Day	Month	Duration	Campaign	Pdays	Previous	Poutcome	Y							
2	30 Unemployed	Married	Primary	No	1787	No	No	Cellular	19 Oct	79	1	-1	0	Unknown	No							
3	33 Services	Married	Secondary	No	4789	Yes	Yes	Cellular	11 May	220	1	339	4	Failure	No							
4	35 Management	Single	Tertiary	No	1350	Yes	No	Cellular	16 Apr	185	1	330	1	Failure	No							
5	30 Management	Married	Tertiary	No	1476	Yes	Yes	U							known	No						
6	59 Blue-Collar	Married	Secondary	No	0	Yes	No	U							known	No						
7	35 Management	Married	Secondary	No	747	No	No	C							pre	No						
8	36 Self-Employed	Married	Tertiary	No	307	Yes	No	C							er	No						
9	39 Technician	Married	Secondary	No	147	Yes	No	C							known	No						
10	41 Entrepreneur	Married	Tertiary	No	221	Yes	No	U							known	No						
11	43 Services	Married	Primary	No	-88	Yes	Yes	Cellular	17 Apr	313	1	147	2	Failure	No							
12	39 Services	Married	Secondary	No	9374	Yes	No	Unknown	20 May	273	1	-1	0	Unknown	No							
13	43 Admin.	Married	Secondary	No	264	Yes	No	Cellular	17 Apr	113	2	-1	0	Unknown	No							
14	36 Technician	Married	Tertiary	No	1109	No	No	Cellular	13 Aug	328	2	-1	0	Unknown	No							
15	20 Student	Single	Secondary	No	502	No	No	Cellular	30 Apr	261	1	-1	0	Unknown	Yes							
16	31 Blue-Collar	Married	Secondary	No	360	Yes	Yes	Cellular	29 Jan	89	1	241	1	Failure	No							
17	40 Management	Married	Tertiary	No	194	No	Yes	Cellular	29 Aug	189	2	-1	0	Unknown	No							
18	56 Technician	Married	Secondary	No	4073	No	No	Cellular	27 Aug	239	5	-1	0	Unknown	No							
19	37 Admin.	Single	Tertiary	No	2317	Yes	No	Cellular	20 Apr	114	1	152	2	Failure	No							
20	25 Blue-Collar	Single	Primary	No	-221	Yes	No	Unknown	23 May	250	1	-1	0	Unknown	No							
21	31 Services	Married	Secondary	No	132	No	No	Cellular	7 Jul	149	1	163	1	Other	No							

Data grouping and summarizing

Data Filtering

1. Ctrl+A. Format as a Table. Green Light.
2. Filter "Success" outcomes. 129 results found. 82-Married, 10-Divorsed, 37-Single. Therefore, success of the campaign is more likely for married individuals.
3. Success outcome has less contacts during the campaign 5-1, while failure outcome has more contacts during the campaign, mostly 8-4 out of 8-1.
4. Entrepreneurs are more likely to have a failure outcome from contacts during the campaign 1:15 ratio success: failure, while all other professions have steady ratio 1:2 or 1:4 of success: failure outcome from a contacts during the campaign.
5. There are a lot of "other" and "unknown" outcomes from a campaign total of 3902 in comparison to "failure" and "success" total of 619. Therefore, we have lot of missing data about campaign outcomes.

Assignment 3

Color scale or data bar to explore and highlight the data

1. From the color formatting it is obvious that people with average negative yearly balances are more likely to have failure outcome from the campaign.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
28	48	Management	Married	Tertiary	No	1175	Yes	No	Telephone	18	Nov	1476	3	-1	0	Unknown	No			
29	40	Management	Single	Tertiary	No	1361	No	No	Cellular	29	Jan	1473	1	-1	0	Unknown	Yes			
30	38	Services	Married	Secondary	No	0	No	Yes	Cellular	16	Jul	1473	6	-1	0	Unknown	No			
31	36	Unemployed	Married	Secondary	No	-872	Yes	Yes	Cellular	20	Nov	153	1	183	1	Failure	No			
32	56	Admin.	Married	Secondary	No	45	No	No	Unknown	5	May	1467	1	-1	0	Unknown	Yes			
33	26	Unemployed	Single	Secondary	No	622	No	No	Cellular	28	Jan	1451	2	-1	0	Unknown	Yes			
34	33	Admin.	Married	Secondary	No	0	Yes	No	Telephone	17	Jul	1448	7	-1	0	Unknown	Yes			
35	31	Management	Married	Tertiary	No	1224	Yes	Yes	Cellular	24	Jul	1441	2	-1	0	Unknown	Yes			
36	35	Self-Employed	Single	Secondary	No	1732	No	No	Telephone	21	Nov	1441	3	-1	0	Unknown	No			
37	39	Technician	Married	Secondary	No	-735	Yes	No	Cellular	30	Jan	144	5	248	2	Failure	No			
38	30	Blue-Collar	Single	Secondary	No	315	No	No	Cellular	23	Jun	1407	1	-1	0	Unknown	Yes			
39	60	Retired	Married	Secondary	No	197	No	No	Cellular	7	Oct	1386	1	-1	0	Unknown	No			
40	37	Technician	Single	Tertiary	No	6101	No	No	Cellular	4	May	60	5	87	1	Success	No			
41	29	Technician	Married	Tertiary	No	858	Yes	No	Cellular	12	Jun	234	7	330	1	Failure	No			
42	42	Management	Married	Primary	No	-679	Yes	Yes	Cellular	6	May	159	1	343	2	Failure	No			
43	51	Self-Employed	Married	Secondary	No	95	Yes	No	Cellular	23	Jul	1309	4	-1	0	Unknown	Yes			
44	32	Management	Divorced	Tertiary	No	874	No	No	Cellular	12	Aug	1282	6	-1	0	Unknown	Yes			
45	32	Blue-Collar	Single	Secondary	No	118	Yes	No	Cellular	17	Jul	1275	1	-1	0	Unknown	Yes			
46	46	Blue-Collar	Married	Secondary	No	668	Yes	No	Unknown	15	May	1263	2	-1	0	Unknown	Yes			
47	39	Admin.	Married	Secondary	No	650	No	No	Cellular	15	May	123	2	364	1	Failure	No			
48	45	Admin.	Married	Secondary	No	1803	Yes	No	Unknown	2	Jun	1350	6	1	0	Unknown	Yes			

2. Smaller contact duration seems to have more failure outcomes.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1106	32	Entrepreneur	Married	Unknown	No	73	Yes	No	Cellular	21	Nov	158	1	135	1	Failure	No			
1109	38	Management	Single	Tertiary	No	8669	Yes	No	Telephone	20	Apr	5	3	150	6	Failure	No			
1115	70	Retired	Married	Tertiary	No	15834	No	No	Cellular	5	Apr	172	3	186	1	Failure	No			
1118	40	Management	Married	Tertiary	No	-1212	Yes	No	Cellular	15	May	1259	2	333	18	Failure	No			
1129	41	Services	Married	Unknown	No	88	Yes	No	Cellular	11	May	105	1	336	2	Failure	No			
1132	46	Blue-Collar	Married	Primary	No	91	Yes	No	Cellular	6	May	29	1	356	2	Failure	No			
1147	59	Admin.	Married	Secondary	No	92	Yes	No	Cellular	12	May	139	2	350	1	Failure	No			
1162	32	Admin.	Single	Secondary	No	92	Yes	No	Cellular	18	May	110	1	315	2	Failure	No			
1165	42	Services	Married	Secondary	No	96	Yes	No	Cellular	5	May	238	3	340	2	Failure	No			
1190	50	Management	Married	Tertiary	No	-470	Yes	No	Cellular	8	May	506	2	170	3	Failure	No			
1194	31	Technician	Single	Secondary	No	102	Yes	No	Telephone	17	Apr	460	2	345	2	Failure	No			
1196	34	Technician	Single	Secondary	No	105	Yes	No	Cellular	7	May	146	2	171	3	Failure	No			
1211	50	Management	Married	Tertiary	No	-470	Yes	No	Cellular	19	Nov	130	2	160	1	Failure	No			
1213	38	Blue-Collar	Married	Secondary	No	121	Yes	Yes	Cellular	6	Feb	168	1	212	2	Failure	No			
1229	43	Blue-Collar	Married	Secondary	No	130	No	No	Cellular	5	May	156	1	337	3	Failure	No			
1231	35	Blue-Collar	Married	Secondary	No	137	No	Yes	Unknown	4	Oct	5	1	808	12	Failure	No			
1242	33	Blue-Collar	Married	Secondary	No	139	Yes	Yes	Cellular	18	May	732	2	291	5	Failure	Yes			
1250	33	Blue-Collar	Married	Secondary	No	142	Yes	No	Telephone	11	May	41	2	300	5	Failure	No			
1260	51	Management	Single	Tertiary	Yes	-55	Yes	No	Cellular	11	May	281	2	266	6	Failure	No			
1285	46	Blue-Collar	Married	Primary	No	147	Yes	No	Cellular	6	May	116	2	169	3	Failure	No			
1303	35	Admin.	Married	Secondary	No	147	Yes	Yes	Cellular	30	Jan	44	1	194	7	Failure	No			

Assignment 3

- Longer contact duration seems to have more success outcomes.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
370	61	Admin.	Married	Unknown	No	4629	Yes	No	Cellular	27 Jan		181	1	92	1	Success	Yes			
377	59	Technician	Married	Primary	No	4198	No	No	Cellular	14 Jan		309	1	91	10	Success	Yes			
389	47	Housemaid	Married	Secondary	No	3940	Yes	No	Cellular	9 Nov		171	2	96	5	Success	No			
448	63	Entrepreneur	Married	Secondary	No	3904	No	No	Telephone	4 Aug		250	2	187	2	Success	Yes			
461	75	Retired	Married	Secondary	No	3771	No	No	Telephone	15 Apr		185	1	181	2	Success	Yes			
471	60	Admin.	Married	Secondary	No	3735	No	No	Cellular	22 Dec		190	1	188	9	Success	Yes			
550	34	Admin.	Married	Tertiary	No	899	Yes	No	Telephone	19 Nov		377	3	7	4	Success	No			
560	47	Admin.	Single	Secondary	No	3696	No	No	Cellular	12 Jul		250	2	181	4	Success	Yes			
577	47	Admin.	Married	Secondary	No	3676	No	No	Cellular	18 Oct		567	1	95	3	Success	Yes			
583	38	Management	Married	Unknown	No	3576	No	No	Cellular	25 Nov		349	2	183	1	Success	Yes			
600	43	Services	Single	Secondary	No	3403	Yes	No	Cellular	12 Nov		707	2	97	2	Success	Yes			
605	36	Blue-Collar	Divorced	Secondary	No	2843	No	No	Cellular	12 Feb		473	1	182	1	Success	No			
650	34	Technician	Single	Secondary	No	2729	Yes	No	Cellular	6 Apr		945	1	139	1	Success	No			
651	55	Management	Married	Primary	No	2587	No	No	Cellular	4 Sep		328	2	94	2	Success	Yes			
681	54	Technician	Married	Tertiary	No	876	No	No	Cellular	27 Oct		269	3	541	3	Success	Yes			
765	31	Technician	Single	Secondary	No	2544	No	No	Cellular	15 Jul		252	1	184	3	Success	Yes			
783	21	Student	Single	Secondary	No	2488	No	No	Cellular	30 Jun		258	6	169	3	Success	Yes			
785	36	Management	Married	Tertiary	No	495	No	No	Cellular	12 Mar		213	3	182	6	Success	Yes			
798	53	Management	Married	Unknown	No	2240	Yes	No	Cellular	2 Jun		282	2	385	1	Success	Yes			
824	58	Admin.	Single	Secondary	No	2223	No	No	Cellular	4 Dec		1139	2	114	1	Success	No			
838	43	Blue-Collar	Married	Secondary	No	2160	No	No	Cellular	8 Sep		361	1	98	1	Success	Yes			

- Larger number of previous contacts before the campaign affect greater success outcomes.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Age	Job	Marital	Education	Default	Balance	Housing	Loar	Contact	Day	Months	Duration	Campaign	Pday	Previous	Poutcome				
40	37	Technician	Single	Tertiary	No	6101	No	No	Cellular	4 May		60	5	87	1	Success	No			
63	32	Technician	Married	Unknown	No	14533	No	No	Cellular	31 Dec		646	3	198	2	Success	No			
72	52	Technician	Married	Tertiary	No	2557	No	No	Cellular	28 Jan		134	5	204	2	Success	Yes			
135	50	Blue-Collar	Married	Primary	No	12519	Yes	No	Cellular	17 Apr		146	2	147	4	Success	No			
146	79	Retired	Married	Primary	No	8556	No	No	Telephone	18 Jan		326	2	272	2	Success	Yes			
175	80	Retired	Married	Secondary	No	8304	No	No	Telephone	6 Apr		681	1	118	11	Success	No			
221	59	Self-Employed	Married	Tertiary	No	2013	No	No	Cellular	19 Feb		222	4	94	2	Success	Yes			
285	29	Management	Single	Tertiary	No	15459	No	No	Cellular	26 May		245	3	97	7	Success	No			
308	38	Technician	Married	Secondary	No	6728	No	No	Cellular	25 Nov		232	1	96	4	Success	Yes			
310	38	Technician	Single	Secondary	No	-242	Yes	No	Cellular	18 May		1173	3	278	6	Success	Yes			
324	35	Technician	Married	Tertiary	No	7050	No	No	Cellular	11 Aug		222	3	182	4	Success	Yes			
334	41	Admin.	Divorced	Secondary	No	6046	Yes	Yes	Telephone	15 Mar		300	6	182	2	Success	Yes			
346	36	Management	Married	Tertiary	No	925	No	No	Cellular	14 Dec		321	3	189	4	Success	Yes			
357	40	Services	Single	Secondary	No	1547	Yes	No	Cellular	12 May		88	2	173	1	Success	No			
370	61	Admin.	Married	Unknown	No	4629	Yes	No	Cellular	27 Jan		181	1	92	1	Success	Yes			
377	59	Technician	Married	Primary	No	4198	No	No	Cellular	14 Jan		309	1	91	10	Success	Yes			
389	47	Housemaid	Married	Secondary	No	3940	Yes	No	Cellular	9 Nov		171	2	96	5	Success	No			
448	63	Entrepreneur	Married	Secondary	No	3904	No	No	Telephone	4 Aug		250	2	187	2	Success	Yes			
461	75	Retired	Married	Secondary	No	3771	No	No	Telephone	15 Apr		185	1	181	2	Success	Yes			
471	60	Admin.	Married	Secondary	No	3735	No	No	Cellular	22 Dec		190	1	188	9	Success	Yes			

Assignment 3

- Lesser number of previous contacts before the campaign are more likely to cause campaign failure.

BankTest

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
121	31	Management	Married	Tertiary	No	22546	Yes	No	Cellular	14	May	8	6	267	4	Failure	No			
124	57	Retired	Married	Primary	No	464	Yes	No	Cellular	4	Feb	75	1	190	6	Failure	No			
134	36	Blue-Collar	Married	Secondary	No	461	No	No	Cellular	11	May	254	2	353	1	Failure	No			
156	42	Technician	Single	Tertiary	No	411	Yes	No	Cellular	12	May	78	5	292	3	Failure	No			
157	60	Admin.	Married	Secondary	No	396	Yes	No	Cellular	11	May	443	1	350	1	Failure	No			
167	42	Admin.	Married	Secondary	No	396	Yes	No	Cellular	11	May	80	1	308	4	Failure	No			
179	40	Services	Married	Secondary	No	378	Yes	Yes	Cellular	13	May	541	1	295	6	Failure	No			
204	35	Technician	Married	Secondary	No	259	Yes	No	Cellular	29	Jan	381	1	247	2	Failure	No			
220	38	Services	Single	Secondary	No	196	Yes	No	Cellular	2	Feb	60	1	271	1	Failure	No			
226	31	Blue-Collar	Married	Unknown	No	120	Yes	No	Cellular	3	Feb	100	3	201	2	Failure	No			
260	42	Blue-Collar	Married	Primary	No	97	Yes	No	Cellular	3	Feb	216	4	271	5	Failure	No			
262	43	Services	Married	Primary	No	88	Yes	Yes	Cellular	17	Apr	313	1	147	2	Failure	No			
272	38	Management	Married	Tertiary	No	0	Yes	No	Cellular	13	May	537	5	303	2	Failure	No			
280	50	Technician	Divorced	Secondary	No	54	Yes	Yes	Cellular	13	May	698	2	292	1	Failure	Yes			
293	60	Retired	Divorced	Unknown	No	49	Yes	No	Cellular	29	Jan	188	1	261	2	Failure	No			
295	59	Self-Employed	Married	Tertiary	No	185	No	No	Cellular	22	Jun	177	5	138	1	Failure	No			
302	30	Blue-Collar	Married	Secondary	No	28	Yes	No	Cellular	18	May	284	2	371	1	Failure	No			
311	51	Unemployed	Married	Secondary	No	13	Yes	No	Cellular	13	May	361	1	364	1	Failure	No			
313	45	Blue-Collar	Single	Secondary	No	0	Yes	No	Cellular	26	Apr	1472	1	192	4	Failure	Yes			
322	54	Blue-Collar	Married	Secondary	No	0	No	No	Telephone	16	Apr	1366	1	150	1	Failure	No			
326	64	Technician	Married	Secondary	No	0	No	No	Telephone	16	May	886	1	04	9	Failure	No			

- Smaller number of days that passed by after the client was last contacted from a previous campaign seems to have more success rate outcomes.

BankTest

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1479	52	Technician	Married	Secondary	No	195	Yes	No	Cellular	18	Feb	220	1	63	5	Failure	Yes			
1494	52	Technician	Married	Secondary	No	196	Yes	Yes	Cellular	24	Feb	116	1	273	3	Failure	No			
1501	49	Admin.	Married	Secondary	No	197	Yes	No	Cellular	13	Apr	223	1	182	2	Failure	Yes			
1504	32	Blue-Collar	Married	Secondary	No	889	Yes	No	Cellular	10	Sep	185	1	99	5	Success	Yes			
1508	48	Management	Married	Tertiary	No	60	No	No	Telephone	7	May	376	2	339	7	Failure	No			
1513	34	Services	Married	Secondary	No	1076	No	No	Cellular	12	May	152	1	182	6	Success	Yes			
1514	48	Unemployed	Married	Secondary	No	817	No	No	Cellular	14	Jan	517	1	282	4	Success	Yes			
1521	64	Housemaid	Married	Primary	No	768	No	No	Cellular	15	Jul	250	2	182	1	Success	No			
1525	40	Management	Divorced	Tertiary	No	69	Yes	No	Cellular	6	May	503	2	280	2	Failure	No			
1532	32	Technician	Single	Tertiary	No	13711	Yes	No	Cellular	14	May	638	1	175	3	Success	No			
1534	47	Housemaid	Married	Primary	No	201	Yes	No	Cellular	7	Aug	58	2	382	1	Failure	No			
1546	32	Technician	Divorced	Secondary	No	202	Yes	No	Cellular	4	Feb	429	1	201	1	Failure	No			
1552	33	Admin.	Married	Tertiary	No	79	Yes	No	Cellular	22	Oct	249	2	335	2	Failure	Yes			
1553	51	Blue-Collar	Divorced	Secondary	No	203	Yes	No	Cellular	8	May	134	1	170	5	Failure	No			
1566	45	Self-Employed	Married	Secondary	No	706	No	No	Cellular	30	Apr	78	2	176	1	Success	No			
1567	26	Student	Single	Secondary	No	203	No	No	Cellular	16	Nov	104	1	196	1	Failure	No			
1573	47	Unknown	Married	Unknown	No	677	No	No	Cellular	7	Aug	384	1	99	4	Success	No			
1575	36	Blue-Collar	Single	Secondary	No	219	Yes	Yes	Cellular	11	May	270	2	347	1	Failure	No			
1582	40	Unemployed	Married	Secondary	No	219	Yes	No	Cellular	17	Nov	204	2	196	1	Failure	No			
1586	32	Technician	Single	Tertiary	No	5514	No	No	Cellular	22	Apr	319	1	182	1	Success	Yes			
1600	35	Services	Divorced	Secondary	No	321	Yes	No	Cellular	14	May	150	1	260	1	Failure	No			

Pivot table showing relationship of column Y vs other factors

1. The lesser the average yearly balance the more likely a client not to subscribe to a term deposit.

	A	B
2		
3	Row Labels	
4	No	
5	-3313	
6	-2082	
7	-1746	
8	-1680	
9	-1400	
10	-1313	
11	-1310	
12	-1224	
13	-1212	
14	-1202	
15	-1164	
16	-1148	
17	-1053	
18	-988	
19	-970	
20	-967	
21	-966	

	A	B
2		
3	Row Labels	
4	No	
5	Yes	
6	-1206	
7	-970	
8	-887	
9	-839	
10	-824	
11	-701	
12	-522	
13	-477	
14	-454	
15	-416	
16	-393	
17	-386	
18	-375	
19	-315	
20	-311	
21	-308	

NO VS YES to term deposit subscription

2. The length of the last contact duration affects subscription decision. The longer the length the more likely a customer to subscribe.

SUBSCRIBED YES

	A	B
1976	No	
1977	1484	
1978	Yes	
1979	1504	
1980	No	
1981	1516	
1982	No	
1983	Yes	
1984	1521	
1985	No	
1986	1529	
1987	No	
1988	1531	
1989	No	
1990	1532	
1991	Yes	
1992	1558	
1993	Yes	
1994	1579	
1995	Yes	

NOT SUBSCRIBED NO

	A	B
80	No	
81	42	
82	No	
83	43	
84	No	
85	44	
86	No	
87	45	
88	No	
89	46	
90	No	
91	47	
92	No	
93	48	
94	No	
95	49	
96	No	
97	50	
98	No	
99	51	

Assignment 3

3. Some age categories have either only YES or only NO to a term subscription.

	A	B
I45	=66	
I46	No	
I47	Yes	
I48	=67	
I49	=68	
I50	Yes	
I51	=69	
I52	=70	
I53	No	
I54	=71	
I55	=72	
I56	=73	
I57	=74	
I58	=75	
I59	=76	
I60	No	
I61	=77	
I62	=78	
I63	=79	
I64	=80	
I65	=81	
I66	No	
I67	=83	
I68	=84	
I69	Yes	
I70	=86	
I71	No	
I72	=87	
I73	Yes	
I74	Grand Total	

4. The more the number of contacts performed during this campaign and for this client the more likely customer not to subscribe to the term deposit.

62	No	
63	=24	
64	No	
65	Yes	
66	=25	
67	No	
68	=28	
69	No	
70	=29	
71	No	
72	=30	
73	No	
74	=31	
75	No	
76	=32	
77	No	
78	=44	
79	No	
80	=50	
81	No	

Assignment 3

5. The more the number of contacts performed before this campaign and for this client the more likely customer to not subscribe to the term deposit.

	A	B
46	<input type="checkbox"/> 15	
47	No	
48	<input type="checkbox"/> 17	
49	No	
50	<input type="checkbox"/> 18	
51	No	
52	<input type="checkbox"/> 19	
53	No	
54	<input type="checkbox"/> 20	
55	No	
56	<input type="checkbox"/> 22	
57	No	
58	<input type="checkbox"/> 23	
59	No	
60	<input type="checkbox"/> 24	
61	No	
62	<input type="checkbox"/> 25	
63	No	
64	Grand Total	