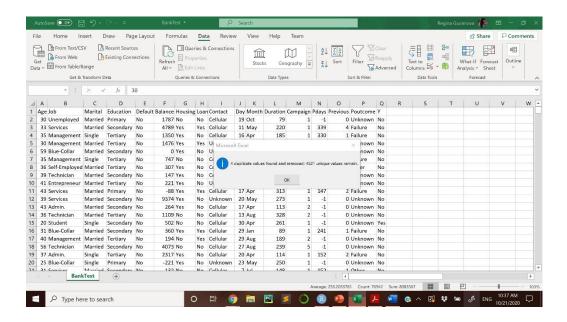
Data Cleaning

- 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.



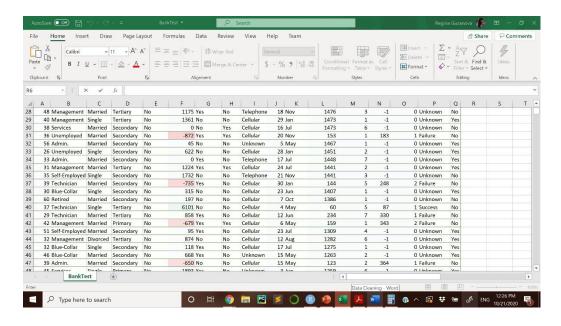
Data grouping and summarizing

Data Filtering

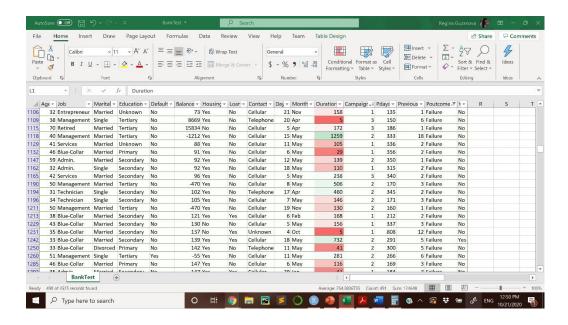
- 1. Ctrl+A. Format as a Table. Green Light.
- 2. Filter "Success" outcomes. 129 results found. 82-Married, 10-Divorsed, 37-Single. Therefore, success of the campaign is more likely for married individuals.
- 3. Success outcome has less contacts during the campaign 5-1, while failure outcome has more 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.

Color scale or data bar to explore and highlight the data

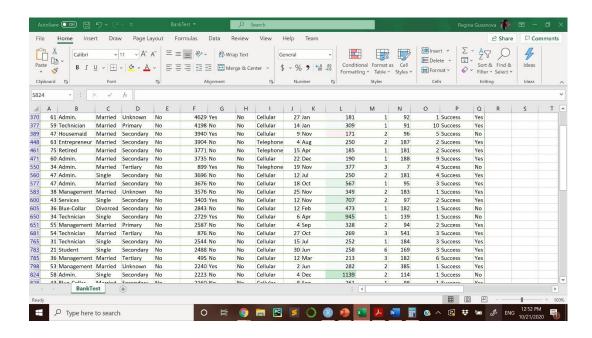
1. From the color formatting it is obvious that people with average negative yearly balances are more likely to have failure outcome from the campaign.



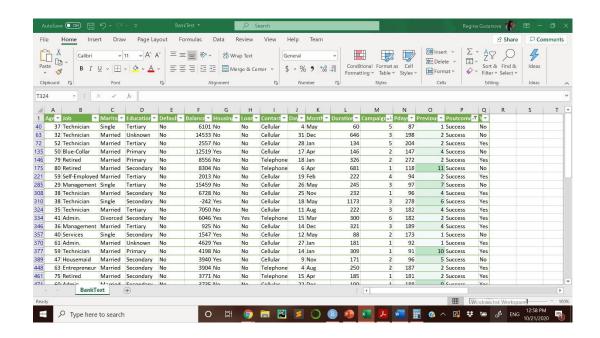
2. Smaller contact duration seems to have more failure outcomes.



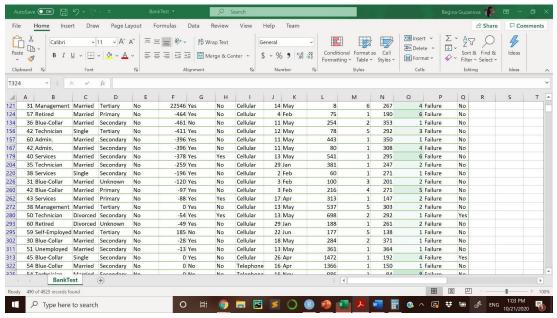
3. Longer contact duration seems to have more success outcomes.



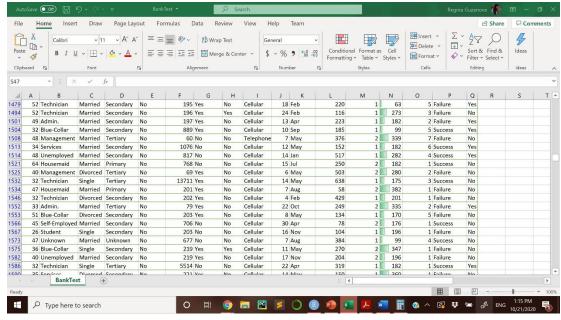
4. Larger number of previous contacts before the campaign affect greater success outcomes.



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

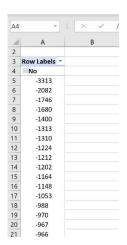


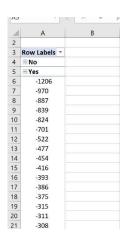
6. Smaller number of days that passed by after the client was last contacted from a previous campaign seems to have more success rate outcomes.



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.





NO VS YES to term deposit subscribtion

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

SUBSCRIBED YES

	Α	В
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

	А	В
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	

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

Some age categories have		
4	A	В
	■66	
	No	
	Yes	
0//00000	⊞67	
	■68	
150	Yes	
151	⊞69	
152	■70	
153	No	
154	⊞71	
155	⊞72	
156	⊞73	
157	⊞74	
158	⊞75	
159	■76	
160	No	
161	⊞77	
162	⊞78	
163	⊞79	
164	⊞80	
165	■81	
166	No	
167	⊞83	
168	■84	
169	Yes	
170	■86	
171	No	
172	■87	
173	Yes	
174	Cuand Takel	

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	

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.

4	A	В
46	15	
47	No	
48	■17	
49	No	
50	■18	
51	No	
52	19	
53	No	
54	■20	
55	No	
56	■22	
57	No	
58	23	
59	No	
60	■24	
61	No	
62	■25	
63	No	
64	Grand Total	