

# Preprocessing

*Admin*

*24 4 2019*

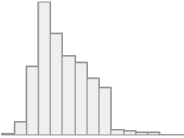


## Raw Dataset

### Data Frame Summary












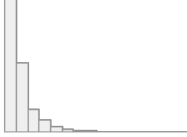







data


**Dimensions:** 4521 x 17

**Duplicates:** 0

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	age [integer]	Mean (sd) : 41.2 (10.6) min < med < max: 19 < 39 < 87 IQR (CV) : 16 (0.3)	67 distinct values		4521 (100%)	0 (0%)
2	job [factor]	1. admin. 2. blue-collar 3. entrepreneur 4. housemaid 5. management 6. retired 7. self-employed 8. services 9. student 10. technician [ 2 others ]	478 (10.6%) 946 (20.9%) 168 ( 3.7%) 112 ( 2.5%) 969 (21.4%) 230 ( 5.1%) 183 ( 4.0%) 417 ( 9.2%) 84 ( 1.9%) 768 (17.0%) 166 ( 3.7%)		4521 (100%)	0 (0%)
3	marital [factor]	1. divorced 2. married 3. single	528 (11.7%) 2797 (61.9%) 1196 (26.5%)		4521 (100%)	0 (0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
4	education [factor]	1. primary 2. secondary 3. tertiary 4. unknown	678 (15.0%) 2306 (51.0%) 1350 (29.9%) 187 ( 4.1%)		4521 (100%)	0 (0%)
5	default [factor]	1. no 2. yes	4445 (98.3%) 76 ( 1.7%)		4521 (100%)	0 (0%)
6	balance [integer]	Mean (sd) : 1422.7 (3009.6) min < med < max: -3313 < 444 < 71188 IQR (CV) : 1411 (2.1)	2353 distinct values		4521 (100%)	0 (0%)
7	housing [factor]	1. no 2. yes	1962 (43.4%) 2559 (56.6%)		4521 (100%)	0 (0%)
8	loan [factor]	1. no 2. yes	3830 (84.7%) 691 (15.3%)		4521 (100%)	0 (0%)
9	contact [factor]	1. cellular 2. telephone 3. unknown	2896 (64.1%) 301 ( 6.7%) 1324 (29.3%)		4521 (100%)	0 (0%)
10	day [integer]	Mean (sd) : 15.9 (8.2) min < med < max: 1 < 16 < 31 IQR (CV) : 12 (0.5)	31 distinct values		4521 (100%)	0 (0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
11	month [factor]	1. apr	293 ( 6.5%)		4521 (100%)	0 (0%)
		2. aug	633 (14.0%)			
		3. dec	20 ( 0.4%)			
		4. feb	222 ( 4.9%)			
		5. jan	148 ( 3.3%)			
		6. jul	706 (15.6%)			
		7. jun	531 (11.7%)			
		8. mar	49 ( 1.1%)			
		9. may	1398 (30.9%)			
		10. nov	389 ( 8.6%)			
		[ 2 others ]	132 ( 2.9%)			
12	duration [integer]	Mean (sd) : 264 (259.9) min < med < max: 4 < 185 < 3025 IQR (CV) : 225 (1)	875 distinct values		4521 (100%)	0 (0%)
13	campaign [integer]	Mean (sd) : 2.8 (3.1) min < med < max: 1 < 2 < 50 IQR (CV) : 2 (1.1)	32 distinct values		4521 (100%)	0 (0%)
14	pdays [integer]	Mean (sd) : 39.8 (100.1) min < med < max: -1 < -1 < 871 IQR (CV) : 0 (2.5)	292 distinct values		4521 (100%)	0 (0%)
15	previous [integer]	Mean (sd) : 0.5 (1.7) min < med < max: 0 < 0 < 25 IQR (CV) : 0 (3.1)	24 distinct values		4521 (100%)	0 (0%)
16	poutcome [factor]	1. failure	490 (10.8%)		4521 (100%)	0 (0%)
		2. other	197 ( 4.4%)			
		3. success	129 ( 2.9%)			
		4. unknown	3705 (82.0%)			

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
17	y [factor]	1. no 2. yes	4000 (88.5%) 521 (11.5%)		4521 (100%)	0 (0%)