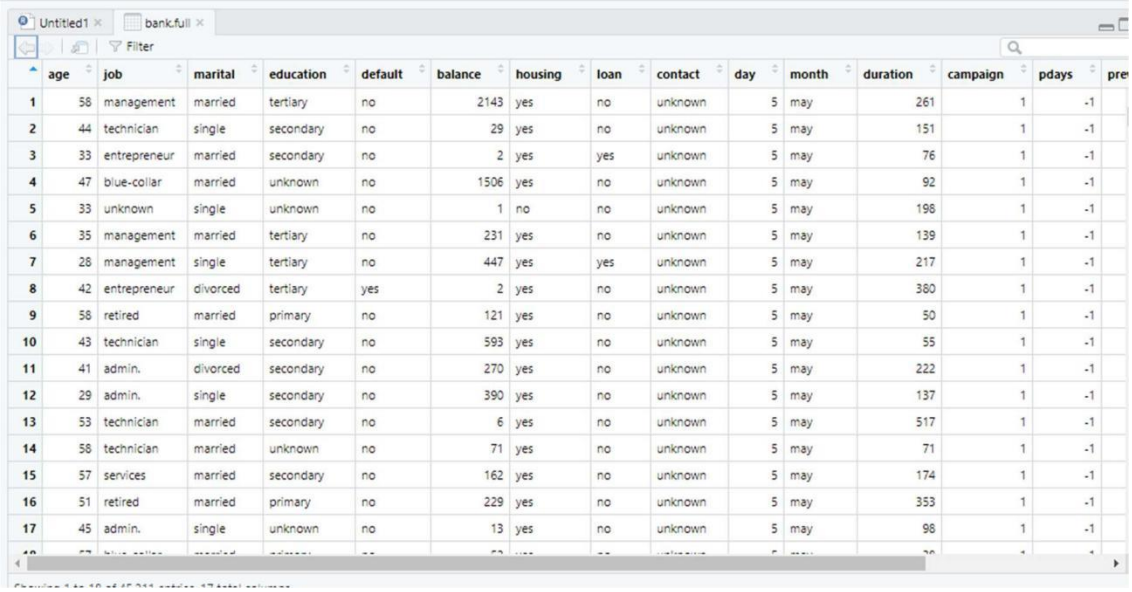
Banking Data Analysis

Koduri Gokul
19BCD7006

Dataset:



Untitled1 × bank.full ×

Filter

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	pre
1	58	management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	
2	44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	
3	33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	
4	47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	
5	33	unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1	
6	35	management	married	tertiary	no	231	yes	no	unknown	5	may	139	1	-1	
7	28	management	single	tertiary	no	447	yes	yes	unknown	5	may	217	1	-1	
8	42	entrepreneur	divorced	tertiary	yes	2	yes	no	unknown	5	may	380	1	-1	
9	58	retired	married	primary	no	121	yes	no	unknown	5	may	50	1	-1	
10	43	technician	single	secondary	no	593	yes	no	unknown	5	may	55	1	-1	
11	41	admin.	divorced	secondary	no	270	yes	no	unknown	5	may	222	1	-1	
12	29	admin.	single	secondary	no	390	yes	no	unknown	5	may	137	1	-1	
13	53	technician	married	secondary	no	6	yes	no	unknown	5	may	517	1	-1	
14	58	technician	married	unknown	no	71	yes	no	unknown	5	may	71	1	-1	
15	57	services	married	secondary	no	162	yes	no	unknown	5	may	174	1	-1	
16	51	retired	married	primary	no	229	yes	no	unknown	5	may	353	1	-1	
17	45	admin.	single	unknown	no	13	yes	no	unknown	5	may	98	1	-1	

Analysis:

```
> head(bank.full, 5)
  age          job marital education default balance housing loan contact day month duration campaign pdays previous
1  58   management married  tertiary      no    2143     yes   no unknown   5   may      261        1    -1        0
2  44   technician  single secondary      no      29     yes   no unknown   5   may      151        1    -1        0
3  33 entrepreneur married secondary      no       2     yes  yes unknown   5   may       76        1    -1        0
4  47  blue-collar married   unknown      no    1506     yes   no unknown   5   may       92        1    -1        0
5  33      unknown  single   unknown      no       1      no   no unknown   5   may      198        1    -1        0
  poutcome Target
1  unknown     no
2  unknown     no
3  unknown     no
4  unknown     no
5  unknown     no
>

>
> dim(bank.full)
[1] 45211    17
>
```

```
>
> str(bank.full)
'data.frame':   45211 obs. of  17 variables:
 $ age      : int  58 44 33 47 33 35 28 42 58 43 ...
 $ job      : chr  "management" "technician" "entrepreneur" "blue-collar" ...
 $ marital  : chr  "married" "single" "married" "married" ...
 $ education: chr  "tertiary" "secondary" "secondary" "unknown" ...
 $ default  : chr  "no" "no" "no" "no" ...
 $ balance  : int  2143 29 2 1506 1 231 447 2 121 593 ...
 $ housing  : chr  "yes" "yes" "yes" "yes" ...
 $ loan     : chr  "no" "no" "yes" "no" ...
 $ contact  : chr  "unknown" "unknown" "unknown" "unknown" ...
 $ day      : int  5 5 5 5 5 5 5 5 5 5 ...
 $ month    : chr  "may" "may" "may" "may" ...
 $ duration : int  261 151 76 92 198 139 217 380 50 55 ...
 $ campaign : int  1 1 1 1 1 1 1 1 1 1 ...
 $ pdays    : int  -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...
 $ previous : int  0 0 0 0 0 0 0 0 0 0 ...
 $ poutcome : chr  "unknown" "unknown" "unknown" "unknown" ...
 $ Target   : chr  "no" "no" "no" "no" ...
>
```

```
>
> colSums(is.na(bank.full))
      age       job   marital education   default   balance   housing      loan   contact       day     month  duration
        0         0         0         0         0         0         0         0         0         0         0         0
 campaign     pdays  previous  poutcome    Target
        0         0         0         0         0
>
```

```
>
> data.frame(sort(table("education" = bank.full$education),decreasing = T))
  education  Freq
1 secondary 23202
2  tertiary 13301
3   primary  6851
4   unknown  1857
>
```

```
>
>
> data.frame(table("loan" = bank.full$loan,
+                  "housing" = bank.full$housing,
+                  "default" = bank.full$default))
  loan housing default  Freq
1   no      no      no 16992
2  yes      no      no  2709
3   no     yes      no 20461
4  yes     yes      no  4234
5   no      no     yes   212
6  yes      no     yes   168
7   no     yes     yes   302
8  yes     yes     yes   133
>
```

```
>
> data.frame(sort(table("marital_status" = bank.full$marital),decreasing = T))
  marital_status  Freq
1        married 27214
2         single 12790
3       divorced  5207
>
```

```
>
> data.frame(sort(table("contact" = bank.full$contact),decreasing = T))
    contact  Freq
1  cellular 29285
2   unknown 13020
3 telephone  2906
>
>
```

```
> data.frame(sort(table("contacted" = bank.full$previous != 0),decreasing = T))
  contacted  Freq
1     FALSE 36954
2      TRUE  8257
>
```