Assignment3_Q2

JL

25 November 2018

Finding #1

Job categories used.

```
job_category <- distinct(bank, job) %>% arrange(job)

kable(job_category, caption="job_category")
```

Table 1: job_category

job
admin.
blue-collar
entrepreneur
housemaid
management
retired
self-employed
services
student
technician
unemployed
unknown

Finding #2

Defaulted Person's total balance for each job category. Interesting finding is that housemaid defaulted but has positive bank balance.

```
default_BalNJob <- filter(bank, default == "yes") %>% group_by(.,job) %>% summarise(balance = sum(balan

kable(default_BalNJob, caption="default_BalNJob")
```

Table 2: default_BalNJob

job	balance
self-employed	-3559
entrepreneur	-3524
technician	-2399
blue-collar	-2054
management	-1609
services	-1280
unemployed	-1019
admin.	-968
retired	-453
student	-230

job	balance
housemaid	1232

Finding #3

Find the age group that has the highest default rate. People mostly tend to default in 20-30 age group & 40-50 age group.

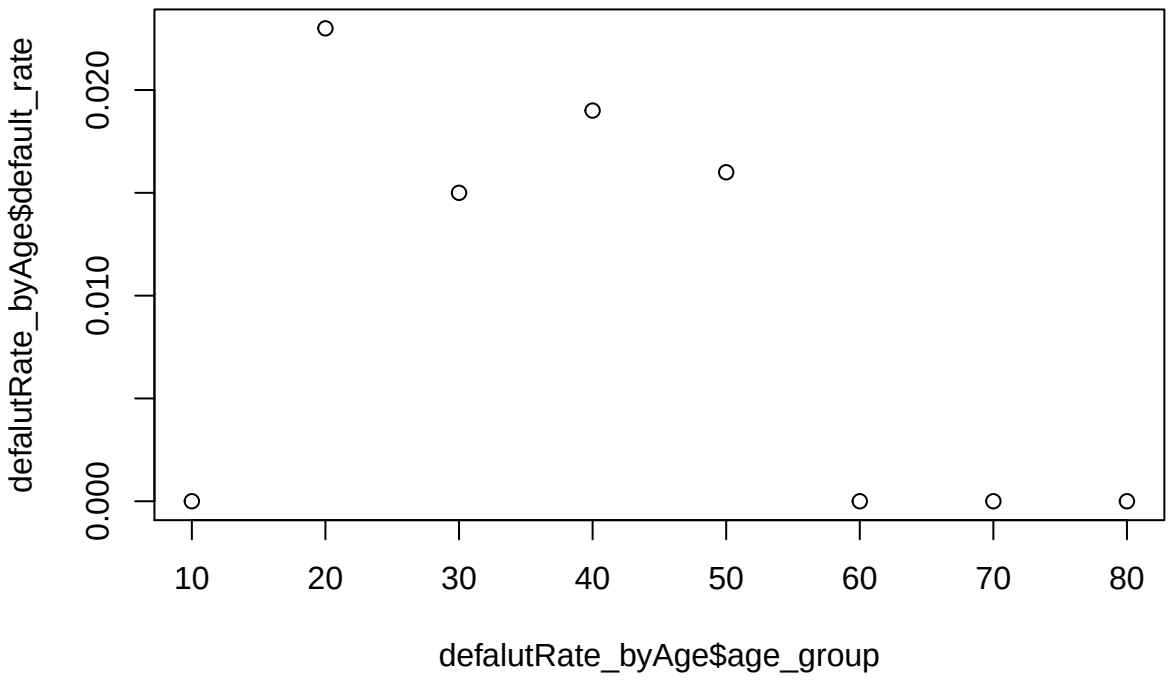
```
defalutRate_byAge <- bank %>% group_by(age_group = (age %/% 10) * 10) %>%
  summarise(default_rate = round(sum(ifelse(default=="yes",1,0))/n(),3)) %>%  arrange(default_rate)

kable(defalutRate_byAge, caption="defalutRate_byAge")
```

Table 3: defalutRate_byAge

age_group	default_rate
10	0.000
60	0.000
70	0.000
80	0.000
30	0.015
50	0.016
40	0.019
20	0.023

```
plot(defalutRate_byAge$age_group,defalutRate_byAge$default_rate)
```



Finding #4

Bank clients have a loan while doesn't have a housing, grouping by job

```
loan_no_housing <- bank %>% filter(loan == "yes" & housing == "no") %>% group_by(.,job) %>% summarise(c

kable(loan_no_housing, caption="loan_no_housing")
```

Table 4: loan_no_housing

job	count
admin.	35
blue-collar	53
entrepreneur	12
housemaid	8
management	59
retired	19
self-employed	15
services	33
student	1
technician	46
unemployed	4

Finding #5

Bank client have a job between 20 to 40, grouping by job.

```
age20to40_withJob <- bank %>% filter(age>=20 & age <= 40 & !(job %in% c("housemaid","retired","student"
  group_by(.,job) %>% summarise(count = n())

kable(age20to40_withJob, caption="age20to40_withJob")
```

Table 5: age20to40_withJob

job	count
admin.	273
blue-collar	515
entrepreneur	78
management	552
self-employed	97
services	269
technician	458
unknown	9

Finding #6

The longest call duration.

```
longest_duration <- filter(bank,duration == max(duration))

kable(longest_duration, caption="longest_duration")
```

Table 6: longest_duration

age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campa
59	unemployed	married	primary	no	0	no	no	cellular	30	jan	3025	

Finding #7

The eldest person.

```
eldest_person <- filter(bank,age==max(age))

kable(eldest_person, caption="eldest_person")
```

Table 7: eldest_person

age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign
87	retired	married	primary	no	230	no	no	cellular	30	oct	144	1

Finding #8

number of person in each job category.

```
jobs_count <- group_by(bank,job) %>% summarise(total = n())

kable(jobs_count, caption="jobs_count")
```

Table 8: jobs_count

job	total
admin.	478
blue-collar	946
entrepreneur	168
housemaid	112
management	969
retired	230
self-employed	183
services	417
student	84
technician	768
unemployed	128
unknown	38

Finding #9

campaign successful rate by job category. entrepreneur has lowest success rate, maybe due to busy work schedule.

```
successRate_byJob <- bank %>% group_by(job) %>% summarize(success_rate = round(sum(ifelse(poutcome=="su
  arrange(success_rate)

kable(successRate_byJob, caption="successRate_byJob")
```

Table 9: successRate_byJob

job	success_rate
entrepreneur	0.006
blue-collar	0.015
self-employed	0.016
services	0.022
management	0.024
unemployed	0.031
housemaid	0.036
technician	0.036
admin.	0.048
student	0.048
retired	0.057
unknown	0.079

Finding #10

Default rate group by Marital status, education level, and average bank balance. Single and primary school educated person has highest default rate.

```
married_group <- filter(bank, marital=="married") %>% group_by(education) %>%
  summarise(aveage_balance=mean(balance),default_rate=round(sum(ifelse(default=="yes",1,0))/n(),3))
single_group <- filter(bank, marital=="single") %>% group_by(education) %>%
  summarize(aveage_balance=round(mean(balance),0), default_rate=round(sum(ifelse(default=="yes",1,0))/n

married_group <- mutate(married_group, marital="married")
single_group <- mutate(single_group, marital="single")

marital_df <- bind_rows(married_group,single_group) %>% arrange(default_rate)

kable(marital_df, caption="marital_education_default")
```

Table 10: marital_education_default

education	aveage_balance	default_rate	marital
unknown	1562.000	0.000	single
tertiary	1860.719	0.007	married
primary	1371.639	0.011	married
secondary	1272.911	0.016	married
unknown	1725.547	0.017	married
tertiary	1755.000	0.017	single
secondary	1154.000	0.020	single
primary	2066.000	0.027	single