```
###Homework 3

library("tidyverse")
library("FNN")


#1
##A
set.seed(30)

tayko = read.csv("Tayko.csv")

tayko = tayko %>%
  mutate(id = 1:nrow(tayko))

train = tayko %>%
  sample_frac(0.8)

validation = tayko %>%
  slice(setdiff(tayko$id, train$id))

train.mlr = lm(spending~freq + last_update + web + gender + address_res + address_us, train)

summary(train.mlr)

validation = validation %>%
  mutate(spending_prediction = predict(train.mlr, validation))
validation = validation %>%
  mutate(error = spending - spending_prediction)

library("forecast")
accuracy(validation$spending_prediction, validation$spending)

##B
tayko= tayko %>%
  mutate(id = 1:nrow(tayko)) %>%
  rename(spending_actual = spending)

n = nrow(tayko)

temp = as_tibble()

for(obs_num in 1:n)
{
```

```r
  train = tayko %>%
    filter(id != obs_num)

  validation = tayko %>%
    filter(id == obs_num)

  train.mlr = lm(spending_actual~freq + last_update + web + gender + address_us +
address_res, train)

  validation = validation %>%
    mutate(spending_prediction = predict(train.mlr, validation))

  am = accuracy(validation$spending_prediction, validation$spending_actual)

  temp = temp %>%
    bind_rows(as_tibble(list(run = obs_num, RMSE = am[2])))

  print(paste("iteration", obs_num, "completed"), sep = " ")

}
#KFOLD

set.seed(30)
library(caret)
fold = createFolds(tayko$spending, k=10)
View(fold)

test = tayko[fold$Fold01, ]
str(test)

train = tayko[-fold$Fold01, ]
str(train)

library(caret)
library(C50)
library(irr)
library(tidyverse)
train.mlr = lm(spending~ freq + last_update + web + gender + address_us + address_res, train)

test = test %>%
  mutate(spending_prediction = predict(train.mlr, test))

am1 = accuracy(test$spending_prediction, test$spending)
```

```
##2
##A
rm(list = ls())
pl = read.csv("personal_loan.csv")
pl  = pl %>%
  mutate(id = 1:nrow(pl))

pl = pl %>%
  mutate(education_undergrad = if_else(education == "undergrad", 1, 0),
      education_graduate = if_else(education == "graduate", 1, 0),
      education_advanced = if_else(education == "advanced", 1, 0))

pl = pl %>%
  mutate(ls_reject = if_else(loan_status == "reject", 1, 0))

set.seed(30)
pl = pl %>%
  mutate(id = 1:nrow(pl))

train = pl %>%
  sample_frac(0.8)

validation = pl %>%
  slice(setdiff(pl$id, train$id))

train.lr = glm(ls_reject~ age + experience + income + family + ccavg + education_graduate +
education_advanced + mortgage + securities_account +
         cd_account + online + credit_card,train, family = "binomial")
summary(train.lr)
```

An increase of $1000 in income, while holding other variables constant, increase the average odds that the customer rejects the loan offer by a factor of ~ 0.94249.
A customer who has an advanced education will reject the offer with an average odds of ~ 0.019507 relative to a customer who does not have an advanced education while holding all other variables constant.

```
##B
library(caret)
validation = validation %>%
  mutate(prob_prediction = predict(train.lr, validation, type = "response"))

validation = validation %>%
```

```r
  mutate(loan_status_prediction = if_else(prob_prediction >= 0.5, 1, 0))

validation = validation %>%
  mutate(ls_reject = factor(ls_reject, c(0,1)),
         loan_status_prediction = factor(loan_status_prediction, c(0,1)))

confusionMatrix(validation$loan_status_prediction, validation$ls_reject)
```