

# R Notebook

```
rm(list=ls())  
graphics.off()
```

## Survey selection

```
survey_nps_tot = df_clean[df_clean$survey %in% c( "4", "16", "18", "52"),]
```

## Variable aggregation

```
survey_nps_tot <- survey_nps_tot %>%  
  mutate(regione_aggregata = case_when(  
    regione_des_ana %in% c(1, 11, 3, 2, 7, 6, 8) ~ as.factor(regione_des_ana),  
    TRUE ~ "Altro"  
  ))  
survey_nps_tot$regione_aggregata = as.factor((survey_nps_tot$regione_aggregata))
```

```
survey_nps_tot <- survey_nps_tot %>%  
  mutate(direzione_aggregata = case_when(  
    direzione %in% c(1, 2, 3, 4, 5, 7, 6, 8, 9) ~ as.factor(direzione),  
    TRUE ~ "Altro"  
  ))  
survey_nps_tot$direzione_aggregata = as.factor((survey_nps_tot$direzione_aggregata))
```

## NA Imputation

```
survey_nps_tot <- kNN(survey_nps_tot, variable = "risk_rating_comm")
```

## Data Cleaning

```
survey_nps_tot_clean <- survey_nps_tot[complete.cases(survey_nps_tot[, c("nps", "segmento_des_comm", "r  
c("nps", "segmento_des_comm", "regione_aggregata", "direzione_aggregata", "r
```

## Linear Model

```
modello_1 <- lm(nps ~ segmento_des_comm + regione_aggregata + direzione_aggregata + sesso_code_ana + f
summary(modello_1)
```

```
##
## Call:
## lm(formula = nps ~ segmento_des_comm + regione_aggregata + direzione_aggregata +
##     sesso_code_ana + fascia_eta_code_ana + vdbank_flg_ana + xntweb_flg_ana +
##     cliente_solo_cc_flg_ana + cliente_investitore_flg_ana + cliente_con_mutuo_flg_ana +
##     in_bonis_flg_comm + multi_flg_comm + fascia_utilizzo_online_comm +
##     risk_rating_comm, data = survey_nps_tot_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.3031 -0.7493  0.2784  1.0638  1.9323
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.218794   0.249816  32.899  < 2e-16 ***
## segmento_des_comm3 -0.088373   0.038857  -2.274  0.022966 *
## segmento_des_comm4 -0.078673   0.207596  -0.379  0.704718
## segmento_des_comm5  0.267013   0.798657   0.334  0.738139
## segmento_des_comm7  0.404108   0.058418   6.918  4.87e-12 ***
## segmento_des_comm9 -0.079021   0.186778  -0.423  0.672252
## regione_aggregata11 -0.049493   0.096306  -0.514  0.607323
## regione_aggregata2 -0.110247   0.152927  -0.721  0.470980
## regione_aggregata3  0.760413   0.338912   2.244  0.024874 *
## regione_aggregata6  0.421428   0.433277   0.973  0.330749
## regione_aggregata7  0.345306   0.318901   1.083  0.278922
## regione_aggregata8 -0.020788   0.170171  -0.122  0.902775
## regione_aggregataAltro 0.232331   0.306827   0.757  0.448945
## direzione_aggregata2  0.341074   0.108866   3.133  0.001735 **
## direzione_aggregata3  0.519786   0.187215   2.776  0.005506 **
## direzione_aggregata4  0.262166   0.094813   2.765  0.005701 **
## direzione_aggregata5  0.339439   0.114898   2.954  0.003141 **
## direzione_aggregata6 -0.404015   0.349447  -1.156  0.247645
## direzione_aggregata7  0.123020   0.326647   0.377  0.706468
## direzione_aggregata8  0.322148   0.189045   1.704  0.088397 .
## direzione_aggregata9  0.082994   0.441408   0.188  0.850864
## direzione_aggregataAltro 0.807705   0.425527   1.898  0.057708 .
## sesso_code_ana2      0.055936   0.028300   1.977  0.048121 *
## fascia_eta_code_ana2  0.004414   0.050055   0.088  0.929732
## fascia_eta_code_ana3  0.040652   0.040515   1.003  0.315701
## fascia_eta_code_ana4 -0.304170   0.066568  -4.569  4.95e-06 ***
## fascia_eta_code_ana5 -0.010992   0.045390  -0.242  0.808661
## fascia_eta_code_ana6 -0.220052   0.072362  -3.041  0.002364 **
## fascia_eta_code_ana7 -0.437948   0.099536  -4.400  1.09e-05 ***
## fascia_eta_code_ana8 -0.045025   0.045903  -0.981  0.326680
## vdbank_flg_ana2      -0.003951   0.065450  -0.060  0.951862
## xntweb_flg_ana2      -0.018517   0.080545  -0.230  0.818176
```

```
## cliente_solo_cc_flg_ana2      0.328369    0.222911    1.473 0.140756
## cliente_investitore_flg_ana2  0.049240    0.032491    1.516 0.129676
## cliente_con_mutuo_flg_ana2   -0.048071    0.038994   -1.233 0.217690
## in_bonis_flg_comm2          0.007747    0.214918    0.036 0.971247
## multi_flg_comm2             0.013750    0.031208    0.441 0.659527
## fascia_utilizzo_online_comm2  0.212868    0.055751    3.818 0.000135 ***
## fascia_utilizzo_online_comm3  0.222015    0.222150    0.999 0.317629
## risk_rating_comm10          0.219437    0.077361    2.837 0.004570 **
## risk_rating_comm11          0.445172    0.221942    2.006 0.044903 *
## risk_rating_comm12          0.454744    0.218860    2.078 0.037754 *
## risk_rating_comm2          -0.051695    0.069913   -0.739 0.459673
## risk_rating_comm3          -0.025655    0.071424   -0.359 0.719455
## risk_rating_comm4          -0.113170    0.074459   -1.520 0.128567
## risk_rating_comm5           0.017683    0.065602    0.270 0.787513
## risk_rating_comm6           0.190256    0.069328    2.744 0.006075 **
## risk_rating_comm7          -0.020936    0.075784   -0.276 0.782351
## risk_rating_comm8          -0.183589    0.071445   -2.570 0.010194 *
## risk_rating_comm9           0.299600    0.099414    3.014 0.002588 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.381 on 10166 degrees of freedom
## Multiple R-squared:  0.03509,    Adjusted R-squared:  0.03044
## F-statistic: 7.544 on 49 and 10166 DF,  p-value: < 2.2e-16
```

## Relationships

```
modello_5 = lm(nps ~ segmento_des_comm + regione_aggregata + direzione_aggregata +
  sesso_code_ana + fascia_eta_code_ana + risk_rating_comm +
  regione_aggregata:risk_rating_comm + fascia_eta_code_ana:risk_rating_comm +
  sesso_code_ana:fascia_eta_code_ana
, data = survey_nps_tot_clean)

# summary(modello_5)
```

## Lasso

### Lasso Linear Model

```
# Identifica le variabili con coefficienti diversi da 0
selected_variables <- rownames(lasso_coefs)[lasso_coefs!= 0]
selected_variables <- selected_variables[selected_variables != "(Intercept)"] # Escludi l'intercetta

# Estrai le colonne corrispondenti alle variabili selezionate
x_selected <- x[, selected_variables]

# Costruisci un nuovo modello lineare con solo le variabili selezionate
modello_finale <- lm(survey_nps_tot_clean$nps ~ x_selected)
```

```
# Visualizza il nuovo modello
summary(modello_finale)
```

```
##
## Call:
## lm(formula = survey_nps_tot_clean$nps ~ x_selected)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.2675 -0.7906  0.2536  1.0984  2.0418
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.314520   0.236805   35.111 < 2e-16
## x_selectedsegmento_des_comm3 -0.072939   0.037373   -1.952  0.05101
## x_selectedsegmento_des_comm5  0.442338   0.801419    0.552  0.58100
## x_selectedsegmento_des_comm7  0.427034   0.054401    7.850 4.59e-15
## x_selectedregione_aggregata3  0.057273   0.049216    1.164  0.24457
## x_selectedregione_aggregata6  0.007198   0.315117    0.023  0.98178
## x_selectedregione_aggregataAltro -0.163673   0.071725   -2.282  0.02251
## x_selecteddirezione_aggregata2  0.102818   0.057840    1.778  0.07550
## x_selecteddirezione_aggregata3  0.116107   0.056674    2.049  0.04052
## x_selecteddirezione_aggregata5  0.060920   0.051632    1.180  0.23808
## x_selecteddirezione_aggregata7  0.245243   0.062026    3.954 7.74e-05
## x_selecteddirezione_aggregata9  0.224871   0.311622    0.722  0.47055
## x_selecteddirezione_aggregataAltro 0.611902   0.387987    1.577  0.11480
## x_selectedsesso_code_ana2      0.040773   0.028093    1.451  0.14671
## x_selectedfascia_eta_code_ana3  0.044205   0.034850    1.268  0.20467
## x_selectedfascia_eta_code_ana4 -0.273765   0.063507   -4.311 1.64e-05
## x_selectedfascia_eta_code_ana6 -0.215638   0.069586   -3.099  0.00195
## x_selectedfascia_eta_code_ana7 -0.390011   0.097487   -4.001 6.36e-05
## x_selectedfascia_eta_code_ana8 -0.036818   0.041204   -0.894  0.37159
## x_selecteddvdbank_flg_ana2      0.104586   0.058189    1.797  0.07231
## x_selectedxntweb_flg_ana2      -0.124844   0.054921   -2.273  0.02304
## x_selectedcliente_solo_cc_flg_ana2 0.363587   0.223241    1.629  0.10341
## x_selectedcliente_investitore_flg_ana2 0.058025   0.031987    1.814  0.06970
## x_selectedcliente_con_mutuo_flg_ana2 -0.017660   0.037280   -0.474  0.63571
## x_selectedin_bonis_flg_comm2     0.203512   0.210249    0.968  0.33309
## x_selectedmulti_flg_comm2        0.013569   0.031045    0.437  0.66206
## x_selectedfascia_utilizzo_online_comm2 0.227917   0.055720    4.090 4.34e-05
## x_selectedfascia_utilizzo_online_comm3 0.226732   0.222924    1.017  0.30914
##
## (Intercept)          ***
## x_selectedsegmento_des_comm3      .
## x_selectedsegmento_des_comm5
## x_selectedsegmento_des_comm7      ***
## x_selectedregione_aggregata3
## x_selectedregione_aggregata6
## x_selectedregione_aggregataAltro  *
## x_selecteddirezione_aggregata2    .
## x_selecteddirezione_aggregata3    *
## x_selecteddirezione_aggregata5
## x_selecteddirezione_aggregata7    ***
```

```
## x_selecteddirezione_aggregata9
## x_selecteddirezione_aggregataAltro
## x_selectedsesso_code_ana2
## x_selectedfascia_eta_code_ana3
## x_selectedfascia_eta_code_ana4      ***
## x_selectedfascia_eta_code_ana6      **
## x_selectedfascia_eta_code_ana7      ***
## x_selectedfascia_eta_code_ana8
## x_selectedvdbank_flg_ana2           .
## x_selectedxntweb_flg_ana2           *
## x_selectedcliente_solo_cc_flg_ana2
## x_selectedcliente_investitore_flg_ana2 .
## x_selectedcliente_con_mutuo_flg_ana2
## x_selectedin_bonis_flg_comm2
## x_selectedmulti_flg_comm2
## x_selectedfascia_utilizzo_online_comm2 ***
## x_selectedfascia_utilizzo_online_comm3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.386 on 10188 degrees of freedom
## Multiple R-squared:  0.02498,    Adjusted R-squared:  0.0224
## F-statistic: 9.669 on 27 and 10188 DF,  p-value: < 2.2e-16
```

```
AIC(modello_finale)
```

```
## [1] 35697.42
```

## Lasso relationships

```
# Aggiunge la variabile target 'nps' a 'x_selected'
x_selected <- as.data.frame(x_selected)
selected_variables <- colnames(x_selected)
# Crea la formula includendo tutte le variabili selezionate e interazioni di secondo ordine
formula_interazioni <- as.formula(paste("y ~ (", paste(selected_variables, collapse = " + "), ")^2"))

# Costruisci il modello lineare con interazioni di secondo ordine
modello_interazioni <- lm(formula_interazioni, data = x_selected)
```