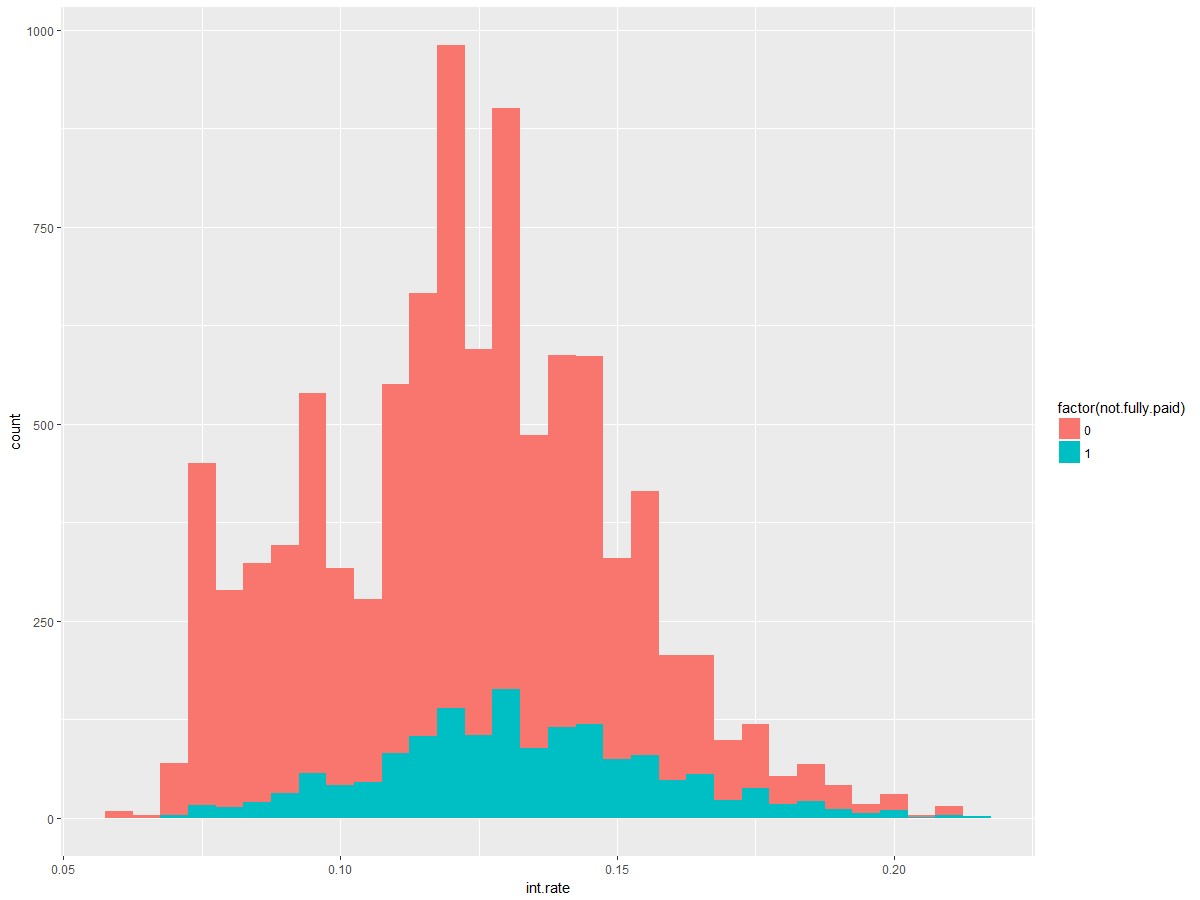
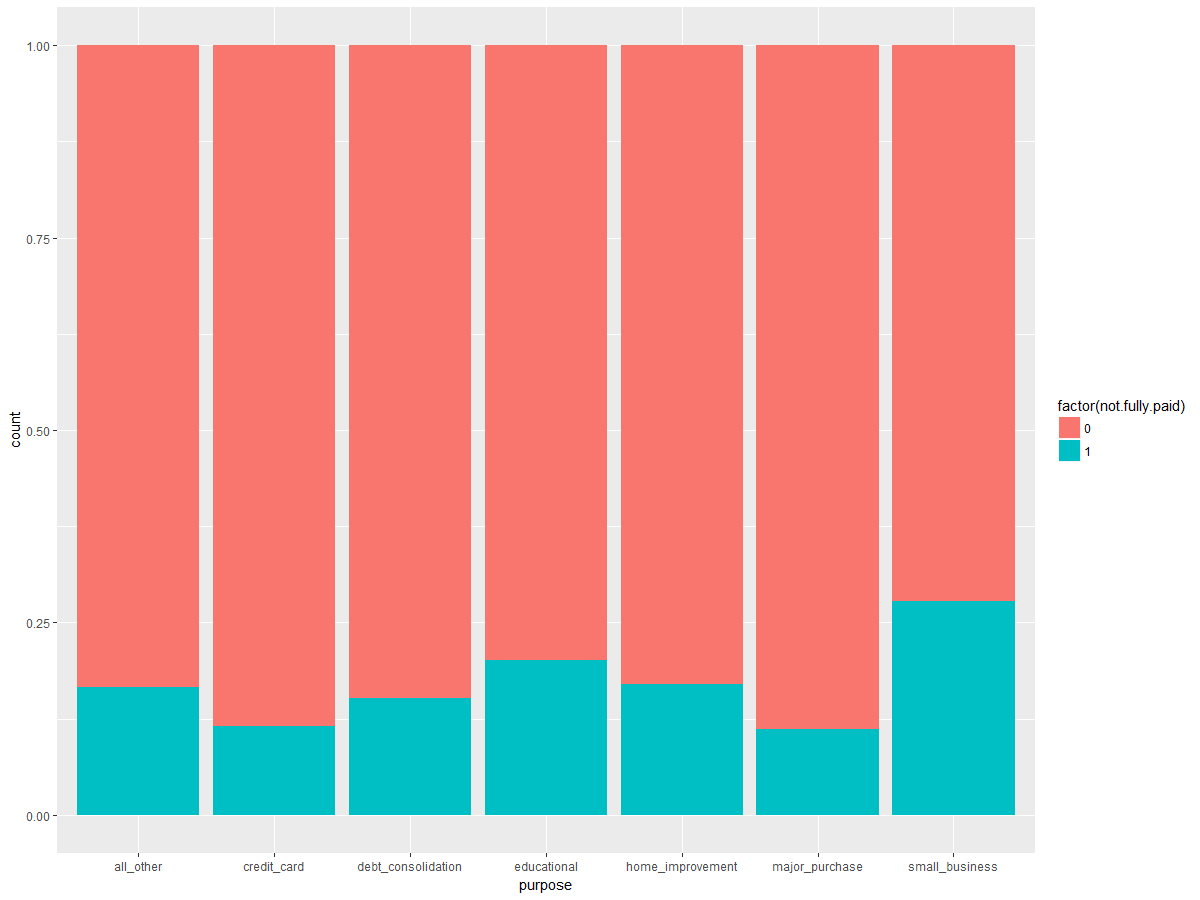
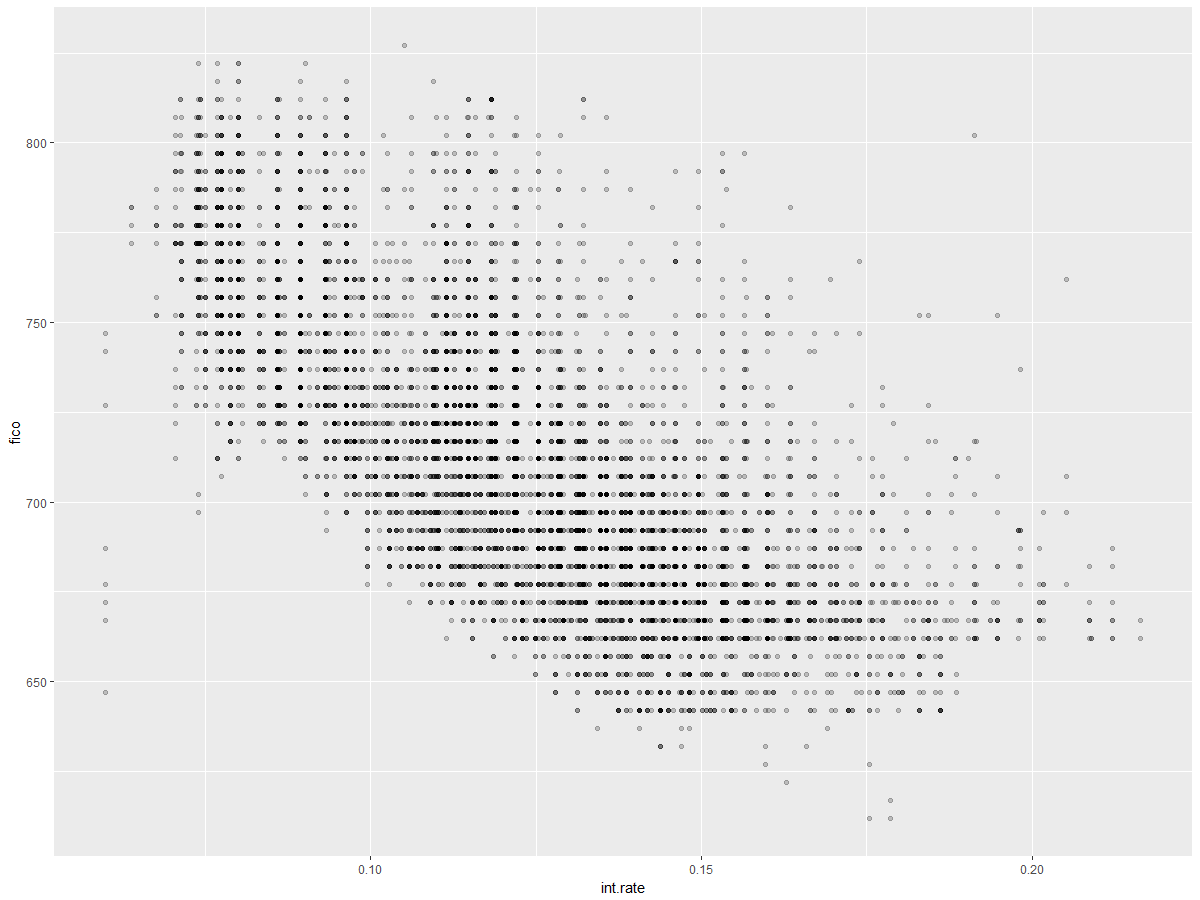
2.א מסקנה: אין כ:"כ קשר בין שיעור הריבית להחזרה, מאחר והתפלגות של מי שחזיר ומי שאינו מחזיר נראית שווה.



2.ב מסקנה ההלוואה הכי מסוכנת היא לעסקים קטנים



שאלה 2.ג מהקשר עולה כי ככל שהדרוג יותר גבוהה הריבית יותר נמוכה.

2.ד

ניראה כי יש התאמה מאוד נמוכה בין 2 העמודות, ולכן זה לא יעזור בחיזוי.

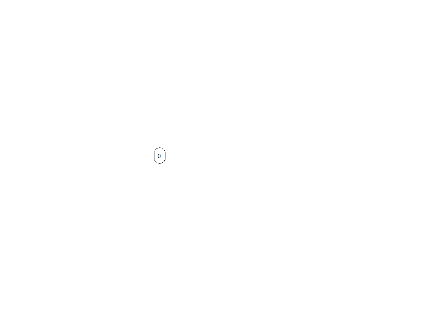
[,1] [,2]

[1,] 1.000000 -0.182402

[2,] -0.182402 1.000000

שאלה 3

לא נוצר עץ!! סיבה אפשרית: לא נמצאה אפשרות לבצע חלוקה של הנתונים. זו התמונה שיצאה.



שאלה 4

סעיף 1

Call:

glm(formula = not.fully.paid ~ ., family = binomial(link = logit),

data = testset)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.2956 -0.6177 -0.4970 -0.3676 2.6118

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 8.588e+00 1.290e+00 6.657 2.79e-11 \*\*\*

credit.policy -3.289e-01 8.419e-02 -3.906 9.37e-05 \*\*\*

purposecredit\_card -5.321e-01 1.092e-01 -4.871 1.11e-06 \*\*\*

purposedebt\_consolidation -3.372e-01 7.731e-02 -4.362 1.29e-05 \*\*\*

purposeeducational 6.800e-02 1.519e-01 0.448 0.65450

purposehome\_improvement 9.308e-02 1.263e-01 0.737 0.46123

purposemajor\_purchase -3.532e-01 1.661e-01 -2.126 0.03350 \*

purposesmall\_business 5.376e-01 1.169e-01 4.599 4.25e-06 \*\*\*

int.rate 1.534e+00 1.727e+00 0.888 0.37461

installment 1.208e-03 1.750e-04 6.900 5.18e-12 \*\*\*

log.annual.inc -4.093e-01 5.959e-02 -6.869 6.47e-12 \*\*\*

dti -3.235e-04 4.579e-03 -0.071 0.94368

fico -9.002e-03 1.421e-03 -6.337 2.34e-10 \*\*\*

days.with.cr.line 1.303e-05 1.330e-05 0.979 0.32735

revol.bal 3.026e-06 9.505e-07 3.184 0.00145 \*\*

revol.util 2.401e-03 1.277e-03 1.880 0.06012 .

inq.last.6mths 8.662e-02 1.371e-02 6.316 2.69e-10 \*\*\*

delinq.2yrs -8.808e-02 5.552e-02 -1.586 0.11267

pub.rec 2.469e-01 9.743e-02 2.534 0.01126 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 8424.0 on 9577 degrees of freedom

Residual deviance: 7852.6 on 9559 degrees of freedom

AIC: 7890.6

Number of Fisher Scoring iterations: 5

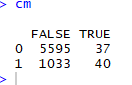
המשתנים המסומנים בצהוב משמעותיים במודל.

סעיף 2

[1] 0.1625479

תוצאת החיזוי אומרת כי ב16% מהמקרים אני יחזה טעות.

סעיף 3 על פי טבלת החיזוי- אפשר ליראות כי אנחנו חוזים בצורה טובה מי יחזיר את ההלואה, אבל לגבי החיזוי של מי לא יחזיר את ההלואה הסיכוי לחזות מאוד נמוך וברוב הפעמים המודל שלנו טועה!!!



סעיף 4- אנחנו לקחנו CUT OFF של 0.5, ניתן ליראות שככל הCUT OFF יותר נמוך נקבל יותר טעיות בחיזוי של מי שיחזיר את ההלואה, מאחר עלות אי החזרת הלוואה גדולה יותר-עדיף לקחת יותר נמוך, (בדוגמא המצורפת לקחנו CUT OFF =0.2, ניתן ליראות כי יש יותר טעיות בחיזוי.

