EXPLANATION ON THE STEPS

1. Imported the packages from library that is required.
2. Setting the working directory and loading the dataset
3. Changed the numerical variable into categorical variables using factors
4. Divided the data into training and testing set (60:40), according to the question
5. Created pivot table, using cast() and melt(),

Cast() is used to reshape the molten data

Melt() is used to convert a data frame with several measurement columns into a data frame in this canonical format, which has one row for every observed (measured) value.

1. This is the probability of loan acceptance (Loan = 1) conditional on having a bank credit card (CC = 1) and being an active user of online banking services (Online = 1) is 72/72+812= 0.0814 (done with the help of pivot table created in part A
2. Created two separate pivot tables for the training data. One will have Loan (rows) as a function of Online (columns) and the other will have Loan (rows) as a function of CC.
3. Compute the following quantities [P(A | B) means “the probability ofA given B”]:

i.P(CC = 1 | Loan = 1) (the proportion of credit card holders among the loan acceptors = 0.267658

)ii.P(Online = 1 | Loan = 1) = 0.5799257

iii.P(Loan = 1) (the proportion of loan acceptors) = 0.08966667

iv.P(CC = 1 | Loan = 0) = 0.297327

v.P(Online = 1 | Loan = 0)= 0.585866

vi.P(Loan = 0) = 0.9103333

1. Naive\_Bayes Probabiliity =(F1\*F2\*F3)/[(F1\*F2\*F3)+(F4\*F5\*F6)]

O.267658\*0.5799257\*0.08966667/(O.267658\*0.5799257\*0.08966667)+(0.297327\*0.585866\*0.9103333)

=0.013918217/0.172492615

=0.08068877

10. The part B value is more accurate as it is computed via bayes theorem.

The reason for the accuracy of the bayes theorem is the fact is calculates probability

based on the given condition. Whereas naive bayes assumes conditional independence among the variables

11. Which of the entries in this table are needed for computing P(Loan = 1 | CC = 1, Online = 1)? Run naive Bayes on the data. Examine the model output on training data, and find the entry that corresponds to P(Loan = 1 | CC = 1, Online = 1). Compare this to the number you obtained in (E).

= On comparison this value to that of E , its exactly the same, as the values in the model output on training data and the entry that corresponds to P(Loan = 1 | CC = 1, Online = 1) is same.