Assignment 3 :

Q1)
Import the affairs.csv file.

It was derived from a survey of women in 1974 by Redbook magazine, in which married women were asked about their participation in extramarital affairs.

Description of Variables

The dataset contains 6366 observations of 10 variables:(modified and cleaned)

rate_marriage: woman's rating of her marriage (1 = very poor, 5 = very good)

age: women's age

yrs_married: number of years married

children: number of children

religious: women's rating of how religious she is (1 = not religious, 4 = strongly religious)

educ: level of education (9 = grade school, 12 = high school, 14 = some college, 16 = college graduate, 17 = some graduate school, 20 = advanced degree)

occupation: women's occupation (1 = student, 2 = farming/semi-skilled/unskilled, 3 = "white collar", 4 = teacher/nurse/writer/technician/skilled, 5 = managerial/business, 6 = professional with advanced degree)

occupation_husb: husband's occupation (same coding as above)

affair: outcome 0/1, where 1 means a woman had at least 1 affair.

   Now, perform Classification using logistic regression and check your model accuracy using confusion matrix and also check the accuracy

NOTE: Perform OneHotEncoding for occupation and occupation_husb, since they should be treated as categorical variables. Careful from dummy variable trap for both!!

   What percentage of total women actually had an affair?

Q2)
Code Challenge :
Dataset - caesarian.csv
Task : Perform Knn on this dataset and show the results with confusion matrix and also print the accuracy score .