Exercise 1:

# Impute missing values (categorical variables)

train.Gender.fillna("Female",inplace=True)

train.Married.fillna("No",inplace=True)

Exercise 2:

# Impute missing values (numerical variables)

train.fillna(train.median(),inplace=True)

Exercise 1:

Text

Description automatically generated

Exercise 2:

Text

Description automatically generatedtrain.isnull().sum()

Text

Description automatically generated

train.median()

Exercise 3:

#Split train data for cross validation

from sklearn.model\_selection import train\_test\_split

x\_train,x\_cv,y\_train,y\_cv = train\_test\_split(X,y,test\_size=0.3)

Exercise 4:

#Predict values using kNN

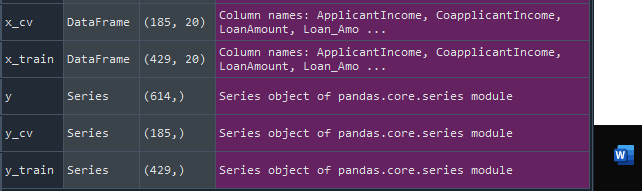
pred\_test=kNN.predict(test)

#Write test results in csv file

predictions=pd.DataFrame(pred\_test,columns=['predictions']).to\_csv('knn\_prediciton.csv')

np.unique(pred\_test, return\_counts=True)

Exercise 3:



Exercise 4:

Graphical user interface, application, table

Description automatically generated

(array(['N', 'Y'], dtype=object), array([ 60, 307], dtype=int64))

Exercise 5:

Text

Description automatically generated

Exercise 5:

the AUC-ROC score for Random Forest is 0.7466867469879518  
  
Chart, line chart

Description automatically generated