**Code :**

#Loading the test segmentation csv file into a data frame for customer segmentation

customer\_segmentation=read.csv("test\_segmentation.csv")

customer\_segmentation

#Performing Exploratory analysis

#Class of data object

class(customer\_segmentation)

#This function is used to Display Internal structure of data

str(customer\_segmentation)

#This function is used to give Summary of data

summary(customer\_segmentation)

#This function is used to give Column names

names(customer\_segmentation)

#This function is used to give Dimensions of data

dim(customer\_segmentation)

#This function is used to give the Data of the top

head(customer\_segmentation)

#This function is used to give Data from the top

tail(customer\_segmentation)

#Checking if there are any NA values in the data set

any(is.na(customer\_segmentation))

# So from the output it is understood that there are NA values in the data set

#Let us extract the count of NA values in the data set

sum(is.na(customer\_segmentation))

#Lets check the NA values column wise because there columns of both Numeric and Character

any(is.na(customer\_segmentation$ID))

any(is.na(customer\_segmentation$Gender))

any(is.na(customer\_segmentation$Ever\_Married))

any(is.na(customer\_segmentation$Age))

any(is.na(customer\_segmentation$Graduated))

any(is.na(customer\_segmentation$Profession))

any(is.na(customer\_segmentation$Work\_Experience))

any(is.na(customer\_segmentation$Spending\_Score))

any(is.na(customer\_segmentation$Family\_Size))

any(is.na(customer\_segmentation$Var\_1))

any(is.na(customer\_segmentation$Segmentation))

#So from the output it is evident that Ever\_Married , Graduated , Profession and Work experience

#Family\_size , Var\_1 have NA values

class(customer\_segmentation$Ever\_Married)

#So Ever\_Married is column which contains character values

#So we cannot replace with mean value

customer\_segmentation$Ever\_Married[is.na(customer\_segmentation$Ever\_Married)]="No"

class(customer\_segmentation$Graduated)

#So Graduated is column which contains character values

customer\_segmentation$Graduated[is.na(customer\_segmentation$Graduated)]="Yes"

class(customer\_segmentation$Profession)

#So Profession is column which contains character values

customer\_segmentation$Profession[is.na(customer\_segmentation$Profession)]="Engineer"

class(customer\_segmentation$Work\_Experience)

#So Experience is column which contains integer values

x=mean(customer\_segmentation$Work\_Experience,na.rm = TRUE)

x

# From the output we can see that we got a numeric value(i.e we got a decimal value)

#But work experience cannot be such a value. So lets either use floor or ceiling

y=floor(x)

y

customer\_segmentation$Work\_Experience[is.na(customer\_segmentation$Work\_Experience)]=y

class(customer\_segmentation$Family\_Size)

#So Family\_Size is column which contains integer values

z=mean(customer\_segmentation$Family\_Size,na.rm = TRUE)

z

# From the output we can see that we got a numeric value(i.e we got a decimal value)

w=ceiling(z)

w

customer\_segmentation$Family\_Size[is.na(customer\_segmentation$Family\_Size)]=w

class(customer\_segmentation$Var\_1)

#So Var\_1 is column which contains character values

customer\_segmentation$Var\_1[is.na(customer\_segmentation$Var\_1)]="Cat\_2"

customer\_segmentation

#Checking if there are any NA values now

any(is.na(customer\_segmentation))

#From the result it is understood that there are NA values in the data set

#So the data is now cleaned

#Writing the updated ones into new csv file named credit\_cards\_details

write.csv(customer\_segmentation,"customer\_segmentation\_cleaned.csv")





























