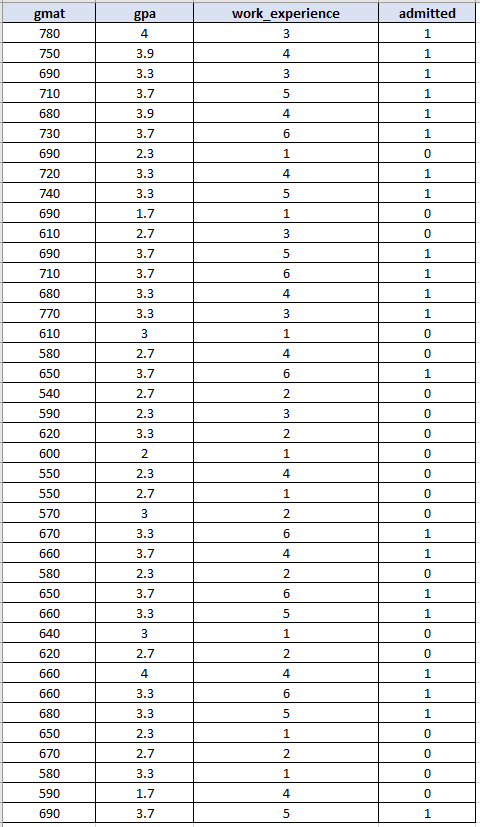
**Practical – 3: Problems related to Logistic Regression in Data Analytics**

**Description:** Our goal is to build a logistic regression model in Python in order to determine whether candidates would get admitted to a prestigious university. Here, there are two possible outcomes: Admitted (represented by the value of ‘1’) vs. Rejected (represented by the value of ‘0’). We can then build a logistic regression in Python, where:

* The dependent variable represents whether a person gets admitted; and
* The 3 independent variables are the GMAT score, GPA and Years of work experience

This is how the dataset would look like:

****

**Code:**

import pandas as pd

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

from sklearn.linear\_model import LogisticRegression

from sklearn import metrics

import seaborn as sn

candidates = {'gmat': [780,750,690,710,680,730,690,720,740,690,610,690,710,680,770,610,580,650,540,590,620,600,550,550,570,670,660,580,650,660,640,620,660,660,680,650,670,580,590,690],

'gpa': [4,3.9,3.3,3.7,3.9,3.7,2.3,3.3,3.3,1.7,2.7,3.7,3.7,3.3,3.3,3,2.7,3.7,2.7,2.3,3.3,2,2.3,2.7,3,3.3,3.7,2.3,3.7,3.3,3,2.7,4,3.3,3.3,2.3,2.7,3.3,1.7,3.7],

'work\_experience': [3,4,3,5,4,6,1,4,5,1,3,5,6,4,3,1,4,6,2,3,2,1,4,1,2,6,4,2,6,5,1,2,4,6,5,1,2,1,4,5],

'admitted': [1,1,1,1,1,1,0,1,1,0,0,1,1,1,1,0,0,1,0,0,0,0,0,0,0,1,1,0,1,1,0,0,1,1,1,0,0,0,0,1]

}

df = pd.DataFrame(candidates,columns= ['gmat', 'gpa','work\_experience','admitted'])

X = df[['gmat', 'gpa','work\_experience']]

y = df['admitted']

X\_train,X\_test,y\_train,y\_test = train\_test\_split(X,y,test\_size=0.25,random\_state=0)

logistic\_regression= LogisticRegression()

logistic\_regression.fit(X\_train,y\_train)

y\_pred=logistic\_regression.predict(X\_test)

confusion\_matrix = pd.crosstab(y\_test, y\_pred, rownames=['Actual'], colnames=['Predicted'])

sn.heatmap(confusion\_matrix, annot=True)

print('Accuracy: ',metrics.accuracy\_score(y\_test, y\_pred))

**Output:**

Accuracy: 0.8

