**SET B SPLITTING DATASET INTO TRAINING AND TESTING 2ND**

**import** pandas **as** pd

dataset**=**pd.read\_csv("Data1.csv")

dataset

|  | **Region** | **Age** | **Income** | **Online Shopper** |
| --- | --- | --- | --- | --- |
| **0** | India | 49.0 | 86400.0 | No |
| **1** | Brazil | 32.0 | 57600.0 | Yes |
| **2** | USA | 35.0 | 64800.0 | No |
| **3** | Brazil | 43.0 | 73200.0 | No |
| **4** | USA | 45.0 | NaN | Yes |
| **5** | India | 40.0 | 69600.0 | Yes |
| **6** | Brazil | NaN | 62400.0 | No |
| **7** | India | 53.0 | 94800.0 | Yes |
| **8** | USA | 55.0 | 99600.0 | No |
| **9** | India | 42.0 | 80400.0 | Yes |

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

x**=**dataset.iloc[:,:**-**1].values

y**=**dataset.iloc[:,3].values

x\_train,x\_test,y\_train,y\_test**=**train\_test\_split(x,y,test\_size**=**0.2,random\_state**=**0)

print(x\_train)

[['India' 53.0 94800.0]

['India' 40.0 69600.0]

['USA' 55.0 99600.0]

['Brazil' 43.0 73200.0]

['Brazil' 32.0 57600.0]

['Brazil' nan 62400.0]

['India' 42.0 80400.0]

['USA' 35.0 64800.0]]

print(x\_test)

[['USA' 45.0 nan]

['India' 49.0 86400.0]]

print(y\_test)

['Yes' 'No']

print(y\_train)

['Yes' 'Yes' 'No' 'No' 'Yes' 'No' 'Yes' 'No']

print(y\_test)

['Yes' 'No']

​

​