

Q7: Consider a data frame containing data about students i.e. name, gender and passing division:

```
In [2]: import numpy as np
import pandas as pd

df = pd.read_csv('F:\CS\sem 5\Data Analysis and Visualisation\Practicals\Student
df
```

```
Out[2]:
```

	Name	Birth_Month	Gender	Pass_Division
0	Mudit Chauhan	December	M	III
1	Seema Chopra	January	F	II
2	Rani Gupta	March	F	I
3	Aditya Narayan	October	M	I
4	Sanjeev Sahni	February	M	II
5	Prakash Kumar	December	M	III
6	Ritu Agarwal	September	F	I
7	Akshay Goel	August	M	I
8	Meeta Kulkarni	July	F	II
9	Preeti Ahuja	November	F	II
10	Sunil Das Gupta	April	M	III
11	Sunali Sapre	January	F	I
12	Rashmi Talwar	June	F	III
13	Ashish Dubey	May	M	II
14	Kiran Sharma	February	F	II
15	Sameer Bansal	October	M	I

a. Perform one hot encoding of the last two columns of categorical data using the `get_dummies()` function.

```
In [3]: df1 = pd.get_dummies(df, columns=['Gender', 'Pass_Division'])
df1
```

Out[3]:

	Name	Birth_Month	Gender_F	Gender_M	Pass_Division_I	Pass_Division_II	Pass_Division
0	Mudit Chauhan	December	0	1	0	0	
1	Seema Chopra	January	1	0	0	1	
2	Rani Gupta	March	1	0	1	0	
3	Aditya Narayan	October	0	1	1	0	
4	Sanjeev Sahni	February	0	1	0	1	
5	Prakash Kumar	December	0	1	0	0	
6	Ritu Agarwal	September	1	0	1	0	
7	Akshay Goel	August	0	1	1	0	
8	Meeta Kulkarni	July	1	0	0	1	
9	Preeti Ahuja	November	1	0	0	1	
10	Sunil Das Gupta	April	0	1	0	0	
11	Sunali Sapre	January	1	0	1	0	
12	Rashmi Talwar	June	1	0	0	0	
13	Ashish Dubey	May	0	1	0	1	
14	Kiran Sharma	February	1	0	0	1	
15	Sameer Bansal	October	0	1	1	0	

b. Sort this data frame on the "Birth Month" column (i.e. January to December). Hint: Convert Month to Categorical.

```
In [4]: months = ["January", "February", "March", "April", "May", "June", "July",  
                 "August", "September", "October", "November", "December"]  
  
df1['Birth_Month'] = pd.Categorical(df1['Birth_Month'], categories=months, order  
df1.sort_values(by='Birth_Month', inplace=True)  
df1
```

Out[4]:

	Name	Birth_Month	Gender_F	Gender_M	Pass_Division_I	Pass_Division_II	Pass_Division
<b>1</b>	Seema Chopra	January	1	0	0	1	
<b>11</b>	Sunali Sapre	January	1	0	1	0	
<b>4</b>	Sanjeev Sahni	February	0	1	0	1	
<b>14</b>	Kiran Sharma	February	1	0	0	1	
<b>2</b>	Rani Gupta	March	1	0	1	0	
<b>10</b>	Sunil Das Gupta	April	0	1	0	0	
<b>13</b>	Ashish Dubey	May	0	1	0	1	
<b>12</b>	Rashmi Talwar	June	1	0	0	0	
<b>8</b>	Meeta Kulkarni	July	1	0	0	1	
<b>7</b>	Akshay Goel	August	0	1	1	0	
<b>6</b>	Ritu Agarwal	September	1	0	1	0	
<b>3</b>	Aditya Narayan	October	0	1	1	0	
<b>15</b>	Sameer Bansal	October	0	1	1	0	
<b>9</b>	Preeti Ahuja	November	1	0	0	1	
<b>0</b>	Mudit Chauhan	December	0	1	0	0	
<b>5</b>	Prakash Kumar	December	0	1	0	0	