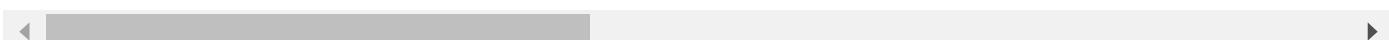


```
In [3]: #B.sruthi reddy  
#21btrcd029  
  
df=pd.read_csv("C:/Users/User/Downloads/billionaire.csv")  
df
```

Out[3]:

	Unnamed: 0	name	last_name	age	gender	time	permanent_country	company	main_in
0	675	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Divi
1	676	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Divi
2	677	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Divi
3	678	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Divi
4	679	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Divi
...	...	...	...	...	...	...	...	...	...
352364	358313	Zygmunt Solorz-Zak	Solorz-Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta
352365	358314	Zygmunt Solorz-Zak	Solorz-Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta
352366	358315	Zygmunt Solorz-Zak	Solorz-Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta
352367	358316	Zygmunt Solorz-Zak	Solorz-Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta
352368	358317	Zygmunt Solorz-Zak	Solorz-Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta

352369 rows × 18 columns

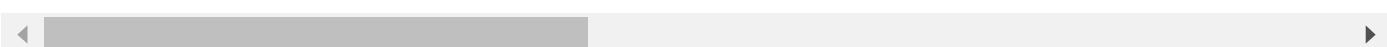


In [3]: df=pd.read\_csv("C:/Users/User/Downloads/billionaire.csv")
df

Out[3]:

		Unnamed: 0	name	last_name	age	gender	time	permanent_country	company	main_in
0	675	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
1	676	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
2	677	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
3	678	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
4	679	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
...	...	...	...	...	...	...	...	...	...	...
352364	358313	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352365	358314	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352366	358315	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352367	358316	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352368	358317	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	

352369 rows × 18 columns

In [7]: `df.dropna()  
df.head()`

Out[7]:

	0	name	last_name	age	gender	time	permanent_country	company	main_industry
0	675	Ahmad Al Ghurair	Abdulla bin	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Diversified
1	676	Ahmad Al Ghurair	Abdulla bin	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Diversified
2	677	Ahmad Al Ghurair	Abdulla bin	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Diversified
3	678	Ahmad Al Ghurair	Abdulla bin	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Diversified
4	679	Ahmad Al Ghurair	Abdulla bin	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Diversified

In [ ]:

In [8]: `df.dropna(axis=1)`

Out[8]:

		Unnamed: 0	name	age	main_industry	annual_income	daily_income	birth_year
			Abdulla bin Ahmad Al Ghurair	67	Diversified	95604000	261928	1955 abdulla_bi
0	675		Abdulla bin Ahmad Al Ghurair	67	Diversified	95604000	399254	1955 abdulla_bi
1	676		Abdulla bin Ahmad Al Ghurair	67	Diversified	95604000	536021	1955 abdulla_bi
2	677		Abdulla bin Ahmad Al Ghurair	67	Diversified	95604000	410391	1955 abdulla_bi
3	678		Abdulla bin Ahmad Al Ghurair	67	Diversified	95604000	558904	1955 abdulla_bi
4	679		Abdulla bin Ahmad Al Ghurair	67	Diversified	95604000	219780	1957
...	...	...	...	...	...	...	...	...
352364	358313	Zygmunt Solorz-Zak	65	Media & Entertainment	74820000	207493	180887	1957
352365	358314	Zygmunt Solorz-Zak	65	Media & Entertainment	74820000	231452	1957	
352366	358315	Zygmunt Solorz-Zak	65	Media & Entertainment	74820000	204986	1957	
352367	358316	Zygmunt Solorz-Zak	65	Media & Entertainment	74820000	180887	1957	
352368	358317	Zygmunt Solorz-Zak	65	Media & Entertainment	74820000	219780	1957	

352369 rows × 8 columns

In [10]: `df.head()`

Out[10]:

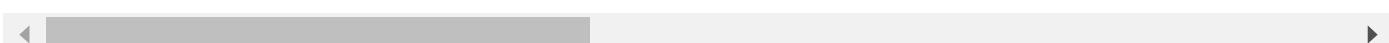
	0	name	last_name	age	gender	time	permanent_country	company	main_industry
0	675	Ahmad Al Ghurair	Abdulla bin	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Diversified
1	676	Ahmad Al Ghurair	Abdulla bin	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Diversified
2	677	Ahmad Al Ghurair	Abdulla bin	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Diversified
3	678	Ahmad Al Ghurair	Abdulla bin	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Diversified
4	679	Ahmad Al Ghurair	Abdulla bin	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Diversified

In [12]: `df.dropna(thresh=6)`

Out[12]:

		Unnamed: 0	name	last_name	age	gender	time	permanent_country	company	main_in
0	675	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
1	676	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
2	677	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
3	678	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
4	679	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
...	...	...	...	...	...	...	...	...	...	...
352364	358313	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352365	358314	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352366	358315	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352367	358316	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352368	358317	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	

352369 rows × 18 columns



In [14]: df.fillna(200)

Out[14]:

		Unnamed: 0	name	last_name	age	gender	time	permanent_country	company	main_in
0	675	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
1	676	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
2	677	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
3	678	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
4	679	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
...	...	...	...	...	...	...	...	...	...	...
352364	358313	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352365	358314	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352366	358315	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352367	358316	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352368	358317	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	

352369 rows × 18 columns

In [18]: <code>x=df["age"].mean()</code> x	

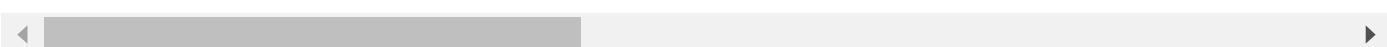
Out[18]: 72.13032644755924

In [20]: df.fillna(x)

Out[20]:

		Unnamed: 0	name	last_name	age	gender	time	permanent_country	company	main_in
0	675	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
1	676	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
2	677	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
3	678	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
4	679	Abdulla bin Ahmad Al Ghurair	Al Ghurair	67	M	2013.0	United Arab Emirates-Dubai	Mashreq Bank	Dive	
...	...	...	...	...	...	...	...	...	...	...
352364	358313	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352365	358314	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352366	358315	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352367	358316	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	
352368	358317	Zygmunt Solorz- Zak	Solorz- Zak	65	M	2022.0	Poland-Warsaw	Cyfrowy Polsat	M Enterta	

352369 rows × 18 columns


In [21]: 

```
x=df["age"].median()
x
```

```
Out[21]: 72.0
```

```
In [22]: x=df["age"].mode()  
x
```

```
Out[22]: 0    72  
Name: age, dtype: int64
```

```
In [23]: d=df["age"].nunique()  
d
```

```
Out[23]: 87
```

```
In [24]: d=df["age"].unique()  
d
```

```
Out[24]: array([ 67,  39,  53,  63,  74,  66,  77,  57,  86,  61,  84,  83,  69,  
        44,  73,  71,  104,  87,  62,  60,  85,  52,  76,  68,  45,  54,  
        94,  56,  79,  78,  72,  55,  75,  101,  99,  51,  70,  59,  89,  
        65,  80,  82,  58,  93,  64,  81,  98,  49,  36,  91,  48,  50,  
        46,  41,  43,  38,  95,  90,  42,  30,  47,  25,  96,  97,  33,  
        92,  35,  37,  40,  29,  32,  102,  31,  27,  88,  103,  100,  106,  
        34,  28,  26,  107,  117,  19,  24,  110,  111], dtype=int64)
```

```
In [5]: import pandas as pd  
df=pd.read_csv("C:/Users/User/Downloads/billionaire.csv")  
df  
y=df.replace(2,3)  
print(y)
```

	Unnamed: 0		name	last_name	age	gender	\
0	675	Abdulla bin Ahmad Al Ghurair	Al	Ghurair	67	M	
1	676	Abdulla bin Ahmad Al Ghurair	Al	Ghurair	67	M	
2	677	Abdulla bin Ahmad Al Ghurair	Al	Ghurair	67	M	
3	678	Abdulla bin Ahmad Al Ghurair	Al	Ghurair	67	M	
4	679	Abdulla bin Ahmad Al Ghurair	Al	Ghurair	67	M	
...	...		...	...	...	...	
352364	358313	Zygmunt Solorz-Zak	Solorz-Zak	Solorz-Zak	65	M	
352365	358314	Zygmunt Solorz-Zak	Solorz-Zak	Solorz-Zak	65	M	
352366	358315	Zygmunt Solorz-Zak	Solorz-Zak	Solorz-Zak	65	M	
352367	358316	Zygmunt Solorz-Zak	Solorz-Zak	Solorz-Zak	65	M	
352368	358317	Zygmunt Solorz-Zak	Solorz-Zak	Solorz-Zak	65	M	
	time	permanent_country		company			\
0	2013.0	United Arab Emirates-Dubai		Mashreq Bank			
1	2013.0	United Arab Emirates-Dubai		Mashreq Bank			
2	2013.0	United Arab Emirates-Dubai		Mashreq Bank			
3	2013.0	United Arab Emirates-Dubai		Mashreq Bank			
4	2013.0	United Arab Emirates-Dubai		Mashreq Bank			
...	...		...	...			
352364	2022.0	Poland-Warsaw	Cyfrowy Polsat				
352365	2022.0	Poland-Warsaw	Cyfrowy Polsat				
352366	2022.0	Poland-Warsaw	Cyfrowy Polsat				
352367	2022.0	Poland-Warsaw	Cyfrowy Polsat				
352368	2022.0	Poland-Warsaw	Cyfrowy Polsat				
	main_industry		wealth_source_details				\
0	Diversified		diversified				
1	Diversified		diversified				
2	Diversified		diversified				
3	Diversified		diversified				
4	Diversified		diversified				
...	...		...				
352364	Media & Entertainment	TV broadcasting; TV station					
352365	Media & Entertainment	TV broadcasting; TV station					
352366	Media & Entertainment	TV broadcasting; TV station					
352367	Media & Entertainment	TV broadcasting; TV station					
352368	Media & Entertainment	TV broadcasting; TV station					
	industry	state	headquarters				\
0		Diversified;Banking	NaN				
1		Diversified;Banking	NaN				
2		Diversified;Banking	NaN				
3		Diversified;Banking	NaN				
4		Diversified;Banking	NaN				
...	...	...	...				
352364	Media & Entertainment; Media;Media;Entertainment		NaN				
352365	Media & Entertainment; Media;Media;Entertainment		NaN				
352366	Media & Entertainment; Media;Media;Entertainment		NaN				
352367	Media & Entertainment; Media;Media;Entertainment		NaN				
352368	Media & Entertainment; Media;Media;Entertainment		NaN				
	countries	annual_income	daily_income	birth_year			\
0	are	95604000	261928	1955			
1	are	95604000	399254	1955			
2	are	95604000	536021	1955			
3	are	95604000	410391	1955			
4	are	95604000	558904	1955			
...	...	...	...	...			
352364	pol	74820000	219780	1957			

352365	pol	74820000	207493	1957
352366	pol	74820000	180887	1957
352367	pol	74820000	231452	1957
352368	pol	74820000	204986	1957

	name_cleaned
0	abdulla_bin_ahmad_al_ghurair
1	abdulla_bin_ahmad_al_ghurair
2	abdulla_bin_ahmad_al_ghurair
3	abdulla_bin_ahmad_al_ghurair
4	abdulla_bin_ahmad_al_ghurair
...	...
352364	zygmunt_solorz_zak
352365	zygmunt_solorz_zak
352366	zygmunt_solorz_zak
352367	zygmunt_solorz_zak
352368	zygmunt_solorz_zak

[352369 rows x 18 columns]

```
In [6]: c1 = pd.Categorical([1, 2, 3, 1, 2, 3])
print (c1)
c2 = pd.Categorical(['e', 'm', 'f', 'i',
                     'f', 'e', 'h', 'm'])
print (c2)
```

```
[1, 2, 3, 1, 2, 3]
Categories (3, int64): [1, 2, 3]
['e', 'm', 'f', 'i', 'f', 'e', 'h', 'm']
Categories (5, object): ['e', 'f', 'h', 'i', 'm']
```

```
In [10]: import pandas as pd
import numpy as np
```

```
In [15]: g=pd.read_excel("practice-3-03.xlsx")
gf=pd.DataFrame(g)
gf
```

Out[15]:

sno	name	qualification
0	juoo	p
1	ree	f
2	tea	p
3	gof	f
4	vgd	p

```
In [29]: #gf["qualification"].value_counts()
#gf
gf["qualification"].value_counts()
gf
```

Out[29]:

	sno	name	qualification
0	1	juoo	p
1	2	ree	f
2	3	tea	p
3	4	gof	f
4	5	vgd	p

```
In [24]: encode={"qualification":{"p":1,"f":0}}
encoded_gf=gf.replace(encode)
encoded_gf
```

Out[24]:

	sno	name	qualification
0	1	juoo	1
1	2	ree	0
2	3	tea	1
3	4	gof	0
4	5	vgd	1

```
In [32]: output=encoded_gf.to_excel("info.xlsx")
output
```

```
In [41]: info=pd.read_excel("info.xlsx")
info
```

Out[41]:

	Unnamed: 0	sno	name	qualification
0	0	1	juoo	1
1	1	2	ree	0
2	2	3	tea	1
3	3	4	gof	0
4	4	5	vgd	1

```
In [39]: type(info["qualification"])
```

Out[39]:

pandas.core.series.Series

```
In [43]: info["qualification"]=info["qualification"].astype("category")
info.dtypes
```

```
Out[43]: Unnamed: 0          int64
sno            int64
name           object
qualification   category
dtype: object
```

```
In [46]: info["qualification_new"] = info["qualification"].cat.codes
info
```

```
Out[46]:   Unnamed: 0  sno  name  qualification  qualification_new
0          0    1  juoo           1              1
1          1    2   ree            0              0
2          2    3   tea            1              1
3          3    4   gof            0              0
4          4    5   vgd            1              1
```

```
In [47]: pd.get_dummies(info, columns=["qualification"]).head()
```

```
Out[47]:   Unnamed: 0  sno  name  qualification_new  qualification_0  qualification_1
0          0    1  juoo           1              0              1
1          1    2   ree            0              1              0
2          2    3   tea            1              0              1
3          3    4   gof            0              1              0
4          4    5   vgd            1              0              1
```

```
In [8]: import pandas as pd
info=pd.read_excel("info.xlsx")
from sklearn.preprocessing import OrdinalEncoder
```

```
In [9]: obj=OrdinalEncoder()
```

```
In [16]: info["encode"] = obj.fit_transform(info[["qualification"]])
print(info["encode"])

0    1.0
1    0.0
2    1.0
3    0.0
4    1.0
Name: encode, dtype: float64
```

```
In [14]: info.value_counts()
```

```
Out[14]:   Unnamed: 0  sno  name  qualification  encode
0          1    1  juoo           1        1.0      1
1          2    2   ree            0        0.0      1
2          3    3   tea            1        1.0      1
3          4    4   gof            0        0.0      1
4          5    5   vgd            1        1.0      1
dtype: int64
```

In [ ]: 2) Write short description of encoding & its methods

Encoding in Python refers to the process of converting data from one format to another. There are several encoding methods available in Python, including:

1.ASCII encoding: This method converts ASCII characters to binary format.

2.Unicode encoding: This method represents characters **in** a standard format, allowing **for** multilingual support.

3.UTF-8 encoding: This **is** a variable-width encoding method that can represent any Unicode character.

4.Base64 encoding: This method **is** used to convert binary data into ASCII characters, making it easier to transmit over text-based protocols

In [ ]:

In [ ]:

In [ ]: