```
'''
This program calculates the entropy and gain of the Outlook subset using datastructures and d

'''
```

```
    '\nThis program calculates the entropy and gain of the Outlook subset using datastructur
```

```
import pandas as pd
import numpy as np
```

```
df = pd.read_csv('D:\DS\Data\Tennis.csv')
```

```
df
```

	Day	Outlook	Temp	Humidity	Wind	PlayTennis
0	D1	Sunny	Hot	High	Weak	No
1	D2	Sunny	Hot	High	Strong	No
2	D3	Overcast	Hot	High	Weak	Yes
3	D4	Rain	Mild	High	Weak	Yes
4	D5	Rain	Cool	Normal	Weak	Yes
5	D6	Rain	Cool	Normal	Strong	No
6	D7	Overcast	Cool	Normal	Strong	Yes
7	D8	Sunny	Mild	High	Weak	No
8	D9	Sunny	Cool	Normal	Weak	Yes
9	D10	Rain	Mild	Normal	Weak	Yes
10	D11	Sunny	Mild	Normal	Strong	Yes
11	D12	Overcast	Mild	High	Strong	Yes
12	D13	Overcast	Hot	Normal	Weak	Yes
13	D14	Rain	Mild	High	Strong	No

```
S=df[['Outlook', 'PlayTennis']]
S
```

	Outlook	PlayTennis
0	Sunny	No
1	Sunny	No
2	Overcast	Yes
3	Rain	Yes
4	Rain	Yes
5	Rain	No
6	Overcast	Yes
7	Sunny	No
8	Sunny	Yes
9	Rain	Yes
10	Sunny	Yes
11	Overcast	Yes

```
# Entropy for PlayTennis
```

13   Rain   No

```
A=S.groupby(['PlayTennis']).count()

A
```

	Outlook
PlayTennis	
No	5
Yes	9

```
B= df['PlayTennis'].value_counts(normalize=True, sort=False)
B
```

```
No     0.357143
Yes    0.642857
Name: PlayTennis, dtype: float64
```

```
A['Probability']=B
A
```

	Outlook	Probability
PlayTennis		

```
A['Entropy']= (-(A['Probability'])* np.log2(A['Probability'])-(1-(A['Probability']))*np.log2(
A
```

	Outlook	Probability	Entropy
PlayTennis			
No	5	0.357143	0.940286
Yes	9	0.642857	0.940286

```
# NEXT ENTROPIES for Outlook or other subsets
```

```
C=df.groupby('Outlook')['PlayTennis'].value_counts(normalize=True, sort=False)
```

```
C=C.unique()
C
```

```
array([1. , 0.4, 0.6])
```

```
W=S.groupby(["Outlook"]).count()
W
```

	PlayTennis
Outlook	
Overcast	4
Rain	5
Sunny	5

```
W['Probability']=C
W
```

	PlayTennis	Probability
Outlook		
Overcast	4	1.0
Rain	5	0.4
Sunny	5	0.6

```
W['Entropy']= (-(W['Probability'])* np.log2(W['Probability'])-(1-(W['Probability']))*np.log2(
W
```

```
C:\Users\fakhi\AppData\Local\Programs\Python\Python310\lib\site-packages\pandas\core\arr
  result = getattr(ufunc, method)(*inputs, **kwargs)
```

	PlayTennis	Probability	Entropy
Outlook			
Overcast	4	1.0	NaN
Rain	5	0.4	0.970951
Sunny	5	0.6	0.970951

```
for i in range(len(W)):


    if (W.iat[i, 1]) == 1:
        (W.iat[i, 2]) = 0
```

```
W.reset_index()
```

	Outlook	PlayTennis	Probability	Entropy
0	Overcast	4	1.0	0.000000
1	Rain	5	0.4	0.970951
2	Sunny	5	0.6	0.970951

```
T=len(df)
T
```

```
    14
```

```
total=0

for i in range(len(W)):
    # print(W.iat[i, 0])
    # print(W.iat[i, 2])
    # print(T)
    # print(A.iat[0, 2])

    total = total + (W.iat[i, 0])/T*(W.iat[i, 2])
    # print(total)
```

```
Gain = (A.iat[0, 2]) - total
print('Gain', Gain)
```

    Gain 0.24674981977443933