```
In [21]: import numpy as np
         import pandas as pd
         import math
```

```
In [4]: predict_df  = pd.read_csv("data\PredictingSignupsTrain.csv")
        predict_df
```

Out[4]:

	Referrer	Location	Read_FAQ	Pages_Viewed	Service_Chosen
0	Slashdot	USA	Yes	Mid	None
1	Google	France	Yes	High	Premium
2	Digg	USA	Yes	High	Basic
3	Kiwitobes	France	Yes	High	Basic
4	Google	UK	No	Mid	Premium
5	(direct)	'New Zealand'	No	Low	None
6	(direct)	UK	No	Mid	Basic
7	Google	USA	No	High	Premium
8	Slashdot	France	Yes	Mid	None
9	Digg	USA	No	Mid	None
10	Google	UK	No	Mid	None
11	Kiwitobes	UK	No	Mid	None
12	Digg	'New Zealand'	Yes	Low	Basic
13	Google	UK	Yes	Mid	Basic
14	Kiwitobes	France	Yes	Mid	Basic
15	Google	UK	No	Mid	Premium
16	Digg	USA	No	Low	Basic
17	Slashdot	'New Zealand'	Yes	High	None

In [18]:
```python
class_none = predict_df[predict_df["Service_Chosen"]=='None']
print("Class None:\n", class_none['Location'].value_counts())
class_none = predict_df[predict_df["Service_Chosen"]=='Basic']
print("Class Basic:\n", class_none['Location'].value_counts())
class_none = predict_df[predict_df["Service_Chosen"]=='Premium']
print("Class Premium:\n", class_none['Location'].value_counts())
```

```
Class None:
 USA               2
'New Zealand'     2
UK                2
France            1
Name: Location, dtype: int64
Class Basic:
 France            2
USA               2
UK                2
'New Zealand'     1
Name: Location, dtype: int64
Class Premium:
 UK          2
France      1
USA         1
Name: Location, dtype: int64
```

In [19]:
```python
class_none = predict_df[predict_df["Service_Chosen"]=='None']
print("Class None:\n", class_none['Read_FAQ'].value_counts())
class_none = predict_df[predict_df["Service_Chosen"]=='Basic']
print("Class Basic:\n", class_none['Read_FAQ'].value_counts())
class_none = predict_df[predict_df["Service_Chosen"]=='Premium']
print("Class Premium:\n", class_none['Read_FAQ'].value_counts())
```

```
Class None:
 No     4
Yes    3
Name: Read_FAQ, dtype: int64
Class Basic:
 Yes     5
No      2
Name: Read_FAQ, dtype: int64
Class Premium:
 No      3
Yes     1
Name: Read_FAQ, dtype: int64
```

```
In [20]: class_none = predict_df[predict_df["Service_Chosen"]=='None']
         print("Class None:\n", class_none['Pages_Viewed'].value_counts())
         class_none = predict_df[predict_df["Service_Chosen"]=='Basic']
         print("Class Basic:\n", class_none['Pages_Viewed'].value_counts())
         class_none = predict_df[predict_df["Service_Chosen"]=='Premium']
         print("Class Premium:\n", class_none['Pages_Viewed'].value_counts())
```

```
Class None:
 Mid     5
Low      1
High     1
Name: Pages_Viewed, dtype: int64
Class Basic:
 Mid     3
Low      2
High     2
Name: Pages_Viewed, dtype: int64
Class Premium:
 Mid     2
High     2
Name: Pages_Viewed, dtype: int64
```

```
In [29]: info_ref_s = -(3/3)*math.log2(3/3)
         info_ref_g = -(1/6)*math.log2(1/6) - (1/6)*math.log2(1/6) - \
             (4/6)*math.log2(4/6)
         info_ref_d = -(1/4)*math.log2(1/4) - (3/4)*math.log2(3/4)
         info_ref_k = -(1/3)*math.log2(1/3) - (2/3)*math.log2(2/3)
         info_ref_di = -(1/2)*math.log2(1/2) - (1/2)*math.log2(1/2)
         info_ref = (3/18)*info_ref_s + (6/18)*info_ref_g + 4/18 * info_ref_d + \
             3/18 * info_ref_k + 2/18 * info_ref_di
         info_ref
```

Out[29]: 0.8616541669070521

```
In [33]: info_loc = [0,0,0,0]
         info_loc[0] = -(2/5)*math.log2(2/5) - (2/5)*math.log2(2/5) - \
             (1/5)*math.log2(1/5)
         info_loc[1] = -(1/4)*math.log2(1/4) - (2/4)*math.log2(2/4) - \
             (1/4)*math.log2(1/4)
         info_loc[2] = -3*(2/6)*math.log2(2/6)
         info_loc[3] = -(2/3)*math.log2(2/3) - (1/3)*math.log2(1/3)
         info_loc = (5/18)*info_loc[0] + (4/18)*info_loc[1] + 6/18 * info_loc[2] + \
             3/18 * info_loc[3]
         info_loc
```

Out[33]: 1.4374612767181787

```
In [34]: info_service = -2*(7/18)*math.log2(7/18) - (4/18)*math.log2(4/18)
         info_service
```

Out[34]: 1.541982284286398