

解题代码

```
import pandas as pd

def Task():
    # 使用pandas库的read_excel方法读入数据中医数据
    #***** BEIGN *****

    data = pd.read_excel('data/data.xls')
    answer_1 = data.head(5)

    #***** END *****

    #***** BEIGN *****

    #观察数据属性类型是否符合算法要求
    info = data.info()
    answer_2 = info

    #***** END *****

    #***** BEIGN *****

    # 使用合适的策略对数据进行离散化处理
    typelabel = dict(zip(data.columns[:-1], 'ABCDEF'))
    #***** END *****

    keys = list(typelabel.keys())
    answer_3 = keys

    # 等频离散
    k = 4

    datas = list()

    #***** BEIGN *****

    for j in keys:
        label = [typelabel[j]+str(i+1) for i in range(k)]
        d = pd.qcut(data[j], k, labels=label)
        datas.append(d)

    #***** END *****

    datas.append(data['TNM分期'])
    # 经过离散化处理后的数据集
    datas = pd.DataFrame(datas).T
    answer_4 = datas.head(5)
```

#将离散化后的数据集存储到apriori.txt文件中

```
filepath = 'data/apriori.txt'
```

```
datas.to_csv(filepath, header=0, index=0, sep=',')
```

```
return answer_1, answer_2, answer_3, answer_4
```