In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
var=pd.read_csv('C://Users/Gopi/Desktop/santa-2019-revenge-of-the-accountants/data.csv')
```

In [3]:
```
print(var.shape)
```

```
(6000, 12)
```

In [4]:
```
var.columns
```

Out[4]:
```
Index(['family_id', 'choice_0', 'choice_1', 'choice_2', 'choice_3', 'choice_4',
       'choice_5', 'choice_6', 'choice_7', 'choice_8', 'choice_9', 'n_people'],
      dtype='object')
```

In [5]:
```
sample=var
```

In [6]:
```
y=sample['n_people']
```

In [9]:
```
del var['n_people']
```

In [16]:
```
del var['family_id']
```

In [17]:
```
X=var
```

In [18]:
```
print(X.shape)
print(y.shape)
```

```
(6000, 10)
(6000,)
```

In [19]:
```
from sklearn.tree import DecisionTreeClassifier
tree = DecisionTreeClassifier(criterion = 'entropy', random_state = 0)
tree.fit(X,y)
```

Out[19]:
```
DecisionTreeClassifier(class_weight=None, criterion='entropy', max_depth=None,
                       max_features=None, max_leaf_nodes=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, presort=False,
                       random_state=0, splitter='best')
```

In [20]:
```
y_pred=tree.predict(X)
print(y_pred)
```

```
[6 2 2 ... 3 3 5]
```

In [21]:
```
pred=pd.DataFrame(y_pred)
sub_df=pd.read_csv('C://Users/Gopi/Desktop/santa-2019-revenge-of-the-accountants/sample_submission.csv')
datasets=pd.concat([sub_df['family_id'],pred],axis=1)
datasets.columns=['family_id','assigned_day']
datasets.to_csv('sample_submission.csv',index=False)
```