In [1]:
```
import pandas as pd
```

In [2]:
```
df=pd.read_csv("asmoutputfile.csv")
```

In [3]:
```
df.columns
```

Out[3]:
```
Index(['ID', 'HEADER:', '.text:', '.Pav:', '.idata:', '.data:', '.bss:',
       '.rdata:', '.edata:', '.rsrc:', '.tls:', '.reloc:', '.BSS:', '.CODE',
       'jmp', 'mov', 'retf', 'push', 'pop', 'xor', 'retn', 'nop', 'sub', 'inc',
       'dec', 'add', 'imul', 'xchg', 'or', 'shr', 'cmp', 'call', 'shl', 'ror',
       'rol', 'jnb', 'jz', 'rtn', 'lea', 'movzx', '.dll', 'std::', ':dword',
       'edx', 'esi', 'eax', 'ebx', 'ecx', 'edi', 'ebp', 'esp', 'eip'],
      dtype='object')
```

In [4]:
```
df.shape
```

Out[4]:
```
(10868, 52)
```

In [5]:
```
df1=pd.read_csv("Bytes_final_bigram.csv")
```

In [6]:
```
df1.columns
```

Out[6]:
```
Index(['Unnamed: 0.1', 'FileName', '?? ??', 'Unnamed: 0', 'FileName.1',
       '79 4f', '4f 79', '79 2a', '2a 79', '79 6c',
       ...
       '7c 06', '5d 17', '7a f5', '58 34', '17 5d', '7a 7c', '4b 4c', '86 04',
       '04 df', '47 cb'],
      dtype='object', length=2002)
```

In [9]:
```
df1.rename(columns={'FileName':'ID'})
df1.drop(columns=['Unnamed: 0.1','Unnamed: 0','FileName.1'],axis=1,inplace=True)
```

In [11]:
```
df1.rename(columns={'FileName':'ID'},inplace=True)
```

In [12]:
```
df1.columns
```

Out[12]:
```
Index(['ID', '?? ??', '79 4f', '4f 79', '79 2a', '2a 79', '79 6c', '04 0d',
       '73 79', '6c 79',
       ...
       '7c 06', '5d 17', '7a f5', '58 34', '17 5d', '7a 7c', '4b 4c', '86 04',
       '04 df', '47 cb'],
      dtype='object', length=1999)
```

In [13]:
```
df1.shape
```

Out[13]:
```
(10868, 1999)
```

In [14]:
```
df2=pd.read_csv("result_with_size.csv")
```

In [15]:
```
df2.columns
```

Out[15]:
```
Index(['Unnamed: 0', 'ID', '0', '1', '2', '3', '4', '5', '6', '7',
       ...
       'f9', 'fa', 'fb', 'fc', 'fd', 'fe', 'ff', '??', 'size', 'Class'],
      dtype='object', length=261)
```

In [17]:
```
df2.drop(['Unnamed: 0'],axis=1,inplace=True)
```

In [18]:
```
df2.columns
```

Out[18]:
```
Index(['ID', '0', '1', '2', '3', '4', '5', '6', '7', '8',
       ...
       'f9', 'fa', 'fb', 'fc', 'fd', 'fe', 'ff', '??', 'size', 'Class'],
      dtype='object', length=260)
```

In [19]:
```
df2.shape
```

Out[19]:
```
(10868, 260)
```

In [20]:
```
output_lables=pd.read_csv("trainLabels.csv")
```

In [25]:
```
output_lables.rename(columns={'Id':'ID'},inplace=True)
output_lables.columns
```

Out[25]:
```
Index(['ID', 'Class'], dtype='object')
```

In [22]:
```
df3=pd.read_csv("pixel_asm_file.csv")
```

In [23]:
```
df3.shape
```

Out[23]:
```
(10866, 201)
```

In [24]:
```
df3.columns
```

Out[24]:
```
Index(['pixel1', 'pixel2', 'pixel3', 'pixel4', 'pixel5', 'pixel6', 'pixel7',
       'pixel8', 'pixel9', 'pixel10',
       ...
       'pixel192', 'pixel193', 'pixel194', 'pixel195', 'pixel196', 'pixel197',
       'pixel198', 'pixel199', 'pixel200', 'ID'],
      dtype='object', length=201)
```

In [32]:
```
df4=pd.merge(df1,df2,on='ID',how='inner')
```

In [33]:
```
df5=pd.merge(df3,output_lables,on='ID',how='inner')
```

In [34]:
```
final_df=pd.merge(df4,df5,on='ID',how='inner')
```

In [35]:
```
final_df.shape
```

Out[35]:
```
(10866, 2459)
```

In [39]:
```
final_df.to_csv("microsoft_final_file.csv")
```

In [ ]:
```

```