- Name : Neeraj Boyapati
- Unique Id : E0320031
- Subject : Advanced Python Programming (Lab Assessment)
- Subject Code : CSE 180

-- Breast Cancer -- (SVM)

In [44]:
```
import pandas as pd
```

In [45]:
```
import numpy as np
```

In [46]:
```
import matplotlib.pyplot as plt
```

In [47]:
```
import seaborn as sns
```

In [5]:
```
# Import Dataset
df=pd.read_csv("brestcancer.csv")
df.head()
```

Out[5]:

	Unnamed: 0	id	clump_thickness	uniform_cell_size	uniform_cell_shape	marginal_adhesion
0	0	1000025	5	1	1	1
1	1	1002945	5	4	4	5
2	2	1015425	3	1	1	1
3	3	1016277	6	8	8	1
4	4	1017023	4	1	1	3

In [7]:
```
df['bare_nuclei'] = df['bare_nuclei'].replace('?','0')
df['bare_nuclei'].unique()
```

Out[7]:
```
array(['1', '10', '2', '4', '3', '9', '7', '0', '5', '8', '6'],
      dtype=object)
```

In [8]:
```
df.bare_nuclei=pd.to_numeric(df.bare_nuclei).astype("int64")
df.dtypes
```

Out[8]:
```
Unnamed: 0                int64
id                        int64
clump_thickness           int64
uniform_cell_size         int64
uniform_cell_shape        int64
marginal_adhesion         int64
single_epithelial_size    int64
bare_nuclei               int64
bland_chromatin           int64
normal_nucleoli           int64
mitoses                   int64
class                     int64
dtype: object
```

In [9]:
```
# Check for null values
pd.isnull(df).sum()
```

Out[9]:
```
Unnamed: 0                0
id                        0
clump_thickness           0
uniform_cell_size         0
uniform_cell_shape        0
marginal_adhesion         0
single_epithelial_size    0
bare_nuclei               0
bland_chromatin           0
normal_nucleoli           0
mitoses                   0
class                     0
dtype: int64
```

In [10]:
```
# Outcome Feature (Class) values
df['class'].unique()
```

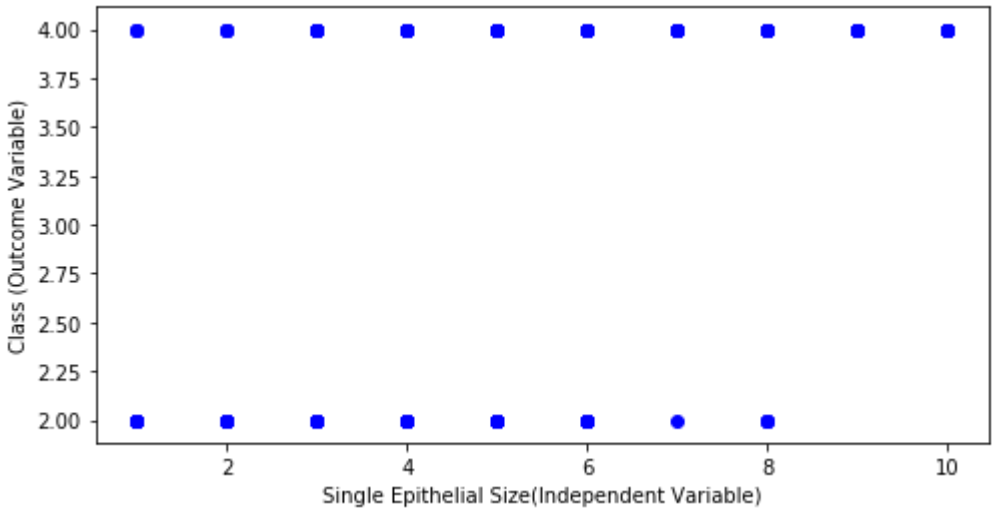
Out[10]:
```
array([2, 4], dtype=int64)
```

In [11]:
```
# Data Description
df.describe()
```

Out[11]:

	Unnamed: 0	id	clump_thickness	uniform_cell_size	uniform_cell_shape	marginal_
count	699.000000	6.990000e+02	699.000000	699.000000	699.000000	6
mean	349.000000	1.071704e+06	4.417740	3.134478	3.207439	
std	201.928205	6.170957e+05	2.815741	3.051459	2.971913	
min	0.000000	6.163400e+04	1.000000	1.000000	1.000000	
25%	174.500000	8.706885e+05	2.000000	1.000000	1.000000	
50%	349.000000	1.171710e+06	4.000000	1.000000	1.000000	
75%	523.500000	1.238298e+06	6.000000	5.000000	5.000000	
max	698.000000	1.345435e+07	10.000000	10.000000	10.000000	

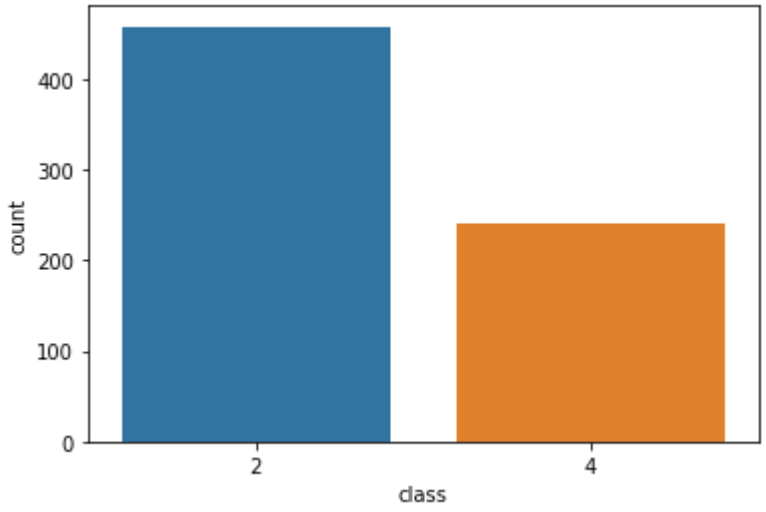
In [12]:
```
# Scatter Plot
X=df["clump_thickness"]
Y=df["class"]
plt.figure(figsize=(8,4))
plt.scatter(X,Y, marker='o', color='blue')
plt.xlabel("Single Epithelial Size(Independent Variable)")
plt.ylabel("Class (Outcome Variable)")
plt.show()
```



- This Scatter Plot depicts Class values with respect to their respective Clump Thickness values

In [13]:
```
# Count Plot
sns.countplot(x="class",data=df)
```

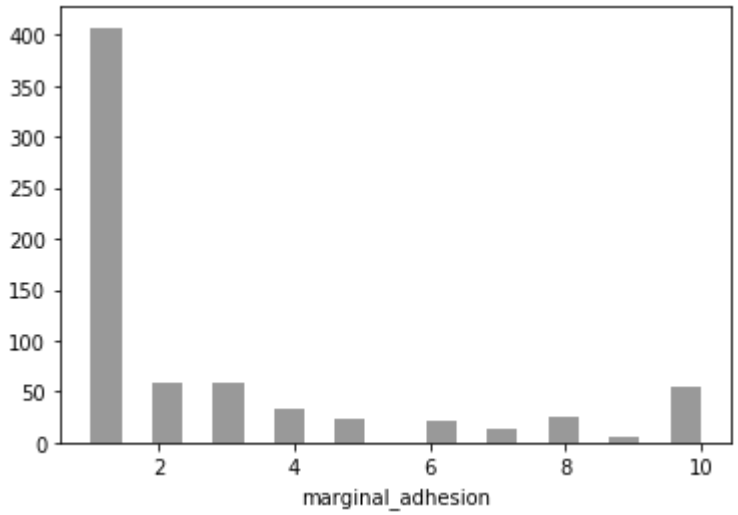
Out[13]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x9176c58160>
```



- This is the Count Plot which shows the total number of Benign and Malignant Cases which are shown as 2 and 4 respectively.

In [49]:
```
sns.distplot(df['marginal_adhesion'],color='black',bins=20,kde=False)
```

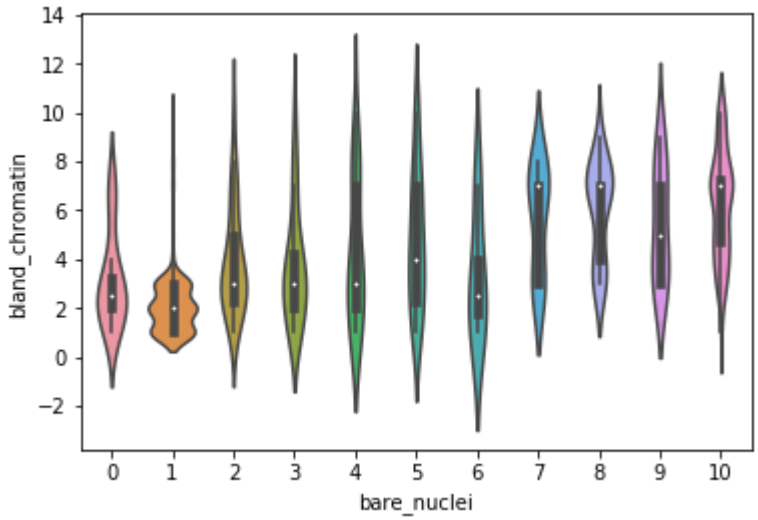
Out[49]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x9170b9ba20>
```



- This Distplot displays the variation of the Class values with respect to their respective Marginal Adhesion Values

In [51]:
```
# Violin Plot
sns.violinplot(x='bare_nuclei',y='bland_chromatin',data=df)
```

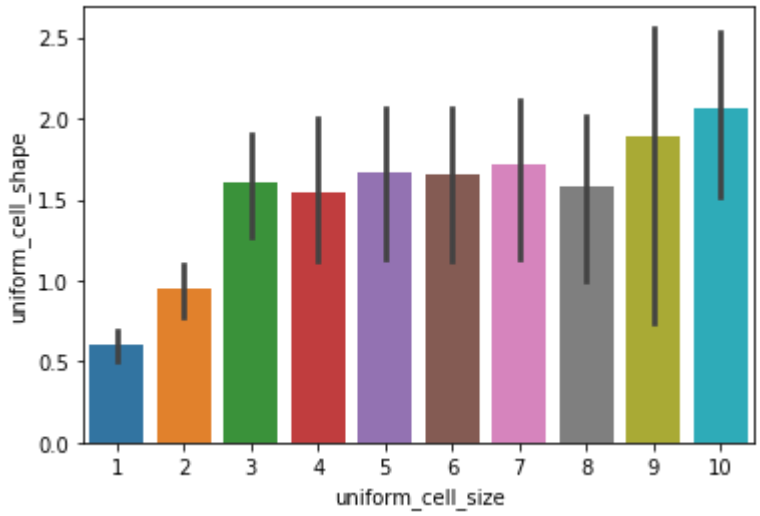
Out[51]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x91773989e8>
```



- This Violin Plot displays the relation between the Bare Nuclei and Bland Chromatin

In [52]:
```
# Barplot
sns.barplot(x='uniform_cell_size',y='uniform_cell_shape',data=df,estimator=np.st
```

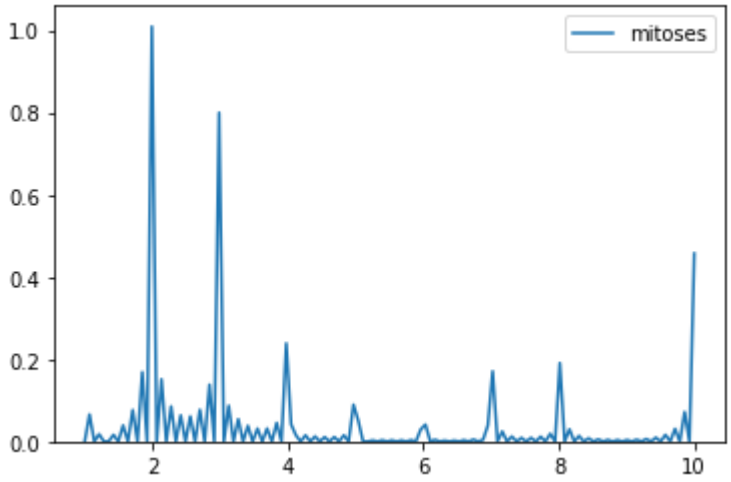
Out[52]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x9177453fd0>
```



- This Bar Plot displays the relation between the Uniform Cell Size and Uniform Cell Shape

In [54]:
```
sns.kdeplot(df['mitoses'])
```

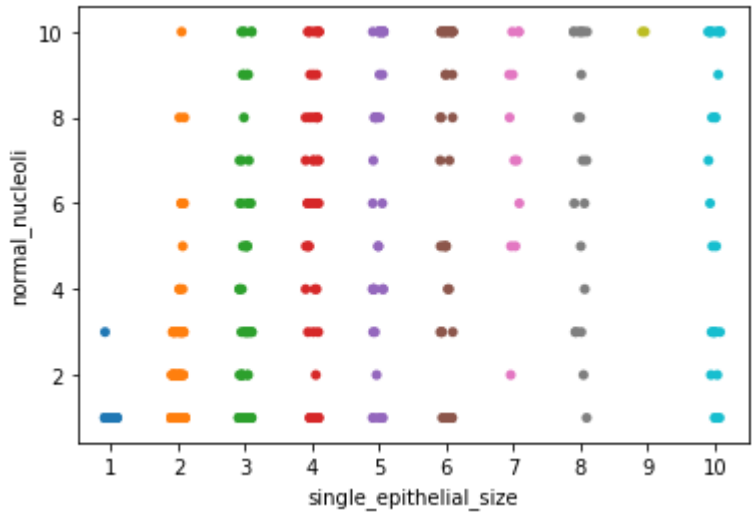
Out[54]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x917757ca20>
```



- This Distplot displays the variation of the Class values with respect to their respective Mitoses Values

In [55]:
```
# Strip Plot
sns.stripplot(x='single_epithelial_size',y='normal_nucleoli',data=df,jitter=True
```

Out[55]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x91775de828>
```



- This Bar Plot displays the relation between the Single Epothelial Size and Normal Nucleoli

In [57]:
```
from sklearn.model_selection import train_test_split
```

In [58]:
```
# Separating the Dependent and Independent Features
x=df[["clump_thickness","uniform_cell_size","uniform_cell_shape","marginal_adhes
      "single_epithelial_size","bare_nuclei","bland_chromatin","normal_nucleoli"
y=df["class"]
```

In [32]:
```
# x_train & y_train for Train the model
# x_test & y_test for Test/ Predict model
x_train,x_test,y_train,y_test = train_test_split(x,y,train_size=0.8,random_state
```

In [33]:
```
# Check train and test data shapes
print(x_train.shape)
print(x_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(559, 9)
(140, 9)
(559,)
(140,)
```

In [34]:
```
from sklearn import svm
```

In [35]:
```
# Build SVM Classifier Model (Support Vector Classifier)
svc = svm.SVC(kernel= 'linear')
```

In [36]:
```
# Train Model
svc.fit(x_train,y_train)
```

Out[36]:
```
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
    kernel='linear', max_iter=-1, probability=False, random_state=None,
    shrinking=True, tol=0.001, verbose=False)
```

In [37]:
```
# Predicting Outcome Class
y_pred = svc.predict(x_test)
```

In [38]:
```
from sklearn.metrics import confusion_matrix, accuracy_score
```

In [39]:
```
# Confusion Matrix from predicted and actual class values
cm = confusion_matrix(y_test, y_pred)
print("Confusion Matrix :")
print(cm)
```

```
Confusion Matrix :
[[82  3]
 [ 1 54]]
```

In [40]:
```
# Accuracy Score of Class prediction
acc=accuracy_score(y_test, y_pred)*100
print("Accuracy = ",acc,"%")
```

```
Accuracy =  97.14285714285714 %
```

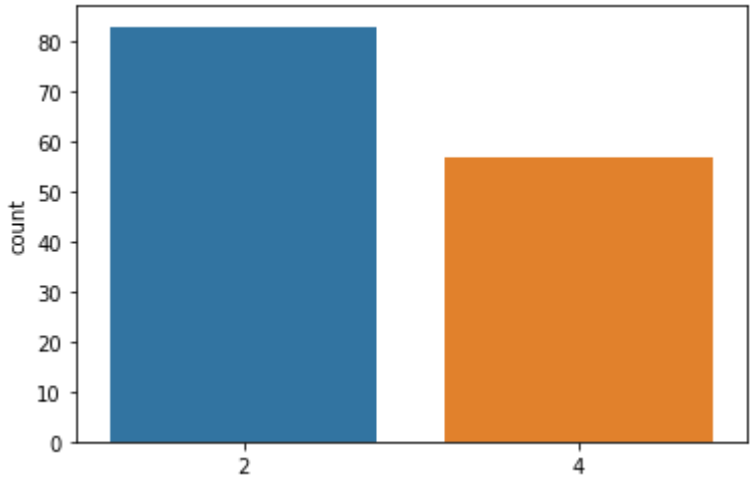
In [41]:
```
# Comparision of Real and Predicted Class values
out = pd.DataFrame({'Real_class': y_test, 'Predicted_class': y_pred})
out.head(10)
```

Out[41]:

	Real_class	Predicted_class
476	2	2
531	2	2
40	2	4
432	2	2
14	4	4
157	2	2
266	4	4
31	2	2
251	4	4
103	4	4

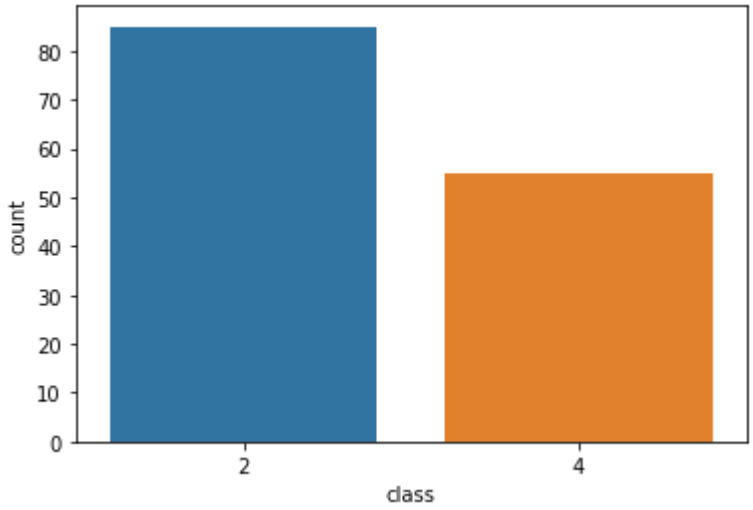
In [42]:
```
sns.countplot(y_pred,data=df)
```

Out[42]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x9170ae5cf8>
```



In [43]:
```
sns.countplot(y_test,data=df)
```

Out[43]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x9170b2d048>
```



In [ ]:

In [ ]: