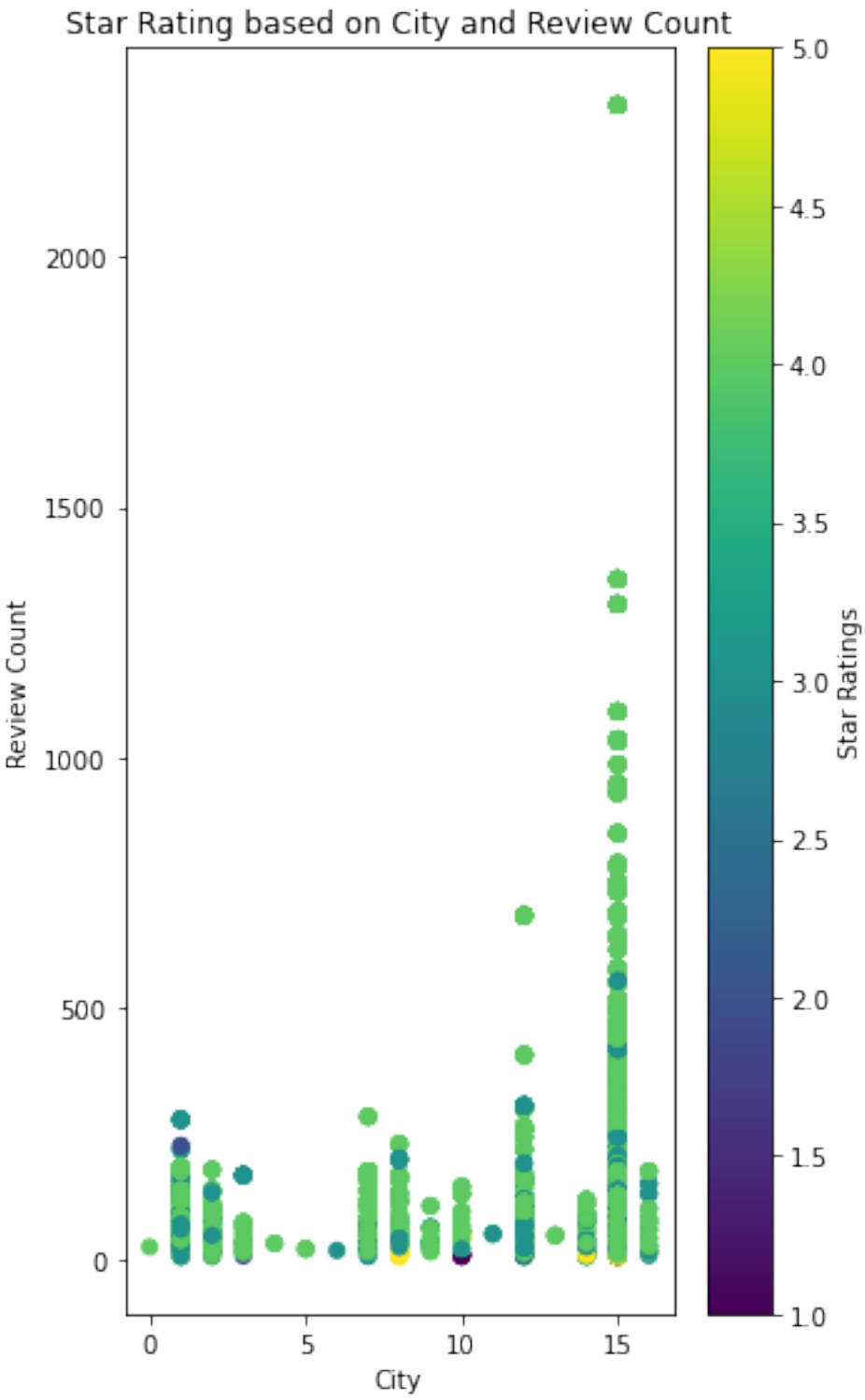
ScatterPlots

December 6, 2021

[1]:
```
#### Scatter Plot Classification
```

[2]:
```
from matplotlib import pyplot as plt
import pandas as pd
from sklearn.preprocessing import LabelEncoder

df = pd.read_csv ('YelpDataset.csv', index_col = 0)

x = df.drop('stars', axis = 1)
y = df['stars']

def convert(data):
    encode = LabelEncoder()
    data['business_id'] = encode.fit_transform(data.business_id)
    data['city'] = encode.fit_transform(data.city)
    data['state'] = encode.fit_transform(data.state)
    data['categories'] = encode.fit_transform(data.categories)
    data['review_id'] = encode.fit_transform(data.review_id)
    data['date'] = encode.fit_transform(data.date)
    data['text'] = encode.fit_transform(data.text)
    data=data.fillna(-999)
    return data

x = convert(x)

plt.figure(figsize=(5,8))
plt.scatter(x['city'], x['review_count'], c = y)
c_bar = plt.colorbar()
c_bar.set_label('Star Ratings')
plt.xlabel("City")
plt.ylabel("Review Count")
plt.title("Star Rating based on City and Review Count")

plt.tight_layout()
plt.show()
```


Star Rating based on City and Review Count
Review Count
City
Star Ratings
0
500
1000
1500
2000
0
5
10
15
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

[ ]: