

[Type text]

Antriksh Kukreti

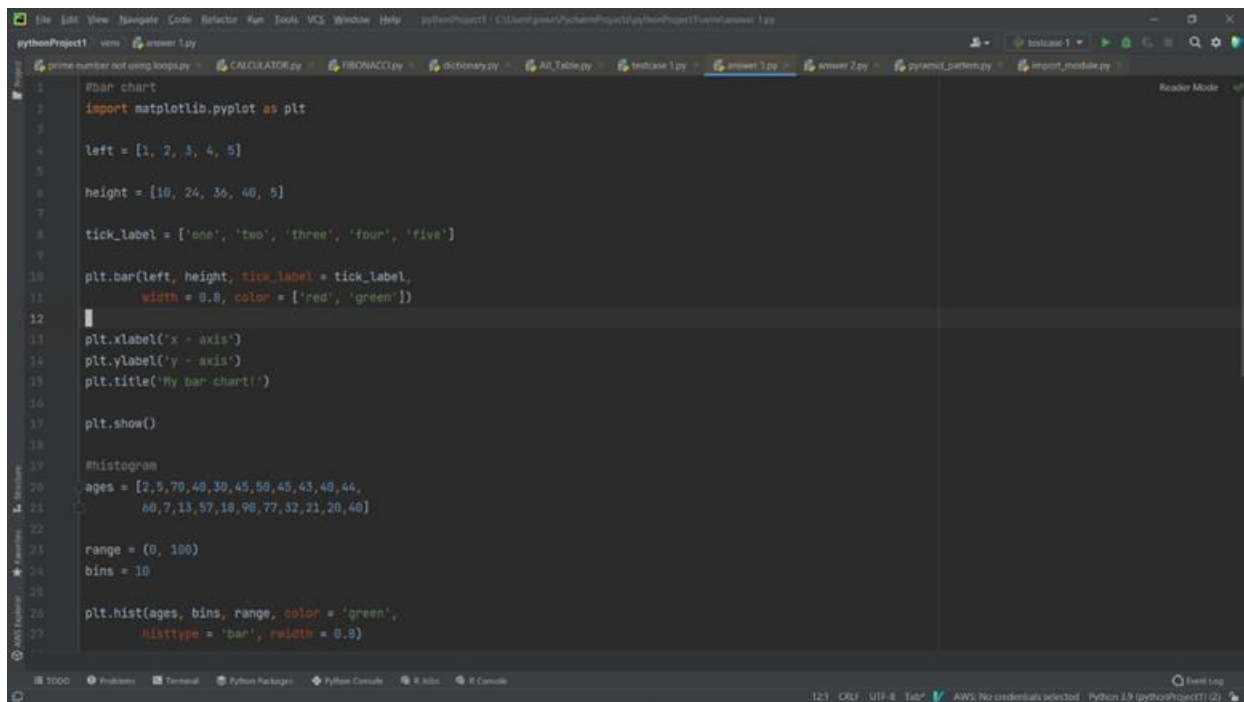
MCA3c

2001034

ML-Practical

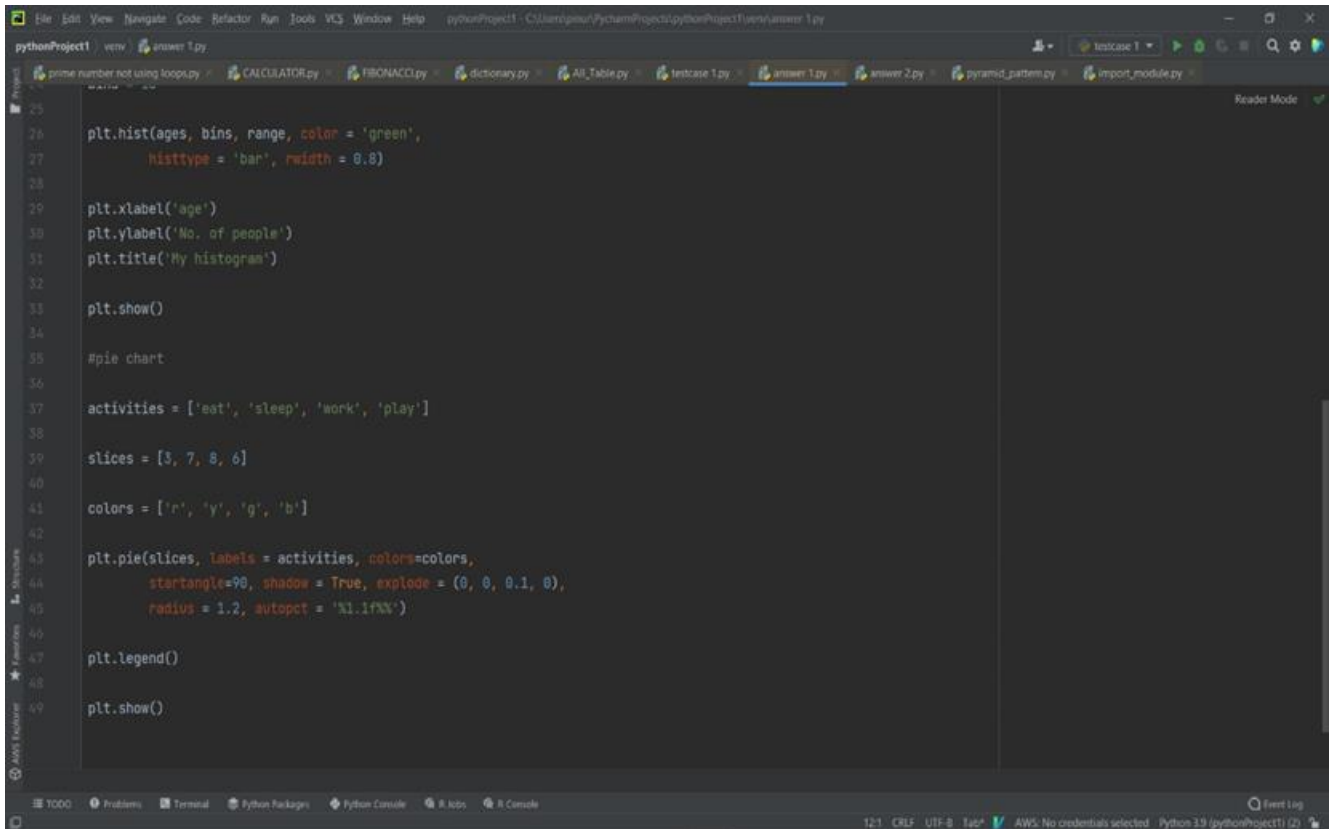
Ques01-A folder "dataset" consists of a file pollution.csv, representing the air pollution data. You have to perform the following tasks on that:

- Data Exploration (with the help of different kinds of Plots)
- Split the dataset in such a way that 80% is kept for training purpose and 20% is kept for test purpose. Build a Decision Tree, SVM and KNN models to classify categories of "AQI Column" and report the model performance on the test dataset with accuracy.



```
1 #bar chart
2 import matplotlib.pyplot as plt
3
4 left = [1, 2, 3, 4, 5]
5
6 height = [10, 24, 36, 48, 5]
7
8 tick_label = ['one', 'two', 'three', 'four', 'five']
9
10 plt.bar(left, height, tick_label = tick_label,
11         width = 0.8, color = ['red', 'green'])
12
13 plt.xlabel('x - axis')
14 plt.ylabel('y - axis')
15 plt.title('My bar chart!')
16
17 plt.show()
18
19 #histogram
20 ages = [2,5,70,40,30,45,50,45,43,40,44,
21        60,7,13,57,18,98,77,32,21,20,40]
22
23 range = (0, 100)
24 bins = 10
25
26 plt.hist(ages, bins, range, color = 'green',
27         histtype = 'bar', width = 0.8)
```

[Type hint]



The screenshot shows a PyCharm IDE window with a Python script. The script uses Matplotlib to create a histogram and a pie chart. The histogram is titled 'My histogram' and shows the distribution of ages. The pie chart shows the distribution of activities: eat, sleep, work, and play.

```
pythonProject1 - view - answer1.py
pythonProject1 - view - prime_number_not_using_loop.py - CALCULATOR.py - FIBONACCI.py - dictionary.py - All_Table.py - testbase1.py - answer1.py - answer2.py - pyramid_pattern.py - import_module.py
25
26 plt.hist(ages, bins, range, color = 'green',
27         histtype = 'bar', width = 0.8)
28
29 plt.xlabel('age')
30 plt.ylabel('No. of people')
31 plt.title('My histogram')
32
33 plt.show()
34
35 #pie chart
36
37 activities = ['eat', 'sleep', 'work', 'play']
38
39 slices = [3, 7, 8, 6]
40
41 colors = ['r', 'y', 'g', 'b']
42
43 plt.pie(slices, labels = activities, colors=colors,
44        startangle=90, shadow = True, explode = (0, 0, 0.1, 0),
45        radius = 1.2, autopct = '%1.1f%%')
46
47 plt.legend()
48
49 plt.show()
```

```
import pandas as pd
from pandas import read_csv
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error

# load dataset
dataframe = pd.read_csv("pollution.csv")
data = dataframe.values

# split into inputs and outputs
X, y = data[:, :-1], data[:, -1]
print(X.shape, y.shape)

# split into train test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=1)
print(X_train.shape, X_test.shape, y_train.shape, y_test.shape)
```

[Type text]

