[Type text]

Antriksh Kukreti

MCA3c

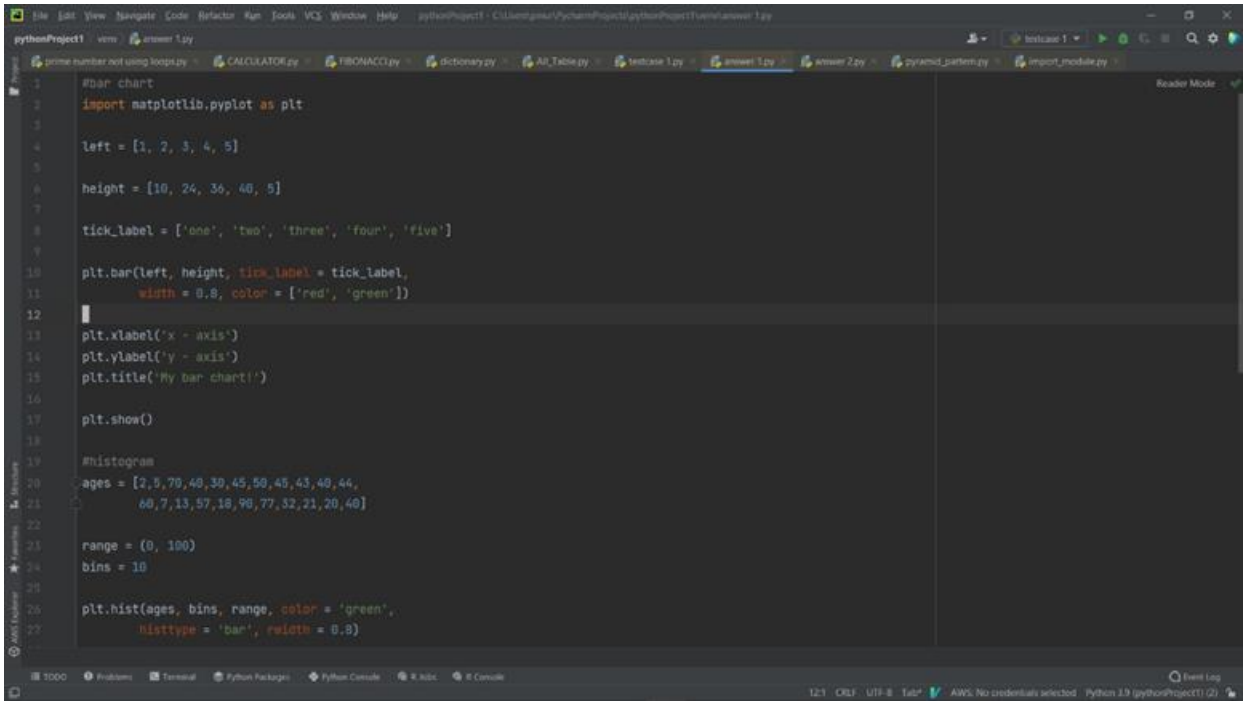
2001034

ML-Practical

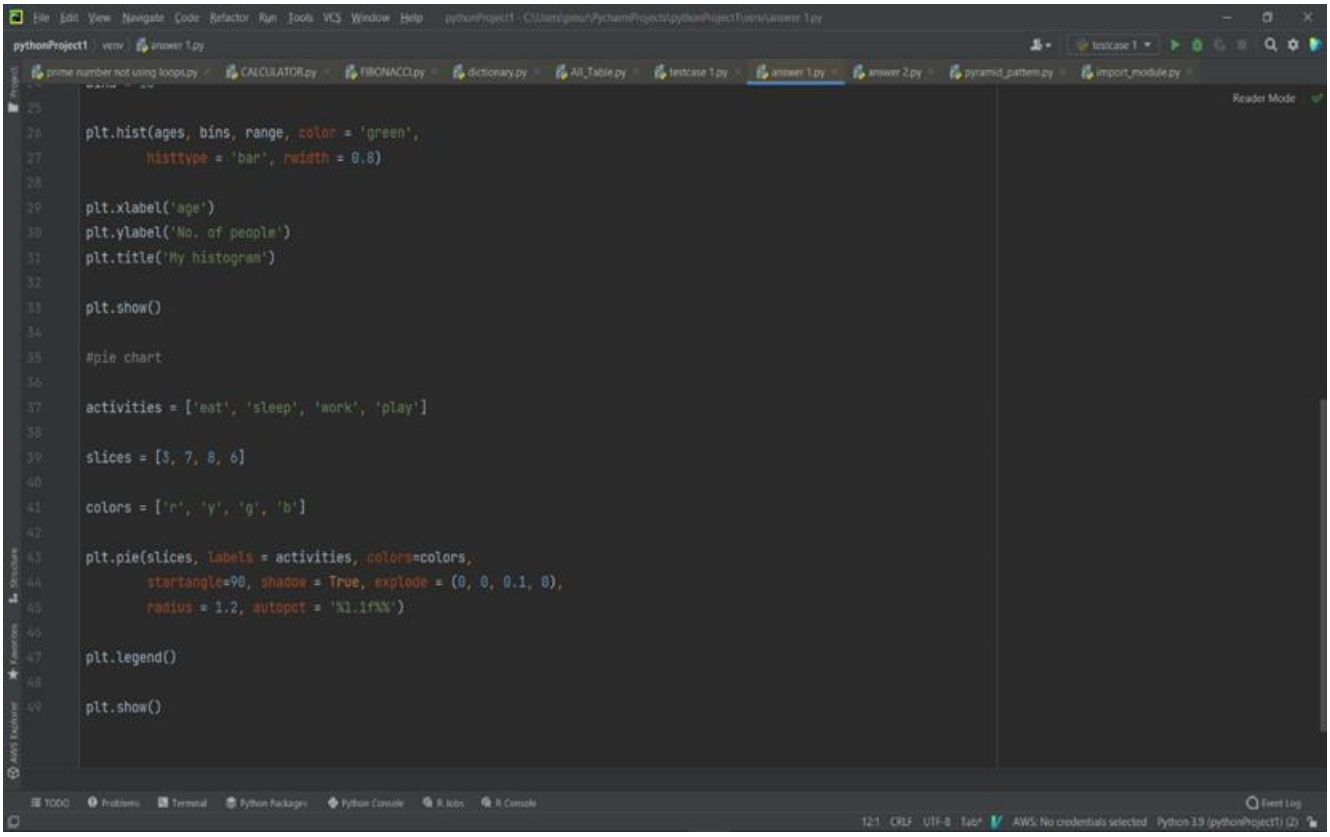
Ques01-A folder "dataset" consists of a file pollution.csv, representing the air pollution data. You have to perform the following tasks on that:

i. Data Exploration (with the help of different kinds of Plots)

ii. 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.

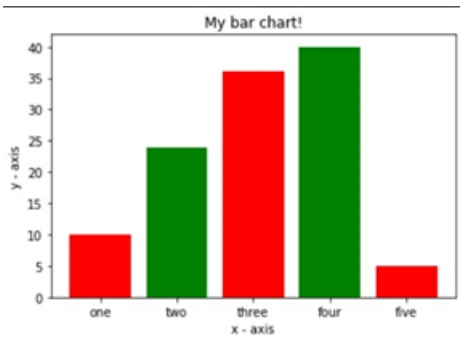


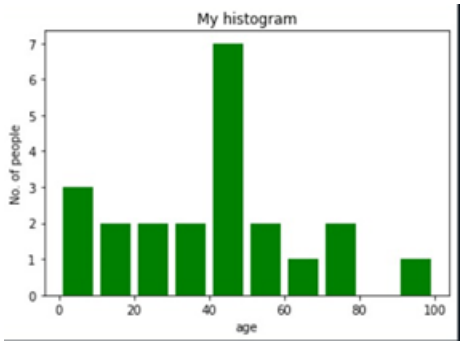
[Type text]

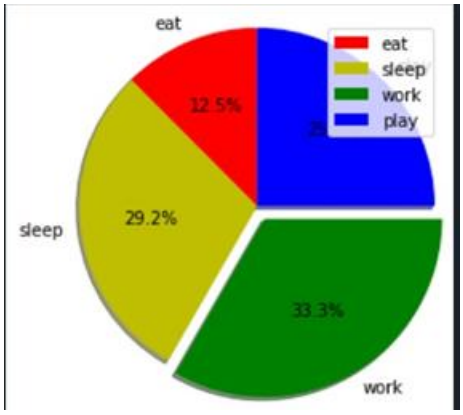


```
plt.hist(ages, bins, range, color = 'green',
        histtype = 'bar', rwidth = 0.8)

plt.xlabel('age')
plt.ylabel('No. of people')
plt.title('My histogram')

plt.show()

#pie chart

activities = ['eat', 'sleep', 'work', 'play']

slices = [3, 7, 8, 6]

colors = ['r', 'y', 'g', 'b']

plt.pie(slices, labels = activities, colors=colors,
        startangle=90, shadow = True, explode = (0, 0, 0.1, 0),
        radius = 1.2, autopct = '%1.1f%%')

plt.legend()

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)
```


My bar chart!
y - axis
x - axis
one
two
three
four
five


My histogram
No. of people
age


eat
sleep
work
play
12.5%
29.2%
33.3%