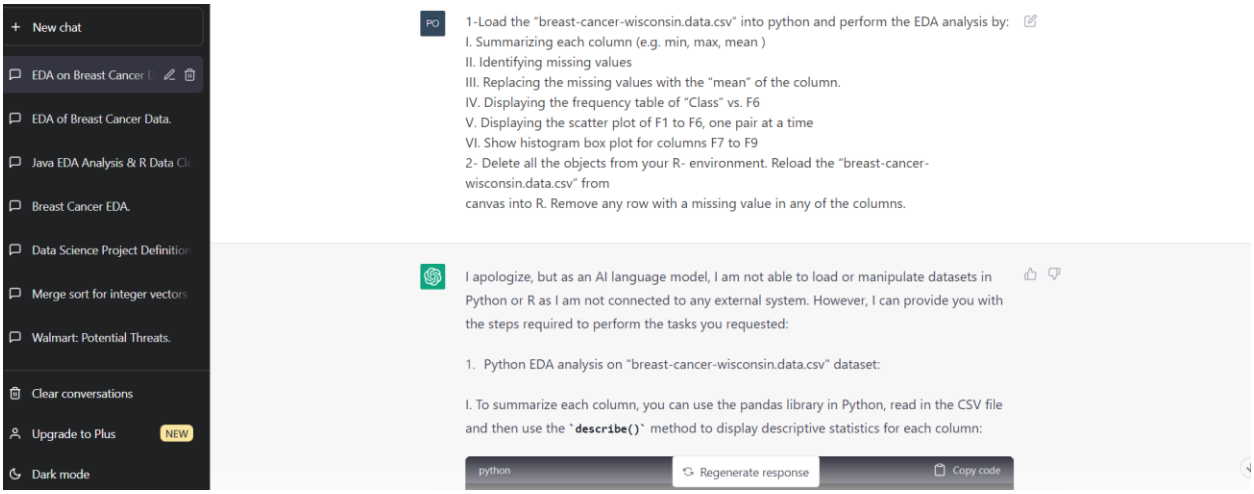
Poorvi_Raut_HW02X_CHATGPT

Poorvi Raut -20009560

Below are screenshots of code generated of HW02_EDA in python language using CHATGPT



I. To summarize each column, you can use the pandas library in Python, read in the CSV file and then use the `describe()` method to display descriptive statistics for each column:

python — Copy code

```
import pandas as pd

# read in the CSV file
data = pd.read_csv('breast-cancer-wisconsin.data.csv')

# summarize each column
print(data.describe())
```

II. To identify missing values, you can use the `isnull()` method to find any null values in the dataset:

python Copy code

```python
import pandas as pd

# read in the CSV file
data = pd.read_csv('breast-cancer-wisconsin.data.csv')

# check for missing values
print(data.isnull().sum())
```

III. To replace the missing values with the mean of the column, you can use the `fillna()` method in combination with the `mean()` method:

python Copy code

```python
import pandas as pd

# read in the CSV file
data = pd.read_csv('breast-cancer-wisconsin.data.csv')

# replace missing values with the mean of the column
data.fillna(data.mean(), inplace=True)
```

IV. To display the frequency table of "Class" vs. F6, you can use the `crosstab()` method from pandas:

python Copy code

```python
import pandas as pd

# read in the CSV file
data = pd.read_csv('breast-cancer-wisconsin.data.csv')

# replace missing values with the mean of the column
data.fillna(data.mean(), inplace=True)

# display frequency table of "Class" vs. F6
print(pd.crosstab(data['Class'], data['F6']))
```

V. To display the scatter plot of F1 to F6, one pair at a time, you can use the `scatter_matrix()` function from pandas:

python Copy code

```python
import pandas as pd
import matplotlib.pyplot as plt

# read in the CSV file
data = pd.read_csv('breast-cancer-wisconsin.data.csv')

# replace missing values with the mean of the column
data.fillna(data.mean(), inplace=True)

# create scatter matrix of F1 to F6
pd.plotting.scatter_matrix(data[['F1', 'F2', 'F3', 'F4', 'F5', 'F6']])
plt.show()
```

VI. To show the histogram box plot for columns F7 to F9, you can use the `**hist()**` and `**boxplot()**` methods from pandas:

python Copy code

```python
import pandas as pd
import matplotlib.pyplot as plt

# read in the CSV file
data = pd.read_csv('breast-cancer-wisconsin.data.csv')

# replace missing values with the mean of the column
data.fillna(data.mean(), inplace=True)

# create histogram of F7 to F9
data[['F7', 'F8', 'F9']].hist()
plt.show()

# create box plot of F7 to F9
data[['F7', 'F8', 'F9']].boxplot()
plt.show()
```

PO

Delete all the objects from your python- environment. Reload the “breast-cancer-wisconsin.data.csv” from
canvas into Python. Remove any row with a missing value in any of the columns.

To delete all objects from your Python environment, you can use the `**del**` command followed by the variable name or object you want to delete. To reload the CSV file and remove any rows with missing values, you can use the following code:

python Copy code

```python
# delete all objects in Python environment
del data

import pandas as pd

# read in the CSV file
data = pd.read_csv('breast-cancer-wisconsin.data.csv')

# remove any rows with missing values
data.dropna(inplace=True)
```