1) Write a python program to find all null values in “Students data” dataset and replace them with zero.

Soln

```
import pandas as pd

df=pd.read_csv("D:/Data science CSV files/student-dataset.csv")

df

df.isnull()

df.dropna()

df.fillna(0)
```

---

4)Write a Python program to extract dependent and independent variables from diabetes dataset.

Soln->

```
from sklearn import datasets

import pandas as pd

diabetes_data = datasets.load_diabetes()

diabetes_data

df = pd.DataFrame(diabetes_data.data, columns=diabetes_data.feature_names)

 df['target'] = diabetes_data.target

 print("Diabetes Dataset:")

print(df.head())

X = df.iloc[:, :-1]

X

y = df['target']

y
```

---

5) Write a python program to create Scatter plot of Sepal length against Petal length.

Soln->

```
import matplotlib.pyplot as plt

import seaborn as sns

from sklearn import datasets

import pandas as pd

iris = datasets.load_iris()

df = pd.DataFrame(data=iris.data, columns=iris.feature_names)
```

```
df.columns = ['sepal_length', 'sepal_width', 'petal_length', 'petal_width']
plt.title('Scatter plot of Sepal Length vs Petal Length')
plt.xlabel('Sepal Length (cm)' )
plt.ylabel('Petal Length (cm)' )
plt.scatter(df['sepal_length'], df['petal_length'], color='blue', edgecolor='k')
```

7) Write a python program to print the number of outliers in following dataset

Soln->

```
import numpy as np
data = [1, 2, 2, 2, 3, 1, 1, 15, 2, 2, 2, 3, 1, 1, 2]
data=np.array(data);
data
Q1 = np.percentile(data, 25)
Q3 = np.percentile(data, 75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
outliers = data[(data < lower_bound) | (data > upper_bound)]
print(f"Number of outliers: {len(outliers)}")
print(f"Outlier values: {outliers}")
```

9) Write a python program to find all null values in Play Tennis dataset and replace them with star(*)

Soln->

```
import pandas as pd
df=pd.read_csv("D:/Data science CSV files/play_tennis.csv")
df
df.isnull()
df.fillna("*")
```

10) Write a python program to find all null values in “Students data” dataset and replace null values mean of the column.

Sol->

```
import pandas as pd
df=pd.read_csv("D:/Data science CSV files/student-dataset.csv")
```

```
df
df.mean
df
df.fillna("")
```

---

11) Write a Python program to build SVM model to iris dataset. The Dataset is available in the scikit-learn library.

Soln->

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
iris = datasets.load_iris()
X = iris.data
y = iris.target
df = pd.DataFrame(data=X, columns=iris.feature_names)
df['species'] = y; print(df.head())
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
svm_model = SVC(kernel='linear')
svm_model.fit(X_train, y_train)
y_pred = svm_model.predict(X_test)
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
accuracy = accuracy_score(y_test, y_pred)
print(f"\nAccuracy: {accuracy * 100:.2f}%")
```

---

13) Write a Python program to create data frame of following information using pandas

Soln->

```
import pandas as pd
data = {
   'Company': ['TATA', 'MG', 'KIA', 'HYUNDAI'],
   'Model': ['Nexon', 'Astor', 'Seltos', 'Creta'],
   'Year': [2017, 2021, 2019, 2015]
}
df = pd.DataFrame(data)
df
```

14) Write a python program to rescale the data between 0 and 1. (use inbuilt dataset)
soln->

```
import pandas as pd
from sklearn.datasets import load_iris
from sklearn.preprocessing import MinMaxScaler
iris = load_iris()
data = pd.DataFrame(iris.data, columns=iris.feature_names)
print(data.head())
scaler = MinMaxScaler()
rescaled_data = scaler.fit_transform(data)
rescaled_df = pd.DataFrame(rescaled_data, columns=data.columns)
print(rescaled_df.head())
```

18) Write a Python program to Standardized the ‘Wholesale customers’ dataset

Soln->

```
import pandas as pd
from sklearn.preprocessing import StandardScaler
data = pd.read_csv("D:/Data science CSV files/Wholesale customers data.csv")
data
features = data.columns
scaler = StandardScaler()
data_standardized = scaler.fit_transform(data[features])
```

```
data_standardized_df = pd.DataFrame(data_standardized, columns=features)

print("\nStandardized Dataset:\n", data_standardized_df.head())
```

---

22) Write a python program to find all null values in “Students data” dataset and remove them.

Soln->

```
import pandas as pd

In [2]:

data = pd.read_csv("D:/Data science CSV files/student-dataset.csv")

data

data.isnull()

Out[3]:

data.fillna(-)

data
```

---

23) Write a Python program to find mean and standard deviation of Annual income column of Mall_Customers dataset

Soln->

```
import pandas as pd

In [11]:

data = pd.read_csv("D:/Data science CSV files/Mall_Customers.csv")

data

mean_income = data['Annual Income (k$)'].mean()

print(f"Mean Annual Income: {mean_income:.2f}")

std_income = data['Annual Income (k$)'].std()

print(f"Standard Deviation of Annual Income: {std_income:.2f}")
```

---

24) Write a Python program to rescale data between (0, 2) the ‘Wholesale customers’ dataset

Soln->

```
import pandas as pd

from sklearn.preprocessing import MinMaxScaler

df=pd.read_csv("D:/Data science CSV files/Wholesale customers data.csv")

df

colums_to_rescale=df.columns

scaler=MinMaxScaler(feature_range=(0,2))
```

```
scaler

df_rescaled=pd.DataFrame(scaler.fit_transform(df[colums_to_rescale]),
columns=colums_to_rescale)

df_rescaled
```

---

26) Write a Python program to extract dependent and independent variables using iloc function of any dataset from final dataset folder

Soln->

```
import pandas as pd

df=pd.read_csv("D:/Data science CSV files/Wholesale customers data.csv")

df.head

x=df.iloc[:,:-1]

x

y=df.iloc[:,:-1]

y

print("Independent variable\n")

x

print("dependent variable\n")

y
```