Assignment 1

Course: Python For Data Science

Submitted by: Saanvi Sinha SRN:PES2UG22EC113

MCQ Questions:

Q1: Which of the following functions can be used to split the data into train and test?

a. pandas.train_test_split( )
b. numpy.train_test_split( )
c. sklearn.model_selection.train_test_split( )
d. sklearn.train_test_split( )

Answer: c. sklearn.model_selection.train_test_split( )

Q2: The Gini coefficient ranges from

a. 0 to 1

b. -1 to 0

c. -1 to 1

d. None of the above

Answer: a. 0 to 1

5 marker questions

Q1. What is type coercion in Python? Illustrate with an example where coercion is possible and one where it fails.

A1. Type coercion means converting one data type into another using syntax datatype(object). It can be done with int(), float(), str(), etc. Possible only if the value is compatible with the target data type.

Example for possible case is as follows:
```
salary_tier = "500"     # converting numeric string to integer
num = int(salary_tier)          # "500" → 500
```

Example for failure case is as follows:
```
salary_tier = "five"
num = int(salary_tier)   # Error: cannot convert non-numeric string
```

Q2. Explain the role of the hue parameter in Seaborn scatter plots with an example.

A2. hue adds a third (categorical or continuous) variable to a 2D plot by mapping that variable to color.In other words, the hue parameter is used to add an additional variable by representing its categories (or values) through different colors. It makes it possible to compare groups or see structure inside the scatter by color-coding points.

Example:
```
import seaborn as sns
sns.scatterplot(x='Age', y='Price', hue='FuelType', data=cars_data)
plt.legend(title='Fuel Type')
plt.show()
```

Here, points are colored based on the fuel type (e.g., Petrol, Diesel, CNG). This allows easy visualization of how different categories behave in relation to Age and Price.

Thus, hue makes the plot more informative by showing group differences or clusters within the same scatter plot.