KNN Assignment, Week-8

The dataset "bank_marketing.csv" is related with direct marketing campaigns (phone calls) of a banking institution. The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).

Important: When you read the csv file, you might need to manipulate the columns somehow for better printing.

For more information about the dataset, please check out the link below:
https://archive.ics.uci.edu/dataset/222/bank+marketing

In [ ]:
```
# (5pts) Step 1: Import necessary libraries and Load the dataset
```

In [ ]:
```
df.head()
```

In [ ]:
```
df.tail()
```

In [ ]:
```
# (10 pts) Step 2: Preprocessing
# Convert categorical variables into numerical; one way to do is Label Encodin
```

In [ ]:
```
# (10 pts) Step 3: Define features (X) and target (y)
```

In [ ]:
```
# (10 pts) Step 4: Scale the features (important for KNN)
```

In [ ]:
```
# (5 pts) Step 5: Split into training and test sets (80&20)
```

In [ ]:
```
# (10 pts) Step 6: Find the optimal K using 5-fold cross-validation
```

In [ ]:
```
# (10 pts) Step 7: Plot the error rate for different K values
```

In [ ]:
```
# (10 pts) Step 8: Find and print the optimal K (K with the minimum error rate
```

In [ ]:
```
# (10 pts) Step 9: Train the final KNN model with the optimal K
```

In [ ]:
```
# (10 pts) Step 10: Make predictions
```

In [ ]:
```
# (10 pts) Step 11: Model evaluation
```