Instructions

1. Follow the instructions in question carefully.
2. A Jupyter notebook along with the output for each cell is expected. Please do not submit zip or rar files.
3. Any assignment submitted using other python IDEs are not considered for grading.

**Problem statement**:

In this assignment, you will be working with a dummy advertising data set, which indicates whether an internet user clicked on an advertisement on a company website. Create a model that will predict whether a user will click on an ad based on the profile of that user.

This data set contains the following features:

\* 'Daily Time Spent on Site': consumer time on site in minutes

\* 'Age': cutomer age in years

\* 'Area Income': Avg. Income of geographical area of consumer

\* 'Daily Internet Usage': Avg. minutes a day consumer is on the internet

\* 'Ad Topic Line': Headline of the advertisement

\* 'City': City of consumer

\* 'Male': Whether or not the consumer was male

\* 'Country': Country of consumer

\* 'Timestamp': Time at which consumer clicked on Ad or closed window

\* 'Clicked on Ad': 0 or 1 indicated clicking on Ad

For the dataset (Advertising dataset), implement the Decision Tree classifier using Python. [9M]

Dataset: [advertising.csv](https://bits-pilani.instructure.com/courses/741/files/139482?wrap=1)[Preview the document](https://bits-pilani.instructure.com/courses/741/files/139482?wrap=1)

**QUESTIONS:**

Q-1: Load the dataset and print the metadata in the notebook. 1M

Q-2: Print a heatmap to check NULL values. 1M

Q-3: Perform stratified splitting of train and test data. 1M

Q-4: Build a classifier model using the Decision Tree algorithm. 2M

Q-5: Print confusion matrix and classification report before and after pruning the Decision tree. (1+1)M

Q-6 Plot the final decision tree model. 1M

Q-7: Find out the stratified cross-validation accuracy 1M.

For any Queries, Contact: [murtuza.dahodwala@wilp.bits-pilani.ac.in](mailto:murtuza.dahodwala@wilp.bits-pilani.ac.in)