**DATA ANALYTICS FOR ARTIFICIAL INTELLIGENCE**

Module Code: H9DAI

Module Level: Level 9

**Programme**

MSCAI\_JAN25

**Lecturer**

Dr. Anu Sahni

# Continuous Assessment

## Datasets

Student must use the last digit in their student ID to download a dataset. If the last digit is an odd number, use ‘used cars.csv’. If the last digit is an even number, use ‘WineQT.csv’.

For example, if the last digit is 1, download ‘used cars.csv’ ; if the last digit is 2, download ‘WineQT.csv’.

## Questions

1. Data characteristics (20 points)

Present the data statistics and characteristics in your dataset:

* + Data population (4 points)
  + Data attributes (4 points)
  + Types of attributes (5 points)
  + Visualise the data using the top group\* such as histogram (7 points)

1. Statistcal Analysis (30 points)
   * Calculate feature correlations in your data (10 points)
   * Use graph from the bottom group\* such as heat map to show the correlations (10 points)
   * Draw a graph from the middle group\* such as scatter plot to show relationship between any two features and draw line model (10 points)

1. Data Manipulation (30 points)

Apply 2 data manipulation techniques on your dataset.

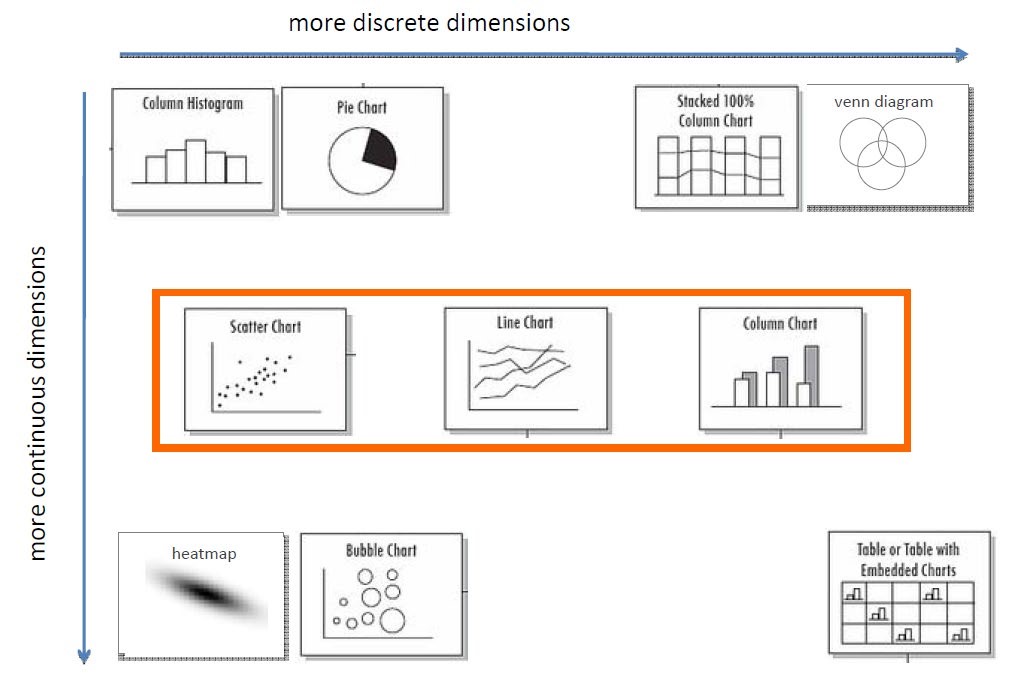
* + Technique dealing with Missing Values (5 points)
  + Show graphically how replacing missing values by central value (mean/median/mode) changes the central value (10 points)
  + Technique dealing Categorical Data where appropriate (10 points)
  + Detecting and Removing Outliers (5 points)

1. Data Preprocessing (10 point)

Apply 2 data preprocessing techniques on your data:

* + Scaling / Normalisation techniques where appropriate (10 points)

\*Data Visualisation: 3 types of graphs from the table below



* + 1 graph from the top group: histogram, pie chart, stacked column chart, venn diagram
  + 1 graph from the middle group: scatter chart, line chart, multidimension column chart
  + 1 graph from the bottom group: heat map, bubble chart

### Dataset-Specific Analysis (10 points)

For each dataset, perform an in-depth analysis of key factors affecting the target variable:

* For the Used Cars dataset, identify the top factors influencing the car’s price. Provide detailed justification using statistical evidence (e.g., correlation analysis, boxplots) for your findings.

* For the Wine Quality dataset, determine the most significant chemical properties influencing wine quality. Provide detailed justification using statistical evidence (e.g., correlation analysis, boxplots) for your findings.