

---

## Assignment Overview

This assignment consists of two case studies testing your understanding of machine learning, EDA, and model evaluation using classification and regression techniques.

## Submission Requirements:

- Jupyter notebook/.ipynb file with complete analysis for both case studies
- 

## CASE STUDY 1: CUSTOMER CHURN PREDICTION (CLASSIFICATION)

### Dataset Overview

**File:** customer\_churn\_data.csv (10,000 customers, 24 features)

The dataset contains customer demographics (age, gender, senior citizen status, partner, dependents), service information (tenure, phone service, internet type, additional services), account details (contract type, billing method, charges, satisfaction scores), and support interaction data. The target variable is binary churn status (Yes/No).

### TASKS FOR CLASSIFICATION

#### Task 1: Data Loading and Exploration

- Load dataset and check basic information, data types, shape
- Handle missing values and display summary statistics

#### Task 2: Exploratory Data Analysis

- Analyse churn distribution and calculate churn rate
- Examine demographic patterns and service usage impact on churn
- Investigate financial factors and payment method effects
- Create correlation analysis for numerical features

---

## Task 3: Data Preprocessing

- Encode categorical variables and scale numerical features
- Perform feature engineering if beneficial
- Split data

## Task 4: Model Development

- Implement 4 classification algorithms: Logistic Regression, Decision Tree, Random Forest, SVM
- Train models and tune hyperparameters using validation set

## Task 5: Model Evaluation

- Evaluate using Accuracy, Precision, Recall, F1-Score
  - Create confusion matrices and compare model performances
  - Select best model and test on test set
-

## CASE STUDY 2: REAL ESTATE PRICE PREDICTION (REGRESSION)

### Dataset Overview

**File:** house\_price\_data.csv (8,000 properties, 25 features)

The dataset includes location features (neighbourhood, distances to city centre/metro/schools, crime rates), property characteristics (age, area, rooms, floors, amenities), quality indicators (construction quality, renovation status, energy efficiency), and market information (listing duration, sale season/year). The target variable is sale price in thousands of dollars.

### TASKS FOR REGRESSION

#### Task 1: Data Loading and Exploration

- Load dataset and check for missing values, outliers
- Display summary statistics and analyse data quality

#### Task 2: Exploratory Data Analysis

- Analyse price distribution and check for skewness
- Examine location impact and property feature relationships
- Investigate quality factors and market timing effects
- Create correlation analysis and identify multicollinearity

#### Task 3: Feature Engineering and Preprocessing

- Handle categorical variables and create new features
- Address outliers and multicollinearity issues
- Split data

---

## Task 4: Model Development

- Implement Linear Regression Model

## Task 5: Model Evaluation

- Evaluate using MAE, MSE, RMSE, R-squared
- Create residual plots and compare model performances
- Select best model and validate on test set

---

## DELIVERABLES

### Required Submissions:

**Jupyter Notebook/ .ipynb file:** Complete analysis with visualizations

### Submission Guidelines:

- Submit ZIP file: **SRN\_StudentName\_ML\_Lab1.zip**
  - The ZIP file must include two files named **SRN\_Case1.ipynb** and **SRN\_Case2.ipynb**.
-