**GTSC2143 Machine Learning for Business**

**Tutorial 5**

**Please write down your answers in this document and submit it at iSpace by the end of this tutorial.**

## Activity 1. Data Loading and Preprocessing

1. Load the Dataset
   a) Load the house prices dataset and filter out house id '1925069082'

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression, Lasso, LogisticRegression
from sklearn.metrics import mean_squared_error, r2_score, accuracy_score, classification_report,
confusion_matrix

# Load and filter data
data = pd.read_csv("GTSC2143-Lecture 4 predicting-house-prices-assignment_home_data.csv")
filtered_data = data.query("id!= 1925069082")
```

   b) Split into train (80%) and test (20%) sets using `random_state=42`

## Activity 2. Predicting House Price - Model Comparison

1. Feature Selection and Model Training
   a) Select as many as possible meaningful variables for predicting house prices from all the variables
   b) Clearly specify which features you include and provide rationale for excluding certain variables
   c) Analysis: Write 2-3 sentences explaining your feature selection decisions.

2. Train and Compare Models
   a) Train a Linear Regression model using your selected features
   b) Train a Lasso Regression model using the same features (use alpha=1.0)
   c) For both models, calculate: MSE, RMSE, $R^2$ Score
   d) Display coefficients for both models
3. Model Comparison Analysis
   a) Create a comparison table showing performance metrics for both models
   b) Count non-zero coefficients in each model
   c) Analysis: Write 2-3 sentences comparing model quality and explaining which performs better and why.

## Activity 3. Predicting 'quick_sold' - Logistic Regression

1. Logistic Regression Model
   a) Train a logistic regression model using features: 'price', 'bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'floors'

b) Calculate and display:
    - Accuracy score
    - Classification report
    - Model coefficients
c) Analysis: Write 2-3 sentences interpreting what the coefficients tell us about factors affecting quick sales.

### Activity 4. Prediction for Excluded House

1. Predict for House ID '1925069082'
    a) Use your best price prediction model to predict its price
    b) Use your logistic regression model to predict its probability of quick sale
    c) Display:
        - Predicted price vs actual price
        - Predicted probability of quick sale
        - Final quick_sold classification
    d) Analysis: Write 2-3 sentences evaluating both predictions and their business implications.

- End of Tutorial 5 -