Student Group Task:

Predicting Car Insurance Claims Objective Work in groups to analyze the dataset and build models to predict whether a policyholder will file a claim in the next 6 months.

Task Overview

1. Data Cleaning
    - Handle missing values.
    - Encode categorical features.
    - Normalize or scale variables if needed.
2. Exploratory Data Analysis (EDA)
    - Plot distributions of variables.
    - Explore relationships with 'is_claim'.
3. Model Building
    - Use Correlation to carry out a Principal Component Analysis (PCA) to determine relevant variables for the objective.
    - Build and compare at least two models:
        i. Logistic Regression
        ii. Decision Tree Classifier
        iii. (Optional) Try Random Forest
4. Model Evaluation
    - Evaluate using: Accuracy, Precision, Recall, F1-Score
    - Discuss performance and select the best model.
5. Reporting
    - Introduction
    - Data Cleaning Summary
    - Modeling Approach
    - Model Results & Comparison
    - Limitations & Recommendations
6. Executive Summary- A two-page summary with key insights and recommendations.

A draft R script has been included in your project pack to aid your assessment. You can choose to either build and modify the template or develop one from scratch.