Data Analysis & Data Science with Python - Task 5 Report

Task 1: Student Pass/Fail Prediction

Objective:

Predict whether a student passes (1) or fails (0) based on Study Hours and Attendance.

Model: Logistic Regression

Evaluation:

- Accuracy: ~85%

- Confusion Matrix: Shows good separation between pass and fail.

Insights:

- Both study hours and attendance are strong predictors.

- Logistic regression performed well with only two features.

Tools Used:

- pandas, seaborn, matplotlib, scikit-learn

Task 2: Sentiment Analysis with NLP

Objective:

Classify customer reviews as positive or negative using logistic regression.

Model: Logistic Regression with TF-IDF Vectorization

Text Preprocessing:

- Lowercasing

- Removing punctuation/special characters

- TF-IDF vectorization (scikit-learn)

Evaluation:

- Precision, Recall, and F1-score all strong

Data Analysis & Data Science with Python - Task 5 Report

- Model correctly classified most simple examples

Insights:

- Positive reviews use words like: "love", "excellent", "amazing"

- Negative reviews include: "hate", "worst", "awful"

Tools Used:

- pandas, re, scikit-learn