



Activity 7			
Topic:	Topic 6: Regression	Week No.	9
Course Code:	CSEL302	Term:	2 <sup>nd</sup> Semester
Course Title:	Introduction to Intelligent Systems	Academic Year:	2023-2024
Student Name		Section	
Due date	March 27, 2024 12:00 PM	Points	

**Assessment Task:** Netflix Userbase Analysis

**Objective:**

Leverage linear regression to predict Monthly Revenue and logistic regression to classify customers based on a positive or negative feedback proxy, using the Netflix Userbase dataset. This task will encompass data preprocessing, exploratory data analysis (EDA), model building, evaluation, and visualization.

**Dataset:**

"Netflix Userbase.csv", containing user demographics, subscription details, and other relevant information.

**Requirements:**

**Part 1: Data Preprocessing**

- Load the Dataset: Import the dataset using Pandas and display its structure.
- Missing Values: Identify and manage any missing values appropriately.
- Encode Categorical Variables: Transform categorical variables into numerical representations suitable for regression analysis.
- Feature Selection: Decide which features to include in the regression models. Explain your choice.

**Part 2: Exploratory Data Analysis (EDA)**

- Descriptive Statistics: Provide a summary of the data using the `.describe()` method.
- Visualizations: Create visualizations (e.g., distribution of Monthly Revenue, user demographics) to understand the data better.

**Part 3: Linear Regression Model (Predicting Monthly Revenue)**

- **Build the Model:** Develop a linear regression model to predict the Monthly Revenue based on selected features.
- **Model Evaluation:** Evaluate the model using R-squared, RMSE, or other relevant metrics.



#### **Part 4: Logistic Regression Model (Predicting Customer Feedback)**

- **Model Building:** Assuming a binary feedback variable (positive/negative), construct a logistic regression model to predict feedback based on user demographics and subscription details.
- **Evaluation:** Use accuracy, precision, recall, F1-score, and the confusion matrix for model assessment.

#### **Part 5: Comparative Analysis and Visualization**

- **Feature Importance:** Examine the significance of different features in both regression models.
- **Insights:** Draw insights from the models' predictions, focusing on how various factors might influence Monthly Revenue and feedback.

#### **Evaluation Criteria:**

- **Completeness and Correctness:** All parts of the task must be completed accurately, following data science best practices.
- **Code Quality and Comments:** The code should be well-organized, commented, and easy to follow.
- **Insightfulness:** Demonstrated ability to extract meaningful insights from the analysis and model results.
- **Presentation:** Effectively communicate findings through visualizations and summaries.

This task is designed to test practical knowledge of linear and logistic regression, as well as data manipulation, visualization, and analytical skills in a real-world-like scenario. It offers a hands-on experience with a dataset that includes user demographics, subscription details, and revenue information, providing a comprehensive overview of data science applications in customer behavior analysis.

#### **Submission Instruction:**

- Share the Google Collab Activity to [markbernardino@lspu.edu.ph](mailto:markbernardino@lspu.edu.ph)
- Filename Format: **2A-BERNARDINO-EXER7**

Inability to follow this instruction will be deducted 5 points each for filename format and late submission per day. Also, cheating and plagiarism will be penalized.