Republic of the Philippines
Laguna State Polytechnic University
Province of Laguna

Activity 6			
Topic:	**Topic 6: Logistic Regression**	**Week No.**	8
Course Code:	**CSST104**	**Term:**	2nd Semester
Course Title:	**Machine Learning**	**Academic Year:**	2023-2024
Student Name		**Section**	
Due date	**March 27, 2024 12:00 PM**	**Points**	

Assessment Task: Predicting Feedback on Online Food Orders Using Logistic Regression

Objective: Develop a logistic regression model to predict the customer feedback (positive/negative) on online food orders. This task involves data loading, preprocessing, exploratory data analysis (EDA), model building, evaluation, and visualization.

Dataset:
You will use a dataset that contains information on online food orders, including customer demographics, order details, and feedback. The dataset includes features like Age, Gender, Marital Status, Occupation, Monthly Income, and Feedback.

Requirements

Part 1: Data Loading and Preprocessing

1. **Load the Dataset:** Import the dataset using Pandas and display the first few rows to understand its structure.
2. **Handle Missing Values:** Identify and handle any missing values in the dataset. Choose an appropriate strategy (e.g., imputation or removal) based on the context.
3. **Encode Categorical Variables:** Convert categorical variables into a numeric format suitable for logistic regression. Consider techniques like one-hot encoding or label encoding.
4. **Feature Selection:** Identify which features to include in the model. Justify your selections based on the dataset's context and preliminary analysis.

Part 2: Exploratory Data Analysis (EDA)

1. **Descriptive Statistics:** Use .describe() to summarize the numeric columns. Highlight any interesting findings.
2. **Visualizations:** Create visualizations to understand the relationships between features and the target variable. Suggestions include:
    - Distribution of Age and its impact on Feedback.
    - Proportions of Feedback across different levels of Monthly Income.
    - Correlation matrix heatmap to identify any interesting correlations between features.

Part 3: Logistic Regression Model

1. **Build the Model:** Implement a logistic regression model using scikit-learn. Split your data into training and test sets to evaluate the model's performance.
2. **Model Evaluation:** Assess your model's performance using appropriate metrics, such as accuracy, precision, recall, and the confusion matrix. Discuss the results.

Part 4: Data Analysis and Visualization

1. **Feature Importance:** Analyze and visualize the importance of different features in your logistic regression model. Discuss how each feature influences the prediction of Feedback.
2. **Prediction Insights:** Visualize the distribution of predicted probabilities. Discuss any patterns or insights you can derive from how the model makes predictions.

Evaluation Criteria

- **Completeness and Correctness:** All parts of the task are completed correctly, following data science best practices.
- **Code Quality and Comments:** The code is well-organized and commented, making it easy to follow your analysis process.
- **Insightfulness:** Demonstrates the ability to extract meaningful insights from the analysis and model results.
- **Presentation:** Effectiveness in communicating findings through visualizations and summary.

This task is designed to assess skills in handling a real-world data science problem, from initial data preprocessing to deriving insights from a logistic regression model. It tests practical knowledge of logistic regression in Python, along with data manipulation, visualization, and analytical skills.

Submission Instruction:

- Share the Google Collab Activity to markbernardino@lspu.edu.ph
- Filename Format: **2A-BERNARDINO-EXER6**

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.