**DevelopersHub Corporation**

# Data Science & Analytics Internship Tasks

**Due Date:** 2nd August, 2025

# Overview

As part of your Data Science & Analytics Internship at DevelopersHub Corporation, you are required to complete **at least 3 out of the 5 tasks** listed below. You are encouraged to complete all of them if you are interested in gaining more hands-on experience.

These tasks are designed to help you develop core skills in data science, such as data exploration, visualization, model building, and performance evaluation using Python. You will use libraries like pandas, matplotlib, seaborn, and scikit-learn.

# Task 1: Exploring and Visualizing a Simple Dataset

**Objective:**

Understand how to read, summarize, and visualize a dataset.

**Dataset:**

Iris Dataset (CSV format, available through seaborn or other open sources) **Instructions:**

* Load the dataset using the pandas library.

* Display dataset structure using .shape, .columns, and .head().

* Create basic visualizations:

○ Scatter plot to analyze relationships between variables.

○ Histogram to examine data distribution.

○ Box plot to detect outliers and spread of values.

* Use matplotlib and seaborn for visualizations.

**Skills:**

* Data loading and inspection using pandas

* Basic data summarization

* Visualization using matplotlib and seaborn

# Task 2: Credit Risk Prediction

**Objective:**

Predict whether a loan applicant is likely to default on a loan.

**Dataset:**

Loan Prediction Dataset (available on Kaggle)

**Instructions:**

* Handle missing data appropriately.

* Visualize key features such as loan amount, education, and income.

* Train a classification model like Logistic Regression or Decision Tree.

* Evaluate the model using accuracy and a confusion matrix.

**Skills:**

* Data cleaning and handling missing values

* Exploratory Data Analysis (EDA)

* Binary classification using machine learning

* Model evaluation using confusion matrix and accuracy

# Task 3: Customer Churn Prediction (Bank Customers)

**Objective:**

Identify customers who are likely to leave the bank.

**Dataset:**

Churn Modelling Dataset

**Instructions:**

* Clean and prepare the dataset.

* Encode categorical features such as geography and gender.

* Train a classification model.

* Analyze feature importance to understand what influences churn.

**Skills:**

* Categorical data encoding (Label Encoding / One-Hot Encoding)

* Supervised classification modeling

* Understanding and interpreting feature importance

# Task 4: Predicting Insurance Claim Amounts

**Objective:**

Estimate the medical insurance claim amount based on personal data.

**Dataset:**

Medical Cost Personal Dataset

**Instructions:**

* Train a Linear Regression model to predict charges.

* Visualize how BMI, age, and smoking status impact insurance charges.

* Evaluate model performance using MAE and RMSE.

**Skills:**

* Regression modeling

* Feature correlation and visualization

* Error evaluation using MAE and RMSE

# Task 5: Personal Loan Acceptance Prediction

**Objective:**

Predict which customers are likely to accept a personal loan offer.

**Dataset:**

Bank Marketing Dataset (UCI Machine Learning Repository)

**Instructions:**

* Perform basic data exploration on features such as age, job, and marital status.

* Train a Logistic Regression or Decision Tree classifier.

* Analyze the results to identify which customer groups are more likely to accept the offer.

**Skills:**

* Data exploration and visualization

* Classification modeling

* Business insight extraction from data

# Submission Requirements (Checklist for Each Task)

To receive credit for each completed task, ensure the following items are included:

## 1. Jupyter Notebook

* Introduction and problem statement

* Dataset understanding and description

* Data cleaning and preparation

* Exploratory Data Analysis (EDA) with graphs

* Model training and testing

* Evaluation metrics (e.g., accuracy, confusion matrix, MAE, RMSE, etc.)

* Conclusion summarizing key insights

**2. Code Quality**

● Code should be clean, well-structured, and include comments explaining each step.

## 3. GitHub Repository

* Create a dedicated GitHub repository for your internship tasks.

* Give the repository a clear and descriptive name.

* Add a README.md file summarizing:

○ The task objective

○ Your approach

○ Results and insights

## 4. Submission on Google Classroom

● Submit the link to your GitHub repository for each completed task.

# Important Note

* **Complete at least 3 out of 5 tasks** before the due

date: **2nd August 2025**.

* You may complete all 5 tasks for more practice and

stronger portfolio.

* Ask for help whenever needed. Mentors are here

to guide you through.