University Assignment: Data Analytics Using Python

Assignment Brief

Course: Data Analytics

Submission Deadline: [Insert Date]

Level: Undergraduate / Postgraduate

Instructor: [Insert Instructor Name]

Objective:

To apply data analytics techniques using Python to explore, clean, analyze, and visualize a dataset, and derive insights using libraries like Pandas, NumPy, Matplotlib, and Scikit-learn.

Instructions:

1. Use Python (Jupyter Notebook preferred).

2. Choose a public dataset (e.g., from Kaggle, UCI ML Repository, or data.gov).

3. Submit:

   - Jupyter Notebook (.ipynb)

   - PDF export of notebook

   - Brief report (2-3 pages) summarizing your findings

Tasks:

Task 1: Data Loading & Cleaning (20 Marks)

- Load the dataset using pandas

- Handle missing/null values

University Assignment: Data Analytics Using Python

- Check and correct data types

- Remove duplicates (if any)

Task 2: Exploratory Data Analysis (EDA) (30 Marks)

- Use summary statistics (.describe())

- Plot distributions using matplotlib/seaborn

- Use box plots, scatter plots, and correlation heatmaps

Task 3: Feature Engineering (15 Marks)

- Create at least 2 new features

- Normalize/scale variables where needed

- Encode categorical variables (if any)

Task 4: Basic Modeling (20 Marks)

- Split data into training/testing sets

- Apply regression or classification (depending on dataset)

- Use scikit-learn for modeling

- Evaluate with appropriate metrics (accuracy, RMSE, confusion matrix, etc.)

Task 5: Insights and Reporting (15 Marks)

- Highlight at least 3 key insights from your analysis

- Visualize your findings

- Discuss limitations or assumptions

University Assignment: Data Analytics Using Python

Bonus (Optional, +5 Marks):

- Deploy your notebook using Streamlit or Flask

- Use GitHub for version control and submission