Statistics Course Project Guidelines

Project Title: Data Visualization and Statistics Using Python

---

General Instructions

1. **Team Formation:**
   - Each team must consist of exactly **two members**.
   - Provide the names and student IDs of both members in the submission document.
2. **Project Overview:**
   - The project is worth **15 marks**.
   - Use the dataset provided to create visualizations and calculate data statistics using Python.
   - All analyses must be completed through original Python code.
3. **Plagiarism Policy:**
   - Any use of AI-generated code (e.g., ChatGPT or similar tools) will result in a **grade of zero**.
   - Any form of plagiarism (copying from external sources) will result in a **grade of zero**.
4. **Submission Requirements:**
   - Submit a ZIP or RAR file containing:
     - The Python code (jupyter notebook) used for the project.
     - A well-prepared **Word document** (details provided below).
5. **Discussion Slot:**
   - Each team will be assigned a slot for discussion after the submission deadline.
   - Be prepared to explain your project, methodology, and findings.

Submission Document Structure

The document must include the following sections:

1. **Title Page:**
    - Project title.
    - Team details (Student 1 Name & ID, Student 2 Name & ID).
2. **Visualizations:**
    - Include each visualization created using Python.
        - Describe the type of visualization used (e.g., bar chart, scatter plot).
        - Explain the insights and information provided by the visualization in the context of the dataset.
3. **Data Statistics:**
    - Highlight the data statistics calculated (e.g., mean, median, standard deviation).
    - Provide insights gained from these statistics and their significance.
4. **Conclusion (Optional):**
    - Summarize the key findings of your project.

Project Guidelines

- Use only Python libraries for visualization and data statistics (example and suggestions → Pandas, Matplotlib, Seaborn, NumPy). Feel free to use any extra library.
- The Python code must be well-structured and thoroughly tested.
- The document must clearly explain the analysis without including the Python code itself.

Grading Criteria

Criteria	Marks
Correlation Matrix (Plot)	1 marks
Scatter Plot (between BloodPressure and BMI)	2 marks
Distribution Plot For the following features (each feature has its own plot alone): - Pregnancies - SkinThickness - Age	2 marks
Pie Charts For the following features (each feature has its own plot alone): - Outcome	1 marks
Boxplots For the following features (each feature has its own plot alone): - SkinThickness - DiabetesPedigreeFunction - BloodPressure	2 marks
Central Tendency: (for all features) - Mean - Median - Mode	1
Dispersion: (for all features) - Standard - Deviation - Variance	1
Minimum and Maximum Values (for all features)	1
Document and teamwork	3
Total	15

Important Notes

- Late submissions will not be accepted.
- Ensure your analysis and documentation are original and comply with the instructions provided.
- You must be prepared to defend your findings and code during the discussion slot.