Introduction to Data Science Project Instructions

You are required to complete the following project as part of the **Introduction to Data Science** course. Please read the instructions carefully and make sure to incorporate the tools and skills learned during the course.

Objective

1. Select a **unique dataset** (no two students can work on the same dataset).
2. Perform **Exploratory Data Analysis (EDA)** with at least **10 to 15 different analyses**.
3. **Preprocess** the dataset to prepare it for machine learning.
4. Apply any **machine learning model** of your choice to the preprocessed dataset.
5. Build an **interactive web application** using **Streamlit** to present your work.

Step-by-Step Guide

1. Dataset Selection

- Choose an interesting dataset with **enough features and records** to perform meaningful analysis.
- Add your dataset name and link in the provided Excel sheet.
- Verify that your dataset is unique (first-come, first-served basis).

---

2. Exploratory Data Analysis (EDA)

Perform the following analyses:

- Summary statistics (mean, median, mode, etc.)
- Visualizations (e.g., histograms, scatter plots, box plots)
- Correlation analysis
- Missing value analysis
- Outlier detection
- Feature distribution analysis
- Data types and unique value counts
- Trend analysis (if applicable)
- Grouped aggregations (e.g., by category or time)
- Pairwise feature relationships
- Any other relevant analysis based on your dataset

---

3. Data Preprocessing

Prepare the data for modeling:

- Handle missing values (e.g., imputation or removal)
- Encode categorical variables
- Normalize or scale numerical features
- Split into training and testing sets
- Apply any additional necessary transformations

---

4. Machine Learning Model

- Select an appropriate model (supervised or unsupervised) based on your dataset (e.g., regression, classification, clustering).
- Train the model and evaluate it using appropriate metrics (accuracy, RMSE, F1-score, etc.).
- ⚠️ **Important:** Your model **must be able to make predictions at runtime** within the Streamlit application based on user input. If this feature is missing, **it may result in loss of marks**.

---

5. Streamlit Application

Use **Streamlit** to present your project interactively. Include the following sections:

- **Introduction**: Brief overview of your dataset and project goals
- **EDA Section**: Visualizations, tables, and insights
- **Model Section**: Description of the model used, runtime predictions, and performance results
- **Conclusion**: Key takeaways from your project

Make sure your app is **interactive, informative, and user-friendly**.

Additional Notes

- Use skills learned during the course such as:
  - **Data preprocessing**
  - **EDA and visualization**
  - **Model training and evaluation**
  - **Building Streamlit apps**
- The **final deadline will be announced soon**, and it **will be before the summer vacation**.
- This is an **individual project**.

- 🚨 **Start working on your project from today. No excuses about having less time will be entertained later.**

If you have any questions at any stage, feel free to reach out. Good luck!
