**PyNEXT Project**

**Project instructions**

1. This project carries **50% weight** in the PyNEXT course evaluation.
2. You must submit the project as a **.ipynb file (Jupyter Notebook)**.
3. You have to submit this project **individually.**
4. A **demo of this project** will be provided for better understanding.
5. You can use **any dataset** of your choice.
6. You may use **multiple datasets**, but they must be similar. Example: if you are working with House prices then you can use multiple house price dataset
7. You shall submit the project in Google Classroom

**Demo project**: [House price regression](https://www.kaggle.com/code/miftahuladib/fork-of-housing-price-regression-top-8)

**Recommended Datasets:**

[House price dataset](https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data)

[Titanic Dataset](https://www.kaggle.com/competitions/titanic/data)

**Project deadline: 5th October,2025**

**Project overview:**

### **1) Import a Dataset**

### You can choose any dataset you like (e.g., Titanic, Iris, Housing Prices, Health Dataset). If unsure, seek suggestions from instructors or use one of the recommended datasets.

### **2) Import All the Necessary Libraries**

### **3) Perform Exploratory Data Analysis (EDA)**

### Understand the dataset using descriptive statistics .info(), .describe(), .shape(), .head(), etc.

### Create visualizations (histograms, scatter plots, correlation heatmaps, bar plots, etc.).

### After each plot/graph/curve, write an explanation in a Markdown cell describing what you observed.

### **4) Check for Outliers**

### Use boxplots to check if there are any outliers in the columns.

### If any outliers exist, remove them.

### **5) Handle Missing Values**

### Check for null values in the dataset and display them.

### If any null values exist, fill them with appropriate values such as 0, mean, or median.

### **6) Feature Engineering (For extra marks )**

### Create new features from existing variables (e.g., BMI = weight / height²).

### Drop unnecessary columns that do not add value (e.g., id, redundant text fields, etc.).

### 

### **7) Key Insights & Findings**

### At the end of your notebook, write a structured summary of your main findings in Markdown.

### Example Structure:

### Data Overview (size, features, types)

### Outlier Analysis (what was removed)

### Missing Value Treatment

### Important EDA Findings (patterns, distributions, correlations)

### Feature Engineering Steps

### Final Cleaned Dataset Ready for Modeling

Also write your personal opinion on the overall project experience

### **8) Any kind of creative preprocessing or data science operation, not mentioned in the instructions, will be appreciated.**

### **Marks Distribution**

### Dataset Selection: **8 Marks**

### Importing Libraries: **2 Marks**

### Exploratory Data Analysis (EDA): **40Marks**

### Outlier Detection & Treatment: **10 Marks**

### Missing Value Handling: **10 Marks**

### Key Insights & Findings: **20 Marks**

### Overall Documentation: **10 Marks**

### **Total: 100 Marks**

### 