**Multiple Linear Regression Project Report**

**1. Project Title:**

Predicting Insurance Premium Using Multiple Linear Regression

**2. Introduction**

This project applies **Multiple Linear Regression**, a supervised learning algorithm, to predict insurance premiums based on multiple factors: **age, height, and weight**. The model learns the relationship between these features and the premium amount to make accurate predictions.

**3. Dataset Description**

The dataset (Insurance.csv) consists of four columns:

* **Age (Independent Variable - X1):** The age of the insured person.
* **Height (Independent Variable - X2):** The height of the person (in cm).
* **Weight (Independent Variable - X3):** The weight of the person (in kg).
* **Premium (Dependent Variable - Y):** The insurance premium to be paid.

**4. Methodology**

**Step 1: Data Loading and Exploration**

* Used pandas to read and display the dataset.
* Checked for missing values in the dataset.

**Step 2: Data Preprocessing**

* Found missing values in the height column.
* Replaced missing values with the **mean height** using fillna().

**Step 3: Model Training**

* Used sklearn.linear\_model.LinearRegression() to create a **multiple linear regression model**.
* Trained the model using the fit() method with **age, height, and weight** as input features and premium as the target variable.

**Step 4: Prediction**

* Used reg.predict([[27, 165.56, 60]]) to predict the **insurance premium for a 27-year-old person, 165.56 cm tall, weighing 60 kg**.

**5. Results & Interpretation**

* The model successfully established a relationship between **age, height, weight, and premium**.
* The prediction provides an estimated premium based on these input features.
* **Handling missing values correctly** improved data consistency.

**6. Conclusion**

* **Multiple Linear Regression** is effective for predicting numerical values when multiple factors influence the outcome.
* More features like **health conditions, income, or location** could further improve accuracy.
* Proper handling of missing data ensures better model performance.

**7. Future Enhancements**

* Evaluate model performance using **Mean Absolute Error (MAE), R-squared score**.
* Apply **feature scaling** for better accuracy.
* Consider **Polynomial Regression** if relationships are non-linear.

**Project Completed by:** M.Iqbal  
**Tools Used:** Python, Pandas, Scikit-Learn