**Executive Summary**

Diabetes is a medical condition that is caused due to insufficient production and secretion of insulin from the pancreas in case of Type-1 diabetes and defective response of insulin Type-2 diabetes. Diabetes is one of the most prevalent medical conditions in people today

Hospital readmission for diabetic patients is a major concern in the United States. Over $250 million dollars was spent on treatment of readmitted diabetic inpatients in 2011 alone. Diabetes is chronic and does not have any specific cure.

**Objective:-**

Hospital readmission rates for certain conditions are now considered an indicator of hospital quality, and also affect the cost of care adversely. Hospital readmissions of diabetic patients are expensive as hospitals face penalties if their readmission rate is higher than expected and reflects the inadequacies in health care system. For these reasons, it is important for the hospitals to improve focus on reducing readmission rates. Identify the key factors that influence readmission for diabetes and to predict the probability of patient readmission.

**Approach:-**

The dataset chosen is that available on the UCI website which contains the patient data for the past 10 years for 130 hospitals. The code has been written in Python using different libraries like scikit-learn, seaborn, matplotlib etc. Different machine learning techniques for classification and regression like Logistic regression, Random forest etc have been used to achieve the objective.

**Keywords:**

Machine Learning, Python, scikit-learn, EDA, Healthcare