



Introduction to Statistical Data Analysis Project

Project Deliverables

You will be required to provide the following deliverables.

- A python notebook with your solution.

Instructions

Background Information

The management of a certain NGO Hospital would like to have a product developed that would make predictions on whether a person has diabetes or not.

The data were collected and made available by “National Institute of Diabetes and Digestive and Kidney Diseases” as part of the Pima Indians Diabetes Database. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here belong to the Pima Indian heritage (a subgroup of Native Americans) and are females of ages 21 and above.

Problem Statement

Your task for this project will be to perform univariate and bivariate analysis in an effort to prepare your data for modeling in the later stages.

You can use the following guiding template [\[Link\]](#).

Dataset

Datasets for this project can be found here [\[https://bit.ly/3eoAbDS\]](https://bit.ly/3eoAbDS).

Project Source: [\[Link\]](#)